



[Tobias Kraft und Christian Thomas]

Datenmanagementplan des Akademienvorhabens „Alexander von Humboldt auf Reisen – Wissenschaft aus der Bewegung“

Stand: März 2022

Persistent Identifier: [urn:nbn:de:kobv:b4-opus4-37252](https://nbn-resolving.org/urn:nbn:de:kobv:b4-opus4-37252)

Die vorliegende Datei wird Ihnen von der Berlin-Brandenburgischen Akademie der Wissenschaften unter einer Creative Commons Namensnennung 4.0 International Lizenz zur Verfügung gestellt.



Datenmanagementplan des Akademienvorhabens „Alexander von Humboldt auf Reisen – Wissenschaft aus der Bewegung“

Kurzbeschreibung

Das Akademienvorhaben „Alexander von Humboldt auf Reisen – Wissenschaft aus der Bewegung“ (AvH-R) verfolgt das Ziel einer Open Science und strebt durch die Veröffentlichung des Datenmanagementplans (DMP) eine nachhaltige und transparente Dokumentation während und nach Abschluss des Forschungsprozesses an. Der Datenmanagementplan in dieser ersten publizierten Version beschreibt den Umgang mit den erzeugten sowie gesammelten Forschungsdaten im laufenden Akademienvorhaben (Stand: Mai 2022).

Der DMP wurde über die BBAW-Instanz des „Research Data Management Organisers“¹ erstellt, einer Open Source Software zur kollaborativen Erstellung von Datenmanagementplänen. Dieses Papier basiert auf dem Fragenkatalog der BBAW-Initiative „Forschungsdatenmanagement“².

Das Dokument teilt sich in die Abschnitte „Allgemein“, „Inhaltliche Einordnung“, „Datensicherung“, „Datennutzung“, „Metadaten und Referenzierung“, „Rechtliche und ethische Fragen“ sowie „Speicherung, Publikation und Langzeitarchivierung“. Die Forschungsdaten des Akademienvorhabens sind wiederum eingeteilt in die Datensätze „Bilddigitalisate“, „Texte“, „Register“ und „Chronologie“, für welche die Fragen des Datenmanagementplans jeweils gesondert beantwortet werden.

¹ <https://github.com/rdmorganiser>

² <https://forschungsdatenmanagement.bbaw.de>

Allgemein

Thema

Wie lautet die primäre Forschungsfrage des Projektes?

Das Akademienvorhaben „Alexander von Humboldt auf Reisen – Wissenschaft aus der Bewegung“ umfasst die vollständige Edition der Manuskripte Alexander von Humboldts zum Themenkomplex Reisen. Das Korpus der Edition beinhaltet Tagebücher, Dokumente und Korrespondenzen im Kontext der Humboldt'schen Forschungsreisen. Im Zentrum des Vorhabens stehen die Manuskripte sowohl der Amerikanischen als auch der Russisch-Sibirischen Reisetagebücher. Darüber hinaus werden umfangreiche Materialien aus dem Humboldt-Nachlass an der Staatsbibliothek zu Berlin-PK und der Biblioteka Jagiellońska in Krakau inhaltlich erschlossen und nach Themenschwerpunkten ediert. Die Reihe *edition humboldt* wird sechzehn Bände umfassen, die im Sinne eines Hybrid-Projektes sowohl als Printausgaben (*edition humboldt print*) als auch als digitale Edition (*edition humboldt digital*) veröffentlicht werden. Das Forschungs- und Editionsprojekt erfüllt seine Aufgaben in enger Kooperation mit der Universität Potsdam, der Staatsbibliothek zu Berlin - PK, der Technischen Universität Berlin und anderen Forschungseinrichtungen in der Region Berlin-Brandenburg. Gemeinsam mit der Universität Potsdam wird die Internet-Zeitschrift *HiN – Alexander von Humboldt im Netz* (ISSN 1617-5239) zweimal jährlich herausgegeben.

Wie lautet das offizielle Akronym oder Kürzel des Projektes?

AvH-R

Disziplin

Welcher Disziplin / welchen Disziplinen ist das Projekt zuzuordnen?

- Geistes- und Sozialwissenschaften / Geschichtswissenschaften
- Geistes- und Sozialwissenschaften / Literaturwissenschaft
- Geistes- und Sozialwissenschaften / Sprachwissenschaften

Welchem Zentrum der BBAW ist das Projekt zuzuordnen?

Zentrum Preußen - Berlin

Verbindungen

Welche Art von Beziehung besteht zwischen dem Projekt und der angegebenen Verbindung?

Verbindung "*edition humboldt digital (ehd)*": Digitale Ressource

Verbindung "*Zotero*": Literaturdatenbank (z.B. Zotero, Citavi)

Verbindung "*Twitter*": Wissenschaftskommunikation / Social Media

Verbindung "*Alexander-von-Humboldt-Forschung*": Vorgängerprojekt

Verbindung "*HiN – Internationale Zeitschrift für Humboldt-Studien*": Andere: internationales peer-reviewed Journal

Verbindung "*Die unselbständigen Schriften Alexander von Humboldts*": Digitale Ressource (Vorgängerprojekt)

Verbindung "*Deutsches Textarchiv (DTA)*": Digitale Ressource

Verbindung "*avhumboldt.de*": Wissenschaftskommunikation / Social Media

Verbindung "*Twitter-Bot AvHChrono*": Wissenschaftskommunikation / Social Media

Verbindung "*edition humboldt print (ehp)*": Andere: Druck- und ebook-Ausgabe der Berliner edition humboldt

Link zur digitalen Ressource / dem Webauftritt / dem Social-Media-Profil

Verbindung "*edition humboldt digital (ehd)*": <https://edition-humboldt.de/>

Verbindung "*Zotero*": <https://www.zotero.org/groups/667230/avh-r/>

Verbindung "*Twitter*": https://twitter.com/avhr_bbaw

Verbindung "*Alexander-von-Humboldt-Forschung*":
<https://webarchive.bbaw.de/default/20201029141512/http://avh.bbaw.de/>

Verbindung "*HiN – Internationale Zeitschrift für Humboldt-Studien*": <https://www.hin-online.de>;
https://www.hin-online.de/index.php/hin/archiving_policy

Verbindung "*Die unselbständigen Schriften Alexander von Humboldts*":
<https://webarchive.bbaw.de/default/20201029141512/http://telota.bbaw.de/avh/uns/>

Verbindung "*Deutsches Textarchiv (DTA)*":
<https://www.deustextarchiv.de/search/metadata?corpus=avh>

Verbindung "*avhumboldt.de*": <https://www.avhumboldt.de/>

Verbindung "*Twitter-Bot AvHChrono*": <https://twitter.com/AvHChrono>

Verbindung "*edition humboldt print (ehp)*": <https://www.springer.com/series/16345>

Ggf. Erläuterung

Verbindung "*edition humboldt digital (ehd)*": Ausgehend von einem „digital first“-Ansatz und auf Grundlage eines in TEI-XML erstellten Textkorpus' erarbeitet das Vorhaben sowohl eine Print- als auch eine digitale Edition. Die Printedition der Tagebücher konzentriert sich auf eine Rekonstruktion des originären Reiseverlaufs und ist als Lesefassung konzipiert. Die digitale Edition zielt auf eine möglichst umfassende textorientierte Transkription und Kommentierung der Handschriften sowie auf eine intelligente Nutzung normdateibasierter Webdienste und Informationsangebote (s. Tobias Kraft und Stefan Dumont: „The Humboldt Code. On creating a hybrid digital scholarly edition of a 19th century globetrotter“, Wiener Digitale Revue 1 (2020), <https://doi.org/10.25365/wdr-01-03-02>). Die digitale Edition wird in Umfang und Darstellungsform über die Möglichkeiten des Drucks hinausgehen und bereits vor Drucklegung den aktuellen Stand der Arbeit an den Texten und Datensätzen als work in progress präsentieren.

