

DIKUSA-Teilprojekt: „Sorbische Kulturdenkmale in Sachsen. Entwicklung einer virtuellen Arbeitsumgebung für ihre digitale Erfassung und Präsentation“

Datenmanagementplan

Erstellt durch: Robert Lorenc, Alf-Christian Schering, Konstantin Freybe

Datum: 07.07.2023

Version: 20230707_1

Administrative Angaben

Gesamtprojekttitel: Vernetzung digitaler Kulturdaten in Sachsen – Aufbau einer technischen Infrastruktur für die Forschung zu Mobilität, Migration und Transformation von Orten, Personen und Artefakten (in zeitlicher und räumlicher Perspektive) – DIKUSA

Antragsnummer Teilprojekt (SAB): 100613649

Kontonummer (SAB): - ; Bewilligung vom 25.02.2022

Kundennummer SI (SAB): 2000002349

Clusternummer (SAB): 4192

Fördergeber: Landesmittel des Freistaats Sachsen, Titelgruppe 70. Die Zuwendung stammt aus Steuermitteln. Diese Steuermittel werden auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes zur Verfügung gestellt.

Zeitraum: 01.02.2022-31.12.2025

Mitarbeitende/Datenurheber SI: Dr. Robert Lorenz, Dr. Friedrich Pollack (Leiter), Alf-Christian Schering

Koordinierung: KompetenzwerkD an der SAW - Dr. Dirk Goldhahn, Mag. Peter Mühleder, PD Dr. Franziska Naether

Hauptantragsteller: Sächsische Akademie der Wissenschaften zu Leipzig, vertreten durch den Präsidenten Prof. Dr. Hans Wiesmeth und den Generalsekretär Dr. Christian Winter

Kontakt: kompetenzwerkD@saw-leipzig.de

Kontakt SI: friedrich.pollack@serbski-institut.de

Kurzbeschreibung/Abstract Gesamtvorhaben:

Im Forschungsalltag stellt die niedrighschwellige digitale Wissenserfassung und -integration die außeruniversitären geisteswissenschaftlichen Forschungseinrichtungen Sachsens vor unterschiedlich große Herausforderungen. Ziel dieses Verbundvorhabens ist es, eine entsprechende technische Infrastruktur zu entwickeln und die Kompetenz der Häuser für die digitale Aufnahme von Archivmaterial und Objektdaten, deren Verlinkung, Visualisierungen sowie den Abgleich mit Normdatensätzen zu ermöglichen. Dafür wird zusammen mit dem KompetenzwerkD und der SLUB eine technische Basis bestehend aus mehreren Komponenten erstellt: eine zentrale Wissensbasis mit Referenzierungs- und Reconciliation-Diensten als Datenhub, eine „Toolbox“ mit nachnutzbaren Softwarekomponenten für die Datenarbeit und Dienste für die Visualisierung auf historischem Kartenmaterial. Dies ermöglicht zugleich eine attraktive Außendarstellung der Teilprojekte. Die

technische Infrastruktur wird im Rahmen von sechs Teilprojekten aus dem Bereich der Geisteswissenschaften und der Digital Humanities entwickelt, die Kulturdaten über Orte, Personen und Artefakte in der zeitlichen Perspektive seit dem Mittelalter bis zur Gegenwart in spezifischen Fragestellungen in den Blick nehmen.

Erster Pressebericht: <https://www.saw-leipzig.de/de/aktuelles/neues-verbundprojekt>

Kurzbeschreibung/Abstract Teilprojekt:

Ziel des Teilprojekts am Sorbischen Institut ist die Digitalisierung und Präsentation des Inventars sorbischer Kulturdenkmale für Sachsen auf einer virtuellen Plattform. Es markiert die materiellen Zeugnisse der sorbischen Geschichte als besonderen und zugleich integralen Teil der sächsischen Kulturlandschaft und macht sie einem breiten Publikum erstmals im bisher größten Umfang zugänglich. Im Rahmen des Projekts werden die infrastrukturellen Grundlagen für die strukturierte Erfassung und Verschlagwortung, die inhaltliche Darstellung sowie die kartografische Visualisierung sorbischer Kulturdenkmale geschaffen. Schnittstellen stellen die im Projekt erfassten Norm- und Kulturdaten nach offenen und verbreiteten Standards wie Linked Data systemübergreifend zur freien Verfügung.

