

■ ONE SCHEMA TO RULE THEM ALL. THE INNER WORKINGS OF THE DIGITAL ARCHIVE ARCHE

by Martina Trognitz & Matej Ďurčo

Abstract: A Resource Centre for the HumanitiEs (ARCHE) is a digital archive provided by the Austrian Centre for Digital Humanities, which is part of the Austrian Academy of Sciences. The archive welcomes data from all humanities fields and offers persistent hosting as well as dissemination of resources. The technical setup of ARCHE is described, focusing on the bespoke metadata schema that was created to help in finding, accessing, and reusing data. Beyond that the schema binds together the different system components, contains mappings to other well-known schemas, provides multilingual labels for GUI display, and is used to generate a metadata form.

Keywords: archiving; metadata; metadata schema; system architecture

EIN SCHEMA SIE ALLE ZU BINDEN. DAS INNENLEBEN DES DIGITALEN ARCHIVS ARCHE

Zusammenfassung: A Resource Centre for the HumanitiEs (ARCHE) ist ein digitales Archiv, das vom Austrian Centre for Digital Humanities an der Österreichischen Akademie der Wissenschaften betrieben wird. Das Archiv speichert Daten aus dem gesamten Bereich der Geisteswissenschaften und bietet eine persistente Aufbewahrung und Bereitstellung von Ressourcen. Es wird der technische Aufbau von ARCHE beschrieben, wobei das Augenmerk auf das für die Auffindbarkeit, den Zugriff und die Nachnutzung von Daten maßgeschneiderte Metadatenchema gelegt wird. Darüber hinaus verbindet das Schema die verschiedenen Systemkomponenten, enthält Mappings auf andere wohlbekannte Schemata, stellt mehrsprachige Label für die GUI-Ansicht bereit und wird verwendet, um ein Metadatenformular zu generieren.

Schlüsselwörter: Archivierung; Metadaten; Metadatenchema; Systemarchitektur

DOI: <http://doi.org/10.31263/voebm.v71i1.1979>



Dieses Werk ist lizenziert unter einer
Creative-Commons-Lizenz Namensnennung 4.0 International

Content

1. *About ARCHE*
2. *Definition of Terms*
3. *System Architecture*
4. *Metadata Schema*
5. *Outlook*

1. About ARCHE

ARCHE (A Resource Centre for the HumanitiEs) is a digital archive that offers stable and persistent hosting as well as the dissemination of digital research data and resources for the Austrian humanities community. It is a service provided by the Austrian Centre for Digital Humanities at the Austrian Academy of Sciences in Vienna (ACDH-OeAW).

ARCHE is the successor of a repository service established in 2014 as the CLARIN Centre Vienna / Language Resources Portal (CCV/LRP). The mission of CCV/LRP was to provide depositing services and easy and sustainable access to digital language resources created in Austria. ARCHE replaced CCV/LRP in 2017 and extends its mission by offering advanced and reliable data management and a depositing service open to a broader range of humanities fields in Austria.

The archive implements the tasks and functions of the Open Archival Information System reference model.¹

All mandatory responsibilities described in the OAIS reference model² are already implemented in ARCHE. Negotiations for appropriate information that take place during the deposition process are based on a dedicated Collection Policy³ and further auxiliary documentation detailing the requirements⁴ for deposition of research data. A Deposition Agreement,⁵ ensures that ARCHE obtains sufficient control of the information for long term preservation. The designated community is likewise specified in the Collection Policy. The independent understandability of data is ensured by a set of metadata requirements and data quality checks. The preservation procedures to follow are already documented internally and will be formalised and made publicly available in a Preservation Policy. Lastly, the preserved information is already disseminated via ARCHE's web interface.

The preservation strategy employed by ARCHE is performing migration of formats as opposed to providing software emulation.⁶ To enhance sustainability, the use of open access and open data policies is promoted,

including the application of the FAIR Data Principles⁷ to provide Findable, Accessible, Interoperable and Reusable data and metadata.

2. Definition of Terms

This section provides basic definition of concepts related to semantic web technologies for readers unfamiliar with it. The semantic web⁸ is a part of the World Wide Web that makes use of a range of standards specified by the World Wide Web Consortium (W3C). These standards help in sharing and linking data, as well as making it machine readable and processible.