Verbindung "Zotero": Die Bibliographie der Edition wird in der Literaturverwaltungssoftware Zotero gepflegt. Die öffentlich einsehbare Zotero-Gruppe ermöglicht die kollaborative Pflege der Daten und kann von allen Interessierten abgerufen werden – auch in verschiedenen Zitierstilen und Exportformaten. Eine Kopie der Bibliographie befindet sich in eXistdb.

Verbindung "Twitter": Twitter-Account des Akademienvorhabens AvH-R

Verbindung "Alexander-von-Humboldt-Forschung": Die Arbeitsstelle edierte den handschriftlichen Nachlass und die unselbständig erschienenen Schriften Alexander von Humboldts (1769–1859) und dokumentiert dessen Leben und Werk. Am 31. Dezember 2014 wurde die zum Zentrum Preußen – Berlin gehörende Arbeitsstelle Alexander-von-Humboldt-Forschung planmäßig geschlossen. Seit 1968 waren in der Reihe „Beiträge zur Alexander-von-Humboldt-Forschung“ 20 Bände mit Briefwechseln, sechs Tagebucheditionen, zwei Bibliographien, fünf Monographien und vier Bände mit Aufsätzen erschienen. Die Forschungsstelle publizierte zahlreiche kürzere Beiträge in der Reihe „Berliner Manuskripte zur Alexander-von-Humboldt-Forschung“ und war Mitherausgeberin der Internationalen Zeitschrift für Humboldt-Studien HiN – Humboldt im Netz (ISSN 1617-5239, bisher 29 Hefte). Die Sammlungen der Arbeitsstelle sind bis heute das weltweit reichste Archiv mit Materialien von und über A. v. Humboldt. Die Projektwebsite ist archiviert.

Verbindung "HiN – Internationale Zeitschrift für Humboldt-Studien": HiN ist eine internationale Open Access Zeitschrift (CC BY-NC 4.0 Lizenz) und veröffentlicht seit 2000 kontinuierlich aktuelle Forschung zu Alexander von Humboldt in Deutsch, Englisch, Spanisch und Französisch. Das halbjährlich erscheinende, digitale Periodikum ist eine Publikation der Universität Potsdam und der Berlin-Brandenburgischen Akademie der Wissenschaften. Seit 2018 erscheint die Zeitschrift sowohl online als auch gedruckt (ISSN online 1617-5239, print 2568-3543).

Verbindung "Die unselbständigen Schriften Alexander von Humboldts": Neben seinen in Buchform publizierten Schriften hat Alexander von Humboldt zwischen 1789 und 1859 eine Vielzahl kleinerer Schriften, Abhandlungen, Akademiereden, Gutachten, Buchauszüge und Vorworte veröffentlicht. Seit 2008 liegt eine Datenbank dieser Schriften vor. Diese wurde am 29.10.2020 archiviert und durch die Bereitstellung der Texte im Deutschen Textarchiv (DTA) ersetzt.

Verbindung "Deutsches Textarchiv (DTA)": 2015-2016 wurden in Kooperation mit dem Deutschen Textarchiv (DTA) 164 der unselbständigen Schriften aufwendig digital bearbeitet und als Volltexte veröffentlicht (<https://www.deutschestextarchiv.de/dtae/avh>). Alle Texte wurden vom DTA in TEI-XML (DTA-Basisformat) kodiert und um die entsprechenden Bilddigitalisate ergänzt. Darüber hinaus finden sich im DTA weitere, für das Akademienvorhaben AvH-R relevante Texte, die ggfs. mit der *edition humboldt* verlinkt werden können. So wurde der Volltext des publizierten Reiseberichts von Gustav Rose, Humboldts Begleiter in Russland und Sibirien 1829 im DTA publiziert (vgl. dazu <https://www.avhumboldt.de/?p=15955>), eben Nachschriften von Humboldts Kosmos-Vorträgen 1827/28 (vgl. <https://www.deutschestextarchiv.de/doku/textquellen#avhkv>).

Verbindung "avhumboldt.de": avhumboldt.de – Alexander von Humboldt Informationen online ist die Online-Informationenplattform zu Humboldt-Aktivitäten weltweit und das Nachfolgeprojekt zu www.humboldt-im-netz.de. Seit 1999 werden auf dieser Seite Informationen und Materialien zu Leben und Werk von Alexander von Humboldt zusammengestellt und kommentiert. avhumboldt.de ist ein gemeinsames Projekt der Universität Potsdam und der Berlin-Brandenburgischen Akademie der Wissenschaften.

Verbindung "Twitter-Bot AvHChrono": @AvHChrono, Beiträge basieren auf den Tageseinträgen der Alexander von Humboldt-Chronologie der edition humboldt digital (<https://edition-humboldt.de/X0000001>). Initiiert und gepflegt von Jürgen Hermes, Universität Köln.

Verbindung "edition humboldt print (ehp)": edition humboldt print (J.B. Metzler / Springer Nature), Hrsg. von Ottmar Ette und den MitarbeiterInnen des Akademienvorhabens AvH-R; Umfang (geplant): 16 Bde. in 3 Reihen. Print & eBook [Paywall].

Projektablauf

Wann beginnt oder begann die Projektlaufzeit?

1. Januar 2015

Wann endet die Projektlaufzeit?

31. Dezember 2032

Wann findet die nächste Evaluation statt?

23.-24. August 2022

Wann findet die nächste Durchführungskontrolle statt?

1. Januar 2025

Ggf. Erläuterung

Der genaue Termin der Durchführungskontrolle wird erst in der 2. Jahreshälfte 2022 bekannt gegeben.

Projektkoordination

Name

Beteiligte Person "OE": Ottmar Ette

Beteiligte Person "TK": Tobias Kraft

Welche Rolle hat diese Person im Projekt?

Beteiligte Person "OE": Projektleitung

Beteiligte Person "TK": Arbeitsstellenleitung

Mitarbeitende

Besteht eine Zusammenarbeit zwischen Ihrem Projekt und TELOTA?

Ja

Wie viele Mitarbeitende beschäftigt das Projekt aktuell und welche Aufgaben erfüllen diese?

1,0 VZÄ Arbeitsstellenleiter; 2,8 VZÄ WiMi AvH-R; 1 SHK AvH-R; 0,5 VZÄ WiMi TELOTA; weitere Personen ehrenamtlich beteiligt; zu den Aufgaben Einzelner vgl. <https://edition-humboldt.de/H0002928>.

Projektpartner

Name des Projektpartners

Projektpartner "SBB": Staatsbibliothek Berlin Preußischer Kulturbesitz

Projektpartner "Biblioteka Jagiellońska": Biblioteka Jagiellońska, Krakau

Projektpartner "Uni Potsdam": Universität Potsdam

Ggf. Erläuterung

Projektpartner "SBB": Es werden umfangreiche Materialien aus dem Humboldt-Nachlass an der Staatsbibliothek zu Berlin-PK und der Biblioteka Jagiellońska in Krakau in fünf Materialbänden inhaltlich erschlossen und nach Themenschwerpunkten ediert.

Projektpartner "Biblioteka Jagiellońska": Es werden umfangreiche Materialien aus dem Humboldt-Nachlass an der Staatsbibliothek zu Berlin-PK und der Biblioteka Jagiellońska in Krakau in fünf Materialbänden inhaltlich erschlossen und nach Themenschwerpunkten ediert.

Projektpartner "Uni Potsdam": Das Forschungs- und Editionsprojekt erfüllt seine Aufgaben in enger Kooperation mit der Universität Potsdam (Lehrstuhl Prof. Dr. Ottmar Ette), der Staatsbibliothek zu Berlin – PK und anderen Forschungseinrichtungen in der Region Berlin-Brandenburg. Gemeinsam mit der Universität Potsdam wird die Internet-Zeitschrift „HiN – Humboldt im Netz“ zweimal jährlich herausgegeben.

Förderung

Wer fördert das Projekt?

Akademienprogramm (Bund und Länder)

Gibt es von Seiten der Forschungsförderer Vorgaben oder Richtlinien bezüglich des Umgangs mit den im Projekt erhobenen Forschungsdaten?

