eCH-0205 Linked Open Data

Name	Linked Open Data
eCH-number	eCH-0205
Category	Accessory Document
Quality stage	Defined
Version	1.0
Status	Approved
Decision on	2018-03-06
Date of issue	2018-03-13
Replaces version	-
Requirements	-
Annexes	-
Languages	English (original)
Authors	Members of the eCH Specialized Group "Open Government Data" Main Author: Beat Estermann, Berner Fachhochschule beat.estermann@bfh.ch For a list of further contributors, see Annex B.
Editor / Distribution	eCH registered association, Mainaustrasse 30, *Postfach* [P.O. Box], 8034 Zürich T 044 388 74 64, F 044 388 71 80 www.ech.ch / info@ech.ch

Summary

This document provides the Swiss Linked Data community with a shared vision of the state of linked open data publication in the public and heritage sectors in Switzerland and gives people who are new to the community a first overview of previous and ongoing activities in the area of data publication, data use, and know-how exchange. The document contains a short introduction to linked (open) data, gives a detailed account of what linked data publication is about, provides an overview of the present state of linked data publication by Swiss public and heritage sector organizations, and presents a series of exemplary use cases that serve as test and study cases to tackle current challenges and demonstrate the usefulness of linked (open) data in practice.

Table of Contents

1 Introduction

1.1 Status

Approved: This document was approved by the Experts' Committee. It has normative power for the defined field of application in the determined scope of application.

1.2 Background of this Document

The members of the Swiss Linked Open Data community regularly meet at the annual Linked Open Data Switzerland Workshop organized by HES-SO Valais/Wallis (Data Semantics Lab) in cooperation with the Swiss Federal Archives and the eCH Specialized Group "Open Government Data". The community mainly consists of specialists active in the area of linked data publication in the public sector (including research and heritage sector). This Whitepaper is the result of an effort of several community members to formalize their know-how in the area of linked data publication in order to foster the exchange within the existing community, to highlight current challenges, and to provide guidance for newcomers in the field.

The information presented in this document has been gathered through practical experience, desk research, the online exchange among communities of practices, such as the Wikidata+GLAM Community, as well as in the context of workshops at various real-life events, including the annual Linked Open Data Switzerland Workshops in Bern, the E-Government Fokus Event in fall 2016 in Bern, the LODLAM Summit 2017 in Venice, and Wikimania 2017 in Montreal.

1.3 Purpose of this Document

The purpose of this document is to provide the Swiss Linked Data community with a shared vision of the state of linked open data publication in the public and heritage sectors in Switzerland and to give people who are new to the community a first overview of previous and ongoing activities in the area of data publication, data use, and know-how exchange.

Concretely, the present document:

- provides a short introduction into linked (open) data;
- gives a detailed account of what linked open data publication is about;
- provides an overview of the present state of the publication of public and heritage sector information as linked (open) data in Switzerland;
- presents a series of exemplary use cases that serve as test and study cases to tackle current challenges and demonstrate the usefulness of linked (open) data in practice;
- points to existing case reports regarding the publication of linked data by Swiss public and heritage sector organizations.

Please note that this document does not provide a thorough introduction to linked data, as this aspect has been covered by many other publications and presentations. It also does not replace the personal involvement with the linked data community, as the direct exchange of know-how and ideas as well as the agreement on a common course of action is key to the successful publication and use of linked open data and cannot readily be replaced by guidelines or other formalizations of knowledge.

2 Brief Introduction to Linked Data

2.1 Linked Data in a Nutshell

The traditional World Wide Web, as we know it, allows us to click ourselves from one online document to the other by means of hyperlinks. The resulting network of documents distributed across many different websites has been termed the “Web of Documents”. The vision of linked data consists in extending this concept of a network of decentralized, but interlinked resources to the domain of data. The goal is to create a giant, decentralized database allowing computers to answer queries based on information found on many different websites, the “Web of Data”.