To represent data the Resource Description Framework (RDF)⁹ can be used. Information is then represented in subject-predicate-object triples, which in turn are called graphs. An example for an RDF triple could be “photo-wasTakenIn-Vienna”, which simply describes that a the photograph described with the triple was taken in Vienna. Another triple could be “Vienna-isCapitalOf-Austria”, which states that Vienna is the capital of Austria.

RDF triples can be stored in static files in e.g. RDF/XML¹⁰ or Turtle (ttl)¹¹ format. In order to not only store triples, but to also be able to retrieve and query them, a specialised kind of database was developed: the triplestore. Information from a triplestore can be queried using a query language, such as SPARQL.¹²

To avoid having triples such as “photo-isCapitalOf-Vienna” an ontology can be used. An ontology contains rules that describe how subjects, predicates and objects can be combined. For example a rule to avoid the aforementioned triple could state that the predicate “isCapitalOf” can only be used to combine subjects and predicates that are both geographic entities. The W3C provides a standard for representing ontologies: the Web Ontology Language (OWL).¹³ In this article the term “metadata schema” is synonymous with ontology.

In order to be able to distinguish between subjects, predicates or objects that have the same name (e. g. distinguish between Vienna, the city in Austria, and Vienna, a village in Canada), a controlled vocabulary can be used. This vocabulary may contain different terms (e. g. a list of cities) along with their respective descriptions (e. g. coordinates) and alternative names (e. g. city names in different languages). For this means the Simple Knowledge Organization System (SKOS)¹⁴ is recommended by the W3C.

3 System Architecture

The system runs in a Docker¹⁵ environment on one of the virtual machines hosted on dedicated servers maintained by the computing centre of

the Austrian Academy of Sciences. It is based on the open-source repository software Fedora Commons version 4¹⁶ which provides a sound technological basis for implementing the Open Archival Information System (OAIS) reference model. Fedora Commons performs fine-grained versioning of digital objects and allows for exporting, archiving and migrating of digital objects in XML-based formats. The technical setup of ARCHE is depicted in Figure 1.

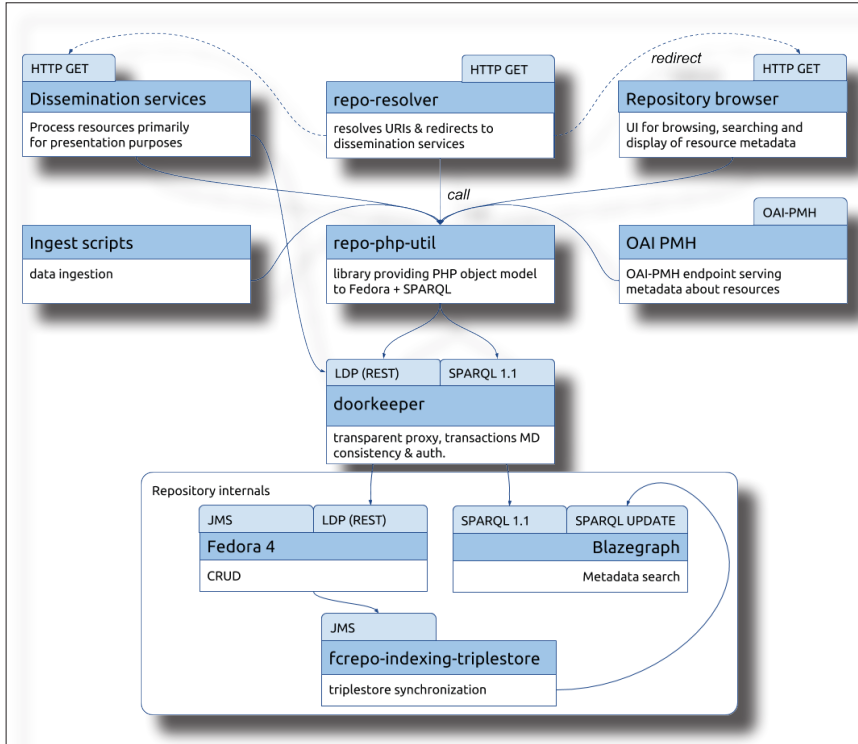


Fig. 1: The technical setup of ARCHE

Fedora is synchronised with Blazegraph,¹⁷ a graph database and triplestore we use for indexing and querying the metadata using SPARQL queries.¹⁸ The synchronisation is managed by the “Fedora Indexing Service (Triplestore)”¹⁹, a plugin for Fedora 4.