Dem Vorhaben sind das Leitbild Open Science und die Lizenzierungsleitlinien der BBAW bekannt. Das Vorhaben orientiert sich zur möglichst offenen Bereitstellung der eigenen Forschungsdaten an diesen institutionellen Vorgaben.

Weitere Anforderungen

Gibt es von weiteren Seiten (z.B. von der Fachcommunity) Anforderungen an das Datenmanagement, die beachtet werden müssen?

Ja

Welche Anforderungen an das Datenmanagement sind dies?

Das Datenmodell der *edition humboldt digital* basiert in weiten Teilen auf dem Basisformat des Deutschen Textarchivs (DTABf) und folgt damit einem von der DFG empfohlenen Standard der Textauszeichnung und -archivierung in TEI-XML (Deutsche Forschungsgemeinschaft (2015): Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft. Bonn, S. 6). In Einzelfällen, die in der Spezifik des Korpus begründet sind, weicht die Kodierung vom DTABf ab bzw. geht über dessen Datenmodell hinaus.

Inhaltliche Einordnung

Datensätze

Um was für Daten handelt es sich?

Datensatz "Bilddigitalisate": Digitale Faksimiles der historischen Quellen (Original-Handschriften).

Datensatz "Texte": Die Hauptaufgabe des Vorhabens besteht in der vollständigen Herausgabe der Reisetagebücher der zwei großen Reisen Humboldts durch die Amerikas und Russland sowie weiterer Reisen insbesondere nach England, Südfrankreich, Spanien und Italien. Von überragender Bedeutung sind gerade bei den beiden transkontinentalen Reisen die bislang fast völlig unberücksichtigt gebliebenen literarischen Reiseskizzen, Messergebnisse, Laborprotokolle, wissenschaftlichen Essays und Exzerpte, die einen noch nicht gehobenen Schatz transdisziplinär vernetzter Erkenntnisse an der Schnittstelle von Kultur- und Naturwissenschaften darstellen. Ziel des Forschungsprojektes ist die erstmalig vollständige Edition dieses Quellenmaterials, an dem Humboldt auch nach den Reisen wie an einem lebenslangen „Projekt Reisetagebücher“ unablässig weitergearbeitet hat. Die Edition wird durch die Herausgabe derjenigen ausgewählten Briefwechsel Humboldts abgerundet, die innerhalb seines weltweiten wissenschaftlichen Netzwerks mit den Forschungsreisen in unmittelbarer Beziehung stehen und zusätzliche Perspektiven eröffnen.

Datensatz "Register": Alle Personen-, Orts- und Institutionsnamen sowie Literaturangaben in den Quellentexten (und ihren Metadaten) werden mit ihren jeweiligen Registereinträgen verknüpft. Darüber hinaus entsteht ein mit externen taxonomischen Datenbanken verknüpftes Pflanzenregister sowie ein vom Vorhaben angelegtes Glossar.

Datensatz "Chronologie": Chronologische Übersicht über wichtige Daten des Lebens Alexander von Humboldts.

Bitte kategorisieren Sie den Datensatz durch die Auswahl eines Datentyps.

Datensatz "Bilddigitalisate": Bilder, Grafiken, Digitalisate

Datensatz "Texte": Annotierter und/oder strukturierter Text

Datensatz "Register": Annotierter und/oder strukturierter Text

Datensatz "Chronologie": Annotierter und/oder strukturierter Text

Mit welcher Methode werden die Daten erhoben oder erstellt?

Datensatz "Bilddigitalisate": Digitalisierung der Manuskripte

Datensatz "Texte": Die XML-Datensätze entstehen durch die grundständige Transkription sowie historisch-kritische Edition des Quellenmaterials (Reisetagebücher, Dokumente und Briefe). Einige der Dokumente bzw. Forschungsdaten (z.B. aus Registern) wurden aus früheren Druck-Editionen retrodigitalisiert, und werden innerhalb der *ehd* grundlegend überarbeitet neu herausgegeben. Alle edierten Texte erscheinen in der digitalen Edition. Eine Auswahl dieser Texte erscheint in einer für den Druck optimierten Auswahl und Darstellung in Buchform.

Print: Für die Edition der Reisetagebücher sind 11 Bände veranschlagt, für die Edition der Themenschwerpunkte aus Nachlass und Korrespondenz 5 Bände.

Digital: Die textuelle Ebene erfasst den eigentlichen Text und wird in drei Darstellungsvarianten angezeigt: „Text mit Faksimile“ beinhalten alle durch die HerausgeberInnen vorgenommenen textkritischen Auszeichnungen; „Lesetext (Einzelseiten)“ und „Lesetext (Gesamtansicht)“ zielt auf eine lesefreundliche Darstellung und vereinfacht das Schriftbild (Auflösung des Zeilenumbruchs, der Abkürzungen und Textkorrekturen durch den Autor und die HerausgeberInnen, Normalisierung der Orthographie, stillschweigende Übernahme von Textergänzungen durch die HerausgeberInnen). Die Erfassung der inhaltlichen Ebene konzentriert sich darauf, möglichst viele der in den Texten enthaltenen Informationen als strukturierte Daten aufzubereiten (Personen-, Orts- und Institutionsnamen, Werktitel, Pflanzennamen, Maßangaben) und, wo als sinnvoll erachtet, durch Stellenkommentare zu erweitern.

Datensatz "Register": Die Register zu Personen, Orten und Institutionen sowie das Glossar und ein Register zu historischen Maßen, Gewichten und Währungen werden in TEI-XML bearbeitet und vorgehalten. Jeder Eintrag verfügt über eine eigene, eindeutige und permanente Identifizierungsnummer. Er wird darüber hinaus mit einer oder ggf. mehreren IDs aus den entsprechenden Normdateien versehen (sofern ein Eintrag in der Normdatei existiert), um eine projektübergreifende Identifizierung der Personen, Orte etc. zu ermöglichen. Zusätzlich zu den Basisdaten eines Eintrags wird außerdem eine Kurzbeschreibung vorgehalten. Darüber hinaus wurden im Rahmen des Projekts die Registerangaben von 25 zwischen 1973 und 2016 veröffentlichten Editionen (Briefe, Dokumente, Tagebuch-Exzerpte) der Alexander von Humboldt-Forschung retrodigitalisiert und zusätzlich bereitgestellt. Die Registerinformationen der gedruckten Editionen sind Teil der zunächst im Akademie-Verlag und schließlich bei De Gruyter veröffentlichten Reihe Beiträge zur Alexander von Humboldt-Forschung, herausgegeben von der Alexander-von-Humboldt-Forschungsstelle (1970–2014) der BBAW. Verknüpfungen zu anderen Registerinträgen und den edierten Materialien werden automatisiert aus der Datenbank ausgelesen. Für das Pflanzenregister werden keine eigenen Einträge in TEI-XML-Registerdateien angelegt (wie bei den anderen Registern). Stattdessen werden die wissenschaftlichen Pflanzennamen in den edierten Texten mit Hilfe der entsprechenden TEI-Kodierung normalisiert und dann automatisch ausgelesen. Das Pflanzenregister wird somit komplett dynamisch erstellt. Zu jedem distinkten Pflanzennamen werden die Belegstellen in den Texten gelistet und automatisiert mit verschiedenen taxonomischen Datenbanken verlinkt. Aufgrund der dynamischen Erstellung dieses Registers können hier bislang noch keine Permalinks angeboten werden. Anstatt eines in XML-kodierten Werkregisters nutzt die *edition humboldt digital* eine Zotero-Bibliographie. Die Projektdatenbank des Vorhabens befindet sich in der Zotero-Gruppe „AvH-R“ unter <https://www.zotero.org/groups/avh-r>.