In recent years, there have been increasing efforts to break up data silos and to integrate data from hitherto isolated databases. This is especially the case for the open data movement with its focus on the secondary use of data and the creation of mashups with data from various sources. On one hand, the call to open up public sector information can be seen as a logical extension of the freedom of information regulations that have been adopted by many countries since the 1990ies in an effort to enhance government transparency. On the other hand, the open data movement is also driven by a technical and economical vision: transaction costs related to the secondary use of data are to be reduced and a semantic web is to be created by linking many open datasets from various sources. Thereby, linked open data will serve as an infrastructure resource for third parties to build value-added services on top of it, such as new combinations of data, visualizations, or other data-driven services. Thus, linked open data is also expected to foster economic development.

Subject	Predicate	Object
Bern (Q70)	is a (P31 - instance of)	municipality of Switzerland (Q70208)
Bern (Q70)	is the capital of (P1376 - is the capital of)	Switzerland (Q39)
Berlin (Q64)	is a (P31 - instance of)	municipality of Germany (Q262166)
Berlin (Q64)	is the capital of (P1376 - is the capital of)	Germany (Q183)
Switzerland (Q39)	is a (P31 - instance of)	country (Q6256)
Germany (Q183)	is a (P31 - instance of)	country (Q6256)
municipality of Switzerland (Q70208)	is a subclass of (P279 - subclass of)	municipality (Q15284)
municipality of Germany (Q262166)	is a subclass of (P279 - subclass of)	municipality (Q15284)

Table 1: Example triples taken from Wikidata

The term “linked data” was coined by Tim Berners-Lee, director of the World Wide Web Consortium, in 2006 and refers to a set of best practices for publishing and connecting structured data on the Web1. These best practices have since been adopted by an increasing number of data providers, leading to the creation of a global data space containing billions of statements – the Web of Data2. Like the “Web of Documents”, the “Web of Data” relies on unique resource identifiers (URIs) to address various resources on the Web, and on hyperlinks to create connections between them. This time, however, it is not documents that are interconnected, but atomic units of data, so called triples. Triples take the form of minimal sentences, consisting of a subject, a predicate and an object. Each triple corresponds to one statement about the world (see table 1); together, they make up a network of interconnected statements.

In his “Linked Data” note of 2006, Tim Berners-Lee outlined the principles of linked data which can be summarized as follows: Data is published in form of triples using open standards such as RDF, whereby URIs are used to name (identify) the different things represented by the elements of the RDF triples. Thanks to the use of the HTTP protocol, the URIs can be looked up on the Web where useful information about them is provided. All aspects of the data are expressed in RDF, including the ontology, so that the data is self-describing. Thus, when browsing data, one can look up the properties and classes one finds in data, and get information from the RDF, RDFS, and OWL ontologies, including the relationships between the terms in the ontology. In order to create links among different datasets, data publishers are encouraged to provide pointers from their data to the HTTP URI-based names contained in other data sets. In sum, these technologies provide an environment where applications can query structured data made available on the Web.

In 2010, Tim Berners-Lee published a 5-star model in order to indicate the road to good “linked open data”. Linked open data (LOD) is linked data which is released under an open licence, which allows the data to be re-used for any purpose for free. Linked data does not in general have to be open – there are many opportunities to use linked data internally, and for personal or group-wide data. The data, however, that is published as open data, can be rated based on a 5-star deployment scheme (see figure 1):

- **One-star data** is available on the Web (in whatever format) under an open license.
- **Two-star data** is in addition available as structured data (e.g. in tabular form instead of an image scan of a table).
- **Three-star data** is in addition available in a non-proprietary open format (e.g. CSV).
- **Four-star data** contains in addition URIs to denote things and uses open standards from W3C (RDF and SPARQL) so that others can point to the data.
- **Five-star data** is in addition linked to other data sources on the Web to provide context.

[1] Tim Berners-Lee (2006-07-27). "Linked Data". *Design Issues*. W3C. Retrieved 2017-11-16.