The access to Fedora and to the triplestore is guarded by a dedicated self-developed component, the doorkeeper, which allows to apply custom business logic, ensure the validity of the ingested data, make sure that the

resources and the metadata fulfill our requirements, and check that the required metadata fields are filled with correct values. The doorkeeper also provides client authentication for both Fedora and the SPARQL endpoint, as well as authorisation for the SPARQL endpoint. Authorisation for Fedora is controlled with the Fedora Commons WebAC module.

To facilitate the communication with Fedora, the PHP library “repo-php-util”²⁰ was developed. It provides high-level methods to manipulate repository resources in a consistent way, building on the basic functions exposed by the native Fedora 4 API, which are quite cumbersome to use directly. This library is used by most other components.

3.1. Components for Curation Tasks

Curation for ARCHE involves all activities related to processing data after its submission by the depositor until its transfer into Fedora. It is a mixture of manual and automated tasks. For initial automated checks of a submitted data collection the component “repo-file-checker”²¹ was developed. It is a command line application to perform a virus scan, to check for conformance of file names and accepted formats, and to validate formats. An overview of the dataset with respect to size, structure and formats used is also provided by this tool. These automated checks serve as a starting point for manual curation.²² The data is ingested into Fedora by using custom PHP scripts, which make use of the methods provided by “repo-php-util”. Upon ingest the doorkeeper executes further checks, which also includes validation of metadata against the metadata schema of ARCHE. The metadata schema is described in detail section 4.

3.2. Components for Dissemination

Dissemination and access to the resources stored in Fedora is provided via a web interface based on Drupal 8.²³ A custom Drupal module, “repo-gui”²⁴, allows the user to navigate, search and view the content. The web interface provides a basic user friendly view of selected metadata associated with the resources as well as an expert view displaying all available metadata. Furthermore the metadata is also available as RDF/XML.

In order to display specific data types or formats a growing set of dissemination services is provided, which allow the users to seamlessly view resources online. Examples include a transformation of TEI files into HTML, a IIIF viewer²⁵ to view high-resolution images, and a 3D viewer²⁶ for viewing 3D content.

The “repo-resolver”²⁷ service is used to resolve a resource URI to a particular representation of the resource, e.g. its default view in the repository browser, its metadata in RDF, its binary content or a PDF generated from the resource. Binding of resources to certain dissemination services is expressed either on the resource level, depending on its type, or in terms of general rules related to the resource’s metadata content (e.g. “apply dissemination service D, if a resource has a metadata property X and its value is Y”). These bindings are stored in Fedora 4 as dedicated repository resources and the metadata schema of ARCHE will include a distinct class to formally describe dissemination services.

ARCHE also exposes a public OAI-PMH endpoint²⁸, which is already regularly harvested by the Virtual Language Observatory (VLO)²⁹, CLARIN’s main metadata catalogue.

4. Metadata Schema

As a principle, all resources in ARCHE must have metadata, in order to help in finding, understanding, and reusing the data. Also all metadata is openly searchable via our web interface. It is encoded in RDF/XML³⁰ and provided with a CC0 license, which maximises the interoperability between ARCHE and the depositors and other service providers, like e.g. VLO,³¹ which harvest our metadata.

4.1. Development

The development of a suitable metadata schema for ARCHE faced some challenges related to the heterogeneity of data and the aim for supporting multiple metadata schemas.

In an initial approach a metadata schema was created by reusing properties from well known schemas, as DCMI Metadata Terms,³² FOAF,³³ SKOS³⁴ or RDFS.³⁵ Custom properties were included, when no other would fit. This resulted in a somewhat messy metadata schema, inconvenient for metadata curators to use, and not capable to properly describe all the resources we had collected until that point.

