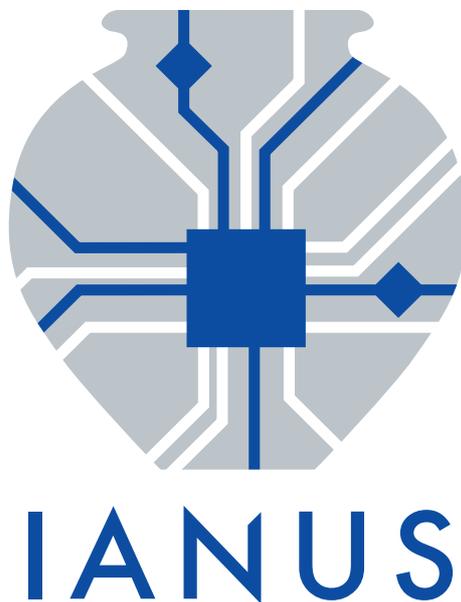


Fach- und Organisationskonzept zum Betrieb eines nationalen Forschungsdatenzentrums für die Archäologien und Altertumswissenschaften in Deutschland

Version 0.95 – 01/2016



Deckblatt

Autoren: Felix Schäfer, Maurice Heinrich
Anne Sieverling, Martina Trognitz, Zoe Schubert

Datum Erstellung: Montag, 03. Juni 2013

Letzte Bearbeitung: Montag, 04. Januar 2016

Version 0.1	2013-06-13	Initiales Brainstorming während des Besuchs beim ADS	Felix Schäfer
Version 0.11	2013-08-12	Erste Einarbeitung Ideen / Vorschläge aus Expertisen Schneider / Schwiengelshohn	Felix Schäfer
Version 0.12	2013-09-05	Diverse Ergänzungen Rechtliches, Arbeitsschritte, Qualitätssicherung	Felix Schäfer
Version 0.13	2013-09-13	Diverse Ergänzungen Zielgruppen, Technische Partner; Nummerierung	Felix Schäfer
Version 0.14	2013-09-17	Rechtschreibung, Erstellung 4.1.3, 4.2, 5, 5.2, Ergänzungen Glossar	Maurice Heinrich
Version 0.15	2013-10-18	Umstrukturierung Kapitel 3, 4 & 5; Überarbeitung Kapitel 4.1 & 4.2.1	Felix Schäfer
Version 0.16	2013-11-05	Überarbeitung Kapitel 3.1, 4.2.1 & 4.2.2	Felix Schäfer
Version 0.17	2013-11-12	Korrekturen, Inhaltsverzeichnis, Erstellung & Strukturierung Kapitel 10–10.4	Maurice Heinrich
Version 0.18	2013-11-20	Kapitel 4.1 und 4.2 finalisiert	Felix Schäfer
Version 0.19	2013-11-22	Überarbeitung 10–10.6, Erstellung 6, Korrekturen, Finalisierung 4.1–4.2, Überarbeitung 10.1–10.2	Maurice Heinrich / Felix Schäfer
Version 0.20	2013-11-29	Umstrukturierung Kapitelreihenfolge, Finalisierung Kapitel 4, 8, 9	Felix Schäfer / Maurice Heinrich
Version 0.21	2013-12-03	Überarbeitung Kapitel 6	Anne Sieverling
Version 0.22	2014-01-13	Einarbeitung Feedback Dally	Felix Schäfer
Version 0.23	2014-01-16	Kapitel 7.4 eingefügt	Martina Trognitz
Version 0.24	2014-01-20	Zusammenfassung (Kapitel 2) überarbeitet	Felix Schäfer
Version 0.40	2014-02-26	Überarbeitung für DFG-Folgeantrag	Felix Schäfer / Maurice Heinrich
Version 0.41	2014-03-12	Einarbeitung Workflows / Arbeitsschritte	Felix Schäfer
Version 0.42	2014-04-11	Erstellung Google-Dok für Bearbeitung im Rahmen Expertise 5.2	Felix Schäfer
Version 0.43	2014-09-16	Überarbeitung Kapitel 5.1	Maurice Heinrich /

			Felix Schäfer
Version 0.44	2014-10-27	Überarbeitung Kapitel 5.3 Neues Kapitel 3.3 (Verbände etc.)	Anne Sieverling / Felix Schäfer
Version 0.45	2014-12-16	Umzug von GoogleDoc zu OwnCloud, Überarbeitung Kapitel 9	Anne Sieverling / Maurice Heinrich
Version 0.46	2015-01-23	Einfügung Kommentare Huth & Thaller; Texte von Fr. Schubert; Kapitel 5 neu	Felix Schäfer
Version 0.47	2015-02-11	Finalisierung Kapitel 6.2 und 7.1–7.3	Felix Schäfer
Version 0.48	2015-03-10	Finalisierung Kapitel 10	Felix Schäfer
Version 0.49	2015-04-24	Überarbeitung Kapitel 5 und 13	Manfred Thaller / Zoe Schubert / Felix Schäfer
Version 0.50	2015-06-16	Finalisierung Kapitel 5, Generalüberarbeitung Kapitel 13	Felix Schäfer
Version 0.51	2015-09-01	Redaktionelle Prüfung	Esther Schneidenbach
Version 0.55	2015-09-16	Einarbeitung redaktionelle Fehler	Felix Schäfer
Version 0.60	2015-10-03	Finalisierung Kapitel 12 Metadaten	Felix Schäfer
Version 0.70	2015-10-16	Finalisierung Kapitel 14 Arbeitsschritte	Felix Schäfer
Version 0.80	2015-10-16	Überarbeitung Kapitel 8 Geschäftsmodell und Kapitel 9 Personaltabelleau	Felix Schäfer
Version 0.82	2015-12-14	Finalisierung Kapitel 13	Felix Schäfer
Version 0.83	2015-12-19	Finale Korrekturen Konzept Teil 1	Maurice Heinrich
Version 0.86	2015-12-27	Finale Korrekturen Konzept Teil 2	Maurice Heinrich
Version 0.88	2015-12-29	Finalisierung Kapitel 14	Felix Schäfer
Version 0.89	2015-12-31	Finalisierung Kapitel 11	Felix Schäfer
Version 0.90	2016-01-03	Korrekturen Kapitel 11 & 14	Maurice Heinrich
Version 0.95	2016-01-04	Zusammenführung Teil 1 und 2, Klärung letzte offene Fragen/Kommentare	Felix Schäfer
Version 0.96		Redaktionelle Endprüfung	Anne Sieverling
Version 0.97		Vervollständigung Glossar	Anne Sieverling / Felix Schäfer / Maurice Heinrich
Version 0.98		Einarbeitung Rückmeldungen Förtsch, Thaller, Huth	Maurice Heinrich / Felix Schäfer
Version 1.0		Finalisierung Konzept zur Veröffentlichung	

Inhaltsverzeichnis

1	Einleitung	9
1.1	Ausgangslage: Moderne Altertumswissenschaften	9
1.2	Maßnahmen zum Aufbau eines Forschungsdatenzentrums	12
1.3	Entstehung und Verwendung des Konzeptes	13
1.4	Akteure an dem Konzept	14
1.4.1	Autoren	14
1.4.2	Qualitätssicherung	14
2	Zusammenfassung	16
2.1	Kernaussagen über IANUS	16
2.2	Mission Statement / Leitbild von IANUS	21
3	Zielgruppen und Stakeholder	23
3.1	Fachrichtungen	23
3.2	Institutionen	24
3.2.1	Außeruniversitäre Forschungseinrichtungen	24
3.2.2	Universitäten und Fachhochschulen	26
3.2.3	Museen und Sammlungen	27
3.2.4	Denkmalfachbehörden	29
3.2.5	Kommerzielle Dienstleister	32
3.3	Fachverbände, Vereine, Arbeitsgemeinschaften	34
3.4	Nutzergruppen	35
3.4.1	Datenproduzenten, Dateneigentümer, Datengeber	35
3.4.2	Datennachnutzer	37
3.4.3	Sonstige Nutzer	37
4	Daten	39
4.1	Zum Begriff Forschungsdaten	39
4.2	Charakteristika archäologisch-altertumswissenschaftlicher Daten	39
4.2.1	Fachlich-methodische Heterogenität (Datengenese)	39
4.2.2	Technische Heterogenität (Datenformate)	40
4.3	Daten im Kontext von IANUS	40
4.3.1	Verwaiste Daten (sog. Altdaten)	40
4.3.2	Aktuelle Projektdaten	41
4.3.3	Kontinuierliche Daten	41
4.3.4	Kriterien	41
5	Anforderungen an ein OAIS-konformes Archiv	43

5.1	Datenmodell und Informationsobjekte	44
5.2	Funktionsmodell	46
6	Zentrale Aufgaben und Dienstleistungen	49
6.1	Kernbetrieb	50
6.1.1	Technische Dienstleistungen	51
6.1.2	Nicht-technische Dienstleistungen	59
6.2	Erweiterter Betrieb	63
6.2.1	Technische Dienstleistungen	63
6.2.2	Nicht-technische Dienstleistungen	64
6.3	Öffentlichkeitsarbeit und Community Building	66
7	Technische Architektur	70
7.1	Drei-Schichten-Architektur	70
7.1.1	Zugriffsschicht	71
7.1.2	Managementschicht	71
7.1.3	Speicherschicht (Bitstream Preservation)	71
7.2	Technische Partner	71
7.2.1	Technische Partner (Rechenzentren) zur Bitstream-Preservation	71
7.2.2	Technische Partner für Exit Strategie	72
7.3	Schnittstellen	73
7.3.1	Datengeber und IANUS	73
7.3.2	Datennachnutzer und IANUS	73
7.3.3	APIs für institutionelle Webseiten und Fachsysteme	74
7.3.4	Einbindung von Referenzsysteme und Normdaten	74
7.4	Persistente Identifier	74
8	Geschäftsmodell und Rechtsform	77
8.1	Gesellschaftsform	77
8.2	Gremien	78
8.2.1	Aufsichtsrat	79
8.2.2	Wissenschaftlicher Beirat	79
8.2.3	Nutzer-Beirat	80
8.3	Finanzierung	81
8.3.1	Finanzierungsmodelle	81
8.3.2	Säule – Basisfinanzierung	83
8.3.3	Säule – Gebühren	85
8.3.4	Säule – Drittmittelprojekte	87
8.4	Kostenmodell	87
8.4.1	Kostenfaktoren	89

8.4.2	Gesamtkosten	90
9	Personaltableau.....	92
9.1	Personalanforderungen.....	92
9.1.1	Leitung und Organisation	92
9.1.2	Datenkuratierung und fachspezifische Dienstleistungen	93
9.1.3	Technische Aufgaben	93
9.1.4	Allgemeine Anforderungen.....	94
9.2	Personalkosten	95
9.2.1	Personal Basisfinanzierung	95
9.2.2	Personal Gebühren	95
9.2.3	Personal Drittmittel.....	96
10	Datenschutz, Lizenzen, Zugangsbeschränkungen	97
10.1	Beziehung Datengeber – IANUS.....	97
10.1.1	Rechtliche Rahmenbedingungen	97
10.1.2	Datenübergabevertrag.....	99
10.2	Beziehung IANUS - Datennachnutzer.....	102
10.2.1	Rechtliche Rahmenbedingungen	102
10.2.2	Nutzungsvereinbarungen	103
10.3	Zugangsbeschränkungen für Daten.....	104
10.3.1	Freier Zugriff („Open Access“)	106
10.3.2	Gruppenspezifischer Zugriff („Group Access“).	107
10.3.3	Personenspezifischer Zugriff („Individual Access“)	107
10.3.4	Datengeberspezifischer Zugriff („Embargo“)	107
11	Qualitätssicherung und IT-Sicherheit.....	108
11.1	Standards und Zertifizierung	108
11.2	Integrität, Authentizität, Vertraulichkeit.....	109
11.3	Zugriffssicherung	110
11.4	Notfallplan und Exit-Strategie	111
11.5	Kontrolle der Fachcommunity.....	112
12	Metadaten.....	114
12.1	Vorgaben nach dem OAIS-Referenzmodell	115
12.2	Kerninformationen für die Verwaltung von Archivdaten.....	116
12.2.1	Verpackungsinformationen	116
12.2.2	Erhaltungsmetadaten.....	117
12.2.3	Erschließungsinformationen	119
12.2.4	Metadaten für Projekte und Datensammlungen	120
12.2.5	Metadaten für Einzeldateien	121

12.3	Kontrollierte Vokabulare	122
12.4	Nachweis und Bereitstellung von Archivdaten	123
12.5	Metadaten für den Datenaustausch und Webservices	124
13	Strukturierung und Spezifikation der Datenpakete	126
13.1	Transferpaket (TP)	126
13.2	Submission Information Package (SIP)	127
13.2.1	Grundstruktur.....	127
13.2.2	<i>data</i> -Verzeichnis.....	128
13.2.3	<i>metadata</i> -Verzeichnis.....	129
13.3	Archival Information Package (AIP)	130
13.3.1	Allgemeine Anforderungen	130
13.3.2	Container-Format	132
13.3.3	Grundstruktur.....	132
13.3.4	<i>data</i> -Verzeichnis.....	133
13.3.5	<i>metadata</i> -Verzeichnis.....	135
13.4	Dissemination Information Package (DIP)	135
13.4.1	Allgemeine Anforderungen	135
13.4.2	Präsentation und Access	137
14	Arbeitsschritte der digitalen Langzeitarchivierung	139
14.1	Vorgaben	140
14.1.1	Akzeptierte und präferierte Datenformate	140
14.1.2	Notwendige Metadaten und Beschreibungen	141
14.1.3	Weitere Vorgaben.....	141
14.2	Vorbereitung der Daten und Datenübertragung	141
14.2.1	Vorbereitung der Daten	141
14.2.2	Anmeldung im Online-System.....	142
14.2.3	Datenübertragung	143
14.2.4	Tabellarische Zusammenfassung.....	144
14.3	Verarbeitung TP zu SIP (Pre-Ingest)	144
14.3.1	Initiale Validierung TP.....	145
14.3.2	Verbesserung TP.....	146
14.3.3	Finalisierung SIP	147
14.3.4	Tabellarische Zusammenfassung.....	147
14.4	Verarbeitung SIP zu AIP (Ingest)	149
14.4.1	Kuratierung SIP	149
14.4.2	Finalisierung AIP.....	151
14.4.3	Tabellarische Zusammenfassung.....	151

14.5	Verarbeitung AIP zu DIP (Access)	152
14.5.1	Erstellung DIP	153
14.5.2	Tabellarische Zusammenfassung.....	154
15	Glossar - Definition relevanter Begriffe	156
16	Literaturverzeichnis	159

Hinweis:

Bei allen Bezeichnungen, die auf Personen bezogen sind, meint die gewählte Formulierung beide Geschlechter, auch wenn aus Gründen der leichteren Lesbarkeit die männliche Form gewählt wurde.

1 Einleitung

1.1 Ausgangslage: Moderne Altertumswissenschaften

Seit der 2. Hälfte des 19. Jahrhunderts haben die Archäologien und die Altertumswissenschaften einen komplexen Prozess der Ausdifferenzierung erfahren, der sich in der Verankerung der Forschung in spezifischen Institutionen wie den Universitäten, dem Deutschen Archäologischen Institut, den Akademien, den Museen oder den Landesdenkmalämtern artikuliert hat. Befördert durch Großgrabungen, die u. a. in Kleinasien, Griechenland und dem Vorderen Orient im letzten Viertel des 19. Jahrhunderts einsetzten, sind in rascher Folge neue Arbeitsfelder entstanden, die wiederum als universitäre Disziplinen Gestalt angenommen haben.

Die altertumswissenschaftliche Forschung umfasst heute ein breites Spektrum verschiedener Archäologien, Philologien und Geschichtswissenschaften, die jeweils räumlich und / oder zeitlich unterschiedliche Kulturen zum Gegenstand haben. Je nach Quellengattung werden sie durch stärker spezialisierte Disziplinen ergänzt, z. B. durch die Bauforschung, die Dendrochronologie, die Epigraphik oder die Numismatik. Schließlich spielen auch Fächer und Methoden eine Rolle, die eigentlich in anderen Fachdomänen beheimatet sind, aber dennoch wichtige Informationen zu historischen Phänomenen beitragen können; hierzu zählen etwa die Anthropologie, Archäobiologie und Archäozoologie, die Geologie und Geographie sowie naturwissenschaftliche und medizinische Untersuchungen.

Die langfristige Sicherung und Verbreitung von Forschungsergebnisse erfolgte zunächst in analoger Form mit spezifischen Ausprägungen der Wissensaufbereitung z. B. in Bibliotheken, Grabungsarchiven, Corpora, Gelehrtennachlässen, Phototheken, gedruckten Publikationen, Gipsabgüssen und Abklatschen oder Sammlungen in Museen. Heute übernehmen diese Aufgabe in einem rasant wachsendem Maße vor allem digitale Daten. Je nach Projekten und Fragestellungen werden entweder '(Primär-)Daten' generiert – z. B. bei Feldforschungen, Untersuchungen mit Originalobjekten oder Texteditionen – oder durch Auswertungen, Analysen und Berechnungen neue '(Sekundär-)Daten' mit Erkenntnissen und Interpretationen gewonnen, die dann in der Regel publiziert werden.¹

Beide Prozesse werden zunehmend durch elektronische Systeme, die unmittelbar digitale Daten erzeugen, verarbeiten, visualisieren und bereitstellen, geprägt. In immer mehr Fällen bilden ausschließlich digital vorliegende Informationen ('digital born') neben schriftlichen Dokumentationen und materiellen Hinterlassenschaften sogar die einzige Grundlage für moderne Forschungsfragen. Die konkreten Einsatzgebiete von Computern in den Altertumswissenschaften sind dabei vielfältig: Datenbanken dienen der Strukturierung und Verwaltung von Informationen zu Fundstellen und Objekten, geographische Informationssysteme helfen bei der Analyse und Darstellung von raumbezogenen Angaben, elektronische Texte werden automatisiert nach sprachlich-syntaktischen Charakteristika durchsucht, Statistikprogramme werten quantitative Beobachtungen aus, 2D- und 3D-Rekonstruktionen helfen wissenschaftliche Ergebnisse zu visualisieren und zu überprüfen, digitale Fotos, Zeichnungen und Messungen dokumentieren vergängliche Kontexte. Je nach Sichtweise verdrängen oder ergänzen diese Verfahren immer

¹ Zu den Unzulänglichkeiten der Begriffe 'Primär- / Rohdaten' und 'Sekundärdaten' siehe J. Klump, Digitale Forschungsdaten, in: nestor. Enzyklopädie (2010) Kap.17:104 f.

stärker die bislang etablierten Praktiken und deren in der Vergangenheit zumeist analogen Resultate.

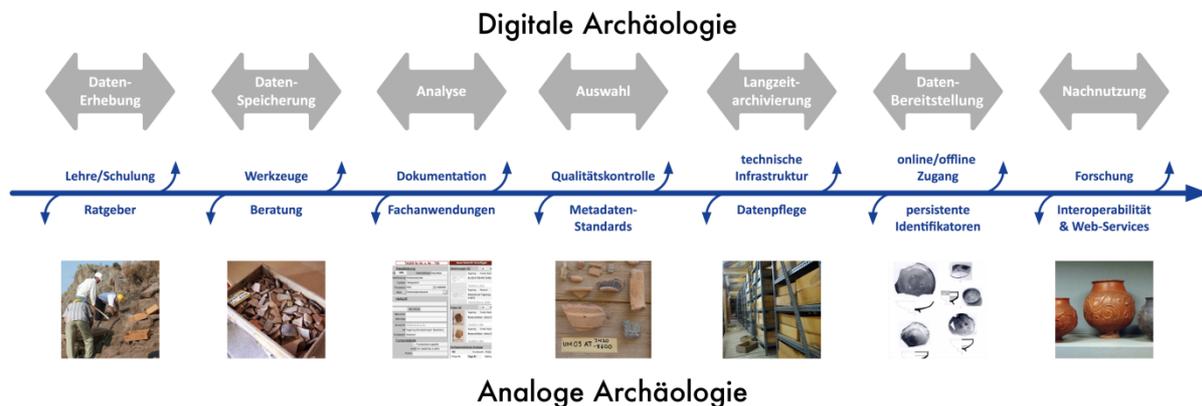


Abb. 1 Idealierte Gegenüberstellung analoger und digitaler Arbeitsschritte in der Archäologie

In solchen Systemen kommt ein generelles Interesse der modernen Forschung zum Ausdruck, das die Untersuchung einzelner Materialgruppen nicht ausschließt, aber den Kontext kultureller Phänomene als vordringliche Grundlage wissenschaftlicher Erkenntnis begreift. Die Wertschöpfung liegt vor allem in der Zusammenschau sinnvoller und logischer Einheiten, die durch automatisierte Verfahren extrahiert und miteinander in Bezug gesetzt bzw. verknüpft werden. Diese Kontexte sind wiederum zunehmend eingebunden in die Betrachtung größerer Zusammenhänge: anspruchsvolle Konzepte (z. B. Raum, Erinnerung und Gedächtnis, Landschaft, Macht, Innovationen), anthropologische Fragestellungen zur Geschichte der Menschheit insgesamt oder auf Regionen bzw. Kulturen bezogene Forschungsschwerpunkte (z. B. Südostasien, Afrika, Pazifik, Germanen, Kelten). Dadurch ergeben sich Schnittmengen u. a. zur Ethnologie, zu den Sozialwissenschaften, den Geschichtswissenschaften, aber auch zu den Naturwissenschaften.

Wie in allen wissenschaftlichen Disziplinen besitzen auch in den Altertumswissenschaften die den Forschungen zugrunde liegenden digitalen Daten einen Lebenszyklus, in dessen Verlauf sich ihre Natur ständig ändert: Rohdaten werden generiert und ergänzt; vorläufige, dynamische Daten werden verändert, prozessiert und analysiert; statische, finalisierte Daten werden abgelegt und veröffentlicht und bilden (alleine oder angereichert mit weiteren Daten) selbst wieder die Basis für neue Erkenntnisprozesse. Jede Phase stellt dabei spezifische Anforderungen an individuelle oder kollektive Akteure, divergierende Arbeitsabläufe, unterschiedliche Hard- und Software-Ressourcen sowie rechtliche wie finanzielle Rahmenbedingungen. Der Übergang von einer Phase in eine andere ist dabei fließend und verläuft keineswegs immer zyklisch, wie der abgebildete idealisierte und stark vereinfachte Kreislauf suggeriert (Abb. 2).² Die Realität in den Altertumswissenschaften zeigt, dass die Übergänge zwischen den einzelnen Phasen oft mit Informationsverlusten und

² Abb. 2 beruht auf dem Modell, das vom UK Data Archive publiziert wurde: <http://www.data-archive.ac.uk/create-manage/life-cycle> [Zugriff am 16.11.2014]. Ein alternative Modell, das stärker den Aspekt der Datenkuratierung berücksichtigt, wurde vom Digital Curation Centre erstellt: <http://www.dcc.ac.uk/resources/curation-lifecycle-model> [Zugriff am 16.11.2014]. Ausführlich siehe S. Rümpel, Der Lebenszyklus von Forschungsdaten, in: Forschungsdatenmanagement (2011) S. 25–34.

Brüchen einhergehen und nur einige wenige Daten ihren Weg in eine Publikation finden.



Abb. 2 Vereinfachtes Modell des Data-Life-Cycles

Diese tiefgreifenden Veränderungen prägen nicht nur immer stärker den Arbeitsalltag von Wissenschaftlern und erfordern eine immer größere Expertise im Bereich des Forschungsdatenmanagements, sondern bringen auch eine Vielzahl neuer Möglichkeiten und Chancen mit sich. Gleichzeitig werfen sie eine Reihe von offenen Fragen auf und verursachen neue Probleme, welche die Forschung gleichermaßen in fachlich-methodischer wie institutionell-organisatorischer Hinsicht betreffen. Die informationstechnische Herausforderung für die Altertumswissenschaften angesichts ihrer Positionierung an einer Schnittstelle zwischen Natur- und Geisteswissenschaften besteht in der Entwicklung von disziplinenübergreifenden, langfristig verfügbaren Systemen und Infrastrukturen, welche die Zerstreung von Informationen nach analogen Ordnungskriterien (Bibliothek, Fotothek, Archiv, Grabungsakten, Objektsammlungen, Publikationen etc.) aufhebt. Im Speziellen sind zwei Herausforderungen zu konstatieren:

- Charakteristisch für die archäologischen Disziplinen ist, dass bei einigen Methoden, etwa im Verlauf von Ausgrabungen und Oberflächenbegehungen, historische Befundsituationen irreversibel zerstört werden und der Prozess der Datengenerierung nicht wiederholt werden

kann – einmal ausgegrabene Kontexte können nicht nochmals untersucht werden, zerstörte Bauwerke sind für die Zukunft verloren, undokumentierte Objekte dem wissenschaftlichen Erkenntnisprozess entzogen. Daraus ergeben sich besondere Anforderungen hinsichtlich Vollständigkeit, Nachvollziehbarkeit und langfristigem Zugriff an die Dokumentation solcher Maßnahmen, die zusammen mit den physischen Objekten einmalige, nicht-reproduzierbare Primärquellen mit zeitlich unbegrenzter Gültigkeit darstellen. So wie heute Archäologen noch mit Grabungsunterlagen des 19. Jahrhunderts arbeiten, werden auch künftige Generationen auf die digitalen Daten von heute zurückgreifen wollen und müssen. Das Problem der langfristigen Sicherung von Forschungsdaten betrifft somit eine zentrale Grundlage des Fachs.

- Die zweite Herausforderung besteht in der starken technischen und fachlich-methodischen Heterogenität der Daten und Quellen, die für die Erforschung vergangener Kulturen relevant sind. Für die Beantwortung komplexer Fragen müssen mehrere Disziplinen mit sehr unterschiedlichen Qualitäten und Quantitäten von analogen Quellen und digitalen Informationen zusammenarbeiten und die jeweiligen Inhalte miteinander verknüpft werden. Erst im Gesamtbestand dieser miteinander verknüpften Daten wird das volle Potential eines Forschungsprojekts sichtbar. Archäologen sind einerseits mit der Aufgabe konfrontiert, diese komplexen Arbeitsgebiete in geeigneten technischen Lösungen zur Datenintegration abzubilden. Andererseits wird die Interoperabilität von Daten durch die Vielfalt von Akteuren erschwert, die sich in Deutschland und – bei internationalen Projekten – im Ausland mit der Altertumskunde beschäftigen (Universitätsinstituten, Museen, Akademien, Denkmalfachbehörden der Länder und Kommunen, Forschungseinrichtungen, kommerziellen Firmen und ehrenamtlichen Laien) und jeweils eigene institutionelle, rechtliche und finanzielle Rahmenbedingungen aufweisen.

Beide Aspekte, die Heterogenität und die Singularität von Daten, sind nicht alleine für die Altertumswissenschaften spezifisch³ und sind keine neuen Phänomene des digitalen Zeitalters. Auch wenn in anderen Disziplinen und Institutionen bereits vereinzelt Lösungen zum Umgang mit digitalen Forschungsdaten existieren⁴, ist für die Altertumswissenschaften in Deutschland festzuhalten, dass bislang eine zentrale Einrichtung fehlt, welche die wichtigsten Fragen fachspezifisch adressiert sowie Daten für die gesamte Fachcommunity nachhaltig archiviert und langfristig bereitstellt. Als Gründe für diese Situation sind die hohe Komplexität der Primärdaten und die damit verbundenen vielschichtigen Metadaten zu vermuten, die beide ein Spezifikum der altertumswissenschaftlichen Forschung darstellen.

1.2 Maßnahmen zum Aufbau eines Forschungsdatenzentrums

Aus der Erkenntnis heraus, dass die skizzierten digitalen Arbeitsprozesse mit ihren Formaten, Metabeschreibungen, Interoperabilitätsfragen und Archivierungsproblemen mit den derzeitigen technischen, personellen und organisatorischen Infrastrukturen in den unterschiedlichen

³ z. B. ist die Nicht-Reproduzierbarkeit von Daten auch in der Klimaforschung ein Thema, die u. a. auf einmaligen Wettermessungen basieren oder in den Sozialwissenschaften, in denen viele Umfragen nur eine nicht wiederholbare Momentaufnahme wiedergeben.

⁴ Siehe hierzu verschiedene Beiträge in nestor. Bestandsaufnahme (2012).

altertumswissenschaftlichen Institutionen und Disziplinen nicht mehr beherrschbar sind und dass dadurch zentrale Qualitäten digitaler Daten wie die Austauschbarkeit, Verfügbarkeit und Nachnutzbarkeit zur Zeit kaum gesichert werden können, wurde 2008 durch die DFG eine Arbeitsgruppe mit repräsentativen Fachvertretern initiiert.

Im April 2010 wurde stellvertretend für diese Arbeitsgruppe durch Prof. Ortwin Dally (damals DAI Berlin, jetzt DAI Rom) und Prof. Frederike Fless (damals TOPOI Berlin, jetzt DAI Berlin) ein Antrag bei der DFG zur „Entwicklung eines Kompetenzzentrums für altertumswissenschaftliche Forschungsdaten“ eingereicht. Grundlage dafür waren die zuvor geführten Diskussionen zu Bedarf und Notwendigkeit eines nationalen Zentrums, das langfristig die Fragen und Probleme in den Altertumswissenschaften im Umgang mit digitalen Daten adressiert und maßgeblich zu deren Lösung beiträgt. Mit Hilfe des Antrags wurde die Förderung einer ersten, dreijährigen Projektphase angestrebt, in der die spezifischen Anforderungen und Rahmenbedingungen formuliert und ein Konzept für einen langfristigen Betrieb erarbeitet wurden.

Diese Ergebnisse bildeten die Basis für einen zweiten, erfolgreichen Antrag an die DFG, mit dem der Aufbau zu einer funktionierenden Organisationseinheit und die Überführung in einen Regelbetrieb gefördert werden. Erklärtes Ziel aller Beteiligten ist es, dass am Ende der sechsjährigen Projektlaufzeit ein nationales fachspezifisches Forschungsdatenzentrum existiert, welches für die Altertumswissenschaften in Deutschland die zentralen Aufgaben eines professionellen, OAIS-konformen Archives erfüllen und sukzessive um weitere Dienstleistungen ergänzt werden kann.⁵ Insgesamt soll auf diese Weise eine institutionen-, projekt- und fachübergreifende Infrastruktur für die nachhaltige Archivierung, Bereitstellung, Nachnutzung und Vernetzung von qualitätsvollen und interoperablen digitalen Inhalten geschaffen werden.⁶

Mit dem Vorhaben wird versucht, eine der zentralen Forderungen des Wissenschaftsrates zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020 umzusetzen.⁷

1.3 Entstehung und Verwendung des Konzeptes

Das vorliegende Dokument stellt das Konzept für den infrastrukturellen Kern des Forschungsdatenzentrums IANUS dar, das im Wesentlichen während der ersten dreijährigen Förderung (09.2011 – 09.2014) durch die DFG verfasst wurde. Das Dokument diente in der Version 0.4 als Anlage für einen Folgeantrag, der im Februar 2014 bei der DFG eingereicht und positive bewilligt wurde. In der aktuellen, finalen Version 1.0 beschreibt das Fach- und Organisationskonzept in einem idealtypischen Ansatz die grundlegenden Aufgaben, Ziele, Rahmenbedingungen und Komponenten von IANUS, das als eine nationale Infrastruktur für die fachspezifische Langzeitarchivierung von digitalen Forschungsdaten aufgebaut werden soll. Zum einen werden hierfür die fachlichen, technischen und administrativen Anforderungen beschrieben,

⁵ Siehe unten **Kapitel 6**.

⁶ Die beiden **Kapitel 1.1** und **1.2** sind auch publiziert als Teil des Aufsatzes O. DALLY – F. FLESS – R. FÖRTSCH – M. HEINRICH – F. SCHÄFER, IANUS. Die Konzeption eines nationalen Forschungsdatenzentrums für die Archäologie und die Altertumswissenschaften, in: S. WINGHARDT (HRSG.), Archäologie und Informationssysteme. Vom Umgang mit archäologischen Fachdaten in Denkmalpflege und Forschung, Arbeitshefte zur Denkmalpflege in Niedersachsen 42 (2013) S. 118–127.

⁷ Wissenschaftsrat (2359-12), insbes. Abschnitt BIII.4 (Langzeitarchivierung und Langzeitverfügbarkeit), S. 58–61.

zum anderen aber auch die Informationsobjekte, Prozesse und Arbeitsschritte ausführlich diskutiert.

Mit diesem Dokument werden vor allem vier Ziele verfolgt:

- Es bildet die Grundlage für die zweite Projektphase zur konkreten technischen, organisatorischen und personellen Umsetzung von IANUS. Dabei dient es sowohl als gemeinsame Diskussionsgrundlage für alle am Projekt beteiligten Personen als auch als Basis für weitere konkrete Entscheidungen, etwa zu den technischen Systemen, sinnvollen Arbeitsabläufen, notwendigen Strategien, zu berücksichtigenden Standards und notwendigen Dokumentationen.
- Neben den wichtigsten Zielen und Aufgaben, die bis zur Etablierung eines Regelbetriebes spätestens 2017 kurzfristig bis mittelfristig umzusetzen sind, werden auch langfristige Weiterentwicklungen und Ausbaustufen benannt. Das Konzept beschreibt das Forschungsdatenzentrum daher in seiner grundsätzlichen, langfristigen Organisationsstruktur und Arbeitsweise, die auch über die unmittelbare, bis 2017 geplante 2. Projektphase gültig bleiben.
- Den Entscheidungsträgern, Mittelgebern, Betreibern von Dateninfrastrukturen und interessierten Stakeholdern soll das Dokument einen umfassenden Überblick über IANUS vermitteln, um im Bedarfsfall auf dieser Basis eigene Planungen und Entwicklungen vorbereiten zu können.
- Schließlich soll das Konzept auch für eine Transparenz gegenüber der Fach-Community sorgen, indem es entweder als Ganzes oder in ausgewählten Teilen veröffentlicht wird.

Je nach der Gesamtentwicklung von IANUS einerseits, aber auch fachlicher wie technischer Veränderungen andererseits ist zu überlegen, in welchen Situationen das Konzept auch künftig auf seine Aktualität hin überprüft und fortgeschrieben werden sollte.

1.4 Akteure an dem Konzept

1.4.1 Autoren

- Felix Schäfer, Projektkoordination (Themen Zielgruppen, Dienstleistungen, Datenmodell, Metadaten, Workflows)
- Maurice Heinrich, Projektkoordination (Themen Geschäftsmodell, PR, Lizenzen, Dienstleistungen)
- Zoe Schubert, WissMA (Themen OAIS, Datenmodell, Workflows)
- Martina Trognitz, WissHK (Themen Testdaten, PID, IT-Empfehlungen)
- Anne Sieverling, WissHK (Themen Öffentlichkeitsarbeit, Rechtliche Aspekte)

1.4.2 Qualitätssicherung

Version 1.0 des Konzepts wurde im Rahmen eines Werkvertrages durch die folgenden Experten inhaltlich redigiert und kommentiert, auf Vollständigkeit und Konsistenz überprüft sowie hinsichtlich aktueller Entwicklungen im Bereich des Forschungsdatenmanagements und der Langzeitarchivierung ergänzt:

- Manfred Thaller, Professur für Historisch-kulturwissenschaftliche Informationsverarbeitung,

Universität zu Köln

- Karsten Huth, Sächsisches Staatsarchiv, Dresden
- Reinhard Förtsch, Wissenschaftlicher Direktor IT-Referat, DAI

Die redaktionelle Überarbeitung des Textes haben übernommen

- Anne Sieverling, WissHK
- Esther Schneidenbach, WissHK

2 Zusammenfassung

2.1 Kernaussagen über IANUS

Die inhaltliche Basis für das folgende Konzept bilden die Aktivitäten, die im Rahmen der ersten Förderphase von IANUS von September 2011 bis September 2014 durchgeführt wurden. Dies waren insbesondere:

- Gespräche mit Mitarbeitern und Vor-Ort-Begehungen von Infrastrukturen, die bereits digitale Forschungsdaten archivieren und bereitstellen
- fünf Expertisen mit Empfehlungen zu spezifischen Aspekten von externen Fachleuten
- regelmäßige Treffen von vier verschiedenen Arbeitsgruppen
- eine Online-Befragung der relevanten Zielgruppe (Stakeholder-Analyse)
- Vorträge und Gespräche im Rahmen von Konferenzen und Verbandstagungen
- Durchführung eines Workshops zu lizenzrechtlichen Fragen
- Durchführung von drei Testbeds zu technischen Aspekten
- Analyse von 19 Test-Datensammlungen aus verschiedenen Institutionen und Projekten
- Einbeziehung maßgeblicher Veröffentlichungen
- Berücksichtigung der Hinweise der in **Kapitel 1.4.2.** genannten Personen

Aus der Summe der Ergebnisse und Erkenntnisse ergeben sich folgende Kernaussagen über IANUS, die in den Kapiteln des Konzeptes ausführlicher beschrieben werden:

1. IANUS ist ein **gemeinsames Projekt** der altertumswissenschaftlichen Fachgemeinschaft in Deutschland, das von nationalen und internationalen Akteuren getragen und geleitet wird. Diese werden in die Entwicklung von IANUS eingebunden durch die Mitarbeit in mehreren Arbeitsgruppen und Gremien, die Mitarbeit bei den IT-Empfehlungen, die Durchführung von Umfragen und gemeinsamen Veranstaltungen, eine konsequente Öffentlichkeitsarbeit sowie Präsentationen und Diskussionen auf Konferenzen und in Projekten.
2. Als **Zielgruppen** werden alle archäologischen und altertumswissenschaftlichen Disziplinen adressiert, unabhängig von ihrer institutionellen Verankerung (Museen, Denkmalfachbehörden, Universitäten/Hochschulen, außeruniversitäre Institutionen, kommerzielle Dienstleister, Akademien) und ihrer geographischen, zeitlichen, thematischen und methodischen Ausrichtung. Dies schließt explizit kulturhistorisch arbeitende Natur- und Geowissenschaften mit ein. Darüber hinaus richten sich die Angebote prinzipiell auch an eine interessierte Öffentlichkeit.
3. Während der Aufbauphase von IANUS übernimmt das **Deutsche Archäologische Institut** als zentrale Einrichtung der deutschen Altertumswissenschaften stellvertretend eine koordinierende Rolle und setzt sich für dessen nachhaltigen Erfolg ein. Damit IANUS auch langfristig einen zuverlässigen, vertrauenswürdigen organisatorischen Rahmen besitzt, **wird vom Wissenschaftsrat 2015 eine dauerhafte institutionelle Anbindung an das DAI empfohlen.** Dies ermöglicht zudem die Schaffung von Synergien mit dort bereits vorhandenen IT-Infrastrukturen (Hard- und Software-Ausstattung, Kooperationsverträge, iDAI.welt).

4. Das Community Building und die **Öffentlichkeitsarbeit** werden als zentrale Bestandteile der Arbeit von IANUS begriffen, da nur ein kontinuierlicher Kontakt und Austausch mit den Zielgruppen gewährleistet⁸, dass Nutzer von neuen Entwicklungen unmittelbar profitieren, Datenproduzenten und -eigentümer von der Notwendigkeit der Langzeitverfügbarkeit von Forschungsdaten überzeugt und die Angebote auf neue Anforderungen angepasst werden können. Zum Einsatz kommen dabei neben der Homepage verschiedene Arten der Kommunikation wie etwa Newsletter per Mail, Social-Media-Kanäle, gedruckte Infomaterialien, Blogs, RSS-Feeds, Pressemitteilungen, Online-Umfragen oder direkte Kontakte zu Forschern auf einschlägigen Veranstaltungen.
5. Zentrale Aufgabe wird die professionelle **langfristige Archivierung, Bereitstellung und Vernetzung** von Forschungsdaten sein, insbesondere von sog. unpublizierten Rohdaten. Dies beinhaltet die aktive und passive Akquise von Datensammlungen, deren archivalische Aufbereitung, qualitätsvolle Dokumentation, kontinuierliche Kuratierung, physikalisch Erhaltung und Online-Bereitstellung, um eine langfristige und dauerhafte Nachnutzung durch eine designierte Fachgemeinschaft zu ermöglichen. 'Langfristig' bedeutet dabei für die Bestandserhaltung digitaler Ressourcen nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können, wobei für konkrete Planungszwecke zunächst ein zeitlicher Rahmen von 30–50 Jahren zugrunde gelegt wird.⁹
6. Der Kernbetrieb von IANUS als institutionalisiertes Datenzentrum und als Langzeitarchiv für die Forschung wird als eine **infrastrukturelle Aufgabe** verstanden, die einerseits einen Dienstleistungscharakter besitzt (vergleichbar mit Bibliotheken und Archiven), andererseits aber auch wissenschaftliche Tätigkeiten und Profile umfasst.¹⁰ Auch wenn Forschungsdaten der primäre Gegenstand der Tätigkeit sein werden und alle Phasen des Datenlebenszyklus adressiert werden sollen, ist es nicht das Ziele von IANUS, diese in fachwissenschaftlicher Hinsicht zu analysieren, auszuwerten oder zu bewerten.
7. Die Konzepte und Vorgaben des ISO-Standard 14721:2003 für ein **Offenes Archiv-Informationssystem (OAIS)** sollen größtmögliche Anwendung finden. Ebenso werden die Kriterien für vertrauenswürdige digitale Archive (DIN 31644) des Nestor-Kompetenznetzwerkes¹¹, die Vorgaben für Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)¹² und die Anforderungen des Data Seal of Approval¹³ berücksichtigt, um nach Abschluss der Entwicklungs- und Aufbauphase eine erfolgreiche Zertifizierung des Archives anstreben zu können.

⁸ Ergebnisse der Begehung in IANUS 1 <http://www.ianus-fdz.de/projects/uebersicht/wiki/Vor-Ort-Begehungen>

⁹ Ute Schwens - Hans Liegmann, Langzeitarchivierung digitaler Ressourcen, in: Rainer Kuhlen - Thomas Seeger - Dietmar Strauch (Hrsg.), Grundlagen der praktischen Information und Dokumentation, 5. Ausgabe (München 2004) S. 567. IANUS. SCHWIEGELSHOHN (2013) S. 7. 11.

¹⁰ IANUS. Schwiegelshohn (2013) S. 30.

¹¹ <http://nbn-resolving.de/urn:nbn:de:0008-2008021802>

¹² http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

¹³ <http://datasealofapproval.org/>

8. Neben der Langzeitarchivierung von Forschungsdaten sollen mittel- bis langfristig **weitere Dienstleistungen** je nach Bedarf für die Fachgemeinschaft angeboten werden:
 - a. Ein zentral betriebener Online-Nachweiskatalog zu Fundstellen, Maßnahmen bzw. Aktivitäten, Forschungsprojekten, Grabungsdokumenten bzw. Archiven und archäologische Objekten in Deutschland.
 - b. Publikation von IT-Empfehlungen zum nachhaltigen Umgang mit digitalen Daten und zu den verschiedenen Phasen des Datenlebenszyklus
 - c. Schulungen, Summer Schools und Online-Materialien zum praktischen Umgang mit IT-Technologien in Forschung und Lehre sowie zu allen Aspekten des Forschungsdatenmanagements
 - d. Beratung von Wissenschaftlern und Institutionen bei Planung, Durchführung und Abschluss von Forschungsprojekten, insbesondere beim Datenmanagement und bei der Speicherung von Daten mittels Sync-&Share-Services
9. Die **Inanspruchnahme von Dienstleistungen** von IANUS, insbesondere die Langzeitarchivierung, erfolgt auf einer freiwilligen Entscheidung, es sei denn Institutionen oder Mittelgeber formulieren andere Vorgaben, auf die IANUS keinen Einfluss hat. Die Archivierung von Forschungsdaten erfolgt nicht-exklusiv, d.h. die Verwertungs- und Nutzungsrechte bleiben bei dem Dateneigentümer, der somit die Daten auch bei anderen Einrichtungen archivieren und bereitstellen kann.
10. Daten, die über IANUS archiviert und bereitgestellt werden, sollten einen konkreten **Bezug zu Deutschland** aufweisen, d.h. entweder unter Beteiligung deutscher Forscher und Institutionen erhoben worden sein, durch Mittelgeber in Deutschland mitfinanziert worden sein oder sich geographisch auf Deutschland beziehen. Ein essentieller Teil der Metadaten und der Dokumentation muss in deutscher Sprache vorliegen.
11. IANUS begreift sich als Teil eines nationalen und internationalen Netzwerkes zum **Austausch von Forschungsdaten**. Dazu wird einerseits ein enger Verbund und projektbezogene Kooperationen mit dem Archaeology Data Service (ADS) in UK, dem e-depot for nederlandse archaeologie (edna c/o DANS) in NL und weiteren thematisch relevanten Dateninfrastrukturen angestrebt. Andererseits will IANUS seine Daten so frei wie rechtlich möglich und auf Basis etablierter Standards über semantische Schnittstellen (u.a. OAI-PMH, CIDOC-CRM), Linked Open Data und Web-Services interoperabel zur Verfügung stellen. Dazu ist z.B. die Übermittlung von Metadaten an Portale wie Europeana, Deutsche Digitale Bibliothek und Propylaeum vorgesehen.
12. Um die archivierten Datensammlungen und Dateien nachhaltig zitieren und referenzieren zu können, erhalten diese sogenannte **persistente Identifikatoren (PID)**, die als DOIs (*Digital Object Identifier*) über *da|ra*¹⁴ registriert werden. Standardmäßig erhält jede Datensammlung einen einzigen DOI, während die Vergabe weiterer DOIs für einzelne Dateien oder kleinere Dateigruppen durch den Datengeber im Rahmen der Datenübergabe vorgeschlagen bzw. durch einen Datennachnutzer bei Bedarf beantragt

¹⁴ Registrierungsagentur für Sozial- und Wirtschaftsdaten, betrieben von GESIS. Leibniz-Institut für Sozialwissenschaften und dem Leibniz-Informationszentrum Wirtschaft ZBW: <http://www.da-ra.de>

werden kann. Für jedes Objekt in IANUS, das durch eine DOI referenziert werden kann, wird eine eigene Landing-Page bereitgestellt.

13. IANUS bekennt sich zu einem barrierefreien Austausch von Forschungsdaten und fühlt sich den Zielen des **Open Access** Gedankens verpflichtet. Auch wenn die Entscheidung über die Lizenzierung von Forschungsdaten, die über die Angebote von IANUS der Öffentlichkeit zur Verfügung gestellt werden, dem Dateneigenümer obliegt, setzt sich IANUS dafür ein, dass möglichst freie und offene Zugangs- und Nachnutzungsregelungen regelhaft Anwendung finden. Insofern ist der Zugriff auf und die Nutzung der bereitgestellten Daten kostenlos.
14. Für bestimmte Daten, insbesondere Geoinformationen zu gefährdeten Denkmälern und Angaben zu Personen, müssen **Zugriffsbeschränkungen** beachtet werden. Deren Einhaltung soll durch ein detailliertes Rechte-Rollen-Konzept sowie dessen Umsetzung durch technische Schutzmechanismen sichergestellt werden. Vier Stufen der Zugriffsbeschränkung sind vorgesehen und können von Dateneigentümern für ganze Datensammlungen oder einzelne Dateien festgelegt werden:
 - a. unbeschränkter Zugriff für alle Datennachnutzer
 - b. Zugriff nur für eine designierte Gruppe oder Individuen
 - c. Zugriff nur nach Anfrage bei und Genehmigung durch den Dateneigentümer
 - d. Zugang nur für Dateneigentümer als zeitlich befristetes Embargo

Die Metadaten sind dabei über alle Zugriffslevel hinweg immer offen, frei durchsuchbar und nachnutzbar, es sei denn übergeordnete Rechte zum Schutz von Personen oder Kulturgütern machen Einschränkungen notwendig.

15. Zur langfristigen **finanziellen Verstetigung** wird ein Drei-Säulen-Modell angestrebt.
 - a. Eine Basisfinanzierung, idealerweise unter Beteiligung des Bundes und der Länder, schafft die notwendige finanzielle Sicherheit, damit der institutionelle Kern von IANUS langfristig existieren und die grundlegenden Aufgaben (langfristige Verfügbarmachung von Forschungsdaten) dauerhaft ausführen kann.
 - b. Fallbezogene Gebühren, die Datenproduzenten einmalig bzw. Institutionen regelmäßig bei der Übergabe von Daten an das Archiv entrichten und als Unkosten bei (Dritt-)Mittelgebern geltend machen können¹⁵, bilden einen Beitrag für Ausgaben, die datenspezifisch für Personal und IT-Infrastruktur entstehen. Die Höhe der Gebühren richtet sich nach dem jeweiligen Speicherbedarf und dem Aufwand zur Kuratierung, der von der Anzahl, Heterogenität, Dokumentation und gewünschten Bereitstellung der Dateien abhängt.
 - c. Für die dynamische Weiterentwicklung von IANUS und die Durchführung gezielter Projekte, etwa die Aufarbeitung von Altdaten, die technische Aktualisierung von Systemen, die Verbesserung der Interoperabilität der Daten oder die Einführung neuer Dienstleistungen, müssen Drittmittel eingeworben werden.

16. IANUS soll als ein nationales Zentrum für die gesamten Altertumswissenschaften und

¹⁵ http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

Archäologen in Deutschland fungieren. Insofern wird für den Zusammenschluss der Fach-Community und die Institutionalisierung des infrastrukturellen Kerns eine **rechtliche Verstetigung** angestrebt, welche die Partizipation und Rechte der Stakeholder unter Berücksichtigung ihrer föderalen Strukturen eindeutig definiert und die besondere Rolle des DAI berücksichtigt. Die gemeinnützige GmbH erscheint derzeit als die am besten geeignete Form, da diese einerseits eine Steuerung durch die Mittelgeber und die Vertreter der Fach-Community erlaubt, andererseits aber auch eine Qualitätssicherung durch externe Experten (z. B. Mitglieder anderer Forschungsdatenzentren) in einem wissenschaftlichen Beirat ermöglicht. Des Weiteren erlaubt diese Rechtsform die Einnahme von aufwandbezogenen Gebühren.

17. Mit Personen, die Daten zur Archivierung und Bereitstellung an IANUS übergeben, wird ein **Datenübergabevertrag** geschlossen, der alle rechtlichen und finanziellen Fragen, insbesondere zu Urheberrecht, Haftung, Verwertung, Nutzung, Schutz sensibler Daten und die Höhe der Gebühr, für beide Seiten verbindlich regelt. Datennachnutzern, die digitale Ressourcen und Fachinformationen über IANUS beziehen bzw. downloaden, müssen vorher die Rechte und Pflichten gegenüber den Dateneigentümern (z. B. zur Zitation der Quelle) und IANUS zur Kenntnis nehmen und **Nutzungsbedingungen** zustimmen.
18. Da IANUS kein eigenes Rechenzentrum für den langfristigen physischen Erhalt der Datenobjekte, die sog. **Bitstream Preservation**, betreiben wird, erfolgt dies bei spezialisierten, vertraglich verbundenen Partnern, die entsprechendes Wissen und die notwendigen Hard- und Software-Systeme zur physischen Archivierung von Daten besitzen. Um eine erhöhte Ausfallsicherheit zu erzielen, wird eine redundante Datenhaltung mit mindestens zwei Rechenzentren angestrebt, von denen das Rechenzentrum der Universität Köln als ein technischer Partner bereits feststeht.
19. Für den Fall, dass IANUS nach einem Zeitraum des erfolgreichen Betriebs seine Tätigkeit einstellen muss (z. B. aufgrund einer fehlenden Basisfinanzierung) wird eine **Exit-Strategie** formuliert, damit die dann vorhandenen Archivdaten an eine dritte Institution übergeben und dort für die Zukunft archiviert werden können. Zur Umsetzung der Exit-Strategie werden regelmäßig entsprechende Finanzmittel zurückgelegt.
20. Für die Kuratierung von Datenbeständen, die IANUS zur Langzeitarchivierung anvertraut werden, werden verbindliche **Workflows** und Verantwortlichkeiten definiert. Diese legen die einzelnen Arbeitsschritte und Kriterien fest, um ein initiales *Transferpaket (TP)* in ein valides *Submission Information Package (SIP)* und ein *Archival Information Package (AIP)* zu transformieren, aus dem dann ein oder mehrere *Dissemination Information Package (DIP)* erzeugt werden. Sie werden durch eine extensive Dokumentation ergänzt, die zugunsten einer transparenten Vorgehensweise der Öffentlichkeit zur Verfügung gestellt wird. Hierzu gehören auch Vorgaben an Datengeber, welche Datenformate zur Übernahme in das Archiv von IANUS akzeptiert bzw. präferiert werden, und Festlegungen darüber, in welche Archivierungs- und Bereitstellungsformate diese konvertiert werden. Sollten bereits archivierte digitale Objekte technisch veralten, werden **Migrationspläne** entwickelt und durchgeführt.
21. Neben technischen Vorgaben werden auch Vorgaben bzgl. der Beschreibung und der Dokumentation der Daten formuliert. Diese von Datengebern eingeforderten **Metadaten** beziehen sich sowohl auf eine ganze Datensammlung als auch auf einzelne Dateien. Sie

dienen als Grundlage für Einträge im Nachweiskatalog von IANUS, damit Datennachnutzer in IANUS vorhandene Datenbestände durchsuchen und auffinden können, als inhaltlich-methodische Dokumentation, um eine wissenschaftliche Nachnutzung von Dateien zu ermöglichen und als technisch-praktische Informationen für die Verwaltung und Kuratierung der Daten. Die Angaben müssen in deutscher Sprache vorliegen. Kerninformationen werden jedoch auch in Englisch eingefordert oder übersetzt, um einen internationalen Nutzerkreis anzusprechen. Bei Daten, die zu einem in Ausland durchgeführten Projekt gehören, werden die Metadaten zusätzlich in der jeweiligen Landessprache eingefordert.

22. Zur Erfassung und Verwaltung von Datensammlungen, digitalen Objekten und zugehöriger Dokumentation wird ein „**Archiv-Management-System**“ benötigt. In diesem werden alle inhaltlichen, technischen und rechtlichen Metadaten gespeichert, bearbeitet und für unterschiedliche Zwecke aufbereitet. Ferner werden hier zur Nachvollziehbarkeit der Arbeitsschritte der Datenkuratierung alle Veränderungen an Dateien und zugehörigen Dokumenten protokolliert. Um eine Verbesserung der Qualität der Metadaten zu erzielen, werden online verfügbare Informationen und Webservices eingebunden, z. B. der Bibliothekskatalog iDAI.bibliography (ZENON)¹⁶, das Ortsregister iDAI.gazetter¹⁷ oder weitere Thesauri und Normdateien.
23. Das Archiv-Management-System von IANUS dient ferner als Backend für das **Online-Datenportal**, über das Datengeber neue Datensammlungen oder Beschreibungen hochladen und Datennachnutzer Dateien auffinden und herunterladen können.
24. Die Ablage der aufbereiteten Informationsobjekte (SIP, AIP, DIP) erfolgt auf einer für alle Datensammlungen **einheitlichen Verzeichnisstruktur**, in der auch die Metadaten als XML-Dateien gemäß der Standards METS, PREMIS und DC enthalten sind. Sofern Datenformate es technisch ermöglichen, werden die Metadaten zusätzlich direkt in den einzelnen Dateien gespeichert.
25. Ferner sollen folgende technische Komponenten für die zu entwickelnde **Arbeitsumgebung** von IANUS zum Einsatz kommen:
 - a. ein internes Ticket-System für die Kommunikation mit Nutzern (z. B. Jira)
 - b. ein öffentliches Wiki für die Dokumentation von Workflows (z. B. Confluence)
 - c. eine Archivierungssuite für die technische Formatkonvertierung und Dokumentation von Dateien
26. Alle für die Anforderungen von IANUS selbst entwickelten Softwaresysteme werden als **OpenSource** auf entsprechenden Plattformen (z. B. github, sourceforge) frei zur Verfügung gestellt, um eine Nachnutzung durch Dritte zu ermöglichen.

2.2 Mission Statement / Leitbild von IANUS

Ziel von IANUS ist es, auf die Möglichkeiten und Herausforderungen zu reagieren, die sich durch

¹⁶ <http://zenon.dainst.org/>

¹⁷ <http://gazetteer.dainst.org/>

die Arbeit mit digitalen Daten und Methoden ergeben und die zunehmend den Alltag in den Altertumswissenschaften prägen. Da in Deutschland bislang für die Altertumswissenschaften eine zentrale, disziplinspezifische Adresse fehlt, die Fragen und Antworten zum Forschungsdatenmanagement sowie zum Lebenszyklus von forschungsrelevanten Informationen bündelt und fachlich angepasste technische Lösungen umsetzt, wird eine Daten- und Serviceinfrastruktur aufgebaut. Diese soll – nach einer Konzept- und Aufbauphase – über die Grenzen von Bundesländern, Institutionen und Projektlaufzeiten hinweg im Regelbetrieb IT-Dienstleistungen anbieten.

Geplant ist, digitale Forschungsdaten aus der Archäologie, den Altertumswissenschaften und weiteren verwandten Nachbardisziplinen zu sammeln, zu beschreiben, zu katalogisieren, zu migrieren, zu archivieren und – soweit möglich – online frei verfügbar bereitzustellen. Dadurch soll eine Nachprüfbarkeit von Forschungsergebnissen im Sinne der guten wissenschaftlichen Praxis erleichtert, eine langfristige Nachnutzung von Forschungsdaten ermöglicht, dem drohenden Verlust von primären und sekundären Inhalten entgegengewirkt sowie der Austausch von Fachinhalten verbessert werden.

Darüber hinaus werden Forscher und Institutionen bei der Erstellung, Beschreibung, Verarbeitung, Erhaltung und Weitergabe von digitalen Daten unterstützt, u.a. durch die Formulierung und Publikation von Mindeststandards, Ratgebern und Best-Practice-Beispielen, Tools, Anleitungen und Lehrmaterialien. Sofern bereits nationale oder internationale fachlich relevante Archivsysteme und Datenspeicher existieren, wird IANUS eine kooperative Zusammenarbeit mit diesen anstreben und seine Angebote entsprechend abstimmen.¹⁸

Durch die langfristige Archivierung und die dauerhafte Bereitstellung von interoperablen digitalen Daten, die vergangene Kulturen beschreiben und dokumentieren, will IANUS zudem einen Beitrag zum Erhalt des historischen Erbes in Deutschland und weltweit leisten.

¹⁸ Veröffentlicht als Mission Statement unter http://www.ianus-fdz.de/attachments/download/470/2012-12_17_MissionStatement.pdf.

3 Zielgruppen und Stakeholder

Im Folgenden werden die Zielgruppen beschrieben, für die IANUS Dienstleistungen anbieten will und an deren Bedürfnissen und Anforderungen sich die Konzeption maßgeblich orientiert. Der langfristige Betrieb und Erfolg von IANUS als Forschungsdatenzentrum wird wesentlich von der Höhe der Akzeptanz in diesen unterschiedlichen Gruppen abhängen.

Die Zielgruppen lassen sich dabei sowohl nach fachlichen und institutionellen Gesichtspunkten gruppieren als auch als funktionale Nutzergruppen in Bezug auf IANUS beschreiben. Langfristig besteht der Anspruch, außer für Wissenschaftler und Fachleute aus den unten aufgeführten Disziplinen auch für die interessierte Öffentlichkeit Angebote zu entwickeln.

3.1 Fachrichtungen

IANUS wird Dienstleistungen für alle Fachrichtungen anbieten, die sich mit der Erforschung der Altertumskunde beschäftigen. Der fachliche Schwerpunkt liegt dabei auf den Kulturen des Vorderen Orients, des Mittelmeerraumes, Afrikas und Mitteleuropas. Geographisch weiter entfernt liegende Kulturen, wie etwa im eurasischen Raum, in Mittel- und Südamerika oder Ozeanien sollen dabei ebenfalls berücksichtigt werden, soweit ausreichende Kompetenzen vorhanden sind, um eine adäquate Datenkuratierung anbieten zu können. Die gleiche Einschränkung gilt für zeitlich ältere oder jüngere Epochen, wie die frühe Menschheitsgeschichte oder die Industriearchäologie.

Die Grenzen zu anderen Fachrichtungen und entsprechenden Einrichtungen sind dabei fließend, z. B. in zeitlicher Hinsicht zum Mittelalter und der Neuzeit oder in methodischer Hinsicht zur Sinologie, Afrikanistik oder Ethnologie. Ein besonderer Fokus liegt auf den archäologisch arbeitenden Fachdisziplinen, da in ihnen in großem Umfang neue und einzigartige Daten erzeugt werden, die aufgrund der mangelnden Reproduzierbarkeit in besonderem Maße gefährdet sind.

Als Disziplinen mit einem Fokus auf den materiellen Hinterlassenschaften sind zu nennen:

- Klassische Archäologie
- Provinzialrömische Archäologie
- Urgeschichte / Prähistorische Archäologie
- Frühgeschichte
- Mittelalterarchäologie
- Vorderasiatische Archäologie
- Orientalische Archäologie
- (Früh-)Christliche und byzantinische Archäologie
- Biblische Archäologie
- Islamische Archäologie
- Ägyptische Archäologie
- Altamerikanistik
- Afrikanische Archäologie
- Montanarchäologie

Auf die Analyse von Text- und Schriftzeugnissen spezialisiert sind die Sprach- und Literaturwissenschaften:

- Lateinische Philologie
- Griechische Philologie
- Altorientalistik / Assyriologie
- Byzantinistik
- Ägyptische Philologie
- Koptologie
- Hethitologie

Die Auswertung und Interpretation historischer Quellen erfolgt in den Geschichtswissenschaften:

- Alte Geschichte
- Mittelalterliche Geschichte
- Papyrologie
- Epigraphik
- Numismatik

In einigen anderen Fachdomänen haben sich spezialisierte Disziplinen entwickelt, die ebenfalls für die Erforschung vergangener Kulturen relevant sind:

- Geoarchäologie
- Archäometrie
- Dendrochronologie
- (Antike) Architekturgeschichte /
Bauforschung
- Anthropologie
- Paläobotanik
- Archäozoologie
- Paläogenetik
- Geophysik

3.2 Institutionen

Wie in vielen anderen Disziplinen auch, werden in den Altertumswissenschaften Forschungsdaten nicht nur unmittelbar an Universitäten oder außeruniversitären Forschungseinrichtungen erzeugt, sondern ebenso in verschiedenen privaten und staatlichen Einrichtungen. Aufgrund des öffentlichen Interesses an der Thematik und ihrer Bedeutung für den Erhalt des kulturellen Erbes werden Museen eingerichtet, die neben der Aufbewahrung und Präsentation von Objekten auch die Aufgabe haben, Inhalte der Archäologie und der Altertumswissenschaften der Bevölkerung zu vermitteln, z. B. im Rahmen von Ausstellungen, Publikationen oder Führungen. Daneben existieren Landesdenkmalämter und weitere Fachbehörden, die in einem festgelegten geographischen Raum die gesetzlichen Aufgaben der Denkmalpflege und des Kulturgüterschutzes wahrnehmen und dabei teilweise auch einen expliziten Forschungsauftrag besitzen. Je nach gesetzlichen Rahmenbedingungen können die Durchführung und Dokumentation von Ausgrabungen, geophysikalischen Prospektionen oder Materialanalysen an kommerzielle Firmen vergeben werden, die forschungsnahe Dienstleistungen erbringen und wiederum eine eigenständige Gruppe bilden.

Die jeweiligen Akteure besitzen dabei unterschiedliche Forschungs- und Verwaltungsaufträge, technische Infrastrukturen, Workflows und Standards für die Datenverarbeitung, finanzielle Möglichkeiten und rechtliche Vorgaben. Da nur wenige von ihnen bereits eigene Ansätze zum nachhaltigen Management von Forschungsdaten und zur Archivierung ihrer digitalen Daten entwickelt oder diese bereits erfolgreich umgesetzt haben, sind die meisten als potentielle Datenlieferanten für IANUS einzustufen und daher sollen sie im Folgenden kurz charakterisiert werden¹⁹.

3.2.1 Außeruniversitäre Forschungseinrichtungen

Zu den außeruniversitären Forschungseinrichtungen zählen Einrichtungen, die fast ausschließlich

¹⁹ Vgl. IANUS. SCHWIEGELSHOHN (2013) S. 11–13.

