

Aus der Praxis zur digitalen Archivierung in Wien und Berlin.

11. Tagung des Arbeitskreises der Kirchlichen Hochschulbibliotheken
(AKKH)

19.02.2020 - Martina Trognitz, ACDH-CH, ÖAW
martina.trognitz@oeaw.ac.at

Zu mir

- Computerlinguistik und Klassische Archäologie in Heidelberg
- 2012-2017: Projekt IANUS am DAI in Berlin
 - <https://www.ianus-fdz.de/>
 - <https://www.ianus-fdz.de/it-empfehlungen/>
- Seit 2017 am ACDH-CH der ÖAW in Wien
 - ARCHE: <https://arche.acdh.oeaw.ac.at>

Archiv & Archivierung

- *«Ein Archiv (lat. archivum ‚Aktenschrank‘; aus altgr. ἀρχεῖον archeíon ‚Amtsgebäude‘) ist eine Institution oder Organisationseinheit, in der Archivgut zeitlich unbegrenzt im Rahmen der Zuständigkeit des Archivs oder des jeweiligen Sammlungsschwerpunktes aufbewahrt, benutzbar gemacht und erhalten wird (Archivierung).»* » [[Wikipedia, Archiv](#)]

Der Datenlebenszyklus



ianus-fdz.de/it-empfehlungen/lebenszyklus
ukdataservice.ac.uk/manage-data/lifecycle

Archiv & Archivierung

- Analog
 - Objekte
 - Dokumentationsmaterial (Daten)

 - Informationen nur vor Ort einsehbar
 - Online Findbücher

 - Physischer Platz und spezielle klimatische Bedingungen

• Analog

- Objekte
- Dokumentationsmaterial (Daten)

- Informationen nur vor Ort einsehbar
- Online Findbücher

- Physischer Platz und spezielle klimatische Bedingungen

• Digital

- Dokumentationsmaterial (Daten)
 - born digital
 - digitalisiert

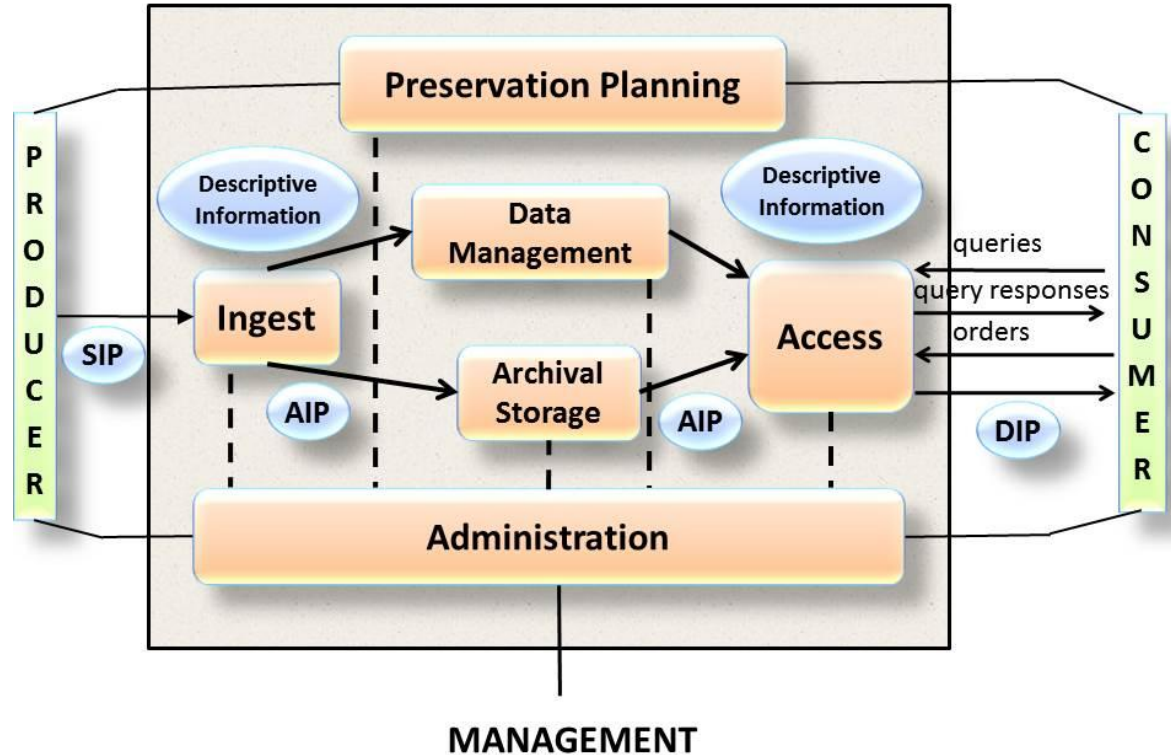
- Informationen ortsunabhängig abrufbar
- Online Kataloge