It was decided to replace the initial draft with a bespoke metadata schema with custom classes and properties. In a first step, a total of 15 already submitted data collections were analysed with respect to the metadata provided along. The collections already were following metadata standards like Dublin Core,³⁶ the TEI Header,³⁷ Component Metadata of

CLARIN (CMDI),³⁸ and the metadata recommended by IANUS.³⁹ Additionally the DataCite Metadata Schema⁴⁰ was taken into account, to be able to meet criteria set by DataCite, an organisation specialised in minting DOIs for research data.⁴¹

The resulting generic metadata schema is published in OWL⁴²-format on GitHub.⁴³ It contains 14 main classes, 71 datatype properties, and 56 object properties (about half of them are inverse properties) to describe projects (acdh:Project), data collections (acdh:Collection), and even single files (acdh:Resource). Description of persons (acdh:Person), organisations (acdh:Organisation), places (acdh:Place), and publications (acdh:Publication) related to the resources is also possible with the schema. In case rich metadata is available for resources in a dedicated format (e.g. CMDI, EAD, LIDO etc.), which cannot be losslessly encoded in the metadata schema of ARCHE, the corresponding metadata records are stored as separate binary objects and linked to the resources with the dedicated metadata property “acdh:hasMetadata”. The presence of the inverse property “acdh:isMetadataFor” allows to link between the resource and the metadata, and vice versa.

4.2. Documentation

OWL syntax not only provides means to describe an ontology, but also to document it. Thus we were able to create a self contained schema with general documentation and more specific descriptions and natural language labels for each property and class. Furthermore multilingual support is present by using the xml:lang attribute. In this way descriptions and labels can be served in English and German. Figure 2 illustrates how the property “acdh:hasUpdatedDate” is described and documented in the schema.

```
2322 <!-- https://vocabs.acdh.oew.ac.at/schema#hasUpdatedDate -->
2323
2324 <owl:DatatypeProperty rdf:about="https://vocabs.acdh.oew.ac.at/schema#hasUpdatedDate">
2325   <rdfs:subPropertyOf rdf:resource="https://vocabs.acdh.oew.ac.at/schema#hasDate"/>
2326   <rdfs:domain rdf:resource="https://vocabs.acdh.oew.ac.at/schema#RepoObject"/>
2327   <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#date"/>
2328   <rdfs:comment xml:lang="de">Datum der letzten Aktualisierung.</rdfs:comment>
2329   <rdfs:comment xml:lang="en">The date of the last update/change to the resource.</rdfs:comment>
2330   <rdfs:label xml:lang="en">has updated date</rdfs:label>
2331   <skos:altLabel xml:lang="en">Date of last update</skos:altLabel>
2332   <skos:altLabel xml:lang="de">Datum der letzten Aktualisierung</skos:altLabel>
2333 </owl:DatatypeProperty>
```

Fig. 2: The OWL-representation of the property “acdh:hasUpdatedDate”

How the properties are used in the actual objects is specified by the use of property restrictions with cardinality constraints in the class definition, as well as by respecting the domain and range included in each property definition. The cardinality information is important for both the web interface and the doorkeeper, to be able to validate ingested metadata and display it correctly. By setting cardinality constraints we are able to state if a property is optional or mandatory and if it may only be used once per object. Mandatory metadata was selected to ensure the inner workings of ARCHE and only applies to a very small set of metadata properties, comparable to the Dublin Core Metadata Element Set.⁴⁴

Additionally we selected recommended metadata properties to increase the understandability of data. User friendly information about mandatory, recommended, and optional metadata properties is available for depositors on ARCHE's website.⁴⁵ To include this information in the schema, the annotation property “acdh:recommendedClass” was introduced. In this way properties can not only be marked as mandatory or optional, but also as recommended for specific classes, which is not possible by only using cardinality constraints. An example is shown in Figure 3.

```

3173 <owl:Axiom>
3174 <owl:annotatedSource rdf:resource="https://vocabs.acdh.oeaw.ac.at/schema#Collection"/>
3175 <owl:annotatedProperty rdf:resource="http://www.w3.org/2000/01/rdf-schema#subClassOf"/>
3176 <owl:annotatedTarget>
3177 <owl:Restriction>
3178 <owl:onProperty rdf:resource="https://vocabs.acdh.oeaw.ac.at/schema#hasSubject"/>
3179 <owl:minCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">0</owl:minCardinality>
3180 </owl:Restriction>
3181 </owl:annotatedTarget>
3182 <acdh:ordinality>recommended</acdh:ordinality>
3183 </owl:Axiom>
3184 <owl:Axiom>