Forschung in einem speziellen Wissensgebiet betreiben und nur in reduziertem Umfang Lehr- oder Ausbildungsaufgaben wahrnehmen²⁰. Ihr wesentliches Alleinstellungsmerkmal ist die thematische Konzentration, so dass die Anzahl an fachspezifischen Wissenschaftlern, die an einer bestimmten Frage arbeiten, höher als z. B. an Universitäten ist (abgesehen von kooperativen Arbeitsformen wie Exzellenzinitiativen oder Sonderforschungsbereichen). Die Forschungsaktivitäten können dabei gleichermaßen im Inland wie auch im Ausland erfolgen. Die meisten außeruniversitären Institute sind in größere Forschungsorganisationen eingebunden, wie etwa die Leibniz-Gemeinschaft, die Max-Planck-Gesellschaft oder die Union der deutschen Akademien der Wissenschaften, und werden überwiegend vom Bund finanziert teilweise unter Beteiligung der jeweiligen Bundesländer.²¹

Die Konzentration auf einzelne Forschungsfragen hat zur Folge, dass einerseits der Bedarf an Forschungsdaten um ein vielfaches höher ist als an anderen fachspezifische Einrichtungen und dass andererseits deutlich mehr Forschungsleistung erbracht wird. Diese Anforderung führt in vielen dieser Einrichtungen dazu, dass bereits strukturierte Verfahren für das Datenmanagement eingesetzt werden, um die Menge an Forschungsdaten effizient verwalten zu können. Vor allem Einrichtungen, die Teil der o.g. größeren Organisationen sind, besitzen dabei auch eigene mehr oder minder ausgereifte und spezialisierte Strategien zur Archivierung und Veröffentlichung digitaler Daten.

Dies hängt auch damit zusammen, dass an außeruniversitären Forschungseinrichtungen Langzeitprojekte betrieben werden, die teilweise bereits vor über hundert Jahren begonnen wurden (z. B. die Edition von Inschriften, Wörterbücher historischer Sprachen oder die Ausgrabung antiker Stätten), so dass sich im Idealfall eine Prozesskultur mit internen Regeln und einer kontinuierlichen Datenverarbeitung etablieren konnte. Häufig sind solche Vorhaben durch eigene Finanzierungsmodelle langfristig abgesichert und werden durch feste Mitarbeiter über längere Zeiträume betreut. Ein anderer Teil der Forschung an diesen Einrichtungen unterliegt dagegen den klassischen Projektzyklen der kurz- bis mittelfristigen Drittmittelförderung mit wechselnden Fragestellungen, Personaltableau und Kooperationspartnern. Zu den außeruniversitären Forschungseinrichtungen mit einer besonderen Relevanz für die Archäologie und Altertumswissenschaften gehören:

- das Deutsches Archäologisches Institut (DAI) mit Hauptsitz in Berlin als nachgeordnete Bundesanstalt des Auswärtigen Amtes
- die wissenschaftlichen Akademien in München, Mainz, Heidelberg, Berlin, Göttingen und Düsseldorf
- das Max-Planck-Institut für Evolutionäre Anthropologie in Leipzig und das Max-Planck-Institut für Wissenschaftsgeschichte in Berlin
- das Zentrum für Baltische und Skandinavische Archäologie in Schleswig
- das Curt-Engelhorn-Zentrum Archäometrie gGmbH in Mannheim, gleichzeitig An-Institut der Universität-Heidelberg

²⁰ Nicht berücksichtigt in diesem Abschnitt werden Museen und Denkmalfachbehörden, die zwar auch zu den außeruniversitären Facheinrichtungen zu zählen sind, aber unten in eigenen Abschnitten charakterisiert werden.

²¹ Vgl. IANUS. SCHWIEGELSHOHN (2013) S. 8.

3.2.2 Universitäten und Fachhochschulen

In Deutschland gibt es knapp 170 Seminare und Institute, an denen Forschung und Lehre in den unter **Kapitel 3.1** genannten Fachrichtungen betrieben wird. Die dort tätigen Professoren, wissenschaftlichen Mitarbeiter, technischen Angestellten und Studierenden bilden damit eine zentrale, wenn nicht sogar die umfangreichste Gruppe, die sich hauptamtlich mit der systematischen Erforschung und Vermittlung vergangener Kulturen beschäftigt. Aufgrund ihrer absoluten Größe und der Doppelfunktion ihrer Mitglieder sowohl als Datenproduzenten als auch Datennachnutzer spielt sie eine besondere Rolle für IANUS.

Neben der Ausbildung und Qualifizierung von Studierenden und Promovenden als wesentliches Merkmal der Universitäten und Fachhochschulen, werden dort parallel unzählige Projekte durchgeführt, je nach Fachrichtung oftmals im Ausland. Meist handelt es sich hierbei um Aktivitäten einzelner Institute oder Seminare, die in Kooperation mit nationalen oder internationalen Partnern durchgeführt werden. Gleichzeitig gibt es mehrere Beispiele für die Zusammenarbeit thematisch verwandter Fachbereiche innerhalb einer einzigen oder zwischen mehreren Hochschulen, die der Etablierung institutioneller oder regionaler wissenschaftlicher Schwerpunkte dienen. Entweder dienen sie als inneruniversitärer Fächerverbund primär zur Verbesserung der Lehre und Forschung oder es werden dezidierte Forschungsthemen adressiert, deren Untersuchung primär durch Drittmittel finanziert und vor allem in Form von Exzellenzclustern, Graduiertenschulen/-kollegs, Sonderforschungsbereichen oder Schwerpunktprogrammen organisiert werden²².

Die Prozessketten und der Umgang mit Daten in Forschungsprojekten, die an Hochschulen angesiedelt sind, lassen sich etwa folgendermaßen zusammenfassen:

- Ein Institut beantragt alleine oder in Kooperation mit Partnern bei einem Drittmittelgeber ein Forschungsvorhaben²³. Bereits in diesem Stadium sind in der Regel Angaben zum geplanten Datenmanagement und zur langfristigen Sicherung der Ergebnisse zu machen²⁴. Im späteren Verlauf gibt es jedoch kaum Mechanismen, mit denen die Einhaltung der hier gemachten Angaben kontrolliert und ggf. sanktioniert werden können.
- Ein Drittmittelgeber genehmigt ein Forschungsvorhaben, häufig verbunden mit Auflagen und/oder Mittelkürzungen, ohne ein adäquates Datenmanagement vom Antragsteller einzufordern.
- Ein bewilligtes Forschungsprojekt wird eigenverantwortlich von dem beantragenden Institut und den Partnern durchgeführt. Dabei ist insgesamt zu beobachten, dass trotz eines zunehmenden Einsatzes digitaler Techniken in den meisten Projekten explizite IT-

²² Im Jahr 2013 wurden vom BMBF zwei Exzellenzcluster, drei Graduiertenschulen und ein internationales Kolleg sowie von der DFG vier Graduiertenkollegs, drei Schwerpunktprogramme und acht Sonderforschungsbereiche mit Bezügen zu den Altertumswissenschaften gefördert

²³ Der wichtigste Projektförderer in den Altertumswissenschaften ist die DFG. Daneben besitzen auch das BMBF, die Fritz-Thyssen-Stiftung, die Gerda-Henkel-Stiftung, die Volkswagenstiftung und die Akademien eine größere Relevanz als Drittmittelgeber.

²⁴ Z. B. verlangt die DFG in ihrem „Leitfaden für die Antragstellung“ (DFG-Vordruck 54.01 – 11/13) unter Punkt 2.4 explizite Angaben zum „Umgang mit den im Projekt erzielten Forschungsdaten“:
<http://www.dfg.de/antragstellung/forschungsdaten/>

Kompetenzen und Pläne zum Forschungsdatenmanagement fehlen oder nur unzureichend umgesetzt werden²⁵. Des Weiteren ist zu konstatieren, dass es in den allermeisten Disziplinen der Altertumswissenschaften keine verbreiteten Standards und fest etablierten Systeme gibt, so dass als Folge häufig individuelle, kurzfristige Projektlösungen entwickelt werden.

- Nach Abschluss eines Projektes und nach Veröffentlichung der Ergebnisse²⁶, bleibt die Frage nach dem weiteren Umgang mit den vorhandenen, unpublizierten digitalen Forschungsdaten meist unbeantwortet. Wenn überhaupt, werden sie auf Systemen des durchführenden Institutes gespeichert, so dass zumindest ein Zugriff unabhängig von einzelnen Projektmitarbeitern möglich ist, aber deren Existenz nur einem kleinen Kreis von Wissenschaftlern bekannt ist. Im günstigeren Fall werden sie an eine lokale technische Infrastruktur übergeben, d.h. in der Regel ein eigenes Rechenzentrum oder eine Bibliothek, wo zumindest eine Bitstream-Erhaltung der Daten professionell durchgeführt werden kann. Beide Arten von Einrichtungen sind in der Regel jedoch aufgrund ihres übergreifenden institutionellen Auftrages nicht in der Lage, eine fachspezifische Datenkuratierung zu leisten, die Qualität der inhaltlich Dokumentation zu überprüfen und die Informationen für disziplinspezifische Bedürfnisse anzubieten.

Sowohl bei den Einzelaktivitäten als auch bei den Verbundforschungen verhindert der grundsätzliche Projektcharakter meist die nachhaltige Pflege von Forschungsdaten und entwickelten Anwendungssystemen. Sobald ein Projekt beendet ist und die Förderung ausläuft, können die meisten Hochschulen die Forschungen mit den eigenen finanziellen Ressourcen nicht weiterführen. Üblicherweise führt ein Ende der Mittel auch zu einem Verlust von Forschungsdaten, da gerade junge Forscher nach dem Projektabschluss die Universität verlassen und damit unbeabsichtigt auch Kenntnisse über den Zugriff auf die Daten und deren Entstehungskontext verloren gehen, sofern diese nicht ausreichend dokumentiert wurden. Auch wenn von Drittmittelgebern verstärkt ein nachhaltiger Umgang mit Projektdaten und die Verstetigung von Ergebnissen eingefordert werden, besonders auch bei Verbund- und Schwerpunktförderung, wird dieses Ziel bislang in der Realität nur von vereinzelt Hochschulen erreicht.

Eine ähnliche Problematik ergibt sich auch bei Forschungsdaten, die im Rahmen von Qualifizierungsarbeiten einzelner Personen (also B.A.- oder M.A.-Abschlussarbeiten, Promotionen, Habilitationen) erzeugt werden. Zwar kann hier die Veröffentlichung der Ergebnisse ein Teil des Prüfungsverfahrens sein, doch gibt es nur in den seltensten Fällen Vorschriften zum weiteren Verbleib der zugrunde liegenden Rohdaten. Diese sind im Regelfall und in Abhängigkeit der persönlichen Einstellung und den technischen Möglichkeiten des Einzelnen ebenfalls für eine weitere Nachnutzung verloren, obwohl es sich oft um umfassende Materialvorlagen handelt.

3.2.3 Museen und Sammlungen

²⁵ Die Gründe für diese Situation sind unterschiedlich: entweder werden bereits bei der Antragstellung keine entsprechenden Ressourcen eingeplant; wenn sie mitbeantragt werden, werden sie gelegentlich vom Drittmittelgeber gestrichen und als Teil der Eigenleistung angesehen; oder es fehlt de facto an geeignetem Personal, das eine solche Aufgabe an der Schnittstelle zwischen einer Fachdisziplin und der Informatik übernehmen kann.

²⁶ Die häufigste Form der Publikation in den Altertumswissenschaften sind immer noch gedruckte Monographien, Artikel/Beiträge und Rezensionen. Elektronische Veröffentlichungen nehmen zwar in den letzten Jahren zu, spielen aber im wissenschaftlichen Diskurs bislang nur eine untergeordnete Rolle.

In Deutschland existieren etwa 100 Museen, die sich mit Objekten antiker Kulturen und mit archäologischen Befunden beschäftigen, d. h. diese sammeln, bewahren, erforschen, ausstellen und vermitteln. Die Größen, die Trägerschaften und die Museumstypen sind dabei genauso vielfältig und heterogen wie in der deutschen Museumslandschaft insgesamt²⁷:

- große, international bekannte Museen mit langen Sammlungstraditionen, z. B. die Staatlichen Museen in Berlin, die Antikensammlungen und Glyptothek in München, das Roemer- und Pelizaeus-Museum in Hildesheim, das Römisch-Germanische Zentralmuseum in Mainz, das Museum Weltkulturen in den Reiss-Engelhorn-Museen in Mannheim oder das Liebieghaus in Frankfurt am Main
- zentrale Landesmuseen, die oft eng mit der Kultur- und Bodendenkmalpflege verbunden sind und einen staatlichen Sammlungsauftrag besitzen, z. B. die Stiftung Schleswig-Holsteinische Landesmuseen Schloss Gottorf in Schleswig, das Landesmuseum für Vorgeschichte in Halle, das Römisch-Germanische-Museum in Köln, die Archäologische Staatssammlung in München, das Badische Landesmuseum in Karlsruhe oder die Rheinischen Landesmuseen in Bonn und Trier.
- Stiftungsmuseen, die hinsichtlich des Profils, der finanziellen Ausstattung und der rechtlichen Rahmenbedingungen eigenständiger sind, allerdings eher die Ausnahme darstellen, z. B. das Neanderthal Museum in Mettmann oder das Museum für Kunst und Gewerbe in Hamburg
- verschiedene mittelgroße Freiluftmuseen und Parks, die an besonderen historischen Orten mit einer überdurchschnittlich guten archäologischen Überlieferung Bodenbefunde freilegen, konservieren, rekonstruieren und erklären, z. B. das Pfahlbaumuseum in Unteruhldingen, die archäologischen Parks in Xanten und Kempten, das Museum und Park Kalkriese, die Slawenburg Raddusch, das Keltenmuseum Heuneburg oder das Römerkastell Saalburg in Bad Homburg v.d.H.
- diverse kommunale Museen, die sich unterhalb der Ebene der Landesmuseen um die Lokalgeschichte kümmern, sowie private Museen, die die individuellen Sammlungen einzelner Personen der Öffentlichkeit zugänglich machen

Hinzu kommen rund 80-100 Sammlungen und Museen an deutschen Universitäten²⁸, die als Lehr- und Forschungssammlungen angelegt wurden und neben originalen Objekten auch Gipsabgüsse, Abklatsche und Repliken enthalten können. In den meisten Fällen sind sie bereits wissenschaftlich erforscht und publiziert und da in der Regel kein Etat für Neuerwerbungen vorhanden ist, ändert sich der Bestand – abgesehen von gelegentlichen Schenkungen – heutzutage kaum noch.

Ebenso vielfältig wie die Landschaft der Museen stellt sich auch deren individuelles Forschungsinteresse dar, das sich im Vergleich zu den universitären und außeruniversitären Einrichtungen hinsichtlich Fragestellungen und Motivationen oftmals unterscheidet. Es kann sowohl die Folge von direkten Aufträgen sein als auch im Zusammenhang mit der Wissensvermittlung an die Öffentlichkeit stehen, z. B. als Vorbereitung einer Ausstellung oder als

²⁷ Vgl. die Angaben auf der Homepage des deutschen Museumsbundes:
http://www.museumbund.de/de/das_museum/

²⁸ <http://www.universitaetssammlungen.de/search/swp/Archäologie>

Grundlage für einen Wiederaufbau eines Gebäudes. Die Forschungsthemen konzentrieren sich dabei zumeist auf Fragen, die in unmittelbarem Zusammenhang mit eigenen Objekten und Befunden stehen.

Die allermeisten forschungsrelevanten Daten, die in Museen erhoben und verwaltet werden, beziehen sich unmittelbar auf Einzelobjekte und werden in Fachsystemen (überwiegend Datenbanksystemen) vorgehalten. Zu den Informationen, die in der Regel für ein einzelnes Artefakt erfasst werden, gehören eine Beschreibung und wissenschaftliche Einordnung, Angaben zur Geschichte, Provenienz, Restaurierung und zugehörigen Dokumenten, aber auch administrative und rechtliche Aspekte. Allein aufgrund der Vielzahl an wissenschaftlichen Objekten stellen Museen in Bezug auf IANUS unverzichtbare Eigentümer von Forschungsrohdaten dar – auch dann, wenn nur ein Teil dieser Daten bislang in digitaler Form vorliegt.

Dank einer weitgehenden Vernetzung und einem hohen Organisationsgrad der Museen, etwa im Deutschen Museumsbund oder dem Internationalen Museumsrat (ICOM), existieren mehrere Standards und Absprachen, die den Datenaustausch vereinfachen. So hat das aus diesen Institutionen heraus entwickelte und ISO-zertifizierte Referenzmodell CIDOC-CRM²⁹ auch weit über die Welt der Museen hinaus eine große Bedeutung für die Dokumentation des kulturellen Erbes erlangt. Ein konkretes Ergebnis dieser Bemühungen ist das vielfach genutzte Online-Portal digiCULT, in dem Museen aller Arten und unterschiedlicher Bundesländer ihre Bestände über einheitliche Schnittstellen online bereitstellen können³⁰. Auch in der Deutschen Digitalen Bibliothek³¹ als nationaler und in Europeana³² als europäischer digitaler Katalog zu Kultur und Wissen werden in zunehmenden Maße Beständen aus archäologischen Museen nachgewiesen.

Neben diesen positiven Entwicklungen ist aber gleichzeitig zu betonen, dass in den meisten Museen die zur Verfügung stehenden Ressourcen für IT-Belange oft sehr begrenzt sind und vor allem in kleineren und mittleren Häusern eine eigene, fachspezifische Unterstützung fehlt. Insofern ist zum gegenwärtigen Zeitpunkt wohl für die Mehrheit der Museen davon auszugehen, dass es noch keine Überlegungen oder gar praktische Lösungen gibt, die die Fragen der langfristigen Archivierung und Bereitstellung von digitalen Daten adressieren, sofern nicht lokale Staatsarchive diese Aufgabe übernehmen.

3.2.4 Denkmalfachbehörden

In Deutschland unterliegen alle bekannten und noch unbekanntes archäologischen (Boden-) Denkmäler dem Denkmalschutz. Dessen Ausgestaltung und Durchführung obliegt aufgrund der Kulturhoheit den einzelnen Bundesländern, die individuelle gesetzliche Regelungen erlassen haben. Dies hat zur Folge, dass die Aufgaben des Denkmalschutzes und der Denkmalpflege jeweils mit unterschiedlichen Akteuren, finanziellen Mitteln, rechtlichen Rahmenbedingungen und praktischen Ausprägungen realisiert werden. Gemeinsam ist allen Bundesländern, dass als oberste staatliche Fachbehörde die ‚Landesämter für Archäologie und Denkmalpflege‘ fungieren³³, denen

²⁹ <http://www.cidoc-crm.org>

³⁰ <http://www.digicult-verbund.de>

³¹ <https://www.deutsche-digitale-bibliothek.de/>

³² <http://www.europeana.eu/>

³³ Insgesamt 17, da in Nordrhein-Westfalen zwei existieren (Westfalen und Rheinland).

auch die Koordination aller in einem Bundesland archäologisch aktiven Akteure (Museen, Universitätsinstituten, Forschungseinrichtungen, Firmen etc.) obliegt.

Sie werden durch verschiedene untergeordnete Denkmalschutzbehörden ergänzt und unterstützt, die bei den Regierungspräsidien, Landkreisen und Kommunen angesiedelt sein können. Als gemeinsame Interessenvertretung der Landesarchäologien dient der „Verband der Landesarchäologen e.V.“³⁴, innerhalb dessen eine Kommission „Archäologie & Informationssystem“ sich mit dem Einsatz von IT in der archäologischen Denkmalpflege beschäftigt. Eine der Hauptaufgaben dieser Kommission ist es, Richtlinien, Empfehlungen und Lösungen zu erarbeiten, die den Datenaustausch zwischen den verschiedenen Fachämtern unterschiedlicher Bundesländer vereinfachen, und auf Entwicklungen zu reagieren, welche die gesamtdeutsche Denkmalpflege betreffen. Auf nationaler Ebene vertritt der Verband der Landesarchäologen die Interessen der archäologischen Bodendenkmalpflege im Rahmen der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland und gegenüber dem Beauftragten der Bundesregierung für Kultur und Medien („Kulturstaatsminister“).

Die Aufgaben der archäologischen Denkmalpflege umfassen die Untersuchung (meistens in Form von Ausgrabungen und Prospektionen), die Dokumentation und die Konservierung von beweglichen und unbeweglichen Zeugnissen der Vergangenheit, die sich im Boden oder in Gewässern erhalten haben. Die Denkmalschutzbehörden haben zudem den Auftrag, die erhobenen Daten zu sammeln, aufzubewahren und der interessierten Öffentlichkeit zu vermitteln, woraus sich häufig eine enge institutionelle Verbindung mit den Museen, in der Regel den jeweiligen Landesmuseen, ergibt³⁵. Charakteristisch für die Fachämter ist ferner die Tatsache, dass einige spezialisierte Arbeitseinheiten unterhalten werden, wie etwa Labore für Archäobotanik und Dendrochronologie oder Restaurierungswerkstätten.

Die Erfüllung einer hoheitlichen Aufgabe bedeutet, dass die Arbeit der Fachämter im weitesten Sinne eine Verwaltungstätigkeit darstellt, die durch ausgebildete Wissenschaftler und Techniker ausgeübt wird. So werden die Denkmalbehörden z. B. bei neuen Infrastrukturprojekten oder kommunalen Planungsverfahren in die Entscheidungsprozesse mit einbezogen oder müssen vor Bauvorhaben die im Boden befindlichen archäologischen Überreste bergen und dokumentieren (sog. Rettungs- oder Notgrabungen). Zur Durchführung dieser administrativen Tätigkeiten unterhalten vor allem die Landesdenkmalämter umfangreiche Datenbanken und geographische Informationssysteme, in denen zentrale, z. T. auch sensible personenbezogene Informationen zusammenfließen, vor allem:

- Fundstellen mit ihrer geographischen Lage
- abgeschlossene und laufende Maßnahmen (z. B. Prospektion, Ausgrabung, Restaurierung, Bergung etc.)
- beteiligte Akteure (z. B. Eigentümer, Investoren, Bauunternehmen, Grabungsfirmen, Gutachter, Labore etc.)
- zugehörige Dokumente (z. B. Fotos, Zeichnungen, Berichte, Tabellen, Pläne, Messreihen etc.)

³⁴ <http://www.landesarchaeologen.de/>

³⁵ Immer wieder gibt es die Konstellation, dass die Leitung eines Denkmalfachamtes und die Leitung eines Landesmuseums durch dieselbe Person wahrgenommen werden.

- geborgene Objekte und ihr aktueller Aufbewahrungsort (z. B. Depots, Museen, Archive)

Grob können die verwendeten Systeme in zwei Gruppen aufgeteilt werden, da sie entweder nach Fundstellen oder nach Aktivitäten organisiert sind. Aufgrund der Hoheit der Länder über die jeweiligen Daten und der Heterogenität der einzelnen Systeme fehlt bislang eine Plattform, welche die Informationen zu den rund eine Millionen Fundstellen in Deutschland an einer einzigen Stelle bündelt und gemeinsam zur Verfügung stellt. Das Interesse an der Standardisierung der unterschiedlichen Metadaten ist im Kreis der Landes- und Kommunalarchäologen sehr groß und kommt u.a. in einer eigenen Arbeitsgruppe innerhalb der o.g. Kommission „Archäologische Informationssysteme“ zum Ausdruck. Dort wurde u. a. der ADEX-Standard, ein basales Metadatenschema zur Beschreibung von Fundstellen entwickelt, um eine bundesländerübergreifende Zusammenarbeit zu verbessern³⁶. Das Format ist dabei so einfach gehalten, dass es auch von Fachinstitutionen außerhalb der archäologischen Landesämter genutzt werden kann. Für IANUS besitzt es daher eine hohe Relevanz zur Integration und zum Austausch heterogener ortsbezogener Metadaten.

Die Arbeitsschritte und die erhobenen Daten unterliegen dabei einer festen Prozesskette, die auch dazu geführt hat, dass bei den verschiedenen Ämtern individuelle Richtlinien und Standards zur Beantragung und Durchführung von Ausgrabungen, zur Aufbewahrung geborgener Objekte sowie zur Dokumentation und Konservierung von Bodenbefunden entwickelt wurden. Dies gilt sowohl dann, wenn die Arbeiten von angestellten Mitarbeitern der Fachämter erbracht werden, als auch wenn sie – wie in manchen Bundesländern üblich – als Aufträge an kommerzielle Grabungs- und Prospektionsfirmen vergeben werden. Auch wenn viele Ausgrabungen aufgrund gesetzlicher Verpflichtungen zur Dokumentation und Bergung von Kulturgütern durchgeführt werden und die erzeugten Informationen teilweise spezifische Verwaltungsvorgänge widerspiegeln, besitzen die Daten insgesamt ein hohes Potential für die Forschung. Dieses kann durch die Fachämter alleine nur zum Teil ausgeschöpft werden, da aufgrund oftmals zu geringer Personalstellen nicht in allen Fällen eine vollständige wissenschaftliche Aufarbeitung und Publikation der Dokumente geleistet werden kann. In einigen Ländern und Kommunen wird versucht, dieses Desiderat durch die Vergabe von Master- und Promotionsvorhaben für universitäre Qualifizierungsarbeiten zu kompensieren.

Die Art der Akteure, die an dem Prozess einer denkmalpflegerischen Maßnahme beteiligt sind, ist zumeist heterogener als bei reinen Forschungsprojekten, da neben den technischen und wissenschaftlichen Mitarbeitern des Fachamtes auch die Eigentümer eines Grundstücks, die Verursacher einer Maßnahme (z. B. Baufirmen oder Investoren), Kollegen aus anderen Behörden (z. B. Vermessungs-, Planungs- und Umweltamt) und ggf. beauftragte Dienstleistungsfirmen involviert sein können.

Auch sind die Fragen, wem nach Abschluss die geborgenen Objekte und die zugehörige Dokumentation gehören sowie wer die Urheber- und Nutzungsrechte an den Daten besitzt, je nach Bundesland unterschiedlich geregelt. Alle Denkmalfachbehörden haben nach landeseigenen Richtlinien den Auftrag, die Objekte zu konservieren und die analogen wie digitalen Dokumente zu

³⁶ ADeX® ist beim deutschen Patent- und Markenamt unter der Nr. 30 2008 030 421 als Wort-/Bildmarke für den Verband der Landesarchäologen eingetragen. <http://www.landesarchaeologen.de/verband/kommissionen/archaeologie-und-informationssysteme/projektarbeitsgruppen/adex/>

archivieren. Der allergrößte Teil der (Alt-)Dokumentation liegt dabei in analoger Form vor und wird von den einzelnen Fachämtern lokal aufbewahrt und kuratiert. Je nach Möglichkeiten wurden Maßnahmen zur Retrodigitalisierung dieser Datenbestände unternommen, um sie in die jeweiligen Fachinformationssysteme integrieren zu können. Daneben gibt es einen kontinuierlich wachsenden Anteil an digital born data, da immer stärker und umfangreicher digitale Techniken bei Ausgrabungen, Prospektionen und Vermessungen zum Einsatz kommen. Soweit erkennbar fehlen für die Langzeitarchivierung von digitalen Datenbeständen der Denkmalämter bislang etablierte Lösungen oder sind erst im Aufbau³⁷. In allen bekannten Fällen, in denen Archivstrukturen für digitale Daten existieren, handelt es sich entweder um institutionelle, generische oder fachunabhängige Einrichtungen, die in der Regel keine fachspezifische Kuratierung und Bereitstellung der Daten leisten. Aufgrund der Hoheit der Denkmalämter über diese Daten, die aus rechtlichen Bestimmungen nicht aufgegeben werden kann, und der Verpflichtung, digitale Daten in Landesarchiven aufzubewahren, spielt die von IANUS geplante Langzeitarchivierung von Daten für die Denkmalbehörden eine eher untergeordnete Rolle.

Eine neue Entwicklung hinsichtlich der Zugänglichkeit zu Datenbeständen bewirkte die INSPIRE-Initiative der EU, mit der die Veröffentlichung von raumbezogenen Daten aus öffentlich-rechtlichen Einrichtungen vorangetrieben wird. Nach der überwiegenden Auffassung der Landesarchäologen fallen auch archäologische Schutzflächen unter diese Verordnung. Das führt zwar dazu, dass vereinzelte Denkmalämter Informationen für die Geoportale ihrer Länder³⁸ zur Verfügung stellen, aber es fehlt weiterhin eine fachspezifische und länderübergreifende Lösung. Die Veröffentlichung von Geoinformationen mit hoher Genauigkeit der Fundstellen und Schutzflächen wird dabei intensiv diskutiert, da hierdurch eine Zunahme von Grabräubern und Vandalismus bei noch nicht ausgegrabenen Denkmälern befürchtet wird.

Je nach personellen und finanziellen Kapazitäten (oftmals durch Drittmittelgeber unterstützt) unternehmen die Denkmalbehörden auch Forschungsprojekte, in denen selbstgewählte wissenschaftliche Fragestellungen im Vordergrund stehen und die Tätigkeit weniger durch externe Sachzwänge, sondern durch ein Forschungsinteresse geleitet wird. Bei diesen Vorhaben werden häufig auch Kooperationen mit Museen und Universitätsinstitutionen geschlossen. Der Umfang der Forschungstätigkeit bei den Denkmalfachbehörden insgesamt fällt dabei wesentlich geringer aus als bei den Universitäten und Akademien.

3.2.5 Kommerzielle Dienstleister

Eine weitere Gruppe relevanter Akteure für die Archäologie und Altertumswissenschaften bilden kommerzielle Dienstleister. Auch diese erzeugen und verarbeiten wissenschaftlich relevante Fachdaten, häufig im Rahmen eines öffentlichen Auftrages von Museen und Denkmalämtern, aber auch für Projekte, Forschungseinrichtungen, Investoren oder Privatpersonen. Zu dieser Gruppe gehören vor allem kleine und mittlere Unternehmen, die in der Mehrheit Dienstleistungen in folgenden Bereichen anbieten:

³⁷ Ein aktuelles Beispiel ist das "Digitale Archiv-NRW", das auch von der dortigen Denkmalpflege genutzt werden soll.

³⁸ Ein besonders weit entwickeltes Beispiel ist der „Bayernviewer-Denkmal“:
http://www.geodaten.bayern.de/tomcat_files/denkmal_start.html

- Ausgrabung
- Fotografie und Zeichnung (Objekte, Gebäude, Umwelt)
- geophysikalische Prospektion
- Digitalisierung analoger Materialien (Bücher, Zeichnungen, Fotos, Texte etc.)
- Restaurierung (Objekte, Gebäude)
- Visualisierung und virtuelle Rekonstruktion (Objekte, Gebäude)
- Vermessung
- Materialuntersuchungen und Analysen (Objekte, Proben)

In der Regel arbeiten die Vertreter dieser Gruppe unter anderen rechtlichen und ökonomischen Voraussetzungen als die bisher beschriebenen vier Arten von Institutionen, die alle im weitesten Sinne durch öffentliche Mittel finanzierte Einrichtungen sind. Ein wesentlicher Unterschied besteht darin, dass kommerzielle Dienstleister immer auch wirtschaftliche Interessen an ihren Daten besitzen, zumindest für einen bestimmten Zeitraum, und diese in einem klar definierten, finanziell vergüteten Auftragsverhältnis entstanden sind.

Bei öffentlichen Auftraggebern sind die Firmen häufig dazu verpflichtet, die von Ihnen generierten Daten nach bestimmten Vorgaben nach Abschluss der Arbeiten abzugeben, wobei die Details jeweils im Vorhinein verbindlich festgehalten werden. Im Regelfall gehen dabei die Nutzungs- und Verwertungsrechte auf den Auftraggeber als finanziellen Investor über, es sei denn es gibt davon abweichende individuelle Vertragsgestaltungen. Da das Urheberrecht dagegen immer bei dem Ersteller, also dem Dienstleister als Auftragnehmer, verbleibt, wird diesem in vielen Fällen das Erstpublikationsrecht eingeräumt, um wissenschaftliche Veröffentlichungen als Werbung für die eigene Arbeit und die Beteiligung an Projekten nutzen zu können. Sofern Dienstleistungsaufträge im Kontext der Bodendenkmalpflege vergeben werden und daher zusätzliche Fachämter (etwa zur Erteilung einer Grabungsgenehmigung) involviert sind, haben diese keine eigene Rechte an den Daten, können aber Auflagen festlegen, z. B. dass ein Kopie der Daten für die amtsinterne Nutzung abgegeben werden muss.

Während die langfristige Archivierung und Bereitstellung auch von unveröffentlichten Daten kommerzieller Dienstleister eher ein Desiderat der Forschung ist, liegt die Verantwortung für diese Aufgabe meistens nicht bei den Datenproduzenten, sondern bei den Auftraggebern. Inwieweit die Fragen einer nachhaltigen Datenhaltung auch bei kommerziellen Dienstleistern eine Rolle spielen, besonders dann wenn keine Übergabe von Dokumenten an Dritte stattfindet und wie praktische Umsetzungen hierfür aussehen, ist bislang unbekannt. Die Tatsache, dass einige Akteure Mitglieder des Bundesverbands freiberuflicher Kulturwissenschaftler sind und einem Kodex folgen, in dem die Grundprinzipien und ethischen Normen für den Erhalt und die Pflege des historischen Erbes definiert werden ³⁹, belegt wie sehr sie sich ihrer gesellschaftlichen, kulturellen und wissenschaftlichen Verantwortung bewusst sind. Gleichzeitig zeigen Erfahrungen in England, dass auch die Bereitstellung unprozessierter Daten von kommerziellen Dienstleistern als eine Möglichkeit zur öffentlichkeitswirksamen Darstellung der eigenen Arbeiten wahrgenommen wird. Trotz der genannten Einschränkungen ist diese Gruppe für IANUS insofern ein interessanter

³⁹ <http://www.b-f-k.de/firmen/kodex.php><http://www.b-f-k.de/firmen/kodex.php>

Ansprechpartner, weil sich deren Mitglieder alltäglich mit der adäquaten Vorbereitung von Daten für eine Übergabe an Dritte und der Einhaltung von Standards auseinandersetzen müssen.

3.3 Fachverbände, Vereine, Arbeitsgemeinschaften

Neben den institutionellen Akteuren in den Altertumswissenschaften gibt es eine Reihe von Fachverbänden, Vereinen und Arbeitsgemeinschaften, die sich über gemeinsame inhaltliche Methoden, geographische Räume (sowohl historische Kulturregionen als auch moderne politische Gebiete), Fachdisziplinen oder andere Merkmale definieren. Einige der wichtigsten, da stark überregional agierende Vereinigungen sind:

- AG für Christliche Archäologie (AGCA)
- AG Computeranwendungen und Quantitative Methoden in der Archäologie e.V. (CAA-DE)
- Archäologische Kommission für Niedersachsen e.V. (AK-NDS)
- Dachverband Archäologischer Studierendenvertretungen e.V. (DASV)
- Deutsche Gesellschaft zur Förderung der Unterwasserarchäologie e.V. (DEGUWA)
- Deutsche Gesellschaft für Archäologie des Mittelalters und der Neuzeit e.V. (DGAMN)
- Deutsche Gesellschaft für Ur- und Frühgeschichte e.V. (DGUF)
- Deutsche Orient Gesellschaft e.V. (DOG)
- Deutscher Archäologenverband e.V. (DArV)
- Deutscher Verband für Archäologie e. V. (DVA)
- Deutscher Verein zur Erforschung Palästinas
- Fachgruppe „Archäologische Museen“ im Deutschen Museumsbund (DMB)
- FemArc – Netzwerk archäologisch arbeitender Frauen e.V.
- Gesellschaft für Archäologie in Bayern e.V.
- Gesellschaft für Naturwissenschaftliche Archäologie ARCHAOMETRIE e.V. (GNAA)
- Hugo Obermaier-Gesellschaft für Erforschung des Eiszeitalters und der Steinzeit e.V.
- Koldewey-Gesellschaft – Vereinigung für baugeschichtliche Forschung e.V.
- Mittel- und Ostdeutscher Verband für Altertumforschung e.V. (MOVA)
- Mommsen-Gesellschaft e.V. – Verband der deutschen Forscher auf dem Gebiete des griechisch-römischen Altertums
- Nordwestdeutscher Verband für Altertumforschung e.V. (NWDV)
- Spätantike Archäologie und Byzantinische Kunstgeschichte e.V.
- Sudanarchäologische Gesellschaft zu Berlin e.V.
- Verband der Landesarchäologen e.V. (VLA)
- West- und Süddeutscher Verband für Altertumforschung e.V. (WSVA)

Die besondere Relevanz dieser Vereinigungen für IANUS beruht auf verschiedenen Aspekten. Zunächst kann eine gezielte Informationspolitik betrieben werden, da diese Gruppierungen hinsichtlich ihrer Mitglieder in der Regel relativ homogen zusammengesetzt sind. Es ist daher eine themen-, disziplinen- oder organisationsspezifische Vorstellung von IANUS und seinen Dienstleistungen möglich und es können individuelle Bedürfnisse und Hindernisse in Bezug auf

IANUS ermittelt werden, etwa im direkten Austausch auf Verbandstagungen oder Sitzungen von Arbeitsgruppen. Umgekehrt kann IANUS auch von der gebündelten Fachkompetenz profitieren, z. B. bei der Ausarbeitung oder Aktualisierung bestimmter Kapitel in den IT-Empfehlungen oder bei der Aufbereitung von Daten für eine Langzeitarchivierung, um eine optimale fachspezifische Nachnutzung ermöglichen zu können. Ein anderer Mehrwert ist darin zu sehen, dass die Vereinigungen bzw. deren Leitungsgremien als Multiplikatoren zu den einzelnen Mitgliedern fungieren und Informationen über IANUS weiterleiten können. Darüber hinaus ist sogar denkbar, dass explizite Empfehlungen zur Archivierung von Datenbeständen ausgesprochen werden und die Vertrauenswürdigkeit des Forschungsdatenzentrums unterstrichen wird.

3.4 Nutzergruppen

Nach dem OAIS-Standard sind in einem Datenarchiv zwei Nutzergruppen zu unterscheiden:

- Datenproduzenten⁴⁰, die für bestimmte Daten verantwortlich sind und diese an eine Einrichtung zum Ziel der Kuratierung, Archivierung und Bereitstellung übergeben
- Datennachnutzer, welche die Daten, die ein Archiv über seine technischen Plattformen zur Verfügung stellt, nachnutzen wollen

Da die Rolle des Datenproduzenten in dem Referenzmodell relativ weit aufgefasst und unscharf definiert ist, wird sie in dem Kontext von IANUS mittels zweier weiterer Nutzergruppen – die Dateneigentümer und die Datengeber – stärker ausdifferenziert. Diese Ergänzung zu OAIS ist für das Konzept und den Betrieb von IANUS notwendig, um verschiedene Akteure eindeutig beschreiben und um vor dem Hintergrund der rechtlichen Verantwortung und Haftung klar hervorheben zu können, mit welcher Art von Nutzern zu einem gegebenen Zeitpunkt innerhalb des gesamten Kuratierungsprozesses Informationen ausgetauscht werden. Dabei kann eine reale Person mehrere Rollen in sich gleichzeitig vereinen.

Neben diesen vier Gruppen gibt es für IANUS eine weitere Gruppe, deren Beziehung nicht über den Austausch von Daten definiert wird, sondern durch die Inanspruchnahme von nicht-technischen Dienstleistungen, die Sonstigen Nutzer.

3.4.1 Datenproduzenten, Dateneigentümer, Datengeber

Als Datenproduzenten im Sinne von IANUS werden sowohl Individuen und Organisationen als auch Systeme, für die legitime Vertreter existieren, aufgefasst, die direkt und unmittelbar neue Forschungsdaten generieren und im juristischen Sinne das nicht abtretbare Recht des Urhebers an diesen besitzen. Daneben kann es Individuen und Organisationen als separate Dateneigentümer geben, denen in einem eigentumsrechtlichen Sinne die Daten gehören und die etwa über die Nutzungs- und Verwertungsrechte entscheiden. So kann z.B. eine Institution Eigentümer von Daten sein, ohne dass diese Daten persönlich durch deren Repräsentanten erhoben wurden, wenn in einem Arbeitsvertrag geregelt ist, dass die während eines Anstellungsverhältnisses generierten Daten nicht in den Besitz des individuellen Datenproduzenten, sondern in das Eigentum der

⁴⁰ Soweit wie möglich werden in dem Konzept die Begriffe und Definitionen des OAIS-Referenzmodells in der deutschen Übersetzung verwendet: OAIS-de (2012).

Institution übergehen. Die Urheberrechte an den Daten besitzen Dateneigentümer nur dann, wenn sie auch selbst die Daten produziert haben. Mit der Rolle des Datengebers wird die natürliche Person bezeichnet, welche direkten Kontakt zu IANUS aufnimmt, um Forschungsdaten bei IANUS kuratieren, archivieren und bereitstellen zu lassen. Ein Datengeber kann identisch mit dem Dateneigentümer und/oder dem Datenproduzent sein, muss es aber nicht zwangsläufig, wenn er z.B. im Auftrage Dritter handelt. In der Praxis werden in vielen Fällen die Datengeber mindestens auch die Datenproduzenten sein.

Da nur ein Dateneigentümer die Archivierung und Bereitstellung von Daten veranlassen und den Datenübergabevertrag (siehe **Kapitel 10.1.2**) verbindlich unterzeichnen kann, obliegt IANUS die essentielle Sorgfaltspflicht⁴¹ bei neu angebotenen Datensammlungen kritisch zu prüfen, welcher Akteur welche Beziehung zu den Daten besitzt und welche Rechte und Pflichten sich daraus ableiten lassen. Dabei ist zu beachten, dass häufig auch bei den Institutionen und Wissenschaftlern nicht klar ist, wer die Rechte an den jeweiligen Daten tatsächlich besitzt.

Alle drei Akteure – Datenproduzenten, Dateneigentümer und Datengeber – bilden die entscheidende Zielgruppe für IANUS, da erst dadurch, dass diese ihre Daten zur langfristigen Archivierung und Bereitstellung anvertrauen, das Forschungsdatenzentrum seine primäre Zielsetzung erfüllen kann. Erst mit einer kritischen Masse an Daten erfährt IANUS in den Fach-Communities die notwendige Relevanz, die dessen Aufbau und Betrieb rechtfertigt. Insofern gilt es, diese Gruppe kontinuierlich zu adressieren, führende Akteure in die Aktivitäten von IANUS einzubinden und die Angebote entsprechend sich verändernder Erwartungshaltungen bzw. fachlicher Anforderungen anzupassen. Kriterien für eine erfolgreiche Ansprache sind:

- klare Vermittlung der Vorteile für eine professionelle Datenarchivierung und Datenbereitstellung
- zuverlässige und professionelle Langzeitarchivierung über festgelegte Zeiträume
- effiziente Mechanismen zur Datenübertragung
- klare und einfache Nutzeroberflächen der Systeme
- keine oder nur geringe Einschränkungen für die aktuelle Forschungsarbeit (z. B. bzgl. bestimmter Datenformate)
- Unterstützung durch Standards, Tools und Vorlagen, um den (zusätzlichen) Aufwand für die Dokumentation und Datenaufbereitung so gering wie möglich zu halten
- moderate, transparente und kalkulierbare Kosten, deren Höhe durch die Qualität der Daten bei Übergabe an IANUS und die Anforderungen an die Bereitstellung beeinflusst wird
- Sicherheitsmechanismen und Rechte-Rollen-Konzepte, um einen unbefugten Zugriff auf die Daten zu verhindern
- die Hoheit über und Verantwortung für die Daten verbleiben beim Dateneigentümer⁴², die Urheberrechte verbleiben bei den Datenproduzenten
- Gestaltungsmöglichkeiten bzgl. der eigenen Metadaten (z. B. eigenständige Vergabe von

⁴¹ Vgl. die Rechtsexpertisen von P. Klimpel (<http://dx.doi.org/10.13149/000.3h7mtr-d>) und K. Hoeren (<http://dx.doi.org/10.13149/000.0nc98i-f>) sowie die Ergebnisse eines 2013 durchgeführten Workshops zum Thema Lizenzen (<http://hdl.handle.net/11858/00-1780-0000-0022-DBE5-C>).

⁴² Vgl. IANUS. SCHWIEGELSHOHN (2013) S. 30.

Zugriffsrechten)

- Effiziente Unterstützung und Beratung beim Forschungsdatenmanagement

Welche allgemeinen Kriterien die Daten von Datenproduzenten erfüllen müssen, damit sie von IANUS akzeptiert werden können, ist in **Kapitel 14.1** beschrieben. Sofern Datenproduzenten ihre Daten IANUS anvertrauen, wird ein Datenübergabevertrag abgeschlossen, um die Rechte und Pflichten beider Parteien rechtsverbindlich zu regeln. (siehe **Kapitel 9.2**) Dieser enthält u.a. eine Einwilligung zur Kuratierung und Migration der übergebenen Daten und legt eventuelle Nutzungs- und Zugriffseinschränkungen, die Art der Präsentation auf der IANUS-Homepage und die Höhe der Gebühren für die Kuratierung fest.

3.4.2 Datennachnutzer

Alle Mitglieder der in **Kapitel 3.1** und **3.2** beschriebenen Fachrichtungen und Institutionen sind als primäre Datennachnutzer von IANUS zu bezeichnen, d. h. die Dokumentation und Aufbereitung von Daten und Informationen erfolgt in der Weise, dass diese Personenkreise die Informationen ohne zusätzliche Hilfsmittel und auf Basis des eigenen Fachwissens und in Kenntnis der deutschen Sprache verstehen und nutzen können. Dateneigentümer, Datenproduzent und Datengeber können dabei ebenfalls die Rolle eines Datennachnutzers ausüben.

Eine sekundäre Gruppe bildet die interessierte Öffentlichkeit und Angehörige anderer Wissenschaften, die aus individuellem Interesse, zum Zwecke der Lehre und Ausbildung oder aus sonstigen Gründen über IANUS bereitgestellte Daten nutzen wollen. Die Möglichkeiten der Anzeige und Nutzung einzelner Datensätze kann dabei durch Zugriffsrechte eingeschränkt sein.

Die Nutzung der Daten ist dabei prinzipiell für alle Datennachnutzer kostenfrei. Vor einer Nutzung muss den Nutzungsbedingungen zugestimmt werden. (siehe **Kapitel 9.3**)

Auch wenn die Datennachnutzer erst dann eine größere faktische Bedeutung für IANUS besitzen, wenn eine gewisse Quantität von Daten (eine „kritische Masse“) bereitgestellt sein wird, also Daten gesucht und gefunden werden können, gilt es diese Gruppe frühzeitig in der Planungsphase zu berücksichtigen, da sie das Qualitäts- und Risikomanagement beeinflussen.

Die Praxis bestehender Datenzentren zeigt, dass für eine erfolgreiche Nachfrage der Daten durch Datennachnutzer folgende Aspekte berücksichtigt werden müssen:

- Schnelles und gutes Auffinden von relevanten Forschungsdaten („zielgerichtete Suche“)
- Schneller Überblick über bestimmte Fachgebiete („nicht-zielgerichtetes Browsen“)
- Übersichtliche Darstellung der Ergebnisse
- Eindeutige Lizenz- und Nutzungsbedingungen für die Daten
- Möglichkeiten zum lokalen Download von Daten
- Einbindung von Daten in lokale Systeme über technische Schnittstellen
- Nachweis der Daten in IANUS in unterschiedlichen Portalen, z. B. Propylaeum, Europeana und Deutsche Digitale Bibliothek

3.4.3 Sonstige Nutzer

Alle weiteren Personen, die Dienstleistungen von IANUS in Anspruch nehmen und dies nicht in den o.g. Rollen tun, lassen sich in einer weiteren Gruppe zusammenfassen, die tendenziell sehr

heterogen ist. Zu ihnen gehören z. B. folgende Personenkreise

- Autoren und Leser der IT-Empfehlungen
- Nutzer von Online-Tutorials und Präsentationen
- Teilnehmer an Veranstaltungen wie Summer Schools, Workshops, Schulungen etc.
- Empfänger bzw. Abonnenten von Social-Media-Kanälen, Newslettern und elektronischer Kommunikation
- Partner in Drittmittelprojekten

4 Daten

4.1 Zum Begriff Forschungsdaten

Gegenstand der Bemühungen von IANUS sind Forschungsdaten aus den Archäologien und Altertumswissenschaften. Der Begriff Forschungsdaten wird dabei in einem weiten Sinne verstanden, der alle Daten umschließt, die für die Forschung in diesen Disziplinen interessant sein können. Danach sind sowohl Primärdaten relevant, die zur gezielten Beantwortung wissenschaftlicher Fragen z. B. auf Ausgrabungen als Fotos und Zeichnung, bei chemisch-physikalischen Analysen als Messreihen und Graphen oder bei Editionen in Form von Textannotationen erhoben werden, als auch solche, die z. B. im Rahmen von Verwaltungs- und Inventarisationsvorgängen in Denkmalfachbehörden und Museen oder bei Digitalisierungsmaßnahmen generiert werden.

Die Daten haben dabei im Allgemeinen eine deutlich längere Lebensdauer als die Projekte und Systeme, in denen sie entstanden sind und werden nur zum Teil bzw. lediglich als Auswahl über Veröffentlichungen in den wissenschaftlichen Diskurs eingebunden. Dies erfolgt in den allermeisten Fällen in den Altertumswissenschaften immer noch in Form von gedruckten Monographien, Zeitschriften und Kongressbeiträgen. Die digitalen Grundlagen, wie auch alle unveröffentlichten Daten, geraten auf den lokalen Speichermedien in Vergessenheit oder werden komplett gelöscht und stehen somit einer Nachnutzung nicht mehr zur Verfügung. Aus diesem Grund liegt der Schwerpunkt von IANUS auf den unbearbeiteten, originalen Primärdaten wie auf den prozessierten Sekundärdaten unterschiedlicher Arbeitsstände, da diese nahezu vollständig dem Datenlebenszyklus entzogen sind (Abb. 2) und in besonderer Weise für die Zukunft gefährdet sind. Des Weiteren sind auch die digitalen Versionen von gedruckten Ergebnissen relevant, werden berücksichtigt und zur Kuratierung angenommen.

4.2 Charakteristika archäologisch-altertumswissenschaftlicher Daten

4.2.1 Fachlich-methodische Heterogenität (Datengense)

Da unter dem Begriff Altertumswissenschaften ganz unterschiedliche Disziplinen aus den Geistes- und Naturwissenschaften zusammengefasst werden, weisen die für IANUS relevanten Daten eine hohe fachlich-methodische Heterogenität auf. Typische Arbeitsweisen sind etwa:

- archäologische Ausgrabungen
- Oberflächenbegehungen (Survey)
- geophysikalische Prospektionen
- Entnahme geologischer Bohrkerne
- architektonische Bauaufnahmen und Rekonstruktionen
- naturwissenschaftliche Materialuntersuchungen
- Bestimmung von Knochen- und Pflanzenresten
- Vermessung und Geodäsie
- Anfertigung von Laserscans und 3D-Dokumentation
- Analyse geographischer Räume und Landschaftsmodellierung
- Auswertung von Luftbildern und Satellitenaufnahmen

- syntaktische und semantische Quellenkritik
- Annotation von Texten

4.2.2 Technische Heterogenität (Datenformate)

Genauso vielfältig wie die Methoden und Disziplinen sind auch die verwendeten Datenformate. Eine statistische Auswertung von Testdaten aus 19 Projekten im Umfang von ca. 700 GB Speicher und knapp 230.000 Dateien ergaben 462 unterschiedliche Dateiformate, die produziert und verarbeitet werden⁴³. Auch wenn IANUS zur Archivierung und Kuratierung nicht alle Formate bei der Datenübergabe akzeptieren wird (siehe **Kapitel 14.1**), so existieren doch mehrere Datenarten, die eine erhöhte Relevanz für die Altertumswissenschaften besitzen:

- Vektorgrafiken / Zeichnungen / CAD
- Rastergrafiken, insbesondere digitale Fotos
- freie Texte und beschreibende Dokumente
- strukturierte Daten wie Tabellen, Listen, Kataloge
- Mark-Up-Dateien und Webseiten
- statistische Daten
- Datenbanken
- Koordinaten und raumbezogene Daten, insbesondere aus geographischen Informationssystemen
- geophysikalische Messdaten
- naturwissenschaftliche Messreihen
- Punktwolken von Laserscannern
- Rekonstruktionen / 3D-Daten / Virtual Reality

4.3 Daten im Kontext von IANUS

4.3.1 Verwaiste Daten (sog. Altdaten)

In mehreren Einrichtungen gibt es erhebliche Mengen an sogenannten Altdaten oder verwaisten Daten. Darunter werden Datenbestände aus abgeschlossenen Projekten verstanden, für die keine regulären Finanzmittel mehr vorhanden, keine Personen aktuell tätig und interne Mitarbeiter als Kenner der Daten häufig nicht mehr erreichbar sind. Als Ansprechpersonen fungieren oftmals nur Angestellte der Institutionen, an denen das Projekt durchgeführt wurde, die aber nicht persönlich bei der Genese und Verarbeitung dieser Datenbestände beteiligt waren und diese daher in der Regel nur unzureichend kennen. Gelegentlich liegen die Projekte bereits mehrere Jahre zurück, so dass zusätzlich die Gefahr besteht, dass die Dateien technisch nicht mehr oder nur mit unverhältnismäßig großem Aufwand gelesen werden können und darüber hinaus erforderliche Metainformationen zum Verständnis fehlen bzw. über die Zeit verloren gegangen sind.

⁴³ M. Trognitz (2014), Statistische Auswertung der Testdaten für IANUS [Version 1.0], Hrsg. von IANUS.
doi:10.13149/000.j5ecae-l

Dennoch sind auch diese Daten für die zukünftige Forschung oft von großer Bedeutung. Wenn diese in einem wesentlichen Umfang auf die Dauer unberücksichtigt bleiben, werden sie als wichtige Grundlagen dem Erkenntnisprozess in den Archäologien und den Altertumswissenschaften dauerhaft entzogen sein. Es ist daher das langfristige Ziel von IANUS, möglichst viele Altdaten in das Langzeitarchiv zu übernehmen. Jeder Bestand ist dabei individuell zu behandeln und wird überdurchschnittlichen Aufwand an Kuratierung, Dokumentation und Erschließung erfordern. Diese hängen jeweils von der Kategorie der Altdaten ab:

- Daten, die nicht in digitalisierter Form vorliegen
- digitalisierte Daten, die sich auf veralteten Speichermedien befinden
- digitalisierte Daten, die in nicht mehr unterstützten Dateiformaten vorliegen
- digitalisierte Daten, für die keine bzw. nicht ausreichende Dokumentation existiert oder deren Metadatenformat nicht den unterstützten Formaten entspricht
- digitalisierte Daten, deren Metadaten im gewünschten Zielformat vorliegen

4.3.2 Aktuelle Projektdaten

Einfacher in der Handhabung sind Daten aus aktuellen oder erst kürzlich abgeschlossenen Projekten, bei denen Verantwortliche und Datenbeauftragte erreichbar sowie Gelder für die Kuratierung und Archivierung vorhanden sind. Im Idealfall, d. h. bei einer frühzeitigen Kontaktaufnahme mit IANUS, kann bereits während der Durchführung eines Projektes eine Unterstützung bei der Erhebung, Verarbeitung und Analyse der Daten erfolgen und die Vorbereitungen für die Nachnutzung und Langzeitarchivierung der Daten begleitet werden. Ein weiterer Vorteil ist, dass Personen existieren, welche den Kontext der Datenerhebung kennen und somit die Dokumentation der Daten als auch die Auszeichnung der Daten mit Metadaten vornehmen können. Ähnlich ist es bei kürzlich abgeschlossenen Projekten, bei denen es möglich ist, Verantwortliche zu kontaktieren und diese Rückfragen beantworten können.

4.3.3 Kontinuierliche Daten

Eine dritte Gruppe von Daten, die für IANUS relevant sein wird, sind solche aus laufenden, mehrjährigen Projekten, bei denen es sinnvoll ist, zu einem bestimmten Zeitpunkt einen Teil der Daten zu archivieren, z. B. Corpora-Projekte oder Langzeitgrabungen, die vergleichbar mit Langzeitstudien in den Sozial- und Umweltwissenschaften sind. Ähnliches gilt auch für laufende Fachsysteme, die kontinuierlich gepflegt und aktualisiert werden und für die es keinen definierten Abschluss wie bei Projekten mit definierten Laufzeiten gibt. In beiden Fällen sind zusätzliche Arbeiten wie Versionierungen und Verweise auf ältere Archivierungspakete erforderlich.

4.3.4 Kriterien

Ganz generell wird IANUS bereit sein, für diejenigen digitalen Daten die technische Verantwortung zu übernehmen, die eine oder mehrere der folgenden Kriterien erfüllen:

- Daten, die statisch sind und sich nicht mehr ändern
- Daten, die unter Beteiligung von deutschen Forschern erzeugt wurden
- Daten, die durch deutsche Mittelgeber (mit)finanziert wurden

- Daten, deren Dokumentation (auch) in deutscher Sprache vorliegt
- Daten, die zu einer der in **Kapitel 3.1** genannten Fachrichtungen zuzuordnen sind
- Daten, die von internationalen Forschern erzeugt und internationalen Mittelgebern finanziert wurden, sich aber auf Deutschland beziehen
- Daten, die den in **Kapitel 14.1** skizzierten inhaltlichen und technischen Vorgaben entsprechen

Für angebotene Daten, die diese Kriterien nicht erfüllen, muss jeweils eine individuelle Entscheidung getroffen werden, ob eine Übernahme in die Archivbestände von IANUS möglich ist und eine dauerhafte Kuratierung gewährleistet werden kann. Nähere Details werden in einer Sammlungsstrategie (*collections policy*) festgelegt. (siehe auch **Kapitel 6.1.1**)

5 Anforderungen an ein OAIS-konformes Archiv

Aufgrund der Zielsetzung von IANUS kommt dem OAIS-Referenzmodell (ISO-Standard 14721:2012 Open Archive Information System) eine besondere Relevanz zu⁴⁴. Es beschreibt in einem theoretischen Funktionsmodell alle organisatorischen und technischen Prozesse der digitalen Archivierung und in einem Informationsmodell die zu verarbeitenden Daten. Die formulierten Konzepte und Vorgaben dienen als Leitlinie für die Entwicklung und den Betrieb eines Archives und werden bei der Planung und Umsetzung von IANUS in hohem Maße Anwendung finden, auch um eine später angestrebte Zertifizierung (siehe **Kapitel 11.1**) zu erleichtern. Allerdings gibt das OAIS-Referenzmodell nicht vor, wie die Ziele der dauerhaften Erhaltung und Bereitstellung von digitalen Informationen praktisch umgesetzt werden sollen bzw. können. Es wird davon abgesehen, eine Einschränkung bezüglich Dateitypen, -formaten oder Systemarchitekturen vorzunehmen. Darin liegt eine der wesentlichen Stärken des Modells: Es ist gänzlich funktional und zeichnet sich durch seine Allgemeingültigkeit und seinen modularen Aufbau aus. Gleichzeitig erfordert die konkrete Umsetzung und technische Implementierung der abstrakten Funktionsvorgaben entsprechende Überlegungen, die das Modell wiederum bewusst ausklammert. Ein weiterer Vorteil des Referenzmodells ist, dass es zur Etablierung einheitlicher Fachtermini beiträgt, die zwischen IT-Lösungen und archivalischen Fachaufgaben vermitteln.

Nach dem OAIS-Standard muss eine Organisation folgende Aufgaben verbindlich erfüllen⁴⁵:

- Einwerbung, Verhandlung und Annahme von geeigneten Datenbeständen von Datenproduzenten und Dateneigentümern
- Erlangung der erforderlichen Kontrolle über die Daten, um deren Langzeiterhaltung sicherzustellen
- Definition der Stakeholder bzw. Zielgruppen, die entweder als Datengeber oder als Datennachnutzer die angebotenen Informationen verstehen sollen
- Sicherstellung der uneingeschränkten und unabhängigen (Nach-)Nutzbarkeit der Daten durch die vorgesehenen Zielgruppen, welche die Informationen unmittelbar (d. h. ohne spezielle Hilfsmittel oder den Kontakt mit den Datenproduzenten) verstehen sollen
- Erstellung und Umsetzung von Richtlinien und Dokumentation von Abläufen, wobei alle potentiell relevanten und bekannten Gefahren identifiziert werden müssen, um den Erhalt von bereits archivierten Daten sicherzustellen. Das betrifft zum Beispiel das unbeabsichtigte Löschen von Daten, aber auch die Sicherstellung, dass bei Schließung des Archivs die vorhandenen Forschungsdaten nicht verloren gehen (sog. Exit-Strategie).
- Verbreitung und Verfügbarmachung der Informationen für die definierten Zielgruppen, entweder als Kopie der ursprünglich anvertrauten Datenobjekte oder als Kopie der im Kuratierungs-/Archivierungsprozess geänderten Originaldateien, welche in jedem Fall eine nachvollziehbare Referenz auf das archivierte Ursprungsobjekt enthalten muss, um die Authentizität der Inhalte zu garantieren.

Neben dem OAIS-Referenzmodell existieren zwar einige weitere zertifizierte Vorgaben, de facto-

⁴⁴ OAIS-DE (2012). Eine Zusammenfassung bietet: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_368.pdf

⁴⁵ Vgl. OAIS-DE (2012) S. 27.

Standards und gesetzliche Vorschriften, doch sind diese häufig mit Blick auf das klassische Archivgut oder das Records Management entstanden, wie es für Geschäftsprozesse der Verwaltung oder in Wirtschaftsunternehmen üblich ist⁴⁶. Sie sind daher nur sehr bedingt für den Kontext von IANUS anwendbar, da die Archivierung und Bereitstellung von Forschungsdaten sich signifikant von der Langzeitarchivierung der genannten Dokumente und Datengruppen unterscheidet.

Im Folgenden sollen die technischen Anforderungen, die sich aus dem OAIS-Standard ableiten lassen, für IANUS konkretisiert werden. Im Wesentlichen geht es dabei um das enthaltene Datenmodell, das gemeinsam mit dem Funktionsmodell die beiden Kernkomponenten des OAIS bildet. Es stellt Anforderungen an Form und Beschreibung der digitalen Archivalien, die verwaltet werden sollen und kann nie losgelöst von dem Funktionsmodell betrachtet werden. Letzteres definiert wiederum Prozesse und Arbeitsabläufe innerhalb eines digitalen Archivs, die unabdingbar für die Übernahme, die Erhaltung und die Zugänglichkeit digitaler Objekte sind. Die entsprechende Bereitstellung für eine bestimmte Zielgruppe mündet dabei in unterschiedlichen, zusätzlichen Anforderungen.

Das gemeinsame Ziel aller Aspekte ist es, eine künftige Nachnutzbarkeit von Daten zu gewährleisten, wofür sowohl die physische Aufbewahrung von Daten (*Bitstream Preservation*) als auch die Interpretierbarkeit der Informationen (*Content Preservation*) sichergestellt werden müssen.

5.1 Datenmodell und Informationsobjekte

Die im OAIS-Datenmodell vorgegebenen Inhalte bilden den theoretischen Kern für jegliche technische Konkretisierung. Danach werden digitale Archivalien als Informationsobjekte definiert, die als logischer Container immer aus zwei Elementen zusammengesetzt sind:

- Originaldaten bzw. Inhaltsinformationen (*content information*): Sie bestehen aus den übernommenen, ggf. aufbereiteten digitalen Ursprungsobjekten (*data objects*), die es zu erhalten gilt sowie beschreibenden Repräsentationsinformationen (*representation information*), also notwendigen Angaben, um Dateien für eine Zielgruppe sichtbar und verstehbar zu machen.
- Archivierungsangaben bzw. Erhaltungsmetadaten (*preservation description information – PDI*): Sie enthalten Angaben über Verfahren und Techniken, die auf Inhaltsinformationen angewandt wurden, z. B. für Veränderungen, Sicherung, Identifizierung und Kontextualisierung. Diese Information, die für die dauerhafte Aufbewahrung benötigt werden, können in die fünf Kategorien *Provenienz*, *Referenz*, *Beständigkeit*, *Kontext* und *Zugriffsrechte* eingeteilt werden. (ausführlich **Kapitel 12.1**)

Identifizierbar sind beide Elemente durch Verpackungsinformationen (*packaging information*), die das Verhältnis der einzelnen digitalen Objekte und Metadaten innerhalb eines Informationspakets zueinander bestimmen und damit als Bindeglied zwischen ihnen fungieren. Durch diese Informationen lassen sich alle Komponenten als eine einzige logisch zusammenhängende Einheit innerhalb eines Archives identifizieren und adressieren.

⁴⁶ Z. B. der ISO-Standard 15489-1 (http://www.iso.org/iso/catalogue_detail?csnumber=31908) oder das Model Requirements for Records Systems (MoReq2010) (<http://sysresearch.org/moreq/>)

Zur Wiederauffindbarkeit und Verwaltung eines Informationspakets existieren zusätzliche strukturierte Metadaten als Erschließungsinformationen (*descriptive information*), die Paketbeschreibung (*package description*) genannt werden und in Findmitteln (z. B. dem Archiv-Management-System) vorgehalten werden.

