



PROTOKOLL 3. TREFFEN DER ARBEITSGRUPPE INTEROPERABILITÄT

02. Juli 2014, 10.00 – 15.30 Uhr,

DAI Berlin, Wiegandhaus

TEILNEHMER

Dr. Reiner Göldner, LA für Archäologie, Sachsen, reiner.goeldner@lfa.sachsen.de
Prof. Dr. Günther Görz, Universität Erlangen-Nürnberg, goerz@informatik.uni-erlangen.de
Guido Heinz, RGZM Mainz, heinz@rgzm.de
Dr. Leif Isaksen, University of Southampton, leifuss@googlemail.com
Wibke Kolbmann, DAI Berlin, wk@dainst.de
Jens Ludwig, SUB Göttingen, ludwig@sub.uni-goettingen.de
Dominik Lukas, TOPOI, Freie Universität Berlin, dominik.lukas@topoi.org
Dr. Allard Mees, RGZM Mainz, mees@rgzm.de
Rainer Simon, AIT Austrian Institute of Technology, rainer.simon@ait.ac.at
Sabine Thänert, DAI Berlin Zentrale, sthaenert@dainst.de

DURCHFÜHRUNG

Dr. Felix Schäfer, DAI Berlin Zentrale
Maurice Heinrich, DAI Berlin Zentrale
Martina Trognitz, DAI Berlin Zentrale

VERHINDERT

Hans-Georg Becker, TU Dortmund, hans-georg.becker@udo.edu
Tom Brughmans, University of Southampton, tom.brughmans@yahoo.com
Prof. Dr. Kai-Christian Bruhn, FH Mainz, i3mainz bruhn@geoinform.fh-mainz.de
Prof. Dr. Gregory Crane, Tufts University, gregory.crane@tufts.edu
Frank Förster, Universität Kiel, ffoerster@gshdl.uni-kiel.de
Prof. Dr. Reinhard Förtsch, DAI Berlin Zentrale, CoDArchLab Uni Köln, rf@dainst.de
Prof. Dr. Monika Hagedorn-Saupe, Staatliche Museen zu Berlin, m.hagedorn@smb.spk-berlin.de
Irmela Herzog, LVR Bonn, i.herzog@lvr.de
Prof. Dr. Eric Kansa, University of California/Berkeley, ekansa@ischool.berkeley.edu
Rainer Komp, DAI Berlin, rainer.komp@dainst.de
Robert Kummer, Universität zu Köln, rokummer@gmail.com
Undine Lieberwirth, TOPOI, Freie Universität Berlin, undine.lieberwirth@fu-berlin.de
Dr. Michael Merkel, Helms-Museum Hamburg, michael.merkel@helmsmuseum.de
Kathrin Mertens, Helms-Museum Hamburg, kathrin.mertens@helmsmuseum.de
Matteo Romanello, DAI Berlin, mro@dainst.de
Prof. Dr. Frank Schwarzbach, Professor für Geoinformatik, schwarzbach@htw-dresden.de
Regine Stein, Universität Marburg, r.stein@fotomarburg.de
Arne Weiser, HTW Berlin, arne.weiser@htw-berlin.de

Nr.	Beschreibung	Akteure	To-Dos
1	BEGRÜßUNG UND FORMALIEN		
	Begrüßung der Teilnehmer und Formalien	Schäfer	