```

Fig. 3: The property “acdh:hasDescription” is marked as recommended for the class “acdh:Collection” with the custom annotation property “acdh:recommendedClass”

4.3. Term Lists

For some properties the use of term lists or controlled vocabularies is necessary, to ensure consistency and homogeneity in data description, as well as to increase findability. We decided against including these directly in the OWL file, to avoid unnecessary bloating and ease the maintenance of both the ontology and the vocabularies.

Instead the controlled vocabularies are kept in “ACDH Vocabularies”, the vocabulary repository service of ACDH.⁴⁶ The vocabularies are stored

in SKOS-format⁴⁷ allowing for multilingualism, modelling of hierarchies, and linking to related concepts from other vocabularies.

Properties for which a controlled vocabulary exists are annotated with the custom annotation property “acdh:vocabs” and the URI to the respective vocabulary as shown in Figure 4. Currently only three vocabularies exist for the properties “hasCategory”,⁴⁸ “hasLifeCycleStatus”,⁴⁹ and “hasAccessRestriction”.⁵⁰ Further vocabularies for the properties “hasLanguage”, “hasAppliedMethod”, “hasCharacterEncoding”, “hasRelatedDiscipline”, “hasSubject”, “hasLicense”, “hasFormat”, and “hasTransferMethod” are planned.

```
1679 <!-- https://vocabs.acdh.oeaw.ac.at/schema#hasCategory -->
1680
1681 <owl:DatatypeProperty rdf:about="https://vocabs.acdh.oeaw.ac.at/schema#hasCategory">
1682   <rdfs:domain rdf:resource="https://vocabs.acdh.oeaw.ac.at/schema#RepoObject"/>
1683   <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
1684   <rdfs:comment xml:lang="en">Type of resource, e. g. corpus. Choose from list. Can be re
1685   <rdfs:label xml:lang="en">has category</rdfs:label>
1686   <skos:altLabel xml:lang="en">Category</skos:altLabel>
1687   <skos:altLabel xml:lang="de">Kategorie</skos:altLabel>
1688   <acdh:vocabs>https://vocabs.acdh.oeaw.ac.at/archecategory/en/</acdh:vocabs>
1689 </owl:DatatypeProperty>
```

Fig. 4: For the property “acdh:hasCategory” the annotation property “acdh:vocabs” indicates which controlled vocabulary applies

4.4. Interlinking

The metadata set includes properties to establish relations to other publications and sources as well as to related data.⁵¹ Linking to previous versions of resources is also anchored in the metadata schema with dedicated properties (“acdh:isNewVersionOf” and “acdh:isPreviousVersionOf”).

Mappings to other well known and established metadata schemas can directly be stored in the OWL file by using “owl:equivalentProperty” or “owl:equivalentClass”. A mapping to Dublin Core is already present in the schema and served via the OAI-PMH endpoint mentioned in section 3.2.⁵² Further mappings are planned.⁵³

4.5. Display of Metadata

When objects are viewed in ARCHE, not all available metadata is displayed right away to the user. Instead a predefined set containing the most relevant information is displayed in a visually pleasing way, as demonstrated in

Figure 5. All metadata values can be accessed by switching from the “Basic-View” to the “Expert-View” or when using the API.