- Redundanter Speicherplatz und spezielle Workflows

- OAI Referenzmodell
<http://www.oais.info>

OAIS Referenzmodell

- Open Archival Information System
- [ISO 14721:2012](#)
- [Magenta Book](#)
- SIP, AIP, DIP



Digitale Archive

- Fachübergreifend
 - [DANS in Den Haag \(NL\)](#)
 - [ARCHE in Wien \(AT\)](#)
 - [Swedish National Data Service \(SWE\)](#)
 - [Zenodo \(CERN\)](#)
 - ...
- Spezialisierung ‘Archäologie’
 - [Archaeology Data Service in York \(UK\)](#)
 - [the Digital Archaeological Record](#) & [Open Context](#) (USA)
 - [Archaeological Information System of the Czech Republic \(CZ\)](#)
 - ...

Digitale Archive

- Fachübergreifend
 - [DANS in Den Haag \(NL\)](#)
 - [ARCHE in Wien \(AT\)](#)
 - [Swedish National Data Service \(SWE\)](#)
 - [Zenodo \(CERN\)](#)
 - ...
- Spezialisierung 'Archäologie'
 - [Archaeology Data Service in York \(UK\)](#)
 - [the Digital Archaeological Record](#) & [Open Context](#) (USA)
 - [Archaeological Information System of the Czech Republic \(CZ\)](#)
 - ...

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

OpenDOAR

<http://v2.sherpa.ac.uk/opensoar>

ROAR

<http://roar.eprints.org>

Was also ist digitale Archivierung?



[IMGFLIP]

Was also ist digitale Archivierung?

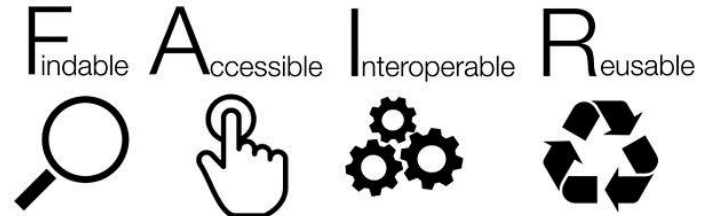
- Aufbewahrung von Materialien in digitalen Formaten, um:
 - sie zu lesen
 - sie zu öffnen
 - damit zu arbeiten
- Idealerweise langfristig (> 10 Jahre)
- Verschiedene Ansätze
 - Migration
 - Emulation
- Außerdem Hilfsmittel, wie Metadaten, um das Material:
 - zu finden
 - zu lokalisieren
 - abzurufen

Beispiele

- IANUS
 - [Reflectance Transformation Imaging \(RTI\) im Projekt Tayma](#) (22 775 Objekte)
DOI: 10.13149/001.pi46vf-z
- ADS
 - [Whitechapel Station \(Crossrail XSH10\)](#) (123 Objekte)
DOI: 10.5284/1055118
- ARCHE
 - [Glaser Abklatsche \(2015-2016\)](#) (4271 Objekte)
Handle: 21.11115/0000-000B-C63A-5
- DANS
 - [Mesopotamian cylinder seals: description and photographs of five seals from the NINO collection](#) (1072 Objekte)
DOI: 10.17026/dans-x2d-cexg

Gemeinsamkeiten

- Ein Stapel Dateien
 - In ausgewählten Formaten
- Metadaten
 - Unterschiedliche Attribute & Lizenzangabe
- Persistente URI
- Vorbereitender Arbeitsablauf
- Redundante Datenspeicherung