2	AKTUELLE ENTWICKLUNGEN IM IANUS-PROJEKT		
	Präsentation der Entwicklungen im IANUS-Projekt seit dem vergangenen Treffen. (siehe Präsentation)	Schäfer	
	<p>Diskussion:</p> <p>F: Anzahl an beantragten Stellen für die 2. Phase von IANUS A: <i>Gesamt 6-7 Vollzeitäquivalente</i></p> <p>F: Wie soll IANUS zum Projektende aussehen? Wie wird die Finanzierung aussehen? Welche Rechtsform wird IANUS haben? A: <i>Rechtsform: Empfehlung der Rechtsexpertise ist gGmbH Finanzierung: Basisfinanzierung (ideal wäre Beteiligung durch Bund), Gebühren der Datengeber, Drittmittelprojekte Für die Nutzer der Daten sollen keine Kosten entstehen. Rolle DAJ: wichtiger Partner (Mitgesellschafter) eines Verbundes, könnte Räume, Verwaltung und IT-Infrastruktur zur Verfügung stellen</i></p> <p>F: Wie sieht die Rückmeldung aus der Community aus? Sind bestimmte Anforderungen, Bedürfnisse oder Wünsche bekannt? A: <i>Hinweis auf Umfrage 2013 (Stakeholderanalyse) und die Ergebnisse daraus, die zeigen, dass größter Bedarf bei der Datenarchivierung besteht. Anforderungen auch variieren von der Art der Institution. Außerdem gibt es eine Reihe an Unterstützungsschreiben für das Projekt und es sind bereits Testdaten eingegangen, die zeigen, dass verschiedene Akteure bereit sind, Daten zur Archivierung abzugeben.</i></p>	<p>Mees Schäfer</p> <p>Isaksen Schäfer</p> <p>Ludwig Schäfer</p>	

3	VORSTELLUNG MASTERARBEIT „DIGITALES VERZEICHNIS FÜR FUNDSTELLEN UND GRABUNGSDOKUMENTATION“		
	Konzept für einen Nachweiskatalog, Schwerpunkt: Metadaten und Metadatenschemata. (Siehe Präsentation und Masterarbeit)	Kolbmann	
	<p>Diskussion:</p> <p>Aktuell fehlt ein Bundesländer übergreifender Fundstellenkatalog für ganz Deutschland. Auch der internationale Datenaustausch über Fundstellen und laufende Forschungsprojekte wäre wünschenswert. In IANUS geplant ist ein Katalog, der zwischen Fundstellen, Aktivitäten und Datenslg./Archiven unterscheidet.</p> <p>Lizenzen und Lizenzmodelle</p> <ul style="list-style-type: none"> • Tendenzen von bestimmten Institutionen bestimmte Lizenzierungsmodelle zu verwenden: in Bibliotheken überwiegend CC-BY-NC. Forscher verwenden meist überhaupt keine Lizenz und müssen überzeugt werden, sich für eine Lizenz zu entscheiden. Anwälte empfehlen nicht zu restriktive Lizenzen zu verwenden, da die Lizenzgeber auch tatsächlich gewillt sein müssen Verstöße zu ahnden (notwendige Zeit, Kosten, Ressourcen) • Nichtkommerzielle Lizenzen sind nicht immer einfach zu benutzen und können selbst für gemeinnützige Einrichtungen problematisch werden. 	<p>Schäfer</p> <p>Mees / Kolbmann</p> <p>Göldner</p>	