Overview [Copy Resource Link](#) [Switch to Expert-View](#)

amc: Austrian Media Corpus

amc
austrian media corpus

Type: Collection

Contact(s): Hannes Pirker

Creator(s): Hannes Pirker, Matej Ďurčo, Jutta Ransmayr, Daniel Schopper

Curator(s): Hannes Pirker

Contributor(s): Austria Press Agency

Available Date: 2013-06-01

Extent: 40159000 documents, 305580000 paragraphs, 667000000 sentences, 10091000000 tokens

Number of Item(s): 4

Binary Size: 56.19 KB

License: restricted

Access Restriction: public

[Download Collection](#) [GUI access](#) [RDF access](#)

Summary

Spatial Coverage: Austria

Description: The AMC was created as part of a public-private cooperation between the Austrian Academy of Sciences and the Austria Press Agency (APA). Thanks to the efforts of APA, the AMC covers a great portion of the Austrian media landscape of the past three decades, comprising a wide range of text types which can be classified as journalistic prose (Austrian newspapers, magazines, press releases). The texts in the AMC are lemmatized and tagged with part-of-speech labels. The corpus can be accessed via a corpus search engine (SketchEngine) on the

Fig. 5: Basic metadata for the collection “amc: Austrian Media Corpus” as displayed in the “Basic-View” in ARCHE. Full metadata can be accessed by switching to “Expert-View” with the toggle in the upper right corner

4.6. Metadata Editor

Presently depositors can provide metadata in Turtle (ttl) or RDF/XML format or fill in spreadsheets provided by ARCHE. The latter option is preferred by depositors, because providing a correct Turtle or RDF/XML file is not only painstaking and error-prone, but also requires special knowledge. The spreadsheets are then transformed into a Turtle file by the data curator,⁵⁴ to allow for automated ingesting of data with its metadata. For future deposits a browser-based metadata editor is currently under development.

The editor will provide forms with autocomplete functions and be able to output valid Turtle or RDF/XML files. The forms will be constructed automatically from the metadata schema, which will be provided by server-side components of ARCHE via dedicated endpoints in JSON Schema⁵⁵ format. Automated creation and validation of the forms is done with a Javascript application. This ensures that the forms always reflect the latest version of the metadata schema in use without the need to change the source code of the metadata editor.

While creating the initial prototype (illustrated in Figure 6), we found that further adjustments and annotations in the metadata schema were required, in order to facilitate the correct and automatic generation of forms. Since many metadata properties do not apply only to one object class, but to unions of classes (e.g. a collection, a project and a resource can have a contributor), we introduced 28 additional classes. This was needed to avoid anonymous classes. These classes are all children of the class “acdh:Helper” to be able to distinguish them from the classes used to represent content information in ARCHE. Additionally the annotation property “acdh:ordering” was introduced to define how the properties will be ordered in the form. Basic validation criteria for property values depend on the ranges defined in the metadata schema and are respected by the editor.

The screenshot shows the MetaDataEditor interface. The top navigation bar includes 'HOME', 'DATA', 'EDIT', and 'ONTOLOGY'. The left sidebar shows a tree view of 'Collections / Resources' with a list of identifiers. The main content area displays a form for a 'collection' with the following fields:

- hasCreator***: Bibliothek der Österreichischen Akademie der Wissenschaften
- hasSpatialCoverage***: ad-Duralb, al-Ahjur
- hasCurator***: Christoph Hoffmann
- collection**: (Section header)
- Title**: AT-OeAW-BA-3-27-A-GL1000b_01
- Title or name of Project, Collection or Resource. Is also used as name for an organisation. Alternative title**: (Empty text input)
- Alternative title or name of Project, Collection or Resource. Can also be used as alternative name for an organisation. Identifier**: (Empty text input)
- hasIdentifier***: (Empty text input)

Footnote: * Unique identifier given by ACDH and used in ACDH systems, as well as identifiers with a stable URL, or URI assigned by other parties, e.g. a DOI, a

Fig. 6: A form to enter metadata for a collection to be ingested in ARCHE

5. Outlook

Though ARCHE has become operational in 2017, a lot still remains to be done.

On the curation side, further improvements and expansions in regard to long-term preservation strategies are intended to be worked on, implemented, and formalised in a dedicated Preservation Policy. We also plan to consolidate and further automate the curation workflow. A part of this task is the installation and improvement of the metadata editor described in section 4.6 to support depositors and curators in preparing the collections for ingest.

With respect to the metadata schema, we aim to further formalise the schema based on the experiences we gathered in the initial phase. We will also ensure translations for all labels and comments are provided, in order to make the schema and consequently the user interface fully multilingual.