Konkrete Aufgaben

- Ein Stapel Dateien
 - In ausgewählten Formaten
- Metadaten
 - Unterschiedliche Attribute & Lizenzangabe
- Persistente URI
- Vorbereitender Arbeitsablauf
- Redundante Datenspeicherung
- Formatempfehlungen
 - [ADS](#), [IANUS](#), [ARCHE](#) ...
- Metadatenschemata
 - [Dublin Core](#), [DataCite](#), [ADeX](#), [IANUS](#), [ARCHE](#) ...
- Handle, DOI, URN, ARK
 - IANUS: [10.13149/000.ttfe5f-0](#)
- Datenmanagementplan
- Rechenzentren

- Forschungsdatenzentrum
Archäologie & Altertumswissenschaften
<https://www.ianus-fdz.de/>
- DFG Projekt
 - 3 Jahre Planung
 - 3 Jahre Umsetzung
- Ergebnisse
 - <https://www.ianus-fdz.de/projects/ergebnisse/wiki>
 - Fach- und Organisationskonzept
 - IT-Empfehlungen
 - Vor-Ort-Begehungen
 - Expertisen (technisch & rechtlich)



IANUS - Technisches System

- Eigene Entwicklung (Java)
- Angelehnt an Archivematica
- Weboberfläche mit Cake PHP
 - <http://datenportal.ianus-fdz.de>
- Verfügbar auf GitHub
 - <https://github.com/dainst/IANUS-Archive-Management-System>



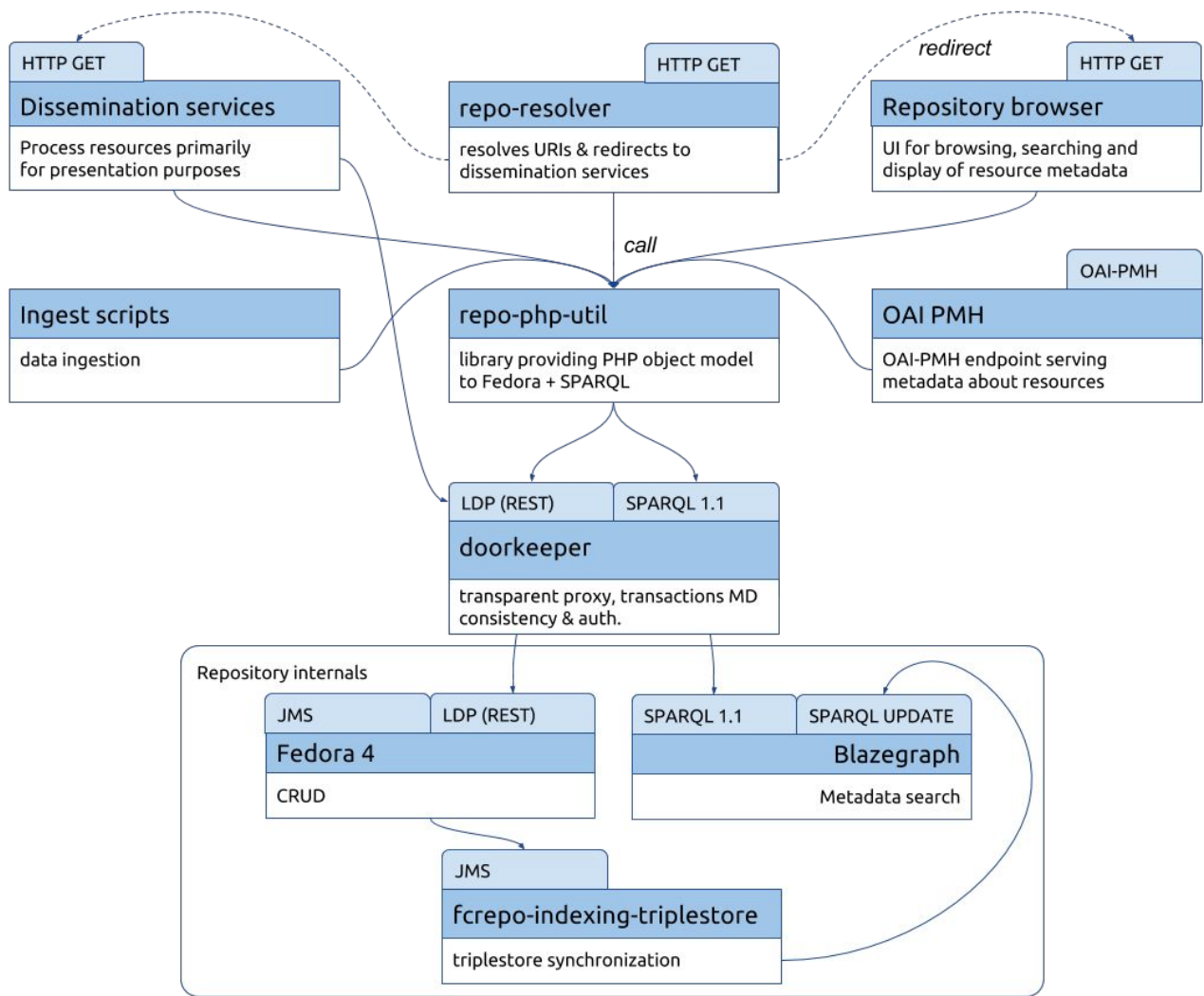
- A Resource Centre for the HumanitiEs
<https://arche.acdh.oeaw.ac.at>
- Nachfolger von CLARIN Centre Vienna / Language Resources Portal (CCV/LRP)
 - 2014-2017
- CoreTrustSeal (2018)
 - <https://www.coretrustseal.org/wp-content/uploads/2018/03/ARCHE.pdf>