<ul style="list-style-type: none"> • Klarstellung der verschiedenen Lizenzmodelle, z.B. in den IT-Empfehlungen wäre hilfreich für die Nutzer. Trotz länderspezifischer Aspekte von Lizenzen sind diese auch für eine internationale Nutzung von Daten von Bedeutung. <p>Open Data, Zugriffsrechte</p> <ul style="list-style-type: none"> • Notwendig ist breite Diskussion, wie mit „open data“ umgegangen werden soll. Aktuell bestehen verschiedene Institutionen auf ihren eigenen, „privaten“ Daten. Es muss verdeutlicht werden, worin die Vorteile von „open data“ liegen. • Seit 2003 existiert die „Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities“, deren Umsetzung bei den unterzeichnenden Institutionen noch stockt. • Zumindest die Informationen über die Daten (Metadaten) sollten offen zugänglich sein. Einschränkungen könnten Fundortangaben und vom Datenschutz betroffene Informationen sein. Mittels Zugangsberechtigungen könnten Einschränkungen auf Informationen wieder aufgehoben werden. • Abhängig von Bundesländern gibt es unterschiedliche Vorgaben zum Umgang mit den Daten und deren Verfügbarkeit. Außerdem muss deutlich werden, wann über Metadaten und wann über die Daten selbst gesprochen wird. • Ein Blick auf Fallbeispiele wäre interessant, wie etwa auf Europeana: gibt es Tendenzen, die Daten zunächst sehr restriktiv zur Verfügung zu stellen und diese später zu lockern? • In Europeana müssen Metadaten mit CC-0 zur Verfügung gestellt werden, was bei manchen Institutionen jedoch zu einem Konflikt führt, da verwendete Vokabulare andere Lizenzen besitzen. Außerdem verdienen Museen an hochauflösenden Bildern, weshalb diese nicht an Europeana weitergegeben werden. • Wie offen mit den Daten umgegangen wird hängt auch von der Konkurrenz in dem jeweiligen Fachgebiet ab. Eine Strategie zum Umgang wären Embargos für eine gewisse Zeitspanne, die danach aufgehoben werden. • Embargos und andere Zugriffsbeschränkungen sollen erlaubt werden, wobei Workflow/Policy zu definieren sind, um aktiv und regelmäßig die Notwendigkeit für Restriktionen bei Datengebern zu überprüfen. • Strategie für Anfangszeit von IANUS: wichtig ist Menge an offenen Daten, um Nutzern den Mehrwert von Archivierung und Bereitstellung zu verdeutlichen; umgekehrt sind differenzierte Zugriffsrechte wichtig, da „Schutz der eigenen Daten“ zentral für Datenproduzenten ist, um überhaupt Daten zu übergeben • Es sollte aktiv versucht werden, größere Datenbestände zu erhalten, um kritische Masse zu erlangen und Relevanz aufzuzeigen <p>Suche und Präsentation von Daten</p> <ul style="list-style-type: none"> • In welcher Art und Weise werden Daten gesucht und präsentiert? • Zunächst grob das Thema und sobald etwas Spezifisches gefunden wird so viel wie möglich zu dem einen Aspekt und dessen Quelle. • Für Nachweiskatalog sind vor allem Metadaten für eine ganze Datensammlung wichtig, Metadaten für die einzelnen Ressourcen sind zunächst nicht relevant • Hinweis auf zwei Vorträge bei der CAA in Paris, wie Forscher nach Daten suchen und wie sie sie verwenden. • IANUS sollte eine API anbieten, um anderen Systemen die Einbettung von Widgets zu erlauben, damit Daten von IANUS eingebunden werden können. --> Suchen und Finden von Daten nicht nur über IANUS-Homepage, sondern auch dezentral über institutionelle Webseiten 	<p>Isaksen</p> <p>Göldner</p> <p>Görz</p> <p>Kolbmann</p> <p>Schäfer</p> <p>Isaksen</p> <p>Kolbmann</p> <p>Ludwig / Kolbmann</p> <p>Schäfer</p> <p>Kolbmann / Ludwig</p> <p>Mees</p> <p>Kolbmann Isaksen</p> <p>Schäfer</p> <p>Göldner</p> <p>Isaksen</p>	<p>Aufnahme von Lizenzen in IT-Empfehlungen (2.3.3)</p> <p>Policy für Zugriffsbeschränkungen formulieren</p> <p>Definition und Entwicklung von APIs</p>
---	---	---