As described in section 4.4 mapping to other schemas is possible directly within the ontology. Further mappings to other schemas will be introduced. Moreover we plan to map the schema to the CIDOC Conceptual Reference Model (CRM)⁵⁶, to achieve better interoperability with other systems and aggregators.

Next to continuous improvements of the user interface, including the deployment of further specialised dissemination services, a language switch will be implemented.

The whole system will be continuously invigorated. And new functionalities shall be developed in response to challenges brought about by new collections to be deposited, be it in terms of new data types, internal complexity or the sheer size of the collections.

Mag.^a Martina Trognitz
Austrian Academy of Sciences,
Austrian Centre for Digital Humanities (ACDH)
E-Mail: martina.trognitz@oeaw.ac.at

DI Matej Ďurčo
ORCID: <http://orcid.org/0000-0002-5274-8278>
Austrian Academy of Sciences,
Austrian Centre for Digital Humanities (ACDH)
E-Mail: matej.durco@oeaw.ac.at

- 1 Consultative Committee on Space Data Systems, Reference Model for an Open Archival Information System (OAIS): recommended practice CCSDS 650.0-M-2, Magenta Book, Issue 2 (2012). <https://public.ccsds.org/Pubs/650x0m2.pdf> (15.03.2018)
- 2 Consultative Committee on Space Data Systems, Reference Model for an Open Archival Information System (OAIS): recommended practice CCSDS 650.0-M-2, Magenta Book, Issue 2 (2012),. <https://public.ccsds.org/Pubs/650x0m2.pdf> (15.03.2018), 3-1
- 3 <https://arche.acdh.oeaw.ac.at/browser/collection-policy>
- 4 <https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-meta-data>
- 5 <https://arche.acdh.oeaw.ac.at/browser/deposition-agreement>
- 6 S. E. Funk, Migration, in: H. Neuroth - A. Oßwald - R. Scheffel - S. Strathmann - K. Huth (eds.) nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (2010), Kap.8:10-8:15. <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949> (15.03.2018)
- 7 Force11, Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data, Publishing version b1.0. <https://www.force11.org/fairprinciples> (15.03.2018)
- 8 <https://www.w3.org/standards/semanticweb/> (23.05.2018)
- 9 R. Cyganiak - D. Wood - M. Lanthaler, RDF 1.1 Concepts and Abstract Syntax (2014). <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/> (17.05.2018)
- 10 F. Gandon - G. Schreiber, RDF 1.1 XML Syntax (2014). <http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/> (17.05.2018)
- 11 D. Beckett - T. Berners-Lee, Turtle - Terse RDF Triple Language (2011). <http://www.w3.org/TeamSubmission/2011/SUBM-turtle-20110328/> (17.05.2018)
- 12 S. Harris - A. Seaborne, SPARQL 1.1 Query Language (2013). <http://www.w3.org/TR/2013/REC-sparql11-query-20130321/> (17.05.2018)
- 13 S. Bechhofer - F. van Harmelen - J. Hendler - I. Horrocks - D. L. McGuinness, P. F. Patel-Schneider - L. A. Stein, OWL Web Ontology Language (2004). <http://www.w3.org/TR/2004/REC-owl-ref-20040210/> (17.05.2018)
- 14 A. Miles - S. Bechhofer, SKOS Simple Knowledge Organization System (2009). <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> (17.05.2018)
- 15 <http://docker.com/>
- 16 <http://fedorarepository.org/>

