

# Archaeology Data Service York/UK



Begehung am 09.-11.05.2012  
teilweise aktualisiert Juli 2013 und Nov. 2014

Aufgaben	
Hauptaufgaben:	Archivierung von Forschungsdaten, ausschließliche Archäologie mit Bezug zu UK (geographisch, finanziell oder institutionell)
weitere Aufgaben:	Beratung, Forschung und Entwicklung im Rahmen von Drittmittelprojekten
nicht Teil der Aufgaben:	Bereitstellung von Forschungswerkzeugen, wissenschaftliche Auswertung von Daten
relevante Fachrichtung:	Archäologie

Allgemeine Angaben	
Rechtsform:	ist bei der Universität von York angegliedert
Gründung/Geschichte:	1996
Anzahl Mitarbeiter:	<p>15 (Stand Juni 2013) 14 FTE, 2 FTE/2</p> <p>Anzahl FTE (Planstellen) ca. 6 (finanziert durch Einnahmen/Gebühren)</p> <p>Anzahl Hilfskraftstellen (Planstellen) 0</p> <p>Anzahl FTE (befristet, Drittmittel) ca. 8 (finanziert durch Projektmittel)</p> <p>Anzahl Hilfskraftstellen (befristet, Drittmittel) 0</p> <p>Anzahl WissMA 5</p> <p>Anzahl IT 1 System-Admin 3 Softwareentwickler</p>
Zertifizierung:	Data Seal of Approval 2014-2015, Winner Digital Preservation Awards 2012, Data Citation Index
Referenzmodell:	OAIS (SIP, AIP, DIP)

Datenbestand	
Verpflichtung Daten zu sichern?	Ja. Alle in UK durchgeführten Grabungen, müssen Daten archivieren. AHRC- und NERC finanzierte Projekte müssen dies beim ADS machen. Projekte von English Heritage, British Academy, Council for British Archaeology und Society of Antiquaries of London wird empfohlen, dies beim ADS zu tun.
Findet gezielte Datenakquise statt?	Nein
Anzahl an Projektarchiven:	383 Collections (Stand Mai 2012) 423 Collections (Stand Juli 2013) 515 Collections (Stand Nov. 2014)
Anzahl an „grauer Literatur“:	16.021 (Stand Mai 2012) 20.058 (Stand Juli 2013) 28,857 (Stand Nov. 2014) (ca. 320 verschiedene Datenprovider)

Anzahl an Doktorarbeiten:	19 (Stand Nov. 2014)
Anzahl an retrodigitalisierten Zeitschriften:	19 (Stand Nov. 2014)
Anzahl an Einträgen Nachweiskatalog:	1.132.000 (Stand Mai 2012) 1.326.357 (Stand Nov. 2014) (aus 25 verschiedenen Ausgangskatalogen)
Gesamtgröße Archiv:	13.000 GByte (13 TB)
Datensatzgröße:	50% der Daten 100-200 MB
Datenzuwachs jährlich:	letzten vier Jahre: 30% - 50%
Altdaten vorhanden:	Ja, aber nur vereinzelt. Wenn, dann oft von Institutionen/Firmen, die diese aufbereiten und dokumentieren können. Alte Formate werden so weit wie möglich in neue migriert.

## Datengenerierung

technische Unterstützung bei Datenerhebung (z. B. durch Tools, Software, etc.):	Nein
fachliche Unterstützung bei Datenerhebung:	Ja

## Datenupload

Online-Upload möglich?	Datenübertragung bislang nur offline. Ab Sommer 2013 über neues Web-Tool („ADS-easy“) für kleine bis mittlere Projektarchive (bis ca. 300 Dateien) möglich.
Auswahlkriterien für Aufnahme von Daten:	Angebote Daten werden evaluiert hinsichtlich - fachlich relevanter Inhalt und potentielle Nutzbarkeit - Aufwand der Aufbereitung, Kuratierung und Veröffentlichung - Prüfung, ob geeignete, alternative Archiven existieren. Bei Erfüllung der Kriterien werden Daten akzeptiert.  personenbezogene, vertrauliche oder sensible Daten nur nach Anonymisierung oder expliziter Erlaubnis zur Archivierung. In Ausnahmen ist ein zeitlich befristetes Embargo zur Nicht-Veröffentlichung möglich.
Werden angebotene Daten abgelehnt? (Gründe)	- personenbezogener, sensibler Inhalt - Dateien zu groß für Online-Download - Wunsch des Datengebers
Vertragsverhältnis Datengeber - Datenarchiv:	- personenbezogener, sensibler Inhalt - Dateien zu groß für Online-Download - Wunsch des Datengebers
Software/Collection Registry:	Selbst entwickeltes „Collection Management System“, Oracle DB als Backend, Coldfusion für Frontend