Beschreibung der Daten und Metadaten

Beschreibung & Ursprung der Daten

Die im Rahmen des DIKUSA-Teilprojekts genutzten bzw. erstellten Daten stammen aus einem Archivbestand physischer Aktenordner, in denen Informationen zu Sorbischen Kulturdenkmälern zusammengetragen wurden. Die formale und inhaltliche Erschließung dieses Bestands bildet die Grundlage für deren Digitalisierung. Große Teile dieser Archivalien sind bereits i.w.S. in digitaler Form verfügbar. Hier gilt es die Kongruenz beider Bestände zu ermitteln und ggf. Diskrepanzen zu berichtigen.

- physische Dokumente
 - 61 Aktenordner erstellt von Trudla Malinkowa
- digitale Dateien
 - MS Access-DB
 - hierarchisch geordnet abgelegte Word-Dateien
 - autom. erstellte Sicherungskopien obiger Word-Dateien (ca. 50%)
 - überwiegend große Schnittmenge zwischen Access-DB & Word-Dateiensammlung

Verwendete Datentypen und -formate

- analoge Fotografien, incl. Negativstreifen
- 2019 Textdateien, in den Formaten .doc (1902), .docx (1), .rtf (56), .wps (59), .pdf (1)
- vier Tabellendateien, im Format .wks
- 1158 Backup-Dateien, in den Formaten .wbk (1099), .bps (59)
- 948 Bilddateien, in den Formaten .jpg (936), .tif (11), .pdf (1)
- Vortragsfolien im Format .ppt

Umfang

- Textdaten: 100 MB
- analoge Fotografien: 6.300 Einzelmotive

- Bilddaten: 3.29 GB

Dazugehörige Metadaten

- Geografische Informationen/Verknüpfungen ins Ortsregister
- Amtliches sorbisches Siedlungsgebiet: ja/nein
- Personendaten/-verknüpfungen
- Zeitliche Einordnung (Entstehung, Verlust, etc.)
- Denkmalinschriften (Sprachvarianten)
- Objektinformationen
- Denkmalschutzstatus ja/nein

Lokale Wissensbasis:

Beschreibung:

Die lokale Wissensbasis liegt als PostgreSQL Datenbank vor und wird auf einem Webserver des Vertragspartners Strato betrieben. Sie orientiert sich im Wesentlichen an den in der vorliegenden Datenbank vorhandenen Daten/Metadaten. Neben der Hauptdatenbank „tmd.pomnik“ existieren noch weitere Tabellen, in denen Links über Fremdschlüssel auflösbar gehalten werden, die kontrollierte Vokabular für Denkmaltypen und Klassifikationen durch Fachwissenschaftler enthalten.

Datenmodell:

- relationales Datenmodell (PostgreSQL)
- Verwendung von Normdatenmodellen, wie CIDOC, LIDO
- Datenmodell „**tmd.pomnik**“:
- Wesentliche Attribute der Denkmaldatensätze:
ID, Objekt, Standort, Entstehungsjahr, Entstehung, Verlust, Verlustjahr, Verlustbeschreibung, Material, Zustand, Erfassung, Erfasser, Attribute nach dem Eventmodell (LIDO): Personen, Zeiten, Beschreibungen, Inschriften, Basisdokument, ID-Landesamt, Siedlungsgebiet, Sprachgebiet
- zugeordnete Tabellen:
 - „**tmd.category**“: Mapping-Tabelle, kontrolliertes Vokabular
 - „tmd.image“: Ablageorte und Dateinamen von Bilddateien zu Denkmaldatensätzen
 - „tmd.pomnik_bak“: autom. Backup
 - „tmd.pomnik_category“: kontrolliertes Vokabular für Denkmaltypen
 - „tmd.predicate_workflow“: kontrolliertes Vokabular für Revision durch Fachwissenschaftler
 - „tmd.textfile“: Ablageorte und Dateinamen zu Textdateien, aus denen Daten extrahiert werden/wurden