[2] Christian Bizer, Tom Heath, and Tim Berners-Lee (2009). Linked data-the story so far. Semantic services, interoperability and web applications: emerging concepts (2009): 205-227.

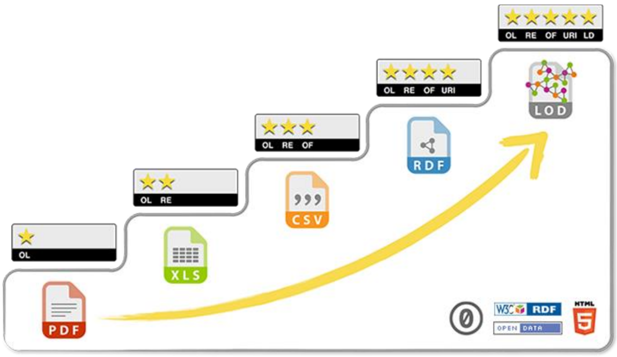

Figure 1: 5-star deployment scheme for open data[3]

In order to deploy linked data, the following infrastructure layers are required:

- a technical infrastructure (e.g. server, triple store, SPARQL endpoint);
- data models / ontologies (i.e. the semantic infrastructure);
- name services / authority files (i.e. registers of proper names);
- the actual "data".

While data providers may either run their own technical infrastructure or use a shared service, the semantic infrastructure and the registers of proper names are typically shared infrastructures. This means that data publishers usually do not manage all the data models / ontologies or run all the name services required to represent their data by themselves, but rely at least in part on the semantic infrastructure and registers of proper names provided by third parties. The actual data may be published as separate graphs, keeping the data from different providers apart, or it may be ingested and managed in one shared data space, as is the case of collaborative data management platforms like Wikidata[4].

2.2 Further Reading / Introductory Videos

Below are some pointers to introductory material to linked data (including Wikidata):

- Christian Bizer, Tom Heath, and Tim Berners-Lee (2009). **"Linked data-the story so far".** Semantic services, interoperability and web applications: emerging concepts (2009): 205-227 (Link).
- Tom Heath and Christian Bizer (2011). Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136, Morgan & Claypool Publishers (Link).
- Linked Data Tools: **"Introducing Linked Data and the Semantic Web"** (Link).
- Asav Bartov: **"A Gentle Introduction to Wikidata"** (Video).

[3] http://5stardata.info/en/. See also: Tim Berners Lee (2009) Linked Data, W3C Design Issues.

[4] Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation. It provides a common source of data which can be used by Wikimedia projects such as Wikipedia, and by anyone else. The data is made available under a "public domain" (CC 0) license.

3 Core Aspects of Linked Data Publication

Figure 2 gives an overview of the core aspects of data publication as linked data. It differentiates between the core processes of data publication and data use on one hand and various contextual aspects on the other hand. The latter comprise social aspects, given the fact that linked data has a lot to do with building communities and collaborating across organizational and disciplinary boundaries, as well as the aspect of technical functionalities in the form of various platforms and tools. The present section provides a short description of each of the aspects as well as the main challenges related to them[5].