## Kuratierung von Daten

Arbeitsschritte zur Datenarchivierung:	Ja. Alle in UK durchgeführten Grabungen, müssen Daten archivieren. AHRC- und NERC-finanzierte Projekte müssen dies beim ADS machen. Projekte von English Heritage, British Academy, Council for British Archaeology und Society of Antiquaries of London wird empfohlen, dies bei ADS zu tun.
Werkzeuge für Datenarchivierung:	Eigenentwicklung verfügbare Software (kommerziell, OpenSource) Collection Management System und Web-frontend (java-based-application on top of/around of the Fedora-Framework) Software „Coldfusion“ (bereitgestellt durch die Uni York)

Versionskontrolle:	Nein
Verhältnis manuelle/auto-matisierte Datenkuratierung:	16.021 (Stand Mai 2012) 20.058 (Stand Juli 2013) (ca. 320 verschiedene Datenprovider)
Aufwand Datenkuratierung insgesamt (Personen, Zeit):	19
wissenschaftliche Bewertung von Daten:	18
Verwendung Repository-System:	1.132.000 (aus 25 verschiedenen Ausgangskatalogen)
Software zur Verwaltung/Ablage der Archivdateien:	strukturierte, verzeichnisbasierte Dateiablage Eigenentwicklung verschiedene Scripts für Synchronisation, Checksummen, etc. verfügbare Software (kommerziell, OpenSource) verschiedene Tools für Konvertierung von Datenformaten
Software zur Verwaltung Metadaten/Rechte/Versionen:	Web-Oberfläche des CMS Eigenentwicklung Collection Management System auf Basis von Oracle & Coldfusion verfügbare Software (kommerziell, OpenSource) Tools für technische Informationen, v. a. DROID

## Archivierung von Daten

technischer Partner für LZA:	Universität York, UK Data Archive an der Universität Essex
Anzahl redundanter Datenkopien:	2
Administration Archivsystem inhouse/extern:	inhouse
garantierter Zeitraum:	kein fester Zeitraum garantiert
Erfahrungen mit technischer Migration von Archivdateien:	Ja

## Bereitstellung von Daten

Arbeitsschritte zur Datenbereitstellung:	- Aufbereitung Daten für Bereitstellung (→ DIP) - Eintragung Metadaten und Arbeitsschritte in CMS - Erstellung Projektseite für Website - Transfer Daten auf Web-Server und Archiv-Server
Werkzeuge für Datenbereitstellung:	Eigenentwicklung keine verfügbare Software (kommerziell, OpenSource) verschiedene Tools für Konvertierung von Datenformaten
Registrierung bzw. Login für Nutzer notwendig?	Nein (aber möglich)
Wird die IP und der Zugriff geloggt?	Ja
Gibt es verschiedene Authentifizierungsstufen?	Nein
Zugriff:	Open Access Embargo (Ausnahme): Falls erforderlich, werde im Einzelfall für Daten individuelle Embargo-Fristen ausgehandelt; eine moving wall existiert für manche Zeitschriften.

Wie wird ein Missbrauch von Daten verhindert/kontrolliert (z. B. wirtschaftliche Nutzung)?	keine aktive Kontrolle
Rollenkonzept:	Datengeber, Datennutzer, Datenarchivar, Administrator
Qualitätssicherung:	Zertifizierung, Data Seal of Approval, diverse Auszeichnungen und Preise
Schutzmechanismen für bestimmte Informationen:	keine personenbezogene Metadaten/Dateien nicht relevant raumbezogene Metadaten/Dateien Nein