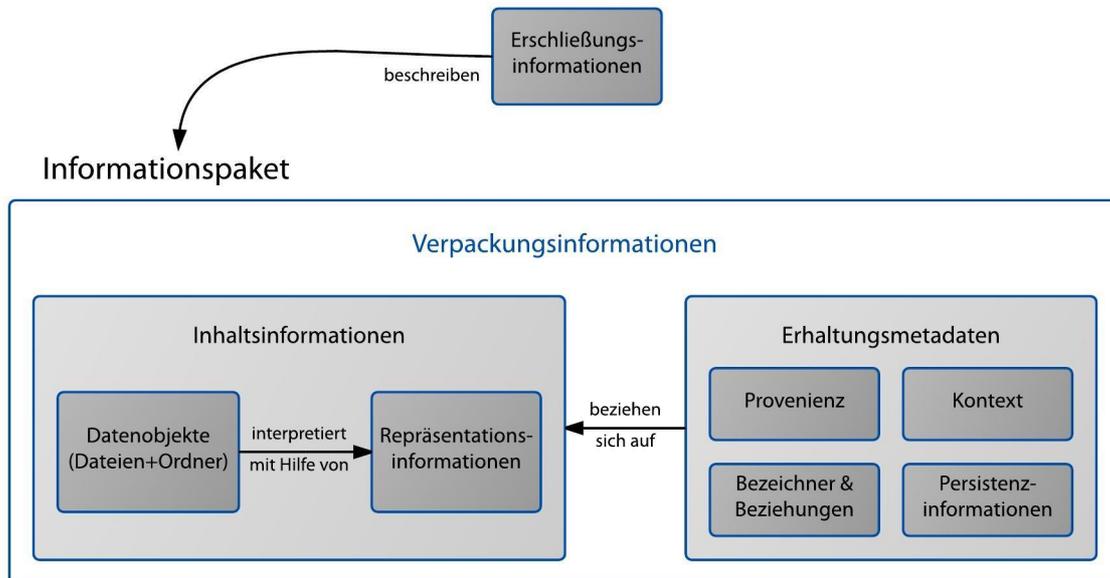


Abb. 3 Bestandteile und Beziehungen eines Informationsobjektes nach dem OAIS-Referenzmodell

Im OAIS-Referenzmodell werden drei in Verbindung stehende und aufeinander bezogene Informationsobjekte unterschieden, so dass die verschiedenen Prozesse der digitalen Archivierung getrennt betrachtet werden können. Das ermöglicht eine klare Beschreibung zum Umgang und zur Verarbeitung während einzelner Phasen der Archivierung:

- Das Übergabepaket (*submission information package – SIP*) bezeichnet die Form, in der valide Daten in eine Archivinfrastruktur übernommen werden. Die konkrete Art und der Umfang des Pakets werden zwischen dem Datenproduzenten und Datenarchiv ausgehandelt. Für den Fall, dass die eingelieferten Originaldaten, im Kontext von IANUS als *Transferpakete (TP)* bezeichnet, nicht unmittelbar für die weitere Verarbeitung zu einem AIP geeignet sind (z. B. aufgrund fehlender Dokumentationen oder nicht berücksichtigter Vorgaben), müssen sie in einem sog. Pre-Ingest-Prozess in ein valides und konsistentes SIP überführt werden.
- Das Archivpaket (*archival information package – AIP*) dient zur Aufbewahrung der übergebenen Daten im Archivspeicher. Die Konzeption des Pakets übernimmt das Digitale Archiv unter Berücksichtigung der Anforderung seiner Zielgruppe. Auch wenn das AIP in einem Repository zwischen der Erstellung und der tatsächlichen Nutzung existiert, spielt es für die Gesamterhaltung von Daten eine zentrale Rolle. Während SIPs und DIPs mehrfach erzeugt und umgewandelt werden können, um in unterschiedlichen Archivsystemen und Umgebungen zur Datenbereitstellung nutzbar zu sein, ist die Umwandlung von Daten in ein AIP in aller Regel nur ein einziges Mal erforderlich. Insofern lautete die generelle Anforderung an ein AIP ein digitaler Container zu sein, der digitale Inhalte über lange Zeiträume hinweg speichern kann, herstellerunabhängig ist, in mehreren Kopien auf

unterschiedlichen Speichersystemen existieren kann und bei katastrophalen Ereignissen eine Wiederherstellung gewährleistet.

- Das Auslieferungspaket (*dissemination information package* – DIP) stellt die übergebenen und archivierten Daten für externe Datennachnutzer bereit. Um eine Spezifizierung des DIP zu ermöglichen und entsprechende Werkzeuge für den Zugang digitaler Objekte umzusetzen, müssen in der designierten Zielgruppe die gängigen Methoden und die Bedürfnisse der Nutzer möglichst genau analysiert werden. Der Zugang zu einem DIP bildet die Brücke zwischen einem Archiv und den Nutzern, da dadurch eine Datensammlung sichtbar und nachnutzbar wird.

Ein SIP kann in ein oder mehrere AIPs überführt werden; auch kann ein DIP aus einem oder mehreren AIPs gebildet werden. Alle Details zur Strukturierung der Informationspakete und zur Ablage der Daten im konkreten Kontext von IANUS werden in **Kapitel 13** beschrieben.

5.2 Funktionsmodell

Neben den Informationsobjekten werden in dem OAIS-Referenzmodell auch die Prozesse und Arbeitsabläufe innerhalb eines digitalen Archivs anhand eines Funktionsmodells definiert. Dabei werden sechs Einheiten unterschieden, die dank des modularen Aufbaus an unterschiedliche Umgebungen angepasst werden können: Übernahme (*ingest*), Archivspeicher (*archival storage*), Datenverwaltung (*data management*), Administration (*administration*), Erhaltungsplanung (*preservation planning*), Zugriff (*access*).

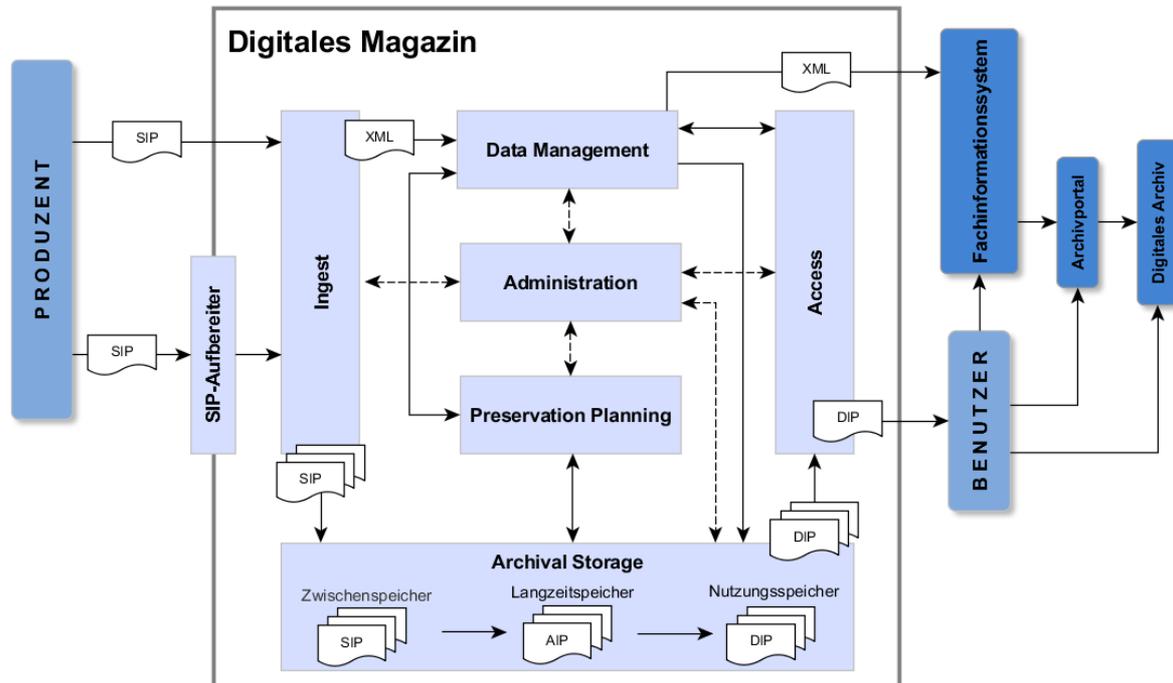


Abb. 4 Die Funktionseinheiten und Informationspakete nach dem OAIS-Referenzmodell⁴⁷

⁴⁷ Abb. angefertigt von dem Projekt „Digitales Magazin Thüringen“:

Bei der *Übernahme (ingest)* der zu archivierenden Datenbestände muss das Archiv bereits in der Vorbereitung zur Einlagerung Kontakt mit den Datengebern aufnehmen und technische Unterstützung leisten. Zentral ist dabei die Qualitätssicherung der SIPs nach definierten Kriterien und die anschließende Herstellung valider, den Anforderungen genügender AIPs. Dazu muss nach einer Bewertung der Daten eines TP gegebenenfalls auch deren Qualität verbessert werden, wodurch ein erster verändernder Eingriff nötig sein kann. Dem OAIS-Referenzmodell folgend können sich SIPs der Datengeber hinsichtlich des Formats und der ursprünglichen Systemumgebung stark voneinander unterscheiden. Das digitale Archiv muss mit dieser Divergenz bei der Übernahme umgehen, sie anschließend für die Einlagerung nach standardisierten Regeln umformen und letztlich gültige AIPs generieren. Die Dokumentation aller notwendigen Erschließungsinformationen sowie die erste Planung für Updates im Magazin und zur Organisation des Datenmanagements gehören dabei ebenfalls zur Übernahme.

Der Aufbau und die Organisation des digitalen Magazins werden dem nächsten Bereich zugerechnet, dem *Archivspeicher (archival storage)*. Der Aufgabenbereich umfasst hier die Übernahme und anschließende Einlagerung der AIPs, sowie deren Wartung und eine Gewährleistung für die Wiederauffindbarkeit der einzelnen Pakete. Zu diesem Zweck müssen einerseits die genutzten Datenträger des Archivs regelmäßig auf Lesbarkeit überprüft werden (redundante Mehrfachspeicherung ist ausdrücklich empfohlen), andererseits aber auch die Lesbarkeit der darauf gespeicherten AIPs sichergestellt werden (Obsoleszenz vermeiden). Für die Wiederauffindbarkeit muss eine magazinteknische Lagerungshierarchie aufgebaut werden.

Die *Datenverwaltung (data management)* hat zum Ziel, die Informationen zu den eingegangenen und archivierten Datenbeständen zu pflegen, zugänglich zu halten, kontinuierlich zu ergänzen und aufzubereiten. Ebenso fällt die Verwaltung und Wartung des Archiv-Management-Systems zur Sicherstellung der Lesbarkeit, Verständlichkeit und Nutzbarkeit der Daten in diesen Bereich.

Um die Organisation und Leitung der Beziehungen zwischen Archivaren und Nutzern sowie den Software- und Hardwaresystemen geht es im Aufgabenfeld der *Administration (administration)*. Unter Einbeziehung von Qualitätssicherungskriterien muss geregelt werden, welche Arbeiten des Systems automatisiert und damit maschinell erledigt werden können und wo vom Menschen durchgeführte Arbeitsschritte notwendig sind. Die Zuständigkeiten müssen dabei genau festgelegt werden. Außerdem sind hier die rechtlichen Aspekte (Verträge zur Übernahme, Archivierung und Nutzung) verortet. Zuletzt fallen auch die Prüfung der Informationspakete und der Betrieb der notwendigen technischen Infrastrukturen in diesen Bereich.

Die fünfte Einheit im Funktionskonzept wird als *Erhaltungsplanung (preservation planning)* bezeichnet. Darunter ist eine vorausschauende Planung und Reaktion auf künftige technische Entwicklungen zu verstehen, die u.a. die Empfehlung von Migrationszyklen für Updates und die Migration in besser geeignete Formate betrifft. Zudem müssen die Analyse des Nutzerverhaltens und der Anforderungen an das System durch die Nutzer berücksichtigt werden. Hierfür ist es notwendig, Strategien und Vorlagen für die *Information Packages* zu entwickeln.

Der letzte Bereich *Zugriff (access)* soll potentiellen Datennachnutzern den Zugang zu den

archivierten elektronischen Informationen unter Berücksichtigung der jeweiligen Zugriffsrechte ermöglichen und erleichtern. Für diesen Zweck werden DIPs generiert und für die Nutzer auf Anfrage und nach entsprechender Prüfung verteilt.

Insgesamt sollte neben diesen geschilderten Aufgaben auch das Digitale Magazin eines OAIS-konformen Archivs als gesamtes verwaltet werden. Darunter fällt die Softwareverwaltung, um das Archiv betreiben zu können und der Aufbau eines sicheren und stabilen Netzwerks. Wichtig ist die Gewährleistung der Sicherheit des OAIS insbesondere durch den Schutz vor unberechtigtem Zugriff von außen.

Die Umsetzung der funktionalen Anforderungen als konkrete Arbeitsschritte im Kontext von IANUS wird in **Kapitel 14** beschrieben.

6 Zentrale Aufgaben und Dienstleistungen

Im Folgenden werden die wichtigsten geplanten Dienstleistungen und Aufgaben von IANUS in knapper Form beschrieben. Diese werden getrennt dargestellt nach zentralen Aufgaben eines Kernbetriebes und möglichen weiteren Dienstleistungen eines erweiterten Betriebes, die je nach Erfolg, Finanzierung und Bedarf perspektivisch nach Etablierung von IANUS in Angriff genommen werden können oder sollen. Innerhalb dieser beiden Einteilungen wird nochmals nach technischen und nicht-technischen Dienstleistungen unterschieden.

Unter dem Begriff Kernbetrieb werden die Hauptkompetenzen von IANUS zusammengefasst. Zu diesen werden solche Leistungen gezählt, die nach dem OAI-Referenzmodell erforderlich sind, damit eine Einrichtung ihre primäre Zielsetzung als Langzeitarchiv für Forschungsdaten erfüllen kann. Dies beinhaltet alle technischen und nicht-technischen Arbeitsschritte und Systeme, mit denen Forschungsdaten für eine zukünftige Nachnutzung dokumentiert, migriert, archiviert und zur Verfügung gestellt werden. Dazu gehören aber auch flankierende Angebote, um die Bedeutung des Forschungsdatenmanagements und der Nachnutzung von Daten in der Fachcommunity zu verankern.

Unter dem Begriff erweiterter Betrieb werden dagegen Dienstleistungen zusammengefasst, die auf dem Kernbetrieb aufbauen und diesen um optionale Funktionen erweitern, aber im eigentlichen Sinne nicht kritisch für den Kernbetrieb selbst sind. Sie erzeugen einen Mehrwert, entweder bezogen auf die in IANUS verfügbaren Daten und Systeme oder bezogen auf die Nutzer, welche die Dienstleistungen von IANUS in Anspruch nehmen. Da die Realisierung dieser zusätzlichen Angebote wesentlich von den künftigen Ressourcen von IANUS abhängt, werden sie hier nur in knapper Form skizziert, um mögliche Entwicklungsperspektiven aufzuzeigen.

Alle folgenden Beschreibungen der Dienstleistung geben den Diskussionsstand zum aktuellen Zeitpunkt während der Konzeptionsphase von IANUS wieder. Je nach Entwicklung der finanziellen Ausstattung, technischen Systeme, rechtlichen Rahmenbedingungen und fachlichen Anforderungen sind langfristig Verschiebungen und Neubewertungen nicht auszuschließen. Als konstantes Prinzip gilt jedoch, dass IANUS seine Dienstleistungen als freiwillig nutzbare Angebote gestalten und von sich aus keine Verpflichtungen zu deren Nutzung formulieren wird. Unabhängig davon bleibt es Institutionen, Verbänden und Drittmittelgebern freigestellt, sich interne Verpflichtungen bzgl. IANUS aufzuerlegen bzw. die Nutzung einzelner Angebote zu empfehlen.

Die anvisierten Aufgaben von IANUS stehen dabei in einem offenen Verhältnis zu den Dienstleistungen, die von anderen Infrastrukturen und Projekten angeboten werden, um im nationalen Kontext einerseits Doppelentwicklungen zu vermeiden und andererseits Synergieeffekte nutzen zu können. Die meisten dieser Datenzentren, Repositorien, virtuellen Forschungsumgebungen und Informationssysteme sind dabei entweder in ihrer Zielsetzung generischer und allgemeiner, da sie häufig einen institutionellen Rahmen oder keine fachspezifische Ausrichtung besitzen und daher verschiedenartige Daten aus ganz unterschiedlichen Wissensdomänen verarbeiten. Oder aber sie sind spezifischer als IANUS, da sie nur konkrete Fragestellungen oder einzelne Quellengattungen adressieren. Aufgrund der institutionellen Einbindung in das DAI und der hohen fachlichen Relevanz ergeben sich besonders enge Verknüpfungen und Abstimmungen zu dessen zentralen Wissensressourcen, die im Rahmen von dessen iDAI.welt-Strategie entwickelt und angeboten werden.

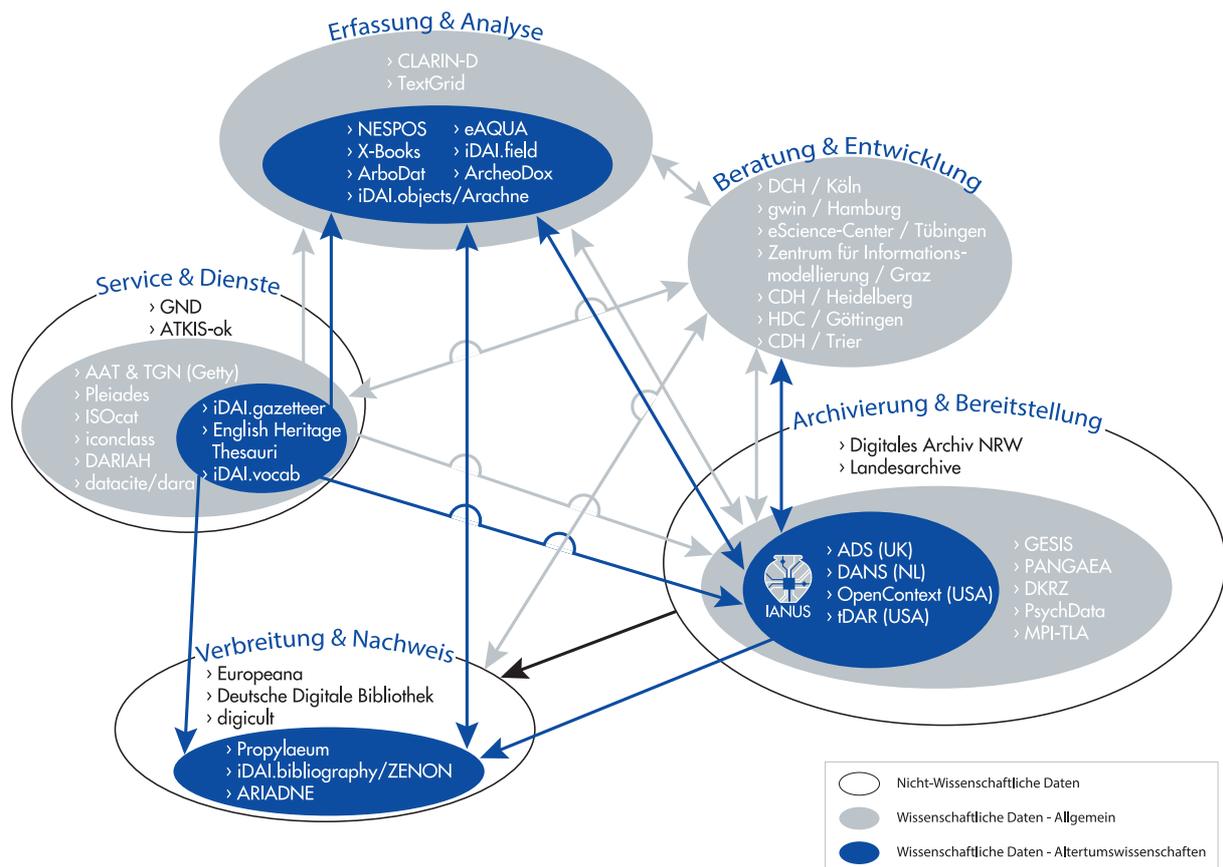


Abb. 5 Schematische funktionale Einbindung der Dienstleistungen von IANUS im Verhältnis zu anderen Infrastrukturen und Projekten

Von Bedeutung für IANUS sind ferner die geisteswissenschaftlichen Forschungs- und Datenzentren, die zur Zeit im Kontext der allgemeinen Digital-Humanities-Entwicklungen unter anderem vor allem an deutschen Universitäten entstehen (z. B. DCH in Köln, eScience-Center in Tübingen, gwin in Hamburg, GCDH und HDC in Göttingen). Gemeinsam ist ihnen, dass sie die Gesamtheit der Geisteswissenschaften, die an den jeweiligen Universitäten vertreten sind, adressieren und keinen fachspezifischen Fokus verfolgen. Ihre Relevanz in Bezug auf IANUS beruht vor allem darin, dass die jeweiligen Mitarbeiter mit den Wissenschaftlern und Projekten vor Ort in engem Kontakt stehen und diese direkt im Umgang mit digitalen Daten beraten und unterstützen können, insbesondere in der Phase der Datengenerierung und -analyse. Daher erscheint es mittelfristig für IANUS lohnend, individuelle Kooperationen und Kontakte zu etablieren, um die Aufgaben zu übernehmen, welche in den lokalen Zentren nicht geleistet werden können – in den meisten Fällen ist dies eine professionelle Langzeitarchivierung. Die möglichen Mehrwerte für IANUS ergeben sich daraus, dass die universitätsbasierten Zentren zum einen als kompetente Multiplikatoren für die Angebote fungieren können und zum anderen Daten, die langfristig nicht oder nicht nur in einer Universität archiviert werden können bzw. sollen, für eine Übergabe vorbereiten können.

6.1 Kernbetrieb

Die Kernaufgaben des Forschungsdatenzentrums IANUS werden die Kuratierung, die Langzeitarchivierung und die Verfügbarmachung von digitalen Forschungsdaten und wissenschaftlich relevanten Dokumenten sein. Diese Zielsetzung folgt ganz wesentlich auch den

Ergebnissen einer 2013 durchgeführten Stakeholder-Analyse in den Altertumswissenschaften. In dieser wurden die professionelle Langzeitarchivierung sowie die dauerhafte Online-Bereitstellung von Forschungsdaten als diejenigen Dienstleistungen genannt, welche die größte Relevanz und Notwendigkeit für die altertumswissenschaftliche Fachgemeinschaft besitzen.⁴⁸ Insbesondere die Langzeitnutzbarkeit von Daten stellt dabei eine besondere technische und organisatorische Herausforderung dar, da sich Dateiformate, Hardwareprodukte und Softwaresysteme in kurzen Zeitintervallen ändern und bestehende Lösungen immer wieder angepasst werden müssen.

6.1.1 Technische Dienstleistungen

1. Langfristige Kuratierung und Archivierung von statischen Forschungsdaten

Statische, weitgehend unveränderliche Daten aus Projekten und Institutionen können nach dem Ende einer Förderung oder nach Projektabschluss bei IANUS abgegeben werden, wo sie nach definierten Workflows und Standards formal überprüft, semantisch vervollständigt, technisch aufbereitet und ausführlich dokumentiert werden, bevor sie in ein Offline-Storage-System zur Bitstream-Preservierung überführt werden. Die Entscheidung und Auswahl, welche Daten fachlich relevant sind und daher archiviert werden sollen, liegt zunächst beim Datenproduzenten selbst. Aber auch IANUS wird überprüfen, ob eine Übernahme entsprechend der eigenen Sammlungsrichtlinie (*collections policy*) und gemäß der technisch-inhaltlichen Vorgaben erfolgen kann oder ob die Übergabe an ein anderes, besser geeignetes Datenzentrum sinnvoll ist. Die öffentlich zugängliche Sammlungsrichtlinie beschreibt dabei, nach welchen Kriterien IANUS Forschungsdaten annimmt oder ablehnt (siehe auch **Kapitel 4.3.4**), wie diese verwaltet und archiviert werden und auf welche Weise eine Weitergabe an Dritte erfolgen kann⁴⁹. Bei der Umsetzung der Strategie soll mittel- bis langfristig sowohl eine aktive als auch passive Akquise erfolgen, d.h. IANUS wird sowohl Datengeber, die sich auf Eigeninitiative bei IANUS melden, begrüßen als auch aktiv Eigentümer forschungsrelevanter Datensammlungen kontaktieren.

Die Archivierungsstrategie von IANUS orientiert sich an folgenden Leitlinien:

- Primäre Ziele sind die Aufbereitung und die Erhaltung von elektronisch vorliegenden, fachlich relevanten Informationen für eine künftige Nachnutzung, etwa in neuen Anwendungen oder für neue Fragestellungen. Um dies zu gewährleisten, steht die langfristige technische Interpretierbarkeit von Dateien und die semantische Nachvollziehbarkeit von Dateiinhalten im Vordergrund.
- Die Grundsätze der Erhaltungs- und Management-Strategien zielen vor allem darauf ab, für alle Daten, die IANUS anvertraut werden, die Authentizität, Funktionsfähigkeit und logische Integrität sicherzustellen.
- Für alle Informationspakete (SIPs, AIPs, DIPs) müssen sowohl Erhaltungsmetadaten, die Informationen zu Referenzen, Kontext, Provenienz, Persistenz und Zugriffsrechten⁵⁰ umfassen, als auch Erschließungsinformationen für das Suchen, Finden und Verstehen der

⁴⁸ IANUS-Stakeholderanalyse 2013. Teil 1, S. 18

⁴⁹ Vgl. die Collections Policy des Archaeology Data Service, 6th Edition vom 01. April 2014: <http://archaeologydataservice.ac.uk/advice/collectionsPolicy>

⁵⁰ Zu diesen Informationen als Teil des Preservation Description Information (PDI) ausführlich OAIS-de (2012) S. 59 f.

digitalen Objekte erhoben und verwaltet werden.

- Zu jedem Zeitpunkt muss eine ausreichende Dokumentation vorhanden sein, um kontinuierlich Erhaltungsmaßnahmen durchführen und technisch aktuelle Informationspakete bereitstellen zu können.
- Da das primäre Ziel der Archivierung der Erhalt von einmaligen Fachinhalten und nicht der Erhalt des Verhaltens der ursprünglich verwendeten Softwaresysteme ist, wird der Verlust von Funktionalitäten oder Darstellungsformen (z. B. bei Datenbanken, GIS-Systemen, 3D-Daten, Webapplikationen) durch die Erhaltungsmaßnahmen bewusst in Kauf genommen. Ausnahmen von dieser Maßgabe sind nur in Einzelfällen möglich, wenn eine hohe fachliche Relevanz vorliegt, der technische Aufwand vertretbar ist und die Arbeitsaufwände adäquat finanziert werden. Ggf. muss das ursprüngliche Aussehen und die Benutzung zum Zeitpunkt der Datenerzeugung durch statische Repräsentationen (z. B. Screenshots, PDF-Exporte o.ä.) und verbale Beschreibung dokumentiert werden.
- Auf alle Fälle werden die vom Datenproduzenten anvertrauten Originaldaten als validierte SIPs unverändert aufbewahrt, um ggf. Fehler bei der primären Kuratierung und bei Formatmigrationen nachvollziehen und beheben zu können.

Zur Erreichung der Erhaltungsziele kommen vor allem folgende Methoden zum Einsatz:

- Normalisierung: Umwandlung (Konvertierung) von Dateien in offene internationale Standardformate (z. B. ASCII-Texte, TIFF)
- Versions-Migration: Dateien werden sukzessive und kontinuierlich in neuere Versionen von Dateiformaten konvertiert. Dies gilt insbesondere dann, wenn Dateien in proprietären Dateiformaten vorliegen, die nicht oder nur mit hohem Verlust in offene Dateiformate umgewandelt werden können, und das proprietäre Dateiformat eine besondere Relevanz in der Fach-Community besitzt.
- Format-Migration: Neben der Normalisierung können weitere Format-Migrationen notwendig sein, z. B. für spezielle Formate für die Bereitstellung.
- Hardware-Erneuerung: Übertragung von Dateien von einem Speichermedium auf ein anderes ohne Veränderung der Bit-Reihenfolge (z. B. bei veralteten Backup-Systemen)

Dateien, die nicht normalisiert und/oder migriert werden können, sind für eine Langzeitarchivierung nach dem Verständnis von IANUS nicht geeignet. Auch ist eine funktionale Archivierung technischer Systeme und Anwendungen mittel Emulation zum gegenwärtigen Zeitpunkt nicht vorgesehen. Zur Umsetzung der o.g. Methoden werden für die akzeptierten und präferierten Datenformate (siehe **Kapitel 14.1.1**) sogenannte erhaltenswerte bzw. signifikante Eigenschaften festgelegt.⁵¹ Diese Festlegungen betreffen den Inhalt, den Kontext, die Erscheinung, die Struktur und das Verhalten von digitalen Objekten und hängen wesentlich von den Nutzungsszenarien der angestrebten Zielgruppe ab. Sie müssen regelmäßig überprüft und aktualisiert werden (z. B. hinsichtlich neu entwickelter und besser geeigneter Dateiformate). Da

⁵¹ Ausführlich zu diesem Thema Nestor-AG „Digitale Bestandserhaltung“ (Hrsg.), Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung, Version 2.0, nestor-materialien 15 (2012): <http://nbn-resolving.de/urn:nbn:de:0008-2012092400> (Zugriff am 30.06.2015). Relevant ist auch das Projekt „Inspect. Investigating significant properties of electronic content“: <http://www.significantproperties.org.uk/> (Zugriff am 20.08.2015)

diese Eigenschaften die Erzeugung bzw. Erneuerung sowohl von Archivierungspaketen (AIPs) als auch von Bereitstellungspaketen (DIPs) betreffen, lassen sich von ihnen die jeweilige Erhaltungsstrategie (das sog. *preservation planning*) ableiten.

Praktisch kommt das Konzept der signifikanten Eigenschaften bei IANUS an zwei Punkten zum Tragen. Zum einen müssen bei der Erstellung eines validen SIP bzw. vor dem Ingest von Archivdaten zusammen mit dem Datenproduzenten individuell für jede Datensammlung bzw. Datei(typ) signifikante Eigenschaften festgelegt werden⁵². Die Entscheidungen (z. B. ob bei Zeichendateien die Struktur einzelner Ebenen erhalten werden soll oder nicht oder ob bei Textdokumenten das Layout wichtig für das Verständnis des Inhaltes ist oder nicht) und deren Umsetzungen müssen dabei nachvollziehbar dokumentiert werden. Zum anderen müssen für einzelne Dateitypen signifikante Eigenschaften definiert werden, damit diese im Bedarfsfall automatisiert in neuere Zielformate migriert werden können und der Erfolg der Migration objektiv überprüft werden kann, z. B. weil eine Tabelle vor und nach einer Dateimigration die gleiche Anzahl an Spalten und Zeilen aufweist.

Das Angebot der Archivierung von Daten erfolgt nicht-exklusiv, d.h. Datenproduzenten können Ihre Daten z. B. aufgrund rechtlicher Verpflichtungen parallel auch an anderen Stellen (Anbietern, Portalen etc.) archivieren und veröffentlichen. Sofern erwünscht kann ein Datenproduzenten eine Kopie der kuratierten und validierten Informationspakete (SIP, AIP) erhalten, um dieses in eigenen Systemen und alternativen Einrichtungen vorzuhalten. Eine Gewährleistung für die langfristige Lesbarkeit und Funktionalität von Daten übernimmt IANUS allerdings nur für diejenigen Informationspakete, die sich im eigenen Archivbestand befinden, nicht aber für diese Kopien, die an Dritte übergeben werden.

IANUS kann keine Garantien dafür geben, wie lange die ihm anvertrauten digitalen Objekte überleben werden. Es verspricht aber, nach bestem Wissen, unter Berücksichtigung des aktuellsten Standes der Technik und mit anerkannten, offenen und transparenten Strategien sich darum zu bemühen, dass die Daten - solange IANUS existiert - erhalten, gefunden, gelesen und nachgenutzt werden können.

2. Online-Bereitstellung von archivierten Forschungsdaten

Das primäre Ziel der Kuratierung und Archivierung von Forschungsdaten ist deren Nachnutzung durch neue Projekte und Personen. Insofern ist der langfristige Zugang (*access*) zu diesen Daten eine zentrale Aufgabe von IANUS. Dabei besteht der Anspruch, den designierten Zielgruppen möglichst qualitätsvolle und verlässliche digitale Ressourcen zum Zweck der Forschung und Lehre anzubieten und den Zugriff möglichst niedrigschwellig zu gestalten. Auch Schulen, Bildungseinrichtungen und die breite Öffentlichkeit werden mit diesem Service der Bereitstellung adressiert.

Bei der Bereitstellung von archivierten Forschungsdaten ist davon auszugehen, dass deren Akzeptanz in dem Maße steigt, je weniger das Ausgangssystem des Nutzers beim Zugriff eine limitierende Rolle spielt (z. B. aufgrund des Betriebssystems oder fehlender Zusatzsoftware) und je

⁵² THALLER (2013) S. 32–36.

benutzerfreundlicher die Angebote – insbesondere zum Suchen und Auffinden der digitalen Inhalte – ausgestaltet sind⁵³. Das System von IANUS wird also so konzipiert und realisiert, dass der Zugriff auf die bereitgestellten Daten durch die Nutzer möglichst unabhängig von deren individuellen IT-Systemen erfolgen kann. Aufgrund dieser Anforderungen ist für die Bereitstellung vor allem ein Online-Datenportal vorgesehen, das als einfachste Zugriffsform den direkten Download von Dateien anbietet. Außerdem lassen sich dort entsprechende Plugins und Viewer integrieren, die für die häufigsten Standard-Dateitypen (z. B. Texte, Rastergrafiken, Dokumente etc.) eine Vorschau-Darstellung unmittelbar im Browser ermöglichen. Das Angebot muss also kompatibel zu den gängigsten Browsertypen entwickelt und bei der Benutzung in die Homepage von IANUS integriert werden.

Bei den Daten, die über ein solches Online-Datenportal weitergegeben werden, handelt es sich nicht um die AIPs selbst, da diese ohne unmittelbaren Zugriff in einem *deep storage* bei externen Partnern archiviert werden (siehe **Kapitel 7.2.1**) und ein Zugriff auf diese Datenbestände nur in fest definierten Ausnahmefällen vorgesehen ist. Stattdessen werden im Rahmen der Datenkuratierung entsprechende DIPs vorbereitet (siehe **Kapitel 5.1 und 13.4**) – je nach Bedarf für unterschiedliche Nutzergruppen mit unterschiedlichen Nutzungs- bzw. Zugriffsrechten unterschiedliche Varianten – und über die IANUS-eigene Server online vorgehalten.

Je nach Art der Daten, individuellen Wünschen von Datenproduzenten und Beteiligungen an Personalkosten sind auch aufwändigere Formen der Bereitstellungen jenseits eines einfachen Datei-Downloads über das Online-Datenportal vorstellbar. So ist denkbar, dass Datenbanken oder strukturierte Inhalte über spezielle Formulare direkt online durchsucht, mehrere Abfragen mit projektspezifischen Wertelisten kombiniert oder einfache GIS-Oberflächen mit Layern und interaktiven Karten dargestellt werden können. Dabei geht es nicht darum, ursprüngliche Erfassungssysteme nachzubauen, sondern in standardisierter, langfristig pflegbarer und kostengünstiger Weise einfache Zusatzfunktionalitäten anzubieten. Eine andere, vom Standardfall abweichende Variante, die von einem Datennachnutzer explizit beantragt werden müsste, wäre die Bereitstellung besonders großer Datenmengen auf speziellen FTP- oder Cloud-Services oder auf postalischem Wege mittels externer Datenträger.

Im derzeitigen Planungsstadium ist die Unterstützung mobiler Endgeräte nicht unmittelbar vorgesehen. IANUS wird jedoch darauf achten, seine Präsentationsstrategie nach möglichst generischen Lösungen auszurichten⁵⁴, um diese Plattformen möglichst bruchlos unterstützen zu können.

Auch wenn IANUS die Ideen des Open Access und des Open Data begrüßt und aktiv unterstützt, ist es sich bewusst, dass in bestimmten begründeten Situationen, insbesondere bei personenbezogenen Daten, Geoinformationen zu gefährdeten Denkmälern und nicht abgeschlossenen Projektdaten, Schutzmechanismen notwendig sind. Für die Bereitstellung der anvertrauten und archivierten Forschungsdaten über das IANUS-Portal können daher Zugriffsbeschränkungen festgelegt werden, so dass z. B. im Bedarfsfall eine persönliche Anmeldung erforderlich ist, bevor eine Datei angesehen und/oder lokal gespeichert werden kann. Mit Hilfe eines ausdifferenzierten Rechte-Rollen-Konzeptes werden die Vorgaben der Dateneigentümer, welche Personenkreise in welcher

⁵³ IANUS-SCHWIEGELSHOHN (2013) S. 13.

⁵⁴ d.h. nach dem aktuellen Stand der Technik HTML5.

Granularität auf die bereitgestellten Datensammlungen und Einzeldateien zugreifen und sie nachnutzen dürfen, festgehalten, umgesetzt und kontrolliert. Die Zugriffsmöglichkeiten auf die Dateien können dabei individuell gesteuert werden, so dass die Sichtbarkeit von ganz freien Publikationen bis hin zu Datensammlungen, die zeitlich befristet unzugänglich und geschützt bleiben sollen (sog. Embargo), variieren kann (ausführlich siehe **Kapitel 10.2**).

Damit Nutzer relevante Forschungsdaten finden und auswerten können, sind Erschließungsinformationen erforderlich, die in Form von strukturierten und standardisierten Metadaten gemeinsam durch die Datenproduzenten und IANUS erhoben werden. Während für Datensammlungen als Ganzes verpflichtende Attribute festgelegt werden (siehe **Kapitel 12.4**), werden für einzelne Dateiformate nur wünschenswerte Attribute definiert, die je nach Beschaffenheit der Daten und manuellem bzw. technischem Aufwand bereitgestellt werden können. Insgesamt wird angestrebt, die Metadaten so detailliert, so vollständig und – mit Hilfe kontrollierter Vokabulare – so einheitlich wie möglich zu erfassen, damit Nutzer über allgemeine und spezialisierte Suchmasken Inhalte finden, Ergebnismengen filtern und mittels sog. facettiertem Browsing erschließen können. Um auch internationale Nutzer zu erreichen und einen internationalen Datenaustausch zu unterstützen, werden die Metadaten zu einer Datensammlung grundsätzlich nicht nur in deutscher, sondern auch in englischer Sprache gespeichert. Sofern eine Datensammlung zu einem im Ausland durchgeführten Projekt gehört, werden die Metadaten zusätzlich in der relevanten Landessprache erfasst, damit sie auch dort wahrgenommen werden können.

Wie bei den Dateien selbst können auch bei einzelnen Metadaten aus rechtlichen Schutzgründen Nutzungseinschränkungen vorliegen (insbesondere sind hier sensible Geoinformationen zu berücksichtigen). Für eine spätere Phase ist auch die Indizierung und automatisierte Erschließung der eigentlichen Dateiinhalte (z. B. mittels Data-/Text-Mining-Verfahren) vorgesehen, um eine noch tiefergehende Recherche zu ermöglichen. Zentral wird bei der Präsentation aller Suchergebnisse sein, dass neben technischen und fachlichen Informationen die eigentlichen Datenproduzenten und Rechteinhaber deutlich erkennbar sind.

Zur eindeutigen Referenzierbarkeit von digitalen Ressourcen in IANUS, werden persistente Identifikatoren (PIDs)⁵⁵ vergeben, für ganze Datensammlungen automatisiert im Rahmen der Aufbereitung durch IANUS, für einzelne Dateien auf individuelle Anfrage durch Datennachnutzer (siehe ausführliche **Kapitel 7.4**).

3. Zentraler Nachweiskatalog archäologischer Fachinformationen

Vorgesehen ist ein Webportal, in dem nicht nur Datensammlungen, die im Archivbestand von IANUS enthalten sind, angezeigt werden, sondern auch solche Informationen, die bei unterschiedlichen Institutionen und in heterogenen Systemen in Deutschland verteilt sind. Die Idee ist daher, an einer zentralen Stelle so viele Angaben wie möglich über archäologische

⁵⁵ PIDs können als ISBN-Nummern für digitale Ressourcen im Internet verstanden werden, ausführlicher K. Schroeder, Persistent Identifier (PI) – ein Überblick, in: H. Neuroth u. a. (Hrsg.), nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3 (2010) Kap.9:22–45 (http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf [Zugriff am 2.12.2013])

Fundstellen, Denkmäler, wissenschaftliche Projekte und Akteure sowie relevante Dokumentensammlungen (analoge und digitale) zu bündeln und nachzuweisen.⁵⁶ Ähnlich wie bei anderen Onlineangeboten, bei denen Wissen digital aggregiert wird – wie im Bibliotheksbereich die Verbundkataloge und virtuellen Fachbibliotheken (insbesondere Propylaeum⁵⁷), für das materielle kulturelle Erbe DigiCult⁵⁸, die Deutsche Digitale Bibliothek⁵⁹ und Europeana⁶⁰ sowie für Datensammlungen die Collection Registry von DARIAH-DE⁶¹ und re3data⁶² – steht auch im Nachweiskatalog von IANUS die Zusammenführung, Integration und Anreicherung von wenigen fachspezifischen, aber heterogen modellierten Metadaten im Vordergrund. Durch das Angebot und die Pflege eines einzigen, durchsuchbaren Web-Interfaces soll die Recherche nach fachspezifischen Inhalten, die nicht Teil des Datenarchives von IANUS sind, für Wissenschaftler erleichtert werden. Fachliches und technisches Vorbild für den Nachweiskatalog von IANUS sind verschiedene nationale Portale, in denen Metadaten zu Monumenten und Fundstellen zusammengeführt werden:

- ArchSearch⁶³ vom Archaeology Data Service (ADS), York (UK)
- Riksantikvarieämbetet - Fornsök⁶⁴ des Swedish Heritage National Board, Stockholm (SE)
- Fund og fortidsminder⁶⁵ der Danish Agency for Culture, Kopenhagen (DK)

Zentrale Eigenschaften sind dabei, dass die Einzelnachweise für alle Ressourcen, die nicht auch bei IANUS archiviert werden, nur als Metadatensatz vorgehalten werden, dass explizit auf die eigentlichen Dateneigentümer hingewiesen wird und dass direkte Verlinkungen auf die Ursprungssysteme (sofern diese öffentlich erreichbar sind) erfolgen. Die jeweiligen Dateneigentümer z. B. Landesdenkmalämter, Museen, Archiven oder andere wissenschaftliche Einrichtungen, behalten dabei alle Rechte an und die Zuständigkeit zur Bereitstellung und Pflege der Originaldaten. Sie sind auch für die Qualität und Aktualität der Metadaten, die im Nachweiskatalog angezeigt werden, verantwortlich.

Die Art der Zugänglichkeit zu den nachgewiesenen Informationen kann dabei variieren:

- Bei nur lokal bzw. offline vorliegenden analogen und/oder digitalen Daten werden die besitzende Einrichtung, ein Aufbewahrungsort mit Anschrift und eine zuständige Kontaktperson genannt.
- Bei Daten, die entweder über IANUS selbst oder über externe Systeme online verfügbar sind, erfolgt zusätzlich zu den o.g. Angaben entweder eine Weiterleitung mittels Referenz (z. B. URL, PID) auf einen zugehörigen Internetauftritt oder ein direkter Zugriff auf eine externe digitale Ressource.

⁵⁶ IANUS - Kolbmann (2014).

⁵⁷ <https://www.propylaeum.de/>

⁵⁸ <http://www.digicult-verbund.de/>

⁵⁹ <https://www.deutsche-digitale-bibliothek.de/>

⁶⁰ <http://www.europeana.eu/>

⁶¹ <http://colreg.de.dariah.eu/>

⁶² <http://www.re3data.org>

⁶³ <http://archaeologydataservice.ac.uk/archsearch/>

⁶⁴ <http://www.fmis.raa.se>

⁶⁵ <http://www.kulturarv.dk/fundogfortidsminder/>

Die zentralen Kategorien, die das Datenmodell des Nachweiskataloges und damit auch dessen Benutzeroberfläche bestimmen, sind:

- **Archäologische Fundstellen** bzw. **Archäologieflächen** wie sie im ADEX-Standard definiert werden: "Eine Archäologiefläche ist eine Fläche, d.h. ein im Gelände abgrenzbarer oder lokalisierbarer Bereich, an der mindestens ein archäologisch qualifiziertes bzw. relevantes Ergebnis vorliegt/vorgelegen hat oder vermutet wird. [...] Eine Archäologiefläche umfasst idealerweise den gesamten Bereich, für den archäologische Informationen vorliegen, unabhängig von der heutigen Erhaltung."⁶⁶ Der Vorteil des ADEX-Standards liegt einerseits darin, dass er von allen Landesdenkmalämtern unterstützt wird und Daten in diesem Format erzeugt werden können. Andererseits ist er bewusst so einfach modelliert, dass er auch in anderen Institutionen und Bereichen leicht anwendbar ist. Zu Archäologieflächen gehören im weitesten Sinne auch Monumente und Bauwerke.
- **Aktivitäten** wie notwendige Maßnahmen, wissenschaftliche Untersuchungen und Forschungsprojekte, für die eindeutig zuweisbare Akteure, eine feste Zeitstellung (Anfang und Ende einer Aktivität) und eine spezifische Aufgaben- bzw. Fragestellung bekannt sind. Im Bereich der Archäologie besitzen Aktivitäten häufig (aber nicht zwingend) einen geographischen Raumbezug, in dem eine Maßnahme stattfindet oder der konkret untersucht wird. Hierfür wurde im ADEX-Standard das Konzept der Untersuchungsfläche definiert: Eine Untersuchungsfläche umschließt den Bereich, der archäologisch untersucht wurde. Sie sind ergebnisneutral bzw. archäologieneutral. Sofern archäologische Ergebnisse in einer Untersuchungsfläche beobachtet werden, werden diese als Archäologiefläche geführt, so dass also normalerweise eine oder mehrere Archäologieflächen zu einer Untersuchungsfläche Bezug nehmen (und umgekehrt).
- **Datensammlungen** können analoge Dokumente in Archiven (z. B. Grabungsunterlagen) und/oder digitale Ressourcen jeglicher Art umfassen, die im Rahmen von Projekten entstanden sind, aber auch im Rahmen regulärer Tätigkeiten aufgebaut werden. Wichtig ist, dass eine Dokumentensammlung immer mit einem Akteur als Eigentümer oder Ansprechperson verbunden sein muss. Über diese Kategorie werden auch die digitalen Bestände von IANUS nachgewiesen und auffindbar gemacht.
- **Akteure**, also Institutionen, Gruppen oder Personen, die Aktivitäten und Projekte durchführen, für Archäologieflächen verantwortlich sind und Datensammlungen besitzen oder erzeugen.

Alle Kategorien können dabei mehrfach untereinander miteinander verbunden sein, z. B. dass an einer Fundstelle mehrere Aktivitäten durchgeführt wurden, die wiederum mehrere Dokumentsammlungen hervorgebracht haben, oder eine Institution an mehreren Fundstellen tätig ist oder eine thematisch zusammenhängende Datensammlung auf verschiedene Institutionen verteilt ist. (ausführlich zu den Metadatenmodellen **Kapitel 12.4 und 12.5**)

Um eine möglichst konsistente und hohe Qualität der Metadaten zu erhalten, sollen so viel wie möglich relevante externe Referenzsysteme eingebunden werden. (siehe **Kapitel 12.3**) Hierfür

⁶⁶ http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_ADeX/ADeX_2-0_Doku.pdf

müssen entsprechende Schnittstellen implementiert und Möglichkeiten zur Referenzierung im Sinne von Linked Open Data vorgesehen werden.

Grundsätzlich wird das Ziel verfolgt, die in dem Nachweiskatalog zusammengeführten basalen Metadaten möglichst frei und ohne Einschränkungen zugänglich zu machen. Dennoch existieren auch sensible und schützenswürdige Informationen (Angaben zu Einzelpersonen aufgrund des persönlichen Datenschutzes; Geoinformationen zu Archäologie- und Untersuchungsflächen aus Schutz vor Raubgräbern und mutwilliger Zerstörung), für die gestaffelte Zugriffsrechte erforderlich sind. Für bestimmte Nutzergruppen müssen einzelne Metadatenattribute ein- oder ausgeblendet und die Genauigkeit der Darstellung von Geoinformationen angepasst werden können. So ist z. B. denkbar, dass für anonyme, nicht angemeldete Benutzer nur ungefähre Koordinaten angezeigt und detaillierte Ortsbezeichnungen ausgeblendet werden, während z. B. Benutzer, die durch eine Anmeldung als Mitarbeiter einer bestimmten Einrichtung authentifiziert sind, exakte Lokalisierungen betrachten dürfen. Welche Rollen existieren sollen und welche Rechte jeweils damit verbunden sein werden, ist mit den Bereitstellern der Daten zu definieren.

Die Art und Weise, wie die Metadaten von Dateneigentümern technisch an IANUS übermittelt werden, kann dabei sehr unterschiedlich ausfallen und hängt von den jeweiligen Systemen ab. Denkbare Szenarien sind:

- eine direkte Anbindung an die Systeme, insbesondere der iDAI.welt, von IANUS zur Erfassung von Akteuren und Beschreibung von Datensammlungen
- Übertragung mittels Web-Services und Schnittstellen, die z. B. im Rahmen der Geodateninfrastrukturen der einzelnen Bundesländer angeboten oder von großen Institutionen (z. B. den Akademien) zum Metadata-Harvesting bereitgestellt werden
- ein- oder mehrmalige (regelmäßige) Importe von offline zugesandten Dateien (z. B. dateibasierte Exporte aus Fachsystemen)

Die Funktionalitäten des Nachweiskataloges sollen bewusst einfach gehalten sein, um eine klare Benutzeroberfläche zu ermöglichen, und sich auf folgende Aspekte beschränken:

- Authentifizierung
- Verschiedene Suchmöglichkeiten über Eingabe von Zeichenketten, Facetted-Browsing und Kartenbasiert, d.h. einfache Funktionalitäten wie Zoom und Auswahl-Polygon
- Anzeige von Suchergebnissen mit Metadaten, Kartendarstellung und Verlinkungen zu externen Systemen für weitere Informationen
- Export von Ergebnissen als Listen oder ggf. als Karten zur lokalen Weiternutzung

4. Bereitstellung von Metadaten via (Web-)Services

Neben einer benutzerfreundlichen Darstellung der von IANUS gesammelten Fachinformationen in einem browserbasierten Nachweiskatalog werden diese als vierte technische Dienstleistung des Kernbetriebes auch maschinenlesbar über SOAP-basierte Web-Services und RESTful-interfaces zur Einbindung in die Portale Dritter zur Verfügung gestellt. Zu einem späteren Zeitpunkt ist auch die Entwicklung von entsprechenden Apps denkbar. Die Metadaten werden dafür soweit harmonisiert

und aufeinander gemappt, dass sie entsprechend maßgeblicher Standards, insbesondere CARARE⁶⁷ und EDM⁶⁸, weitergegeben und geharvestet werden können. Gleichzeitig sollen die Daten des Nachweiskataloges als Linked Open Data zur Verbreitung im Semantic Web zur Verfügung gestellt werden. Langfristig sollen insbesondere folgende Services und Portale über OAI-PMH Schnittstellen entweder direkt von IANUS oder indirekt über die vernetzten Services der iDAI.welt bedient werden:

- für digitale Ressourcen: Deutsche Digitale Bibliothek⁶⁹, Europeana⁷⁰, b2find.EuDat⁷¹
- für Datensammlungen: DARIAH-Collection Registry⁷² und re3data⁷³
- für antike Objekte: iDAI.objects (= Arachne)⁷⁴, DigiCult⁷⁵, ARIADNE⁷⁶
- für bibliographische Nachweise: iDAI.bibliography (= Zenon)⁷⁷, Propylaeum⁷⁸, Bibliotheksverbände

6.1.2 Nicht-technische Dienstleistungen

1. Publikation von fachspezifischen IT-Empfehlungen zum Umgang mit Forschungsdaten

Neben der fachlichen Kuratierung und der technischen Archivierung von digitalen Ressourcen stellt die Informierung und Beratung einer Fach-Community einen wichtigen Aspekt eines ganzheitlichen Forschungsdatenzentrums dar. Eine von IANUS durchgeführte Umfrage zeigt, dass es in den Altertumswissenschaften einen hohen Bedarf an fachspezifischen Empfehlungen, Mindeststandards, Best-Practice-Examples und praktischen Hinweisen rund um das Thema des nachhaltigen Managements von Forschungsdaten gibt⁷⁹.

Daher wurde begonnen, eine frei zugängliche Online-Plattform "IT-Empfehlungen zum nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften" aufzubauen⁸⁰. Die einzelnen Kapitel adressieren vor allem Aspekte, die für die langfristige Archivierung, Bereitstellung, Nutzbarkeit und Interoperabilität von digitalen Daten von zentraler Bedeutung sind, und fokussieren daher auf technische Formate und die inhaltliche Beschreibung von Daten und Methoden. So behandeln sie z. B. Fragen, wie digitale Daten für die Zukunft langfristig aufbewahrt werden sollten; was bei der Benennung, Ablage und Speicherung von Dateien zu beachten ist; welche Metadaten zu Dateien zum Verständnis für Dritte erforderlich sind; auf welche Weise eine

⁶⁷ <http://www.carare.eu/eng/Resources/CARARE-Documentation>

⁶⁸ <http://pro.europeana.eu/edm-documentation>

⁶⁹ <https://www.deutsche-digitale-bibliothek.de/>

⁷⁰ <http://www.europeana.eu/>

⁷¹ <http://b2finde.eudat.eu>

⁷² <http://colreg.de.dariah.eu/colreg/colreg>

⁷³ <http://www.re3data.org>

⁷⁴ <http://www.arachne.uni-koeln.de>

⁷⁵ <http://www.digicult-verbund.de>

⁷⁶ <http://http://www.ariadne-infrastructure.eu>

⁷⁷ <http://zenon.dainst.org/>

⁷⁸ <https://www.propylaeum.de/>

⁷⁹ Siehe die entsprechenden Ergebnisse in der IANUS - STAKEHOLDERANALYSE (2014).

⁸⁰ <http://www.ianus-fdz.de/it-empfehlungen>

bessere Austauschbarkeit erzielt werden kann; oder welche Vorgaben zur Archivierung von Daten (insbesondere bei IANUS) einzuhalten sind. Spezifische Vorschriften zu einzelnen fachlichen Methoden, Fragestellungen, Workflows oder Anwendungen sind dagegen nicht Gegenstand der Empfehlungen. Jeder Abschnitt gliedert sich in eine knapp gehaltene Übersicht mit Empfehlungen, vertiefende Hintergrundinformationen, praktische Tipps und Tools sowie Referenzen zu weiterführenden Quellen.

Stellvertretend für die o.g. Fach-Communities übernimmt IANUS im Kontext der IT-Empfehlungen dabei vier Aufgaben:

1. Implementierung und langfristige Pflege einer technischen Plattform, um die Inhalte zentral und online präsentieren zu können.
2. Koordination und Redaktion der Inhalte und Empfehlungen, die im Wesentlichen von anerkannten Spezialisten, externen Arbeitsgruppen oder ganzen Fachverbänden formuliert werden sollen; nur zu einem geringen Teil wird IANUS selbst Texte verfassen.
3. Kontinuierliche Aktualisierung unter Berücksichtigung neuester Entwicklungen und des Feedbacks von Nutzern, wiederum mit Einbindung externer Fachleute.
4. Öffentlichkeitsarbeit, um die Notwendigkeiten und Möglichkeiten eines nachhaltigen Forschungsdatenmanagements in den Altertumswissenschaften bekannt zu machen.

Ein zusätzliches Ziel ist es, durch die IT-Empfehlungen und dort genannten Mindeststandards die Anwendung der von der DFG herausgegebenen Regeln zur guten wissenschaftlichen Praxis⁸¹ zu vereinfachen und aufwendig erhobene Forschungsdaten für die Zukunft zu erhalten. Es besteht die Hoffnung, dass bei Beachtung und Einhaltung der IT-Empfehlungen nachfolgende Generationen die heutigen, individuellen Erkenntnis- und Datenentstehungsprozesse nachvollziehen können und so einem drohenden Verlust von einmaligen Informationen vorgebeugt werden kann.

2. Angebote zur Aus- und Weiterbildung im Bereich Forschungsdatenmanagement und Digitale Altertumswissenschaften

Da ein grundlegendes Verständnis und Bewusstsein für die Relevanz der Pflege von Forschungsdaten bei vielen Altertumswissenschaftlern nicht vorhanden ist und häufig erst geschaffen werden muss, stellt das Thema "Ausbildung und Qualifizierung" eine weitere zentrale Kernaufgabe für IANUS dar⁸². Die Erfahrungen etablierter Forschungsdatenzentren und Archive zeigen, dass der Betrieb einer technischen Infrastruktur, die Kuratierung von Dateien und die Bereitstellung von Online-Datenportalen als Maßnahmen alleine nicht ausreichen, um Forschungsdaten im Sinne des Data-Life-Cycles zu erhalten und ihre Nachnutzbarkeit zu gewährleisten. Ein zentraler und in der Regel einschränkender Faktor sind die Kenntnisse und Erfahrungen der Datenproduzenten selbst, deren Verhalten und Arbeitsweisen maßgeblich über die Zukunft von Daten entscheiden. Hier sind im Bereich der Altertumswissenschaften große Defizite zu konstatieren, da nur in wenigen (universitären) Studiengängen die relevanten

⁸¹ http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf

⁸² Siehe dazu auch das Arbeitspakt 4 der ersten Phase von IANUS, in der auch eine eigene Arbeitsgruppe etabliert wurde: <http://www.ianus-fdz.de/projects/ap4-ausbildung>

Grundkenntnisse zum nachhaltigen Umgang mit digitalen Daten vermittelt werden.

Für die notwendige Aufklärungsarbeit dienen einerseits die oben genannten IT-Empfehlungen, die allerdings eher den Charakter eines Nachschlagewerkes ohne didaktischen Anspruch besitzen. Darüber hinaus will IANUS allgemeine Basisinformationen zu theoretischen und praktischen Aspekten bewusst einfach und in kleinen, aufeinander aufbauenden Themeneinheiten aufbereiten, so dass sie auch von Wissenschaftlern und Studierenden der Altertumskunde ohne spezielle IT-Kenntnisse verstanden werden können. Als mögliche Präsentationsformen sollen dabei beispielsweise anwendungsbezogenen Praxisratgeber, Best-Practice-Beispiele, Vorlagen/Templates, Online-Tutorials, eLearning-Module etc. entwickelt und online zur Verfügung gestellt werden, die sowohl zum Selbststudium als auch als Hilfestellung und Vorlagen für Dozenten geeignet sein werden. Idealerweise können zusammen mit den Materialien auch geeignete Testdaten bereitgestellt werden, um die erworbenen Kenntnisse praktisch anwenden zu können. Für Wissenschaftler und Studierende mit IT-Erfahrungen sind auch Vertiefungsmodule denkbar, die bestimmte Probleme und Methoden als spezialisierte Lehreinheiten adressieren.

Neben der Erstellung von neuen Dokumenten durch IANUS selbst sollen so weit wie möglich bereits existierende Präsentationen, Vorträge u. a., die von anderen Institutionen oder Projekten sowie von externen Dozenten für Kurse an einzelnen Universitäten erstellt worden sind, gesammelt und ggf. nach einer individuellen Anpassung (z. B. aufgrund von Bildrechten oder zur Einbindung fachtypischer Beispiele) einem größeren Kreis zur Verfügung gestellt werden.

Parallel zu diesen digitalen Formaten sind auch Veranstaltungen wie Summer Schools, Workshops oder Vorträge geplant, um einen direkten, kritischen Austausch zwischen den Nutzern der Materialien, den Autoren bzw. Referenten und IANUS zu ermöglichen. Diese sollen zusammen mit externen Spezialisten als Vortragende und Institutionen als Mitorganisatoren durchgeführt werden. Im Einzelfall ist zu klären, inwieweit auf andere Aktivitäten aufgebaut bzw. mit diesen kooperiert werden kann⁸³ und diese für die Bedürfnisse der IANUS-spezifischen Fach-Community geeignet sind. Insbesondere bei der Vermittlung von Kenntnissen im Bereich des Forschungsdatenmanagements, der Langzeitarchivierung und der Weitergabe von Daten sind viele existierende Angebote stärker allgemein, generisch oder theorieorientiert ausgerichtet und sollten für eine praxisnahe Anwendung in den Altertumswissenschaften aufbereitet werden.

Und schließlich soll eine Bestandsaufnahme über bestehende e-Learning-Plattformen durchgeführt werden, um Studierenden und Lehrenden einen technisch wie inhaltlich einfachen Überblick über zugängliche Ressourcen anzubieten, die IT-bezogene Verfahren in den Altertumswissenschaften erläutern und darstellen⁸⁴. Mit dem gleichen Ziel werden veröffentlichte Schriften, Empfehlungen, Tutorials, Tools, Module etc. in einer öffentlichen Bibliographie und Linkliste zum Thema „Archäoinformatik“ zusammengestellt und über bestehende Publikationsplattformen zur Verfügung gestellt⁸⁵.

⁸³ In diesem Kontext von besonderer Relevanz für IANUS sind die Angebote von:

NESTOR (<http://www.nestor.sub.uni-goettingen.de/education/index.php>)

DARIAH-DE (<https://de.dariah.eu/forschung-und-lehre>)

ARIADNE (<http://http://www.ariadne-infrastructure.eu/Services>)

⁸⁴ Hierfür eignet sich vor allem die Virtuelle Fachbibliothek Propylaeum:

<http://www.propylaeum.de/altertumswissenschaften/internetressourcen/>

⁸⁵ Z. B. über die Bibliographie von DARIAH-DE: <https://de.dariah.eu/bibliographie>

Der Bereich "Ausbildung und Qualifizierung" verfolgt zudem das Ziel, ein Netzwerk zwischen den beteiligten Akteuren, d.h. vor allem IT-Fachleuten, Dozenten, Wissenschaftlern und Studierenden, zu etablieren, um im Rahmen eines informellen Austausches neue Ideen und sich ändernde Anforderungen zu bündeln und die gemeinsamen Bedürfnisse zu ermitteln, auf die IANUS dann durch seine Dienstleistungen reagieren kann. Hierzu ist eine Arbeitsgruppe vorgesehen, die in regelmäßigen Arbeitstreffen die relevanten Fragen diskutiert und künftige Handlungsfelder skizziert. Anknüpfungspunkte ergeben sich dabei zu der „AG Digitale Geschichtswissenschaften“ im Verband der Historiker und Historikerinnen Deutschlands⁸⁶, in der ebenfalls die Auswirkungen und das Potential der digitalen Veränderungen auf die historische Forschung und Lehre adressiert werden.

3. Aktive Unterstützung von Projekten beim Forschungsdatenmanagement

Die Beratung von Forschern und Forschergruppen zum Umgang mit digitalen Daten im Verlauf eines Projektes wird eine weitere wichtige Dienstleistung von IANUS sein. Ausgehend von der Überzeugung, dass die Langzeitarchivierung und Nachnutzbarkeit von digitalen Daten nicht erst nach Abschluss eines Projektes eine Relevanz besitzen, sondern bereits vor der ersten Erzeugung von Daten mitgedacht werden sollten, ist es das Ziel, Wissenschaftler so früh wie möglich zu unterstützen – also bereits in der Phase der Planung und Antragstellung eines neuen Projektes. Sofern erwünscht und finanziell darstellbar kann eine aktive und praktische Hilfe beim Forschungsdatenmanagement auch über einen längeren Zeitraum projektbegleitend erfolgen.

IANUS wird aktiv versuchen, möglichst frühzeitig in neue oder laufende Forschungsprojekte eingebunden zu werden, um das Datenmanagement zielführend und erfolgreich betreuen zu können. Werden die Daten von Anfang an unter Berücksichtigung einer späteren Archivierung und Nachnutzung erhoben, können im Idealfall Projekte auf diese Weise zusätzliche Arbeit, Zeit und Geld einsparen.

Die Beratungsleistungen umfassen beispielsweise die Beantwortung von Fragen rund um einen Datenmanagementplan und die Hilfe, diesen entsprechend der Bedingungen eines Drittmittelgebers oder einer Institution zu formulieren. Hierzu zählt ebenfalls, Antragsteller bei der Planung neuer Forschungsvorhaben über bereits vorhandene Datenressourcen zu informieren, so dass diese nicht erneut erhoben werden müssen und ggf. weniger Mittel beantragt werden können. Während der Durchführung und Erzeugung von Daten können bei der Auswahl geeigneter Dateiformate, der Verwaltung von großen Dateimengen oder bei der kontinuierlichen Erfassung von Metadaten Tipps und Best-Practices vermittelt werden.