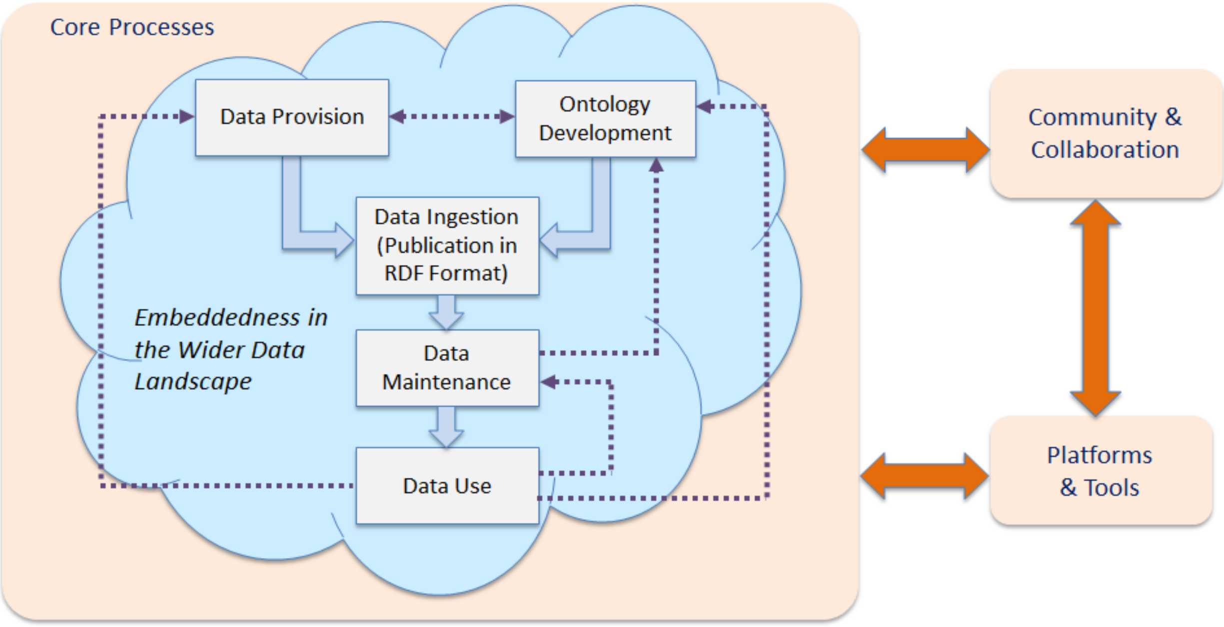

Figure 2: Core aspects of data publication

3.1 Data Provision

Data provision refers to the selection of the data and the right clearance in view of their future release to the public. Thereby, at least three perspectives need to be taken into account, two of which are similar to the ones that are at the core of the open data publication process, while the third one is specific to linked data:

1. **Legal and economic perspective:** To what extent can the data be released to the public without infringing on any property rights, privacy regulations, secrecy laws, or other legal provisions, and without hampering the intended economic exploitation of

[5] See also: Best Practices for Publishing Linked Data, W3C Working Group Note 09 January 2014, as well as Sören Auer et al. (2012) "Managing the life-cycle of linked data with the LOD2 stack." *International semantic Web conference*. Springer, Berlin, Heidelberg.

the data? – All these factors may affect the decision whether or not to publish a given data set or to make it available as open data. “Open” in the context of open data means that the data can freely be reused by anybody for any purpose, be it commercial or not.
For the provision of structured data as open data, the use of the Creative Commons Zero Waiver[6] or a similar legal instrument is recommended, as it greatly facilitates the exchange and the integration of data across organizational and jurisdictional boundaries. The Creative Commons Zero Waiver is also a requirement for data ingestion into Wikidata. Note however that the use of the Creative Commons Zero Waiver requires the data provider to make the data available without any further restrictions or conditions. By applying the Creative Commons Zero Waiver or a similar license to their data, data providers waive all copyrights and related or neighboring rights that they may have in all jurisdictions worldwide, such as their moral rights (to the extent waivable), their publicity or privacy rights, rights they have protecting against unfair competition, as well as database rights (e.g. in EU countries) and rights protecting the extraction, dissemination and reuse of data.

It is possible to publish linked data under certain restrictions; this requires however adequate legal instruments (such as laws or contracts) and most often also specific technical functionalities, such as suitable frameworks for rights management as well as identity and access management.

In our experience, the question of releasing data as open data is not only a legal one, but also one related to the mindset of individual decision makers as well as to organizational culture. Often, releasing data into the open is accompanied by fears of loss of control or the fear of giving some economic advantage to third parties without being able to recoup some of the expenses. These issues often need to be addressed within a given organization along with more technical and legal issues.