Datensatz "Chronologie": Die in dieser Edition vorgelegte Alexander von Humboldt-Chronologie basiert auf der Schrift:

ALEXANDER VON HUMBOLDT. Chronologische Übersicht über wichtige Daten seines Lebens. Bearbeitet von Kurt-R. Biermann, Ilse Jahn und Fritz G. Lange. 2., vermehrte u. berichtigte Auflage, bearbeitet von Kurt-R. Biermann unter Mitwirkung von Margot Faak und Peter

Honigmann. Berlin: Akademie-Verlag 1983 (Beiträge zur Alexander-von-Humboldt-Forschung, Bd. 1).

Die Chronologie wurde im Akademienvorhaben „Alexander-von-Humboldt-Forschung“ inhaltlich erweitert und auf eine elektronische Publikation umgestellt. Seitdem wurden zahlreiche Anregungen, Korrekturvorschläge, Ergänzungen etc. in die Chronologie aufgenommen. 2017 wurden alle Einträge der bisherigen Chronologie als TEI-XML-Datensätze neu angelegt und im Mai 2017 in der aktuell vorliegenden Fassung veröffentlicht. Personen- und Ortsnamen wurden in vielen Fällen bereits mit den jeweiligen Registereinträgen der *edition humboldt digital* verknüpft und werden derzeit weiter ausgezeichnet. Erstmals ist eine Volltextsuche über den gesamten Datenbestand der Alexander von Humboldt-Chronologie möglich. Ein entscheidender Vorteil der jetzt erfolgten Datenkuratierung von HTML zu XML ist die Einbindung externer Ressourcen.

Datenursprung

Wird der Datensatz selbst erzeugt oder nachgenutzt?

Datensatz "Bilddigitalisate":

- Erzeugt
- Nachgenutzt

Datensatz "Texte":

- Erzeugt
- Nachgenutzt

Datensatz "Register":

- Erzeugt
- Nachgenutzt

Datensatz "Chronologie":

- Erzeugt
- Nachgenutzt

Wer hat den Datensatz erzeugt?

Datensatz "Bilddigitalisate": Digitalisierung durch die bestandshaltenden Institutionen, vor allem die Staatsbibliothek zu Berlin-PK und die Biblioteka Jagiellońska in Krakau. Digitalisate aus weiteren Einrichtungen werden entweder nachgenutzt oder, wo diese noch nicht digitalisiert wurden, deren Digitalisierung durch das Akademienvorhaben *AvH-R* beauftragt.

Datensatz "Texte": Die Texte der historisch-kritischen Edition werden durch das Akademienvorhaben erarbeitet. Einige der Dokumente bzw. Forschungsdaten (z.B. aus Registern) wurden aus früheren Druck-Editionen retrodigitalisiert, und werden innerhalb der *ehd* grundlegend überarbeitet neu herausgegeben. Zusätzlich werden Metadaten zu Manuskripten aus dem Verbundkatalog Kalliope importiert und nach Bedarf ergänzt.

Datensatz "Register": Die Register zu Personen, Orten und Institutionen werden zum Großteil durch das Akademienvorhaben erstellt. Darüber hinaus wurden im Rahmen des Projekts die Registerangaben von 25 zwischen 1973 und 2016 veröffentlichten Editionen (Briefe, Dokumente, Tagebuch-Exzerpte) der Alexander von Humboldt-Forschung retrodigitalisiert und zusätzlich bereitgestellt.

Datensatz "Chronologie": Die Alexander von Humboldt-Chronologie basiert auf der Schrift:
ALEXANDER VON HUMBOLDT. Chronologische Übersicht über wichtige Daten seines Lebens. Bearbeitet von Kurt-R. Biermann, Ilse Jahn und Fritz G. Lange. 2., vermehrte u. berichtigte Auflage, bearbeitet von Kurt-R. Biermann unter Mitwirkung von Margot Faak und Peter Honigmann. Berlin: Akademie-Verlag 1983 (Beiträge zur Alexander-von-Humboldt-Forschung, Bd. 1).

Durch das Akademienvorhaben werden alle Einträge als TEI-XML-Datensätze neu angelegt, inhaltlich erweitert und veröffentlicht. Es werden außerdem externe Ressourcen in die Chronologie eingebunden: 1.) Briefe aus dieser Edition (Chronologische Anzeige und Volltextsuche), 2.) Unselbständige Schriften Alexander von Humboldts im „Deutschen Textarchiv“ (Chronologische Anzeige und Volltextsuche), 3.) Briefe von und an Humboldt aus „correspSearch“ (Chronologische Anzeige), 4.) Selbständige Schriften Alexander von Humboldts aus der Digitalisate Bibliographie „Humboldt Digital“ (Chronologische Anzeige).

Ist der nachgenutzte Datensatz bereits unter einer Adresse, PID oder URL verfügbar?

Datensatz "Bilddigitalisate": Ja: Für die meisten Digitalisate: Ja, siehe für SBB-PK und Biblioteka Jaiellońska Krakau (<https://humboldt.staatsbibliothek-berlin.de/werk/>), weitere hier aufzuführen, würde zu kleinteilig werden.

Datensatz "Texte": Nein

Datensatz "Register": Nein

Datensatz "Chronologie": Nein

Wenn nachgenutzt, unter welcher Lizenz wurden die Daten bereitgestellt?

Datensatz "Bilddigitalisate": unterschiedlich; AvH-R verweist daher für die Nachnutzung der Scans jeweils zurück auf die bestandshaltenden Institutionen (vgl. <https://edition-humboldt.de/impressum.xql?l=de> unter "Bilddigitalisate")

Datensatz "Texte": i.d.R. N/A, da die Daten i.d.R. vorab nicht in nachnutzbarer Form publiziert wurden.

Nachnutzung

Für welche Personen, Gruppen oder Institutionen könnte dieser Datensatz (für die Nachnutzung) von Interesse sein?

Datensatz "Bilddigitalisate": Alexander von Humboldt hat über einen Zeitraum von 70 Jahren eigene (und fremde) Reiseprojekte konzipiert, gefördert, durchgeführt, reiseliterarisch dargestellt und wissenschaftlich ausgewertet. Auf seinen eigenen Reisen – vor allem durch die Amerikas, Russland, Spanien, und Italien – führte er konsequent Reisetagebücher bzw. Reisejournale, die in einzigartiger Weise seine Arbeitsweise und Wissensorganisation dokumentieren. Diese Aufzeichnungen verbinden

Beschreibungen des Reiseverlaufs mit Messergebnissen, literarischen Reiseskizzen, wissenschaftlichen Essays, Laborprotokollen, Zeichnungen und Skizzen, Gutachten und Exzerpten, die für die Wissenschaftsgeschichte, Philologie, Geschichtswissenschaft, Kulturgeschichte, Anthropologie, Klimatologie, Geophysik, Vulkanologie, Religionsgeschichte, Kunstgeschichte, Ökonomie und nicht zuletzt Ökologie von höchster Aktualität sind.

Datensatz "Texte": Alexander von Humboldt hat über einen Zeitraum von 70 Jahren eigene (und fremde) Reiseprojekte konzipiert, gefördert, durchgeführt, reiseliterarisch dargestellt und wissenschaftlich ausgewertet. Auf seinen eigenen Reisen - vor allem durch die Americas, Russland, Spanien, und Italien - führte er konsequent Reisetagebücher bzw. Reisejournale, die in einzigartiger Weise seine Arbeitsweise und Wissensorganisation dokumentieren. Diese Aufzeichnungen verbinden Beschreibungen des Reiseverlaufs mit Messergebnissen, literarischen Reiseskizzen, wissenschaftlichen Essays, Laborprotokollen, Zeichnungen und Skizzen, Gutachten und Exzerpten, die für die Wissenschaftsgeschichte, Philologie, Geschichtswissenschaft, Kulturgeschichte, Anthropologie, Klimatologie, Geophysik, Vulkanologie, Religionsgeschichte, Kunstgeschichte, Ökonomie und nicht zuletzt Ökologie von höchster Aktualität sind. Über die API <https://edition-humboldt.de/about/index.xql?id=api&l=de> können die Daten flexibel nachgenutzt, weiter angereichert und mit anderen Datensätzen verknüpft werden (Linked Open Data).