In einem erweiterten Betrieb kann durch die Bereitstellung von technischen Infrastrukturen zur zentralisierte Datenablage von aktuellen dynamischen Projektdaten (sog. Sync-&Cloud-Services, siehe **Kapitel 6.2.1.1**) zusätzlich eine direkte und unmittelbare Unterstützung beim Datenmanagement erfolgen.⁸⁷ Durch den gemeinsamen Zugriff von Datenproduzenten und IANUS auf aktuelle Dateien aus Projekten und Institutionen kann deren strukturierte Speicherung und Dokumentation bereits während des Prozesses ihrer Entstehung und Auswertung aktiv begleitet und ggf. verbessert werden.

⁸⁶ <http://www.historikerverband.de/arbeitsgruppen/ag-digitale-gw>

⁸⁷ IANUS-Schneider (2013) S. 23.

Da analoge und digitale Altdaten in den Altertumswissenschaften eine hohe Relevanz besitzen (vgl. **Kapitel 4.3.1**), wird die Unterstützung nicht nur für diejenigen Projekte angeboten, die neue Daten erheben, sondern auch solche Initiativen, die analoge Altdaten (retro-)digitalisieren wollen oder digitale Altdaten zur Langzeitarchivierung und Nachnutzung aufbereiten wollen (z. B. durch die Ergänzung fehlender Metadaten und die Konvertierung von Datenformaten).

6.2 Erweiterter Betrieb

6.2.1 Technische Dienstleistungen

1. *Automatisierte Erschließung der Datenbestände*

Die inhaltliche Erfassung der Archivbestände von IANUS wird zunächst primär über diejenigen Metadaten erfolgen, die für eine Datensammlung als Ganzes verpflichtend sind. Weitere Metadaten, etwa zu einzelnen Dateien, werden nur dann herangezogen, wenn sie bereits strukturiert vorliegen oder im Rahmen der initialen Kuratierung mit vertretbarem Aufwand erzeugt werden können. Dieses Vorgehen bedeutet u.a., dass Suchanfragen zunächst ausschließlich in den vorhandenen Metadaten, nicht aber in den Dateien selbst durchgeführt werden, so dass Ergebnismengen vor allem von der Qualität der Metadaten abhängen werden.

Um diese Einschränkung zu reduzieren, ist eine tiefer gehende Erschließung der archivierten Dokumente notwendig, was bei einem wachsenden Gesamtbestand realistisch aber nur durch teil- oder vollautomatische Verfahren erfolgen kann. Insofern wird mittelfristig angestrebt, über bestehende Such- und Indizierungs-Lösungen hinaus (z. B. Lucene mit Apache Solr, Elasticsearch) die Möglichkeiten einer rechnergestützten Analyse von Daten auszunutzen, etwa durch die systematische Erschließung von Textinformationen innerhalb von Dateien (insbesondere bei Dokumenten, Datenbanken und Tabellen) oder die Anwendung von Algorithmen aus dem Bereich Text Mining, Data Mining oder Knowledge Discovery in Databases.

2. *Zentrale Datenablage von aktuellen, dynamischen Projektdaten (sog. Sync & Share-Services)*

Projekte, die durch Angehörige verschiedener Einrichtungen durchgeführt werden, sind häufig mit dem Umstand konfrontiert, dass der Zugriff auf einen gemeinsamen Speicherort mit allen aktuellen Projektdaten aufgrund institutioneller, technischer oder rechtlicher Hindernisse erschwert oder verhindert wird. Beispielsweise ist in vielen Fällen der praktische Zugriff aus dem Netzwerk einer Institution auf Dateien, die sich auf einem System einer anderen Institution befinden, gar nicht oder nur mit Mehraufwand möglich. Als Folge daraus erfolgt der Austausch von Daten dann oftmals in individueller direkter Kommunikation oder die Daten werden parallel in mehreren System vorgehalten, so dass deren Konsistenz und Aktualität immer wieder überprüft und abgeglichen werden müssen.

An diesem Punkt möchte IANUS ansetzen und vertrauenswürdige Cloud-Ressourcen mit entsprechenden Sync-&Share-Services zur Verfügung stellen, die von Projekten über die Grenzen von Institutionen hinweg für einen bestimmten Zeitraum und in einer festgelegten Größenordnung in Anspruch genommen werden können. Ziel sind dabei Datenbestände, bei denen Dateien in absehbaren Zeiträumen bearbeitet, hinzugefügt oder gelöscht werden und sich daher dynamisch verändern.

Ein schneller Zugriff auf ein vollständiges und aktuelles Projektverzeichnis soll dabei nicht nur über browserbasierte GUIs erfolgen, wie es gängige Content- bzw. Dokument-Management Systeme bieten. Stattdessen sollen vor allem Systeme mit downloadbareren Clients genutzt werden, die in der Lage sind, externe Cloud-Speicher als Laufwerke in einem lokalen Rechner einzubinden, und entsprechende Synchronisationsmechanismen besitzen. Es wird dabei nicht angestrebt, eigene neue Softwarelösungen zu entwickeln, sondern bestehende Lösungen (zur Zeit z. B. PowerFolder, OwnCloud etc.) auf für IANUS verfügbaren Hardwarekomponenten zu installieren und anzubieten.

Die Verantwortung für die Strukturierung und die Verwaltung von Zugriffsrechten auf die Daten obliegt dem jeweiligen Projekt- bzw. den Dateneigentümern, die selbstständig entscheiden, welche Personen wie auf die Daten zugreifen dürfen und wie die Dateien im Einzelnen abgelegt werden. Je nach Wunsch und personellen Möglichkeiten kann auch eine Beratung und/oder technische Unterstützung zur nachhaltigen und konsistenten Datenstrukturierung erfolgen, auch um eine mögliche spätere Archivierung eines abgeschlossenen Datenbestandes dank einer gut strukturierten, dokumentierten und den Vorgaben entsprechenden Datenablage zu vereinfachen.

Trotz dieses positiven Effektes auf dynamische Daten, die ggf. zu einem späteren Zeitpunkt (z. B. nach Projektabschluss) als statische Daten an ein Archiv überführt werden, ist zu betonen, dass die Langzeitarchivierung als eine von den Sync-&-Share-Services unabhängige Dienstleistung verstanden wird, die separate Anforderungen, Workflows und Systeme benötigt. Darüber hinaus ist zu hoffen, dass durch dieses Angebot das Forschungsdatenmanagement bei mehreren Kooperationspartnern in unterschiedlichen Einrichtungen erleichtert und die Datensicherheit aufgrund regelmäßiger automatischer Backups der Cloud-Speicher erhöht werden kann.

3. Entwicklung, Erprobung und Bereitstellung von fachspezifischen Tools

Zur weiteren Unterstützung der Fach-Community sollen Tools, die insbesondere das Forschungsdatenmanagement und die Aufbereitung von Daten für eine Nachnutzung unterstützen (z. B. durch die Überprüfung von Dateiformaten und Benennungskonventionen, Extraktion von Metadaten, Protokollierung von Dateiänderungen bzw. Dateiversionen), gesammelt und bereitgestellt werden. Diese Tools werden sowohl bereits existierende Softwareprogramme umfassen als auch neue, die mit Unterstützung von IANUS entwickelt werden. Erstere werden hinsichtlich ihrer Relevanz und Funktionalität geprüft, im Rahmen der IT-Empfehlungen (siehe **Kapitel 6.1.2**) kurz vorgestellt und mit Links zur Anwendung für interessierte Nutzer präsentiert. Letztere sind je nach Anforderung, finanziellen Möglichkeiten und kompetenten Partnern ganz neu zu programmieren oder können Weiterentwicklungen von bestehenden Anwendungen darstellen. Sofern erwünscht, ist es auch denkbar, dass die Benutzung von Tools in Workshops erläutert und gemeinsam Vor- und Nachteile diskutiert werden.

6.2.2 Nicht-technische Dienstleistungen

1. Individuelle, Projekt und Institutionen bezogene Beratung

Als Ergänzung und Weiterentwicklung der in **Kapitel 6.1.2** beschriebenen Dienstleistungen „Angebote zur Aus- und Weiterbildung im Bereich Forschungsdatenmanagement und Digitale Altermumswissenschaften“ und „Aktive Unterstützung von Projekten beim Forschungsdatenmanagement“ ist perspektivisch auch ein stärker individualisiertes Angebot angedacht, mit dem

einzelne Personen, Projekt-/Arbeitsgruppen oder Institutionen beraten und unterstützt werden können. Gegenwärtig sind hierfür ganz unterschiedliche Formate denkbar:

- E-Mail-Helpdesk für Datengeber und -nachnutzer
- elektronische Fragestunde
- Schulungen und Beratungen vor Ort in einzelnen Einrichtungen

Welches Format konkret umgesetzt werden soll, kann zuverlässig erst nach einer entsprechenden Stakeholder-Analyse entschieden werden, in der eruiert werden muss, wofür welche Nachfrage in der Fach-Community besteht. Ferner hängen die Details, wie bei allen Dienstleistungen des erweiterten Betriebes, von den künftigen Ressourcen, realistischen Kosten-Nutzen-Rechnungen und noch zu definierenden Betriebsmodellen ab. Wie bei den oben genannten Dienstleistungen wird auch hier eine Abstimmung mit anderen Projekten und Institutionen angestrebt⁸⁸.

2. Aufbereitung und Archivierung von Altdaten mit hoher fachlicher Relevanz

Ein besonderes Desiderat der altertumswissenschaftlichen Forschung ist die Erhaltung von digitalen Daten aus Projekten, die in den vergangenen Jahren abgeschlossen wurden und nun nicht weiter finanziert werden. Aufgrund fehlender Infrastrukturen und Verfahren und eines mangelnden Problembewusstseins ist die Lesbarkeit und Interpretierbarkeit dieser Datenbestände in hohem Maße gefährdet, da sie z. B. nur auf veralteten Speichermedien gesichert wurden. Wenn nicht die durchführende Organisation eigene Mittel bereitstellt, sind in diesen Fällen normalerweise keine Ressourcen mehr vorhanden, um die Daten für eine Langzeitarchivierung vorzubereiten. Oftmals sind auch die Mitarbeiter und Verantwortlichen, welche die Dateien noch aus der Projektzeit kennen, nicht mehr greifbar, so dass eine nachträgliche Dokumentation bzw. inhaltlichen Erschließung und Metadatenvergabe besonders schwierig ist. Je länger das Erstellungsdatum der Dateien zurück liegt, desto mehr ist auch mit veralteten Datenformaten zu rechnen, so dass eine technische Kuratierung ebenfalls zusätzliche Aufwände erfordern wird⁸⁹.

IANUS will daher zusammen mit Institutionen, die Eigentümer solcher verwaister Altdaten sind, versuchen, für die Forschung relevante Bestände zu identifizieren und für die Zukunft zu „retten“, d.h. sie für eine Archivierung und Bereitstellung durch IANUS aufzubereiten. Aufgrund der skizzierten Problematiken und Mehrarbeiten wird dies im Regelfall nur durch gezielte gemeinsame Anträge bei Drittmittelgebern möglich sein. Je nach Datensammlung wird man unterschiedliche Kriterien und Strategien für die Kuratierung und Datenmigration definieren und über neue Wege der Erschließung nachdenken müssen, etwa durch die Mithilfe von Studierenden in betreuten Seminaren, die explizite Beauftragung einzelner externer Experten oder Verfahren für eine (semi-) automatisierte Erzeugung von Metadaten.

3. Bewusstseinsbildung und Interessenvertretung der „Digitalen Altertumswissenschaften“

Auch wenn zum gegenwärtigen Zeitpunkt zu beobachten ist, dass in den Altertumswissenschaften das Bewusstsein für einen nachhaltigen Umgang mit digitalen Forschungsdaten langsam wächst, ist bei vielen Akteuren immer noch ein Beratungs- und Aufklärungsbedarf zu konstatieren. Da die

⁸⁸ Vgl. oben Anm. 83 mit Verweisen auf NESTOR, DARIAH und ARIADNE.

⁸⁹ Der Einsatz von IT-Technologien in den Altertumswissenschaften in nennenswertem Umfang reicht bis in die 1980er Jahre zurück.

Wichtigkeit des Datenmanagements und der Datenarchivierung, die vor dem Verlust von Informationen und Wissen schützen können, oftmals unterschätzt oder überhaupt nicht erkannt wird, werden diese Aufgaben nicht oder nur selten als Bestandteile der wissenschaftlichen Arbeit verstanden und zumeist weder in Zeit- noch Finanzplänen von Projekten integriert. Häufig liegt dies an mangelnden IT-Kenntnissen der beteiligten Akteure, sowohl in den wissenschaftlichen Einrichtungen selbst als auch bei relevanten Verbänden, Mittelgebern oder in politischen Gremien. IANUS setzt sich daher das langfristige Ziel, die Relevanz von fachbezogener, nachhaltiger IT für Fachkollegen und externe Gruppen zu vermitteln und will zu einer umfassenden Aufklärung über die adäquate Pflege digitaler Daten beitragen. Über die oben beschriebenen Dienstleistungen hinaus können gezielte Informationskampagnen durchgeführt werden, die z. B. Geldgeber und Stiftungen oder Projektverantwortliche adressieren, damit in den Zeit- und Finanzplänen für neue Projekte die Datenverarbeitung und -archivierung als fester Bestandteil integriert wird. Eine andere Möglichkeit wäre die Empfehlung und Zertifizierung von Fachanwendungen, Systemen, Einrichtungen und Projekten, die in besonderer Weise offene, dokumentierte, standardisierte und langfristig nachnutzbare Daten generieren oder vorhalten. Ein solches Zertifikat ist dabei nicht als Gegensatz zu etablierten, besonders auf Repositorien und Langzeitarchive zielende Zertifikate⁹⁰ zu verstehen, sondern kann vielmehr eine fachspezifische Ergänzung darstellen, welche auch die Prozesse der Datengenerierung und -analyse stärker in den Blick nimmt.

Und schließlich kann IANUS auch eine aktive Interessenvertretung für die „Digitalen Altertumswissenschaften“ insgesamt leisten, sofern es hier aus den Fach-Communities einen entsprechenden Auftrag erhält. Obwohl der Einsatz von Computern in den Altertumswissenschaften, insbesondere in den Archäologien bereits eine lange Tradition hat⁹¹, spielen die „Digitalen Altertumswissenschaften“ in den aktuellen Diskussionen und Entwicklungen im Kontext der „Digital Humanities“ nur eine untergeordnete Rolle. Hier könnten durch Empfehlungen für Module und Inhalte in neuen Studiengängen (z. B. Archäoinformatik und vergleichbare), die Durchführung von Veranstaltungen für eine breite geisteswissenschaftliche Zielgruppe oder die Mitarbeit in Gremien und Verbänden⁹² Impulse im Sinne der „Digitalen Altertumswissenschaften“ gesetzt werden.

6.3 Öffentlichkeitsarbeit und Community Building

Aktuelle Ergebnisse und Entwicklungen, neue Datensammlungen und Services, begonnene und abgeschlossene Projekte, Angebote im Bereich Ausbildung und Qualifizierung sowie Veranstaltungs- und Literaturhinweise sollen konsequent und ausführlich über die Homepage von IANUS verbreitet werden. Hier werden neben Meldungen zu Projektfortschritten und veröffentlichten Daten auch ausführliche Beschreibungen, Ergebnisse, grundsätzliche Strategien,

⁹⁰ Siehe zu NESTOR, TRAC und dem Data Seal of Approval die Anm. 11–13.

⁹¹ Eine erste Konferenz mit dem Titel „Computer Applications in Archaeology“ wurde bereits 1973 in Birmingham/UK durchgeführt und findet seitdem jährlich sowohl auf mehreren nationalen als auch auf internationalen Veranstaltungen statt.

⁹² Z. B. Interdisziplinärer Forschungsverbund Digital Humanities in Berlin (<http://www.ifdhberlin.de>), Digital Humanities im deutschsprachigen Raum (<http://dig-hum.de/>), DINI/nestor „AG Forschungsdaten“ (http://www.forschungsdaten.org/index.php/AG_Forschungsdaten)

Präsentationen von Veranstaltungen⁹³ und Berichte angeboten sowie Kooperationspartner und Mitarbeiter vorgestellt. Durch ‚Kurzporträts‘ soll auf neue und ältere Datenbestände, die von IANUS archiviert und bereitgestellt werden, aufmerksam gemacht werden, um nicht nur die Aktivitäten des Forschungsdatenzentrums, sondern auch die Datenbestände vielfältig bekannt zu machen. Praktische Anleitungen zum Datenmanagement und zur Vorbereitung von Daten vor der Übergabe an IANUS werden über die IT-Empfehlungen publiziert. Dort werden ebenfalls Hintergrundinformationen, z. B. zur Kalkulation von Gebühren und über die Arbeitsschritte der Langzeitarchivierung, bereitgestellt.

Um eine möglichst hohe Erreichbarkeit bei den unterschiedlichen Zielgruppen zu erlangen, werden die Nachrichten und Neuigkeiten auf unterschiedliche Weisen kommuniziert. Dabei sind vor allem digitale Formate vorgesehen, die je nach Anlass in Abstimmung mit dem Referat für Öffentlichkeitsarbeit des DAI zum Einsatz kommen sollen:

- regelmäßige Newsletter (vierteljährlich, elektronisch)
- Pressemitteilungen über externe Verteiler (idw, DAI)
- RSS-Feed und Blogs
- Social-Media-Kanäle (Twitter, Facebook etc.)
- elektronische Nachrichten: paper.li
- Artikel in Wikipedia
- Nachrichten per E-Mail an Institute und Personen, die als System-Nutzer registriert sind

Für das Monitoring und die Steuerung der eingesetzten elektronischen Formate, die Nutzung der Homepage und bereitgestellten Daten sollen entsprechende Tools wie Google Analytics und Piwik zum Einsatz kommen.

Neben den digitalen Medien sollen auch analoge, gedruckte Materialien wie Flyer, Broschüren und Newsletter über IANUS informieren, um auch weitere, nicht IT-affine Zielgruppen anzusprechen⁹⁴. Durch einen postalischen Versand dieser Erzeugnisse an Universitäten, Institute, Museen, Akademien und weitere Einrichtungen, damit sie dort lokal ausgelegt werden können, oder die Verteilung bei Veranstaltungen lassen sich andere Wissenschaftler und Personen erreichen als dies über rein digitale Kanäle möglich ist. Prinzipiell kann hier zwischen zwei Arten unterschieden werden: solchen Materialien, welche die Grundprinzipien, Dienstleistungen und mittel- bis langfristige Entwicklungen von IANUS in prägnanter Form darstellen (Häufigkeit ca. alle 4–5 Jahre), und solchen, die eher über kurzfristige Ereignisse, Projektergebnisse oder neue Datenbestände berichten (Häufigkeit idealerweise ca. alle 6–12 Monate)⁹⁵.

Ein weiterer wichtiger Weg der Öffentlichkeitsarbeit ist die aktive Teilnahme an Veranstaltungen, wie z. B. wissenschaftlichen Konferenzen, Verbandstagungen, Kolloquien, Workshops und Treffen von Arbeitsgruppen. Auf diesen können z. B. durch Präsentationen nicht nur die Ziele und

⁹³ Die in der 1. Projektphase von IANUS besuchten Konferenzen und Tagungen im Überblick: http://www.ianus-fdz.de/projects/veranstaltungen/wiki/Besuchte_Konferenzen

⁹⁴ Bisherige Beispiele sind ein IANUS-Flyer und die Broschüre „Antike Digital“, die sowohl gedruckt vorliegen, als auch digital auf der Homepage heruntergeladen werden können:
http://www.ianus-fdz.de/attachments/download/514/RZ-DAI-IANUS-flyer_Final_web_2013-10-17.pdf
http://www.ianus-fdz.de/attachments/download/563/ianus_Broschuere_web.pdf

⁹⁵ Vgl. die gedruckte Ausgaben der ADS news: <http://archaeologydataservice.ac.uk/about/newsletter>

Dienstleistungen von IANUS vorgestellt und gemeinsam mit den Fach-Communities diskutiert werden, sondern auch zu allgemeinen Themen von IT in den Altertumswissenschaften Stellung genommen und ein Problembewusstsein vermittelt werden. Durch eine starke Präsenz auf derartigen Veranstaltungen ist einerseits ein direkter persönliche Austausch mit aktuellen und künftigen Nutzern von IANUS möglich, können Fragen und Unsicherheiten unmittelbar beantwortet werden und kann eine Vertrauensbasis für eine potentielle Zusammenarbeit geschaffen werden. Andererseits lassen sich neue technische Trends oder sich ändernde Anforderungen in den Fachbereichen wahrnehmen und bei der Weiterentwicklung von IANUS berücksichtigen.

Für den unmittelbaren Kontakt mit Fachkollegen und Studierenden ist aber auch die Vorbereitung und Durchführungen eigener Veranstaltungen vorgesehen, je nach Thema und Zielgruppe zusammen mit anderen Partnern⁹⁶. Realisiert werden sollen vor allem (aber nicht ausschließlich) Formate, die sich an einzelne Forschungsgruppen, Studierende, Lehrende, Projektmitarbeiter und -leiter an Universitäten und Forschungseinrichtungen richten und das Curriculum ergänzen bzw. als fachspezifische Fortbildungen fungieren. Denkbare Themen und Formate umfassen u.a.:

- Vorträge bzw. Vortragsreihen zu rechtlichen und finanziellen Aspekten der Langzeitarchivierung
- eintägige Workshops zur Benutzung relevanter Tools
- mehrtägige Schulungen zur nachhaltigen Datengenerierung
- 1–2-wöchige Summer Schools zum Datenmanagement von Forschungsdaten

Für eine Öffentlichkeitsarbeit, die gezielt die breite, interessierte Öffentlichkeit und nicht nur ein spezifisches Fachpublikum adressiert, sollen zwei Wege verfolgt werden. Der Fokus der Informationen, die dabei vermittelt werden sollen, liegt auf Datenbeständen, die bekannte Denkmäler, Orte und Texte in und außerhalb Deutschlands behandeln und insofern ein über die Fachgrenzen hinausgehendes, allgemeines Interesse erwarten lassen. Einzelne Aktivitäten und Entwicklungen von IANUS als Forschungsdatenzentrum erscheinen dagegen für Personen, die nicht zum Kreis der Stakeholder gehören, eher von geringerer Relevanz zu sein und spielen daher in diesem Kontext nur eine untergeordnete Rolle.

Zum einen sollen kurze Beschreibungen in populärwissenschaftlichen Magazinen publiziert werden, vor allem in der *Antiken Welt* des Verlages Philipp-von-Zabern, Mainz, und in *Archäologie in Deutschland* des Konrad-Theiss-Verlages, Darmstadt/Stuttgart. Zum anderen wird eine enge Zusammenarbeit mit Institutionen und Verbänden angestrebt, die traditionell bereits Öffentlichkeitsarbeit für ein breites (Laien-)Publikum betreiben, also vor allem Museen und Bodendenkmalämter der Länder und Kommunen. So soll z. B. versucht werden, Daten in IANUS, die thematisch zu einer bestimmten Ausstellung eines Museums oder einer aktuellen Ausgrabung eines Amtes passen, gleichzeitig ebenfalls bekannter zu machen, etwa als Beiträge in Ausstellungskatalogen, als ergänzende Materialien in Online-Auftritten oder als Flyer zum Verteilen vor Ort.

⁹⁶ Vgl. den Lizenzworkshop, der im September 2013 zusammen mit DARIAH-DE durchgeführt wurde: hdl.handle.net/11858/00-1780-0000-0022-DBE5-C

7 Technische Architektur

7.1 Drei-Schichten-Architektur

Für die digitale Langzeitarchivierung der Forschungsdaten durch IANUS wird eine Drei-Schichten-Architektur (3-tier-architecture) geplant⁹⁷, die aus drei getrennten Ebenen besteht:

- Zugriffsschicht
- Managementschicht
- Speicherschicht (Bitstream-Preservation)

Alle drei Schichten kommunizieren mit der jeweils darüber- und ggf. darunterliegenden Schicht über definierte Schnittstellen, so dass sie als organisatorisch und technisch eigenständige Einheiten betrieben werden können. Auf diese Weise werden auch die Voraussetzungen für ein Out-Sourcing an spezialisierte Dienstleister, beispielsweise für die Bitstream-Preservation, geschaffen. Insbesondere die Zugriffsschicht wird als eigenständiger Geschäftsbereich von IANUS angesehen und soll mehr als nur die Bereitstellung einfacher Schnittstellen und Beratungsleistungen umfassen. Um bei den Mitgliedern der Fach-Community eine möglichst hohe Akzeptanz zu erlangen und die effektive Nachnutzung von Daten zu vereinfachen, sollen die Daten über ein eigenes Web-Portal (vor allem zum direkten Download von Dateien) und über verschiedene Web-Services verfügbar gemacht werden. Ein solches Angebot bedeutet einen zusätzlichen Aufwand, der über die reine Definition und Implementierung von Schnittstellen hinausgeht und kontinuierliche Anpassungen an neue Anforderungen und technische Entwicklungen erforderlich macht. Im Gegensatz dazu wird die Ausgliederung der Speicherebene als organisatorisch und technisch einfacher eingeschätzt und ist aufgrund der angestrebten überschaubaren Hardware-Ressourcen von IANUS auch notwendig. Weder Datennachnutzer noch Datengeber werden typischerweise direkt mit der Speicherschicht interagieren, so dass ein Anbieterwechsel, Anpassungen der Schnittstellen oder veränderte technische Anforderungen lediglich zwischen IANUS und dem bzw. den Partnern vereinbart werden müssen. Ein solcher Aufbau kann in Ausnahmefällen aber auch auf besondere Anforderungen einzelner institutioneller Datengeber reagieren, z. B. wenn es für diese konkrete Vorgaben bezüglich der tatsächlichen Speicherorte von archivierten Daten geben sollte.



⁹⁷ Folgender Abschnitt baut auf den Empfehlungen von IANUS. SCHNEIDER (2013) S. 30–31 auf.

7.1.1 Zugriffsschicht

Der Zugriff auf die in Form von DIPs bereitgestellten Datenbestände von IANUS soll vor allem über zwei Mechanismen erfolgen: einmal über ein Online-Datenportal für individuelle Nutzer, damit diese möglichst unabhängig von ihrem jeweiligen Betriebssystem und ohne die Installation zusätzlicher Programme Dateien finden, ansehen und nachnutzen können. Zum anderen sind aber für den automatisierten Austausch von Metadaten verschiedene Web-Services und Schnittstellen vorgesehen, um die Informationen möglichst weit verbreiten und ggf. mit externen Inhalten anreichern zu können. Aufgrund der heterogenen Daten und Benutzergruppen ist ein differenziertes Rechte-Rollen-Konzept erforderlich, das sowohl die individuellen Zugriffe auf Metadaten, einzelne Dateien und ganze Datensammlungen durch Datennachnutzer über ein Web-Portal regelt als auch die automatisierten Zugriffe über Web-Schnittstellen berücksichtigt und in **Kapitel 11.3** skizziert wird.

7.1.2 Managementschicht

Die Managementschicht stellt die wesentliche Schicht von IANUS dar, da in ihr das Fachwissen, definierte Workflows, die Verwaltung und Verarbeitung der Daten und weitere zentrale Aspekte organisiert werden. Diese Schicht stellt die Verknüpfung von Objekten in der Speicherschicht mit den Metadaten, die von IANUS direkt verwaltet werden, her. Zudem liefert es die Relationen zwischen digitalen Objekten untereinander, erlaubt eine Versionierung von Datensätzen, verknüpft die jeweiligen Objekte mit unterschiedlichen Darstellungs- und Zugriffsmechanismen (definiert und betrieben in der Zugriffsschicht), die Abbildung von Zugriffsrechten (Identity Management) und den Schutz der Daten.

7.1.3 Speicherschicht (Bitstream Preservation)

Auf der untersten Schicht, der Speicherschicht, werden die digitalen Objekte einer Datensammlung, d.h. also die SIPs und AIPs bestehend aus den eigentlichen Dateien zusammen mit den zugehörigen Metadaten, gespeichert und archiviert. Auf dieser Ebene soll dabei eine einfache dateibasierte Ablage erfolgen, die eine Rekonstruktion aller Informationen allein anhand der Dateien und Metadaten erlaubt und nicht auf weitere Komponenten wie das Archiv-Management-System angewiesen ist. Aufgrund des Umfangs und der Größe der Datensammlungen werden diese zur Erfüllung der Storage-Anforderungen an (externe) Dienstleister in ein Datenzentrum bzw. Rechenzentrum ausgelagert. Alle direkt mit der Bitstream-Preservation verbundenen Aufgaben, wie Integritätstests, Datensicherheit, Datenreplikation, Tape-Backup usw. obliegen dann diesen Dienstleistern und werden durch Service-Level-Agreements vertraglich abgesichert. Aufgaben von IANUS bleiben alle vorbereitenden Maßnahmen, insbesondere die Erstellung valider Informationspakete, die Überwachung der Services der technischen Partner sowie deren Risikomanagement.

7.2 Technische Partner

7.2.1 Technische Partner (Rechenzentren) zur Bitstream-Preservation

Da IANUS aus Kosten-, Platz- und Kompetenzgründen nicht anstrebt, zu archivierenden Datensammlungen auf eigenen, in-house Hardware-Architekturen vorzuhalten, wird es für die physische Langzeitarchivierung, die sog. *bitstream preservation*, auf Services von professionellen Anbietern zurückgreifen. Dabei sollte nach gängigen Empfehlungen zur Erhöhung der Ausfallsicherheit eine redundante und verteilte Datenvorhaltung erfolgen. Dies bedeutet, dass mehrere Kopien von AIPs an mehreren, ausreichend weit entfernten Orten gespeichert werden, um die Zerstörung der Daten durch ein einzelnes katastrophales, räumlich begrenztes Ereignis unwahrscheinlich zu machen – auch wenn dies zu finanziellen Mehrbelastungen führt. Die Vorhaltung der AIPs in absichtlich unterschiedlichen (Speicher-) Systemen und Repositorien bietet zudem den Vorteil, dass sie nicht von einer unentdeckten Fehlfunktion in einem der Systeme gefährdet werden.

Es wurde daher entschieden, dass die Datenbestände von IANUS bei mindestens zwei spezialisierten Infrastruktureinrichtungen, also vor allem Rechenzentren, archiviert werden sollen. Das DAI vertraut beim Aufbau und Betrieb seiner IT-Infrastruktur im Bereich des Storage-Managements seit mehreren Jahren auf die Expertise des Regionale Rechenzentrum der Universität zu Köln (RRZK) als externem Partner, so dass auch IANUS dank seiner institutionelle Anbindung von dieser Kooperation profitieren und entsprechende *deep storage*-Services (mit)nutzen kann. Ein zweiter Partner für die *bitstream preservation* ist noch zu suchen, wobei bei der Auswahl die für IANUS relevanten Anforderungen zu definieren und bestehende Kriterienkataloge⁹⁸ zu berücksichtigen sind.

Zwischen IANUS und den technischen Partnern müssen verbindliche Service-Level-Agreements (SLA) geschlossen werden, in denen u.a. festgehalten wird, dass IANUS für die Kontrolle des Ingest-Prozesses von Informationspaketen verantwortlich ist, es den vollen Zugriff auf die Daten behält und es die notwendigen Workflows bereitstellt. Auch muss vereinbart werden, in welcher Form IANUS die Einhaltung des SLAs überprüfen kann und welche Konsequenzen sich bei Nicht-Einhaltung von Zusagen ergeben.

7.2.2 Technische Partner für Exit Strategie

Für ein nachhaltiges Konzept eines Forschungsdatenzentrums, das primär die Langzeitarchivierung von digitalen Daten betreibt, ist es auch erforderlich, Vorbereitungen für den Fall zu treffen, dass der Betrieb vollständig eingestellt werden muss, z. B. weil die erforderlichen Mittel fehlen oder die Akzeptanz in der Fach-Community ausbleibt. Diese Überlegungen werden in einer sogenannten Exit-Strategie festgehalten, deren wesentliche Kernpunkte in **Kapitel 11.6** skizziert werden. Einen zentralen Aspekt dabei betrifft die Fragen, wer die Verantwortung für die bereits vorhandenen archivierten Datenbestände übernimmt, sich um deren zukünftige *bitstream preservation* kümmert und weiterhin ihre Bereitstellung für Endnutzer gewährleistet. Insofern werden für dieses Szenario externe technischer Partner benötigt, mit denen zusammen eine potentielle Abwicklung von IANUS durchgeführt werden kann. Dies können entweder diejenigen Einrichtungen sein, mit denen bereits Kooperationen zur *bitstream preservation* bestehen und

⁹⁸ Z. B. verschiedene, von dem Kompetenznetzwerk NESTOR publizierte Materialien
http://www.langzeitarchivierung.de/Subsites/nestor/DE/Publikationen/Materialien/materialien_node.html

deren Auftrag entsprechend ausgeweitet wird, oder aber es sind ganz neue Partner, die sich nach dem Ende von IANUS der Kuratierung und Bereitstellung der existierenden Datenbestände annehmen. Denkbare Einrichtungen wären etwa bestehende Forschungsdatenzentren aus einer anderen Wissenschaftsdisziplin oder aber eines der zahlreichen Länder- und Bundesarchive.

7.3 Schnittstellen

Als wesentliche Komponenten der technischen Gesamtarchitektur von IANUS werden verschiedene Schnittstellen und Web-Services angesehen, die auf unterschiedliche Weise die Weitergabe und den Austausch von Inhalten gewährleisten, sowohl für IANUS als Empfänger von (Meta-)Daten als auch für IANUS als Bereitsteller von Informationen. In besonderem Maße gilt diese Anforderung für die vom DAI selbst betriebenen Fachsysteme und Anwendungen der iDAI.welt⁹⁹, mit denen das Archiv-Management-System von IANUS in besonderem Maße vernetzt sein wird. Soweit wie rechtlich möglich und technisch sinnvoll, sollen dabei die Prinzipien von Linked Open Data umgesetzt und die Daten in maschinell verarbeitbarer und ausreichend dokumentierter Form über eine API zur Verfügung gestellt werden.

7.3.1 Datengeber und IANUS

Schnittstellen, damit Datengeber Ihre Daten an IANUS übertragen können:

- webbasierte Upload-Formulare
- FTP-Protokolle
- Cloud-Services des DAI oder kommerzieller Anbieter
- postalisch versandte oder direkt übergebene externe Datenträger
- im erweiterten Betrieb: eigene Sync-&-Share-Systeme
- Harvesting von anderen größeren Repositorien

7.3.2 Datennachnutzer und IANUS

Web-Interfaces, damit Datennachnutzer die Daten von IANUS finden und nutzen können

- Web-Portal mit Such- und Download-Möglichkeiten und integriertem Nachweiskatalog
- Im erweiterten Betrieb: Web-Mapping- / Web-Feature-Service für freigegebene Geodaten

Bereitstellung von Metadaten mittels OAI-PMH zur Nutzung in externen Web-Systemen und Metadaten-Aggregatoren (siehe **Kapitel 6.1.14**):

- Propylaeum, DARIAH-DE
- Deutsche Digitale Bibliothek, Europeana
- International: ADS (UK), eDNA (NL), ARIADNE, Fasti-Online¹⁰⁰...

⁹⁹ <http://www.dainst.org/idai.welt>

¹⁰⁰ <http://www.fastionline.org>

7.3.3 APIs für institutionelle Webseiten und Fachsysteme

Ferner ist geplant, eine API anzubieten, um anderen Systemen die Einbettung von Widgets zu erlauben, in die Metadaten von IANUS eingebunden werden können und somit nicht nur über die IANUS-Homepage, sondern auch dezentral über institutionelle Webseiten gefunden und genutzt werden können. Als Fachsysteme, die nach der gegenwärtigen Einschätzung von einer solchen Möglichkeit der Einbindung profitieren würden, sind zu nennen

- Fachsysteme der Landesdenkmalämter: ADAB-Web (Baden-Württemberg, Niedersachsen), ArchaeoDox (Schleswig-Holstein, Hamburg) und weitere
- Spezialisierte und etablierte Fachanwendungen wie NESPOS¹⁰¹ (Mettmann), ArboDat¹⁰² (Wiesbaden), xBook¹⁰³ (München) und weitere

7.3.4 Einbindung von Referenzsysteme und Normdaten

Um möglichst einheitliche Metadaten zu erzeugen und zu verwalten, sollen relevante Normdaten und Thesauri in das Archiv-Management-System eingebunden werden¹⁰⁴, entweder über entsprechende automatisierte Schnittstellen oder über manuelle Import-/Export-Funktionalitäten:

- für Ortsnamen: iDAI.gazetteer, Pleiades, Geonames
- für bibliographische Angaben: iDAI.bibliography / ZENON, Propylaeum
- für Personen: Gemeinsame Normdatei (GND), OrcID, VIAF
- für Schlagworte: iDAI.vocab, Getty Arts and Architecture Thesaurus (de), Thesaurus von Propylaeum, Codelisten aus Fachsystemen der Landesarchäologen, English Heritage Data
- für Zeitbegriffe: Perio.do, iDAI.chronontology

7.4 Persistente Identifier

Damit archivierte Dateien und Dokumente nachhaltig zitiert werden können, werden persistente Identifikatoren (PID) benötigt. Ein PID ist eine einzigartige Zeichenkette, die einer Datei oder einer Datensammlung zugeordnet werden kann, um diese eindeutig zu identifizieren oder zu referenzieren. PIDs sind vergleichbar mit ISBN-Nummern für Bücher.

Als Ergebnis eines Testbeds, in dem die aktuell verfügbaren Systeme und Dienste unter Berücksichtigung der Anforderungen von IANUS untersucht wurden¹⁰⁵, wird das ISO-zertifizierte DOI-System empfohlen, für das die weite Verbreitung und der hohe Bekanntheitsgrad sprechen. Zudem wird es auch von anderen Einrichtungen, die Forschungsdaten langfristig archivieren und bereitstellen, verwendet, so dass vielfältige Erfahrungen vorliegen. Für das System spricht außerdem, dass durch die hierarchische Struktur der Agenturen mit dem IDF an der Spitze

¹⁰¹ <https://www.nespos.org/display/openspace/Home>

¹⁰² <http://www.hessen-archaeologie.de/Archaobotanik/Datenbankprogramm/datenbankprogramm.html>

¹⁰³ http://xbook.vetmed.uni-muenchen.de/wiki/Main_Page

¹⁰⁴ Siehe auch die Angaben in **Kapitel 12.3** mit entsprechenden URLs.

¹⁰⁵ IANUS – Testbed PID (2014).

gewährleistet wird, dass tatsächlich alle registrierten DOIs über die DOI-Resolver auflösbar sind.

Für die Registrierung von DOIs wurde im März 2014 ein Service-Level-Agreement mit da|ra abgeschlossen, eine von vier Registrierungsagenturen für DOIs in Deutschland, die gemeinsam von GESIS und die ZBW betrieben wird vor. Der Service richtet sich vor allem an die Sozial- und Wirtschaftswissenschaften, steht aber auch anderen Interessenten kostenlos zur Verfügung. Die an da|ra weitergeleiteten Metadaten zur Registrierung von DOIs erhalten eine CC0-Lizenz.

Jede Datensammlung bzw. jedes Informationspaket, das über IANUS archiviert und bereitgestellt wird, erhält mindestens einen eigenen Identifikator. Dieser wird IANUS-intern zugewiesen, sobald die Daten übergeben werden. Eine verbindliche Registrierung des Identifikators erfolgt jedoch erst zu einem späteren Zeitpunkt, wenn die Daten für die Archivierung und Bereitstellung vorbereitet wurden und der Datenproduzenten seine Zustimmung gegeben hat. Werden nach der Registrierung Veränderungen an den Daten durchgeführt, so muss eine neue Version mit einem neuen PID erzeugt und zugewiesen werden. Wichtig ist in jedem Fall, dass für den jeweiligen Datensatz der passende Zitierhinweis in der Weboberfläche (auf der Landing-Page) gut sichtbar platziert wird, um auch das Zitieren von Einzeldateien einer mit einem PID versehenen Datensammlung zu erleichtern.

Jeder DOI wird folgende Struktur aufweisen: 10.13149/aaa.bbbbbb-c,

- Das Präfix (10.13149) wurde von da|ra einmal für IANUS insgesamt festgelegt
- Das Suffix (aaa.bbbbbb) wird von IANUS selbst definiert, gibt keine semantischen Informationen preis, besteht aus alphanumerischen Zeichen und darf keine Sonderzeichen oder Umlaute enthalten. Groß- und Kleinbuchstaben werden nicht unterschieden. Das Suffix ist zweigeteilt, wobei der erste dreistellige Teil aus einem abstrakten Kürzel für einen Dateneigentümer bzw. eine Institution steht, der zweite nach einem trennenden Punkt eine Zeichenkombination für den Datensatz darstellt.
- Nach dem Suffix folgt getrennt durch einen Bindestrich eine einstellige Prüfziffer (c), die aus dem Präfix und Suffix (ohne Trennzeichen) mit dem Verhoeff-Algorithmus berechnet wird und ebenfalls alphanumerisch ist.

Die Vergabe von PIDs für alle einzelnen Objekte oder Dateien, die Teil einer größeren Datensammlung sind, wird aufgrund der reinen Quantität als nicht praktikabel erachtet. Jedoch erscheint es sinnvoll, dem Nutzer eine Möglichkeit zur Verfügung zu stellen, mit der neue PIDs für einzelne digitale Ressourcen beantragt werden können („PID-on-Demand“), z. B. zur Referenzierung eines einzelnen Bildes einer Ausgrabung. Dabei müssen die notwendigen Metadaten ergänzt werden, was teilweise durch die Übernahme der übergeordneten Datensammlung erfolgen kann, jedoch auch manuell geprüft werden muss. Eine Zuweisung von DOIs an kleinere Informationseinheiten, z. B. einzelne Dateiteile, ist zunächst nicht vorgesehen.

Für jedes bei IANUS registrierte Objekt mit einem eigenen DOI, egal ob Datensammlungen oder einzelne Dateien, soll eine sog. Landing-Page erstellt werden, die als Konzept aus Bibliothekskatalogen bekannt ist. Eine übersichtliche Darstellung der Informationen zu dem aufgerufenen Objekt ermöglicht es dem Nutzer zu überprüfen, ob das Objekt den Erwartungen entspricht. Insofern werden hier folgende Informationen angezeigt:

- Volle Zitation der Daten
- Zugriffsmöglichkeiten auf die Daten und Informationen über eventuelle Beschränkungen

- zugehörige Metadaten
- Informationen über Software oder Kontext, um die Daten zu öffnen und zu verstehen
- Informationen zu anderen Versionen der Daten oder Hinweise auf verwandte Daten

Die mit einem DOI verbundenen Metadaten sollen generell uneingeschränkt sichtbar sein. Einzige Ausnahmen bilden hierbei personenbezogene Daten und, wenn notwendig, auch geografische Angaben. Ein Teil der Metadaten werden Prüfsummen sein, um dem Nutzer eine weitere Möglichkeit zur Überprüfung des Objektes zu geben. Prüfsummen eignen sich aber nur für in sich abgeschlossene einzelne Dateien. Auch wird ein Datumsfeld integriert werden, in das ein Ablaufdatum oder ein Datum zur Wiedervorlage eingetragen werden kann, damit das System eigenständig auf Metadatensätze hinweisen kann, bei denen etwa die URL oder die Zugriffsrechte überarbeitet werden müssen.

Zukünftige Nutzer werden die Möglichkeit erhalten, defekte PIDs zu melden und bei Fehlern eine informative Fehlermeldung zu bekommen. Dies stärkt die Vertrauenswürdigkeit und sichert die Qualität des Dienstes. Durch eine Veröffentlichung über den Umgang und die Arbeitsabläufe für PIDs in IANUS, kann die Vertrauenswürdigkeit zusätzlich gestärkt werden.

Da das DOI-System eine erweiterte Implementierung des Handle-Systems ist, kann im Falle der Auflösung der DOI-Foundation ein eigenes System erstellt werden. Für den Fall, dass auch da|ra seine Dienste einstellt, müssen die von da|ra eingefügten Metadaten ebenfalls für die registrierten Datensätze lokal bei IANUS gespeichert werden.

Die Verwendung von PIDs muss ebenfalls in der Exit-Strategie von IANUS berücksichtigt werden.

8 Geschäftsmodell und Rechtsform

8.1 Gesellschaftsform

Zur Verstetigung und Institutionalisierung des infrastrukturellen Kerns von IANUS ist eine Rechtsform notwendig, die den Zusammenschluss der Fach-Community für ein nationales Forschungsdatenzentrum in den Altertumswissenschaften und Archäologien zum Ausdruck bringt und die die Partizipation und Rechte der einzelnen Akteure und Stakeholder eindeutig definiert. IANUS wird aus verschiedenen Gründen (z. B. zur Gewährleistung langfristig zuverlässiger Rahmenbedingungen und zur Nutzung von finanziellen, technischen, organisatorischen und infrastrukturellen Synergieeffekten) institutionell am DAI angebunden sein, soll aber dennoch von der Fach-Community für die Fach-Community gelenkt werden. Die Zugehörigkeit zu einer bzw. der Betrieb durch eine Bundesinstitution sieht Schwiegelshohn als Notwendigkeit für den Erfolg eines nationalen Forschungsdatenzentrums an.

„[...] Ein solches nationales Forschungsdatenzentrum für die Altertumswissenschaften kann und muss von einer Einrichtung innerhalb der Fachgemeinschaft aufgebaut und betrieben werden, welche die fachliche Reputation hat, erfahren und langlebig genug ist und eine vom Bund finanzierte Einrichtung ist, um diese finanziell herausfordernde Aufgabe langfristig und nachhaltig zu gewährleisten.“¹⁰⁶

Darüber hinaus führt er aus, dass es aus Gründen der Nachhaltigkeit sowie der für sich genommenen geringen Forschungsaktivität nicht sinnvoll erscheint, ein Forschungsdatenzentrum als technische Infrastruktur an ein einzelnes Universitätsinstitut anzugliedern¹⁰⁷.

Ausgehend von diesen Prämissen kommen laut einer Expertise durch die Kanzlei Mock (Berlin), in der verschiedene mögliche Gesellschaftsformen für den Betrieb eines nationalen Datenzentrums juristisch geprüft wurden, nur die Modelle der gemeinnützigen GmbH (gGmbH) sowie die der Vereinsform in Frage. Andere Gesellschafts- und Rechtsformen wie Aktiengesellschaft, GbR, Stiftung und Kommission sind aufgrund verschiedener formaler Kriterien nicht zu empfehlen oder können nicht gewählt werden.¹⁰⁸ So hätte z. B. die Etablierung von IANUS als eigenständige Kommission/Abteilung innerhalb DAI zur Folge, dass alle durch Gebühren oder Dienstleistungen erwirtschafteten Einnahmen direkt an das Bundesfinanzministerium abgeführt werden müssten und nicht wie beabsichtigt in die Kuratierung von Forschungsdaten investiert werden könnten.

Der Interessenausgleich zwischen den verschiedenen Stakeholdern der Fach-Community, in rechtlicher Hinsicht überwiegend Anstalten des öffentlichen Rechtes mit unterschiedlichen Aufträgen, ist am ehesten zu erreichen, wenn IANUS als gGmbH gegründet wird. Die maßgeblich relevanten Akteure können dann in verschiedenen Gremien (Aufsichtsrat und wissenschaftlicher Beirat, siehe **Kapitel 8.2**) angemessen beteiligt werden, um die strategischen und operativen Entscheidungen und Entwicklungen von IANUS mit zu steuern und zu beeinflussen. Die folgenden Ausführungen beschreiben daher IANUS in der Rechtsform einer gGmbH.

¹⁰⁶ IANUS. Schwiegelshohn (2013) S. 11.

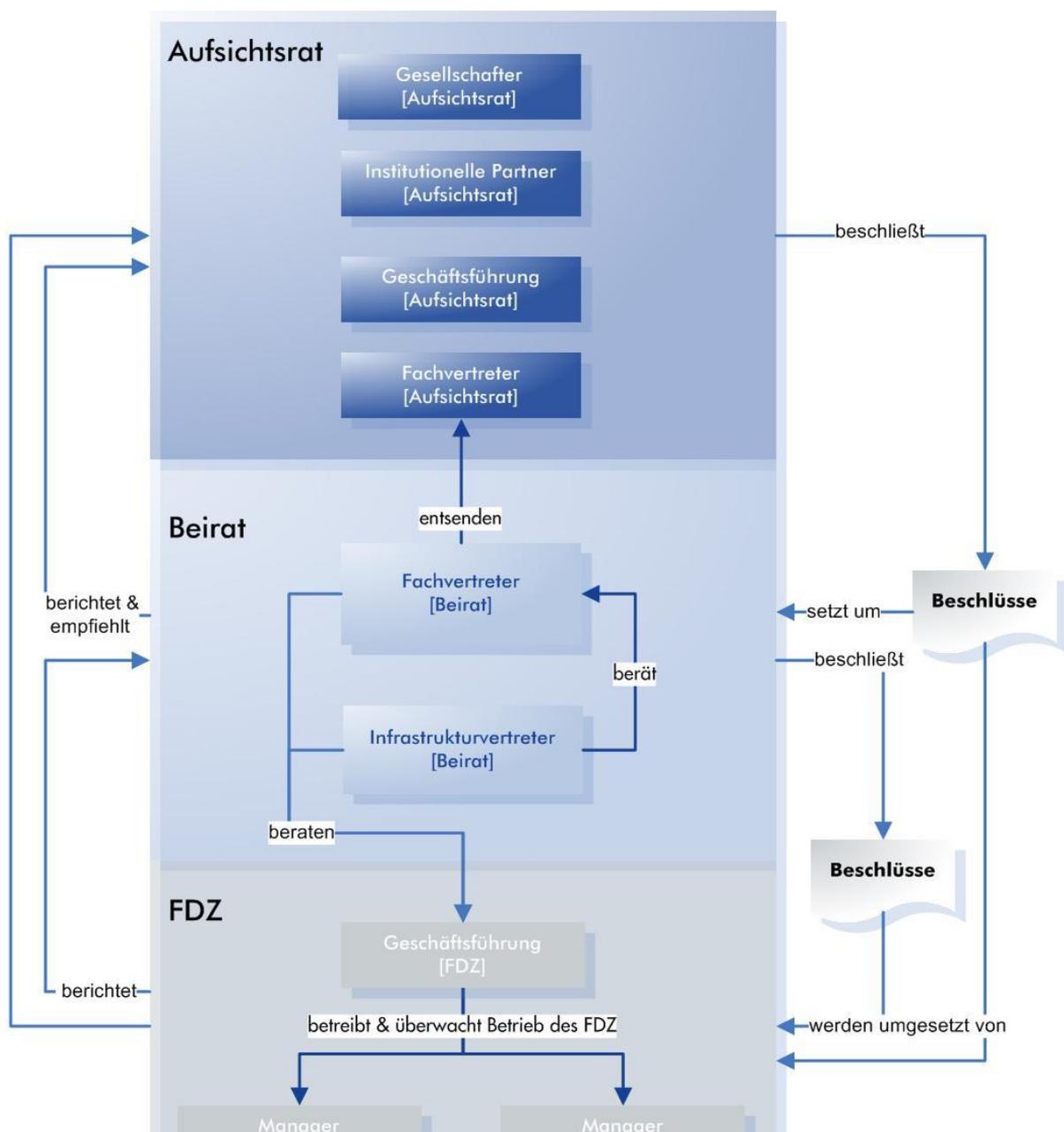
¹⁰⁷ IANUS. Schwiegelshohn (2013) S. 10.

¹⁰⁸ Vgl. IANUS. Mock (2014) S. 19.

Wenn es zu einer maßgeblichen finanziellen Beteiligung des Bundes an IANUS kommt und der Bund daher einer von mehreren Gesellschaftern wird, ist es aus juristischen Gründen sinnvoll, IANUS in den Geschäftsbereich eines Bundesministeriums einzugliedern, wobei dies entweder durch eine unmittelbare Beteiligung des Bundes oder durch eine mittelbare (z. B. durch eine dem Geschäftsbereich eines Ministeriums zugeordnete Einrichtung wie dem DAI) erfolgen kann.

8.2 Gremien

Die Grundlage für die Aufgaben, die Funktionen, die Besetzung, die Entscheidungen, die Verfahren etc. der beiden Lenkungsorgane in IANUS wird durch einen Gesellschaftsvertrag (Satzung) sowie eine Geschäftsordnung von IANUS geregelt. Dabei gilt als Leitgedanke für die Governance-Struktur, dass IANUS als explizites Projekt der Fachgemeinschaft auch von einem die Fachgemeinschaft repräsentierenden Gremium gesteuert wird.



8.2.1 Aufsichtsrat

Der Aufsichtsrat stellt das wichtigste Instrument für die strategische Ausrichtung von IANUS dar. Seine Mitglieder sind zum einen Wissenschaftler, teils gewählt, teils von kooperierenden Forschungseinrichtungen entsandt, zum anderen Vertreter der Zuwendungsgeber von IANUS¹⁰⁹. Er erstellt einen Arbeitsplan über jeweils fünf Jahre, begleitet dessen Umsetzung und kontrolliert die zur Verfügung stehenden Finanzmittel. Zu seinen Aufgaben gehören:

- Festlegung der strategischen Ausrichtung von IANUS
- Erstellung, Diskussion und Verabschiedung eines langfristigen Entwicklungsplanes auf Basis von Vorschlägen der beiden Beiräte und der Geschäftsführung
- Festlegung von Meilensteine innerhalb des Arbeitsplans
- Verabschiedung des Finanzierungsplans und der Mittel für den Arbeitsplan
- Einsetzung und Kontrolle der Geschäftsführung
- Genehmigung der Protokolle und Berichte über aktuelle Entwicklungen
- Kontrolle der Umsetzung der Beschlüsse des Aufsichtsrates
- Entscheidung über Vorschläge der beiden Beiräte, die ein im Gesellschaftsvertrag festgelegtes Kostenvolumen übersteigen
- Entwicklung, Verabschiedung und Änderungen des Gesellschaftsvertrages und der Geschäftsordnung

Der Aufsichtsrat tagt zweimal innerhalb von fünf Jahren, idealerweise zur Mitte und am Ende eines Arbeitsplanes. Er wird besetzt durch Vertreter der Gesellschafter bzw. der Institutionen, die sich zur finanziellen Beteiligung an IANUS verpflichtet haben, durch die Geschäftsführung sowie institutionelle Partner, deren Anzahl variieren kann:

- Hauptgesellschafter (DAI, Bund und ggf. Länder)
- Institution(en), welche maßgeblich zur Basisfinanzierung von IANUS beitragen
- Geschäftsführung IANUS
- gewählte Fachvertreter aus dem wissenschaftlichen Beirat

Das rechtliche Konstrukt der gemeinnützigen GmbH muss nach § 35 GmbHG gerichtlich wie auch außergerichtlich durch eine von den Gesellschaftern einzusetzende Geschäftsführung vertreten werden. Die innerbetriebliche Ausgestaltung der Rechte und Pflichten der Geschäftsführung wird in der Geschäftsordnung der gGmbH festgehalten.¹¹⁰ Auch ist hier die Anbindung an das DAI zu spezifizieren, bei dem die konkrete Fach- und Dienstaufsicht für die Mitarbeiter von IANUS anzusiedeln ist.

8.2.2 Wissenschaftlicher Beirat

Neben dem Aufsichtsrat gibt es einen wissenschaftlichen Beirat, der zu den langfristigen

¹⁰⁹ Die hier aufgeführten Mitglieder des Aufsichtsrates und des wissenschaftlichen Beirates sind lediglich als exemplarische Vorschläge zu verstehen, die bei der Ausarbeitung des Gesellschaftsvertrages diskutiert und festgelegt werden müssen.

¹¹⁰ Vgl. IANUS. Mock (2014) S. 4. 14 f.

Planungen von IANUS Stellung bezieht, bei der Umsetzung der Vorgaben des Steuerungsgremiums berät und diese überwacht sowie konkrete Empfehlungen für den fünfjährigen Arbeitsplan formuliert. Zu seinen Aufgaben gehören:

- Koordinierung und Kontrolle der Umsetzung der Beschlüsse des Aufsichtsrats
- Beratung und Unterstützung der Geschäftsführung bei operativen Aufgaben
- Empfehlungen für und Diskussion über zukünftige Dienstleistungen und technische Erweiterungen von IANUS
- Entscheidung für Maßnahmen, die ein im Gesellschaftsvertrag festgelegtes Kostenvolumen nicht übersteigen
- Empfehlungen für Änderungen des Gesellschaftsvertrages und der Geschäftsordnung
- Entscheidung über die personelle Zusammensetzung des Nutzer-Beirates

Der Beirat tagt mindestens einmal jährlich, bei Bedarf auch häufiger. Er wird maßgeblich durch Personen besetzt, die einerseits die verschiedenen Akteure der deutschen Altertumswissenschaften (siehe **Kapitel 4.1, 4.2**) vertreten und andererseits technischen Infrastruktureinrichtungen bzw. Forschungsdatenzentren angehören. Der Beirat unterteilt sich entsprechend den Stimmrechten in die folgenden zwei Statusgruppen:

- Gruppe der Fachvertreter mit Stimmrechten
 - drei Vertreter Bodendenkmalpflege/Landesarchäologie
 - zwei Vertreter Museen
 - ein Vertreter Außeruniversitäre Forschungseinrichtungen
 - drei Vertreter Universitäre Lehre- und Forschung unterschiedlicher Fachdisziplinen
 - ein Vertreter Wissenschaftliche Verbundprojekte
 - ein Vertreter Kommerzielle Dienstleister
 - ein Vertreter Deutsches Archäologisches Institut
 - Geschäftsführung IANUS
- Gruppe der Infrastrukturvertreter ohne Stimmrechte
 - Archaeology Data Service, UK
 - e-depot for niederlandse archaeologie/DANS, NL
 - Datenarchiv für die Sozialwissenschaften/GESIS, Köln
 - PANGAEA, Bremerhaven
 - Digitales Archiv NRW, Köln
 - zwei universitäre Zentren für geisteswissenschaftliche Daten
 - 1. Technischer Partner – Regionales Rechenzentrum der Universität Köln
 - 2. Technischer Partner – N.N.
 - Archivpartner für Exit-Strategie – N. N.

8.2.3 Nutzer-Beirat

Um regelmäßig und im direkten persönlichen Gespräch überprüfen zu können, ob die Zielsetzungen und Dienstleistungen noch den Anforderungen der in **Kapitel 3** definierten Zielcommunity entsprechen, wird neben dem wissenschaftlichen Beirat ein Gremium etabliert, das sich vor allem aus aktiven Datengebern und Datennachnutzern zusammensetzt. Es hat die

Aufgaben, IANUS bei der Entwicklung der operativen Dienstleistungen zu beraten sowie positive wie negative Rückmeldungen zu konkreten Angeboten zu formulieren. Auf diese Weise sollen praktische Nutzerprobleme und -interessen frühzeitig erkannt und berücksichtigt sowie die inhaltliche Ausgestaltung und Qualität der Serviceleistungen kontinuierlich verbessert werden. Um diese Ziele zu erreichen, sollen parallel und unabhängig von diesem Gremium auch die Aktivitäten der Öffentlichkeitsarbeit (**Kapitel 6.3**) sowie Umfragen und Nutzungsanalysen (**Kapitel 11.8**) genutzt werden.

8.3 Finanzierung

Die Finanzierung von IANUS basiert auf einem dreisäuligen Modell, das im Folgenden näher beschrieben wird. Dieses Modell ist modular aufgebaut, um flexibel auf Veränderungen der Einnahmenseite reagieren zu können. Es wurde zudem eine sogenannte Fall-Back-Strategie definiert, welche sicherstellt, dass bei einem kurzfristigen Wegfall von Einnahmen der Kernbetrieb (siehe **Kapitel 6.1**) von IANUS nicht gefährdet ist und die Online-Bereitstellung von bereits archivierten Forschungsdaten weiterhin langfristig gewährleistet werden kann.

8.3.1 Finanzierungsmodelle

Die Begehungen und Gespräche mit vergleichbaren nationalen und internationalen Datenzentren zeigen¹¹¹, dass verschiedene Finanzierungsmodelle existieren, um die anfallenden Kosten für die Langzeitarchivierung von Forschungsdaten abzudecken. Diese werden hier kurz zusammengefasst, um auf dieser Basis und unter Abwägung der Vor- und Nachteile ein Finanzierungsmodell für IANUS zu beschreiben.

1. Gebührenfinanzierte Modell

Die gesamten Kosten eines Datenzentrums werden zu 100% mittels Gebühren auf die Datenproduzenten umgelegt. Da die Nutzung der Angebote und Daten für Datennachnutzer kostenlos sind, finanzieren die Institutionen und Projekte, welche Ihre Daten kuratieren und archivieren lassen wollen, auf diese Weise die gesamte Einrichtung. Da die Gebühren nur einmal bei der Übergabe von Daten erhoben und fällig werden, werden aus diesen entsprechende Rücklagen für die Zukunft gebildet mit denen auch langfristige Folgekosten abgedeckt werden müssen. Insofern sind die Gebühren bei diesem Modell entsprechend hoch. Die Planbarkeit der Einnahmen hängt ausschließlich von diesem Faktor ab (abgesehen von ggf. Drittmittelprojekten oder einmaligen Zuwendungen).

Als Beispiel aus der Praxis ist tDAR – the Digital Archaeological Record¹¹² – zu nennen, das in den USA von Digital Antiquity¹¹³ betrieben wird. Da tDAR seine gesamten Kosten ausschließlich durch Gebühren abdeckt, fallen diese im Vergleich zu den anderen Finanzierungsmodellen am höchsten aus.

¹¹¹ Vgl. IANUS. Übersicht Datenzentren & Fachsysteme (2013)

¹¹² <http://www.tdar.org/about/pricing/> [Zugriff 21.11.2013]

¹¹³ <http://www.digitalantiquity.org/> [Zugriff 21.11.2013]

2. Institutionelle Modelle

Die gesamten Kosten der Langzeitarchivierung werden vollständig von der das Datenarchiv betreibenden Institution getragen. Voraussetzung dafür ist die grundsätzliche Entscheidung einer zentralen, finanzstarken Einrichtung, ein Forschungsdatenzentrum nicht für die eigene Institution zu finanzieren und zu betreiben, sondern für eine ganze Fach-Community bereitzustellen und eine im Regelfall sowohl für Dateneigentümer und Datenproduzenten als auch Datennachnutzer kostenlose Nutzung von Dienstleistungen zu ermöglichen.

Als Beispiele zu nennen sind hier das Datenarchiv für die Sozialwissenschaften¹¹⁴ als Teil des Leibniz-Institutes für Sozialwissenschaften und PANGAEA. Data Publisher for Earth & Environmental Science¹¹⁵, gemeinschaftlich getragen von dem Alfred-Wegener-Institut für Polar- und Meeresforschung der Helmholtz-Gemeinschaft und dem Zentrum für Marine Umweltwissenschaften der Universität Bremen.

3. Mischfinanzierte Modelle

Bei diesem Modell werden Ansätze aus den beiden obigen „Reinformen“ in verschiedener Gewichtung miteinander kombiniert.

Beispielsweise werden bei DANS/eDNA – Data Archiving and Networked Services¹¹⁶ – in den Niederlanden zwei Drittel aller entstehenden Kosten durch zwei große Einrichtungen getragen, die Königliche Akademie der Niederlande (KNAW)¹¹⁷ sowie The Netherlands Organisation for Scientific Research (NWO)¹¹⁸. Die restlichen Ausgaben müssen durch Drittmittelprojekte und moderate Gebühren, die von Datenproduzenten eingefordert werden, finanziert werden.¹¹⁹

4. Finanzierungsmodell IANUS

Unter Berücksichtigung aller Faktoren aus den Begehungen, Stakeholder-Analyse, Einzelgesprächen und Praxisbeispielen wird auch für IANUS ein mischfinanziertes Finanzierungsmodell angestrebt, da dieses die Gesamtkosten auf mehrere Akteure verteilt und somit die finanzielle Belastung für den Einzelnen reduziert. Vor allem werden Datengeber, die Daten bei IANUS archivieren und veröffentlichen, nur in dem Maße an den Kosten beteiligt, wie sie für eine initiale Datenaufbereitung und anteilig für die Langzeitarchivierung entstehen. Auf diese Weise wird auch dem Wunsch der Mehrheit in der Fach-Community Rechnung getragen, für die angemessene Kosten ein wichtiger Faktor für die Bereitschaft zur Datenarchivierung und -bereitstellung darstellt.¹²⁰

Trotz der Unterschiede der verschiedenen Finanzierungsmodelle und der unterschiedlich hohen Beteiligung von Datenproduzenten an den Gesamtkosten, ist allen Ansätzen gemeinsam, dass die wissenschaftliche Nachnutzung von Daten für die Nachnutzer kostenfrei bleibt. Einen offenen und

¹¹⁴ <http://www.gesis.org/das-institut/wissenschaftliche-abteilungen/datenarchiv-fuer-sozialwissenschaften/> [Zugriff 21.11.2013]

¹¹⁵ <http://www.pangaea.de> [Zugriff 21.11.2013]

¹¹⁶ <http://www.dans.knaw.nl/en> [Zugriff 21.11.2013]

¹¹⁷ <https://www.knaw.nl/en> [Zugriff 21.11.2013]

¹¹⁸ <http://www.nwo.nl/en> [Zugriff 21.11.2013]

¹¹⁹ Vgl. IANUS. Übersicht Datenzentren & Fachsysteme (2013)

¹²⁰ IANUS. Stakeholderanalyse 1: Statistische Ergebnisse (2013) Frage 12.7

kostenfreien Zugang zu Forschungsdaten, besonders wenn sie mit öffentlichen Mitteln finanziert wurden, fordert unter anderem auch die Allianz der Wissenschaftsorganisationen ein.¹²¹ Auch IANUS verpflichtet sich dem Ziel eines kostenlosen Datenzugriffs für Datennachnutzer und auf dieser Prämisse beruhen die folgenden Finanzierungs- und Kostenmodelle.

