

tDAR USA



Online-Begehung am 01.12.2013

teilweise aktualisiert Nov. 2014

Der Bericht basiert ausschließlich auf den Informationen der tDAR-Homepage

Aufgaben	
Hauptaufgaben:	digitales Repositorium mit dem Ziel der Archivierung und Bereitstellung von einmaligen digitalen Daten und Dokumenten aus dem Bereich der Archäologie; adressiert vor allem Nutzer und Einrichtungen in den USA
weitere Aufgaben:	Beratung, Forschung und Entwicklung im Rahmen von Drittmittelprojekten; Vergabe von Stipendien zur Datenkuratierung
nicht Teil der Aufgaben:	Bereitstellung von Forschungswerkzeugen, wissenschaftliche Auswertung von Daten
relevante Fachrichtung:	Archäologie

Allgemeine Angaben	
Rechtsform:	tDAR wird betrieben, entwickelt und gesteuert unter dem organisatorischen Dach von Digital Antiquity, einer non-profit Organisation an der Arizona State University, die eigens gegründet wurde, um die langfristige finanzielle, technische und soziale Zukunftsfähigkeit von tDAR zu gewährleisten
Gründung/Geschichte:	2004: National Science Foundation (NSF) finanziert einen Workshop 2006: Entwicklung eines digitalen Repository für archäologische Daten wird geplant 2008: archaeoinformatics.org wird entwickelt, mithilfe eines Stipendiums wird aus einem Prototyp eine Produktivsystem; Digital Antiquity wird als Organisation gegründet; mit Fedora wird ein Repository erstellt; Projekt „Trans-Atlantic-Gateway (TAG)“ zusammen mit Archaeology Data Service, UK 2011: Aufnahme der Nationalen Archäologischen Database (NADB) 2012: Digital Antiquity überarbeitet und aktualisiert tDAR; Neugestaltung der Oberfläche zur Datenerfassung, Management, Datenintegration und Suchfunktion
Anzahl Mitarbeiter:	Insgesamt 9 1 Executive Director; 1 Director of Technology; 1 Sales & Marketing Coordinator; 1 Digital Library Software Engineer; 4 Digital Data Curators; 1 Administrative Support Anzahl FTE (Planstellen) unbekannt Anzahl Hilfskraftstellen (Planstellen) unbekannt Anzahl FTE (befristet, Drittmittel) ca. 50 % Anzahl Hilfskraftstellen (befristet, Drittmittel) unbekannt Anzahl WissMA 5 Anzahl IT 2
Zertifizierung:	keine
Referenzmodell:	OAIS (SIP, AIP, DIP)

Datenbestand	
Verpflichtung Daten zu sichern?	National Science Foundation (NSF) und National Endowment for the Humanities (NEH) empfehlen tDAR und OpenContext für Datenarchivierung/-publikation
Findet gezielte Datenakquise statt?	unbekannt
Anzahl an Projektarchiven:	510 (Stand 12.2013) mit ca. 24.600 Dateien zum downloaden 609 (Stand 11.2014)

Anzahl an „grauer Literatur“:	ca. 3.500 (Stand 08.2013)
Anzahl an Doktorarbeiten:	60 (Stand 08.2013) 67 (Stand 11.2014)
Anzahl an retrodigitalisierten Zeitschriften:	123 downloadbare Artikel aus Zeitschriften
Anzahl an Einträgen Nachweiskatalog:	357.000 (Stand 12.2013) 363.966 (Stand 11.2014) Zitationen/Referenzen auf externe Ressourcen, primär Archiv-/Bibliotheksbestände, Publikationen und unveröffentlichte Berichte
Gesamtgröße Archiv:	180 GByte
Datensatzgröße:	jede einzelne Datei wird als separater Datensatz verwaltet, daher Größe pro Datensatz ca. 0 – 10 MByte
Datenzuwachs jährlich:	unbekannt
Altdaten vorhanden:	Ja

Datengenerierung

technische Unterstützung bei Datenerhebung (z. B. durch Tools, Software etc.):	Nein
fachliche Unterstützung bei Datenerhebung:	Ja, in Form von individuellen Beratungsleistungen zum Datenmanagement