Datensatz "Register": Die Register ermöglichen durch die Anreicherung mit Normdaten eine projektübergreifende Identifizierung der Personen, Orte, Institutionen etc. und geben den Nutzenden neue Einstiegsmöglichkeiten. Über die API <https://edition-humboldt.de/about/index.xql?id=api&l=de> können die Daten flexibel nachgenutzt, weiter angereichert und mit anderen Datensätzen verknüpft werden (Linked Open Data).

Datensatz "Chronologie": Diese Übersicht will nicht nur ein Instrument für die Humboldt-Forschung, sondern darüber hinaus den zahlreichen Humboldt-Verehrern ein Orientierungsmittel sein. Sie stellt viele falsche Daten in der Humboldt-Literatur richtig und korrigiert auch Irrtümer Humboldts selbst, die ihm, meist bei der Erwähnung lange zurückliegender Ereignisse, unterlaufen sind. Hingegen kann und will sie weder eine Humboldt-Biographie sein noch eine Humboldt-Bibliographie ersetzen. Die Chronologie ist als TEI-kodierter Datensatz noch nicht über die API des Vorhabens zugänglich, kann dies jedoch perspektivisch werden. Dann können die Daten flexibel nachgenutzt, weiter angereichert und mit anderen Datensätzen verknüpft werden (Linked Open Data). Bereits jetzt wird die Chronologie als Datensatz für den Twitter-Bot „AvHChrono“ (<https://twitter.com/avhchrono>) genutzt sowie als Grundlage für die in Entwicklung befindlichen „chronotopischen Zugänge“ zur *edition humboldt digital* (vgl. z.B. <https://www.avhumboldt.de/?p=18750>).

Reproduzierbarkeit

Ist der Datensatz reproduzierbar, d. h. ließe sich er sich, wenn er verloren ginge, erneut erstellen oder erheben?

Datensatz "Bilddigitalisate":

- ja, mit mäßigem, aber vertretbarem Aufwand
- nein bzw. nur mit unverhältnismäßig großem Aufwand

Datensatz "Texte":

- nein bzw. nur mit unverhältnismäßig großem Aufwand

- nein, die Daten sind per se nicht reproduzierbar

Datensatz "Register": nein bzw. nur mit unverhältnismäßig großem Aufwand

Datensatz "Chronologie": nein bzw. nur mit unverhältnismäßig großem Aufwand

Ggf. Erläuterung

Datensatz "Bilddigitalisate": Die Digitalisate könnten bei den bestandshaltenden Institutionen erneut angefragt werden. Allerdings wäre eine erneute Digitalisierung für einen Großteil der archivalischen Quellen nur „mit unverhältnismäßig großem Aufwand“ möglich, zu insgesamt innerhalb des Akademienprogramms nicht tragbaren Kosten. Einige Digitalisate ließen sich gar nicht reproduzieren, da der Zugang zu den Originalen nicht (mehr) gegeben ist, darunter beispielsweise das Russisch-Sibirische Reisetagebuch A. v. Humboldts (vgl. dazu <https://edition-humboldt.de/H0018418>).

Werkzeuge

Welche Software, Technologien oder Verfahren werden zur Erzeugung oder Erfassung der Daten genutzt?

Datensatz "Bilddigitalisate": Professionelle Digitalisierung in den bestandshaltenden Einrichtungen, daher kann das Vorhaben AvH-R hier keine Details zu den dabei eingesetzten Software, Technologien oder Verfahren geben.

Datensatz "Texte": Die Publikation *edition humboldt digital* wird von der Erfassung bis hin zur Publikation in ediarum (s. <https://www.ediarum.org/>) erstellt; die *edition humboldt print* basiert auf denselben TEI-XML-Daten (= Single-Source-Prinzip), die durch einen Satzdienstleister für die Herstellung der Printbände sowie der eBook-Fassung aufbereitet werden. Die digitale Arbeitsumgebung ediarum ist eine von der DH-Initiative TELOTA seit 2012 entwickelte Lösung, die es den WissenschaftlerInnen erlaubt, Transkriptionen von Manuskripten in TEI-konformem XML zu bearbeiten, mit einem Text- und Sachapparat zu versehen und anschließend im Druck und im Web zu veröffentlichen. Dabei besteht ediarum aus mehreren Softwarekomponenten, die zur Erarbeitung einer digitalen Edition entsprechend angepasst wurden: Zur Eingabe und Bearbeitung der Daten wird eine eigens angepasste Version des oXygen XML Author benutzt. Die Speicherung der Daten erfolgt in der freien XML-Datenbank eXistdb. Sie dient gleichzeitig – zusammen mit dem Webserver Jetty – auch als Basis der digitalen Edition, die mit XQuery, XSLT und XPath realisiert wurde. Die fertig edierten Texte werden mit Hilfe des „Cascaded Analysis Broker“ des Deutschen Textarchivs (DTA::CAB) linguistisch annotiert, um editionsintern eine lemma-basierte Suche zu ermöglichen, die auch historische Schreibweisen und andere Wortformen findet.

Datensatz "Register": Für die Erzeugung der Daten wird ediarum genutzt.

Datensatz "Chronologie": Für die Erzeugung der Daten wird ediarum genutzt.

Welche Software, Verfahren oder Technologien sind notwendig, um die Daten zu nutzen?

Datensatz "Bilddigitalisate": Die Bilder können über ein Bildbetrachtungsprogramm angesehen werden. Zur Anzeige der Faksimiles in der Online-Edition wird die am MPI für Wissenschaftsgeschichte entwickelte Software digilib eingesetzt.

Datensatz "Texte": Zur Nutzung der erzeugten TEI-XML Daten sind keine speziellen Anwendungen notwendig, ein einfacher Texteditor reicht aus. Für die Darstellung, wie vom Projekt vorgesehen, wird jedoch die Website mit den entsprechenden Layouteinstellungen benötigt.

Datensatz "Register": Zur Nutzung der erzeugten TEI-XML Daten sind keine speziellen Anwendungen notwendig, ein einfacher Texteditor reicht aus. Für die Darstellung, wie vom Projekt vorgesehen, wird jedoch die Website mit den entsprechenden Layouteinstellungen benötigt.

Datensatz "Chronologie": Zur Nutzung der erzeugten TEI-XML Daten sind keine speziellen Anwendungen notwendig, ein einfacher Texteditor reicht aus. Für die Darstellung, wie vom Projekt vorgesehen, wird jedoch die Website mit den entsprechenden Layouteinstellungen benötigt.

Wird die Dokumentation von ggf. zur Nutzung notwendiger Software benötigt, um die Daten zu nutzen?

Datensatz "Bilddigitalisate": Nein

Datensatz "Texte": Nein

Datensatz "Register": Nein

Datensatz "Chronologie": Nein

Datensicherung

Datengröße

Was ist die tatsächliche oder erwartete Größe des Datensatzes?

Datensatz "Bilddigitalisate": 20 GB bis 100 GB

Datensatz "Texte": weniger als ein GB

Datensatz "Register": weniger als ein GB

Datensatz "Chronologie": weniger als ein GB

Wie hoch ist die erwartete Erzeugungsrate der Daten pro Jahr?

Datensatz "Bilddigitalisate": 5-10 GB

Datensatz "Texte": wenige MB

Datensatz "Register": wenige MB

Datensatz "Chronologie": wenige MB

Wie viele Dateien wird der Datensatz letztlich umfassen?

Datensatz "Bilddigitalisate": 10000

Datensatz "Texte": 1500

Datensatz "Register": 15000

Datensatz "Chronologie": 2000

Wie viele Dateien werden pro Jahr hinzukommen?

Datensatz "Bilddigitalisate": 500-1000

Datensatz "Texte": 100

Datensatz "Register": 150-300

Datensatz "Chronologie": 25-50

Formate

In welchen Formaten liegen die Daten vor?