4	KONZEPT & UMSETZUNG VON PERSISTENTEN IDENTIFIKATOREN		
	Vorstellung der Ergebnisse aus dem Testbed zu persistenten Identifikatoren. (siehe Präsentation und Bericht)	Trognitz	
	<p>Diskussion:</p> <p>F: Wird Versionierung berücksichtigt? Spiegelt sich Versionierung in dem Suffix des DOI wider? A: <i>Jede neue Version einer Ressource benötigt einen neuen DOI. Es gibt jedoch Ausnahmen, wie bsp. der DOI für die IT-Empfehlungen, wo der DOI auf die aktuellste Fassung verweist, während ältere archivierte Versionen einen eigenen Identifikator bekommen. In den Metadaten kann zusätzlich auf ältere Versionen hingewiesen werden. In der Syntax des Suffixes sind keine Versionsangaben vorgesehen.</i></p> <p>F: Was passiert, wenn es das DOI-System nicht mehr geben sollte? IANUS sollte dies in einer evtl. Exit-Strategie berücksichtigen. A: <i>Da das System auf Handles beruht, welche offen dokumentiert sind, kann es technisch nachgebaut werden. Die DOI-Foundation hat zudem einige Vorkehrungen getroffen im Ernstfall vorbereitet zu sein: http://www.doi.org/doi_handbook/6_Policies.html#6.5 Wichtig ist, dass die Zuordnung zwischen DOI und Ressource erhalten bleibt und daher ist es sinnvoll eine Kopie bei IANUS zu behalten.</i></p> <p>In den Zitierhinweisen sollte auch eine URL für die Auflösung von DOIs angegeben werden. Problem: URL kann sich ändern. Besser: nur DOI angegeben und Hinweis auf die Auflösung getrennt davon, in digitalen Ressourcen diesen Hinweis maschinenlesbar speichern. Es gibt bereits den Namensraum info:doi als URI und es besteht die Möglichkeit, dass Browser in Zukunft (auch ohne Plugins) mit diesem Schema, wie mit URLs, umgehen können.</p> <p>Problem der Granularität in der Praxis: Wenn beispielsweise URLs für Einzeldateien verwendet werden, wie lässt sich dies mit DOIs der Datensammlungen verknüpfen? Wenn eine Einzeldatei einen DOI bekommen soll, muss Vollständigkeit der Metadaten überprüft und sie ggf. ergänzt werden.</p> <p>Auflösungsdienst führt nicht dazu, dass Links ungültig werden – dafür ist Datenbereitsteller verantwortlich. IANUS sollte transparente Richtlinien veröffentlichen, die aufzeigen, wie die DOIs gewartet und welche Prozeduren verwendet werden, um z.B. die DOIs auf Gültigkeit zu prüfen.</p> <p>F: Wie verhält sich die Menschen lesbare Landing Page zu der maschinenlesbaren XML-Datei? A: <i>Landing Page sollte alle Informationen angeben, die auch in der XML-Datei hinterlegt sind. Die XML-Datei kann die Grundlage für die Erzeugung der Landing Page bilden.</i></p> <p>Vergabe des Suffixes: Wird von IANUS selbst vorgegeben und mit einem Programm erzeugt. Dieser Schritt muss in dem späteren Produktivsystem integriert werden. Wenn bei Registrierung kein eigenes DOI-Suffix angegeben wird, wird dieses von da ra automatisch erzeugt.</p> <p>F: Überprüft da ra die Gültigkeit einer angegebenen URL? A: <i>Nach SLA mit da ra, ist es die Verpflichtung bei IANUS die Gültigkeit der URL sicherzustellen.</i> A: Humboldt-Uni überprüft DOIs automatisiert alle paar Tage.</p>	<p>Lukas Trognitz</p> <p>Isaksen Trognitz</p> <p>Isaksen Trognitz</p> <p>Isaksen Trognitz</p> <p>Isaksen Trognitz</p> <p>Isaksen Trognitz</p> <p>Schäfer</p> <p>Heinz</p> <p>Lukas Trognitz</p> <p>Heinz / Trognitz</p> <p>Heinz Trognitz / Schäfer Lukas</p>	<p>Berücksichtigung von DOIs in der Exitstrategie</p> <p>Hinweise für Auflösung von DOIs an Nutzer reichen</p> <p>„DOI-on-demand“ in Konzept beschr.</p> <p>Entwicklung einer Policy zur „Pflege“ von DOIs</p> <p>Testen wie da ra sich bei ungültigen URLs verhält.</p>