Datenupload

Online-Upload möglich?	Ja, nach Anmeldung und nach Bezahlung
Auswahlkriterien für Aufnahme von Daten:	unterstützte Datenformate: JPG, TIF, GIF, PICT, BMP, PNG TXT, PDF, OBJ, DOC/X CSV, TAB, XLS/X, MDB SHP, PRJ, AUX, u. a.
Werden angebotene Daten abgelehnt? (Gründe)	falsche Formate, keine Relevanz
Vertragsverhältnis Datengeber - Datenarchiv:	„normaler“ Vertrag für Inanspruchnahme einer Dienstleistung, keine gesonderte Vereinbarung zw. Datengeber und tDAR
Software/Collection Registry:	Eigenentwicklung (s. u.)

Kuratierung von Daten

Arbeitsschritte zur Datenarchivierung:	unbekannt, erforderliche und durchgeführte Arbeitsschritte nach Upload von Daten durch Datengeber sind nicht dokumentiert
Werkzeuge für Datenarchivierung:	Eigenentwicklung unbekannt verfügbare Software unbekannt (kommerziell, OpenSource)
Versionskontrolle:	unbekannt
Verhältnis manuelle/auto-matisierte Datenkuratierung:	unbekannt

Aufwand Datenkuratierung insgesamt (Personen, Zeit):	unbekannt
wissenschaftliche Bewertung von Daten:	Nein
Verwendung Repository-System:	selbst entwickeltes Backend
Software zur Verwaltung/Ablage der Archivdateien:	Eigenentwicklung eines Repository-Systems auf Basis von PostgreSQL und Java Eigenentwicklung 100 % Eigenentwicklung verfügbare Software (kommerziell, OpenSource) keine
Software zur Verwaltung Metadaten/Rechte/Versionen:	Eigenentwicklung eines Repository-Systems auf Basis von PostgreSQL und Java Eigenentwicklung 100 % Eigenentwicklung verfügbare Software (kommerziell, OpenSource) keine

Archivierung von Daten

technischer Partner für LZA:	Arizona State University Library
Anzahl redundanter Datenkopien:	unbekannt
Administration Archivsystem inhouse/extern:	inhouse
garantierter Zeitraum:	keine Garantie
Erfahrungen mit technischer Migration von Archivdateien:	unbekannt

Bereitstellung von Daten

Arbeitsschritte zur Datenbereitstellung:	Daten werden durch Kuratoren für Bereitstellung vorbereitet
Werkzeuge für Datenbereitstellung:	Eigenentwicklung Online-Portal für Endnutzer ist zu 100 % Eigenentwicklung verfügbare Software (kommerziell, OpenSource) unbekannt
Registrierung bzw. Login für Nutzer notwendig?	unbekannt
Wird die IP und der Zugriff geloggt?	Ja
Gibt es verschiedene Authentifizierungsstufen?	unterschiedliche Rechte für Person/Institution und Author/Editor; eingeschränkter Zugriff auf vertrauliche Daten (confidential, sensitive) und Embargo für einzelne Datensätze
Zugriff:	Online über Portal und über technische Schnittstellen
Wie wird ein Missbrauch von Daten verhindert/kontrolliert (z. B. wirtschaftliche Nutzung)?	keine aktive Kontrolle
Rollenkonzept:	Public Access; Registered User; tDAR Curators; tDAR Administrators
Qualitätssicherung:	Kontrolle der Daten durch Datenkuratoren
Schutzmechanismen für bestimmte Informationen:	Datengeber kann Zugriffsrechte bestimmen und online verwalten personenbezogene Metadaten/Dateien Metadata is licensed under CC(by) raumbezogene Metadaten/Dateien Metadata is licensed under CC(by)