Datensatz "Bilddigitalisate":

- JPG, JPEG (Joint Photographic Experts Group)
- TIF, TIFF (Tagged Image File Format)

Datensatz "Texte":

- XML (Extensible Markup Language)
- HTM, HTML (Hypertext Markup Language)

Datensatz "Register":

- XML (Extensible Markup Language)
- HTM, HTML (Hypertext Markup Language)

Datensatz "Chronologie":

- XML (Extensible Markup Language)
- HTM, HTML (Hypertext Markup Language)

Versionierung

Werden verschiedene Versionen des Datensatzes erzeugt?

Datensatz "Bilddigitalisate": Nein

Datensatz "Texte": Ja (manuell)

Datensatz "Register": Ja (manuell)

Datensatz "Chronologie": Ja (manuell)

Welche Versionierungsstrategie wird für diesen Datensatz angewandt?

Datensatz "Texte": Jeder publizierte Zustand (1-2 mal im Jahr) wird von TELOTA versioniert und dauerhaft in der öffentlichen eXistdb-Instanz vorgehalten. In Zukunft ist von Seiten TELOTA die zusätzliche Vorhaltung der versionierten Daten in einem Git angedacht.

Datensatz "Register": Jeder publizierte Zustand (1-2 mal im Jahr) wird von TELOTA versioniert und dauerhaft in der öffentlichen eXistdb-Instanz vorgehalten. In Zukunft ist von Seiten TELOTA die zusätzliche Vorhaltung der versionierten Daten in einem Git angedacht.

Datensatz "Chronologie": Jeder publizierte Zustand (1-2 mal im Jahr) wird von TELOTA versioniert und dauerhaft in der öffentlichen eXistdb-Instanz vorgehalten. In Zukunft ist von Seiten TELOTA die zusätzliche Vorhaltung der versionierten Daten in einem Git angedacht.

Falls der Datensatz versioniert wird, welche Technologie bzw. welches Tool wird zur Versionierung verwendet?

Datensatz "Texte": Einfaches Kopieren

Datensatz "Register": Einfaches Kopieren

Datensatz "Chronologie": Einfaches Kopieren

Backup

Wie häufig wird dieses Backup ausgeführt?

Datensatz "Texte": täglich

Datensatz "Register": täglich

Datensatz "Chronologie": täglich

Um was für ein Backup handelt es sich?

Datensatz "Texte": volles Backup

Datensatz "Register": volles Backup

Datensatz "Chronologie": volles Backup

Welche Organisationseinheit ist für die Erstellung der Backups verantwortlich (z.B. das Projekt selbst, die IT-Abteilung)?

Datensatz "Bilddigitalisate": IT

Datensatz "Texte": TELOTA

Datensatz "Register": TELOTA

Datensatz "Chronologie": TELOTA

Wo befindet sich das Backup?

Datensatz "Texte":

- lokal
- außer Haus

Datensatz "Register":

- lokal

- außer Haus

Datensatz "Chronologie":

- lokal
- außer Haus

Wer initiiert das Backup?

Datensatz "Texte": automatisch

Datensatz "Register": automatisch

Datensatz "Chronologie": automatisch

Wie lange wird dieses Backup jeweils aufbewahrt?

Datensatz "Texte": andere

Datensatz "Register": andere

Datensatz "Chronologie": andere

Weitere Angaben

Datensatz "Texte": Autom. Backup der eXistdb, der alle vier Stunden inkrementell erstellt wird, einmal am Tag mit vollständigem Backup. Dieser Dump wird durch die Standard-BBAW-Routine für die Server, auf denen die eXistdb laufen, extern gesichert. Die Backups werden teilweise länger aufbewahrt, allerdings nach längerem Zeitraum nur noch ein Teil (z.B. nach einem Jahr: 1 Backup eines Monats).

Datensatz "Register": Autom. Backup der eXistdb, der alle vier Stunden inkrementell erstellt wird, einmal am Tag mit vollständigem Backup. Dieser Dump wird durch die Standard-BBAW-Routine für die Server, auf denen die eXistdb laufen, extern gesichert. Die Backups werden teilweise länger aufbewahrt, allerdings nach längerem Zeitraum nur noch ein Teil (z.B. nach einem Jahr: 1 Backup eines Monats).

Datensatz "Chronologie": Autom. Backup der eXistdb, der alle vier Stunden inkrementell erstellt wird, einmal am Tag mit vollständigem Backup. Dieser Dump wird durch die Standard-BBAW-Routine für die Server, auf denen die eXistdb laufen, extern (konserviert) gesichert. Die Backups werden teilweise länger aufbewahrt, allerdings nach längerem Zeitraum nur noch ein Teil (z.B. nach einem Jahr: 1 Backup eines Monats).

Datennutzung

Datenorganisation

Wo wird der Datensatz während des Projektes gespeichert?

Datensatz "Bilddigitalisate": digilib-Server

Datensatz "Texte": <https://xmledit.bbaw.de/ediarum/rest/db/projects/avhr>

Datensatz "Register": <https://xmledit.bbaw.de/ediarum/rest/db/projects/avhr>

Datensatz "Chronologie": <https://xmledit.bbaw.de/ediarum/rest/db/projects/avhr>

Gibt es projektinterne Richtlinien zur einheitlichen Ablage und Benennung der Daten?

Datensatz "Bilddigitalisate": Noch nicht

Datensatz "Texte": Ja

Datensatz "Register": Ja

Datensatz "Chronologie": Ja

Falls es eine Richtlinie gibt oder eine solche in Planung ist, bitte skizzieren Sie diese kurz und verlinken Sie ggf. zu einer ausführlicheren Dokumentation.

Datensatz "Texte": Die Arbeit im Akademienvorhaben folgt den Editionsrichtlinien der *edition humboldt digital*: <https://edition-humboldt.de/richtlinien/index.html>

Datensatz "Register": Die Arbeit im Akademienvorhaben folgt den Editionsrichtlinien der *edition humboldt digital*: <https://edition-humboldt.de/richtlinien/index.html>

Datensatz "Chronologie": Die Arbeit im Akademienvorhaben folgt den Editionsrichtlinien der *edition humboldt digital*: <https://edition-humboldt.de/richtlinien/index.html>

Datensicherheit

Welche Gruppen haben Zugriff auf den Datensatz?

Datensatz "Bilddigitalisate": Die Projektmitarbeitenden haben Zugriff auf die Digitalisate.

Datensatz "Texte": Das Mitarbeiten des Projekts können auf die Bearbeitungsinstanz zugreifen. Die Publikationsinstanz kann nur durch TELOTA bedient werden.

Datensatz "Register": Das Mitarbeiten des Projekts können auf die Bearbeitungsinstanz zugreifen. Die Publikationsinstanz kann nur durch TELOTA bedient werden.

Datensatz "Chronologie": Das Mitarbeiten des Projekts können auf die Bearbeitungsinstanz zugreifen. Die Publikationsinstanz kann nur durch TELOTA bedient werden.

Welche Maßnahmen zur Gewährleistung der Datensicherheit werden getroffen (z.B. Schutz vor unbefugtem Zugriff, Datenwiederherstellung, Übertragung sensibler Daten)?

Datensatz "Bilddigitalisate": Passwortschutz; Daten auf internen Laufwerken bzw. nubes-Cloud der BBAW.

Datensatz "Texte": Passwortschutz; personalisierte Installation des ediarum-Frameworks inkl. Zugriff auf die eXist-Datenbank durch TELOTA.

Datensatz "Register": Passwortschutz; personalisierte Installation des ediarum-Frameworks inkl. Zugriff auf die eXist-Datenbank durch TELOTA.

Datensatz "Chronologie": Passwortschutz; personalisierte Installation des ediarum-Frameworks inkl. Zugriff auf die eXist-Datenbank durch TELOTA.

Qualitätssicherung

Welche Maßnahmen zur inhaltlichen und technischen Qualitätssicherung werden für diesen Datensatz ergriffen?