	Anregung, Informationen zur Landing Page auch maschinenlesbar darzustellen, insbesondere wenn von DOI ausgegangen wird. Alle Metadaten liegen bei IANUS, ein Teil wird an da ra übermittelt, andere sollen mittels Mappings und APIs für andere Dienste zur Verfügung gestellt werden.	Heinz Schäfer	
--	--	---------------	--

5	ENTWURF ZUR VERWENDUNG VON METADATEN UND STANDARDS IN IANUS		
	Theoretischer Hintergrund zu Metadaten und Präsentation eines Konzeptes für IANUS. (siehe Präsentation)	Schäfer	
	<p>Diskussion:</p> <p>Für Nutzergruppen und Zugriffsberechtigungen werden Informationen über Akteure benötigt, um Legitimierungen zu erhalten. IANUS benötigt internes System zur Verwaltung aller Akteure (inkl. der Online-Nutzer); welche Daten davon auch für externe Nutzer bereitgestellt werden, ist noch zu diskutieren.</p> <p>Es gibt Systeme für Akteure, wie etwa GND oder ORCID. Problem: Aktualität der dortigen Informationen, da Update nur die jeweilige Person selbst tun kann. Aktuelle Personen-/Kontakt-Daten nur für interne IANUS-Verwaltung notwendig. Für Archivierung gelten die Angaben, die zum Zeitpunkt der Übergabe aktuell sind/waren.</p> <p>Wenn Mapping auf CARARE vorliegt, gibt es Dienste, die Mapping auf weitere Metadatenschemata ermöglichen. Es reicht daher aus, wenn nur Mappings auf ausgewählte Metadatenschemata angeboten/umgesetzt werden.</p> <p>Wichtig ist, dass nicht nur DAI-Systeme (ZENON, Gazetteer, Wörterbuch, ...) eingebunden werden, sondern auch andere Systeme (z.B. Pleiades, Geonames) Vorteil eigener Systeme ist, dass Datenqualität, Granularität, Funktionalität beeinflusst werden kann. Wichtig ist, dass Links zwischen Systemen eingepflegt werden.</p> <p>Für Fundorte sind neben Ortsnamen auch Koordinaten von Bedeutung, da allein die Namen ambig sein können oder sich im Laufe der Zeit ändern können. Manche Vokabulare sind sehr speziell und die Bedeutung von bestimmten Konzepten hängt auch von dem Kontext (z.B. Ort) ab. Projekte, die sich mit Raum-Zeit-Konzepten auseinandersetzen: ChronOntology und ProjektO</p> <p>Versuch, den DAI-Thesaurus und das Vokabular von Propylaeum auf DDC zu mappen, DDC ist allerdings gröber. Es wurde auch versucht, die verschiedenen DAI-Thesauri untereinander zu mappen.</p> <p>Es müsste zunächst ein Kern an Metadaten definiert werden, wofür zusätzlich ein allgemeines Vokabular konzipiert werden müsste. Eine Verschränkung von oder Mapping zwischen mehreren Thesauri wäre wünschenswert und notwendig (ähnlich wie bei Metadatenschemata).</p> <p>Metadaten müssen auch Information zur physikalischen Beschaffenheit und Qualität der Daten beinhalten; es muss z.B. , sichergestellt sein, dass eine CSV-Datei in der Nachnutzung genauso viele Spalten hat, wie ursprünglich gewollt.</p>	<p>Kolbmann</p> <p>Schäfer</p> <p>Kolbmann</p> <p>Schäfer</p> <p>Kolbmann</p> <p>Mees</p> <p>Simon</p> <p>Göldner</p> <p>Göldner / Mees Isaksen</p> <p>Thänert</p> <p>Göldner</p> <p>Isaksen / Schäfer /</p> <p>Mees</p>	<p>Entwicklung System zur Admin. der Akteure</p> <p>Berücksichtigung ext. Systeme im Konzept</p> <p>Auswahl Daten-Schemata und Mappings festlegen</p>