Metadaten & Interoperabilität	
Mindestanforderungen an Daten:	Formate müssen den akzeptierten Formaten entsprechen, Metadaten erforderlich entsprechend der Eingabemasken für das Online-Upload
Verwendung von Standards:	Ja, s. u.
Metadatenmodell:	Metadata Object Description Schema; Dublin Core eigene Anpassungen an Metadaten? vor allem auf der Ebene der Vokabulare: Cultural Terms, Material Types, Spatial Terms
vorhandene Schnittstellen:	OAI-PMH; OpenSearch; RSS; TAG Gateway (SOAP); Lookup Controller (REST)
Sichtbarkeit der Metadaten in anderen Portalen/Aggregatoren:	Inhalt von tDAR wird von den großen Suchmaschinen indiziert
Sichtbarkeit der archivierten Dateien in anderen Portalen/Aggregatoren:	Ja
System für Persistente Identifikatoren:	DOI, tDAR-ID Anzahl PIDs insgesamt 25.091 DOIs (insgesamt für Digital Antiquity und tDAR)

Nicht-Technische Dienstleistungen	
Bereitstellung Guidelines, Ratgeber etc.:	Informationen zum eigenen System: Wiki, Data Dictionary, Video Tutorials, User Documentation. Ansonsten zusammen mit ADS Weiterentwicklung der „Guides to Good Practice“
Antragsberatung:	Nein
Schulungen/Workshops:	Nein
Weiterentwicklung Standards:	Ja, zusammen mit ADS (s. o.)
Support Datenmanagement:	Ja, in begrenztem Umfang

Finanzen: Einnahmen																																	
Jahresbudget:	unbekannt davon speziell für das Datenzentrum unbekannt																																
Basisfinanzierung:	keine Zeitraum der vertraglich Finanzierung keine nachhaltige Finanzierung? Nein																																
Drittmittel:	Projektunterstützung durch National Endowment for the Humanities & The Andrew W. Mellon Foundation																																
restliche Finanzierung:	abgesehen von Drittmittelprojekten ausschließlich durch Gebühren																																
Gebühren für das Archivieren von Daten:	Gebühren richten sich vor allem nach Anzahl und Speichergröße der Dateien <table border="0"> <thead> <tr> <th>Item/Service</th> <th>Cost</th> <th>Item/Service</th> <th>Cost</th> </tr> </thead> <tbody> <tr> <td>1-9 Files</td> <td>\$50/file</td> <td>Data Curation</td> <td>\$40/hour</td> </tr> <tr> <td>10-49 Files</td> <td>\$40/file</td> <td>Consultation</td> <td>\$90/hour</td> </tr> <tr> <td>50-99 Files</td> <td>\$30/file</td> <td>Programming</td> <td>\$95/hour</td> </tr> <tr> <td>100-499 Files</td> <td>\$25/file</td> <td></td> <td></td> </tr> <tr> <td>500-999 Files</td> <td>\$20/file</td> <td></td> <td></td> </tr> <tr> <td>1.000-4.999 Files</td> <td>\$15/file</td> <td></td> <td></td> </tr> <tr> <td>5.000-9.999 Files</td> <td>\$5/file</td> <td></td> <td></td> </tr> </tbody> </table>	Item/Service	Cost	Item/Service	Cost	1-9 Files	\$50/file	Data Curation	\$40/hour	10-49 Files	\$40/file	Consultation	\$90/hour	50-99 Files	\$30/file	Programming	\$95/hour	100-499 Files	\$25/file			500-999 Files	\$20/file			1.000-4.999 Files	\$15/file			5.000-9.999 Files	\$5/file		
Item/Service	Cost	Item/Service	Cost																														
1-9 Files	\$50/file	Data Curation	\$40/hour																														
10-49 Files	\$40/file	Consultation	\$90/hour																														
50-99 Files	\$30/file	Programming	\$95/hour																														
100-499 Files	\$25/file																																
500-999 Files	\$20/file																																
1.000-4.999 Files	\$15/file																																
5.000-9.999 Files	\$5/file																																
Gebühren für die Nutzung von Daten:	keine																																

Finanzen: Ausgaben		
Kostenstruktur nach Kostenträger:	Personalkosten	unbekannt
	Planstellen	unbekannt
	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:	unbekannt
nicht gelöste Probleme:	unbekannt