Datensatz "Bilddigitalisate": Qualitätsmanagement nach intern abgestimmten Editionsstufen; Daten-Lektorat; Satzspiegel-Lektorat.

Datensatz "Texte": Inhaltliche Qualitätssicherung: ausführliche editorische und Kodierungs-Richtlinien; Kollation nach 4-Augen-Prinzip; Qualitätsmanagement nach intern abgestimmten Editionsstufen; Daten-Lektorat; Satzspiegel-Lektorat. technisch: XML-Schema

Datensatz "Register": Inhaltliche Qualitätssicherung: ausführliche editorische und Kodierungs-Richtlinien; Kollation nach 4-Augen-Prinzip; Qualitätsmanagement nach intern abgestimmten Editionsstufen; Daten-Lektorat; Satzspiegel-Lektorat. Technische Qualitätssicherung: XML-Schema

Datensatz "Chronologie": Inhaltliche Qualitätssicherung: ausführliche editorische und Kodierungs-Richtlinien; Qualitätsmanagement nach intern abgestimmten Editionsstufen; Daten-Lektorat; Satzspiegel-Lektorat. Technische Qualitätssicherung: XML-Schema

Metadaten und Referenzierung

Metadaten

Welche Metadatenstandards werden zur Beschreibung der Daten und Kontextinformation genutzt?

Datensatz "Texte":

- CMIF (Correspondence Metadata Interchange-Format)
- TEI P5 (Text Encoding Initiative)
- DTABf (DTA-Basisformat)
- Andere: DITA

Datensatz "Register": TEI P5 (Text Encoding Initiative)

Datensatz "Chronologie": TEI P5 (Text Encoding Initiative)

Welche kontrollierten Vokabulare werden zur Beschreibung der Daten und Kontextinformation genutzt?

Datensatz "Texte":

- GND (Gemeinsame Normdatei)
- GeoNames

Datensatz "Register":

- GND (Gemeinsame Normdatei)

- GeoNames
- Andere: Encyclopedia of Life, Tropicos - Missouri Botanical Garden, The International Plant Names Index (IPNI), englische Wikipedia, Catalogue of Life, Biodiversity Heritage Library, Global Biodiversity Information Facility

Datensatz "Chronologie":

- GND (Gemeinsame Normdatei)
- GeoNames

Weitere Angaben zu Metadatenstandards und kontrollierten Vokabularen:

Datensatz "Texte": Siehe dazu <https://edition-humboldt.de/H0016212>: „Die Metadaten und Texte der zu edierenden Schriften (Reisetagebücher, Dokumente und Briefe) werden nach den Richtlinien der Text Encoding Initiative (TEI) kodiert. Dabei wird weitgehend das Basisformat des Deutschen Textarchivs (DTABf) verwendet, das u.a. im Rahmen dieses Editionsvorhabens für die Kodierung von handschriftlichen Texten erweitert wurde. Für einige wenige projektspezifische Belange wurden Ergänzungen und Erweiterungen zum Tagset vorgenommen, die konform zu den TEI-Richtlinien sind. Diese Erweiterung betrifft insbesondere projektspezifische Anforderungen, die sich durch das Material ergeben, wie z.B. die Auszeichnung von Maßangaben. Alle Personen-, Orts- und Institutionsnamen sowie Literaturangaben in den Quellentexten (und ihren Metadaten) werden mit ihren jeweiligen Registereinträgen verknüpft.“

Begleittexte: Alle weiteren Texte, wie z.B. die einführenden Forschungsbeiträge zum Tagebuch-Fragment „Isle de Cube. Antilles en général“, wurden ebenfalls in TEI-XML gemäß dem Basisformat des Deutschen Textarchivs erfasst. Die Editionsrichtlinien des Vorhabens werden in DITA, aufbauend auf der ediarum.BASE-Doku gepflegt.

Datensatz "Register": Siehe dazu <https://edition-humboldt.de/H0016212>: „Die Metadaten und Texte der zu edierenden Schriften (Reisetagebücher, Dokumente und Briefe) werden nach den Richtlinien der Text Encoding Initiative (TEI) kodiert. Dabei wird weitgehend das Basisformat des Deutschen Textarchivs (DTABf) verwendet, das u.a. im Rahmen dieses Editionsvorhabens für die Kodierung von handschriftlichen Texten erweitert wurde. Für einige wenige projektspezifische Belange wurden Ergänzungen und Erweiterungen zum Tagset vorgenommen, die konform zu den TEI-Richtlinien sind. Diese Erweiterung betrifft insbesondere projektspezifische Anforderungen, die sich durch das Material ergeben, wie z.B. die Auszeichnung von Maßangaben.“ Ein Registereintrag soll darüber hinaus mit einer oder ggf. mehreren IDs aus den entsprechenden Normdateien versehen werden (sofern ein Eintrag in der Normdatei existiert), um eine projektübergreifende Identifizierung der Personen, Orte etc. zu ermöglichen.

- GeoNames & OpenStreetMap: Die im Register verzeichneten Orte und Institutionsstandorte sind in der Regel mit einer ID der freien geographischen Datenbank GeoNames versehen. So können die Orte projektübergreifend identifiziert werden, was die Nachnutzung der Daten erheblich vereinfacht. Darüber hinaus werden anhand der GeoNames-ID Koordinaten bezogen, auf deren Basis mit Hilfe des freien Kartendienstes OpenStreetMap eine entsprechende Karte bei den Orts- und Institutionseinträgen angezeigt werden kann.

- Gemeinsame Normdatei & BEACON: Anhand der in den Registereinträgen zu Personen notierten ID, die in der Regel der Gemeinsamen Normdatei (GND) entnommen wird, können weitere Daten über den GND-Webservice „Entity Facts“ bezogen werden. So können vor allem die (familiären) Beziehungen zwischen Personen im Register angezeigt werden. Die GND-ID erlaubt darüber hinaus, weitere Editionen, Lexika und Projekte zu verlinken, die im Themenbereich der *edition humboldt*

digital relevant sind. Hierfür werden die BEACON-Schnittstellen dieser Projekte genutzt. So ist es etwa möglich, thematische Überschneidungen mit anderen Projekten des Zentrums Preußen-Berlin an der BBAW über die Verbindungen in den Registerdaten zu identifizieren und der Forschung bereitzustellen (Beispiel Wilhelm von Humboldt oder Aloys Hirt).

- Globale Indizes für wissenschaftliche Namen: Im Pflanzenregister (siehe oben) werden verschiedene Webservices und APIs benutzt, um wissenschaftliche Pflanzennamen aus Texten der *edition humboldt digital* automatisiert mit passenden Einträgen aus taxonomischen Datenbanken zu verknüpfen. Mithilfe des Webservices Global Names Resolver verlinkt das Pflanzenregister wissenschaftliche Namen mit Einträgen in der Encyclopedia of Life, Tropicos - Missouri Botanical Garden, The International Plant Names Index (IPNI) oder der englischen Wikipedia. Darüber hinaus werden der Catalogue of Life, die Biodiversity Heritage Library und die Global Biodiversity Information Facility anhand ihrer eigenen Schnittstellen abgefragt und ebenfalls verknüpft. Weitere Datenbanken können, sofern sie über entsprechende technische Schnittstellen verfügen, in Zukunft ergänzt werden. Die Abfrage der verschiedenen Datenbanken und automatisierte Verlinkung erfolgt dabei auf Basis des wissenschaftlichen Pflanzennamens, der im edierten Text entsprechend kodiert und ggf. normalisiert wird.

Datensatz "Chronologie": Die Chronologie wird als eigenständiger Datensatz (bzw. ein Dokument je Jahr) im TEI-XML-Format (teiCorpus) kodiert.

Persistente Identifikatoren (PIDs)

Sollen für diesen Datensatz persistente Identifikatoren (PIDs) genutzt werden?

Datensatz "Bilddigitalisate": Nein

Datensatz "Texte": Ja

Datensatz "Register": Ja

Datensatz "Chronologie": Ja

Welches System von persistenten Identifikatoren soll genutzt werden?