	<p>Auswahl der vorgestellten Metadatenstandards prinzipiell sinnvoll</p> <p>Einzeldateien in Datensammlungen sollten Metdaten von übergeordneter Sammlung erben können, um Aufwand klein zu halten.</p> <p>Spezifische Ablage-/Ordnerstruktur von Grabungsdokumentation sollte für Metadaten genutzt werden (vgl. ArchaeoDoks); Angaben eignen sich auch für Präsentation und Verwendung der Daten</p> <p>eDNA präsentiert Dateien zum Download innerhalb einer Verzeichnisstruktur</p>	<p>Isaksen</p> <p>Schäfer</p> <p>Göldner</p> <p>Schäfer</p>	<p>Metadatenfeld für relativen Speicherort</p>
	<p>Vorstellung der Tabelle mit Metadaten (siehe Tabelle)</p> <p>Es können sehr schnell sehr viele Metadaten werden, weshalb auf die Benutzbarkeit geachtet werden sollte.</p> <p>Die Tabelle stellt zunächst Idealform (Maximale Ausprägung) dar, die auch in IT-Empfehlungen aufgenommen werden soll, um die Nutzer zu sensibilisieren. Dort muss auch herausgestellt werden, warum bestimmte Metadaten wichtig sind und angegeben werden sollten.</p> <p>Praxis von IANUS, d.h. für Archivierung und Nachnutzung vorgegebene Metadaten, muss noch entwickelt werden und kann davon abweichen.</p> <p>→ Menge der Metadaten so weit wie möglich reduzieren und gut begründen. Von Nutzern möglichst kleine Menge an Informationen verpflichtend verlangen, möglichst viel optional, da Nutzer nicht alle Metadaten ausfüllen werden.</p> <p>Auch wenn Technische Metadaten „fest“ zu einer Datei gehören und auch zukünftig ausgelesen werden können, sind sie als explizit gespeicherte Information für Verwaltung und Migration von Datenformaten wichtig.</p> <p>Für einige Dateitypen, z.B. Bilder oder 3D-Daten, besitzen Metadaten zur Provenienz, Entstehungs- und Verarbeitungsgeschichte hohe Relevanz, vor allem wenn sowohl Rohdaten als auch daraus abgeleitete Daten bereitgestellt werden</p> <p>Primäres Ziel von Metadaten ist die Unterstützung des Nutzers bei der Nachnutzung von Dateien; insofern muss entschieden werden, welche Angaben für einen Nutzer hilfreich sind und welche nicht.</p> <p>Auf irgendeine Weise muss Qualitätsmanagement betrieben werden, um qualitative Daten vorzuhalten.</p> <p>Bei der Migration von Daten ist es wichtig zu wissen, welche Änderungen an den Daten erlaubt sind, ohne die Intention des Autors zu verfälschen. Dies müsste ebenfalls in den Metadaten adressiert werden.</p> <p>Zu METS: ist eher ein „packaging format“ (wrapper). Es wird wahrscheinlich andere Standards geben, die relevanter für IANUS sein könnten, z.B. OAI-PMH für Datenaustausch</p>	<p>Schäfer</p> <p>Heinz</p> <p>Schäfer / Isaksen</p> <p>Schäfer</p> <p>Mees / Schäfer</p> <p>Heinz / Görz</p> <p>Schäfer</p> <p>Heinz</p> <p>Göldner</p> <p>Ludwig</p>	<p>Optionale und notwendige Metadaten festlegen</p> <p>Relevanz von METS für IANUS überprüfen</p>
<p>6</p>	<p>ZUSAMMENFASSUNG UND ABSCHLUSS</p>		
	<p>Keine weiteren Anmerkungen.</p>		