2. **User perspective:** To what extent is the data useful to potential users? – This question concerns not only the thematic area of the data, but also aspects like data quality, data completeness, timeliness of data publication, etc. Some of these aspects mainly depend on the circumstances of data generation and management on the side of the data provider. Others, such as data completeness, may also depend on the data publication activities by providers of complementary data. Furthermore, the usability of the data is strongly affected by the conditions under which users are allowed to use the data, the quality of documentation, and the responsiveness of the data provider to requests by data users.

3. **Linkability of the data:** Linked data is about interlinking datasets from various providers by pointing to the same named entities and classes and/or to the same properties. This implies that the various datasets that are meant to be linked to each other should share some aspects on which they can be aligned to each other, such as the same set of named entities representing municipalities or the same set of properties used to express relationships between organizational units. While a dataset can be

[6] https://creativecommons.org/publicdomain/zero/1.0/

published in RDF format without creating links to other datasets, the full power of linked data only unfolds when opportunities for interlinking are created. When selecting data for publication, it therefore makes sense to ask which links may be created to other data that have already been published as linked data elsewhere.

In some cases, some of the data may not be available in form of structured data, but in form of unstructured data (images, text, etc.). In this case, structured data first needs to be extracted from the unstructured data (e.g. through named entity recognition in archival documents, through face recognition in photographs, etc.).

3.2 Ontology Development

Ontologies (sometimes also referred to as data models) are shared languages that are used to represent aspects of the world in form of data. They typically consist of classes and properties for which definitions are provided as well as rules how the classes and properties may be combined among each other.

Besides shared sets of named entities (authority files), ontologies provide the links between different datasets within the linked data cloud. Like natural languages, they exist within their communities which develop and use them. Ontological structures and definitions are typically a product of negotiation within a given community, and each of these communities has its own set of rules that govern this negotiation process. In practice, the rules and processes may vary a lot from one community to the other: while some ontologies rely on formal processes and well-established expert committees in order to establish the authority and consistency of the data, others let their data models evolve in a more organic manner, making the ontologies more adaptable, but subject to change over time. The differences that can be observed in this area are reminiscent of Eric Raymond's Cathedral and Bazaar metaphor which was originally applied to two different free software development models[7].

Given the fact that ontologies provide the links between different datasets, it is common practice within the linked data community not to create ontologies from scratch for a given data publication project, but to reuse existing ontologies to a maximum. The reuse of ontologies avoids proliferation of different ontologies and leads to the consolidation of existing ones within a community. From a user point of view, it is easier to handle new data without having to learn a new ontology. On the downside, already existing ontologies may not be able to express the provider's data with the same precision as a customized ontology. In order to increase their expressiveness, existing ontologies can be extended or if existing ontologies prove inadequate for a given data publication project, new ontologies can be created. But even in the latter case it is good practice to provide information how the newly defined ontology can be mapped to ontologies that are widely used within the linked data community,

[7] Eric S. Raymond (1999). The Cathedral and the Bazaar. Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly Media.

such as schema.org[8], Dublin Core[9], vCard[10], or FOAF[11].

When choosing or developing an ontology, one will often need to strike a balance between three sometimes contradicting dimensions, such as:

- **The expressive power of an ontology:** To what extent is it able to express the subtleties of the real-world phenomena within the domain under consideration with the required precision?
- **Its practicability when it comes to the large-scale use by all its potential users:** With what ease will it be discovered, understood and uniformly applied by all the people that may use it to express information about a given domain?
- **Its usability from the perspective of data consumers:** How easy is it to write queries against the data encoded based on the ontology? How resource-intensive are common queries?