## Metadaten & Interoperabilität

Mindestanforderungen an Daten:	Es gibt ausführliche Richtlinien und Formulare, wie Daten dokumentiert sein müssen und welche Formate akzeptiert werden. Für jedes Dateiformat (Bilder, Texte, Zeichnungen etc. gibt es individuelle Vorgaben, welche Metadaten erforderlich sind. Grundsätzlich gilt, dass Datengeber so viele Informationen wie möglich über Datensammlungen, Dateien und Projekte in strukturierter Form zur Verfügung stellen. Datengeber sind für die Qualität der Daten verantwortlich, d.h. je schlechter die Daten aufbereitet sind (technisch, inhaltlich), desto höher sind die Gebühren.
Verwendung von Standards:	Modellierung der Projekt- und Datendokumentation nach Standards z. B. des Digital Curation Centers, Arts and Humanities Research Councils und UK Data Archives. Hinsichtlich Begriffen Verwendung eigener Thesauri, die an andere angelehnt bzw. übernommen sind (z. B. English Heritage, Royal Commission of Scotland etc.)
Metadatenmodell:	Extended Dublin Core (see Guidelines for Depositors) MIDAS Standard, MIDAS2 = XML-Form of MIDAS, mapping zw. MIDAS und CIDOC-CRM eigene Anpassungen an Metadaten? keine; ggf. werden zusätzliche fachspezifische Informationen zu einzelnen Datenformaten erhoben
vorhandene Schnittstellen:	OAI-PMH Z39.50 Linked data Service (RDF): data.archaeologydataservice.ac.uk Webservices of Catalogues WMS/WFS bisher nicht implementiert
Sichtbarkeit der Metadaten in anderen Portalen/Aggregatoren:	Metadaten werden bereitgestellt für/angezeigt in: - Europeana - MEDIN (Marine Environmental Data) - Heritage Gateway - Monument Inventory Gateway (von English Heritage) - Portal von, Universitäten und Geldgebern - Zeitschrift „Internet Archaeology“ - ARENA-Portal - TAG-Portal = Archaeology Trans Atlantic Gateway, Datenaustausch zwischen tDAR und ADS
Sichtbarkeit der archivierten Dateien in anderen Portalen/Aggregatoren:	aus Lizenzgründen downloadbare Dateien nur über ADS-Website
System für Persistente Identifikatoren:	DOI (pro Projektarchive + Journal + Grey Literature Report) Anzahl PIDs insgesamt ca. 420 online Projektarchive (Stand Juni 2013) mit ca. 1.1 Mio Dateien (ohne eigene DOIs) ca. 20.000 Grey-Literature-Reports ca. 50 Zeitschriften (keine DOIs für einzelne Beiträge), DPhils etc. ca. 300 Internet Archaeology (DOIs für einzelne Beiträge)

## Nicht-Technische Dienstleistungen

Bereitstellung Guidelines, Ratgeber etc.:	- Entwicklung und Bereitstellung von "Guides to Good Practice" - Online-Material für Schulungen/Lehre - Begutachtung von Datenmanagementplänen von neuen Anträgen bei AHRC
---	--

Antragsberatung:	Ja (für AHRC)
Schulungen/Workshops:	Ja
Weiterentwicklung Standards:	Ja, laufende Projekte (Juli 2013): - ACE Archaeology in Contemporary Europe - SENESCHAL
Support Datenmanagement:	Ja, persönliche Beratung bei der Erstellung eines Datenmanagementplans (z. B. während einer Antragsstellung und bei der Planung neuer Projekte)

### Finanzen: Einnahmen

Jahresbudget:	unbekannt davon speziell für das Datenzentrum unbekannt
Basisfinanzierung:	keine Zeitraum der vertraglich Finanzierung keine nachhaltige Finanzierung? keine
Drittmittel:	unbekannt
restliche Finanzierung:	unbekannt
Gebühren für das Archivieren von Daten:	Ja
Gebühren für die Nutzung von Daten:	keine

### Finanzen: Ausgaben

Kostenstruktur nach Kostenträger:	Personalkosten Planstellen Drittmittelstellen	unbekannt
	Hardware/Software	unbekannt
	Reisekosten, Gebäudeinstandhaltung	unbekannt
Kostenstruktur nach Kostenstellen:	Daten akquirieren und aufbereiten	unbekannt
	Daten archivieren	unbekannt
	Daten zugänglich machen	unbekannt
	Softwareentwicklung	unbekannt
	Kosten für PIDs	unbekannt
Höhe Durchschnitts-Kosten pro 1 GB Daten:	unbekannt	

### Negativa

nicht funktionierende/ optimierbare Workflows oder Prozesse:	Metadaten für einzelne Dateien können verbessert werden (z. B. Beziehungen/Abhängigkeiten von Dateien zueinander, Dokumentation Archivierungsschritte mit einzelnen Dateien)
nicht gelöste Probleme:	alle für die Archivierung relevanten Fragen/Probleme ausreichend geklärt und technisch umgesetzt