Generell gilt zudem, dass IANUS die Entgelte für die von ihm zu erbringenden Dienstleistungen nach dem Kostendeckungsprinzip kalkuliert und als gemeinnützige Einrichtung ohne Gewinnerzielungsabsicht arbeitet. Außerdem wird das gesamte Finanzierungsmodell durch die Anbindung von IANUS an das DAI beeinflusst, das maßgeblich zur Erwirtschaftung der benötigten Finanzmittel und damit zur Reduzierung der Kosten für die Datengeber beiträgt.

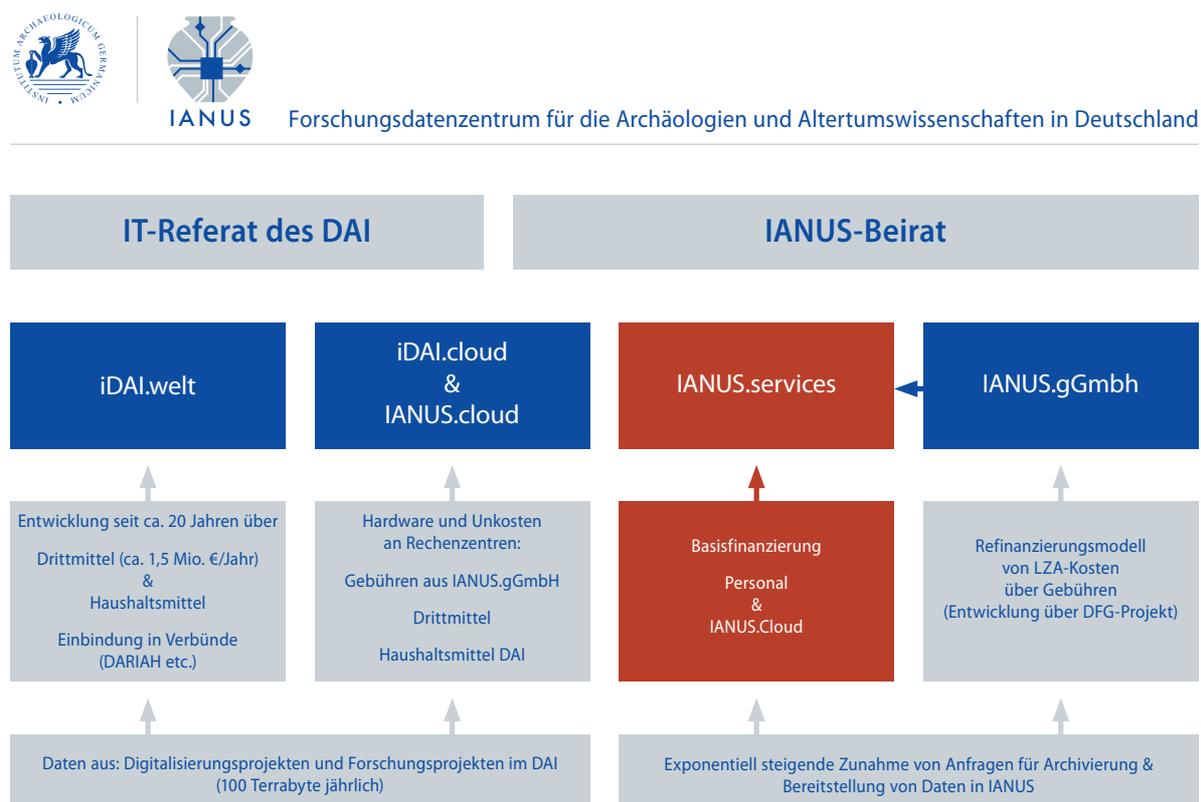


Abb. 6 Gemischtes Finanzierungsmodell für IANUS unter Berücksichtigung der Anbindung an das DAI

8.3.2 Säule – Basisfinanzierung

Die Idee einer Basisfinanzierung ist es, die Aufrechterhaltung der Verfügbarkeit und Funktion aller kritischen Geschäftsprozesse sicherzustellen, so dass bereits archivierte Forschungsdaten weiterhin langfristig kuratiert und zur Verfügung gestellt werden können – unabhängig davon, ob neue Forschungsdaten in das Archiv einfließen oder nicht. Zu den erforderlichen, langfristigen Aufgaben gehört das technische Kernsystem betriebsbereit zu halten, notwendige Updates einzuspielen, die Migrationen archivierter Daten in neue Formate zu überführen, aktuelle IT-Entwicklungen zu verfolgen und darauf zu reagieren sowie alle vertraglichen Verpflichtungen mit den

¹²¹ „Grundsätze zum Umgang mit Forschungsdaten“ der Schwerpunktinitiative "Digitale Information" in der Allianz der Wissenschaftsorganisationen: <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsätze/> [Zugriff 21.11.2013]

Datenproduzenten einzuhalten. Daneben gilt es auch im Rahmen der Basisfinanzierung, neue Projektvorhaben voranzutreiben und neue Forschungsdaten aktiv zu akquirieren, um die Einnahmeseite zu konsolidieren. Für diese Aufgaben fallen im Wesentlichen folgende Fixkosten an:

- Kernmannschaft an Mitarbeitern
- Miete und Einrichtung Arbeitsplätze
- hochperformanter Netzanschluss an das Deutsche Forschungsnetz (DFN)
- Bitstream-Preservation bei zwei technischen Partnern (Rechenzentren)
- Betrieb eines Webportales und Administration von Produktivsystemen
- Rückstellungen von Geldern im Rahmen einer Exit-Strategie
- verwaltungstechnische Aufgaben
- Öffentlichkeitsarbeit
- Sitzungen des Aufsichtsrats und des Beirats

Diese Kosten müssen vollständig mit einer Basisfinanzierung gedeckt werden, die je nach Einnahmen aus Gebühren und Drittmittelprojekte aufgestockt werden kann (siehe **Kapitel 7.3.3, 7.3.4**). Diese Basisfinanzierung sollte von einer oder mehrere Bundeseinrichtungen übernommen werden oder über eine gemeinsame Bund-Länderfinanzierung erfolgen, da IANUS von seinem nationalen Charakter und seiner Zielsetzung her als dauerhafte Einrichtung geplant wird und insofern von einer übergeordneten Instanz getragen werden sollte, die ihrerseits fest etabliert ist.¹²² Auch der Wissenschaftsrat rät in seiner Stellungnahme zur Evaluierung des Deutschen Archäologischen Institutes vom Oktober 2015 die Grundfinanzierung (Säule – Basisfinanzierung) von IANUS durch die zuständigen Bundesministerien langfristig sicherzustellen.¹²³

Für einen Teil der genannten Kosten, insbesondere die für die Bitstream Preservation notwendigen Hardware- und Speicherkomponenten, kann die Basis-Finanzierung durch das DAI übernommen werden, da es für diesen Bereich eine langjährige Kooperation mit dem RRZK in Köln etabliert hat und notwendige Mittel bereits im Haushalt des DAI eingeplant sind. Für den größten Teil, die Personalkosten erscheinen zum gegenwärtigen Zeitpunkt vor allem einige Bundesministerien als primäre Mittelgeber geeignet. Trotz der hohen Relevanz für die Forschung ist eine unmittelbare finanzielle Beteiligung durch die DFG nicht realistisch, da diese nur zeitlich befristete Projekte oder Infrastrukturmaßnahmen fördert.

Eine feste finanzielle Beteiligung könnte derart ausgestaltet sein, dass die Basisfinanzierung langfristig unter dem Vorbehalt der Mittelverfügbarkeit garantiert wird, wobei im Kontext von IANUS unter „langfristig“ ein Zeitraum von mindestens 30–50 Jahren verstanden wird.¹²⁴ Als finanzieller Planungszeitraum wird ein Fünf-Jahres-Turnus vorgeschlagen, wie er auch bei vergleichbaren Einrichtungen mit 4–7 Jahren üblich ist.¹²⁵ Für diesen Zeitraum definiert der Aufsichtsrat die strategische Ausrichtung und die Ziele und garantiert die zur Erfüllung der Aufgaben notwendigen Mittel. Vor Ablauf eines aktuellen Planungszeitraumes muss den

¹²² Vgl. IANUS. SCHWIEGELSHOHN (2013) S. 2. 11–12. 31–32.

¹²³ Vgl. WR Drs. 4907-15, S. 15; <http://www.wissenschaftsrat.de/download/archiv/4907-15.pdf>

¹²⁴ Vgl. Votum der DFG-AG beim 6. Arbeitstreffen 2013

¹²⁵ Vgl. IANUS. Übersicht Datenzentren & Fachsysteme (2013).

Mittelgebern über die erreichten und nicht erreichten Ziele Rechenschaft abgelegt werden, was wiederum den folgenden Planungszeitraum beeinflussen kann. Wenn die Basisfinanzierung als Ganzes oder zu einem Großteil wegfällt und es keine adäquate Ersatzfinanzierung gibt, IANUS also in seiner Existenz gefährdet ist, greift die Exit-Strategie (siehe **Kapitel 11.6**).

8.3.3 Säule – Gebühren

Datenproduzenten, die ihre Daten an IANUS zur Langzeitarchivierung übergeben, sollen mittels eines festgelegten und transparenten Gebührenmodells einmalig und anteilig an den Kosten zur Aufbereitung, Archivierung und Bereitstellung ihrer Daten beteiligt werden.¹²⁶ Dabei wird zwischen fallbezogenen Gebühren für einzelne Projekte und pauschalen Gebühren für Institutionen unterschieden. Die Höhe der jeweiligen Gebühren hängt maßgeblich vom individuellen Kuratierungsaufwand der Daten ab, der wiederum durch folgende Faktoren beeinflusst wird:

- Art der Datenübergabe
- Anzahl und Größe der Dateien
- Heterogenität der verwendeten Datenformate
- Komplexität und Strukturierung der Daten
- Qualität der Dokumentation der Daten
- Gewünschte Art der Präsentation
- Mitwirkung der Datengeber beim Pre-Ingest- und Ingest-Prozess
- Verwaltung und Management

Bei der Kalkulation sind vor allem maßgeblich die initialen und einmaligen Aufwände der Kuratierung zu berücksichtigen. Bei sehr komplexen und heterogenen Datensammlungen ist darüber hinaus angedacht, die Mehrkosten für die dauerhafte Kuratierung und Bereitstellung partiell ebenfalls bei den Gebühren anzurechnen. Generell werden diese also aufwandsabhängig erhoben.¹²⁷

Die Gebühren müssen in einem regelmäßigen Turnus überprüft werden, um sie an praktische Erfahrungen sowie an veränderte Rahmenbedingungen anzupassen. Bei der Erhebung und Kalkulation von Gebühren sind die langjährigen Erfahrungen des Archaeology Data Service in UK ein wichtiger Anhaltspunkt für IANUS¹²⁸. Auch wenn die Zahlen nicht als absolute Preise übertragbar sind, z. B. allein aufgrund anderer Personalkosten, stellt die Zusammensetzung und die Gewichtung der einzelnen Faktoren eine wertvolle Grundlage dar, um von Beginn an realistische Preise anzusetzen.

1. Projektbezogene Gebühren

Die Grundidee der projektbezogenen Gebühren ist, dass die unmittelbaren Dienstleistungen, die

¹²⁶ Vgl. WR Drs. 4907-15, S.15 (<http://www.wissenschaftsrat.de/download/archiv/4907-15.pdf>).

¹²⁷ Vgl. die im Oktober 2015 veröffentlichten „Leitlinien zum Umgang mit Forschungsdaten“ der DFG (http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf), nach der eigene Mittel für die Archivierung und Bereitstellung von Forschungsdaten beantragt werden können.

¹²⁸ <http://archaeologydataservice.ac.uk/advice/chargingPolicy>

von IANUS erbracht werden, wenn ein Datengeber – zumeist am Ende eines Projektes – einen abgeschlossenen Datenbestand an IANUS übergibt, um diese Daten dort archivieren und online zur Verfügung stellen zu lassen, vom Datengeber mitfinanziert werden. Dabei setzt sich IANUS dafür ein, dass diese zusätzlichen, einmaligen Projektkosten („pay once, store forever“), vergleichbar wie die Pauschalen für Publikationen, bei einem Drittmittelantrag geltend gemacht werden können.

Neben der Bereitschaft der Drittmittelgeber derartige Ausgaben in die Projektförderung aufzunehmen (bei der DFG können seit Oktober 2015 Kosten zur Kuratierung und Bereitstellung von Forschungsdaten regulär beantragt werden)¹²⁹, setzt dieses Ziel auch eine frühzeitige Kontaktaufnahme bei neuen Projektvorhaben mit IANUS voraus, um bereits im Vorfeld die voraussichtlichen Kosten überschlagen und entsprechend in einem Antrag einplanen zu können¹³⁰. Die Beantragung und auch die Nutzung bewilligter Gelder zur Archivierung und Veröffentlichung von Forschungsrohdaten würde in besonderer Weise den Empfehlungen für eine gute wissenschaftliche Praxis, wie sie von der DFG formuliert wurden¹³¹, Rechnung tragen – unabhängig davon, ob dies über IANUS oder ein anderes professionelles Archiv erfolgt.

2. Institutsbezogene Gebühren

Da diese fall- bzw. projektbezogene Beantragung der Kuratierungs- und Archivierungskosten nicht für alle Akteure in den Altertumswissenschaften gleichermaßen adressierbar ist, soll darüber hinaus für Institutionen, die an bestimmten Dienstleistungen von IANUS regelmäßig oder längerfristig interessiert sind, die Möglichkeit geschaffen werden, Verträge mit pauschalen Gebühren auszuhandeln. Für die Inanspruchnahme von Dienstleistungen im Bereich der Datenkuratierung und –archivierung dienen als Bemessungsgrundlage die Datenmenge, die Datenarten und die Anzahl der zu kuratierenden Projekte, die für einen festgelegten Zeitraum (als Regellaufzeit werden 2 Jahre vorgeschlagen) erwartet werden. Bei anderen Dienstleistungen kommen dagegen andere Kriterien zum Tragen.

Dieses Instrumentarium der institutionellen Partnerschaft hilft IANUS dabei, seinen Betrieb langfristiger als bei Einzelprojekten mit projektbezogenen Gebühren zu planen und so die vorhandenen Ressourcen effizienter einzusetzen. Umgekehrt können beteiligte Institutionen unabhängig von dem jeweiligen Projektaufkommen feste Beträge einplanen und es kann eine durchgängige Unterstützung beim Datenmanagement angeboten werden. Darüber hinaus können sie sukzessive interne Kompetenzen bezüglich einer optimierten Übergabe von Daten an IANUS aufbauen bzw. IANUS kann umgekehrt sich auf die Charakteristika von Daten einzelner Institutionen einstellen. Unter der Annahme, dass auf diese Weise sich Synergieeffekte ergeben und mittelfristig der Zeitaufwand bei IANUS für die Kuratierung institutioneller Daten reduziert werden kann, können den Gebühren niedrigere Ansätze zugrunde gelegt werden und dadurch fallen diese bezogen auf den Durchschnitt je Projekt günstiger aus als bei individuellen Projektabwicklungen.

¹²⁹ http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

¹³⁰ Vgl. nestor. Bestandsaufnahme (2012) S. 194.

¹³¹ http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf [Zugriff: 14.11.2013]

Für den Fall, dass eine Institution auf diese Weise in überdurchschnittlichem Maße einen finanziellen Beitrag zu IANUS leistet und sich auf den effektiven Betrieb von IANUS verlässt, ist es vorgesehen, einen Vertreter der Institution als institutionelle Partner mit einfachem Stimmrecht in dem Aufsichtsrat zu beteiligen. Während der Laufzeit eines Vertrages bekommt eine Institution somit die Möglichkeit, in IANUS mitzuwirken und strategische Entscheidungen mit zu treffen.

8.3.4 Säule – Drittmittelprojekte

Die letzte Säule umfasst alle Drittmittel, welche von IANUS als Antragsteller eingeworben werden, sowie alle Projekte, an welchen IANUS in jeglicher Form und Ausprägung beteiligt ist. Durch diese zusätzlichen Einnahmen sollen verschiedene Ziele erreicht werden, die sich alleine aus einer Basisfinanzierung und den Gebühren nicht oder nur eingeschränkt realisieren lassen:

- Unterstützung von Drittmittelprojekten beim Forschungsdatenmanagement
- Durchführung von Schulungen, Tutorials, Workshops, Summer Schools
- größere Aktualisierungen der IT-Empfehlungen
- umfangreichere Anpassungen, Updates und Weiterentwicklungen der technischen Systeme und Schnittstellen
- Kuratierung, Archivierung und Bereitstellung von Alt-Daten (siehe Kapitel 5.2), ggf. mit Retrodigitalisierung von Materialien
- Austausch/Interoperabilitätsszenarien von Daten mit anderen Datenzentren und IT-Infrastrukturen
- wissenschaftliche Analyse und Erschließung der archivierten Daten¹³²

Der Ansatz, zur Bewältigung von einmaligen größeren Aufgaben, wie etwa die Weiterentwicklung eines Repository-Systems oder die Übernahme und Aufbereitung von großen, teils unerschlossenen Forschungsdatenbeständen Drittmittel einzuwerben, ist in existierenden Datenarchiven national (z. B. PANGAEA) wie international (z. B. MPI-TLA) üblich und hat sich vielfach bewährt.¹³³

Als Beispiel für große, von der EU geförderte Infrastrukturprojekte, bei welchen IANUS einen Beitrag leisten könnte, sind DARIAH, CLARIN, CARARE oder ARIADNE zu nennen. Neben der inhaltlichen wie auch finanziellen Notwendigkeit der Beteiligung an solchen Vorhaben hilft es darüber hinaus, die Sichtbarkeit der altertumswissenschaftlichen Fachgemeinschaft innerhalb der Geisteswissenschaften bzw. der Digital Humanities zu erhöhen, deren kulturpolitische Relevanz zu stärken und auf deren besondere Bedürfnisse hinzuweisen.

8.4 Kostenmodell

Die konkreten Kosten der Langzeitarchivierung sind jeweils abhängig von

- Anzahl und Komplexität der Workflows in einer Institution
- der Menge, Heterogenität und Komplexität der zu archivierenden Objekte

¹³² Vgl. WR Drs. 4907-15, S.102; <http://www.wissenschaftsrat.de/download/archiv/4907-15.pdf>

¹³³ Vgl. nestor. Bestandsaufnahme (2012) S. 307.

- dem Aufbereitungsgrad der Metadaten
- den gewünschten Zugriffsmöglichkeiten und Schnittstellen
- den Anforderungen Dritter an die archivierende Institution
- den Verpflichtungen der archivierenden Institution gegenüber Dritten¹³⁴

Die Erstellung eines belastbaren Kostenmodells für die Langzeitarchivierung ist eine Herausforderung, da allein die technischen Änderungen hinsichtlich Datenformaten, Systemarchitekturen, Software etc. nicht vorhersehbar und insofern kaum kalkulierbar sind. Wenn nicht alle potentiellen Kostenfaktoren identifiziert wurden, muss das jeweilige Modell sukzessiv und zeitnah im laufenden Betrieb angepasst werden, was immer noch kein Garant für ein funktionierendes Modell ist, aber den Prozess reeller abbildet als vorher und somit die Wahrscheinlichkeit eines belastbareren Modells erhöht.

Dennoch gibt es einige internationale Versuche, das Vorhaben der Entwicklung von Kostenmodellen für die Langzeitarchivierung umzusetzen.¹³⁵ Auch mehrere durch die DFG geförderte Projekte kommen im Rahmen eines Symposium zu Forschungsdaten-Infrastruktur zu dem Schluss, dass u.a. bei Kosten- und Preismodellen im Bereich der Langzeitarchivierung noch ein erheblicher Forschungsbedarf besteht.¹³⁶

Die folgende Auflistung gibt einen Überblick, welche Kostenmodelle für IANUS relevant sind¹³⁷:

- LIFE 3 – Digital preservation costing, and develop costing models¹³⁸
- KRDS 2 – Cost/benefit studies, tools and methodologies focussing on long-lived data¹³⁹
- DANS – costing model for long-term preservation and dissemination of digital research data¹⁴⁰
- CMDP – a new Cost Model for Digital Preservation at the Royal Danish Library & the Danish National Archives¹⁴¹
- DP4lib – a business and cost model for a digital preservation service¹⁴²
- CDL – California Digital Library Cost Modelling¹⁴³

¹³⁴ nestor. Enzyklopädie (2009) Kap. 14:4

¹³⁵ P. Wheatley et al., Digital Preservation and Data Curation Costing and Cost Modelling: <http://wiki.opf-labs.org/display/CDP/Home> [Zugriff 21.11.2013]

¹³⁶ Vgl. Radieschen. Symposium FDI (2013) S. 13

¹³⁷ P. Wheatley et al., Digital Preservation and Data Curation Costing and Cost Modelling: <http://wiki.opf-labs.org/display/CDP/Home> [Zugriff 21.11.2013]

¹³⁸ LIFE - Life Cycle Information for E-Literature, Kooperationsprojekt der UCL und British Library: <http://www.life.ac.uk/> [Zugriff 22.11.2013]

¹³⁹ Ch. Beagrie, Keeping Research Data Safe: <http://www.beagrie.com/krds.php> [Zugriff 22.11.2013]

¹⁴⁰ An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS (2012): doi:10.1007/s00799-012-0092-1 [Zugriff 22.11.2013]

¹⁴¹ Cost Model for Digital Preservation (CMDP): Project at the Royal Danish Library and the Danish National Archives: <http://www.costmodelfordigitalpreservation.dk/contact> [Zugriff 22.11.2013]

¹⁴² DP4lib – Digital Preservation for Libraries: http://dp4lib.langzeitarchivierung.de/index_downloads.php.de [Zugriff 25.11.2013]

¹⁴³ CDL – Cost Modelling <https://wiki.ucop.edu/display/Curation/Cost+Modeling> [Zugriff 25.11.2013]

- ADS – Charging Policies, ADS-easy a cost tool for archiving¹⁴⁴

Nach zukünftiger Prüfung und Evaluierung wird, sofern das jeweilige Modell kompatibel ist und an die Bedürfnisse von IANUS angepasst werden kann, die Entscheidung für ein Modell getroffen werden. Bereits jetzt wird aber bei den verschiedenen Ansätzen zur Kostenermittlung deutlich, dass die Datenkuratierung und -archivierung zeitaufwendig und somit kostenintensiv sind. Dies bestätigen auch die Erfahrungen in existierenden Datenarchiven. So gibt z. B. der Archaeology Data Service in York den Kostenanteil der Datenkuratierung inklusive der aktiven Akquise mit knapp 50% an den Gesamtkosten an, bei PANGAEA sind es sogar knapp zwei Drittel.¹⁴⁵ Auch N. Beagrie kommt in seinem Report „Keeping Research Data Safe 2“ zu dem Schluss, dass die Kuratierung von Forschungsdaten mit 45% an den Gesamtkosten den kostenintensivsten Part der Langzeitarchivierung darstellt.¹⁴⁶

8.4.1 Kostenfaktoren

Zunächst werden die Kosten für den Aufbau und den Betrieb von IANUS nach ihrer Entstehungsart unterschieden. Diese lassen sich in die folgenden Arten unterteilen:¹⁴⁷

- Hardware: Server / Festplatten-Cache (Server) / Speicherplatz / PC-Arbeitsplätze für Mitarbeiter / Neukonfiguration bestehender Systeme
- Software: Lizenzen für kommerzielle Software / Online-Datenportal für den Access / Archive-Management-System / Betriebssystem für Server und Client-Rechner / Systeme für Email, Tickets, Dokumentation / Monitoring-System(e)
- Personal: Geschäftsleitung / Projektmanager / Softwareentwickler / Datenkuratoren / Technischer Support / Assistenz (Verwaltung, Buchhaltung) / Öffentlichkeitsarbeit / Projektakquise
- Räume: Büros der Mitarbeiter / Besprechungsräume / Technik-Serverräume / Betriebskosten
- externe Dienstleistungen: Support-Kosten / SLA für Bitstream-Preservation / Infrastrukturdienste (z. B. Anschluss an DFN) / Versicherungen
- Sonstige: Services / Zertifizierungen / Materialien Öffentlichkeitsarbeit

Es gibt drei Arten von Kosten die jeweils zu unterschiedlichen Zeitpunkten, die o.g. Kosten verursachen: Initiale Kosten, Beschaffungskosten, Betriebskosten.¹⁴⁸ Da die Initialen Kosten, also die Informationsbeschaffung über LZA-Systeme, Projektplanung, Erhebung über Technik, Datenbestand, Zugang und Personalbedarf usw. für IANUS bereits in der ersten bzw. zweiten Phase der DFG-Förderung anfielen, sind sie im Rahmen dieses Konzepts zu vernachlässigen und werden nicht weiter betrachtet.

¹⁴⁴ ADS – Archaeology Data Service, charging policies, 2013, online abrufbar unter: <http://archaeologydataservice.ac.uk/advice/chargingPolicy> [Zugriff 2013-11-25]

¹⁴⁵ Vgl. IANUS. Übersicht Datenzentren & Fachsysteme (2013).

¹⁴⁶ Vgl. Beagrie, N et al., Keeping research data safe 2, Bristol, 2010, Joint Information System Committee (JISC): <http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx> [Zugriff 21.11.2013]

¹⁴⁷ DP4lib – Digital Preservation for Libraries: http://dp4lib.langzeitarchivierung.de/index_downloads.php.de [Zugriff 25.11.2013]

¹⁴⁸ Vgl. NESTOR. ENZYKLODÄDIE (2009) Kap. 14:3 f.

1. Beschaffungskosten – einmalige Kosten

- Arbeitsplätze für Personal (Möbel, PC, Literatur, Geschäftsausstattung etc.)
- Beschaffung von Hardware für ein Produktiv-Systeme (Server, Router, Speicher etc.)
- Softwarelizenzen für Client-Rechner und Server
- Anpassung bestehender Software-Produkte / Eigenentwicklung
- Einstellung neuer Mitarbeiter und/oder Schulungen vorhandener Mitarbeiter
- ggf. Liegenschaft vorbereiten (Brandschutz, Arbeitsschutz, LAN/W-LAN etc.)

2. Betriebskosten – laufende Kosten – jährlich

- Personalkosten (Gehälter, Sonderzahlungen)
- Reisekosten der Mitarbeiter
- Öffentlichkeitsarbeit (Teilnahme an und Durchführung von Veranstaltungen)
- Services externer Partner (Bitstream Preservation, Storage, Datennetz)
- Versicherungen
- Büromaterial, Geschäftspapiere
- laufende Lizenzkosten für Software
- laufende Beitragszahlungen bei Konsortien und Verbänden
- regelmäßige Sicherheitsmaßnahmen (Backups, Wartung, Tests, Software-Upgrades)
- Rückstellungen von Mitteln für Exit-Strategie
- Miete und Nebenkosten

3. Betriebskosten – laufende Kosten – mehrjährig

- Weiterbildungen / Fortbildung der Mitarbeiter
- Hard- und Software-Komplettersatz in Intervallen
- Zertifizierung / Begutachtungen des Datenarchives

8.4.2 Gesamtkosten

Vorbehaltlich der ausstehenden finalen Beurteilung und Entscheidung für ein Kostenmodell können nach jetzigem Kenntnisstand und der in **Kapitel 8.4.1** aufgeführten Faktoren die folgenden Gesamtkosten für den Betrieb von IANUS mit gewissen Unschärfen kalkuliert werden. Die vorliegenden Zahlen sind als vorsichtige Schätzungen auf Basis der Erfahrungen an vergleichbaren Einrichtungen zu verstehen, um einen ersten Eindruck des Finanzvolumens zu vermitteln. Ihre Höhe muss nach der Evaluierung der Kostenmodelle angepasst und damit belastbar kalkuliert werden. Ihnen liegt die Annahme zugrunde, dass bei dem Kernbetrieb von IANUS insgesamt 10 FTE mitarbeiten und bezahlt werden (siehe **Kapitel 9.3**). Je nach dem Aufkommen der Einnahmen aus Gebühren und eingeworbenen Drittmitteln können weitere Personalstellen und damit auch personenbezogene Sachkosten hinzukommen, die dann größtenteils durch diese Sondermittel selbst finanziert, ggf. aber auch durch Eigenanteile unterstützt werden müssen.

1. Beschaffungskosten – einmalige Kosten

Arbeitsplätze (für 10 FTE)	50.000 €
Hardware Produktiv-Systeme	20.000 €
Softwarelizenzen für Client-Rechner und Server	20.000 €
Personaleinstellungen und -schulungen	15.000 €
Softwareentwicklung (Repository-System, Web-Portal etc.)	200.000 €
Liegenschaft vorrichten	100.000 €
<u>Summe</u>	<u>405.000 €</u>

2. Betriebskosten – laufende Kosten – jährlich

Personal (10 FTE)	654.130 €
Reisekosten der Mitarbeiter (insgesamt 20 Reisen á 400 €)	8.000 €
Öffentlichkeitsarbeit / Veranstaltungen	25.000 €
Versicherungen	1.000 €
laufende Lizenzkosten für Software	1.000 €
laufende Beitragszahlungen bei Konsortien, Verbänden	500 €
Büromaterial, Geschäftspapiere	4.000 €
Services externer Partner	2 x 52.000 €
Rückstellungen Exit-Strategie	25.000 €
Miete und Nebenkosten	25.000 €
<u>Summe</u>	<u>847.130 €</u>

3. Betriebskosten – laufende Kosten – mehrjährig

Weiterbildungen Personal (alle 2 Jahre)	15.000 €
Software-Komplettersatz in Intervallen (alle 3–5 Jahre)	20.000 €
Hardware-Komplettersatz in Intervallen (alle 4–5 Jahre)	70.000 €
Hardware-Austausch bei technischen Partnern (alle 5 Jahre)	2 x 260.000 €
Zertifizierung Datenarchiv/Prozesse (alle 5 Jahre)	1.000 €
<u>Summe (ca. pro Jahr)</u>	<u>130.700 €</u>

9 Personaltableau

IANUS benötigt für die Erfüllung der in **Kapitel 6** genannten Aufgaben und Dienstleistungen eine bestimmte Anzahl an Personen mit unterschiedlichen Spezialisierungen, um insbesondere die Archivierungs- und Bereitstellungsaufgaben dauerhaft erfüllen zu können. Für den Kernbetrieb wird in Analogie zu den beiden vergleichbaren fachspezifischen und bereits etablierten Forschungsdatenzentren ADS (UK) und DANS/eDNA (NL)¹⁴⁹ von 10 FTE ausgegangen. Diese Zahl der Mitarbeiter ist als statisch und minimal anzusehen, da die zu leistenden Aufgaben vor allem die langfristige Verfügbarkeit und Nachnutzung adressieren und unabhängig vom konkreten Volumen bereits archivierter oder zu kuratierender Daten sind. Der Wissenschaftsrat kommt ebenfalls in seiner Stellungnahme vom Oktober 2015 zu dem Fazit, „[...] dass für den dauerhaften Betrieb von IANUS nach dem Ende der Projektförderung [...] Dauerstellen am DAI verankert werden sollten. Insgesamt ist hier von einem Bedarf von zehn Stellen unterschiedlicher Wertigkeit auszugehen.“¹⁵⁰

Die Konzeption der beiden weiteren Säulen kann und muss im Gegensatz zur 1. Säule dynamisch an die Bedürfnisse bzw. die aktuellen Gegebenheiten angepasst werden. Die zweite über Gebühren finanzierte Säule hängt von der Anzahl der zu kuratierenden Projekte und deren Umfang, Komplexität, Qualität der Dokumentation und gewünschter Präsentation ab. Aufgrund der Personalstrukturen in anderen Langzeitarchiven erscheinen hier etwa 2-3 weitere FTEs (bzw. äquivalente Aufstockungen des Personals des Kernbetriebes) realistisch zu sein. Sofern weniger Datensammlungen zur Langzeitarchivierung anvertraut werden, muss die Anzahl bzw. der Beschäftigungsumfang der Mitarbeiter entsprechend flexibel reduziert werden. Die dritte Finanzierungssäule durch eingeworbene Drittmittelprojekte wird durch die Ziele, Dauer und Größen der bewilligten Vorhaben bestimmt. Auch hier ist nach der Aufbauphase mit etwa 2–3 zusätzlichen FTEs zu rechnen.

In der Theorie könnte man bei genügend großem Bedarf einfach kurzfristig so viele Mitarbeiter einstellen bzw. wieder entlassen wie die Projektlage aktuell erfordert. Dies scheitert allerdings nicht nur an der tatsächlichen Verfügbarkeit von entsprechend ausgebildeten Datenkuratoren, sondern ist auch in der Praxis anderer Langzeitarchive unüblich, zieht einen proportional großen Verwaltungsaufwand nach sich und ist für einen langfristigen und nachhaltigen und Betrieb eines Forschungsdatenzentrums nicht zielführend.

9.1 Personalanforderungen

Der Betrieb eines Forschungsdatenzentrums mit den Kernaufgaben der Kuratierung, Archivierung und Bereitstellung von Forschungsdaten hat die folgenden Anforderungen an das Personal.

9.1.1 Leitung und Organisation

Für die wissenschaftliche und technische Gesamtentwicklung von IANUS wird eine Führungsperson benötigt, die ihr Handeln einerseits mit den Lenkungsorganen der Fach-Community und andererseits aufgrund der institutionellen Anbindung an das DAI mit dessen

¹⁴⁹ Vgl. IANUS. Übersicht Datenzentren & Fachsysteme (2013).

¹⁵⁰ WR Drs. 4907-15, S.18; <http://www.wissenschaftsrat.de/download/archiv/4907-15.pdf>

Leitungsebenen (IT-Direktor, Generalsekretär, Präsidentin) abstimmt. Zugleich fungiert sie aber auch als Geschäftsführer der gGmbH, die als juristisch sinnvolle Rechtsform zur Einnahme von Nutzungsentgelten empfohlen wurde. (siehe **Kapitel 8.1**) Die Person führt die fachlichen und juristischen Verhandlungen mit (institutionellen) Datengebern, anderen Datenzentren und technischen Partnern (z. B. bzgl. neuer Services), entwickelt Ideen für neue Dienstleistungen und Projekte, unterstützt die Akquise von Drittmitteln, trägt die Verantwortung für die Umsetzung von Beschlüssen und Zielvorgaben der Lenkungsorgane und überwacht die korrekte Durchführung der Arbeitsprozesse.

Unterstützt und vertreten wird sie durch eine weitere Leitungsperson, die für die praktisch-organisatorische Abwicklung des Regelbetriebes und von Projekten zuständig ist. Insofern obliegt ihr die Abwicklung der Datenübergabeverträge mit (individuellen) Datengebern und die Überwachung der Einhaltung von Lizenzen, das Finanzcontrolling mit dem kontinuierlichen Monitoring der Kosten sowie die Anpassung des Gebührenmodells. Weiterhin führt sie Maßnahmen zur Qualitätssicherung (z. B. Zertifizierungen, Stakeholder-Analysen) durch und steuert die Weiterentwicklung der Ausbildungs- und Qualifizierungsangebote von IANUS.

Beide Personen teilen sich Aufgaben für Berichte, Publikationen, Öffentlichkeitsarbeit und Präsentationen auf Veranstaltungen. Für die administrative und finanztechnische Abwicklung der Nutzungsentgelte wird eine Verwaltungsfachkraft benötigt.

9.1.2 Datenkuratierung und fachspezifische Dienstleistungen

Die professionelle Aufbereitung digitaler Forschungsdaten für eine langfristige Nachnutzung erfolgt durch Datenkuratoren, die sowohl fachwissenschaftlich, als auch informationstechnisch geschult sein müssen. Diese kommunizieren mit den Datengebern, entwickeln eigenverantwortlich für jede Datensammlung eine Erhaltungs- und Präsentationsstrategie, bewerten die wissenschaftliche Qualität und Quantität der Dokumentation von zu archivierenden Daten, kontrollieren die Einhaltung von Vorgaben und Vereinbarungen (z. B. zu Zugriffsbeschränkungen), führen die technischen Umwandlungen gemäß OAIS-Standard durch und dokumentieren alle Prozesse in einem Archiv-Management-System. Ferner sollen sie in einem angemessenen Umfang Projekte beim Forschungsdatenmanagement beraten und unterstützen sowie neue Metadaten vor einer Übernahme in den geplanten Nachweiskatalog validieren. Aufgrund der Heterogenität altertumswissenschaftlicher Daten werden Mitarbeiter mit unterschiedlichen, in einem informatischen Hochschulstudium erworbenen Spezialisierungen benötigt, die ihr jeweiliges Wissen dann auch in die IT-Empfehlungen einbringen können.

Ein separates Aufgabenfeld betrifft die Erstellung und Durchführung von Schulungs- und Weiterbildungsangeboten sowie eine effektive Öffentlichkeitsarbeit, die die kontinuierliche Pflege der Social-Media-Kanäle, die Erstellung von analogen und digitalen Publikationen, die Durchführung von Nutzerbefragungen und das Testen neuer Benutzeroberflächen mit einschließt.

9.1.3 Technische Aufgaben

Für die Pflege, Fehlerbehebung und technische Weiterentwicklung der verschiedenen Produktivsysteme (Archiv-Management-System, Online-Datenportal, Migrations-Tools, Nachweiskatalog, Wikis, Ticket-System etc.) werden mehrere Softwareentwickler und ein

Systemadministrator benötigt. Auch nach einer initialen Fertigstellung der Anwendungen müssen diese regelmäßig aktualisiert, auf neue Software-Technologien reagiert, neue funktionale Anforderungen umgesetzt und technische Einstellungen dokumentiert werden. Ebenso ist die Kontrolle von Schnittstellen, über welche Daten mit externen Systemen ausgetauscht werden, eine Daueraufgabe. In Abstimmung mit dem IT-Referat des DAI erfolgt zudem die Evaluierung, Beschaffung und Installation neuer Hard- und Software. Eine besondere Aufgabe besteht darin, die Storage-Systeme bei und Schnittstellen zu externen Partnern zu kontrollieren und die Einhaltung von SLA-Vereinbarung zu überprüfen. In Intervallen sind außerdem die Migrationen von Archivbeständen notwendig.

9.1.4 Allgemeine Anforderungen

Die Auflistung dieser Tätigkeitsprofile verdeutlicht, dass die Anforderungen insbesondere an das Personal für die Softwareentwicklung sowie die Kernaufgaben (Kuratierung und Archivierung) vergleichsweise hoch sind. Damit die Datenkuratoren die Inhalte und Dokumentation der Daten verstehen und formal bewerten können, sind fachspezifische Kenntnisse erforderlich, weshalb hierfür ausgebildete Wissenschaftler benötigt werden. Diese müssen darüber hinaus fundierte IT-Kenntnisse besitzen bzw. in der Lage sein, sie sukzessive zu erwerben. Der Bedarf an derart spezialisierten Arbeitskräften ist groß, gleichwohl es momentan ein Defizit in der Ausbildung solcher dualen Fachkräfte gibt. Schon 2011 hat der Wissenschaftsrat in seinen Empfehlungen zu übergreifenden Informationsstrukturen festgestellt, dass „die Weiterentwicklung der Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften [...] in Deutschland nicht zuletzt durch einen Mangel an Expertinnen und Experten mit entsprechenden Doppelqualifikationen gehemmt“ ist.¹⁵¹

Auch andere Datenzentren haben Schwierigkeiten, geeignetes Personal zu akquirieren. So geben die meisten Datenzentren an, dass ihr Personal die benötigten Qualifikationen nicht durch eine Ausbildung, sondern erst durch praktische Tätigkeiten erworben hat.¹⁵² Auch IANUS hat Ende 2015 seine drei neue Datenkuratoren durch einen zweiwöchigen Trainingsaufenthalt beim ADS in York schulen lassen, um dort die notwendige praktische Expertise zu erlangen.

Darüber hinaus stellt es für Informationsstrukturen eine besondere Herausforderung dar, wissenschaftliches Personal auf dem Bewerbermarkt attraktiv zu adressieren. Der Wissenschaftsrat empfiehlt daher, auch um den Abstand zur privaten Konkurrenz um qualifiziertes Personal zur verringern, die zahlenmäßig und vor allem hinsichtlich der Eingruppierung ausreichende Ausstattung mit wissenschaftlichem Personal dauerhaft im Rahmen der Grundfinanzierung mit einzukalkulieren.¹⁵³ Aufgrund der langen Einarbeitungszeit wird die Übertragung dieser Aufgaben an wissenschaftliche Hilfskräfte mit kürzeren, befristeten Arbeitsverträgen problematisch gesehen. Für die Stellen der Datenkuratoren als auch der Softwareentwicklung wird daher eine Eingruppierung in TVöD E13 vorgeschlagen und die folgenden Personalkosten auf dieser Basis berechnet. Neben der notwendigen finanziellen Ausstattung für hochqualifizierte Stellen

¹⁵¹ WISSENSCHAFTSRAT (10466-11) S. 156

¹⁵² Vgl. nestor. Bestandsaufnahme (2012) S. 308

¹⁵³ Vgl. WISSENSCHAFTSRAT (10466-11) S. 50

empfiehlt der Wissenschaftsrat darüber hinaus, das Personal durch eigene Forschungstätigkeiten, Publikationen und/oder Lehraufträgen zu motivieren und es somit langfristig an eine Informationsinfrastruktur zu binden.¹⁵⁴ Solche Möglichkeiten für den Einzelnen ziehen einerseits die inhaltliche Weiterentwicklung der Infrastruktureinrichtung als Ganzes nach sich und erhöhen andererseits dessen Sichtbarkeit und Wahrnehmung in der wissenschaftlichen Fachgemeinschaft.

9.2 Personalkosten

9.2.1 Personal Basisfinanzierung

Das durch die erste Säule – die Basisfinanzierung (siehe **Kapitel 7.3.2**) – finanzierte Personal für den Kernbetrieb beläuft sich auf neun Vollzeit- und zwei Teilzeitstellen:

• Wissenschaftlich-technische Führung (Geschäftsführer gGmbH)	1 FTE	TVÖD-E14	71.830 €
• Organisatorische Leitung	1 FTE	TVÖD-E14	71.830 €
• Verwaltung/Buchhaltung (gGmbH)	½ FTE	TVÖD-E09	24.690 €
• Datenkuratoren mit Spezialisierung (Geodaten, Raster-/ Vektordaten, Datenbanken/Tabellen, 3D-/ Messdaten)	4 FTE	TVÖD-E13	66.270 €
• Schulungen & Öffentlichkeitsarbeit	1 FTE	TVÖD-E11	59.570 €
• Software-Entwickler	2 FTE	TVÖD-E13	66.270 €
• System-Administrator	½ FTE	TVÖD-E10	28.590 €
		<u>Summe</u>	<u>654.130 €</u>

9.2.2 Personal Gebühren

Die zweite Säule – Gebühren (siehe **Kapitel 8.3.3**) – umfasst dann in Abhängigkeit zur Verfügung stehender Mittel aus Gebühreneinnahmen das folgende weitere Personal. Die Erfahrungen der Kollegen beim Archaeology Data Service in York zeigen, dass ein erfahrener Datenkurator im Schnitt pro Monat eine durchschnittlich anspruchsvolle und umfangreiche Datensammlung kuratiert.¹⁵⁵ Hinzukommen Aufgaben weiterer Mitarbeiter, die sich mit steigender Anzahl an Daten zu vollen Personalstellen summieren können. Diese umfassen Anpassungen an den technischen Systemen, die ggf. für die Migration unbekannter Datentypen oder für neue Funktionen bei der Bereitstellung notwendig sind, Arbeiten zur administrative und rechtlichen Abwicklung eine Vorgangs und Maßnahmen der Öffentlichkeitsarbeit.

Darauf basierend lassen sich als initiale Planungsgröße pro Datensammlung folgende prozentuale Arbeitsumfänge veranschlagen, wobei zu betonen ist, dass dies keine finale Berechnungsgrundlage für die Gebühren darstellt. Arbeitsumfang pro Datensammlung pro Monat, immer bezogen auf eine volle FTE-Stelle:

¹⁵⁴ Vgl. WISSENSCHAFTSRAT (10466-11) S. 51 sowie WISSENSCHAFTSRAT (9035-09) S. 28f

¹⁵⁵ Im Jahr 2013 wurden von insgesamt vier FTE-Datenkuratoren 45 Projektarchive kuratiert und veröffentlicht:
<http://archaeologydataservice.ac.uk/about/CollectionsHistory2013>

- Datenkuration TVöD E13
 - Aufbereitung TP-SIP-AIP-DIP, Dokumentation 80 %
 - Qualitätsicherung 20 %
- Softwareentwickler TVöD E13
 - Betrieb und Anpassung technischer Systeme 10 %
- Abwicklung TVöD E12
 - Vertragsverhandlungen 10 %
 - Buchhaltung 10 %
 - Öffentlichkeitsarbeit 5 %

9.2.3 Personal Drittmittel

In der dritten Säule – Drittmittel (siehe **Kapitel 8.3.4**) – sind in Abhängigkeit der Anzahl, der Ziele und der Dauer der bewilligten Projekte folgende Personalstellen zu erwarten, deren genaue Anzahl nicht realistisch vorausberechnet werden kann, da sie maßgeblich von den Aufgaben der einzelnen Projekte abhängen:

- Datenkuration TVöD E13
- Softwareentwickler – Frontend TVöD E13
- Softwareentwickler – Backend TVöD E13
- IT-Empfehlungen TVöD E13
- System-Administrator TVöD E9-E12

10 Datenschutz, Lizenzen, Zugangsbeschränkungen

Zwei grundsätzliche Konstellationen sind für IANUS als Forschungsdatenzentrum aus datenrechtlicher Sicht relevant. Einerseits die Beziehung zwischen den Dateneigentümern bzw. Datenproduzenten, die gegenüber IANUS als Datengeber auftreten und andererseits zwischen IANUS und den Datennachutzern. Wie in **Kapitel 8** beschrieben wird das Prinzip „pay once store forever“ Anwendung finden, was bedeutet, dass ein Datengeber zwar den initialen Kuratierungsaufwand einmalig und anteilig (mit)finanziert, aber unabhängig davon die Daten in der Nachnutzung für Dritte kostenfrei sein werden. Diese Festlegung stellt ein wesentliches Kriterium dar, da IANUS mit den Forschungsdaten Dritter weder irgendwelche Gewinnerzielungsabsichten verfolgt, noch irgendein Umsatz (jenseits der Aufwandspauschale zur Kuratierung) erwirtschaftet werden wird. Die Bereitstellung von Forschungsdaten erfolgt folglich als dauerhaft unentgeltliche Dienstleistung. Aus dieser Konstellation ergeben sich verschiedene daten- und urheberrechtlichen Konsequenzen für den Umgang mit, die Lizenzierung von und den Zugang auf die IANUS anvertrauten Forschungsdaten. Diese werden in dem folgenden Kapitel beschrieben, das auf einer durch zwei externe Fachjuristen 2013 verfassten Expertise beruht¹⁵⁶.

10.1 Beziehung Datengeber – IANUS

10.1.1 Rechtliche Rahmenbedingungen

Datengeber übertragen mit einem nicht-exklusiven Auftrag ihre digitalen Forschungsdaten an IANUS mit dem Ziel, diese professionell zu kuratieren, langfristig physisch zu sichern, für wissenschaftliche Zwecke aufzubereiten sowie kostenlos und gut referenzierbar zur Verfügung zu stellen, u.a. in einem Online-Datenportal und unter Berücksichtigung von ggf. erforderlichen Zugangsrestriktionen. Im Rahmen dieser Arbeiten werden die Forschungsdaten kopiert, durch Formatmigrationen bearbeitet und neue Versionen erzeugt. Diese technischen Bearbeitungen, Vervielfältigungen und öffentliche Zugänglichmachung einer Datei sind urheberrechtlich relevante Arbeitsschritte, weshalb IANUS sich diese von den Dateneigentümern explizit genehmigen und entsprechende Nutzungsrechte einräumen lassen muss. Die Eigentumsverhältnisse an den Daten werden dadurch nicht berührt, d.h. ein Dateneigentümer bleibt juristisch weiterhin Eigentümer seiner Daten mit allen damit verbundenen Rechten und Pflichten. Das Eigentum an den Daten würde nur dann auf IANUS übergehen, wenn die Überlassung in Form eines Schenkungsvertrags erfolgt, woraus sich ferner für IANUS als Beschenktem insgesamt weniger Verpflichtungen ergeben würden¹⁵⁷.

Wissenschaftliche Forschungsergebnisse sind allerdings nur dann urheberrechtlich geschützt, wenn es sich zugleich um Werke oder Gegenstände von Leistungsschutzrechten handelt (beschreibende Texte, Fotos oder grafische Darstellungen), denen eine gewisse geistige Schöpfungshöhe innewohnt¹⁵⁸. Da die Daten selbst keine Sachen oder Sachgegenstände sind, fallen sie unter das sog. Immaterialgüterrecht, d.h. nicht sie selbst, sondern die Person in Form des

¹⁵⁶ IANUS – Hoehren (2014); IANUS – Klimpel/Weitzmann (2014).

¹⁵⁷ IANUS – Hoehren (2014) S. 22-23. 54.

¹⁵⁸ UrhG §2 Abs. 2, online unter: http://www.gesetze-im-internet.de/urhg/_2.html

Urhebers als Schöpfer des jeweiligen Werkes bzw. der Daten besitzt zu berücksichtigende Rechte. Eine Besonderheit besteht bei Datenbanken, für die zur Erlangung eines urheberrechtlichen Schutzes eine wesentliche Investition (finanziell und/oder zeitlich) erforderlich ist. Generell gilt, dass strukturierte Sammlungen von Daten, wie z.B. aufwendige Datenbanken mit größeren Datenbeständen, nach UrhG §4159 als ganzes Werk geschützt sind, nicht aber das einzelne darin enthaltene Datum, sofern es für sich genommen nicht eine geistige Schöpfungshöhe beanspruchen kann.

Einfache Daten, die nur Fakten oder Faktensammlungen repräsentieren und keine geistige Schöpfungshöhe beinhalten, unterliegen keinem Werk- oder Leistungsschutz und sind insofern urheberrechtlich nicht geschützt. Die Archivierung, Weitergabe und Nutzung solcher Daten darf nicht durch Lizenzen eingeschränkt werden, wobei durch Open-Access-Lizenzen explizit auf diesen Umstand hingewiesen werden kann. Das Archiv-Management-System von IANUS muss deshalb die Möglichkeit vorsehen, dass Inhalte als nicht urheberrechtlich geschützt, also offen gekennzeichnet werden können. Außerdem müssen die Datenkuratoren zusammen mit den Datengebern abwägen, ob bei konkreten Daten eine geistige Schöpfungshöhe vorliegt oder nicht, also eine Lizenzierung notwendig ist oder nicht.

Von Relevanz für IANUS ist weiterhin, welche Legitimation und Vollmachten ein Datengeber besitzt, da z.B. dessen arbeitsrechtliches Verhältnis Auswirkungen darauf haben kann, wer welche Lizenzen für welche Daten erteilen darf und wer nicht. Grundsätzlich muss sich ein Arbeitgeber die Nutzungsrechte an den Werken seiner Arbeitnehmer einräumen lassen, was in der Regel durch entsprechend formulierte Arbeitsverträge geschieht. Ist z.B. ein Arbeitnehmer als Fotograf bei einer Institution angestellt, gehören alle während der Arbeits- und Dienstzeit erstellten Fotos nutzungsrechtlich dem Arbeitgeber und dieser muss über die Formen der Archivierung und Bereitstellung der Bilddateien entscheiden. Ausgenommen davon sind freiberufliche und selbstständige Arbeitsverhältnisse, da hier alle Nutzungs- und Verwertungsrechte bei der jeweiligen Person als Urheber der Daten verbleiben. Ebenfalls abweichend verhält es sich bei Hochschullehrern und Professoren, welche aufgrund der Freiheit der Lehre immer alle Rechte an ihren im Rahmen eines Dienstverhältnisses geschaffenen Werken besitzen. Das betrifft sogar alle Ergebnisse und Daten, die Projektmitarbeiter bei einem entsprechenden Vorhaben unter der Leitung eines Hochschullehrers erzeugen. Universitäten und Hochschulen als Arbeitgeber müssen sich daher im Bedarfsfalle immer von den Lehrkörpern die entsprechenden Nutzungs- und Verwertungsrechte einräumen lassen.¹⁶⁰ Für IANUS bedeutet dies, dass es die Anstellungsverhältnisse und daraus resultierenden Rechte von Datengebern (also Dateneigentümer und Datenproduzent) überprüfen muss, um sicher zu gehen, dass die Verwertungsrechte überhaupt von der Person bzw. der Institut übertragen werden können. Im Falle, dass eine nicht berechnigte Person bzw. Institution einen Datenübergabevertrag (siehe **Kapitel 10.1.2**) unterschreibt, kann IANUS rechtlich für die Weiterverwendung der strittigen Daten haftbar gemacht werden, sofern diese Rechte IANUS nicht ausdrücklich bestätigt wurden.

Ein Datengeber muss also im Rahmen des Datenübergabevertrages bescheinigen, dass er die

¹⁵⁹ http://www.gesetze-im-internet.de/urhg/_4.html

¹⁶⁰ Thomas Hoeren (2014), Expertise - Rechtliche Aspekte bei der Langzeitarchivierung digitaler Daten. 1. Teil: Die Rechte aus dem archäologischen Input. [Version 1.0] Hrsg. IANUS. doi:10.13149/000.0nc98i-f

nötigen Rechte oder eine Vollmacht besitzt, Forschungsdaten mit dem Zweck der Archivierung und Bereitstellung anzubieten und zu übergeben. Dieser Nachweis kann entweder durch Vorlage von Verträgen (z.B. zwischen Drittmittelgebern, Verlagen, kommerziellen Unternehmen etc.) oder durch eine nachträgliche, schriftlich fixierte Zustimmung des Urhebers erfolgen. In begründeten Verdachtsfällen kann IANUS diese Angaben mit angemessenem und nachweisbarem Aufwand prüfen. Sofern Teile einer zu archivierenden Datensammlung offensichtlich nicht durch einen Datengeber selbst erstellt worden sind (z.B. Satellitenbilder oder Werke im Rahmen von Dienstleistungsaufträgen) und kein rechtlicher Nachweis erbracht werden kann, werden diese Dateien nicht in das IANUS-Archiv übernommen. Weniger offensichtliche urheberrechtliche Sonderfälle müssen durch den Datenübergabevertrag abgesichert werden, in dem der Datengeber per Unterschrift versichert, dass jedwede (zivil-)rechtliche Konsequenzen aus gerechtfertigten, urheberrechtlichen Ansprüchen Dritter zu seinen eigenen Lasten gehen. Darüber hinaus liegt die Verantwortung für die Einhaltung von Datenschutzbestimmungen beim Datengeber, welcher seine Datensammlung selbst nach etwaigen schutzwürdigen Informationen untersuchen und deren Existenz IANUS bekannt machen muss.

Es kann nicht ausgeschlossen werden, dass einzelne Datensätze, die IANUS anvertraut und dort gespeichert werden, personenbezogene Daten enthalten. Da ihre Verarbeitung und Verwendung die Erlaubnis der jeweils betroffenen Personen voraussetzt, wird IANUS von dem Datengeber die Vorlage entsprechende Zustimmungen einfordern. Dafür muss gewährleistet werden, dass die jeweils Betroffenen in der Lage sind, im Vorfeld den Anlass, das Ziel und die Folgen der digitalen Verarbeitung ihrer personenbezogenen Daten korrekt einzuschätzen.

Im Todesfall einer Person oder bei Schließung einer Institution gehen nach dem Gesetz alle Rechte, auch die an Forschungsdaten, an die regulären Rechtsnachfolger (Übergeordnete Einrichtung, Nachfolgeinstitution, Erben, Ehepartner etc.) über. Da es beim Ableben bzw. Auflösung von Datenproduzenten bzw. -eigentümern für IANUS nicht möglich sein wird, die jeweilige Rechtsnachfolger zu ermitteln und mit diesen ggf. neue Lizenzbedingungen zu verhandeln, müssen hier alternative praktikable Verfahren entwickelt werden. Es wird daher angestrebt, dass IANUS bereits bei der Datenübergabe das langfristige Recht eingeräumt wird, die Archivierung, Kuratierung und Bereitstellung von Daten auch über den Todesfall bzw. Institutsende hinaus zu betreiben und alle Forschungsdaten, für die noch Zugangsbeschränkungen existieren, ab diesem Zeitpunkt Open Access zur Verfügung zu stellen. Dieses Vorgehen ist bei nationalen (GESIS)¹⁶¹ wie internationalen (DANS)¹⁶² Forschungsdatenzentren gängige Praxis.

10.1.2 Datenübergabevertrag

Ein Datenübergabevertrag dient dem Zweck, alle Rechte und Pflichten zwischen einem Dateneigentümer bzw. von diesem beauftragten Datengeber als Dienstleistungsnehmer und IANUS als Dienstleistungsanbieter juristisch verbindlich zu regeln und für beide Seiten

¹⁶¹ http://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Archivierungsvertrag_GESIS_Datenarchiv_annotiert.pdf

¹⁶² <http://www.dans.knaw.nl/sites/default/files/file/juridisch/DANS-licence-agreement-UK.pdf>

Rechtssicherheit zu schaffen. In dem Vertrag werden vor allem folgenden Aspekte thematisiert¹⁶³:

- Nennung der Vertragsparteien
- Definition des Vertragsgegenstandes, d.h. der Daten und Beschreibung der Dienstleistung. Dabei ist IANUS in der Wahl der zweckdienlichen technischen Mittel, Formate und Methoden frei und darf zur Erreichung der Ziele die Hilfe bzw. Dienstleistungen Dritter in Anspruch nehmen.
- Zusicherung des Datengebers, dass er die Eigentums- und Nutzungsrechte an den übergebenen Daten besitzt oder diese durch Vollmachten nachweisen kann.
- Einräumung von nicht-exklusiven Nutzungsrechten für IANUS, insbesondere zur Vervielfältigung, Bearbeitung, Kuratierung, Archivierung und Bereitstellung.
- Verpflichtung von IANUS, die übertragenen Rechte auszuüben, und Festlegung, dass Weitergabe der eingeräumten Rechte nur mit schriftlicher Zustimmung des Dateneigentümers zulässig ist.
- Berechtigung für IANUS, die Daten im Rahmen von Kooperationsverträgen an vergleichbare Archive und Rechenzentren weiterzugeben, insbesondere bei Schließung des Betriebes und Notwendigkeit zur Umsetzung einer Exit-Strategie (siehe **Kapitel 11.5**).
- Klärung, wie im Rahmen der Kuratierung mit ggf. vorhandenen zu schützenden sensiblen Informationen umgegangen werden soll
- Einwilligung des Datengebers zur Veröffentlichung von und uneingeschränktem Zugang zu bestimmten Kern-Metadaten.
- Einwilligung des Datengebers, dass Dritte die Daten für eigenständige wissenschaftliche Untersuchungen (Sekundäranalysen) nutzen dürfen.
- Klärung der Rechtsnachfolge in Fällen des Ablebens des Datengebers, Schließung der datengebenden Institution oder Nicht-Nachvollziehbarkeit des Verbleibs des Datengebers; bei strittiger Rechtsnachfolge gehen sämtliche Verwertungsrechte auf IANUS als Treuhänder über.
- Umfassende Haftungsfreistellung von IANUS und Regelung zur Streitverkündung. Im Falle einer Gesetzesverletzung oder eine Verletzung Rechte Dritter besteht eine Informationspflicht des Datengebers gegenüber IANUS, das sich seinerseits verpflichtet ggf. unverzüglich die Daten zu entfernen oder den Zugang anzupassen.
- Haftung von IANUS ist auf Vorsatz und grobe Fahrlässigkeit im Zusammenhang mit der Archivierungstätigkeit beschränkt.
- Einwilligung des Datengebers, IANUS bei der Abwehr von Ansprüchen Dritter zu unterstützen und mögliche Kosten zu tragen.
- Versicherung, dass sich aus dem Vertrag keine Partnerschaften oder Gemeinschaftsunternehmen ergeben.
- Verpflichtung aller Beteiligten im Zusammenhang mit der Durchführung des Vertrages die

¹⁶³ Vgl. den Mustervertrag in der Expertise Hoeren 2014, S. 55-56 sowie Verträge beim ADS (http://archaeologydataservice.ac.uk/attach/collectionsPolicy/ads_licence_form.doc), DANS (siehe vorherige Anm.)
GESIS (siehe vorherige Anm.)
Zuse-Institut http://ewig.gfz-potsdam.de/wp-content/uploads/2011/12/ZIB_Muster_Uebnahmevereinbarung.pdf
Ferner: http://files.d-nb.de/nestor/materialien/nestor_mat_10.pdf S. 17.

einschlägigen Datenschutzbestimmungen einzuhalten, insbesondere wenn Teile der Daten im Ausland erhoben wurden.

- Hinweis auf die Unumkehrbarkeit des Rechtsgeschäftes, d.h. nach Unterzeichnung des Übergabevertrages und vollständiger Prozessierung der Daten (inkl. der Online-Stellung) können diese nur noch in Ausnahmefällen (z.B. bei nachgewiesenen Rechtsverletzungen) aus dem Archivbestand von IANUS entfernt werden.
- Erklärung des Datengebers zu schutzwürdigen Informationen und Erlaubnis für IANUS bei datenschutzrechtlichen Bedenken geeignete Maßnahmen vorzuschlagen und durchzuführen (z.B. Anonymisierung).
- Festlegung von Zugangsrestriktionen zu definierten Metadaten und Daten sowie Festlegung von Fristen für deren Ablauf und automatischen Veränderungen.
- Festlegung von exklusiven Sondernutzungsrechten auch für geschützte Inhalte für IANUS zu klar umgrenzten Zwecken, z.B. zur Präsentation von Mustersammlungen im Rahmen von Vorträgen
- Hinweis auf die von IANUS formulierten Vorgaben zur Annahme von Datensammlungen (siehe **Kapitel 14.1**) und deren bestmögliche Berücksichtigung durch den Datengeber
- Verpflichtung des Datengebers, Änderungen von Adress- und Kontaktdaten aktiv an IANUS mitzuteilen, sofern dies für die langfristige Kuratierung einer Datensammlung relevant ist
- Festlegung des Übergabedatums sowie Art und Weise der Übergabe
- Höhe und Berechnungsmethode der durch den Datengeber anteilig zu entrichtenden Gebühren
- Anzuwendendes Recht, Gerichtsstand, Schlussbestimmungen, Ausschluss mündlicher Nebenabreden, salvatorische Klauseln
- Unterschriften aller Rechteinhaber und IANUS

Sobald Datengeber und IANUS den Datenübergabevertrag unterzeichnet haben, die Forschungsdaten übergeben und sie vollständig bis zur Online-Bereitstellung prozessiert wurden, gibt es unter normalen Umständen keine Möglichkeit mehr, die Daten aus dem Archivbestand von IANUS wieder zu entfernen. Einmal veröffentlichte Forschungsdaten erhalten im Zuge der Aufbereitung eine eindeutige, persistente Identifikationsnummer (siehe **Kapitel 7.4**), die mit einem Datensatz fest verbunden bleibt, auch wenn sich einzelne Daten nach der Veröffentlichung bei IANUS als fehlerhaft erweisen sollten. Die Situation ist also vergleichbar mit gedruckten Werken, die nach Erscheinen ebenfalls nicht geändert oder in ihrer physischen Gesamtheit vernichtet werden können. Dem analogen Konzept einer aktualisierten oder korrigierten Neuauflage entspricht im digitalen Bereich die Erstellung einer neuen Daten-Version, die aber nicht zur Löschung bereits vorhandener Versionen führt. Ob und unter welchen Umständen eine neue Daten-Version erstellt werden soll, ist zwischen dem Datengeber und IANUS individuell zu vereinbaren und ist z.B. besonders bei kontinuierlich aktualisierten Daten aus laufenden Fachanwendungen zu berücksichtigen. Die im Rahmen der Vertragsvorbereitung und der initialen Kuratierung erfolgte Kommunikation sowie jede Übernahme von Datenpaketen wird vollständig bis zur finalen Archivierung protokolliert und neben dem Datenübergabevertrag sowie den Datensammlungen selbst als Teil des AIP im Archiv aufbewahrt.