Datensatz "Texte": Permalink (permanente URLs im Webaufttritt)

Datensatz "Register": Permalink (permanente URLs im Webaufttritt)

Datensatz "Chronologie": Permalink (permanente URLs im Webaufttritt)

Welche (Sub-)Entitäten / Untereinheiten sollten sinnvollerweise eigene Identifikatoren erhalten? Welche dieser Identifikatoren sollten dauerhaft und zitierfähig sein?

Datensatz "Texte": Die in *edition humboldt digital* bereitgestellten Texte und Daten sind mit Permalinks versehen, die auf die jeweilige Version referenzieren, z.B.: <http://edition-humboldt.de/v1/H0002656>. Alle Versionen werden von der BBAW zum Abruf vorgehalten. Darüber hinaus sind die einzelnen Texte und Datensätze auch mit einer kanonischen URL versehen, die stets auf die aktuellste Version weiterleitet. Dabei entfällt einfach der Versionshinweis im Pfad, z.B. <http://edition-humboldt.de/H0002656>.

Datensatz "Register": Jeder Registereintrag verfügt über eine eigene, eindeutige und permanente Identifizierungsnummer.

Datensatz "Chronologie": Die in *edition humboldt digital* bereitgestellten Texte und Daten (einschließlich der Chronologie-Einträge) sind mit Permalinks versehen, die auf die jeweilige Version referenzieren, z.B.: <https://edition-humboldt.de/v1/H0014942>. Alle Versionen werden von der BBAW zum Abruf vorgehalten. Darüber hinaus sind die einzelnen Texte und Datensätze auch mit einer kanonischen URL versehen, die stets auf die aktuellste Version weiterleitet. Dabei entfällt einfach der Versionshinweis im Pfad, z.B. <https://edition-humboldt.de/H0014942>.

Rechtliche und ethische Fragen

Recht allgemein

Welche rechtlichen Gegebenheiten müssen beachtet werden?

- Die Digitalisate können nur mit Genehmigung der bestandshaltenden Institutionen und unter deren gewählten Lizenz veröffentlicht werden.
- Alle Texte und TEI-XML-Daten der *edition humboldt digital* bzw. der Website <https://edition-humboldt.de> können – soweit nicht anders vermerkt – unter den Bedingungen der Creative Commons-Lizenz CC BY-SA 4.0 nachgenutzt werden.
- Die Print- bzw. eBook-Edition erscheint im Springer-Verlag: <https://www.springer.com/series/16345?detailsPage=titles> (es bestehen keine rechtlichen Einschränkungen für die Onlinepublikation, da diese der Print- bzw. eBook-Edition vorausgeht (digital first Prinzip) und keine sog. moving wall vereinbart wurde).

Speicherung, Publikation und Langzeitarchivierung

Langzeitarchivierung

Muss dieser Datensatz langfristig aufbewahrt werden?

Datensatz "Bilddigitalisate": Ja

Datensatz "Texte": Ja

Datensatz "Register": Ja

Datensatz "Chronologie": Ja

Aus welchen Gründen müssen die Daten langfristig aufbewahrt werden?

Datensatz "Bilddigitalisate":

- Grundlage einer Publikation / Nachweis guter wissenschaftlicher Praxis
- Nachnutzung in Folgeprojekten oder durch andere

Datensatz "Texte":

- Grundlage einer Publikation / Nachweis guter wissenschaftlicher Praxis

- Nachnutzung in Folgeprojekten oder durch andere
- Dokumentation aufgrund gesellschaftlicher Relevanz
- Selbstverpflichtung

Datensatz "Register":

- Grundlage einer Publikation / Nachweis guter wissenschaftlicher Praxis
- Nachnutzung in Folgeprojekten oder durch andere
- Dokumentation aufgrund gesellschaftlicher Relevanz
- Selbstverpflichtung

Datensatz "Chronologie":

- Grundlage einer Publikation / Nachweis guter wissenschaftlicher Praxis
- Nachnutzung in Folgeprojekten oder durch andere
- Dokumentation aufgrund gesellschaftlicher Relevanz
- Selbstverpflichtung

Wie lange müssen die Daten aufbewahrt werden?

Datensatz "Bilddigitalisate": für immer

Datensatz "Texte": für immer

Datensatz "Register": für immer

Datensatz "Chronologie": für immer

Wo werden die Daten (einschließlich Metadaten, Dokumentation und ggf. relevantem Code bzw. relevanter Software) nach Projektende gespeichert bzw. archiviert?

Datensatz "Texte": Eigene Institution: GitLab (BBAW-Instanz)

Datensatz "Register": Eigene Institution: GitLab (BBAW-Instanz)

Datensatz "Chronologie": Eigene Institution: GitLab (BBAW-Instanz)

Wurde mit dem Repository oder Datenzentrum bereits eine angemessene Lösung zur Übergabe besprochen?

Datensatz "Texte": Nein

Datensatz "Register": Nein

Datensatz "Chronologie": Nein

Wann werden die Daten archiviert?

Datensatz "Texte": 31. Dezember 2032

Datensatz "Register": 31. Dezember 2032

Datensatz "Chronologie": 31. Dezember 2032

Publikation der Daten

Soll dieser Datensatz veröffentlicht oder geteilt werden?

Datensatz "Bilddigitalisate": Ja, extern für alle

Datensatz "Texte": Ja, extern für alle

Datensatz "Register": Ja, extern für alle

Datensatz "Chronologie": Ja, extern für alle

Wenn ja, unter welchen Nutzungsbedingungen oder welcher Lizenz sollen die Daten veröffentlicht bzw. geteilt werden?

Datensatz "Bilddigitalisate": Andere: Die Digitalisate werden unter der Lizenz der Datengebenden bereitgestellt.

Datensatz "Texte":

- Namensnennung (BY)
- Weitergabe unter gleichen Bedingungen (SA)

Datensatz "Chronologie":

- Namensnennung (BY)
- Weitergabe unter gleichen Bedingungen (SA)

Wo werden die Daten (einschließlich Metadaten, Dokumentation und ggf. relevantem Code bzw. relevanter Software) publiziert?

Datensatz "Bilddigitalisate": Eigene Institution: *edition humboldt digital* (<https://edition-humboldt.de>)

Datensatz "Texte":

- Eigene Institution: Volltexte in *edition humboldt digital* mit API (TEI-XML-Schnittstelle), Metadaten über mehrere Schnittstellen (OAI-PMH, CMIF und BEACON) angeboten
- Generisches Datenzentrum: Zenodo

Datensatz "Register":

- Eigene Institution: *edition humboldt digital* (<https://edition-humboldt.de/register/index.xql?!=de>)
- Generisches Datenzentrum: Zenodo

Datensatz "Chronologie":

- Eigene Institution: *edition humboldt digital* (<https://edition-humboldt.de/redirect.xql?id=X0000001>)
- Generisches Datenzentrum: Zenodo

Wann werden die Daten publiziert?

Datensatz "Texte": Die Erstellung der Datenbestände der *edition humboldt digital* ist ein fortlaufender Prozess. Umfang und Genauigkeit der Daten wachsen mit dem Voranschreiten des Vorhabens. Neben den editierten Texten sind auch die Editionsrichtlinien der *edition humboldt digital* online unter der Lizenz CC BY-SA 4.0 veröffentlicht.

Datensatz "Register": Die Erstellung der Datenbestände der *edition humboldt digital* ist ein fortlaufender Prozess. Umfang und Genauigkeit der Daten wachsen mit dem Voranschreiten des Vorhabens.

Sollen die Daten erst nach Ablauf einer Sperrfrist zugänglich gemacht werden?

Datensatz "Texte": Nein.

Datensatz "Register": Nein.

Datensatz "Chronologie": Nein.

Wie lang sollen die Daten nach Projektende (nach)nutzbar sein?

Datensatz "Bilddigitalisate": für immer

Datensatz "Texte": für immer

Datensatz "Register": für immer

Datensatz "Chronologie": für immer