Verwendete Vokabulare / Referenzierte Normdatensätze:

- Register Sorbisches Kulturerbe
 - Orte: Verknüpfungen zu Ortsnormdaten im Register auf sechs variablen Ebenen (Ort, Gemeinde, Amt/Verwaltungsgemeinschaft, Kreis, Region, Land)
 - Personen: Verknüpfung per Event-Modell (siehe oben)
 - Weitere Normdaten im Aufbau
- GND – für Personen
- WikiData
- ISGV – HOV für Ortsnamen
- GeoNames – bisher exemplarisch, flächendeckende Verwendung in Planung

Link zu externen Schemata:

- LIDO: <http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html>
- CIDOC: <https://cidoc.mini.icom.museum/>

Ethische und rechtliche Aspekte

[Dieser Abschnitt bedarf noch weiterer Diskussion in der Arbeitsgruppe. Er wird im Herbst 2023 entsprechend aktualisiert werden.]

Während sich mit Blick auf die bestehenden Datensätze und die ihnen zugeordneten Bilddateien momentan keine akuten Handlungsbedarfe abzeichnen, stellt sich die Situation in absehbarer Zukunft differenzierter dar. Soll der Anspruch weiterverfolgt werden, die Denkmaldaten so verfügbar zu machen, dass Bürger*innen im Sinne einer Citizen Science partizipieren, gilt es zwei Aspekte zu berücksichtigen:

- 1) Von Bürger*innen beigetragene Texte und Bilddateien müssen in ein angemessenes Lizenzierungsmodell eingepasst sein, um rechtlich abgesichert den Informationsdienst betreiben zu können. Die Entscheidung über das konkret verwendete Lizenzmodell steht momentan aus.
- 2) Die von nicht-akademischen Beitragenden produzierte Wissen über Sorbische Kulturdenkmale muss mit den Anforderungen an wissenschaftliche Wissensressourcen in Einklang gebracht werden. Die Formulierung von Gütekriterien und die Entwicklung redaktioneller Workflows steht momentan noch aus.

Metadaten (Datenschutz/Persönlichkeitsrechte):

- bzgl. der Lieferung im Rahmen des DIKUSA-Projekts wurde bereits eine Vorauswahl/Filterung der Attribute der Denkmaldatenbank vorgenommen, womit ggf. bedenkliche Inhalte in der internen Datenbank bleiben und nicht in die Öffentlichkeit gelangen

Bilddatenbank (Datenschutz/Persönlichkeitsrechte):

- Zu klären, wie mit abgebildeten Personen und zu erkennenden Adressen umgegangen werden wird

Speicherung, Archivierung und Sicherung der Daten (und Metadaten)

Arbeitsabläufe

Bestandserschließung

Die Denkmal-Daten stammen aus einer Dokumentensammlung, die sowohl physische Dokumente als auch digitale Textdateien sowie eine Access-Datenbank umfasst. Die Kongruenz zwischen physischen und digitalen Dokumenten wird manuell geprüft. Formale und inhaltliche Erschließung dieser Archivbestände erfolgt durch Robert Lorenc, dessen fachliche Expertise maßgeblich ist für die Qualifikation von Datensätzen als ‚DIKUSA-tauglich‘. Robert Lorenc berichtet über seine Fortschritte in Form interner Vorträge und im Rahmen projektbezogener e-Mail-Korrespondenz.

Qualitätssicherung

Einerseits mussten Digitalisate physischer Dokumente erzeugt werden, andererseits mussten digitale Textdateien zu HTML transformiert, anschließend geparkt und schließlich in eine PostgreSQL-Instanz importiert werden. Dieser Prozess ist bereits weit fortgeschritten. Eine

abschließende inhaltliche Prüfung steht noch aus. Die Konsolidierung der Metadaten erfolgt auf Grundlage der Datenbank. In einem fortlaufenden IT Prozess werden Textdaten in Freitextfeldern nach und nach in kontrollierte Vokabulare überführt. Dies geschieht automatisch, semi-automatisch und mitunter manuell, letzteres auf Grundlage der Expertise von Robert Lorenc.