Im Normalfall können im Nachhinein also nur die aktuellen Zugriffsstufen abgeändert werden, um die Zugänglichkeit zu erleichtern, nicht aber die Daten selbst. Schwerwiegende Gründe von dieser Regel abzuweichen liegen nur dann vor, wenn Forschungsdaten unrechtmäßig von Datengebern angeboten und veröffentlicht wurden oder Dritte die Verletzung von datenbezogenen Rechten

geltend machen und dies nachweisen können. In diesen begründeten Ausnahmen ist IANUS verpflichtet, Inhalte aus seinem Archivbestand komplett zu löschen oder den Zugriff stärker einzuschränken bzw. zu sperren, um einen größeren Schaden abzuwenden.

10.2 Beziehung IANUS - Datennachnutzer

10.2.1 Rechtliche Rahmenbedingungen

Bei der Beziehung zwischen IANUS und den Datennachnutzern geht es um die Frage, in welcher Form IANUS digitale Forschungsdaten, die von ganz unterschiedlichen Dateneigentümern stammen, online bereitstellen darf und wie Datennachnutzer diese rechtmäßig (nach)nutzen können. Damit jeder Datennachnutzer weiß, unter welchen Bedingungen die Forschungsdaten verwendet werden können, müssen für alle Forschungsdaten Lizenzen für die Nutzung vorliegen – entweder in allgemeiner, pauschaler oder in individualisierter Form. IANUS muss den Datengeber also über die Notwendigkeit von Lizenzen aufklären und verdeutlichen, dass Lizenzen für einmal ins Netz gestellte Werke unwiderruflich sind, sobald nur eine einzige lizenzgerechte Nutzung erfolgt ist.

Für die Lizenzierung von Forschungsdaten existieren eine Reihe von etablierten offenen Standardlizenzen, die klar gekennzeichnet und im Idealfall maschinenlesbar sind, dank weit verbreiteter Icons bzw. Logos, welche einen hohen Wiedererkennungswert besitzen sowie ein großes Vertrauen bei den Nutzern genießen¹⁶⁴. Sie kommen ohne individuelle Kommunikation zwischen Lizenzgeber und Nutzer zustande: ein Lizenzgeber bietet sein Werk als Angebot an und ein Nutzer nimmt es durch die Nutzung an. Da die Formulierung eigener Lizenzen schwierig, der administrative Aufwand hoch und die Akzeptanz fraglich wären, sollen diese Public Licences¹⁶⁵ so weit wie möglich auch im Kontext von IANUS Anwendung finden. Bei Dateien, deren Zugriff zeitlich oder inhaltlich eingeschränkt werden soll, ist es nicht empfehlenswert, für diese Restriktionen eigene Lizenzen zu schaffen. Stattdessen werden notwendige Zugangsbeschränkungen technisch umgesetzt, insbesondere durch ein Rechte-Rollen-Konzept und Authentifizierungsmechanismen. (siehe **Kapitel 11.3**) Dieses Vorgehen bietet auch den Vorteil, dass eine Lizenz unverändert bleibt, auch wenn Restriktionen des Zugriffs reduziert werden oder ganz wegfallen. Solche Anpassungen müssen dann nur im Rechte-Rollen-Management, nicht aber in den Metadaten der archivierten Daten eingepflegt werden.

Daher erhalten im Standardfalle alle Forschungsdaten, die nicht einem Embargo unterliegen (siehe **Kapitel 10.3.4**), die Standardlizenz von Creative Commons CC-BY-SA; diejenigen, die einem Embargo unterliegen, erhalten vorerst keine Lizenz, da für sie innerhalb einer festgelegten Frist (in der Regel für maximal zwei Jahre) jeglicher Zugriff gesperrt und eine Nachnutzung zunächst nicht vorgesehen ist. Für Institutionen, welche eigene Lizenzen besitzen und aus bestimmten Gründen Ihre Forschungsdaten nicht unter einer CC-BY-SA-Lizenz stellen können, wird es Ausnahmen geben und möglich sein, die Forschungsdaten unter den entsprechend individuellen und explizit formulierten Auflagen nachzunutzen. Forschungsdaten von Institutionen, die keinerlei

¹⁶⁴ IANUS – Klimpel/Weitzmann (2014).

¹⁶⁵ Eine Übersicht über Open Licences gibt der ARIADNE-Bericht „D3.3: Report on data sharing policies“ (2014) (<http://www.ariadne-infrastructure.eu/Resources/D3.3-Report-on-data-sharing-policies-EU-Reporting>).

Nachnutzung erlauben, wird IANUS ablehnen.

Forschungsdaten, die als Open Access mit einer CC-BY-SA-Lizenz frei verfügbar sind, sind für jeden beliebigen Datennachnutzer ohne weitere Einschränkungen nachnutzbar, d.h. er kann sie in jedwedem Format oder Medium vervielfältigen, weiterverbreiten und verändern – und zwar für beliebige Zwecke, sogar kommerzielle, wenn angemessene Urheber- und Rechteangaben gemacht und wenn die abgeleiteten Daten unter den gleichen Bedingungen weitergegeben werden¹⁶⁶. Bei geschützten Forschungsdaten, bei denen der Zugang auf registrierte und berechnigte Nutzer reduziert ist (siehe **Kapitel 10.3**), können die aus der CC-BY-SA-Lizenz resultierenden Rechte durch die Allgemeinen Nutzungsvereinbarungen von IANUS (siehe **Kapitel 10.2.2**) eingeschränkt werden. Insbesondere ist für diese Daten eine Weitergabe an Dritte nicht zulässig und potentielle weitere Nutzer müssen sich zunächst bei IANUS registrieren, um die entsprechenden Zugangs- und Nutzungsrechte zu erhalten.

Nutzer des Online-Datenportals von IANUS sollen auf vier Arten auf rechtliche Aspekte und Datenlizenzen hingewiesen werden.

1. Zunächst werden sogenannte Click-Wrap-Nutzungsbedingung eingeblendet, die Nutzer über die allgemeinen Nutzungsbedingungen von IANUS aufklären und denen im Sinne eines Vertrages zugestimmt werden muss, bevor Dateien oder Metadaten angesehen und zur Nachnutzung heruntergeladen werden können.
Da bei Open Access-Inhalten eine Anmeldung eines Datennachnutzer nicht erforderlich ist, kann die Identität eines Systemnutzers kaum nachgewiesen werden, so dass bei Click-Wrap-Verträgen in der Regel der Vertragspartner unbekannt bleibt und weitere ergänzende Lizenzhinweise erforderlich sind¹⁶⁷. Auch kann bei diesem Verfahren nicht auf die lizenzrechtlichen Besonderheiten einzelner Datensammlungen eingegangen werden.
2. Daher wird auf jeder einzelnen Webseite, die eine spezifische Datensammlung beschreibt und als Landing-Page für die DOI fungiert (siehe **Kapitel 7.4**), ein eindeutiger und gut erkennbarer Lizenzhinweis platziert, sowohl in Form eines Icons als auch als Kurztext. Generell sind Lizenzhinweise verpflichtende Angaben innerhalb der Metadaten, die zu einer Datensammlungen oder einer Einzeldatei gehören und müssen auch bei Weitergabe von Metadaten an Dritte (z.B. über Webservices) enthalten sein.
3. Werden Forschungsdaten zum Download angefordert, werden nicht nur die angeforderten Daten selbst, sondern automatisch in einer separaten Datei auch die für diese spezifischen Lizenzbedingungen mitausgeliefert.
4. Und schließlich sollen, um eine möglichst untrennbare Verbindung von digitalem Material und Lizenz zu erzielen, in den Dateien selbst die Hinweise hinterlegt werden, wobei die praktische Umsetzung stark von den Dateiformaten abhängig ist, z.B. bei Textdokumenten als Teil des Impressums oder bei Rastergrafiken als Teil der IPTC-Metadaten.

10.2.2 Nutzungsvereinbarungen

Für Datennachnutzer gibt es zwei Arten von relevanten Vereinbarungen, die sich überschneiden bzw. miteinander verschränkt sein können:

¹⁶⁶ <https://creativecommons.org/licenses/by-sa/3.0/de/>

¹⁶⁷ IANUS – Klimpel/Weitzmann (2014).

- zum einen die allgemeinen Nutzungsbedingungen, welche die korrekte Nutzung der Online-Angebote und Dienstleistungen von IANUS regeln und für alle Datensammlungen gleichermaßen gelten;
- zum anderen die auf einzelne Datensammlungen bezogenen Lizenzen, die als positiv formulierte Rechte die Nachnutzung von Daten beschreiben, von Dateneigentümern für Datennachnutzer festgelegt und über IANUS nur weitergegeben werden.

In den allgemeinen Nutzungsbedingungen von IANUS werden Datennachnutzer darauf hingewiesen, dass für jede Datensammlung Lizenzen vorliegen, die es einzuhalten gilt. Während angestrebt wird, so viele Daten wie möglich mittels CC-BY-SA-Lizenzen und einem offenen Zugang frei nachnutzbar zur Verfügung zu stellen, können bei Daten, bei denen aufgrund von Vorgaben der Dateneigentümer Zugangsbeschränkungen vorliegen, auch die Nutzungsmöglichkeiten eingeschränkt sein. In diesen Fällen kann die Weitergabe an Dritte unterbunden, eine kommerzielle Verwendung, insbesondere der Verkauf von Daten, verboten und eine Nachnutzung auf die wissenschaftliche und akademische Forschung und Lehre beschränkt sein.

Der Datennachnutzer verpflichtet sich, alle genutzten Informationen und Quellen den wissenschaftlichen Gepflogenheiten entsprechend zu verwenden. Insbesondere müssen Autoren, die unter Verwendung von bereitgestellten Daten neue eigenständige Ergebnisse und Publikationen erzeugen, die ursprünglichen Datenproduzenten, Institutionen und Mittelgeber benennen und die Datensammlung in IANUS mittels der angegebenen DOI zitieren. Auch ist in dem Werk ein Hinweis einzufügen, der die Erstautoren und IANUS von der Verantwortung für die Neuinterpretation bzw. Sekundäranalyse der Daten freispricht. Ferner werden solche Autoren gebeten, IANUS über neue Veröffentlichungen zu informieren und entsprechende Metadaten (z.B. ISBN, PIDs, URL) zuzusenden.

Des Weiteren werden Datennachnutzer aufgefordert, das Archiv auf Fehler aufmerksam zu machen und ihm im Falle einer Registrierung die Verwaltung seiner persönlichen Angaben zu erlauben und sie zur anonymen statistischen Auswertung zu nutzen, z.B. um zu analysieren, wie oft ein Datensatz heruntergeladen wurde. Bei Verletzung der Nutzungsbedingungen hat IANUS das Recht gegen Datennachnutzer vorzugehen, genauso wie auch Dateneigentümer einen Missbrauch ihrer Daten selbstständig ahnden können.

Zu beachten ist, dass bei einer transformierenden Nutzung durch den Datennachnutzer das neue Werk wieder lizenziert wird, aber die Lizenzen der in dem Werk genutzten Materialien (z.B. ein wiederverwendetes Foto) bestehen bleiben. Sollten bei der Nachnutzung die Lizenzen der ursprünglichen Werke verletzt werden, verfällt die Veröffentlichungslegitimation des neu entstandenen Werks. Da in diesem Fall der Aufwand für ein neues Werk umsonst wäre, wird IANUS versuchen, Datennachnutzer über diese Konsequenzen aufzuklären und für Urheberrechte und die Einhaltung von Lizenzen zu sensibilisieren.

10.3 Zugangsbeschränkungen für Daten

Generell bekennt sich IANUS zum Open-Access-Paradigma und ist bestrebt, die ihm anvertrauten Forschungsdateien und Metadaten nicht nur in technisch sicheren aber abgeschlossenen Speichersystem zu archivieren, sondern auch langfristig so offen wie möglich für die Forschung zur Verfügung zu stellen. Gleichzeitig wird aber auch anerkannt, dass bestimmte Daten im Kontext der Altertumswissenschaften als sensible Daten einzuschätzen sind und daher einen besonderen

Schutz benötigen. Dabei handelt es sich vor allem um Geoinformationen, die sich auf Denkmäler und Fundstellen beziehen, die durch Raubgrabungen gefährdet sind, wissenschaftliche Erkenntnisse mit geistiger Schöpfungshöhe, die noch nicht publiziert wurden, sowie jegliche personenbezogene Angaben. Zu diesem Zweck, aber auch für weitere Anforderungen können Datengeber Zugangsbeschränkungen festlegen, deren Einhaltung IANUS durch ein detailliertes Rechte-Rollen-Konzept sowie technische Schutzmechanismen sicherstellen muss. Vier unterschiedliche Stufen von Zugangsbeschränkungen sind vorgesehen und können für ganze Datensammlungen, einzelne Dateien oder spezifische Metadaten festgelegt werden. Dabei wird bei Festlegung der Zugriffs- und Nachnutzungsrechte prinzipiell von offenen Daten ausgegangen, für die bestimmte Einschränkungen formuliert werden.

Allerdings ist zu beachten, dass für jede Datensammlung ein basaler Satz an Metadaten immer frei und ohne jegliche Einschränkung veröffentlicht wird, auch wenn für zugehörige Teile Zugangsbeschränkungen vorliegen. Nur auf diese Weise können Datennachnutzer vorhandene Datenbestände suchen und Inhalte finden und dann bei Interesse eine Anfrage für einen erweiterten Zugriff stellen. Ebenfalls wird so die Weitergabe von Grundinformationen an andere Systeme, insbesondere externe Nachweisportale und Aggregatoren von Metadaten, über Schnittstellen vereinfacht.

Interessierte Personen, die von IANUS bereitgestellte Forschungsdaten nutzen wollen, für die nur ein gruppen- oder personenspezifischer Zugriff möglich ist (siehe **Kapitel 10.2.2 und 10.2.3**), müssen sich online bei IANUS registrieren. Die Registrierung stellt dabei den technischen Prozess der vertraglich bindenden Vereinbarung über die Nutzung von Forschungsdaten mit Zugangsbeschränkungen dar (siehe **Kapitel 10.4**). Nach einer erfolgreichen Registrierung und Freischaltung werden die Personen entweder automatisch pauschal definierten Gruppen zugewiesen (z.B. eine institutionelle Zugehörigkeit aufgrund von E-Mail-Domains) oder aufgrund von individuellen Zustimmungen durch Dateneigentümern in spezifische Einzelgruppen eingetragen. Mit beiden Arten von Gruppen sind jeweils bestimmte Rechte auf bestimmte Datensammlungen verbunden. Für weitere Forschungsdaten, für die dann noch kein Zugriff möglich ist, können Datennachnutzer Anfragen an die jeweiligen Dateneigentümern stellen und eine persönliche Freischaltung beantragen. Die Identifizierung der individuellen Personen als berechtigter Systemnutzer geschieht über ihren Namen bzw. E-Mailadresse.

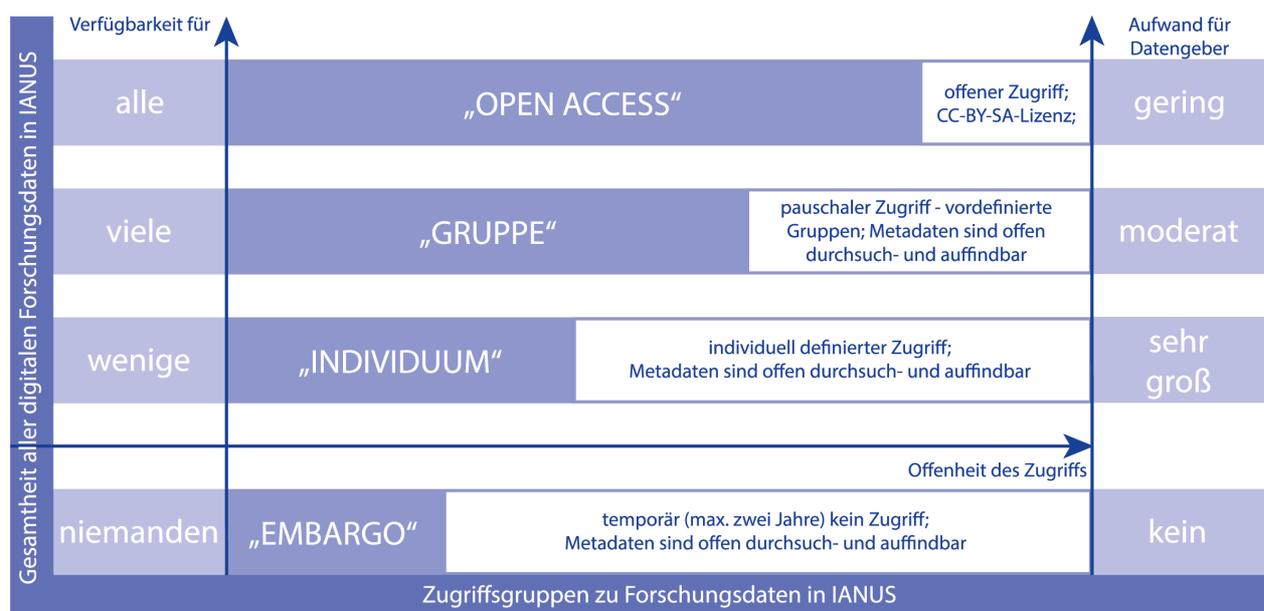
Wie offen oder geschlossen eine Datensammlung letztendlich veröffentlicht wird, legen die jeweiligen Dateneigentümer fest, wobei IANUS auf die unterschiedlichen Möglichkeiten und die Konsequenzen hinweisen wird und dabei auf eine möglichst offene Bereitstellung hinwirken wird. Entscheiden sich Dateneigentümer für einen restriktiven Zugriff auf ihre Daten müssen nachvollziehbar Gründe für die Einschränkung genannt und in den Metadaten festgehalten werden. Sofern diese nicht vorliegen, behält sich IANUS die Entscheidung vor, Daten entweder mit freiem Zugriff zu veröffentlichen oder, sofern eine explizite Zustimmung der Datengeber fehlt, sie ganz abzulehnen. Auch der Wissenschaftsrat kommt in seiner Stellungnahme vom Oktober 2015 im Rahmen der Evaluierung des DAI zu dem Fazit, dass die Zugangsrechte zu den Daten und Informationen in IANUS nur in begründeten Ausnahmen limitiert werden sollten¹⁶⁸. Mögliche, plausible Gründe für eine Beschränkung können nach aktuellem Stand sein, dass

¹⁶⁸ Vgl. WR Drs. 4907-15, S. 90; <http://www.wissenschaftsrat.de/download/archiv/4907-15.pdf>

- Ergebnisse der Daten noch nicht veröffentlicht wurden,
- Personen, Orten und Objekte geschützt werden sollen,
- Rechte von Dritten an den Daten berücksichtigt werden müssen.

Falls Beschränkungen (siehe **Kapitel 10.2**) vereinbart wurde, die nicht von vornherein zeitlich befristet oder von grundsätzlicher Natur sind (z.B. aufgrund von Gesetzen), wird IANUS in regelmäßigen Abständen die Dateneigentümer kontaktieren und aktiv nachfragen, ob die Voraussetzungen für einen eingeschränkten Zugriff noch gegeben sind oder ob die Daten inzwischen für einen größeren Personenkreis geöffnet werden können. Sofern diese nach einer festgelegten Frist keine entsprechende Rückmeldung äußern, wird davon ausgegangen, dass eine Zugriffsbeschränkung nicht mehr notwendig ist und dass die Daten offen unter einer CC-BY-SA-Lizenz bereitgestellt werden können.

Je nach Datengeber und Datensammlung strebt IANUS zudem die Einwilligung von klar definierten Sondernutzungsrechte auch für geschützte Dateninhalte an, so dass diese im Rahmen von nicht-kommerziellen Lehrveranstaltungen, Summer Schools, Schulungen, Vorträgen oder der Öffentlichkeitsarbeit ebenfalls genutzt und ggf. temporär zugänglich gemacht werden können.



10.3.1 Freier Zugriff („Open Access“)

Datennachnutzer können ohne Einschränkungen und ohne Login auf alle Dateien und alle Metadaten zugreifen. Lediglich schützenswerte Geoinformationen können – sofern notwendig und vom Dateneigentümer bzw. Datengeber vorgegeben – mit unvollständigen Ortsangaben, höheren Ungenauigkeiten oder kleineren Maßstäben angezeigt werden, als diese tatsächlich in den Daten vorliegen. Sofern Datennachnutzer präzise Geoinformationen benötigen, müssen sie sich am System anmelden und die nötigen Zugriffsrechte besitzen, die Dateneigentümer gewähren und verwalten können. Insgesamt bedeutet „Open Access“, dass Dateneigentümer und Datengeber keine Kontrollmöglichkeit darüber haben, wer wie auf Ihre Daten zugreift und dass eine gewünschte Verwendung nur durch zugehörige Lizenzbedingungen gesteuert werden kann.

10.3.2 Gruppenspezifischer Zugriff („Group Access“)

Mitglieder einer pauschal definierten Gruppe erhalten nach einem Login einen freien Zugriff und festgelegte Rechte auf ansonsten geschützte Dateien und Metadaten. Diese Gruppe bezieht sich pauschal auf alle Mitglieder von Institutionen, Vereinigungen oder Verbänden und wird durch IANUS automatisiert administriert, etwa durch die Zuweisung von Rechten an eine bestimmte institutionelle E-Mail-Domäne. Unter den restriktiven Zugängen ist der administrative Aufwand für die Dateneigentümer bzw. Datengeber am geringsten.

10.3.3 Personenspezifischer Zugriff („Individual Access“)

Sofern die pauschal definierten Gruppen als Möglichkeit der Zugriffsbeschränkung nicht ausreichen, können auch individuelle Gruppen mit explizit benannten Einzelpersonen erstellt werden (z.B. für ausgewählte Mitarbeiter eines Projektes, die über verschiedene Institutionen verteilt sind). Diesen wird nach einem Login ein freier Zugriff und festgelegte Rechte auf ansonsten geschützte Dateien und Metadaten gewährt. Dateneigentümer bzw. Datengeber entscheiden dabei selbstständig, welches Individuum einer Gruppe hinzugefügt werden soll oder nicht: entweder im Vorhinein durch explizite Benennung von Namen, die bereits als Systemnutzer bei IANUS registriert sind, oder im Nachhinein, wenn Datennachnutzer aktiv mittels einer standardisierten E-Mail eine Anfrage auf Gewährung von Zugriffsrechten stellen („Access on demand“). Die Verwaltung und Verantwortung für die jeweiligen Gruppen liegt dabei bei den Dateneigentümern, die auch die Kommunikation mit den bereits eingetragenen und möglichen neuen Mitgliedern übernehmen. Auch wenn Dateneigentümern auf diese Weise eine höchstmögliche Kontrolle der Zugriffssteuerung erhalten, ist für sie der administrative Aufwand für die Annahme der Anfragen, die Entscheidung darüber und die Mitteilung neuer individueller Zugriffsrechte erhöht.

10.3.4 Datengeberspezifischer Zugriff („Embargo“)

Eine Sonderform stellt das in der Regel auf zwei Jahre befristete Embargo dar, bei dem ausschließlich der Dateneigentümer, -produzent und -geber einen Zugriff auf eine Datensammlung erhält und ein Zugriff für Dritte nicht vorgesehen ist. Lediglich die zugehörigen Metadaten werden sofort offen und für jeden zugänglich gemacht, sofern in diesen keine schützenswerten Informationen zu Personen und Kulturgütern enthalten sind. Diese zeitlich befristete Zugriffseinschränkung ist insbesondere für solche Situationen vorgesehen, in denen Forschungsdaten zwar schon von IANUS kuratiert und archiviert werden können, aber noch nicht für Dritte bereitgestellt werden sollen, da z.B. Publikationen noch ausstehen, Prüfungsverfahren noch nicht abgeschlossen oder rechtliche Fragen noch ungeklärt sind. Grundsätzlich besteht aber das Ziel darin, die Forschungsdaten spätestens mit Ablauf der Embargo-Frist in einer der drei anderen Zugriffskategorien zu veröffentlichen. Insofern muss von einem Datengeber bereits bei der Vereinbarung eines Embargos festgelegt werden, wie die Forschungsdaten nach Ablauf der Frist oder nach Wegfall des Embargogrundes zugänglich gemacht werden sollen.

11 Qualitätssicherung und IT-Sicherheit

Der Betrieb eines digitalen Langzeitarchives stellt besondere Anforderungen an die IT-Sicherheit und erfordert entsprechende Maßnahmen zur Qualitätssicherung, damit archivierte Forschungsdaten tatsächlich für die Zukunft erhalten bleiben. Gemeinsam tragen sie dazu bei, dass Mitglieder der Zielgruppen (siehe **Kapitel 3**) dieses als ein vertrauenswürdigen Archiv wahrnehmen und entsprechend nutzen. Im Folgenden werden daher verschiedene Konzepte und Umsetzungen skizziert, die in diesem Sinne spezifische Fragen adressieren und zum Ziel haben, sowohl die eingesetzten IT-Systeme (inkl. der Übertragungswege) als auch die IANUS anvertrauten Forschungsdaten dauerhaft gegen eine Gefährdung abzusichern¹⁶⁹. Hierzu gehören beispielsweise Ereignisse höherer Gewalt, unbefugte Zugriffe auf Systeme durch Schadsoftware, bewusste oder fahrlässige Manipulationen oder Löschung von digitalen Inhalten, fehleranfällige organisatorische und technische Arbeitsschritte. Vor diesem Hintergrund muss sichergestellt werden, dass die Integrität, Authentizität und Vertraulichkeit der archivierten digitalen Objekte dauerhaft und reversionssicher gewahrt bleiben, der Zugriff durch unbefugte Dritte verhindert wird und alle vorgesehenen Geschäftsprozesse störungsfrei und anforderungsgemäß vollzogen werden können.

11.1 Standards und Zertifizierung

Eine der wichtigsten Vorkehrung zu diesem Zwecke ist die Einhaltung und Anwendung von Standards, insbesondere des OAIS-Referenzmodells (siehe **Kapitel 5**), das vor allem organisatorische Aspekte und Anforderungen an eine nachhaltige Informationsmodell beschreibt und 2012 als ISO-Standard 14721 anerkannt wurde. Die Qualität von dessen Umsetzung kann durch eine Zertifizierung objektiv überprüft und dadurch IANUS-Nutzern nach außen vermittelt werden. Ein solches Verfahren trägt wesentlich zur Vertrauensbildung bei:

- für Dateneigentümer vermittelt es die Sicherheit, dass ihre Daten professionell, nach anerkanntem Stand der Technik und auf Grundlage aktueller Strategien behandelt werden;
- für Datennutzer, dass die Daten inhaltlich authentisch und technisch nachnutzbar sind und bleiben.

Für die Begutachtung und Evaluierung von digitalen Langzeitarchiven haben sich in den letzten Jahren einige Initiativen und Qualitätssiegel etabliert, die auch für IANUS eine Relevanz besitzen¹⁷⁰. Die Akteure verfolgen mit ihren Ansätzen unterschiedliche Ziele, die sich im Aufwand und Umfang des Zertifizierungsprozesses widerspiegeln.

- Das *Data Seal of Approval (DSA)* formuliert 16 Leitsätze für die Beurteilung von Archiven, die mithilfe eines Online-Tools beantwortet werden können. Da das DSA allein auf Basis der eigenen Einschätzung eine Institution vergeben wird, die allerdings nachvollziehbar belegt

¹⁶⁹ Vgl. auch Archivierung digitaler Aufzeichnungen im Kooperationsverbund Digitales Archiv Nord (DAN).

Hauptkonzept [Version 1.0] (2014) S. 21 = <http://www.hamburg.de/contentblob/data/archivierung-digitaler-aufzeichnung-im-dan-hauptkonzept.pdf>

¹⁷⁰ URLs zu den folgenden Zertifikaten sind in Anm. 10–12 angegeben. Siehe auch die EU-Initiative *Trusted Digital Repository* <http://www.trusteddigitalrepository.eu/Welcome.html>

werden müssen, muss es im Vergleich mit den anderen Zertifikaten als eine reine Selbstevaluierung bewertet werden.

- Der von dem Kompetenznetzwerk NESTOR entwickelte „Kriterienkatalog für vertrauenswürdige digitale Langzeitarchive“ (DIN-Norm 31644) ist ein Katalog zur Prüfung der Vertrauenswürdigkeit von digitalen Archiven und besitzt direkte Bezüge zu dem OAIS-Referenzmodell. Aufbauend auf den 34 Kriterien des Kataloges wurde das NESTOR-Siegel entwickelt, das sowohl eine erweiterte Selbstevaluierung als auch ein begutachtetes Prüfverfahren beinhaltet und erst im Falle eines positiven Ergebnisses vergeben wird. Letzteres hängt von den erreichten Punkten ab, die vor allem für reale Umsetzungen von organisatorischen und technischen Konzepten vergeben werden.
- Die dritte und höchste Stufe der Zertifizierung stellt das *Audit and Certification of Trustworthy Digital Repositories* dar, das auf einem eigenen ISO-Standard 16363:2012 beruht und ein vollwertiges, formelles, durch externe Experten überwacht Zertifizierungsverfahren beinhaltet.

Für IANUS sind zum gegenwärtigen Zeitpunkt vor allem die beiden erstgenannten Initiativen von Bedeutung, da die dort zu beantwortenden Fragen während der Aufbauphase als hilfreiche Richtlinien fungieren können. Sobald IANUS als Infrastruktur seinen Regelbetrieb aufgenommen hat, wird eine Zertifizierung durch das *Data Seal of Approval* und das *NESTOR-Siegel* angestrebt und erscheint mit vernünftigem Aufwand darstellbar. Der letztgenannte ISO-Standard dient dagegen nur als weiterer Referenzpunkt, der bei offenen Punkten konsultiert wird, dessen vollständige Einhaltung mittelfristig aber nicht aktiv verfolgt wird.

11.2 Integrität, Authentizität, Vertraulichkeit

Eine zentrale Anforderung an IANUS als fachspezifisches digitales Langzeitarchiv ist die dauerhafte Sicherstellung von Integrität, Authentizität und Vertraulichkeit der anvertrauten Datensammlungen. Um diesem Anspruch gerecht zu werden, müssen Prozesse eingeführt werden, durch die die digitalen Objekte jederzeit sowohl am Speicherort als auch beim Datentransfer gegen den unbefugten Zugriff und unbemerkte Veränderungen durch Dritte abgesichert werden.

Die Integrität der gespeicherten Forschungsdaten wird durch die Implementierung kryptographischer Hashfunktionen als komplexe Prüfsummenverfahren nachgewiesen. Angewandt werden soll entweder der *Message-Digest Algorithm 5* (MD5) oder der *Secure Hash Algorithm 3* (SHA-3). Durch die automatisierte Berechnung und Validierung von diesen Hashwerten kann mit sehr hoher Wahrscheinlichkeit sichergestellt werden, dass Bitfehler, Veränderungen oder unerwünschte Eingriffe frühzeitig erkannt und korrigiert werden können. Daher werden an mehreren Stellen der Workflows (siehe **Kapitel 14.3-14.5**) für jede Datei diese Prüfsummen erzeugt und z.B. vor und nach Kopiervorgängen miteinander verglichen werden. Ebenfalls durch den Abgleich von Hashwerten erfolgt die Validierung der Authentizität und Integrität im Rahmen von Übernahme-, Backup- und Nutzungsprozessen. Da auch Prüfsummenberechnungen und kryptographische Hashfunktionen einer Alterung unterliegen und gebrochen werden können¹⁷¹,

¹⁷¹ D.h. es ist möglich, ein neues Dokument mit dem Hashwert eines bereits vorhandenen Dokumentes zu erstellen.

ist regelmäßig zu überprüfen, ob eine Umstellung auf neue sicherer Methoden und Verfahren erforderlich ist.

Um die Anforderungen an ein vertrauenswürdigen digitales Archiv zu erfüllen, müssen darüber hinaus alle Arbeitsschritte und Prozesse, die eine Veränderung der ursprünglich übertragenen Daten und des Metadatenbestandes bewirken, dokumentiert werden. Ein Großteil der Prozesse wird durch die eingesetzten Systeme und Anwendungen automatisch protokolliert und in einer nach dem PREMIS-Standard aufgebauten XML-Datei (siehe **Kapitel 12.2.2**), die frühzeitig für jede Datensammlung gesondert erzeugt wird, festgehalten. Manuell vorgenommene Arbeitsschritte werden ebenfalls in diesem Dokument eingetragen, dessen Vollständigkeit jeweils am Ende der Pre-Ingest-, Ingest- und Access-Phasen durch einen zweiten Datenkurator gegengeprüft wird. Dadurch wird gewährleistet, dass der Werdegang jedes digitalen Objekts von der Übernahme durch IANUS bis hin zur Bereitstellung für Datennachnutzer lückenlos, transparent und nachvollziehbar bleibt. Auch für diese Logfiles und Protokolldateien werden Prüfsummen erzeugt und vorgehalten, damit sie ebenso wie die eigentlichen Primärdaten gegen Veränderungen und Manipulationen abgesichert sind.

Auf eine Verschlüsselung der Daten im Langzeitspeicher der externen Rechenzentren wird bewusst verzichtet, um auch im Falle eines Schadensereignis einen Datenverlust möglichst gering halten zu können und eine weitgehende Wiederherstellung nicht zusätzlich zu erschweren.

11.3 Zugriffssicherung

Die professionelle Archivierung von Forschungsdaten ist ein arbeitsteiliger Prozess, an dem unterschiedliche Akteure mit unterschiedlichen Aufgaben und Interessen beteiligt sind. Ebenso gibt es Datensammlungen und Metadaten, die je nach den Vorgaben der Dateneigentümer unterschiedlich frei oder beschränkt zugänglich sein können. Für alle Daten, die nicht unmittelbar als Open Access veröffentlicht werden können (siehe **Kapitel 10.3**), muss IANUS also sicherstellen, dass der Zugriff nur autorisiert erfolgen kann. Aber auch vor der eigentlichen Publikation, also während des Datentransfers, der Kuratierung und der Aufbereitung, muss gewährleistet werden, dass nur berechtigte Personen auf für sie bestimmte Daten und Systemkomponenten zugreifen können und ein Missbrauch durch unbefugte Nutzer verhindert werden kann.

Angesichts der Vielzahl der beteiligten Akteure ist ein differenziertes Rechte-Rollen-Konzept erforderlich, das dafür sorgt, dass ausschließlich autorisierten Benutzern der Zugriff auf ausgewählte Informationen und Funktionen gewährt wird, während ihnen gleichzeitig gezielt Inhalte und Systemteile verweigert werden können, die nicht für sie bestimmt sind. Aus Gründen der Sicherheit und Effizienz ist es sinnvoll, die Beziehungen von Personen und ihren Rechten nicht unmittelbar einzeln zu verwalten, sondern stattdessen über die Zuweisung von Gruppen bzw. Rollen zu bündeln. Eine Rolle entspricht dabei der Funktion des Akteurs im Kontext von IANUS nicht jedoch konkreten Einzelpersonen. Eine Person kann daher mehrere Rollen innehaben, z.B. wenn ein Datengeber auch ein Datennachnutzer ist. Inhalt und Umfang der gewährten Berechtigungen orientieren sich an den jeweiligen Nutzungsszenarien und werden über die Zugehörigkeit zu mindestens einer vordefinierten Rolle/Gruppe zugewiesen.

Insgesamt muss das Rechte-Rollen-Konzept drei Ebenen in Bezug zueinander setzen:

- a. die vordefinierten Rollen mit ihren Rechten (s.u.)

- b. die Inhalte (Datensammlungen, Dateien, Metadaten)
- c. die Systemkomponenten (IANUS.pre-ingest, IANUS.ingest, IANUS.access, Online-Datenportal, Nachweiskatalog, Homepage, ...)

Auch wenn viele Prozesse automatisiert ablaufen, muss für jede Systemaufgabe auch eine manuelle Durchführung veranlasst werden können. Aus diesem Grund sind bei der Entwicklung des Konzeptes auch Rechte zu bedenken, die im Regelfall systembasiert, also automatisiert ablaufen. Zum gegenwärtigen Zeitpunkt zeichnen sich folgende Rollen für den Betrieb von IANUS ab, denen unterschiedliche Kombinationen von Berechtigungen zugewiesen werden:

- Administratoren dürfen lesend, schreibend und löschend auf alle Systeme und den gesamten IANUS-Datenbestand zugreifen. Ihnen obliegt das Management der Datenbestände auf Systemebene.
- Datenkuratoren dürfen lesend auf alle Inhalte des Archives zugreifen und schreibend nur auf die Ihnen zur Kuratierung zugewiesenen Datenbestände. Sie können keine bereits vollständig archivierten digitalen Objekte verändern oder löschen, aber Vorschläge machen, die dann von Administratoren ausgeführt werden. Sie übernehmen zusätzlich die Verwaltung der externen Nutzer, indem sie Zugriffsrechte zuweisen und entziehen können.
- Dateneigentümer / Datenproduzent / Datengeber (s.a. **Kapitel 3.4**) sind externe Benutzer, die über das Daten-Onlineportal auf diejenigen Datenbestände und Metadaten lesend zugreifen dürfen, für die sie entsprechende Berechtigungen besitzen. Zusätzlich dürfen sie diejenigen Datensammlungen, die von Ihnen an IANUS eingeliefert wurden, während des Kuratierungsprozesses auch schreibend zugreifen, nach dessen Abschluss allerdings nur noch lesend.
- Datennachnutzer (s.a. **Kapitel 3.4**) sind externe Benutzer, die über das Daten-Onlineportal auf diejenigen Datenbestände und Metadaten lesend zugreifen dürfen, für die sie entsprechende Rechte besitzen.

Das Rechte-Rollen-Konzept im Allgemeinen und die erteilten Zugriffsrechte im Speziellen werden in den Systemen IANUS.DATA-MANAGEMENT und IANUS.USER-MANAGEMENT gespeichert und umgesetzt und nicht – wie alternativ denkbar wäre – als Informationen innerhalb eines AIPs bzw. DIPs. Der Grund für diese Entscheidung liegt darin begründet, dass sich Einschränkungen für die Bereitstellung im Laufe der Zeit ändern können (z.B. weil sich ein Dateneigentümer nach anfänglich restriktiven Zugangsregelungen später für einen offeneren Weg entscheidet), was eine Änderung der betroffenen AIPs und DIPs zur Folge hätte. Dort werden stattdessen nur die unveränderlichen Rechtsaspekte wie etwa der Urheber oder relevante Gesetze zum Zeitpunkt der Datengenese in den Metadaten festgehalten. Die Einhaltung der von einem Dateneigentümer gewünschten Vorgaben wird also erst auf der Ebene der technischen Systeme realisiert, indem z.B. für bestimmte Funktionen eine Registrierung und Anmeldung erforderlich ist oder für einzelne eingeschränkte Datensammlungen eine manuelle Freischaltung durch den Dateneigentümer oder einen Datenkurator notwendig ist, bevor sie für einen individuellen externen Datennachnutzer sichtbar sind.

11.4 Notfallplan und Exit-Strategie

Es gibt zwei denkbare Szenarien von Gefahren, für die IANUS entsprechende Vorkehrungen treffen

muss. Das erste betrifft den Fall einer kurzfristig eintretenden, schwerwiegenden Katastrophe, wie etwa ein kritischer Hardwareausfall, Brand, Überschwemmung oder Diebstahl. Hierfür ist ein Notfallplan (*disaster recovery plan*) zu entwickeln, der entsprechend der Art und der Schwere des eingetretenen Ereignisses folgende Punkte für den Ernstfall regelt bzw. zentrale Informationen bereithält:

- Kontaktdaten der verantwortlichen IANUS-Mitarbeiter
- Zugriff und Nutzbarkeit des Planes in digitaler und analoger Form
- Angaben, wo sich Ausdrücke und digitale Kopien von zentralen Dokumenten befinden (Hardwareinventar, Benutzernamen und Passwörter, Zugriffsdetails zum Remoteserver, Versicherungen etc.)
- Handlungsanweisungen für unterschiedliche Szenarien beschädigter Hard- und Software
- Garantiebedingungen der Server bei Hardwareversagen, ggf. Kontakte zu externen IT-Dienstleistern
- Dokumentation der Sicherungs- und Backup-Strategien der Produktivsysteme, Server und Daten
- Maßnahmen der aktiven Gefahrenprävention

Das zweite Szenario tritt dann ein, wenn IANUS seinen eigenständigen Betrieb nicht aufrecht erhalten kann – etwa weil die für einen Basisbetrieb notwendigen Finanzmittel nicht mehr vorhanden sind – und daher seine Archivtätigkeit einstellen muss. Zur Abmilderung der Konsequenzen dieser Situation wird eine Exit-Strategie benötigt. In dieser wird festgelegt, dass jährlich ein gewisser Prozentsatz des Gesamtbudgets von IANUS beiseitegelegt wird, um die Überführung des Datenbestandes in ein neues Archiv, welches bereit ist, die digitalen Objekte von IANUS zu übernehmen, zu finanzieren. Damit müssen vor allem die Kosten für die Durchführung eines aufwendigen Ingest-Prozesses abgedeckt werden, bei dem die AIPs von IANUS so vorbereitet werden müssen, dass sie als SIPs für ein neues Repository der Nachfolgeeinrichtung geeignet sind, um dann dort nach massiven Konsistenz- und Integritätsprüfungen in neue AIPs umgewandelt werden zu können. Weitere Gelder werden für die künftige Erreichbarkeit und Präsentation der Daten benötigt, die ganz wesentlich von den technischen Möglichkeiten und der Strategie des neuen Archives abhängen.

Es können dann zwar keine neuen Daten mehr kuratiert, archiviert und bereitgestellt werden, aber der vorhandene Datenbestand wäre weiterhin adäquat gesichert und würde zur Nachnutzung für die Datennutzer auch nach der Abwicklung von IANUS vorgehalten werden.

11.5 Kontrolle der Fachcommunity

Da IANUS seine geplanten Dienstleistungen an den Anforderungen und Wünschen der altertumswissenschaftlichen Fachcommunity in Deutschland ausrichtet, ist es erforderlich Kontrollmechanismen zu implementieren, die dafür Sorge tragen, dass dieses Ziel auch weiterhin adäquat erreicht werden kann.

Zum einen sind Gremien mit Vertretern der Fachcommunity geplant (siehe **Kapitel 8.2**), die in regelmäßigen Abständen zusammenkommen und über den aktuellen Status von IANUS informiert werden. Auf diesen Treffen sollen einerseits Ideen zur strategischen Weiterentwicklung von IANUS diskutiert und beschlossen werden, andererseits aber auch direkte nutzerbezogene

Rückmeldungen und Verbesserungsvorschläge zur Ausgestaltung der Angebote formuliert werden. Umgekehrt ist die Leitung von IANUS diesen Gremien gegenüber dazu verpflichtet, über die Tätigkeiten und Entwicklungen innerhalb eines festgelegten Zeitraumes zu berichten.

Zum anderen ist vorgesehen in unregelmäßigen Abständen zu einzelnen Dienstleistungen und Aspekten des Datenlebenszyklus Umfragen und Nutzungsanalysen durchzuführen, um die Qualität und Quantität der Dienstleistungen optimal auf die Bedürfnisse der Stakeholder anpassen zu können.

12 Metadaten

Metadaten spielen für die Archivierung von digitalen Daten eine besondere Rolle. Technische Metadaten über Formate und Dateieigenschaften werden für die Archivierung und für Migrationsprozesse benötigt; inhaltliche Metadaten sind notwendig, um die Auffindbarkeit, Verständlichkeit und Nachnutzbarkeit der Daten zu gewährleisten. Ein weiterer wichtiger Faktor ist die Qualität der Metadaten: Wenn Dokumentation und Metadatenbeschreibungen über den Inhalt und den Kontext des digitalen Objekts unzureichend sind, können Bedeutungen verloren gehen oder es besteht die Gefahr von Missverständnissen. Die notwendige Qualität der Metadaten bemisst sich dabei an den vorausgesetzten Fähigkeiten der Nutzer, d.h. die Metadaten müssen für die definierte Zielgruppe adäquat sein, damit diese die Daten finden, verstehen und gebrauchen können.

Allgemein können Metadaten, die in der Gesamtheit für die Langzeitarchivierung und die Nachnutzung von Datensammlungen und elektronischen Ressourcen relevant sind, in folgende Gruppen unterteilt werden:

- **Deskriptive Metadaten** ermöglichen das Auffinden von Ressourcen mittels Suchfunktionen und liefern Kontextinformationen, um die Ressource verstehen oder interpretieren zu können. Dazu gehören beispielsweise Titel, Schlagworte und Kurzbeschreibung.
- **Strukturelle Metadaten** stellen Verknüpfungen zwischen verschiedenen Versionen der gleichen Ressource oder verwandten Ressourcen her. Auch Hierarchien und Abhängigkeiten können hiermit gekennzeichnet werden. Somit können die Strukturen von Tabellen einer Datenbank oder auch die Anordnung von Seiten eines Buches beschrieben werden.
- **Administrative Metadaten** enthalten die Informationen, die Archive und Repositorien benötigen, um Daten zu übernehmen, im Prozess der Kuratierung und Erhaltung zu verwalten sowie deren Lesbarkeit zu gewährleisten. Denn neben den technischen Metadaten werden auch Informationen über angewandte Prozesse und deren Akteure benötigt. Dazu gehören Angaben über das Erstellungsdatum einer Ressource, wo diese gespeichert ist sowie eindeutige Identifikatoren. Je nach Sichtweise und Implementierung, können die drei folgenden Metadatenarten zu Technik, Recht und Provenienz auch als ‚Untergruppen‘ von administrativen Metadaten aufgefasst werden.
- **Technische Metadaten** beschreiben die digitalen und physikalischen Eigenschaften einer Ressource, die notwendig sind, um sie zu verwenden oder um eine notwendige Migration vorzunehmen. Dazu gehören Dateiformat, Dateigröße, Enkodierung oder speziell von einem Dateiformat abhängige Angaben, wie z.B. bei Rastergrafiken die Farbtiefe und Auflösung.
- **Rechtliche Metadaten** dienen der Verwaltung von Zugriffsrechten auf die Ressource und umfassen auch Urheber- und Lizenzierungsinformationen.
- **Metadaten zur Provenienz** geben darüber Auskunft, mit welchen Quellen die Ressource erstellt wurde und wodurch diese im Laufe der Zeit verändert wurde. Dazu kann beispielsweise eine Ausgangsbilddatei für ein 3D-Modell gehören, aber auch Angaben zu den verschiedenen Anwendungen, mit denen ein Ergebnis erstellt wurde. Provenienzdaten stellen im Sinne einer Dateihistorie eine wichtige Grundlage dar, um die Authentizität von Daten bewerten zu können, und ermöglichen die Reproduzierbarkeit von Ressourcen.

Die folgenden Abschnitte sollen für IANUS konkrete und spezifische Anhaltspunkte darüber liefern,

wann welche Metadaten benötigt werden, wo sie gespeichert und wie sie verwaltet werden sollen. Außerdem werden relevante Standards und Schemata genannt.

12.1 Vorgaben nach dem OAIS-Referenzmodell

IANUS strebt an, ein OAIS-konformes Archiv zu werden, weshalb einerseits die Anforderungen an die drei Informationspakete – Übergabepakete (SIP), Archivpakete (AIP), Auslieferungspakete (DIP) – berücksichtigt werden müssen (siehe **Kapitel 5.1 und 13**), andererseits aber auch die in dem Standard beschriebenen notwendigen Erhaltungsmetadaten (*preservation description information (PDI)*) zu beachten sind. Letztere enthalten alle Angaben, die für die exakte Identifizierung und die Erhaltung der zu archivierenden digitalen Objekte, der sog. Inhaltsinformationen in einem AIP, benötigt werden. Unter anderem beschreiben sie vergangene und gegenwärtige Zustände der jeweiligen digitalen Objekte, um unbemerkte Veränderungen feststellen zu können, was zur Folge hat, dass alle Prozesse und Migrationen, welche die Daten innerhalb eines Archives erfahren, genau protokolliert werden. Im Ernstfall dienen die Erhaltungsmetadaten zusammen mit den Repräsentationsmetadaten (*representation information*) zur Rekonstruktion von archivierten Daten.

Im Einzelnen setzen sich die Erhaltungsmetadaten aus folgenden fünf Aspekten zusammen:

- **Kontextinformationen** (*context information*) benennen Verbindungen zu Informationen außerhalb eines Informationspakets (z. B. zu anderen digitalen Ressourcen) und beschreiben in welcher Beziehung die digitalen Objekte zu ihrer Umgebung stehen, unter anderem auch, warum sie erzeugt wurden.
- **Provenienzinformationen** (*provenance information*) dienen dazu die Geschichte der digitalen Objekte zu dokumentieren und nachvollziehbar zu machen. Dazu gehören Angaben sowohl zum Ursprung bzw. zur Entstehung einer Information, als auch zu allen vorgenommenen Veränderungen und durchgeführten Prozessen sowie zu der Frage, in wessen Obhut die Inhaltsinformationen waren bzw. sind. Die Provenienzinformationen sind ein wesentlicher Teil zum Nachweis der Authentizität von Daten.
- **Referenzinformationen** (*reference information*) tragen mit Hilfe eindeutiger Bezeichner (*unique identifiers*) dazu bei, dass Informationspakete sowie die darin enthaltenen Inhaltsinformationen und Metadaten eindeutig identifizierbar und unterscheidbar werden und zugehörige Beziehungen (*references*) definiert werden können.
- **Persistenzinformationen** (*fixity information*) stellen sicher, dass die Inhaltsinformation nicht unerlaubt verändert werden und tragen zur Stabilisierung bzw. Authentizität eines Informationsobjektes bei, um zu verhindern, dass ein digitales Objekt den Eindruck erweckt etwas zu sein, was es aber tatsächlich nicht ist. Dazu werden Datenintegritätsprüfungen oder Verifizierungsschlüssel zur Verfügung gestellt, es kommen aber auch spezielle Kodierungsverfahren und Fehlererkennungsmethoden zum Einsatz.
- **Informationen über Zugriffsrechte** (*access rights information*) beinhalten alle rechtlichen Vereinbarungen und Regelungen, die das Informationspaket betreffen, insbesondere halten sie fest, wer mit welcher Berechtigung wie auf die Daten zugreifen darf.

Weiterhin werden in dem OAIS-Standard sog. Erschließungsinformationen (*descriptive information*) bzw. Paketbeschreibungen (*package description*) definiert, die vor allem der Auffindbarkeit, Bestellung und Analyse von Informationspaketen dienen. Häufig können sie aus den Inhaltsinformationen und den Erhaltungsmetadaten abgeleitet werden.

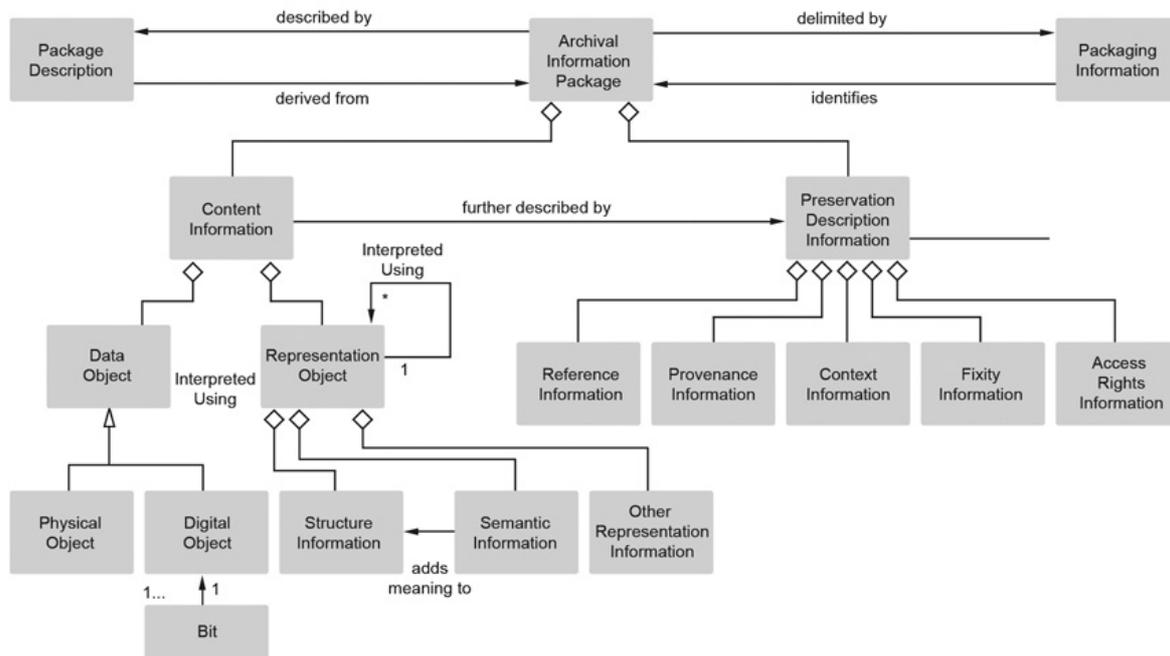


Abb. 7 Übersicht aller Komponenten und Informationseinheiten innerhalb eines AIP nach dem OAIS-Referenzmodell

12.2 Kerninformationen für die Verwaltung von Archivdaten

12.2.1 Verpackungsinformationen

Die Unterscheidung von Metadaten, die relevant für die Langzeitarchivierung von Daten sind (Erhaltungsmetadaten) und solchen, die relevant für das Auffinden und Nachnutzen der Daten sind (Erschließungsinformationen), ist eine grundlegende. Für die Systeme von IANUS ergeben sich daraus einige Anforderungen. Zunächst müssen die Erhaltungsmetadaten und Erschließungsinformationen dauerhaft mit den digitalen Objekten verknüpft werden, wozu die Verpackungsinformation dient. Diese wird in IANUS auf zweierlei Arten festgehalten:

- 1.) Zum einen wird die Zugehörigkeit und Identifizierung von zusammengehörigen digitalen Objekten mit ihren Metadaten durch eine vorgegebene Verzeichnisstruktur sichergestellt, in der alle Komponenten eines vollständigen AIPs enthalten sind (siehe **Kapitel 13.3**). Alle Elemente dieses Verzeichnisses werden in dem Container-Format `.tar` als ein einziges physisches Objekt gespeichert, das eine eindeutige, archivweit einmalige Bezeichnung erhält (siehe **Kapitel 13.3.2**). Innerhalb des Containers gibt es für verschiedene Ordnerhierarchien sowie für einzelne obligatorische Dateien festgelegte Namenskonventionen, so dass jeweils erkennbar ist, ob es sich entweder um Inhaltsobjekte oder um strukturierte Metadaten bzw. unstrukturierte Dokumentationen handelt.
- 2.) Zum anderen werden die Bezüge und Abhängigkeiten zwischen den Inhaltsobjekten und den Metadaten in einer XML-Datei festgehalten, die dem XML-Schema des *Metadata Encoding & Transmission Standard (METS)* entspricht. Dieser Metadaten-Standard wurde speziell für die Beschreibung und Verwaltung von Objekten in digitalen Bibliotheken und Archiven entwickelt und eignet sich in besondere Weise zur Kodierung von Hierarchien und Strukturen innerhalb

von Datensammlungen¹⁷². Ein METS-Dokument besteht aus sieben Hauptabschnitten mit entsprechenden Unterelementen¹⁷³:

- Kopfteil <metsHdr>
- Erschließungsangaben <dmdSec> mit Elementen sowohl für referenzierte <mdREF> als auch für direkt eingetragenen Metadaten <mdWrap>
- Verwaltungsangaben <amd> mit Elementen für technische Metadaten <techMD>, Urheberrechte und Lizenzinformationen <rightsMD>, Angaben zur Quelle <sourceMD> und digitale Herkunftsangaben <digprovMD>
- Dateienabschnitte <fileSec> für Einzeldateien und Dateigruppen <fileGrp>
- Strukturbeschreibungen <structMap>
- Strukturverknüpfungen <smlink>
- Verhalten <behavior>

Der große Vorteil von METS-Dateien ist, dass die deskriptiven und administrativen Metadaten innerhalb eines entsprechenden Abschnittes entweder direkt – als XML-kodierte Metadaten mit unterschiedlichen Namensräumen (z.B. DublinCore, MODS, EAD) bzw. in beliebiger textlicher oder binärer Form – oder indirekt über Verweise auf externe Dokumente und Ressourcen eingebettet werden können. Praktisch hält eine METS-Datei zu jedem Inhaltsobjekt fest, wo sich welche zugehörigen Metadaten befinden und in welcher Form sie jeweils vorliegen, und fungiert damit als ein XML-Container, um unterschiedliche Typen von Metadaten zusammenzuführen.

Da den Verpackungsinformationen eine zentrale Rolle für die gesamte Prozessierung und Kuratierung von digitalen Daten zukommt, müssen beide von IANUS gewählte Formen – die Verzeichnisstruktur und die METS-Datei – möglichst frühzeitig existieren, d.h. im Idealfall bereits von dem Datengeber angelegt werden oder spätestens von einem Datenkurator bei der Erstellung eines validen SIPs (siehe **Kapitel 14.2**).

12.2.2 Erhaltungsmetadaten

Bei den Erhaltungsmetadaten handelt es sich einerseits um rein technische Informationen, die notwendig sind, um digitale Objekte in einem Archiv maschinell zu verarbeiten und zu verwalten. Häufig können sie automatisiert aus den digitalen Objekten ermittelt werden. Sie umfassen entweder dateiunspezifische Angaben – z. B. Bezeichner für Dateiformate in Form von *PRONOM Unique IDs (PUID)*¹⁷⁴, verwendete Softwareversionen, Prüfsummen, Dateigrößen und Zeitstempel – oder formattypische Angaben wie z. B. Bildtiefe bei Rastergrafiken oder Zeichencodierungen bei Textdokumenten. Die Formatidentifikation von einzelnen Dateien spielt dabei aus zwei Gründen eine besondere Rolle: Erstens dient sie zur Gewinnung technischer Metadaten, mit denen die Archivobjekte angereichert werden können, so dass ihre Autarkie gestärkt wird. Und zweitens bildet die Identifikation von Formaten die Grundlage für die Entscheidung, welches Ausgangsformat mit welchem Konvertierungstool in welches Zielformat migriert werden soll.

Andererseits sieht der OAIS-Standard bei den Erhaltungsmetadaten auch weitere Informationen

¹⁷² <http://www.loc.gov/standards/mets/>

¹⁷³ http://www.loc.gov/standards/mets/METSOverview.v2_de.html

¹⁷⁴ Siehe **S. XXX Anm. YYY. (Kapitel 14.4.2)**

vor, welche die Zugriffsrechte, den Entstehungskontext und die Provenienz, die Verarbeitung sowie die Beziehungen zu anderen Ressourcen betreffen. Ein Teil dieser Angaben muss von dem Datengeber geliefert und durch Datenkuratoren kontrolliert, ggf. ergänzt werden. Ein Großteil betrifft jedoch Maßnahmen, die IANUS an den Datensammlungen durchführt und die ab dem Zeitpunkt der Datenübergabe möglichst genau und reproduzierbar protokolliert werden müssen. IANUS wird hierfür den Standard *PREMIS (PREservation Metadata: Implementation Strategies)*¹⁷⁵ umsetzen, der für die vom OAIS vorgegebenen Erhaltungsmetadaten konkrete Implementierungshinweise liefert. Das Datenmodell von PREMIS (Version 3.0)¹⁷⁶ beschreibt fünf unterschiedliche Entitäten, die zur vollständigen Abbildung des Lebenszyklus digitaler Objekte essentiell sind:

- Intellektuelle Entitäten (*intellectual entities*): Zusammengehörige Inhalte, die zu Zwecken des Managements und der Beschreibung als eine einzige intellektuelle Einheit betrachtet werden, z.B. ein Foto, eine Datenbank oder eine Zeichnung.
- Objekte (*object entities*): Die zu archivierenden digitalen Objekte in Form von physischen Dateien, Representationen oder Bitstreams.
- Ereignisse (*events entities*): alle Ereignisse, die während der gesamten Existenz digitaler Daten eintreten können und die ihr Management und ihre Erhaltung betreffen.
- Akteure (*agents entities*): Die Personen, Institutionen und Anwendungen, welche die Ereignisse durchführen und auf die Daten einwirken.
- Berechtigungen (*rights entities*): Berechtigungen und rechtliche Grundlagen (z. B. Copyrights, Lizenzen), auf deren Basis Akteure bestimmte Ereignisse durchführen dürfen.

Für die letzten vier Entitäten stellt PREMIS zudem ein Datenwörterbuch (*data dictionary*) mit sogenannten semantischen Einheiten (*semantic units*) bereit, die als Vorschlagskatalog für ein Kernset an Metadatenelementen verstanden werden können¹⁷⁷. Sofern die Spezifizierungen für eine Entität nicht ausreichend detailliert sind (v.a. bei Akteuren und Berechtigungen), weil sich die PREMIS-Elemente vor allem auf die Archivierungsprozesse beziehen, muss das Schema durch weitere spezialisierte Metadatenschemata erweitert werden.

Da für PREMIS bereits ein XML-Schema existiert und die semantischen Einheiten von PREMIS direkt als XML-Elemente implementiert werden können, wird IANUS für jede Datensammlung eine *premis.xml*-Datei anlegen, um dort die notwendigen Erhaltungsmetadaten zu erfassen. Auch diese sollte wie die METS-Datei möglichst frühzeitig existieren, da sie einen essentiellen Bestandteil eines vollständigen Objektes während des gesamten Lebenszyklus ausmacht und in ihr z.B. die rechtlichen Vorgaben und Lizenzen eines Datengebers zur Veränderung und Bereitstellung seiner Daten festgehalten werden. Zu einem späteren Zeitpunkt werden weitere Daten in der PREMIS-Datei gespeichert, etwa die technischen Formatinformationen, die durch verschiedene Tools automatisch extrahiert werden können (siehe **Kapitel 14.4.2**).

IANUS wird sowohl den METS-Standard als auch den PREMIS-Standard umsetzen. Da beide Standards sich inhaltlich partiell allerdings überschneiden, muss festgelegt werden, an welcher Stelle genau welche Informationen erfasst werden, um eine redundante Speicherung zu

¹⁷⁵ <http://www.loc.gov/standards/premis/>

¹⁷⁶ <http://www.loc.gov/standards/premis/v3/index.html>

¹⁷⁷ <http://www.loc.gov/standards/premis/v3/premis-3-0-datadictionary-only.pdf>

vermeiden. Auch wenn es prinzipiell möglich ist, PREMIS innerhalb von METS zu benutzen¹⁷⁸, wird IANUS sich für einen anderen Weg entscheiden: alle Angaben, die in PREMIS erfasst werden können, werden auch dort in den entsprechenden Elementen erfasst. Sofern es Angaben gibt, die auch durch METS abgedeckt werden, wird dort ein externer Verweis auf den entsprechenden Strukturabschnitt in der PREMIS-Datei eingetragen. Analog gilt, dass etwaige Inhalte, die zunächst in einem METS-Dokument enthalten sind, aber ebenso gut in einem PREMIS-Dokument verwaltet werden können, dorthin überführt werden (z.B. rechtliche und technische Angaben).

12.2.3 Erschließungsinformationen

Erschließungsinformationen dienen der inhaltlichen Beschreibung und zum Verständnis von Datensammlungen und sind ein zentrales Mittel zur Auffindbarkeit von Archivbeständen. Letzteres erfolgt bei IANUS über ein Online-Datenportal und verschiedene Web-Services. Es sind diejenigen Metadaten, die am stärksten disziplinspezifisch sind, also die fachlich-methodischen Besonderheiten der altertumswissenschaftlichen Forschungsdaten abbilden müssen. Als Minimalangaben, die ein Datengeber für eine Datensammlung zur Verfügung stellen muss, werden die 15 Kerninformationen von Dublin Core als qualifizierte Elemente¹⁷⁹ verlangt. Auch wenn dieses ursprünglich ein bibliographisches Datenformat ist, kommt es bei einer Vielzahl anderer, mit IANUS vergleichbaren Datenzentren zum Einsatz (u.a. ADS, eDNA, tDAR und OpenContext) und existieren aufgrund der weiten Verbreitung bereits mehrere Mappings auf andere Schemata.

Neben *Qualified Dublin Core* sind weitere Standards für IANUS relevant und sollen so weit wie möglich und sinnvoll berücksichtigt werden:

- ADeX – Standard für den Austausch archäologischer Fachdaten¹⁸⁰
- Encoded Archival Description (EAD)¹⁸¹
- Lightweight Information Describing Objects (LIDO)¹⁸²
- CARARE Europeana Data Model (CARARE-EDM)¹⁸³
- MIDAS Data Standard for Historic Environment Information¹⁸⁴
- CIDOC Conceptual Reference Model (CIDOC-CRM)¹⁸⁵
- da|ra Metadatenchema 3.1 (für die Registrierung von DOIs)¹⁸⁶

Ähnlich wie die administrativen Metadaten können auch die beschreibenden Metadaten in einem METS-Dokument entweder direkt erfasst werden oder durch Verweise auf externe Ressourcen eingebunden werden. Da innerhalb von IANUS die Angaben nach Dublin Core die Standard- bzw.