ARCHE - Technisches System

- Fedora 4 mit Blazegraph für RDF-Tripel
- PHP Libraries für Zwischenschichten
- Weboberfläche mit Drupal
- <https://arche.acdh.oeaw.ac.at/browser/technical-setup>
- <https://arche.acdh.oeaw.ac.at/browser/api-access>





Repositorien- & Archivsoftware

- Phaidra (Fedora, Java)
 - <https://datamanagement.univie.ac.at>
- Fedora (Java)
 - <https://duraspace.org/fedora>
- DSpace (Java)
 - <https://duraspace.org/dspace>
- Dataverse (Java)
 - <https://dataverse.org>
- ckan (Python)
 - <https://ckan.org>
- Invenio (Python)
 - <https://invenio-software.org>

Repositorien- & Archivsoftware

- Islandora (Fedora mit einer GUI)
 - <https://islandora.ca>
- Archivematica (Java)
 - <https://www.archivematica.org>
- AtOm (Archivbeschreibung, Java)
 - <https://www.accesstomemory.org>
- Preservica
 - <https://preservica.com>
- CyVerse (Tabellarische Daten)
 - <https://www.cyverse.org>
- Omeka (Online Sammlungen, PHP)
 - <https://omeka.org>

Repositorien- & Archivsoftware

- Islandora (Fedora mit einer GUI)
 - <https://islandora.ca>
- Archivematica (Java)
 - <https://www.archivematica.org>
- AtOm (Archivbeschreibung, Java)
 - <https://www.accesstomemory.org>
- Preservica
 - <https://preservica.com>
- CyVerse (Tabellarische Daten)
 - <https://www.cyverse.org>
- Omeka (Online Sammlungen, PHP)
 - <https://omeka.org>

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

OpenDOAR

<http://v2.sherpa.ac.uk/opensoar>

ROAR

<http://roar.eprints.org>

→ Filtern nach Software

Merksätze

- Archivierung beinhaltet mehr als nur Software
 - Sammlungsstrategie, Vorgaben, Kuratierung
 - Rechtlicher Rahmen
 - Digitale Archivierung ist ein Prozess
- Metadaten und Dokumentation ist wichtig
 - Auffindbarkeit, Verständlichkeit & Nachnutzung
- Maschinenlesbarkeit ist wichtig
 - FAIR Data Principles
- Daten sollten möglichst offen zugänglich sein
- Daten sollten eine Lizenz haben
 - Information über die Nachnutzungsmöglichkeiten und -Pflichten, empfohlen: CC-BY



Fragen?