Datenhaltung

Bzgl. der Datenhaltung ist wichtig, dass der Bezug zum Ausgangsmaterial nicht verloren geht. Die favorisierte Strategie besteht darin, ‚Rohdaten‘ (i.w.S.) stets referenziert und digitale Dateien mittels Checksummen überprüfbar zu machen, um eine Übereinstimmung sicherzustellen bzw. feststellbar zu machen.

Die Denkmaldaten enthalten auch Verweise auf das Register Sorbisches Kulturerbe (RSK), das seinerseits Normdaten für Sorbisches Kulturerbe bereitstellt/bereitstellt. Diese Verweise werden über entsprechende Links (semantische Vernetzung) hergestellt.

Recherche

Die testweise Nutzung bereits bestehender, über das Internet zugänglicher Datenbanken zu Geografika dient der Formulierung von Anforderungen an den vom SI zu erstellenden Dienst. Die Auswahl geeigneter Kandidaten erfolgt weitgehend formlos innerhalb des Projektteams. Die Dokumentation von Testnutzungen in Form von ‚User Stories‘ — kurze Mitschriften der Nutzungserfahrung — wird derzeit erprobt und ggf. so standardisiert, dass i.w.S. ein weiterer Forschungsdatentyp im FDM berücksichtigt wird.

Kooperation

- Register Sorbisches Kulturerbe
- Landesdenkmalamt Sachsen
- Citizen-Science-Akteure

Datenspeicherung

- gespeichert auf RSK-Server, wird von Auftragspartner (Strato) betrieben, der eigene Backup-Strategie verfolgt;
- Es werden keine Daten dezentral gehalten (z.B. Laptop, eigene Dateien, etc.)

Datensicherung

- erfolgt auf Grundlage von Cronjobs, rollendes Backup (DB-Dumps), nächtlich mit automatischer Benennung

Archivierung

- Speicherung von Forschungsdaten als .ZIP/ .TAR sowie auf Tape (auf jeden Fall zum Ende des Projektes geplant)
- Erste Tests mit Open Archival Information System (OAIS) ... zu eruieren
- Die Denkmaldaten des Teilprojekts werden nach dessen Abschluss nicht eingemottet, sondern aktiv weiterbearbeitet (im Live-Betrieb)

Datenverfügbarkeit und Dokumentation

Datenverfügbarkeit

Momentan sind alle Pomnik-Daten nur einem bestimmt umgrenzten Personenkreis zugänglich. Im Rahmen des DIKUSA-Projekts werden per RDF-Dump die Daten über die gemeinsame Schnittstelle des Verbund-Projekts verfügbar gemacht. Details zu diesem Prozess, dem

verwendeten Datenmodell und der weiteren Zugänglichmachung der so bereitgestellten Daten finden sich im zentralen Datenmanagementplan des Gesamtprojekts. Deskriptive Metadaten sind (perspektivisch) im Register des Sorbischen Kulturerbes verfügbar und verwaltbar. Perspektivische Bereitstellung der Denkmaldaten im institutseigenen Schaufenster „Sorabicon“.

Aufgrund der Konsolidierung der Datenbasis, werden die Daten nicht sofort veröffentlicht, sondern später im Verlauf des Projektes.

Dokumentation

Gegenwärtig werden große Teile der Dokumentation im Rahmen des Forschungsdatenmanagement dokumentiert. Ziel dieses Ansatzes ist, durch regelmäßige Pflege des Datenmanagementplans den Projektverlauf abbilden und textuelle Ressourcen für die Dokumentation des Projekts in separaten Dokumenten zu erzeugen, so dass spätere Arbeitsschritte zügiger erledigt werden können.

Teile der Dokumentation, insbesondere die der Eingabe- und Verwaltungsschnittstellen, erfolgen ebenda.

Verantwortlichkeiten / Organisatorisches

Robert Lorenc und Alf-Christian Schering sind für das FDM zuständig und tragen vom Projektteam relevante Informationen zusammen, die sie anschließend im DMP aufbereiten. Sie fungieren als Ansprechpartner für diesen Bereich.

Kostenfragen

Die anfallenden Aufgaben werden durch das im Rahmen des Projektes eingestellte Personal erledigt.