¹⁷⁸ <http://www.loc.gov/standards/premis/premis-mets.html>

¹⁷⁹ <http://dublincore.org/documents/dces/> und <http://dublincore.org/documents/dcmi-terms/>

¹⁸⁰ <http://www.landesarchaeologen.de/verband/kommissionen/archaeologie-und-informationssysteme/projektarbeitsgruppen/adex/>

¹⁸¹ <http://www.loc.gov/ead/>

¹⁸² <http://network.icom.museum/cidoc/working-groups/lido/>

¹⁸³ <http://pro.europeana.eu/page/carare-edm>

¹⁸⁴ <http://fishforum.weebly.com/midas-heritage-standard.html>

¹⁸⁵ <http://www.cidoc-crm.org/>

¹⁸⁶ <http://www.da-ra.de/de/fuer-datenzentren/daten-registrieren/doi-und-metadaten/>

Minimalanforderung für archivierte Daten darstellen wird, d.h. für alle Datensammlungen vorliegen müssen, werden diese direkt innerhalb des METS-Dokumentes gespeichert. Für alle darüber hinausgehenden Informationen, die ggf. nur in Einzelfällen vorliegen, wird stattdessen auf separate Metadatendokumente verwiesen.

Da in der Regel vor allem die Datenproduzenten bzw. Datengeber die notwendigen Kenntnisse über ihre Forschungsdaten besitzen, wird ein Großteil der Erschließungsinformationen von diesen beigesteuert werden müssen. Für die Qualitätssicherung, z. B. hinsichtlich der einheitlichen Verwendung von Begriffen, ist eine Überprüfung durch die Datenkuratoren von IANUS notwendig¹⁸⁷. Idealerweise geschieht die Übertragung der geforderten Metadaten über ein in die Homepage von IANUS eingebundenes Online-Formular, das als technische Schnittstelle zwischen externen Datengebern und dem internen Archiv-Management-System dient. Alternativ werden aber auch geeignete Formulare zum Download angeboten, um eine Offline-Bearbeitung der beschreibenden Metadaten zu ermöglichen. Für den Fall, dass bereits Informationen in standardisierter Struktur vorliegen, wird IANUS die Möglichkeit zum (halb)automatisierten Import dieser Daten entwickeln.

Alle Metadaten müssen in deutscher Sprache vorliegen, wobei die Kerninformationen zusätzlich auch in Englisch (inkl. Zusammenfassungen) eingefordert oder übersetzt werden, um einem internationalen Nutzerkreis das Suchen und Finden der IANUS-Datenbestände zu ermöglichen. Angaben, die ein im Ausland durchgeführtes Projekt bzw. dessen Datensammlung beschreiben, sollen darüber hinaus in der relevanten Landessprache abgegeben werden, damit diese über Schnittstellen an landesspezifische Portale und Metadaten-Aggregatoren weitergegeben werden können. Die zukünftige Verwaltung der Metadaten wird also mehrsprachig erfolgen.

12.2.4 Metadaten für Projekte und Datensammlungen

Aus den Spezifikationen von PREMIS und *Qualified Dublin Core* ergeben sich verpflichtende Mindestangaben zur inhaltlichen Beschreibung von Projekten und den durch diese generierten Datensammlungen. Hinzu kommen die Angaben, die zur Registrierung von DOIs über da|ra (siehe **Kapitel 7.4**) notwendig sind, und die Kriterien, welche sich in der Praxis in anderen fachspezifischen Archiven bewährt haben¹⁸⁸. Aus der Summe dieser Anforderungen lässt sich ein verpflichtender Kernsatz an Metadaten zur Beschreibung einer Datensammlung als Ganzes ableiten. Dieser ist von Datenproduzenten bzw. Datengebern vor der Übergabe von digitalen Daten an IANUS bereitzustellen¹⁸⁹:

- Identifizierung – Projekttitle
- Identifizierung – Alternativtitle
- Identifizierung - Projektnummer(n)
- Kurzbeschreibung
- Veröffentlichung – Ergebnisse
- Dauer – Projekt
- Dauer – Datenbestand
- Rechtliches – Urheberrechte

¹⁸⁷ Zu Qualitätskriterien von Metadaten vgl. <http://pro.europeana.eu/taskforce/metadata-quality>

¹⁸⁸ Z. B. beim ADS in England: <http://archaeologydataservice.ac.uk/advice/DatasetlevelMetadata> oder bei EDNA in den Niederlanden: IANUS – KOLBMANN (2014) 20 f. 58 f.

¹⁸⁹ Veröffentlicht unter <http://www.ianus-fdz.de/it-empfehlungen/archivierung>

- Schlagworte – Fachdisziplinen
- Schlagworte – Inhalt
- Schlagworte – Methoden
- Ausdehnung – Geografisch-1
- Ausdehnung – Geografisch-2
- Ausdehnung – zeitlich
- Primärforscher – Person
- Eigentümer – Organisation
- Finanzierung
- Veröffentlichung – Projektdaten
- Rechtliches - Lizenzgeber
- Rechtliches - Datenschutz
- Quellen - Ältere
- Quellen - Zugehörige
- Sprache
- Art der Daten
- Vollständigkeit
- Dateiformate
- Zugriffsrechte
- Signatur Metadaten

Bei der technischen Implementierung zur Erfassung und Verarbeitung der Metadaten werden die genannten Kategorien granularer als hier dargestellt durch relationale Tabellen mit spezifischen Attributen, für die ggf. Codelisten und Normdaten vorgegeben werden (siehe **Kapitel 12.3**), umgesetzt.

12.2.5 Metadaten für Einzeldateien

Die bisher beschriebenen Metadaten beziehen sich jeweils auf eine gesamte Datensammlung eines Projektes und sie sind dementsprechend allgemeiner gehalten. Metadaten besitzen aber auch auf der Ebene von Einzeldateien eine hohe Relevanz. Wenn beispielsweise eine Tabelle in ein anderes Format migriert werden soll, muss sichergestellt sein, dass relevante Eigenschaften wie die Anzahl von Spalten und Zeilen erhalten bleiben. Um verstehen zu können, was in den einzelnen Zellen einer Tabelle eingetragen wurde, muss dokumentiert werden, woher die Daten stammen, mit welchen Methoden sie erzeugt oder wie sie berechnet wurden.

Teilweise können die Metadaten für Einzeldateien von den Angaben der übergeordneten Datensammlung geerbt werden. Wenn es sich um technische Angaben handelt, können sie teilweise automatisch aus einem digitalen Objekt extrahiert werden. So kann im Idealfall z. B. das Aufnahmedatum einer Fotografie mit dem Entstehungsdatum der Bilddatei gleichgesetzt werden oder ist bereits in den IPTC-Daten eingetragen. Allgemeine, für jeden Dateityp gültige technische Metadaten können daher folgende Aspekte umfassen:

- Dateigröße
- Dateiformat
- Erstellende Anwendung / Software (mit Versionsbezeichnung)
- Art der Daten
- Hardware
- Konfigurationseinstellungen / Geräteeinstellungen
- Kalibrierungsdaten
- Erstellungs- und Bearbeitungsdatum

Eine besondere Rolle spielt bei der Erfassung der technischen Merkmale die sogenannte Formatidentifikation, wofür IANUS verschiedene etablierte Dienste nutzen wird (siehe **Kapitel 14.4.2**). Einige, vor allem beschreibende dateibezogene Metadaten müssen aber direkt vom

Datengeber übermittelt werden, da nur dieser oder die ursprünglichen Autoren die Daten ausreichend kennen.

Welche Metadaten für Einzeldateien benötigt werden, hängt von dem jeweiligen Dateityp (z.B. Rastergrafiken, Vektorzeichnungen, Geophysik-Messungen, Fernerkundungsdaten, Datenbanken, Tabellen, GIS, 3D-Daten, Video, Audio) und von der verwendeten Methode bei der Erzeugung ab, so dass an dieser Stelle keine einheitlichen Standards und Minimalanforderungen formuliert werden können. Dieser Tatsache wird jedoch mit den IT-Empfehlungen¹⁹⁰ von IANUS Rechnung getragen, wo in den Kapiteln „Dateiformate“ und „Forschungsmethoden“ formatspezifische Angaben zu erforderlichen Metadaten aufgezählt werden.¹⁹¹

12.3 Kontrollierte Vokabulare

Für Metadaten, die nicht automatisch aus einem Informationspaket bzw. aus Einzeldateien extrahiert werden können, sollen so weit wie möglich kontrollierte Vokabulare und Normdaten zum Einsatz kommen und über entsprechende technische Schnittstellen eingebunden werden. Dies hat mehrere Vorteile:

- schnellere Eingabe von Informationen
- einheitliche Angaben zur besseren Auffindbarkeit und Vergleichbarkeit von Informationen
- Vereinfachung einer mehrsprachigen Datenhaltung und Präsentation
- Möglichkeiten zur Vernetzung von Begriffen und Konzepten

Für PREMIS wird in dem Data Dictionary bei den jeweiligen semantischen Einheiten angegeben, ob ein kontrolliertes Vokabular empfohlen wird oder nicht. Eine Liste von validen und qualitätvollen kontrollierten Vokabularen wird zudem von der Library of Congress¹⁹² veröffentlicht.

Da im Bereich der deutschsprachigen Altertumswissenschaften nur wenige fachspezifische Thesauri und Normdaten existieren, die öffentlich zugänglich sind und damit eine große Akzeptanz besitzen, muss auf verschiedene Ressourcen zurückgegriffen werden. Diese bilden jeweils spezifische thematische Schwerpunkte ab:

- für antike und moderne Ortsangaben: [iDAL.gazetteer](#)¹⁹³, [Pleiades](#)¹⁹⁴, [Geonames](#)¹⁹⁵
- für Personen und Organisationen: Gemeinsame Normdatei (GND)¹⁹⁶, [OrcID](#), [VIAF](#)
- für archäologische Konzepte: [iDAL.vocab](#)¹⁹⁷, [Getty Arts and Architecture Thesaurus \(de\)](#)¹⁹⁸, [English Heritage Data](#)¹⁹⁹²⁰⁰

¹⁹⁰ <http://www.ianus-fdz.de/it-empfehlungen/>

¹⁹¹ Vgl. auch <http://archaeologydataservice.ac.uk/advice/FilelevelMetadata>

¹⁹² <http://id.loc.gov/vocabulary/preservation.html>

¹⁹³ <http://gazetteer.dainst.org>

¹⁹⁴ <http://pleiades.stoa.org>

¹⁹⁵ <http://www.geonames.org>

¹⁹⁶ <http://www.dnb.de/gnd>

¹⁹⁷ <http://archwort.dainst.org/thesaurus/de/vocab>

¹⁹⁸ <http://www.aat-deutsch.de>

¹⁹⁹ <http://www.heritagedata.org/blog/vocabularies-provided>

²⁰⁰ <http://thesaurus.historicengland.org.uk/>

- für bibliographische Angaben: iDAI.bibliography/ZENON²⁰¹, Propylaeum
- für Chronologie-Begriffe: Perio.do²⁰², iDAI.chronontology²⁰³, Zeitstrahl der Landesarchäologien²⁰⁴

Sofern es zugunsten einer einheitlichen Verschlagwortung der Datenbestände notwendig sein sollte, wird IANUS entweder Übersetzungen für bzw. Neueinträge in bestehende Thesauri beisteuern oder ggf. auch eigene kontrollierte Vokabulare erstellen und verwalten. Bei diesem Vorgehen wird auf nationaler Ebene sowohl eine enge Abstimmung mit den Systemen der iDAI.welt²⁰⁵ als auch mit der AG Thesaurusfragen innerhalb des Verbandes der Landesarchäologen²⁰⁶ angestrebt, für einen internationalen Austausch der bestehende Kontakt mit dem Archaeology Data Service in England und edna in den Niederlanden genutzt²⁰⁷.

12.4 Nachweis und Bereitstellung von Archivdaten

Außer für die Kuratierung und Archivierung von digitalen Datensammlungen wird ein Teil der oben beschriebenen Metadaten auch für deren Nachweis und Bereitstellung genutzt. IANUS verfolgt das Ziel, möglichst viele Daten und Inhalte so offen wie möglich für externe Nutzer online zu präsentieren und z.B. zum Herunterladen anzubieten. Hierfür werden sowohl Informationen zum Suchen und Finden als auch Informationen zum Verständnis der Dateien benötigt. Dabei spielen vor allem die deskriptiven Metadaten eine große Rolle, mit denen die primären Fragen zum Was?, Wo?, Wann?, Wer?, Wie? und Warum? beantwortet werden können und die sich z.B. für eine Erschließung mittels facettiertem browsen besonders gut eignen. Wichtig sind darüber hinaus diejenigen Metadaten, welche die Authentizität von Daten belegen, wie etwa Herkunftsangaben, Urheber oder Checksummen, und solche, welche die rechtlichen Rahmenbedingungen für eine Nachnutzung definieren. Administrative oder technische Metadaten sollen weitgehend von der Bereitstellung ausgeschlossen werden, um die Informationsfülle auf die altertumswissenschaftlich relevanten Angaben zu reduzieren. Während für einzelne oder mehrere Dateien in Datensammlungen von Datengebern Zugriffsbeschränkungen festgelegt werden können, sollen die beschreibenden Metadaten für eine Datensammlung prinzipiell immer frei und ohne Einschränkung zugänglich gemacht werden. Eine Ausnahme bilden lediglich sensible Angaben, wie bestimmte geografische oder personenbezogene Angaben, die entweder ganz verborgen bleiben müssen oder nur von registrierten Nutzern mit entsprechenden Rechten eingesehen werden dürfen.

Zusätzlich zu den bei IANUS archivierten Datenbeständen sollen auch verteilte externe Ressourcen

²⁰¹ <http://zenon.dainst.org>

²⁰² <http://perio.do>

²⁰³ <http://chronontology.i3mainz.hs-mainz.de/>

²⁰⁴ http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_Thesauri/Zeitstrahl_V01.pdf

²⁰⁵ <http://www.dainst.org/forschung/forschung-digital/idai.welt>

²⁰⁶ <http://www.landesarchaeologen.de/verband/kommissionen/archaeologie-und-informationssysteme/projektarbeitsgruppen/thesaurusfragen>

²⁰⁷ Siehe auch Appendix A in ARIADNE – D3.2 report on project standards (2013): <http://www.ariadne-infrastructure.eu/eng/Resources/D3.2-Report-on-project-standards>

über einen Online-Katalog zusammengeführt und nachgewiesen werden (siehe **Kapitel 6.1.1, Punkt 3**). Diese Daten, bei denen es sich im Kontext von IANUS im strengen Sinne nicht um Metadaten, sondern um selbständige Daten handelt, enthalten Informationen zu

- Aktivitäten, d.h. Maßnahmen, Untersuchungen und Projekten
- Akteuren, d.h. Personen, Gruppen, Institutionen und Einrichtungen
- Archäologischen Fundstellen, Bauwerken/Monumenten und Archäologieflächen
- Datensammlungen mit ihren analogen und/oder digitalen Ressourcen

Für die zentrale Bündelung solcher Angaben wird ein minimaler Satz an Metadaten benötigt, der sich soweit sinnvoll an den Kerninformationen für die archivierten Datensammlungen orientiert. Daneben existieren verschiedene Standards zur Datenmodellierung, die für einen interoperablen Datenaustausch berücksichtigt werden müssen. Unter Berücksichtigung dieser Bedingungen und existierender fachspezifischer und generischer Lösungen hat Wibke Kolbmann 2013 in ihrer Masterarbeit bereits die wesentlichen Dokumentationsanforderungen und Metadaten-Attribute für die genannten Kategorien des Nachweiskatalog gesammelt, welche sehr gut als Basis für eine Umsetzung genutzt werden können.²⁰⁸

Da die Informationen aus externen Systemen nicht manuell in den Nachweiskatalog eingetragen werden, sondern aus bereits bestehenden Systemen geharvestet bzw. importiert werden sollen, gilt es entsprechende Schnittstellen zu implementieren.

12.5 Metadaten für den Datenaustausch und Webservices

Die Gesamtheit der von IANUS erfassten und verwalteten Metadaten wird vermutlich keinem bestimmten Metadatenschema entsprechen, da die verschiedenen existierenden Standards und Schemata meist nur Teilbereiche abdecken, während in IANUS administrative, technische und inhaltliche Informationen für die Archivierung und Auffindbarkeit von Datenbeständen zusammenfließen.

Um Daten sowohl aus anderen Systemen importieren als auch für andere Systeme zur Verfügung stellen zu können, werden neben technischen Schnittstellen auch Mappings auf bestimmte Metadatenschemata notwendig sein. Aktuell besitzen folgende Infrastrukturen eine besondere Relevanz für IANUS, so dass gegenwärtig geplant ist, die dort verwendeten Metadatenmodelle zu unterstützen:

- iDAI.bibliography/ZENON und Propylaeum für den Nachweis der Archivbestände in Bibliothekskatalogen und –verbänden
- ARIADNE für einen europäischen, fachspezifischen Datenaustausch, insbesondere mit dem ADS und eDNA
- DARIAH-DE zur Registrierung von Daten in der dortigen Collections Registry²⁰⁹
- CARARE-EDM für den Import in die Deutsche Digitale Bibliothek und Europeana als zentrale Aggregatoren von Metadaten zum kulturellen Erbe

²⁰⁸ IANUS – KOLBMANN (2014).

²⁰⁹ <http://colreg.de.dariah.eu>

- ADEX für den Austausch mit den Landesdenkmalämtern
- LIDO für den Austausch mit Museen und Objektsammlungen
- da|ra für die Registrierung von DOIs (siehe **Kapitel 7.4**)

Mittelfristig ist auf technischer Ebene geplant, Schnittstellen und Technologien zu implementieren, die Austauschprotokolle wie OAI-PMH unterstützen und die Publikation der Daten für das Semantic Web sowie als Linked Open Data vereinfachen, so dass andere Projekte und Institutionen automatisiert auf die Metadaten zugreifen und diese weiterverarbeiten können. (siehe **Kapitel 7.3**)

13 Strukturierung und Spezifikation der Datenpakete

13.1 Transferpaket (TP)

Die Hauptaufgabe von IANUS wird die Langzeitarchivierung digitaler Datensammlungen sein, die sich aus einem oder mehreren digitalen Objekten zusammensetzen können. Jedes Datenpaket entspricht bzw. enthält dabei genau eine Datensammlung. Den Umfang und den Inhalt eines Datenpaketes legen in der Regel die Datengeber selbständig fest, d.h. ein Projekt oder eine Institution entscheidet, ob es seine Forschungsdaten in einer einzigen oder in mehreren separaten Datenpaketen archivieren und bereitstellen will. Dabei kann es in manchen Fällen nach Rücksprache mit IANUS sinnvoll sein, ein sehr großes und sehr heterogenes Datenpaket in mehrere kleinere aufzuteilen und diese durch Referenzen miteinander in Beziehung zu setzen.

Eine Grundanforderung an ein Informationspaket lautet, dass die entsprechenden Forschungsdaten zusammen mit zugehörigen Metadaten in geeigneter Form vorliegen, damit sie künftig in technischer wie inhaltlicher Hinsicht nachgenutzt werden können. Da zum gegenwärtigen Zeitpunkt davon auszugehen ist, dass in den meisten Fällen die Datengeber ihre Datensammlungen trotz von IANUS publizierter spezifischer Vorgaben nicht in einer validen Form an IANUS übergeben werden, beginnt demzufolge die Archivierungsarbeit mit dem Verfahren, welches die übertragenen unveränderten Daten, das sog. *Transferpaket (TP)*, in ein Datenpaket überführt, das IANUS sinnvoll und nachhaltig kuratieren, archivieren und bereitstellen kann. Dieser Vorgang zur Erstellung eines validen SIP wird als Pre-Ingest bezeichnet (siehe **Kapitel 14.3**). Die dabei erforderlichen Arbeitsschritte sollen einerseits das Transferpaket nur so wenig wie möglich verändern, um die originale Datenstruktur des Datengebers weitestgehend zu erhalten, andererseits aber so umfangreich wie nötig sein, damit ein valides SIP entsprechend der Vorgaben und Spezifikationen von IANUS entsteht. Wichtige zu beachtende Punkte sind dabei

- die sinnvolle Bereinigung bzw. Auswahl der zu archivierenden Daten
- die nachvollziehbare Strukturierung und Benennung von Ordnern und Dateien
- die Klärung und Dokumentation von rechtlichen Fragen und Lizenzen
- die Vollständigkeit der inhaltlichen Metadaten
- die Unversehrtheit der Daten (bzgl. Viren und Schadsoftware sowie Speicherfehlern etc.)
- die Abwesenheit von Zugriffseinschränkungen und Passwörtern

Auch wenn prinzipiell die Erzeugung eines validen SIP selbständig durch einen Datengeber erfolgen kann, zeigt die Praxis in anderen Forschungsdatenarchiven, dass häufig geschulte Datenkuratoren in enger Rücksprache mit und unter Zuarbeit durch den Datengeber diese Aufgabe übernehmen. Nur durch eine frühzeitige Kontrolle und ggf. Korrektur der Daten durch IANUS-Mitarbeiter ist vor dem eigentlichen Ingest-Verfahren zu gewährleisten, dass trotz der Vielzahl von Fachdisziplinen, Methoden und Datenarten in den Altertumswissenschaften die Einheitlichkeit und die Qualität des wachsenden Archivbestandes gesichert bleibt. Erst für das am Ende des Pre-Ingest-Prozesses aus dem TP erzeugte valide SIP übernimmt IANUS die archivalische Verantwortung und insofern bildet erst das SIP (nicht das TP) den maßgeblichen Gegenstand für den Datenübergabevertrag (siehe **Kapitel 10.1.2**). Vorherige Zustände der TP müssen daher nicht

in der gleichen Sorgfalt und Qualität dokumentiert werden, wie dies für das SIP, das AIP und das DIP gefordert ist.

13.2 Submission Information Package (SIP)

Trotz aller Unterschiede im Einzelnen existieren allgemeine Grundsätze dazu, wie ein valides SIP aussehen sollte. Danach gibt es prinzipiell zwei unterschiedliche Sichtweisen auf ein SIP:

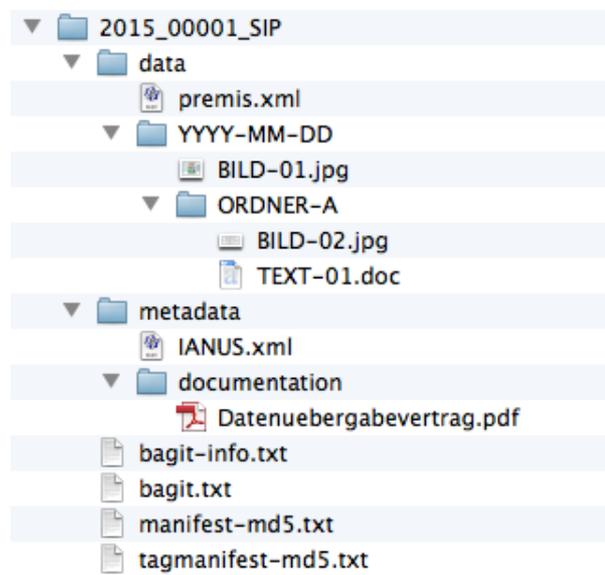
1. Die **physikalische Sicht** betrachtet die Ablage- bzw. Ordnerstruktur, in der die digitalen Dateien physikalisch organisiert und benannt werden. Nach der bewährten Praxis besteht ein SIP aus einer manifestierten XML-Datei, welche dessen physikalische Struktur wiedergibt, einer weiteren XML-Datei, welche die beschreibenden Metadaten enthält, sowie einen Ordner, in dem sich die eigentlichen digitalen Objekte befinden. Die beiden notwendigen XML-Dateien werden dabei im Rahmen des Pre-Ingest-Verfahrens durch IANUS bzw. dessen Systeme erzeugt. Außerdem wird in diesem Schritt ein notwendiger eindeutiger Bezeichner (UUID) vergeben.
2. Die **logische Sicht** gibt Antworten auf die Fragen "Welche Art von Informationen befindet sich in den physischen Ordnern und Dateien?" und "Welche Metadaten-Standards werden verwendet?" Als gemeinsame logische Struktur für die validen IPs wird IANUS sich an den Ergebnissen aus dem Projekt „DA-NRW“ orientieren, die wiederum durch die Standards METS²¹⁰, PREMIS²¹¹ und EAD²¹² beeinflusst sind.

Die erforderlichen und als Teil des SIP mitarchivierten Metadaten werden in **Kapitel 12.1** beschrieben.

13.2.1 Grundstruktur

Im Folgenden wird die Spezifikation und Struktur eines validen SIPs von IANUS beschrieben²¹³, das immer verpackt in einem der zulässigen Containerformate (.tgz | .zip | .tar) vorliegen muss. Der Dateiname des Containers inklusive der Dateierweiterung wird technisch als *ContainerName* bezeichnet.

Das entpackte SIP enthält auf der ersten Hierarchieebene einen Ordner, der denselben Namen trägt wie der Name des SIP ohne die Dateierweiterung. Diese Zeichenkette wird als



²¹⁰ METS provides a means of associating the metadata related to an object and describes the relationships with other objects. <http://www.loc.gov/standards/mets/>

²¹¹ PREservation Metadata: Implementation Strategies: <http://www.loc.gov/standards/premis/>

²¹² Encoded Archival Description, is a non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment. <http://www.loc.gov/ead/>

²¹³ Die Spezifikation des SIP von IANUS ist angelehnt an die Definition eines SIP in der DA-NRW-Suite: https://github.com/da-nrw/DNSCore/blob/master/ContentBroker/src/main/markdown/specification_sip.de.md

OriginalName bezeichnet und besitzt bei IANUS folgendes Schema:

<Jahr (YYYY)>_<5-stellige Nummer (#####)>_SIP z.B. 2015_00001_SIP

Da der *OriginalName* innerhalb der IANUS-Archivsysteme einmalig ist, kommt ihm eine spezielle Bedeutung zu. Wird zu einem späteren Zeitpunkt ein weiteres Paket mit einem bereits vergebenen Originalnamen eingeliefert, wird dieses Paket als Ergänzung zu einem bereits vorhandenen Objekt und nicht als neues, eigenständiges Paket behandelt.

Der Ordner wiederum enthält Checksummen im sog. BagIt-Format²¹⁴. Dabei handelt es sich um vier, in ihrer Struktur festgelegten Dateien, in denen alle Ordner und Dateien mit individuellen Prüfsummen aufgelistet werden. Diese Angaben dienen dem Zweck der Integritätsprüfung, d.h. sie garantieren, dass die Daten beim Transfer vom Datengeber zu IANUS bzw. von IANUS zur Bitstream Preservation bei einem externen technischen Partner (siehe **Kapitel 7.2**) nicht verändert wurden und die Datenübertragung fehlerfrei erfolgte. Diese Prüfsummen sind daher bereits bei einem TP relevant und müssen schon in diesem Kontext erzeugt werden. Zu einem späteren Zeitpunkt, wenn die Kuratierung von Daten abgeschlossen ist und ein finales AIP erzeugt werden soll, werden diese Einträge abschließend aktualisiert bzw. durch neue Werte ersetzt.

Schließlich befinden sich direkt unterhalb der obersten Hierarchieebene ein *data*-Verzeichnis und ein *metadata*-Verzeichnis. Die hier enthaltenen individuellen Datei- und Ordnernamen müssen UTF-8 enkodiert sein und sollten nach Möglichkeit keine Sonderzeichen enthalten. Als Pfadseparator wird ein Unix Style Slash verwendet.

13.2.2 *data*-Verzeichnis

Das *data*-Verzeichnis ist der Ort, in dem die eigentlichen zu archivierenden Forschungsdaten enthalten sind. Für ihn werden nur wenige, allgemeine Vorgaben formuliert. Zunächst muss eine *premis.xml*-Datei existieren, die in ihrem formalen Aufbau den PREMIS-Standard erfüllt. In ihr werden in standardisierter Weise alle für die digitale Langzeitarchivierung erforderlichen Metadaten beschrieben, so dass die Datei einerseits der Nachvollziehbarkeit der Historie von Informationspaketen und andererseits der Übermittlung von formalisiert festgelegten Vereinbarungen zwischen dem Datengeber und IANUS dient. Hier kann z.B. festgehalten werden, wie Daten migriert oder bereitgestellt werden sollen. Die PREMIS-Datei wird vom Archiv-Management-System eingelesen, ausgewertet und die Angaben in entsprechende Handlungen übersetzt. Anschließend wird sie während der Erstellung des AIP mit weiteren Informationen angereichert und in aktualisierter Version mit diesem mitarchiviert.

Ebenfalls eingetragen in die *premis.xml* wird ein eindeutiger technischer Identifier, der während des PreIngest-Prozesses einem SIP zugewiesen wird und der später zur Erzeugung der DOIs als Persistent Identifier (siehe **Kapitel 7.4**) genutzt werden kann. Ein so gebildeter Satz aus Identifiern könnte beispielsweise so aussehen:

- Techn. Identifier: 000.jah37w
- Abgeleitete DOI: 10.13149/000.jah37w-q

Neben der PREMIS-Datei gibt es einen oder mehrere Unterordner, die als Bezeichnung eine Datumsangabe in der Form <YYYY>-<MM>-<DD> tragen. Das Datum entspricht dabei dem Tag

²¹⁴ <https://de.wikipedia.org/wiki/BagIt>

des Eingangs von Daten bei IANUS, die in diesem Unterordner enthalten sind. Sollte es zu einem inhaltlich zusammengehörenden SIP mehrere Teillieferungen geben (z.B. weil ein Teil über ein Web-Protokoll übertragen wurde und ein anderer Teil postalisch auf einem optischen Datenträger zugesandt wurde), lassen sich die Daten in den jeweiligen Datumsordnern ablegen. Sofern es am Ende der Erzeugung eines validen SIP mehrere Datumsordner gibt, muss überprüft werden, dass sich die Einträge in dem BagIt-Manifest, in der PREMIS-Datei und in den zugehörigen Metadaten auf alle Datumsordner beziehen und nicht nur auf einen.

Innerhalb eines Datumsordners kann ein Datengeber seine Inhalte und Ordnerhierarchie völlig frei gestalten. Er muss lediglich hinsichtlich der Dateiformate die Vorgaben bezüglich akzeptierter und präferierter Formate sowie die publizierten Empfehlungen zu den Benennungsregeln beachten (siehe **Kapitel 14.1.1** und **14.1.3**). Lediglich wenn eine Datenlieferung z.B. auf mehreren CDs/DVDs erfolgt, müssen unterhalb der Datumsebene entsprechende Verzeichnisse angelegt werden (z.B. cd1/, cd2/, cd3/ ...), damit die Art des Dateneingangs nachvollziehbar festgehalten werden kann.

Um die langfristige Kuratierung und Archivierung von einzelnen oder mehreren Dateien möglichst effizient und automatisiert durchführen zu können, sollte ein Datengeber nach Möglichkeit ferner dafür sorgen, dass die Zeichenkette aus Dateiname (ohne die Formaterweiterung) und zugehörigem relativen Pfad (d.h. beim SIP ausgehend vom *data*-Ordner) eindeutig ist. Unter dieser Bedingung lässt sich nämlich der zeitliche Verlauf von Dateien über die Lebensdauer eines Digitalen Objektes hinweg einfacher modellieren, da es möglich ist, Nachfolger von Dateien, die z.B. durch Formatkonversionen entstehen, abzubilden. Eine Strukturierung wie

- .../data/1900-01-01/Ausgrabung01/Foto/0001.jpg
- .../data/1900-01-01/Ausgrabung01/Foto/0001.tif

wäre insofern unzulässig, da in beiden Fällen die identifizierende Zeichenkette

- .../data/1900-01-01/Ausgrabung01/Foto/0001

identisch ist. Wenn es gewünscht ist, Dateien mit verschiedenen Formatendungen innerhalb eines SIPs einzuliefern, so muss dies durch Verwendung von Unterordnern explizit kenntlich gemacht werden.

- .../data/1900-01-01/Ausgrabung01/Foto/jpg/0001.jpg
- .../data/1900-01-01/Ausgrabung01/Foto/tif/0001.tif

13.2.3 *metadata*-Verzeichnis

Neben dem *data*-Verzeichnis existiert ein *metadata*-Verzeichnis, in dem verschiedene Dateien abgelegt werden, die in Form von Metadaten, Dokumentation, Korrespondenz u. ä. die zu archivierenden Daten ergänzen. Sein Inhalt wächst im Laufe der Datenverarbeitung, da er durch zusätzliche Informationen der Datengeber und Datenkuratoren oder durch automatisierte Prozesse (z.B. Metadaten-Mappings) angereichert werden kann – sofern die Informationen nicht an anderen Stellen, wie z.B. in der *premis.xml*-Datei, gespeichert werden.

Auf der ersten Ebene werden strukturierte Metadaten in XML-Formaten abgelegt, während in dem Unterordnern *documentation* weitere ergänzende Dokumente zu finden sind. Für ein valides SIP ist lediglich die Existenz zweier Dateien vorgeschrieben:

- *IANUS.xml*, in der alle von IANUS angefragten und von Datengebern zu beantwortenden Metadaten zur Beschreibung einer Datensammlung in einer METS konformen xml-Struktur erfasst sind.²¹⁵ Im Standardfall werden diese Informationen über ein Web-Formular eingegeben, welches Teil des Archiv-Management-Systems von IANUS ist und das automatisiert den Export dieser xml-Datei ermöglicht. Die Datei dient im weiteren Verlauf als Basis für mögliche Mappings und Konvertierungen für unterschiedliche Datenmodelle und Standards. Der Mindestsatz von Metadaten für ein SIP enthält diese Blöcke:
 - zur automatisierten Validierung der Datenübertragung (Integrität, Checksummen, technische Metadaten)
 - zur Beschreibung des SIP-Struktur
 - zur Beschreibung der gesamten Datensammlung und Datenproduzent
 - ggf. zur Beschreibung einzelner Objekte
 - zur Beschreibung aller relevanten Aktionen vor oder während des Ingests
 - zur Beschreibung der Zugangsbeschränkungen
- *Datenebergabevertrag.pdf*, die eingescannte Version des von Datengebern und der für IANUS verantwortlichen Person unterzeichneten Datenübergabevertrages, der als analoges Dokument zusätzlich separat aufbewahrt wird. Die Datei wird in dem Unterordner *documentation* gespeichert.

In dem Unterordner *documentation* können sich noch weitere, vom Datengeber eingelieferte Dokumente befinden, sofern diese zum Verständnis der Datensammlung als Ganzes beitragen und nicht Teil des *data*-Ordners sind. In letzterem, und nicht in *documentation*, sollten z.B. in entsprechenden Unterordnern Listen mit Fotobeschriftungen oder Strukturberichte zu Datenbanken platziert werden.

13.3 Archival Information Package (AIP)

13.3.1 Allgemeine Anforderungen

Im OAIS Funktionsmodell (siehe **Kapitel 5.2**) nimmt das AIP eine zentrale Rolle ein, da es die Verbindung zwischen ursprünglich erstellen Daten und künftig nachnutzbaren Daten herstellt. Während bei einem SIP die ursprüngliche Funktionalität und Struktur der Daten erhalten bleibt, wird IANUS gemäß seiner Archivierungsstrategien (siehe **Kapitel 6.1.1 – 1**) bei einem AIP den Fokus auf die Erhaltung der (Fach-)Inhalt legen. D.h. es sind bei der Erzeugung eines AIPs aus einem SIP auch Migrationen, Exporte, Umstrukturierungen und weitere Maßnahmen erlaubt, durch die Informationen aus ihren ursprünglichen technischen Anwendungskontexten herausgelöst werden können. Diese Maßnahmen sollen dabei so wenig invasiv wie möglich erfolgen und müssen ausführlich dokumentiert werden, um im Sinne der Authentizität die Veränderungen der Daten nachvollziehen zu können. Sind das Aussehen und die Funktion einer Anwendung für das Verständnis von Daten relevant, müssen diese ggf. separat dokumentiert werden (z.B. Form von statischen Screenshots/-casts oder Anleitungen).

²¹⁵ Vgl. das *submission-manifest.txt* in der Übereinkommenvereinbarung des Zuse-Institutes Berlin (siehe Anm. 163)

In IANUS sollen die AIPs mit Hilfe eines Repository-Systems verwaltet und vorgehalten werden. (siehe Kapitel 15.XXX) Damit dieses dem Anspruch einer ‚unbegrenzten Haltbarkeit‘ gerecht wird, müssen die folgenden grundsätzlichen Anforderungen eingehalten werden.

- Bei der Erzeugung eines AIPs aus einem SIP müssen die einzelnen Dateiformate hinsichtlich ihrer Nachnutzbarkeit geprüft und gegebenenfalls durch Datenkonversion in archivwürdige – d.h. in der Regel nicht-proprietäre, offen dokumentierte und weit verbreitete – Formate überführt werden.
- Da die migrierten Dateien in einem Containerformaten zusammengefasst werden, muss im Sinne einer langfristigen Lesbarkeit darauf geachtet werden, dass dieses Containerformat nicht proprietär oder nur für ein einzelnes Repository spezifisch ist. (siehe Kapitel 13.3.2)
- Die einzelnen AIPs müssen auch dann wiederherstellbar und interpretierbar sein, selbst wenn kritische Systemkomponenten nicht mehr funktionieren.
- Bei Veränderungen von Hard- und Software-Komponenten müssen bereits vorhandene AIPs aus einem bestimmten Repository in ein anderes Repository überführt werden können. (vgl. Exit-Strategie Kapitel 11.5)
- Da die AIPs zur Erhöhung der Ausfallsicherheit in mind. zwei Speicher- und Repository-Systemen vorgehalten werden sollen (siehe Kapitel 7.2.1), muss das AIP-Format für die einzelnen Repository-System transparent sein. Nur dann können die Datenbestände einfach und automatisiert zwischen den verteilten Speichern kopiert und synchronisiert werden und es entfällt die Notwendigkeit von mehreren Übernahme-Prozessen.

Aufgrund dieser Prinzipien kann ein ideales AIP als ein digitaler Container definiert werden,

- der in der Lage ist, digitale Inhalte über lange Zeiträume hinweg zu speichern und zu interpretieren (technisch wie inhaltlich),
- der herstellerunabhängig ist,
- der es ermöglicht, mehrere Kopien auf unterschiedlichen Repository-Systemen zu erzeugen
- und der eine Wiederherstellung auch im Falle von Katastrophen gewährleistet.

Im Weiteren ergeben sich zwei weitere Charakteristika eines AIPs:

- Es ist *physikalisch autonom*, d.h. es kann von einem Repository in ein anderes transparent übertragen werden, ohne dass es dabei verändert wird. Für einen verbesserten Zugang in einem neuen Repository darf es vor der Übertragung in ein DIP konvertiert werden, um solche Metadaten zu extrahieren, die ein neues Repository zur Umsetzung seiner Zugangsfunktionalitäten verlangt (dies ist jedoch ein Anliegen des Ziel-Repository, kein Erfordernis an das AIP selbst). Wenn nur ein einzelnes AIP, ohne jede andere Komponente der gesamten Software-Architektur²¹⁶ von IANUS überlebt, muss es immer noch möglich sein, den Inhalt zu extrahieren und es als DIP vorzubereiten, um es mittels dann aktueller Software verarbeiten zu können.
- Es ist *logisch autonom*, d.h. der Inhalt eines einzelnen AIP muss im Ernstfall autark und aus

²¹⁶ Diese kann gleichermaßen kommerzielle Produkte, Open-Source-Lösungen und selbst entwickelte Programme enthalten.

sich selbst heraus verständlich sein, um es in dann aktuelle Informationssysteme inkludieren zu können. Jedes AIP muss daher untrennbar mit seinen Grunddaten verknüpft sein; das schließt alle Metadaten ein, die nötig sind, um die archivierten Dateien fachlich und technisch interpretieren zu können.

13.3.2 Container-Format

Durch die Entscheidung, dass die logische Struktur eines AIP in einem physischen Behälter, der alle Komponenten beinhaltet, gespeichert werden soll (= also eine einzige Container-Datei), und durch die Anforderung, dass dieser Behälter durch weit verbreitete Software-Komponenten und unabhängig von jedem Anbieter verarbeitbar sein soll, wird die Zahl der in Betracht kommenden physischen Archivierungsformate stark reduziert. Dafür werden drei zentrale Kriterien definiert:

- Es muss weit verbreitet, offen und gut dokumentiert sein, so dass Werkzeuge existieren, die das Format für eine sehr lange Zeit unterstützen.
- Es muss eine Integritätsprüfung unterstützen. Diese wird oft verwendet, wenn Daten auf magnetischen Speichermedien gesichert werden, auch wenn die dabei zu erwartende Fehlerzahl durch das Speichern oder das Übertragen heute sehr klein ist. Einige Archivformate enthalten Daten, um solche Fehler zu kennzeichnen, die durch Software erkannt werden, welche zum Lesen der Dateien verwendet wird.
- Es muss Wiederherstellungsinformationen speichern können, d.h. bestimmte redundante Daten in einer Datei vorhalten, die im Falle von Aufbewahrungs- oder Übertragungsfehlern verwendet werden können, um diese mit einer Software zu korrigieren.

Im Vergleich zu anderen Containerformaten (.ace, .rar, .zip) unterstützt das .tar-Format (abgeleitet von tape archive - Bandarchiv) diese Eigenschaften am besten, weshalb es als physisches Containerformat für die Speicherung der AIPs von IANUS zum Einsatz kommen soll. Es handelt sich dabei um ein frühes Unix-Format und wird häufig für die (unkomprimierte) Archivierung verwendet, da es die Speicherung von vielen kleineren Dateien in einer großen Datei unter Beibehaltung verschiedener Dateisysteminformationen erlaubt. Während früher archivierte Dateien nur eine maximale Größe von acht Gigabyte aufweisen durften, gibt es seit 2001 von GNU und POSIX zwei Implementierungen des .tar-Formates, die durch den Einsatz von Basis-256-Kodierung der Dateiinformatoren eine unbegrenzte Dateigröße erlauben und Dateinamen nicht auf eine bestimmte Anzahl von Zeichen eingrenzen.

13.3.3 Grundstruktur

Im Folgenden wird eine Struktur für das AIP von IANUS vorgeschlagen, die möglichst unabhängig von einzelnen Herstellern und spezifischen System ist. Die Praxis in anderen Archiven zeigt, dass diesem Anspruch am ehesten eine einfache Verzeichnisstruktur gerecht wird, d.h. sobald ein physikalischer Container (siehe **Kapitel 13.3.5**) geöffnet wird, können alle Metadaten und Daten-Komponenten getrennt und unabhängig voneinander verarbeitet werden. Selbst wenn eine Komponente – beispielsweise eine veraltete Version eines Metadatenformats aufgrund einer verloren vergangenen Definition – technisch oder inhaltlich nicht mehr interpretiert werden kann, kann der Rest weitergenutzt werden. Es wird daher auf eine XML-Struktur als äußere Schicht des physischen Containers verzichtet, da zunächst diese verstanden und verarbeitet werden muss,

bevor andere Inhalte abgerufen werden können.

Die Struktur des AIP ist in ihrem Gesamtaufbau an die Struktur eines SIP angelehnt, d.h. es enthält ebenfalls die beiden zentralen Verzeichnisse *data* und *metadata* sowie ein BagIt-Manifest und Premis-Dateien. Trotz der Ähnlichkeit und der Tatsache, dass die ursprünglichen Datenobjekte eines validen SIPs als Teil eines neuen AIPs mitarchiviert werden, handelt es sich aber um ein physikalisch neues Informationspaket. Dieses weist in der Summe folgende Eigenschaften auf:

- 1.) *Sicherheit*: Es gibt kein spezifisches Konzept in Bezug auf Sicherheitsmechanismen außer den von der Norm geforderten BagIt-Prüfsummen. Da digitale Signaturen im Kontext der zu erwartenden Forschungsdaten momentan keine Rolle spielen, erscheint dieser reduzierte Aufwand gerechtfertigt. Künftig ist die Frage, wie die Authentizität und Integrität der Daten besser garantiert werden kann, allerdings verstärkt zu adressieren.
- 2.) *Metadaten*: Beschreibende Metadaten werden in einem separaten Unterordner des AIPs gespeichert. Als Metadatenschemata sollen dabei neben einem eigenen IANUS-spezifischen Datenmodell primär die Standards *Qualified Dublin Core*, LIDO, Europeana Data Model, ADEX und Encoded Archival Description zum Einsatz kommen (siehe **Kapitel 12**).
- 3.) *Daten*: In einem AIP werden sowohl die originalen Daten in ihren ursprünglichen Formaten und Ordnerhierarchien als auch die für die Langzeitnutzung kuratierten Daten gespeichert und jeweils in separaten Unterordnern abgelegt.
- 4.) *Identität*: Jedes AIP besitzt seine eigene eindeutige Kennung, die durch den Datenkurator vergeben wird und durch den einmaligen *ContainerName* gekennzeichnet wird.
- 5.) *Beziehungen*: Während die Beziehungen zwischen einzelnen AIPs in den entsprechenden Metadaten beschrieben und durch Referenzen festgehalten werden, gibt es keine technische Unterstützung für die konsequente Erhaltung dieser Angaben. Dies könnte beispielsweise nötig sein, um vor Lösch-Anomalien zu schützen. Auch hier besteht in der Zukunft Optimierungsbedarf.
- 6.) *Historie*: Informationen darüber, was mit dem AIP nach dessen Erstellung passiert, werden in der *premis.xml* festgehalten, da dort die Geschichte der miteinander in Beziehung stehenden AIPs protokolliert wird.
- 7.) *Versionierung*: Nach dem aktuellen Stand der Anforderungen wird es nicht notwendig sein, mehrere Versionen des gleichen AIPs verwalten zu müssen. Sofern der Fall eintritt, dass es zu einem bestehenden AIP ein neueres, ergänzendes Informationspaket gibt, wird dieses als selbständiges AIP archiviert und in dessen Kontext-Metadaten das Verhältnis zu älteren AIPs lediglich beschrieben. Das ältere AIP wird in einem solchen Fall allerdings nicht verändert, etwa um einen Verweis auf ein jüngeres, ergänzendes AIP einzutragen. Diese Information wird lediglich in dem Archiv-Management-System vorgehalten, um sie Datennachutzern in der Präsentationsschicht anzeigen zu können.

13.3.4 *data*-Verzeichnis

In dem *data-Verzeichnis* gibt es zwei bzw. drei Unterverzeichnisse:

data/01_original

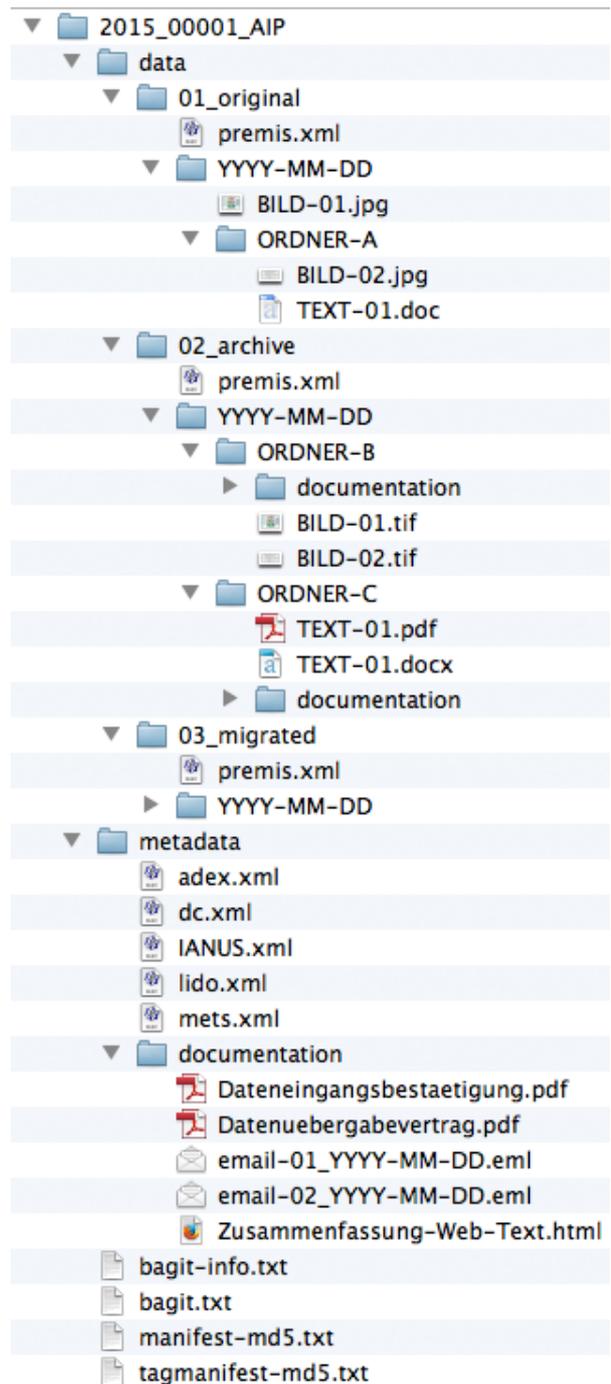
Der Ordner enthält die ursprünglichen Datenobjekte mit unveränderten Dateien und Ordnerhierarchien, wie sie vom Datengeber übergeben und höchstens für die Erzeugung eines validen SIP minimal angepasst wurden.

data/02_archive

In diesem Ordner werden alle Datenobjekte einer Datensammlung gespeichert, die für eine Langzeitarchivierung benötigt werden, d.h. sowohl Dateien, die in einem SIP bereits in einem geeigneten Langzeitformat vorlagen, als auch solche, die aufgrund ungeeigneter Ausgangsformate im Rahmen der Kuratierung aufbereitet werden mussten, z.B. mittels Formatmigrationen oder Datenexporten. Der Umstand, dass es dabei für die erste Gruppe an Dateien zu Dubletten kommen kann (einmal in *01_original* und einmal in *02_archive*), wird bewusst in Kauf genommen, um einen vollständigen und in sich konsistenten Datenbestand innerhalb des Unterordners *02_archive* zu erhalten. Die Gespräche mit anderen fachspezifischen Archiven haben gezeigt, dass bei der Erstellung eines AIPs oftmals die ursprüngliche Ordnerhierarchie für eine bessere Verwaltung und Nachnutzung verändert werden muss. Die Ablagestruktur innerhalb dieses Verzeichnisses kann also von der ursprünglichen Struktur abweichen, so dass mögliche Dubletten in nicht korrespondierenden Ordnern liegen werden.

data/03_migrated

Sofern Dateien aus dem Unterverzeichnis *02_archive* zu einem späteren Zeitpunkt aus in ein neues Langzeit-Format überführt werden müssen, werden diese neu erzeugten Dateien in dem Ordner *03_migrated* abgelegt. Dabei wird für jeden Migrationszyklus ein eigener Ordner mit einer Datumsbezeichnung angelegt, so dass ersichtlich ist, wann Dateien erstellt bzw. obsolet wurden. Die Datenstruktur unterhalb dieser Ebene entspricht der Ablagestruktur unterhalb des Datumsordners von *02_archive*, damit ersichtlich wird, welche jüngere Datei mit welcher älteren korrespondiert. Diese Angaben zu Beziehungen zwischen älteren und neueren



Dateiversionen sowie zu deren Entstehung und Historie werden als automatisch protokollierte und strukturierte Informationen auch in der *premi.xml* festgehalten, damit sie für die Verwaltung der Daten und als Provenienz-Metadaten genutzt werden können.

Im Regelfall liegen in den Datumsordnern von *03_migrated* nur Teilmengen einer archivierten Datensammlung, da Migrationszyklen häufig nur bestimmte Dateiformate betreffen werden und nicht alle. Daher ist es zur Erstellung einer vollständigen, aktuellen Archivversion notwendig, dass alle Datumsordner sowohl in *02_archive* als auch *03_migrated* in chronologischer Folge durchsucht werden und die Dateien entsprechend zusammengestellt werden. In diesem Prozess werden dann ältere Dateiversionen durch jeweils jüngere Dateien – sofern vorhanden – ersetzt.

13.3.5 metadata-Verzeichnis

Wie bei einem validen SIP existiert auch in einem AIP ein eigener Unterordner mit der Bezeichnung *metadata*. Dessen Inhalt speist sich einerseits aus den Informationen und Beschreibungen, die bereits bei der Erstellung eines SIPs vorlagen und bei der Genese eines AIPs aktualisiert wurden oder auch um weitere, neu angelegte Dateien.

- Auf der obersten Ebene befinden sich vor allem xml-formatierte Metadaten, die nach fachspezifischen und allgemeinen Standards, wie DublinCore, Lido oder EDM, strukturiert sind (vgl. **Kapitel 12**) und für unterschiedliche Szenarien des Datenaustauschs genutzt werden sollen. Sie werden nach vorher festgelegten Mappings automatisiert aus der *IANUS.xml*-Datei extrahiert, welche die umfassendsten Metadaten enthält und alle für IANUS relevanten Aspekte und Informationen abdeckt.

In dem Unterordner *documentation* können schließlich solche Dokumente abgelegt werden, die eine Datensammlung als Ganzes betreffen und deren Inhalte nicht bereits in anderer Form in dem Unterordner *data/02_archive* vorhanden sind. Zum Beispiel können bzw. sollen hier abgelegt werden

- die E-Mail-Korrespondenz zwischen Datengeber und Datenkurator, sofern diese relevant ist, etwa weil sie die Daten nochmals erläutern, rechtliche Fragen adressieren oder spezifische Wünsche für die Bereitstellung dokumentieren
- neben dem gescannten Datenübergabevertrag, der bereits als Teil eines validen SIP gefordert wird, eine Dateneingangsbestätigung
- individuelle Erläuterungstexte und Zusammenfassungen, die für die Datenbereitstellung einer Datensammlung über ein Web-Portal mit individualisierten Landing-Pages verwendet werden können

13.4 Dissemination Information Package (DIP)

13.4.1 Allgemeine Anforderungen

Per Definition macht ein OAIS die archivierten und kuratierten Datenbestände für seine vorgesehene Zielgruppe sichtbar und verfügbar. Zu diesem Zweck werden sogenannte Auslieferungspakete (*dissemination information packages* – DIP) definiert, die aus den AIPs generiert und für einen direkten Zugriff zur Verfügung gestellt werden. Sie bestehen grundsätzlich aus den vollständigen Inhaltsinformationen, also Repräsentationen aller archivierten Dateien,

sowie den Erschließungsinformationen als Paketbeschreibungen. Letztere dokumentieren vor allem deskriptive und rechtliche Aspekte und spielen daher für IANUS insbesondere in dem geplanten Nachweiskatalog eine wichtige Rolle, da über diesen als zentralem Online-Findmittel Inhalte gesucht und angefordert werden können (siehe **Kapitel 6.1.1.2**). Beide Informationselemente werden durch die Verpackungsinformationen logisch miteinander verbunden, die auch deshalb notwendig sind, damit ein Datennachnutzer die von ihm abgefragten Informationen klar erkennen kann (siehe **Kapitel 5.1**). Anders als bei der Nutzung klassischer Archive erhält der Datennachnutzer nicht genau das Objekt, das sich tatsächlich in einem Magazin befindet. Vielmehr wird aus dem Archivbestand ein neues Informationspaket erzeugt, das einerseits die Anfrage des Nutzers zielgerichtet beantwortet und dabei andererseits die entsprechenden Rechte zur Nutzung und die Anforderungen an die Langzeitarchivierung berücksichtigt.

Als Teil der Benutzergruppen im Kontext von IANUS ist zunächst an die Datengeber selbst zu denken, welche eine vollständige Kopie ihrer archivalisch aufbereiteten Datensammlung mit aktuellen Dateiformaten und Metadaten in Form eines DIPs erhalten können. In einem solchen Fall wird das DIP weitgehend identisch mit einem AIP sein. Für die allermeisten Datennachnutzer wird ein DIP sich jedoch von dem zugehörigen AIP unterscheiden, da es sich hierbei um eine für die Online-Bereitstellung und Nachnutzung optimierte Instanz eines AIPs handelt. Eine wesentliche Differenz besteht z.B. darin, dass in einem AIP auch die innerhalb eines SIP übertragenen Originaldaten zur dauerhaften und unveränderlichen Aufbewahrung enthalten sind, während diese im Regelfall nicht Teil eines DIPs sind. Ein anderer Unterschied zwischen einem AIP und DIP kann sich daraus ergeben, wenn aus einem AIP mehrere DIPs abgeleitet werden, um für unterschiedliche Zielgruppen z.B. Rastergrafiken mit variierenden Auflösungen bereitzuhalten. Das Verhältnis der bereitgestellten digitalen Objekte als DIPs zu ihren AIPs wird jeweils durch die Vereinbarungen zwischen den Datengebern, IANUS und den Datennachnutzern geregelt (siehe **Kapitel 10**). In ihnen wird vertraglich festgehalten, wie die Präsentationsstrategie für eine individuelle Datensammlung aussehen soll (s.u.).

Prinzipiell hängen die Erstellung und das Format eines DIPs von den Eigenschaften der digitalen Objekte ab, die sich in die folgenden drei Kategorien einteilen lassen:

- rechtliche Bedürfnisse der Datengeber in Bezug auf Sicherheitsstrategien und Zugänglichkeit
- technische Möglichkeiten zur Erhaltung, Darstellung und Übertragung von Dateiformaten
- fachspezifische Anforderungen zur Präsentation und Nachnutzbarkeit

Durch den stetigen Wandel der Technologien werden einerseits Nutzer des Bereitstellungsservices von IANUS immer höhere Ansprüche an das Auffinden und die Nachnutzbarkeit von Daten stellen; andererseits werden über die Zeit Dateiformate, die innerhalb von DIPs verwendet werden, und Präsentationsformen veralten. Die ‚Lebenszeit‘ eines DIPs ist also vergleichsweise kurz und direkt abhängig von technischen Entwicklungen, die außerhalb des Einflussbereiches von IANUS stattfinden. Es wird also eine kontinuierliche Aufgabe sein - zu überprüfen, ob die realisierte Art und Weise der Bereitstellung noch aktuell ist und den Anforderungen der Zielgruppe entspricht. Langfristig wird es notwendig sein, mehr oder minder regelmäßig neue DIPs aus den AIPs zu erzeugen und vorzuhalten. Bei diesem Vorgang muss ein AIP unverändert bleiben und nur als Ausgangsmaterial für die Generierung neuer DIPs dienen.

Um trotz der individuellen Bedürfnisse und der absehbar notwendigen Weiterentwicklungen den Datennachutzern die Archivbestände möglichst homogen präsentieren zu können und um den Verwaltungsaufwand überschaubar zu halten, wird IANUS den Datengebern ein einheitliches Paketbeschreibungsschema für die DIPs und den Zugriff vorschlagen. Des Weiteren wird IANUS die DIPs nach einheitlichen Kriterien (z.B. hinsichtlich der Dateiformate und Metadaten) bereits im Vorhinein erstellen, so dass diese nicht erst spontan bei einer aktuellen Anfrage durch einen Datennachnutzer aus einem AIP erzeugt werden müssen. Ein solches Vorgehen ist auch dadurch begründet, dass die AIPs in externen Rechenzentren zur Bitstream Preservation aufbewahrt werden, ein direkter Zugriff im Falle einer aktuellen Anfrage also nur zeitverzögert stattfinden kann. Stattdessen werden die DIPs am Ende des initialen Kuratierungsprozesses unmittelbar aus den AIPs erzeugt und derart vorbereitet auf lokalen Systemen vorgehalten werden, damit ein unmittelbarer und rascher Zugriff über das Online-Datenportal für die Datennachnutzer möglich ist.

Da der Inhalt und die Organisation jeder Datensammlung individuell sein wird und da die Art der Präsentation eines DIPs von den jeweils aktuellen technischen Möglichkeiten abhängen wird, kann im Rahmen dieses Konzeptes keine allgemeingültige Struktur für DIPs vorweg festgelegt und ausführlich beschrieben werden. Lediglich für ein DIP, das mittels Download bereitgestellt wird (s.u.), werden folgende Bestandteile als minimale Anforderung festgelegt. Danach enthält diese Form des DIPs:

- die eigentlichen angeforderten Inhaltsobjekte als Dateien und/oder Ordnerverzeichnisse
- zugehörige Erschließungsinformationen in Form von strukturierten Metadaten und Dokumentationen
- relevante Lizenzbestimmungen zur Verwendung

Bei einer Download-Anfrage werden diese drei Elemente ad hoc in einem BagIt-Container zusammengeführt und gemeinsam an den Datennachnutzer übertragen.

13.4.2 Präsentation und Access

IANUS wird drei Arten zu Präsentation von Datensammlungen anbieten und neben einem Standardverfahren zur Erzeugung eines DIPs aus einem AIP auch einen flexibel anpassbaren Prozess etablieren, um der Heterogenität altertumswissenschaftlicher Forschungsdaten gerecht zu werden. Die Standard-Variante zur Bereitstellung von DIPs wird eine für jede Datensammlung einzeln adressierbare Webseite sein (siehe **Kapitel 6.1.1.2**), die gleichzeitig auch als notwendige Landing-Page für die persistenten Identifikatoren fungiert (siehe **Kapitel 7.4**) und damit zitierfähig ist. Zentrale Elemente dieser Seite sind:

- die einzelnen Dateien mit den vom Datenproduzenten intendierten semantischen Verzeichnisstrukturen
- Thumbnails oder Vorschauansichten zur schnellen Orientierung
- deskriptive und technische Metadaten zur fachlich-inhaltlichen Beschreibung
- Zitierhinweise, Lizenzangaben und Hinweise zu Möglichkeiten der Nachnutzung
- Angaben und Verweise auf den Dateneigentümer und Datenproduzenten

Die wichtigste Eigenschaft dieser Form der Bereitstellung ist die Möglichkeit, eine oder mehrere Dateien bzw. eine gesamte Datensammlung direkt herunterladen zu können. Im Falle einer

konkreten Downloadanfrage, die im Sinne des OAIS-Modells zugleich die Funktion einer Datenbestellung- und einer Datenauslieferungssitzung übernimmt, werden entsprechend der festgelegten Zugriffsberechtigungen die gewünschten und vorbereiteten digitalen Objekte als DIP in dem oben genannten BagIt-Format ausgeliefert. Sofern technische Einschränkungen vorliegen, z.B. weil die angeforderte Datenmenge ein noch zu definierendes Volumen übersteigt und daher für eine direkte Online-Übertragung ungeeignet ist, werden weitere Möglichkeiten für den Datentransfer angeboten.

Neben dieser für alle Datensammlungen identischen Bereitstellung können darüber hinaus auf expliziten Wunsch von Datengebern und mit deren finanziellen Beteiligung an den erhöhten Kosten auch individuellere Präsentationsformate umgesetzt werden. Denkbar sind z.B. im Falle von Datenbanken angepasste Browsing- und Suchfunktionen, um deren Inhalte schneller anzeigen und effizienter erschließen zu können. Eine solche Umsetzung hängt allerdings immer von den jeweils aktuellen technischen Möglichkeiten ab und muss in einem sinnvollen und wirtschaftlichen Verhältnis zu den fachspezifischen Anforderungen zur Datennachnutzung stehen.

Als dritte Form der Bereitstellung ist die Verbreitung der Metadaten (je nach Art auch der Daten selbst) über Webservices geplant, um einerseits die Auffindbarkeit der Datenbestände über externe Online-Datenportale zu erhöhen, andererseits aber auch die Einbindung in externe Anwendungen zu ermöglichen (siehe **Kapitel 6.1.1.4**).

Auch wenn IANUS aktiv den Open Access-Gedanken unterstützt und möglichst viele Daten offen bereitstellen will, wird es Dateneigentümer geben, welche keinen allgemeinen Zugriff auf alle Metadaten und Daten gestatten können (z. B. aufgrund rechtlicher Einschränkungen). Bei einigen DIPs wird deshalb der Zugriff beschränkt sein und sie werden nur solchen Datennachnutzern angezeigt, die dazu berechtigt sind. IANUS wird also eine Policy für Zugriffsbeschränkungen veröffentlichen und ein technisches Rechtemanagement mit Zugriffskontrolle implementieren, damit die Rechte aller Beteiligten geschützt werden. (siehe **Kapitel 11.3**) Beide Aspekte sind wichtige Voraussetzungen, um notwendiges Vertrauen in der Fachcommunity zu erlangen und die Bereitschaft von Datenproduzenten und Dateneigentümern zu erhöhen IANUS Daten zu übergeben.