Several authors have attempted to describe the characteristics of a good ontology[12]. This is not an easy undertaking, as the quality of an ontology needs to be judged from a multidimensional perspective that takes into account its structure, its semantics, and its usability. Apart from some obvious criteria like the logical consistency of the definitions it contains and the assertions it makes about the real world, the quality criteria that have been suggested so far are hardly clear-cut, require some interpretation, and are usually not backed up by empirical data. Furthermore, attempts to define the quality of ontologies in a domain- and task-independent manner may miss important aspects that may come to the fore in a domain- and task-specific evaluation. In the absence of a clear answer to the question of what constitutes a good ontology, it is advisable to imitate the design patterns of ontologies that are used in similar domains and for similar purposes as the intended one and to put new ontologies to task-specific tests as early in the development process as possible with the above-mentioned dimensions in mind.

3.2.1 Types of Ontologies

Ontologies used in the context of linked data may contain both terminological axioms (terminological knowledge) and facts (assertional knowledge). They can be classified according to their complexity, ranging from the most simple and least expressive to the most complex and most precise[13]:

---

[8] http://schema.org/

[9] http://dublincore.org/specifications/

[10] https://www.w3.org/TR/vcard-rdf/

[11] http://xmlns.com/foaf/spec/

[12] For an overview of various quality criteria that have been identified in the literature, see chapter 3.6 of: Denny Vrandecic (2010). Ontology Evaluation. Doctoral Thesis. https://pdfs.semanticscholar.org/7d97/25cd4367979e56a2f89539d550071eb25f3e.pdf

[13] The classification largely follows the one proposed by: Denny Vrandecic (2010). Ontology Evaluation. Doctoral Thesis. Section 3.3.3 (Semantic spectrum). https://pdfs.semanticscholar.org/7d97/25cd4367979e56a2f89539d550071eb25f3e.pdf

- A **catalog** is an ontology that consists only of label annotations (i.e. a set of URIs with human readable labels).
- A **glossary** is an ontology that only contains annotations (thus, it comprises a set of URIs with human readable labels and definitions of terms).
- A **thesaurus** is an ontology that, besides annotations, also allows the instantiation of classes and properties from the SKOS ontology[14], used to describe how various concepts relate to each other (e.g. by pointing to narrower, broader, or related terms).
- A **formal taxonomy** (or **class hierarchy**) is an ontology that consists of simple subsumptions (class hierarchies), assertions (statements about reality), and annotations.
- A **proper ontology** (or **formal ontology**) is an ontology that contains the entire spectrum of terminological axioms, assertions, and annotations.

Other terms used to refer to ontologies serving a particular purpose are **"authority file"** (assigning unique identifiers to specific entities), **"subject indexing scheme"** (containing a list of index terms used to describe the subject of information artefacts), **"subject headings"** (containing a list of index terms used to assign information artefacts to a particular section of a classification scheme), or **"taxonomy"** (*"a subject-based classification that arranges the terms in the controlled vocabulary into a hierarchy without doing anything further"*[15]). Depending on their exact content these ontologies can be assigned to one of the five above-mentioned types. In order to avoid ambiguity, ontologies usually make use of a **"controlled vocabulary"** – another term that is sometimes used to refer to ontologies of various types (e.g. glossary, thesaurus, or taxonomy). Though formal definitions exist for these terms, at times, use of the concepts can be interchangeable in different professional contexts.

3.3 Data Publication in RDF Format

Data publication in RDF format can take different forms:

- Publication of the data in form of **RDF dumps**, i.e. data files containing RDF data in any of the accepted serializations, such as Turtle, N-Triples, N-Quads, JSON-LD, Notation 3, or RDF/XML.
- Publication of RDF data **through an API other than a SPARQL endpoint**, relying on a search engine with an index (example: data.swissbib.ch).
- Publication of data in a triple store or in a relational database, **queryable through a SPARQL endpoint**. There are several triple stores, such as Apache Jena, Virtuoso, or Stardog, which store the data natively in form of triples. Alternatively, there are

See also: Nicola Guarino, Daniel Oberle, and Steffen Staab (2009). "What is an Ontology?" In S. Staab and R. Studer (eds.), Handbook on Ontologies, International Handbooks on Information Systems, Springer Berlin Heidelberg, 1-17. https://userpages.uni-koblenz.de/~staab/Research/Publications/2009/handbookEdition2/what-is-an-ontology.pdf

[14] SKOS, the Simple Knowledge Organization System, is a W3C standard, based on other Semantic Web standards (RDF and OWL), that provides a way to represent controlled vocabularies, taxonomies and thesauri.