- 17 <https://www.blazegraph.com/>
- 18 ARCHE offers an API endpoint (<https://arche.acdh.oeaw.ac.at/blazegraph/sparql>) and a GUI access (<https://arche.acdh.oeaw.ac.at/blazegraph/#splash>)
- 19 <https://github.com/fcrepo4-exts/fcrepo-camel-toolbox/tree/master/fcrepo-indexing-triplestore>
- 20 <https://github.com/acdh-oeaw/repo-php-util>
- 21 <https://github.com/acdh-oeaw/repo-file-checker>
- 22 The deposition process and curation tasks are described in more detail on: <https://arche.acdh.oeaw.ac.at/browser/deposition-process>
- 23 <http://drupal.org/>
- 24 <https://github.com/acdh-oeaw/repo-gui>
- 25 The Loris IIIF Image Server is used. <https://github.com/loris-imageserver/loris>
- 26 M. Potenziani - M. Callieri - M. Dellepiane - M Corsini - F. Ponchio - R. Scopigno, 3DHOP. 3D Heritage Online Presenter, Computers & Graphics 52, 2015, 129–141. DOI: <http://dx.doi.org/10.1016/j.cag.2015.07.001>
- 27 <https://github.com/acdh-oeaw/repo-resolver>
- 28 <https://arche.acdh.oeaw.ac.at/oai>
- 29 <http://vlo.clarin.eu>
- 30 A human friendly view of the RDF/XML triples stored for each object in ARCHE can be accessed via switching to “Expert-View” or with the button “RDF access”.
- 31 The CLARIN OAI Harvester Web View displays harvested records from ARCHE: <https://vlo.clarin.eu/data/ARCHE.html>
- 32 <http://dublincore.org/documents/dcmi-terms/>
- 33 D. Brickley - L. Miller, FOAF Vocabulary Specification 0.99. <http://xmlns.com/foaf/spec/20140114.html> (15.03.2018)
- 34 A. Isaac - E. Summers (eds.), SKOS Simple Knowledge Organization System Primer. <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/> (15.03.2018)
- 35 D. Brickley - R.V. Guha, RDF Schema 1.1. <http://www.w3.org/TR/2014/REC-rdf-schema-20140225/> (15.03.2018)
- 36 <http://dublincore.org/documents/dcmi-terms/> and <http://dublincore.org/documents/dces>
- 37 The TEI Consortium, 2 The TEI Header, in: The TEI Consortium (ed.), TEI P5: Guidelines for Electronic Text Encoding and Interchange (2018). <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html> (15.03.2018)

- 38 CMDI Taskforce, CMDI 1.2 specification (2016). <https://www.clarin.eu/cmdi1.2-specification> (15.03.2018). Profiles considered: clarin.eu:cr1:p_1345561703620 and clarin.eu:cr1:p_1290431694580
- 39 D. Bibby - P. Gerth - M. Heinrich - S. Jahn - B. Ludwig - A. Posluschny - E. Sieglhoff - A. Sieverling - M. Trognitz, Metadaten in der Anwendung, in: IANUS (Hrsg.), IT-Empfehlungen für den nachhaltigen Umgang mit digitalen Daten in den Altertumswissenschaften. <https://www.ianus-fdz.de/it-empfehlungen/metadaten-anwendung> (15.03.2018)
- 40 DataCite Metadata Working Group, DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0. (2016). DOI: <http://doi.org/10.5438/0012> (15.03.2018)
- 41 <https://www.datacite.org/mission.html>
- 42 B. Motik - P. F. Patel-Schneider - B. Parsia (eds.), OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (2012). <http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/> (15.03.2018)
- 43 <https://github.com/acdh-oeaw/repo-schema>. This article refers to version 1.2 due to be released on 24.05.2018.
- 44 <http://dublincore.org/documents/2012/06/14/dces/>
- 45 <https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata#metadata>
- 46 <https://vocabs.acdh.oeaw.ac.at/>
- 47 A. Miles - S. Bechhofer, SKOS Simple Knowledge Organization System Reference (2009). <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> (15.03.2018)
- 48 <https://vocabs.acdh.oeaw.ac.at/archecategory/Schema>
- 49 <https://vocabs.acdh.oeaw.ac.at/archelifecyclestatus/Schema>
- 50 <https://vocabs.acdh.oeaw.ac.at/archeaccessrestrictions/Schema>
- 51 See the section “Relations to other projects, collections or resources” in the table in <https://arche.acdh.oeaw.ac.at/browser/formats-filenames-and-metadata#metadata>.
- 52 An example query would be: https://arche.acdh.oeaw.ac.at/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=http://hdl.handle.net/21.11115/0000-000B-C715-D
- 53 Candidates are EDM, PE, DCAT, and CMDI.
- 54 This can be automatised, but is still quite error-prone.
- 55 <http://json-schema.org/>
- 56 <http://www.cidoc-crm.org/>