14 Arbeitsschritte der digitalen Langzeitarchivierung

Die digitale Langzeitarchivierung von Forschungsdaten wird nicht nur die wichtigste Dienstleistungen von IANUS sein, sondern auch diejenige mit der höchsten Komplexität. Im Folgenden werden daher die Arbeitsschritte beschrieben, die hierfür erforderlich sind, also die Vorbereitung, Kuratierung, Migration, Archivierung und Bereitstellung von digitalen Ressourcen. Die notwendigen technischen Systeme werden in **Kapitel 15** beschrieben und hier nur in ihren jeweiligen funktionalen Kontexten genannt. Bei der Entwicklung des Ablaufes wurden unterschiedliche Ergebnisse und Erfahrungen mitberücksichtigt:

- Die mehrjährigen Erfahrungen und Praktiken, wie sie sich beim Archaeology Data Service in York/UK und dem e-depot for nederlandse archaeologie in Den Haag/NL bewährt haben.
- Die Vorgaben des OAIS-Referenzmodells und die für eine Zertifizierung erforderlichen Kriterien (siehe **Kapitel 11.1**).
- Die Ergebnisse aus zwei Studien, die im Rahmen des E-ARK Projekt (*European Archival Records and Knowledge*) entstanden sind²¹⁷, einen Überblick zur aktuellen Praxis der digitalen Archivierung in 32 Ländern, insbesondere an Nationalarchiven, geben und real angewandte Methoden zu Arbeitsabläufen, genutzten Formaten und Standards, eingesetzten technischen Systemen und erforderlichen Metadaten beschreiben.
- Der methodische Standard PAIMAS (*Producer Archive Interface Methodology Abstract Standard*)²¹⁸, der eng mit dem OAIS-Referenzmodell verwandt ist und vier Phasen der initialen Datenübergabe und -kuratierung unterscheidet: Kontaktaufnahme und Vorbereitung; formale Definition des SIP; der eigentliche Transfer eines SIP; dessen Validierung durch das Archiv.

In den meisten Fällen wird es sich bei IANUS nicht um institutionelle Datenproduzenten handeln, die regelmäßig oder häufig Daten zur Archivierung abgeben werden, sondern um Individuen mit sehr unterschiedlichen digitalen Erfahrungen und Kenntnissen. Es kann daher zur Zeit nicht davon ausgegangen werden, dass trotz veröffentlichter Vorgaben und Hilfestellungen die Daten immer inhaltlich und technisch valide und direkt für die weitere Kuratierung geeignet sein werden. Auch wenn für die Prozesskette von der ersten initialen technischen Kontrolle bis zur endgültigen Speicherung der AIPs eine weitgehend Automatisierung wünschenswert erscheint, zeigt die Praxis bei anderen Archiven mit Forschungsdaten, dass ein vergleichsweise hoher Aufwand an manueller Arbeit erforderlich ist. Nur so ist zu gewährleisten, dass ausschließlich vollständige, wohlgeformte und für die wissenschaftliche Nachnutzung geeignete Datensammlungen in den Archivbestand von IANUS übernommen werden.

Die folgenden Arbeitsschritte beschreiben das Standard-Verfahren für Datensammlungen aus aktuellen oder kürzlich abgeschlossenen Projektdaten (siehe oben **Kapitel 4.3.2**), bei denen ein

²¹⁷ „D 3.1 E-ARK Report on Available Best Practices (2014)“:

<http://www.eark-project.com/resources/project-deliverables/6-d31-e-ark-report-on-available-best-practices>

„An Overview of Digital Preservation Tools (Draft 02-05-2013)“:

<http://www.eark-project.com/resources/project-deliverables/6-d31-e-ark-report-on-available-best-practices/file>

²¹⁸ CCSDS 651,0-M-1, ISO 20652: 2006 = <http://public.ccsds.org/publications/archive/651x0m1.pdf>

Datenproduzent bzw. -geber noch aktiv eingebunden werden kann. Der Ablauf ist dabei in einer idealtypischen Weise beschrieben, für den es in der Praxis nachvollziehbare und zu dokumentierende Gründe geben mag, davon abzuweichen. Für verwaiste Daten (Altdaten) oder überdurchschnittlich große Datensammlungen sind jeweils individuelle Workflows zu definieren. So sollen etwas in den Fällen, wo eine manuelle Kuratierung aufgrund der Menge an Daten kaum durchgeführt werden kann, zur Erhebung fehlender Metadaten entweder verstärkt automatisierte Verfahren zum Einsatz kommen oder gezielte externe Personenkreise (z.B. Studierende in Universitätskursen) zur Unterstützung angefragt werden.

14.1 Vorgaben

IANUS wird entsprechend seiner Archivierungsstrategie (*archiving policy*, siehe **Kapitel 6.1.1**) inhaltliche und technische Vorgaben machen, die Datengeber berücksichtigen müssen, um Ihre digitalen Daten mit dem Ziel der Archivierung bei IANUS abgeben zu können. Diese beziehen sich einmal auf akzeptierte und präferierte Datenformate sowie auf notwendige Metadaten und Dokumentation. Diese Informationen werden über die Homepage von IANUS sowie ein gesondertes Kapitel in den IT-Empfehlungen bereitgestellt²¹⁹, so dass Datengeber ihre Daten vor der Übergabe an IANUS entsprechend vorbereiten können und ggf. Kosten für die Kuratierung reduzieren können. Sofern diese Vorgaben nicht oder zu einem überwiegenden Teil nicht eingehalten werden, können folgende Szenarien eintreten:

- ein Datengeber muss eigenständig, ggf. mit Beratung durch IANUS, die Daten überarbeiten und nachbessern
- eine Annahme der Daten zum Zweck der Archivierung wird von IANUS abgelehnt
- die Daten werden von IANUS angenommen, aber es wird aufgrund eines erhöhten Kuratierungsaufwandes eine höhere Gebühr berechnet.

14.1.1 Akzeptierte und präferierte Datenformate

Da es weder technisch noch inhaltlich sinnvoll ist, alle in den Altertumswissenschaften gebräuchlichen Dateiformate (siehe **Kapitel 4.2.2**) zu verarbeiten und langfristig zu kuratieren, zumal viele von ihnen als Ausgangsformate für die Langzeitarchivierung ungeeignet sind, werden entsprechende Vorgaben formuliert. Diese beschreiben für unterschiedliche Dateiarnten, welche Dateiformate von Datengeber an IANUS zur Langzeitarchivierung übergeben werden können. Dabei wird zwischen akzeptierten Formaten und präferierten Formaten unterschieden. Zur ersten Gruppe gehören solche Daten, die sich durch einen realistischen und angemessenen Aufwand in geeignete Langzeitformate migrieren lassen; zur zweiten Gruppe gehören solche Dateien, die bereits in einem archivierungstauglichen Format vorliegen und insofern einen geringeren Kuratierungsaufwand erforderlich machen.

Die Vorgaben zu den von IANUS angenommenen Dateiformaten beruhen auf zuvor gemachten Festlegungen zu den jeweiligen signifikanten bzw. erhaltenswerten Eigenschaften, aus denen sich auch die formatspezifischen Migrationsstrategien (*migration policies*) für die Archivierung und

²¹⁹ <http://www.ianus-fdz.de/it-empfehlungen/?q=node/44>

Bereitstellung ergeben. Diese Festlegungen bilden die Grundlage für die weitere technische Verarbeitung der eingegangenen Daten innerhalb von IANUS.

14.1.2 Notwendige Metadaten und Beschreibungen

Um digitale Daten künftig nicht nur im technischen Sinne lesen, sondern auch inhaltlich-fachlich verstehen und interpretieren zu können, ist es erforderlich mit jeder Datenlieferung an IANUS auch zugehörige Metadaten zu übergeben. Diese umfassen folgende Aspekte:

- fachliche Beschreibung der Inhalte
- administrativ-rechtliche Angaben
- technische Spezifikationen
- strukturelle Informationen

und beziehen sich teilweise auf mehrere Daten zusammen (Datensammlungen), teilweise auf einzelne Dateien. Ausführlich werden die von Datengebern zu liefernden Metadaten in **Kapitel 11** beschrieben.

14.1.3 Weitere Vorgaben

Weitere Vorgaben beziehen sich auf

- die Benennung von Dateien und Ordnern, z.B. unerlaubte Sonderzeichen, eindeutige Namen
- die Struktur der Datenablage, z.B. klare Trennung von Daten und Metadaten
- die notwendigen Angaben zu einzelnen Dateien oder Gruppen von Dateien in Dateilisten
- die Auswahl von Dateien, z.B. Entfernung oder Kennzeichnung von Personenbezogenen Inhalten, Dateidubletten, temporären Dateien, Zwischenversionen
- Anforderungen, die sich aus dem Datenübergabevertrag ableiten, u.a. das Einverständnis aller Datenproduzenten zur Archivierung und Bereitstellung von Daten

14.2 Vorbereitung der Daten und Datenübertragung

14.2.1 Vorbereitung der Daten

Bevor Dateneigentümer oder von Ihnen beauftragte Personen Daten mit dem Ziel der Langzeitarchivierung an IANUS übergeben, müssen sie sich vorab über die Modalitäten und die Vorgaben informieren und die Daten entsprechend der online bereitgestellten Informationen vorbereiten. Dieser Prozess kann je nach Bedarf durch IANUS beratend begleitet werden. Als vorbereitende Arbeitsschritte zur Verbesserung der Datenqualität durch den Datengeber können danach ggf. nötig werden:

- Auswahl der für die Archivierung vorgesehenen Dateien, z.B. durch Löschung redundanter oder geschützter Daten und Ergänzung um fehlende Dateien (z.B. Publikationen, Handbüchern von verwendeten Systemen, etc.)
- Erstellung oder Vervollständigung der fachlichen, beschreibenden Dokumentation
- Erstellung oder Vervollständigung strukturierter Metadaten
- Vorbereitung der Dateien durch Migration in die akzeptierten bzw. bevorzugten Dateiformate oder durch Export aus Fachanwendungen

- Eindeutige, nachvollziehbare Strukturierung der Dateiablage
- Erstellung von Datei- und Ordnerlisten
- Klärung von Urheberfragen und Nutzungsrechten
- Je nach Umfang und geplanter Art der Datenübertragung verlustfreie Komprimierung von Daten
- Entfernung von technischen Kopierschutzmaßnahmen (Digital-Rights-Management)

Ein Datenproduzent entscheidet eigenständig darüber, zu welchem Zeitpunkt eines Projektes bzw. der Datenerhebung diese Arbeitsschritte durchgeführt werden sollen, doch gibt es die klare Empfehlung, diese Arbeiten – wie auch den ganzen Aspekt der Langzeitarchivierung und Datennachnutzung – so früh wie möglich im Rahmen eines strukturierten Datenmanagementplanes umzusetzen und nicht erst nach Ende der Datenerhebung, Projektabschluss oder Publikation zu beachten.

14.2.2 Anmeldung im Online-System

Spätestens zu dem Zeitpunkt, an dem ein Datengeber seine Daten ausreichend für die Archivierung von IANUS vorbereitet hat und diese zeitnah übermittelt werden können, ist eine Anmeldung an einem Online-Formular von IANUS erforderlich. Dieser Schritt ist auch bereits zu einem wesentlich früheren Zeitpunkt möglich, z.B. bei Beginn eines Projektes, so dass eine frühzeitige Kontaktaufnahme und Unterstützung erfolgen kann. Für IANUS ergibt sich dadurch die Möglichkeit einer besseren Planbarkeit über zu erwartende Datenzuwächse.

Über das Online-Formular kann und soll ein Datengeber nach einem personalisierten Login Angaben zu folgenden Aspekten machen, die sich auf eine gesamte Datensammlung beziehen (ausführlich zu den Metadaten siehe **Kapitel 12.2**):

- deskriptive Erschließungsdaten, insbesondere
 - Kontaktdaten, Angaben zu geistigen Urhebern
 - Beschreibung des (Forschungs-)Projektes bzw. der Aktivität und Methoden der Datenerhebung, so weit wie möglich und fachlich sinnvoll unter Verwendung von Schlagworten und kontrollierten Vokabularen
 - Beschreibung des Umfangs und der Struktur der digitalen Dateien
 - Angaben zur Art der Dokumentation und vorstrukturierten Metadaten
- rechtliche Metadaten, insbesondere
 - Angaben zu Eigentümern von Daten
 - Lizenzierung und Art der Bereitstellung / Veröffentlichung
 - ggf. zeitlich befristete Einschränkungen des Zugriffs und/oder der Darstellung, die sich auf eine Datensammlung insgesamt, einzelne Ressourcen oder einzelne Metadaten beziehen können
- Angaben für die praktische weitere Verarbeitung
 - Angaben, ob die Strukturen der Verzeichnisse erhalten werden müssen
 - Angaben, ob eine Datensammlung ggf. in kleinere Einheiten aufgeteilt werden kann
 - Einwilligungen zur Datenkuratierung, inkl. Datenmigration, ggf. Benennung von Ausnahmen, die unbedingt unverändert bleiben müssen
 - Akzeptanz der allgemeinen Nutzungsbedingungen

Die eingetragenen Metadaten werden in einem zentralen Archiv-Management-System

gespeichert, können zusätzlich aber auch als XML-Datei im METS-Format²²⁰, in der auch initiale Abschnitte in den Standards EAD²²¹ und PREMIS²²² enthalten sind, heruntergeladen werden. Die Datei kann also von einem Datengeber seiner Datensammlung hinzugefügt werden, so dass für diese im Falle einer Datenübergabe bereits standardisierte Metadaten vorliegen. Ferner wird das Online-System eine Standard-Variante eines Datenübergabevertrages zum Download bereitstellen, damit ein Datengeber sich frühzeitig über die vertraglichen Möglichkeiten informieren kann, die zu einem späteren Zeitpunkt verbindlich ausgehandelt werden.

Angedacht ist weiterhin, dass ein Datengeber im Rahmen einer Anmeldung im Online-System die Möglichkeit erhält, eine grobe Kalkulation der Archivierungskosten bei IANUS vorzunehmen. Als Vorbild hinsichtlich Funktionalität und Gestaltung kann hierbei der *costing calculator* des Archaeology Data Service in UK dienen, der als Teil des dortigen Registrierungs- und Upload-Systems "ADS-Easy" integriert ist²²³.

Sobald ein Datengeber die Eingabe der Metadaten in dem Online-System von IANUS als abgeschlossen markiert, wird automatisch ein Ticket mit einer eindeutigen internen Vorgangsnummer ausgelöst. Bei Annahme des Tickets durch einen Datenkurator von IANUS wird eine automatische Nachricht an den Datengeber mit Informationen zum weiteren Vorgehen und zur konkreten Ansprechperson verschickt. Die weitere Bearbeitung des Tickets erfolgt im Regelfall dann, sobald die beschriebenen Daten tatsächlich bei IANUS eingegangen sind.

14.2.3 Datenübertragung

Der Transfer von Dateien eines Datengebers an IANUS kann auf unterschiedliche Weise erfolgen:

- Kleine Datenmengen (bis 100 MB) können mittels Upload-Formular direkt in dem Online-System auf einen Server von IANUS geladen werden. Dies bietet neben der zeitnahen Übertragung der Daten auch den Vorteil, dass eine automatisierte Überprüfung der akzeptierten Dateiformate erfolgen und dadurch schneller eine Rückmeldung gegeben werden kann. Zur Reduzierung der Übertragungszeit und von Übertragungsfehlern müssen die Dateien vor dem Upload in einem Container-Format (mit den Dateierendungen .zip, .tar oder .tgz) verpackt und komprimiert werden.
- Für mittelgroße Datenmengen, deren Umfang immer noch eine Online-Übertragung zulässt, soll ein ftp-Server bzw. ein eigener Cloud-Service zum Einsatz kommen. Auch hier ist eine frühzeitige Prüfung der eingegangenen Dateien möglich.
- Größere Datenmengen, für die eine Online-Übertragung nicht mehr infrage kommt, müssen postalisch über externe Speichermedien (DVD, Festplatte, USB, ...) an IANUS geschickt werden. Sofern die Datenträger nicht an den Datengeber zurückgesandt werden, müssen sie sachgemäß beschriftet und aufbewahrt werden.

²²⁰ <http://www.loc.gov/standards/mets/>

²²¹ <http://www.loc.gov/ead/>

²²² <http://www.loc.gov/standards/premis/>

²²³ <http://archaeologydataservice.ac.uk/easy/costing>

14.2.4 Tabellarische Zusammenfassung

	Arbeitsschritt	Akteur	System	Benötigte Dokumente/ Informationen
Vo-01	Vorbereitung der Daten bzgl. der Auswahl und des Umfangs	Datengeber	lokaler Rechner	publizierte Vorgaben & Hinweise
Vo-02	Vorbereitung der Daten bzgl. der Dateiformate	Datengeber	lokaler Rechner	publizierte Vorgaben & Hinweise
Vo-03	Vorbereitung der Daten bzgl. der Metadaten/Dokumentation	Datengeber	lokaler Rechner	publizierte Vorgaben & Hinweise
Vo-04	Vorbereitung der Daten bzgl. der Datenstruktur, Benennung etc.	Datengeber	lokaler Rechner	publizierte Vorgaben & Hinweise
Vo-05	Erzeugung von Checksummen	Datengeber	lokaler Rechner	Datei mit Checksummen
Vo-06	ggf. Komprimierung zur schnelleren Datenübertragung (tar)	Datengeber	lokaler Rechner	
Dt-01	Online-Registrierung	Datengeber	IANUS.user-Mgmt.	
Dt-02	Beginn des Prozesses „Datenarchivierung“	Datengeber	IANUS.pre-ingest	TP wird initialisiert
Dt-03	Eingabe der initialen Metadaten für eine Datensammlung als Ganzes	Datengeber	IANUS.pre-ingest	XML-Datei im METS-Format
Dt-04	Datenübertragung an IANUS (falls möglich Online-Upload)	Datengeber	IANUS.pre-ingest / FTP-Server / IANUS.Cloud / Postweg	
Dt-05	während des Uploads Virensan	automatisiert	IANUS.pre-ingest	
Dt-06	Vergleich der Checksummen	automatisiert	IANUS.pre-ingest	Datei mit Checksummen
Dt-07	Erfassung technischer Kerndaten der Einzeldateien und Ordnerstruktur	automatisiert	IANUS.pre-ingest	METS-Datei
Dt-08	Auftragsbestätigung des Datengebers zur Datenarchivierung	Datengeber	IANUS.pre-ingest	Information über neue Daten an Datenkuratoren
Dt-09	ggf. Download der initialen Metadaten (deskriptive, technische)	Datengeber	IANUS.pre-ingest	METS-Datei
Dt-10	automatisierte Eingangsbestätigung von IANUS	automatisiert	IANUS.pre-ingest	vorformulierte email

14.3 Verarbeitung TP zu SIP (Pre-Ingest)

Die primären Ziele der Pre-Ingest-Phase sind die erste Sichtung des initialen, ungeprüften Transferpaketes und die Erstellung eines validierten *Submission Information Package (SIP)*, für das die weitere Kuratierung durch IANUS sinnvoll und lohnend ist. Die in **Kapitel 14.3.4** zusammengefassten Arbeitsschritte dienen folglich zur Umwandlung eines Transferpaketes in ein

SIP, welches den rechtsverbindlichen Gegenstand in dem Übergabevertrag zwischen Datengeber und IANUS darstellt.. Der Arbeitsaufwand während des Pre-Ingest soll dabei möglichst effizient gehalten sein und spätere Maßnahmen nicht vorweg nehmen, so dass auch erfahrene Datengeber in der Lage sind selbständig valide SIPs zu erstellen.

Auch wenn die Daten korrekt nach den veröffentlichten Vorgaben von IANUS erstellt worden sein sollten, ist deren automatisierte technische und manuelle inhaltliche Überprüfung notwendig. Diese Kontrolle kann in einen längeren Prozess inkl. intensiver Kommunikation mit dem Datengeber münden, um etwaige Nachbesserungen zu diskutieren, ungeeignete Dateien auszutauschen, fehlende Angaben und Dokumente einzufordern oder Unklarheiten zu beseitigen. Der Pre-Ingest erfordert daher in hohem Maße das Zusammenspiel und die Abstimmung zwischen den Datengebern und den Datenkuratoren und wird durch das System IANUS.PRE-INGEST unterstützt, das einen gemeinsamen, gleichzeitigen Zugriff auf eine Datensammlung erlaubt. Im negativen Fall, d.h. wenn ein Datengeber nicht auf Verbesserungsvorschläge reagiert, keine Rückfragen beantwortet und notwendige Informationen vorenthält, kann es zu einer Ablehnung der angebotenen Datensammlung durch IANUS kommen.

14.3.1 Initiale Validierung TP

Während die Überprüfung der eingehenden Transferpakete auf Schadsoftware und Viren so früh wie möglich passiert, also unmittelbar vor der Übertragung der Daten in die Serverarchitektur und Arbeitsumgebung von IANUS, erfolgen weitere Schritte der technischen und inhaltlichen Qualitätskontrolle erst wenn alle Daten übertragen, die initialen Metadaten vorliegen und der gesamte Vorgang des Datentransfers durch den Datengeber abgeschlossen wurde. Bei den folgenden Schritten geht es vor allem darum zu prüfen, inwieweit eine Datensammlung die Vorgaben zu akzeptierten und präferierten Dateiformaten und zur notwendigen Dokumentation erfüllt oder nicht. Es gilt ferner die inhaltliche Relevanz der Daten für IANUS zu bewerten, deren Vollständigkeit und Strukturierung einzuschätzen, die Lesbarkeit zu testen und mögliche rechtliche Probleme aufzudecken²²⁴.

Sobald eine neue Datensammlung bei IANUS eingegangen ist und einem zuständigen Datenkurator zugewiesen wurde, werden folgende Schritte zur ersten Validierung durchgeführt:

- Verifizierung der angegebenen Kontaktdaten
- Test der Lesbarkeit des Datenträgers und automatisiertes Öffnen aller Dateien, um korrupte und technisch defekte Dateien zu erkennen
- Feststellung, ob eine oder mehrere Dateien einen Passwortschutz besitzen oder andere Zugriffsbeschränkungen aufweisen
- Kurz-Prüfung der fachlich-wissenschaftlichen Relevanz der Daten für IANUS und möglicher rechtlicher Probleme
- Kurz-Prüfung der Dateiformate anhand der Dateierweiterungen im Hinblick auf die IANUS-spezifischen Vorgaben (akzeptierte und präferierte Formate)
- Kurz-Prüfung, ob eine ausreichende Dokumentation für die gesamte Datensammlung

²²⁴ Vgl. zu verschiedenen Umsetzungen des Pre-Ingest-Prozesses die Ergebnisse in dem E-Ark-Deliverable 3.1 „Report on available best practices“ (Anm. 217) S. 47-58.

vorliegt (z.B. Liste der eingelieferten Dateien, Beschreibungen zu einzelnen Dateitypen) und ob die Daten sinnvoll strukturiert und benannt sind

- Kurz-Prüfung, ob die beschreibenden Metadaten inhaltlich verständlich sind (**Dt-03**) und ob die Angaben mit den tatsächlichen Daten übereinstimmen (z.B. hinsichtlich Anzahl, Datenformaten, Aussagen zu Rechteinhabern)

Zum Abschluss der initialen Validierung eines Transferpaketes informiert ein Datenkurator von IANUS den Datengeber über das Ergebnis der Analyse, erstellt eine Prognose über den Aufwand, die Kosten und den zu erwartenden Erfolg der Archivierung und kommuniziert notwendige Verbesserungen.

14.3.2 Verbesserung TP

In welchem Umfang ein Transferpaket während des Pre-Ingest verändert werden muss, hängt wesentlich von der Qualität der übertragenen Daten ab. Generell gilt der Grundsatz, dass die eigentlichen Originaldaten zu diesem Zeitpunkt nur so wenig wie notwendig verändert werden sollen, da alle für die Archivierung notwendigen Migrationen, Neustrukturierungen, u. ä. erst zu einem späteren Zeitpunkt (**Kapitel 14.4**) erfolgen. Dennoch kann aber auch der Austausch größerer Datenmengen erforderlich sein, wenn z.B. eine übertragene Datensammlung mehrere Dateien enthält, die passwortgeschützt sind, in falschen Formaten vorliegen, rechtlich bedenklich sind, keine inhaltliche Relevanz besitzen und unzureichende Metadaten aufweisen.

Da in dieser Phase alle Arbeitsschritte gemeinsam mit bzw. durch den Datengeber erfolgen, da der Inhalt und Umfang einer zu archivierenden Datensammlung noch veränderbar ist und da IANUS noch keine volle rechtlich verbindliche Verantwortung für eine Datensammlung übernommen hat, müssen Veränderungen an dem Transferpaket nicht in der gleichen Weise vollständig archivalisch dokumentiert und protokolliert werden, wie dies für die folgenden Phasen nach Finalisierung eines SIP und Abschluss eines Datenübergabevertrages notwendig ist.

Zur Verbesserung eines Transferpaketes können folgende Maßnahmen notwendig sein, die vor allem durch einen Datengeber durchzuführen sind:

- Ergänzung von fehlenden digitalen Objekten
- Löschung nicht-archivwürdiger digitaler Objekte
- Änderungen von Dateiformaten durch Export und Konvertierung
- Korrekturen an bestehenden oder Ergänzung von fehlenden Metadaten
- Umbenennung und/oder Umstrukturierungen von Verzeichnissen und Dateien, insbesondere die Ersetzung von Leerzeichen und Sonderzeichen, um eine automatisierte Weiterverarbeitung zu vereinfachen

Bei allen Arbeitsschritten liegt es in der Mitwirkungspflicht des Datengebers darauf zu achten, dass eine Datensammlung in sich konsistent bleibt, also die intendierte Funktionalität der eingelieferten Daten und Referenzen zwischen digitalen Objekten erhalten bleiben, z.B. auch dann wenn Dateien umbenannt werden.

Sobald ein Datengeber alle erforderlichen Zuarbeiten geleistet hat und gemeinsam der zu archivierende Inhalt einer Datensammlung final festgelegt wurde, werden die Schritte der Validierung (**Kapitel 14.3.1**) erneut durchgeführt, damit sichergestellt ist, dass alle Dateien alle Prüfungen bestehen und eine Datensammlung in ihrer letzten Version valide ist.

14.3.3 Finalisierung SIP

Abschließend sind Fragen zu Kuratierung und Bereitstellung der Daten zu klären, die zusammen mit einem Datengeber erörtert und entschieden werden müssen. Sie betreffen im Einzelnen folgende Aspekte:

- Festlegungen zu Art und Weise der Veröffentlichung der Datensammlung über das Online-Datenportal, insbesondere Formulierung eines Kurztexes und Auswahl von geeigneten Abbildungen
- ggf. Klärung individueller Wünsche, die über eine Standardpräsentation hinausgehen (siehe **Kapitel 13.4.2**)
- Beratung und Fixierung von Nutzungsrechten, Lizenzbestimmungen und Zugriffsbeschränkungen für die Nachnutzung
- Entwurf und Abschluss des Datenübergabevertrages

Mit diesen letzten Informationen kann der IANUS-Kurator alle verbleibenden Arbeiten zur Finalisierung eines gültigen SIP ausführen, also

- alle Vereinbarungen und Entscheidungen dokumentieren
- relevante analoge Dokumente einscannen und der Datensammlung hinzufügen (z.B. Korrespondenz, Datenübergabevertrag, Lizenzverträge)
- alle Metadaten in IANUS.DATA-MANAGEMENT importieren
- ggf. originale Datenträger beschriften und aufbewahren oder zurücksenden
- Sicherung des validen SIP auf lokalem Backup-Server
- alle Arbeitsschritte in einer PREMIS-Dokument protokollieren
- die Gegenprüfung des gesamten Prozesses durch einen zweiten Datenkurator

Das so erstellte SIP sollte nun alle zu archivierenden Originaldaten und Metadaten plus die für die Archivierung notwendigen Zusatzinformationen enthalten.

14.3.4 Tabellarische Zusammenfassung

	<i>Arbeitsschritt</i>	<i>Akteur</i>	<i>System</i>	<i>Benötigte Dokumente/ Informationen</i>
PI-1	Annahme einer neuen Datensammlung	Datenkurator	IANUS.pre-ingest	
PI-2	Wenn Datensammlung auf externem Datenträger eingesandt wird: - Test Lesbarkeit des Datenträgers - Einspielen in IANUS.pre-ingest (inkl. Vo-05, Dt-05 & Dt-06) - ggf. Einspielen von Metadaten	Datenkurator	IANUS.pre-ingest	METS-Datei (erzeugt in Dt-09)
PI-3	Bestätigung des Dateneingangs; Verifizierung der Kontaktdaten	Datenkurator	IANUS.pre-ingest	
PI-4	Test der Lesbarkeit durch systematisches Öffnen der Dateien	automatisiert / Datenkurator	IANUS.pre-ingest	
PI-5	Feststellung korrupter Dateien durch Vergleich von Checksummen	automatisiert	IANUS.pre-ingest	Datei mit Checksummen (erzeugt in Dt-06)

PI-6	Prüfung, ob Dateien mit Passwortschutz & Zugriffsbeschränkungen existieren	automatisiert	IANUS.pre-ingest	
PI-7	Kurz-Prüfung der Dateiformate anhand von Dateierweiterungen	automatisiert	IANUS.pre-ingest	Dateiliste
PI-8	Suche nach Leer- und Sonderzeichen bei Datei- & Ordnernamen	automatisiert	IANUS.pre-ingest	Verzeichnis- und Dateiliste
PI-9	Kurz-Prüfung der inhaltlichen Relevanz für IANUS und mögliche rechtliche Probleme	Datenkurator		
PI-10	Kurz-Prüfung der Dokumentation / Metadaten	Datenkurator	IANUS.pre-ingest	METS-Datei (erzeugt in DT-03), Dateilisten, weitere Dokumente
PI-11	Rückmeldung an Datengeber, ggf. mit Verbesserungsvorschlägen	Datenkurator	IANUS.pre-ingest	E-Mail mit Validierungsbericht
PI-12	ggf. Löschung oder Austausch von bestehenden Dateien bzw. Zusendung neuer Dateien	Datengeber	IANUS.pre-ingest / FTP-Server / IANUS.Cloud / Postweg	fehlende oder korrigierte Dateien
PI-13	ggf. Bearbeitung der Metadaten	Datengeber	IANUS.pre-ingest	METS-Datei (erzeugt in Dt-09)
PI-14	ggf. Umbenennung von Dateien und Umstrukturierung von Ordnern	Datengeber	IANUS.pre-ingest	
PI-15	ggf. Prüfung der Integrität der geänderten Datensammlung	Datengeber	IANUS.pre-ingest	
PI-16	ggf. Klärung von Lizenz- und Urheberrechtsfragen	Datengeber		
PI-17	finale Festlegung über Umfang & Inhalt der Datensammlung	Datengeber / Datenkurator	IANUS.pre-ingest	
PI-18	Erzeugung der PREMIS-Datei	automatisiert	IANUS.pre-ingest	XML-Datei im PREMIS-Format
PI-19	Festlegung der Details für Bereitstellung im Online-Datenportal, Klärung individueller Wünsche zur Präsentation	Datengeber / Datenkurator	IANUS.pre-ingest	
PI-20	Kurztexte und Abbildungen für Präsentation der Daten erstellen	Datengeber	IANUS.pre-ingest	Zusammenfassung, Bildvorlagen
PI-21	Festlegung von Nutzungsrechten & Zugriffsbeschränkungen	Datengeber	IANUS.pre-ingest	Beschreibung Lizenzmodelle; PREMIS-Datei
PI-22	Entwurf und Abschluss des Datenübergabevertrages	Datengeber / Datenkurator	Postweg	Datenübergabevertrag
PI-23	Einscannen analoger Dokumentation (z.B. Korrespondenz, Lizenz, Vertrag, sonstige Texte) und Ablage im SIP	Datenkurator	IANUS.pre-ingest	analoge Dokumente
PI-24	ggf. originale Datenträger zur Aufbewahrung beschriften oder zurücksenden	Datenkurator		
PI-25	Protokollierung der Arbeitsschritte	Datenkurator	IANUS.pre-ingest	PREMIS-Datei

PI-26	Kontrolle des gesamten Prozesses durch zweiten Datenkurator	Datenkurator	IANUS.pre-ingest	Checklisten
PI-27	Komprimierung des SIPs im BagIt-Format zur schnelleren Datenübertragung	automatisiert	IANUS.pre-ingest	
PI-28	Sicherung des SIPs auf lokalem Backup-Server und Überprüfung des Kopiervorgang	automatisiert	IANUS.pre-ingest	
PI-29	Metadaten in IANUS.data-management importieren	automatisiert	IANUS.pre-ingest/ IANUS.data-Mgmt.	METS-Datei, PREMIS-Datei

14.4 Verarbeitung SIP zu AIP (Ingest)

Während der Ingest-Phase werden aus dem validierten SIP alle digitalen Objekte und Metadaten extrahiert und in eine für die Langzeitarchivierung geeignete Form überführt. Am Ende soll ein *Archival Information Package (AIP)* vorliegen, das den in **Kapitel 13.3** beschriebenen Anforderungen entspricht und das zur Bitstream Preservation in ein externes Rechenzentrum übertragen werden kann. Die erforderlichen Arbeitsschritte werden primär von den IANUS-Datenkuratoren ausgeführt, die durch das System IANUS.INGEST unterstützt werden, das neben der Protokollierung aller Veränderungen an den Daten vor allem auch die automatisierbaren Workflows steuert. Daneben zeigen die Erfahrungen in anderen Archiven mit altertumswissenschaftlichen Forschungsdaten, dass ein großer Teil der inhaltlichen Kontrolle und Aufbereitung manuell erfolgen muss, da die Qualität der Daten ganz wesentlich von fachlichen und nicht so sehr von technischen Parametern abhängt. Ein Teil der Prozesse erfolgt bereits während der Pre-Ingest-Phase; sie werden nun aber erneut mit dem endgültigen, für die Archivierung und Bereitstellung vorgesehenen Datenbestand durchgeführt und entsprechend vollständig dokumentiert. Für den Ingest-Prozess wird davon ausgegangen, dass die Mitwirkung und Zuarbeit eines Datengebers nur noch bei neuen Problemsituationen oder in Zweifelsfällen erforderlich ist.

14.4.1 Kuratierung SIP

Zunächst werden eine aktuelle Dateiliste erstellt und die Verzeichnisstruktur vollständig als strukturelle Metadaten in der METS-Datei abgebildet. Dabei wird erneut für jede Datei mit Hilfe einer Hashfunktion (z.B. md5 oder SHA-3) eine Checksumme erzeugt, um im weiteren Verarbeitungsprozess ungewollte Fehler rasch ermitteln zu können. Außerdem wird jedem digitalen Objekt ein IANUS-interner Universal Unique Identifier (UUID) zugewiesen, damit Informationen ohne eine zentrale Koordination eindeutig aufgefunden und referenziert werden können. Für die interne, aber auch die externe Ansprache der Datensammlung als Ganzes wird eine DOI als persistenter Identifier (siehe **Kapitel 7.4**) reserviert und die hierfür erforderlichen Metadaten aus den bereits vorliegenden Informationen von dem IANUS-Datenkurator extrahiert. Sofern nicht bereits im Pre-Ingest-Prozess geschehen, werden nun Leer- und Sonderzeichen nach festgelegten Regeln durch unproblematische Zeichen ersetzt, wobei der alte, unveränderte Datei- bzw. Ordnername in der PREMIS-Datei protokolliert wird.

Im nächsten Schritt werden alle formatspezifischen technischen Metadaten aus den digitalen Objekten ausgelesen, um auf dieser Basis über die weiteren Arbeitsschritte entscheiden zu können. Hierbei sollen verschiedene Tools und Webservices zum Einsatz kommen, die im Bereich der digitalen Langzeitarchivierung weit verbreitet und etabliert sind:

- File Format Registry PRONOM²²⁵ des UK National Archives (TNA)
- Digital Record Object Identification (DROID)²²⁶
- Format Identification for Digital Objects (FIDO)²²⁷
- SIEGFRIED²²⁸
- JSTOR/Harvard Object Validation Environment (JHOVE)²²⁹

Als Ergebnis der Formatidentifikation werden in der Regel sog. PRONOM Identifier geliefert, die nicht nur Auskunft über das erkannte Dateiformat selbst geben, sondern auch Informationen über die Version eines Dateiformates enthalten.

Anhand dieser technischen Angaben sowie der zuvor mit dem Datengeber vereinbarten zu erhaltenden Eigenschaften legt ein Datenkurator nun die weitere Erhaltungsstrategie für eine Datensammlung fest und entwickelt einen Migrationsplan (siehe **Kapitel 6.1.1.1**). Er wählt also die am besten für die Archivierung geeigneten Zielformate aus und entscheidet über die Frage, welche digitalen Objekte automatisiert nach den einheitlichen Standardregeln zur Formatmigration von IANUS (die *migration policies*) prozessiert werden können und welche Dateien eine davon abweichende – insbesondere eine manuelle – Kuratierung benötigen, z.B. weil es sich um Spezialformate handelt, die von den o.g. Tools nicht erkannt werden oder weil besondere signifikante Eigenschaften zu berücksichtigen sind. Egal welche Strategie angewandt wird, muss das Ergebnis teils automatisiert, teils manuell überprüft werden. Im Falle eines fehlerhaften Ergebnisses kann auf ein originales digitales Objekt, das in einem SIP unverändert enthalten ist, zurückgegriffen werden und es mit angepassten Parametern erneut verarbeitet werden.

Neben diesen formatbezogenen Aufgaben, die stärker die technischen Aspekte einer Datensammlung betreffen, ist während des Ingests die detaillierte Überprüfung aller inhaltlichen Informationen notwendig, damit eine fachlich sinnvolle Nachnutzung gewährleistet ist. Diese Qualitätssicherung muss überwiegend manuell durch einen Datenkurator geleistet werden, der in einem vernünftigen Maße die vorliegenden Metadaten und zugehörige Dokumente sowohl bezogen auf einzelne Dateien als auch auf die gesamte Datensammlung redigiert, ergänzt und standardisiert. Neben den fachspezifischen Erschließungsinformationen sind auch die übrigen Metadaten (siehe **Kapitel 12.2**) wie administrative und rechtliche sowie Angaben zur Provenienz eingehend zu prüfen und ggf. zu vervollständigen. Weiterhin sind Beziehungen und Referenzen zu anderen Ressourcen – sowohl innerhalb des IANUS-Archivbestandes als auch zu externen Quellen – zu prüfen oder neu anzulegen und in den entsprechenden Abschnitten der METS-Datei

²²⁵ <http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx>

²²⁶ <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

²²⁷ <http://openpreservation.org/technology/products/fido/>

²²⁸ <http://openpreservation.org/blog/2014/09/27/siegfried-pronom-based-file-format-identification-tool/>

²²⁹ <http://jhove.sourceforge.net/>

festzuhalten.

Je nach der Strukturierung der digitalen Objekte innerhalb einer Datensammlung kann es zudem notwendig sein, die Verzeichnisstruktur abzuändern und/oder Dateien umzubenennen, wobei von Fall zu Fall entschieden werden muss, ob eine solche Anpassung eher zur Erhöhung der Archivierbarkeit oder eher zur Verbesserung der Nachnutzbarkeit dient, diese Arbeiten also bereits während des Ingest-Prozesses vorgenommen werden sollten oder erst später als Vorbereitung für die Bereitstellung sinnvoll sind.

14.4.2 Finalisierung AIP

Die anschließenden Arbeitsschritte zur Finalisierung eines AIPs dienen vor allem der Qualitätssicherung des gesamten Kuratierungsprozesses und des Ergebnisses. Daher ist das neu entstandene AIP auf seine Konsistenz und seine Authentizität hin zu überprüfen:

- Ist das Informationspaket in sich schlüssig und sind die zugehörigen Metadaten frei von Widersprüchen?
- Wurden alle Veränderungen nachvollziehbar dokumentiert, so dass sie zu einem späteren Zeitpunkt wiederholt bzw. rückgängig gemacht werden können?
- Sind die Urheber und der Zeitpunkt von Ergänzungen und Neuerungen lückenlos protokolliert?

Zur Beantwortung dieser Fragen dient das System IANUS.INGEST, in dem kontinuierlich die Arbeitsschritte dokumentiert werden sollen und das ggf. mit noch fehlenden Informationen zu vervollständigen ist. Zur Qualitätssicherung dient auch, dass ein 2. Datenkurator unabhängig von dem eigentlichen Bearbeiter einer Datensammlung das SIP als Ausgangsdatenbestand, das Ergebnis der Kuratierung als AIP und die Dokumentation der Arbeitsschritte überprüft und die Korrektheit des gesamten Vorganges bestätigt.

Die letzten Aufgaben ähneln der Finalisierung eines SIP, beinhalten also die Komprimierung des AIPs in ein BagIt-Format zu schnelleren Datenübertragung, die Übertragung des AIPs auf lokale Backup-Server aber auch in externe Rechenzentren zur *bitstream preservation* (siehe **Kapitel 7.2.1**), die Überprüfung der beiden Kopiervorgänge und der Import der finalen Metadaten in das System IANUS.DATA-MANAGEMENT.

14.4.3 Tabellarische Zusammenfassung

	<i>Arbeitsschritt</i>	<i>Akteur</i>	<i>System</i>	<i>Benötigte Dokumente/ Informationen</i>
I-1	Dateiliste erstellen und Verzeichnisstruktur abbilden	automatisiert	IANUS.ingest	Update METS-Datei um strukturelle + technische Metadaten
I-2	Prüfsummen für jede Datei erzeugen	automatisiert	IANUS.ingest	Update METS-Datei
I-3	UUIDs für jede Datei erzeugen	automatisiert	IANUS.ingest	Update METS-Datei
I-4	DOI reservieren und Metadaten erzeugen	Datenkurator	IANUS.ingest	Update METS-Datei
I-5	ggf. Ersetzen von Leer- & Sonderzeichen	automatisiert	IANUS.ingest	Update PREMIS-Datei
I-6	Formatidentifikation zur Erzeugung der	automatisiert	IANUS.ingest	Update METS-Datei

	technischen Metadaten			
I-7	Detail-Prüfung der Metadaten für gesamte Datensammlung	Datenkurator	IANUS.ingest	
I-8	Detail-Prüfung der Metadaten für einzelne Dateitypen	Datenkurator	IANUS.ingest	
I-9	Detail-Prüfung sowie weitere notwendige Dokumentation	Datenkurator	IANUS.ingest	
I-10	Festlegung signifikanter Eigenschaften, Erhaltungsstrategie und Migrationsplan	Datenkurator	IANUS.ingest	
I-11	ggf. Verbesserung der Verzeichnisstruktur für Archivierung & Nachnutzung	Datenkurator	IANUS.ingest	Protokollierung in PREMIS-Datei & METS-Datei
I-12	ggf. Verbesserung der Dateinamen für Archivierung & Nachnutzung	Datenkurator	IANUS.ingest	Protokollierung in PREMIS-Datei & METS-Datei
I-13	Formatkonvertierung der Dateien nach festgelegtem Migrationsplan	automatisiert	IANUS.ingest	Protokollierung in PREMIS-Datei
I-14	ggf. manuelle Kuratierung einzelner Dateien für die Archivierung	Datenkurator	IANUS.ingest	Protokollierung in PREMIS-Datei
I-15	Kontrolle der signifikanten Eigenschaften nach der Konvertierung	automatisiert / Datenkurator	IANUS.ingest	
I-16	Erzeugung der finalen Checksummen für das AIP	automatisiert	IANUS.ingest	Update METS-Datei
I-17	Beziehungen und Referenzen zu anderen Ressourcen prüfen und anlegen	Datenkurator	IANUS.ingest	Update METS-Datei
I-18	Protokollierung der Arbeitsschritte	Datenkurator	IANUS.ingest	Update PREMIS-Datei
I-19	Kontrolle des gesamten Prozesses durch zweiten Datenkurator	Datenkurator	IANUS.ingest	Checklisten
I-20	Komprimierung des AIPs im BagIt-Format zur schnelleren Datenübertragung	automatisiert	IANUS.ingest	
I-21	Sicherung des AIPs auf lokalem Backup-Server und Überprüfung des Kopiervorgangs	automatisiert	IANUS.ingest	
I-22	Sicherung des AIPs zur Bitstream Preservation in externem Rechenzentrum und Überprüfung des Kopiervorgangs	automatisiert	IANUS.ingest	
I-23	Metadaten in IANUS.data-management importieren	automatisiert	IANUS.ingest/ IANUS.data-mgmt.	METS-Datei

14.5 Verarbeitung AIP zu DIP (Access)

Im Folgenden werden nur die Arbeitsschritte für die Standard-Bereitstellung eines DIPs erläutert, also die Vorbereitungen, die notwendig sind, um eine archivierte Datensammlung innerhalb des Online-Datenportals von IANUS präsentieren und zum direkten Download anbieten zu können

(siehe **Kapitel 13.4.2**). Darüber hinausgehende Formate, welche die individuellen Wünsche von Dateneigentümer berücksichtigen oder auf neue technische Möglichkeiten reagieren, sind jeweils im Einzelfall zu beurteilen und können nicht pauschal als festgelegte Workflows vordefiniert werden.

Ganz generell umfasst die Erstellung eines DIPs aus einem AIP vor allem drei Aspekte:

- die Auswahl, welche Datenformate ohne technische Hürden für eine Nachnutzung durch Dritte geeignet sind
- die Überprüfung, inwieweit die vorhandenen beschreibenden Metadaten für Datennutzer verständlich und zielführend sind
- die Umsetzung von Zugriffregelungen und Beschränkungen

Für die allermeisten Entscheidungen, die für die Access-Phase relevant sind, sollten die benötigten Informationen bereits vorliegen, da sie in den vorherigen Arbeitsprozessen erstellt und mit dem Datengebern abgestimmt wurden.

14.5.1 Erstellung DIP

Aus dem AIP wird unter Verwendung des Systems IANUS.ACCESS ein DIP gebildet, indem auf der Basis von Formatidentifikationen, die während des Ingest-Prozesses erfolgen, die generellen Migrationsstrategien (*migration policies*) von IANUS und ggf. individuelle Festlegungen passender Zielformate definiert werden. Während der vergleichbare Prozess bei der Erstellung eines AIPs potentiell auch manuelle Eingriffe erforderlich macht, sollte die Formatmigration von bereits kuratierten digitalen Objekten eines AIPs weitgehend automatisiert erfolgen können bzw. gar nicht erst notwendig sein, wenn eine Archivdatei bereits direkt nachnutzbar ist. Anschließend ist das resultierende DIP in seiner Gesamtheit als auch die signifikanten Eigenschaften einzelner Objekte zu überprüfen. Danach werden für alle Dateien neue Prüfsummen erzeugt, damit Nachnutzer, die diese Daten per Download herunterladen möchten, auch kontrollieren können, ob der Kopiervorgang korrekt und ohne Veränderung des Bitstreams erfolgt ist. Anschließend wird ein DIP auf den IANUS-eigenen Web-Server, der als Backend für das Online-Datenportal dient kopiert, und auch hier der Kopiervorgang mittels Checksummen verifiziert.

Da die Veröffentlichung einer Datensammlung in dem Online-Datenportal nicht nur von Dateiformaten abhängig ist, sondern auch von der inhaltlichen und visuellen Präsentation, sind weitere Arbeitsschritte in diesem Bereich notwendig. Zunächst muss innerhalb des Systems IANUS.ACCESS eine individuelle Webpage erstellt werden, die zwei Funktionen übernimmt. Zum einen ist es diejenige Stelle, an der sich ein interessierter Nachnutzer schnell und präzise über den Inhalt und die fachliche Relevanz einer Datensammlung informieren kann. Die hierfür benötigten Texte und Abbildungen werden im Regelfall von dem Datengeber bereits bei Übergabe einer Datensammlung erstellt und von einem Datenkurator nur noch redaktionell überarbeitet. Zum anderen dient die Webpage in technischer Hinsicht als Landingpage, die als Referenz-URL bei der Registrierung der DOIs zu der jeweiligen Datensammlung angegeben werden muss.

Weiterhin müssen in den Systemen zur Nutzer- und Datenverwaltung von IANUS mit Hilfe des Rechte-Rollen-Konzeptes (siehe **Kapitel 11.3**) die vom Dateneigentümer gewünschten und vereinbarten Zugriffsregelungen, insbesondere mögliche Einschränkungen, umgesetzt werden. Die Nutzungsbedingungen mit den Lizenzangaben, der Hinweis, wie eine Datensammlung zu zitieren ist, und Referenzen auf andere Ressourcen können zwar automatisiert aus den

gesammelten Metadaten abgeleitet werden, ihre korrekte Anzeige und Funktionsweise innerhalb des Online-Datenportal muss aber manuell überprüft werden.

Sobald alle Bestandteile für eine Datenveröffentlichung vorliegen und technisch korrekt in allen IANUS-Systemen umgesetzt wurden, wird der Datengeber informiert, damit dieser die Bereitstellung der von ihm eingelieferten Datensammlung in all ihren Facetten testen und entweder die Freigabe zur Datenpublikation erteilen oder Nachbesserungen formulieren kann. Sobald auch die möglichen letzten Fragen geklärt und Fehler gelöst sind, erfolgt die Freischaltung auf dem öffentlich zugänglichen Produktiv-Server für das Online-Datenportal, so dass alle berechtigten Datennachnutzer eine Datensammlung suchen, finden und nutzen können. Wie bei den anderen Prozessen werden auch bei der Access-Phase die einzelnen Arbeitsschritte dokumentiert bzw. die automatisch protokollierten Angaben vervollständigt und die Metadaten zu einem DIP in IANUS.DATA-MANAGEMENT eingelesen. Abschließend erfolgt – je nach den Regelungen des Datenübergabevertrages – die verwaltungstechnische Abwicklung, d.h. vor allem die Abrechnung der vereinbarten Gebühren und beginnt die Öffentlichkeitsarbeit (siehe **Kapitel 6.3**), um auf den neuen Datenbestand von IANUS hinzuweisen.

14.5.2 Tabellarische Zusammenfassung

	<i>Arbeitsschritt</i>	<i>Akteur</i>	<i>System</i>	<i>Benötigte Dokumente/ Informationen</i>
A-1	Definition der Zielformate für DIP gemäß Formatidentifikation, Migrationsstrategie und individueller Festlegungen	Datenkurator	IANUS.access	
A-2	Konvertierung der Dateien nach den vorherigen Entscheidungen	automatisiert	IANUS.access	Update PREMIS-Datei
A-3	Kontrolle des DIPs nach Konvertierung	automatisiert / Datenkurator	IANUS.access	
A-4	Erzeugung von Prüfsummen für neu erstellte Dateien	automatisiert	IANUS.access	
A-5	Übertragung des DIPs auf Server für Online-Datenportal und Überprüfung des Kopiervorgangs	automatisiert	IANUS.access	
A-6	Sammlung der Informationen für individuelle Webpage (Landingpage)	Datenkurator	IANUS.access	beschreibende Texte und Abbildungen
A-7	Registrierung der DOI	Datenkurator	IANUS.access	METS-Datei
A-8	Umsetzung der Zugriffsrechte nach Festlegungen durch Datengeber	Datenkurator	IANUS.user-mgmt / IANUS.data-mgmt	METS-Datei
A-9	Information an Datengeber zur Prüfung der Präsentation	Datenkurator	Online-Datenportal	E-Mail
A-10	ggf. Nachbesserungen an Präsentation	Datenkurator / Datengeber	IANUS.access	
A-11	Freigabe zur Datenpublikation	Datengeber		Update METS-Datei
A-12	Freischaltung auf Produktivserver	Datenkurator	Online-Datenportal	Update METS-Datei

A-13	Protokollierung der Arbeitsschritte	Datenkurator	IANUS.access	Update PREMIS-Datei
A-14	Metadaten in IANUS.data-management importieren	automatisiert	IANUS.ingest/ IANUS.data-mgmt.	METS-Datei
A-15	verwaltungstechnische Abwicklung (Abrechnung der Gebühren)	Administration		
A-16	Öffentlichkeitsarbeit & SocialMedia-Kanäle	PR		

15 Glossar - Definition relevanter Begriffe

akzeptierte Dateiformate	Diejenigen Dateiformate, in denen Datenproduzenten ihre Dateien an IANUS zur Archivierung übergeben können und die sich durch einen realistischen und angemessenen Aufwand in geeignete Langzeitformate migrieren lassen.
Altdaten	Als Alt-Daten werden Forschungsdaten aus bereits abgeschlossenen Forschungsprojekten bezeichnet, für welche keine Projektmittel mehr zur Verfügung stehen, um den Kuratierungs- und Archivierungsaufwand zu finanzieren, und für welche keine Ansprechpersonen greifbar sind, die Auskunft über Inhalt und Struktur der Daten geben können..
Altertumswissenschaften	Bezeichnung für alle Fachdisziplinen, die sich weltweit mit der Erforschung antiker Kulturen beschäftigen. Innerhalb dieses Konzepts wird der Terminus als Sammelbegriff für alle in Kapitel 4.1 aufgeführten Fachrichtungen verstanden, auch wenn diese in anderen Kontexten nicht zu den Altertumswissenschaften gezählt werden.
Archivierung	To do
BagIt-Format	To do
Bitstream Preservation / Erhaltung	Verfahren und Strategien, die den physischen Erhalt eines gespeicherten Datenobjekts (Bitstreams) auf einem entsprechenden Speichermedium, sicherstellen, z. B. durch redundante Kopien und regelmäßiges Umkopieren von Dateien auf neuere Speichermedien.
Data Seal of Approval	To do
Datengeber	Ein Datengeber ist die Rolle der Person, welche zu archivierende Forschungsdaten bei IANUS abgibt und mit den Datenkuratoren über diese kommuniziert. Ein Datengeber kann der Dateneigentümer und/oder der Datenproduzent der anvertrauten Daten sein, muss es aber nicht. So kann ein Dateneigentümer eine dritte Person dazu beauftragen, sich an seiner Stelle um die Archivierung und Bereitstellung von Daten bei IANUS zu kümmern.
Datenproduzent	Ein Datenproduzent ist diejenige Person oder das System, die bzw. das als Urheber neue Forschungsdaten erzeugt. Dabei kann diese Person auch die Eigentums- und Nachnutzungsrechte an den Daten besitzen, muss es aber nicht zwangsläufig, z.B. wenn diese Rechte aufgrund eines Vertragsverhältnisses auf einen Auftraggeber oder einen Arbeitgeber übertragen werden.
Dateneigentümer	Ein Dateneigentümer ist die Rolle einer Person oder Institution, die die Eigentumsrechte an Forschungsdaten besitzt. Dabei muss diese Person oder Institution nicht zwangsläufig diese Daten auch selbst erhoben und bearbeitet haben, also die Rolle des Datenproduzenten innehaben, kann es aber.
Datennachnutzer	Nach OAIS-de (2012) S. 9: „Die Rolle, die von Personen oder Client-Systemen eingenommen wird, die mit Diensten des OAIS interagieren, um erhaltene Information zu finden und im Einzelnen auf diese Information zugreifen zu können. Das kann andere OAIS sowie OAIS-interne Personen oder Systeme einschließen.“
Datenkurator	To do
Datenlieferung	To do
Datensammlung	Summe aus allen Dateien und zugehörigen Metainformationen, die inhaltlich, rechtlich oder logisch zusammengehören.

Forschungsdaten	To do
institutionelle Partner	Institutionen, die ihre Daten mit IANUS archivieren und hierzu verbindliche, zeitlich befristete Dienstleistungsverträge abschließen; aufgrund der finanziellen Beteiligung und mittelfristigen Bindung an IANUS erhalten diese ein einfaches Stimmrecht im Aufsichtsrat
Kuratierung	To do
METS	<p>Metadata Encoding and Transmission Standard, XML-Format zur Beschreibung von digitalen Sammlungen von Objekten mit Metadaten</p> <ol style="list-style-type: none"> 1. METS Header <i>metsHdr</i>: Metadaten, die das METS-Dokument beschreiben. 2. Descriptive Metadata <i>dmdSec</i>: Die Erschließungsdaten des Objekts. 3. Administrative Metadata <i>amdSec</i>: Verwaltungsangaben wie etwa Herkunft des Objekts, Urheberrechte, Informationen zur Speicherung des Objekts. 4. File Section <i>fileSec</i>: Alle Dateien, die zu dem Objekt gehören. 5. Structural Map <i>structMap</i>: Beschreibung der internen Struktur des Objekts. 6. Structural Links <i>structLink</i>: Beinhaltet Verknüpfungen von Elementen, etwa für die Archivierung von Webseiten. 7. Behavioral <i>behaviorSec</i>: Dieser Abschnitt kann verwendet werden, um ausführbare Elemente (etwa Webservices) mit dem METS-Paket zu verbinden
OAIS	Ein Offenes Archiv-Informationssystem (OAIS) ist ein Archiv, das aus einer Organisation, die Teil einer größeren Organisation sein kann, aus Menschen und Systemen besteht, die die Verantwortung übernommen hat, Information zu erhalten und sie einer vorgesehenen Zielgruppe zugänglich zu machen. Alle hierfür relevanten Aspekte (Aufbau, Arbeitsweise, Struktur, Verantwortlichkeiten etc.) sind in dem OAIS-Referenzmodell beschrieben, das 2003 als ISO-Standard 14721:2003 zertifiziert wurde.
präferierte Dateiformate	Diejenigen Dateiformate, die IANUS bevorzugt von Datengebern zur Archivierung annimmt, da es sich bereits um archivierungstaugliche Formate handelt und insofern der Kuratierungsaufwand im Vergleich zu anderen Formaten geringer ausfällt.
PREMIS	To do
Projekt (DIN 69901)	Vorhaben, das im Wesentlichen durch Einmaligkeit der Bedingungen in ihrer Gesamtheit gekennzeichnet ist, z. B. Zielvorgabe, zeitliche, finanzielle, personelle und andere Begrenzungen, Abgrenzung gegenüber anderen Vorhaben, projektspezifischer Organisation.
Projekt-ID	Fortlaufende eindeutige Nummer eines Projektarchives, d.h. einer inhaltlich oder organisatorisch abgegrenzten Sammlung von Daten; ein Projektarchiv kann dabei Teil eines anderen (größeren) Projektarchives sein.
Provenance-Metadaten	To do
Prüfsumme (Checksum)	To do
Qualified Dublin Core	To do
Sammlung	<p>Bei ADS Überbegriff für folgende Arten von Datenarchiven</p> <ul style="list-style-type: none"> • Zeitschriften & Serien • Grey Literature • Projektarchive

	<ul style="list-style-type: none"> • Bibliographien • Doktorarbeiten
Speicherung	To do
TRAC	Trustworthy Repositories Audit and Certification
Version-ID	Eine Versionsnummer für ein Projektarchiv, diese wird fortgeschrieben, wenn sich maßgebliche inhaltliche Änderungen/Korrekturen an einem Archiv ergeben; sie ändert sich nicht, wenn neue Datenzugänge erfolgen und bestehende unverändert bleiben, und wenn Dateiformate im AIP und DIP technisch migriert werden,
Vorgang-ID	Fortlaufende eindeutige Nummer, die einen Vorgang ab dem ersten Kontakt/Anfrage identifiziert, unabhängig davon, ob es zu einer Archivierung von Daten kommt oder nicht,
Zugang-ID	Fortlaufende eindeutige Nummer, die die Übergabe von Projektdaten beschreibt, die kann z. B. bei fortlaufenden Projekten regelmäßige Ergänzungen von Daten sein.
Zielgruppe/Stakeholder	To do

16 Literaturverzeichnis

ADS IMPACT REPORT (2013)

Neil Beagrie, The Value and Impact of the Archaeology Data Service. A study and methods for enhancing sustainability (Salisbury 2013).

http://repository.jisc.ac.uk/5509/1/ADSReport_final.pdf

FORSCHUNGSDATENMANAGEMENT (2011)

Stephan Büttner, Hans-Christoph Hobohm, Lars Müller (Hrsg.), Handbuch Forschungsdatenmanagement (Bad Honnef 2011).

<http://www.forschungsdatenmanagement.de/>

urn:nbn:de:kobv:525-opus-2412

<http://opus.kobv.de/fhpotsdam/volltexte/2011/241/pdf/HandbuchForschungsdatenmanagement.pdf>

IANUS – HOEHREN (2014)

Thomas Hoeren, Expertise - Rechtliche Aspekte bei der Langzeitarchivierung digitaler Daten. 1. Teil: Die Rechte aus dem archäologischen Input [Version 1.0], Hrsg. IANUS (Berlin 2014)

<http://dx.doi.org/10.13149/000.0nc98i-f>

IANUS – KLIMPEL/WEITZMANN (2014)

Paul Klimpel, John H. Weitzmann, Expertise – Rechtliche Aspekte bei der Langzeitarchivierung digitaler Daten. 2. Teil: Lizenzierung von Materialien und Datenbanken durch das Forschungsdatenzentrum IANUS [Version 1.0], Hrsg. IANUS (Berlin 2014)

<http://dx.doi.org/10.13149/000.3h7mtr-d>

IANUS – KOLBMANN (2014)

Wibke Kolbmann, A Digital Registry for Archaeological Find Spots and Excavation Documentation in IANUS, Masterarbeit HU Berlin, Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft 359 (Berlin 2014).

<http://nbn-resolving.de/urn:nbn:de:kobv:11-100214958>

IANUS - SCHNEIDER (2013)

Gerhard Schneider, Expertise – Aufbau eines Forschungsdatenzentrums Archäologie und Altertumswissenschaften [Version 1.0], Hrsg. IANUS (Berlin 2013)

<http://dx.doi.org/10.13149/000.ekvdia-n>

IANUS - SCHWIEGELSHOHN (2013)

Uwe Schwiegelshohn, Expertise – Aufbau eines Forschungsdatenzentrums Archäologie und Altertumswissenschaften [Version 1.0], Hrsg. IANUS (Berlin 2013)

<http://dx.doi.org/10.13149/000.llwqv2-2>

IANUS - STAKEHOLDERANALYSE (2014)

Maurice Heinrich – Felix Schäfer – Sabine Jahn, Stakeholderanalyse 2013 zu Forschungsdaten in den Altertumswissenschaften. Teil 1: Fragebogen und Ergebnisse [Version 1.0], Hrsg. IANUS (Berlin 2014).

<http://dx.doi.org/10.13149/000.jah37w-g>

IANUS – TESTBED PID (2014)

Martina Trognitz, Abschlussbericht Testbed "Persistent Identifiers" [Version 1.0], Hrsg. IANUS (Berlin 2014).

<http://www.ianus-fdz.de/attachments/download/560/Testbed-Persistent-Identifiers.pdf>

IANUS - ÜBERSICHT DATENZENTREN & FACHSYSTEME (2013)

Fragenkatalog "Vor-Ort-Begehungen" & Berichte zu Datenzentren
(Online-Stellung nach Genehmigung der Datenzentren geplant)

KRDS (2010)

Ch. Beagrie et al., Keeping Research Data Safe: Cost/benefit studies, tools, and methodologies focussing on long-lived data (London 2010).

<http://www.beagrie.com/krds.php>

NESTOR. ENZYKLOPÄDIE (2010)

Heike Neuroth et al. (Hrsg.), nestor Handbuch – Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.3 (Göttingen 2010).

urn:nbn:de:0008-2010071949

NESTOR. BESTANDSAUFNAHME (2012)

Heike Neuroth et al. (Hrsg.), Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme, (Göttingen 2012).

urn:nbn:de:0008-2012031401

OAIS-DE (2012)

Referenzmodell für ein Offenes Archiv-Informationssystem - Deutsche Übersetzung, hrsg. von nestor - Kompetenznetzwerk Langzeitarchivierung, nestor-materialien 16 (Frankfurt a.M. 2012)

urn:nbn:de:0008-2012051101

RADIESCHEN. SYMPOSIUM FDI (2013)

Bericht zum Symposium „Forschungsdaten-Infrastruktur“ am 22.01.2013 am GFZ in Potsdam, durchgeführt von den Projekten Radieschen, re3data.org, KomFor, EWIG und BoKeLa.

http://dx.doi.org/10.2312/RADIESCHEN_002

THALLER (2013)

M. Thaller (Hrsg.), Das Digitale Archiv NRW in der Praxis. Eine Softwarelösung zur digitalen Langzeitarchivierung (Hamburg 2013).

WISSENSCHAFTSRAT (2359-12)

Wissenschaftsrat, Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020, Drs. 2359-12 (Berlin 2011).

<http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>

WISSENSCHAFTSRAT (10465-11)

Wissenschaftsrat, Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften, Drs. 10465-11 (Berlin 2011)

<http://www.wissenschaftsrat.de/download/archiv/10465-11.pdf>

WISSENSCHAFTSRAT (10466-11)

Wissenschaftsrat, Übergreifende Empfehlungen zu Informationsinfrastrukturen (Berlin 2011)

<http://www.wissenschaftsrat.de/download/archiv/10466-11.pdf>

WISSENSCHAFTSRAT (9035-09)

Wissenschaftsrat, Stellungnahme zum Akademienprogramm (Saarbrücken 2009)

<http://www.wissenschaftsrat.de/download/archiv/9035-09.pdf>

WISSENSCHAFTSRAT (4907-15)

Wissenschaftsrat, Stellungnahme zum Deutschen Archäologischen Institut (Bielefeld 2015)

<http://www.wissenschaftsrat.de/download/archiv/4907-15.pdf>