Martina Trognitz
ACDH, ÖAW

martina.trognitz@oeaw.ac.at

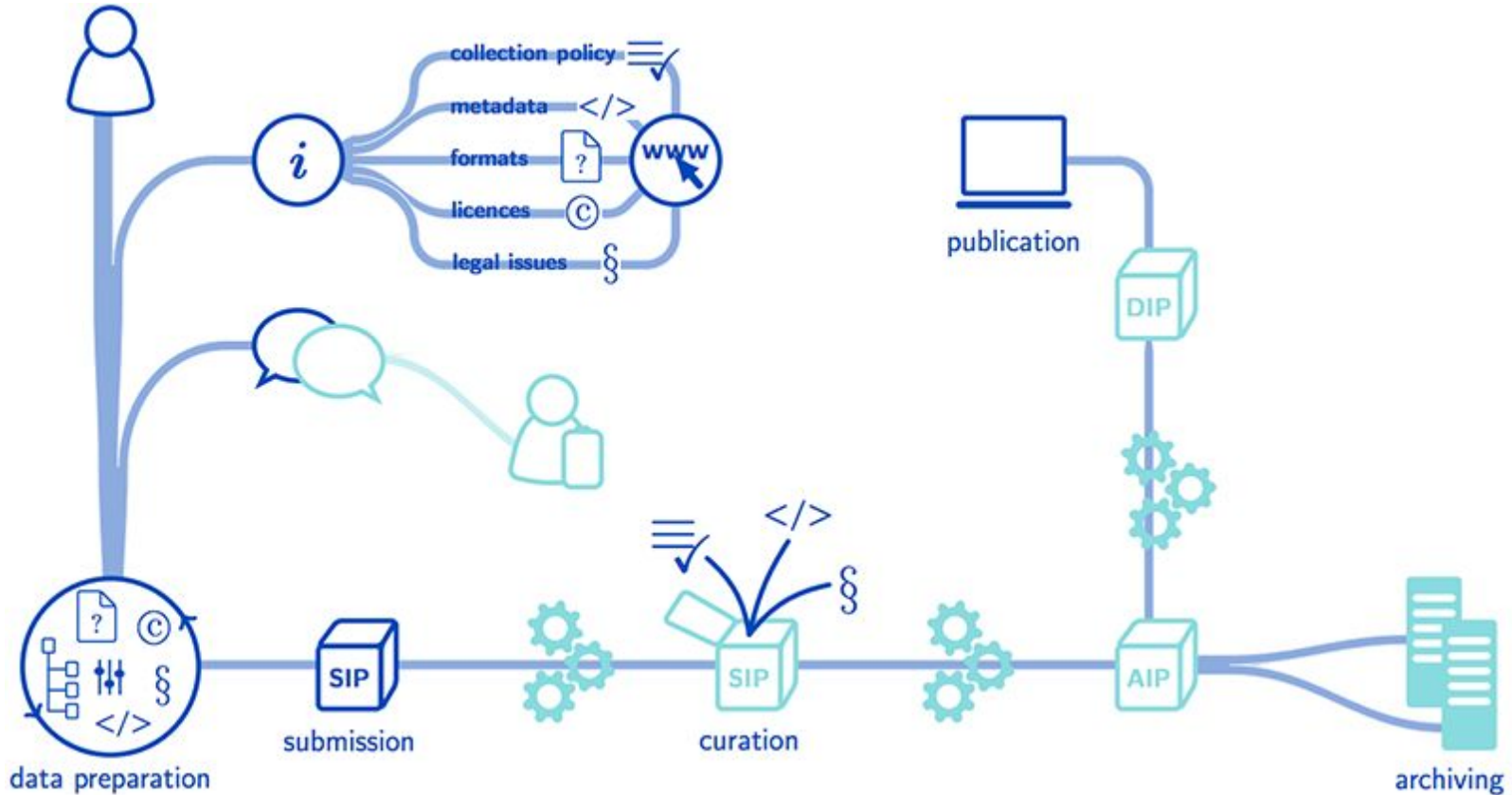
Die Logos stammen von den jeweiligen Projektwebseiten
Quelle und, falls vorhanden, Lizenz von externen Abbildungen sind angegeben



Der übrige Inhalt ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).

Zusatzmaterial - Kuratierungsabläufe

Wie kommen Daten in ARCHE?