[15] Lars Marius Garshol (2004). "Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all." *Journal of information science* 30.4 (2004): 378-391. http://www.themezoom.com/library/pdf/Metadata.pdf

tools for publishing relational databases on the Semantic Web, such as D2R Server, which eliminate the need for replicating the data into a dedicated RDF triple store.

- Ingestion of the data into a **collaborative data management platform**, like Wikidata, from where it can be queried through a SPARQL endpoint; in contrast to the standalone triple stores mentioned above, collaborative platforms like Wikibase (the software solution behind Wikidata) provide a community-sourcing and crowdsourcing environment for the completion and enhancement of data.

Independently of the form of data publication, the URIs contained in the data need to be dereferenceable. This means that users need to be able to use the URIs to retrieve a description of the resources identified by these URIs. This applies both to URIs that are used to identify HTML documents and to URIs that are used in the linked data context to identify real-world objects or abstract concepts. Typically, the descriptions are served in a different format depending on whether the request is made by a human user or a software agent, e.g. HTML for humans and RDF for machines. This can be achieved using a HTTP mechanism called “content negotiation”[16].

A further avenue for linked data publication takes the form of HTML-embedded RDF data (RDFa, Microdata, or JSON-LD, drawing for example on the schema.org vocabulary), and is used mainly for the purpose of search engine optimization. While the data is exposed on a website and can be harvested by third parties, this form of data publication is usually not considered a valid alternative to the publication of structured open data in one of the above-mentioned forms. Its main purpose is to allow search engines to present the data exposed on websites in more meaningful ways. Typically, the set of vocabularies to be used for HTML-embedded data is restricted. Also, data items that are embedded in HTML pages do not systematically have their own identifier, which hampers the assertion of relationships between data items across documents and websites[17].

In addition to the data itself, metadata about published data should be provided, so that data users can assess the quality of published data and choose between different means of access. The metadata should comprise a clear statement about the origin, ownership and terms related to the use of the published data. Furthermore, it should contain information about the frequency of data updates. Datasets should be referenced in an open data catalogue[18]. Data catalogues and metadata entries should conform to the Data Catalogue Vocabulary (DCAT) standard[19].

The following challenges typically need to be tackled when publishing data as linked data:

---

[16] For further information about dereferencing and content negotiation, see: Tom Heath and Christian Bizer (2011). Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136, Morgan & Claypool Publishers. https://pdfs.semanticscholar.org/da63/084ec46c1912c4b4c4cad1ec8104dbef74aa.pdf

[17] Christian Bizer, Tom Heath, and Tim Berners-Lee (2009). Linked data-the story so far. Semantic services, interoperability and web applications: emerging concepts (2009): 205-227.

[18] Such as the one on http://opendata.swiss or https://datahub.io/.

[19] https://www.w3.org/TR/vocab-dcat/

3.3.1 Data Cleansing

Data quality and the degree of tidiness varies a lot between datasets; depending on the initial state of the original data set, data cleansing may be more or less cumbersome. During the data cleansing process, the following aspects should be addressed:

- **Tidiness of the data:** The data in the original datasets should be "tidy"[20]. This means that it should be easy to represent the data in a series of tables where each table contains information only about items of the same class, where each row contains information only about one specific item and where each column contains data for only one variable. With regard to statistical datasets, tidiness means that each variable is a column, each observation is a row, and each type of observational unit is a table.
- **Correctness and coherence of the data:** There may also be incorrect or incoherent data in the datasets that need to be cleaned up. It is not untypical that such issues surface during a data transformation or migration process.