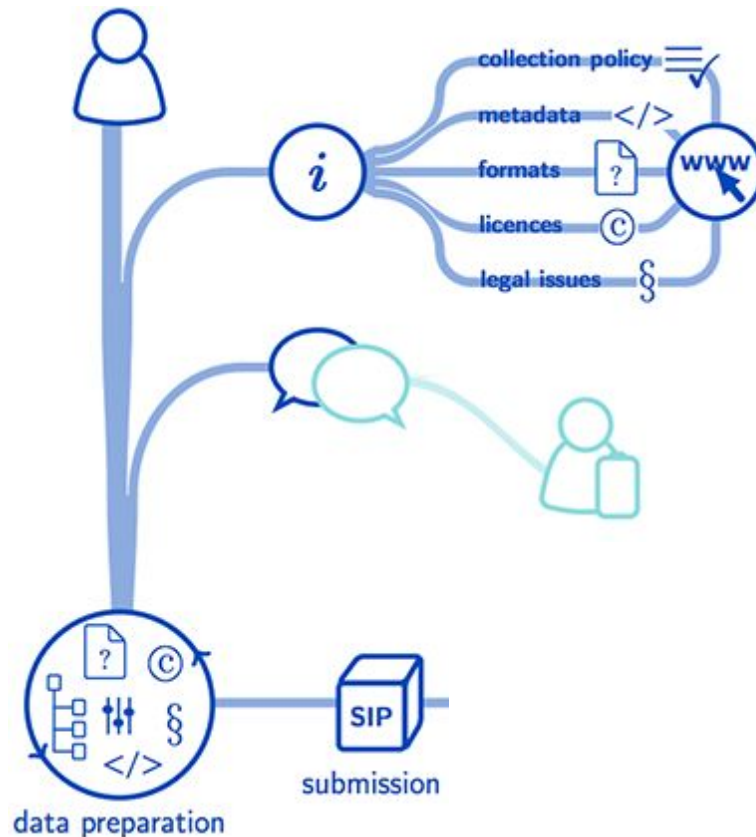


Wie kommen Daten in ARCHE?

- Geregelter Ablauf
 - Vorbereitung auf Datengeberseite
 - Kuratierungsprozess bei ARCHE
 - <https://www.youtube.com/watch?v=mmRfc7fNNSA>

Datengeberseite

- Informationen einholen
- Kontaktaufnahme
- Daten vorbereiten
- Daten übermitteln



[www arche.acdh.oew.ac.at/browser/formats-filenames-and-metadata](http://www.arche.acdh.oew.ac.at/browser/formats-filenames-and-metadata)

- ARCHE speichert
 - Plain text, TextDokumente, PDF-Dokumente, Bilder, Tabellen, Datenbanken, Audio, Video, 3D-Daten, GIS und viel mehr

- ARCHE speichert
 - Plain text, TextDokumente, PDF-Dokumente, Bilder, Tabellen, Datenbanken, Audio, Video, 3D-Daten, GIS und viel mehr
- ... solange die Dateien in Formaten vorliegen, welche für die langfristige Speicherung geeignet sind
- ... und die Namen den Benennungsregeln von ARCHE entsprechen
a-Z, 0-9, -, _

- Warum?
 - verstehen, auffinden und teilen von Daten
 - Wie? Wer? Was?
 - FAIR Data Principles
(Findable, Accessible, Interoperable & Reusable)
- Herausforderungen
 - Heterogenität der Daten
 - Unterstützung mehrerer Metadatenschemata
- Metadaten können über OAI-PMH *geharvested* werden



Wie kommen Metadaten & Daten zusammen?

- ARCHE Metadatenschema

- arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata
- github.com/acdh-oeaw/repo-schema
- Sammlung, Ressource
- Projekt, Personen, Organisationen, Orte, Publikationen

- Datengeberseite

- Tabellen
- [MetaDataEditor](#)

- Kuratierungsseite

- Skripte bei Bedarf

[www arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata](http://www.arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata)

Ein Praxisbeispiel

- Bilder
- CDROM
- KATALOG
- AbbildungsLISTE.doc
- AbbildungstextT1.doc
- AbbildungstextT1_KORR2.doc
- AbbildungstextT2.doc
- BefundLISTE.doc
- BefundLISTE_m_Abb.Nr.doc
- BefundLISTE_NEU.doc
- CecheVEL.doc
- CECHKeramikT1_KORR1.DOC
- CECHKeramikT1_KORR2.doc
- CECHKeramikT2_KORR1.DOC
- CechTiteleiNEU.doc
- Grabungsschnitte für Brigitte Cech.zip
- KatalogLISTE.doc
- Label-Aufdruck.doc
- MPKCechUmschlag.doc
- TafelLISTE.doc

• Dateiformate

- doc -> PDF/A
- zip -> entpacken

• Dateinamen

- Leerzeichen
- Sonderzeichen
- Können sprechender gestaltet werden

• Auswahl der Dateien

- KORR1, KORR2 -> eine Version reicht