3.3.2 Matching Items / Disambiguation

When transforming data into linked data, all entities need to be represented by a unique identifier, which is not necessarily the case in the original dataset. Thereby it is important that every entity get exactly one unique identifier, no matter how many names it has in the original dataset, and that every unique identifier point to exactly one entity. The classical problem of mapping person names to identifiers may serve as an example: If "Igor Strawinsky" was played in Zurich, and "Igor Stravinsky" in Geneva, the data entries should be mapped to the entity representing the Russian composer "Игорь Фёдорович Стравинский (*Igor' Fëdorovič Stravinskij*)", no matter what the actual transliteration of the name happens to be. The same goes for entries about people with several names (pen names, maiden names, etc.). In contrast, there may also be several people with the same name which need to be kept separate. Thus, Hans Caspar Hirzel (1725–1803), a Swiss medical doctor and politician, and Hans Caspar Hirzel (1751–1817), also a Swiss medical doctor and politician, as well as their three namesake ancestors, who also happened to be politicians, would all get separate identifiers.

When publishing data as linked data, the matching of items and their disambiguation needs to take place at two levels: Within the original datasets, and with regard to data that is already available in the linked data cloud. An important instrument for this purpose are so-called name authority files (or base registers), which provide lists of named entities already published as linked data and which many data publishers link against. By linking the entities contained in the original data file to the entries contained in a name authority file, the dataset is automatically linked to all the other datasets that point to the same authority file.

3.3.3 Ensuring Persistence of the Data and the Data Model

A well-considered URI naming strategy and implementation plan, based on HTTP (or

[20] See: Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10), 1-23.

HTTPS) URIs, is crucial for linked data.[21] URI structures used for the publication of linked data should not contain anything that could easily change. In order to maximize the possibilities of reuse that linked data brings to users, URIs need to be stable and reliable. There must be a balance between making URIs readable and keeping them more stable by removing descriptive information that will likely change.

3.3.4 Mapping Between Data Models

As part of the process of linked data publication, a mapping needs to be made between the (implicit) data model of the original dataset and the target ontology that typically needs to be published in RDF format. As mentioned above, the target ontology should have semantic overlaps with the ontologies used for the publication of other datasets. Otherwise, the dataset will not be linked to any other datasets within the linked data cloud.

3.3.5 Updates / Repeated Data Ingestion

In many cases, a one-time data ingest into a triple store will not be sufficient, as the data in the original database will be updated. Updates may take the form of corrections (e.g. a wrong birth date in the original data file), changes due to changes in reality (e.g. name changes, mergers, or dissolutions of organizations), additional time series (e.g. a city's population statistic for the present year), or some other form of additional information. This raises two challenges:

- Some data entries need to be historicized (an official name may be valid from date x to date y, an organization may cease to exist, etc.). In some original datasets, some form of historicization may already be represented (like for example in the inventory of Swiss municipalities published by the Federal Statistical Office); in this case, the issue is mainly one of mapping data models. Other datasets may just represent snapshots without storing information about changes that have occurred in the past (e.g. the database of museums in Switzerland searchable on the Swiss Museums Association's website); in this case, a concept of how to historicize the data needs to be developed: What is the granularity on the timeline that can be supported? How is data quality ensured? Does an absence of a change between two releases of the dataset necessarily mean that no change has happened in reality?
- A process needs to be set up to regularly update the data. If the data in the original dataset is clean and free of any errors, and future updates therefore just involve adding new time series, making regular updates is not so much a technical challenge, but rather an organizational one: a one time process within the organization needs to be implemented as a routine process. The situation gets more complicated if changes to existing data need to be made (e.g. corrections of data that have been ingested earlier) or if differing versions of the same dataset need to be reconciled (e.g. when third parties also have made changes to the target dataset in the meanwhile). In this case,

[21] See *Designing a URL structure for BBC programmes*, Michael Smethurst, 25 September 2014

some form of version history needs to be implemented, and possibly also a version comparison tool.

3.4 Data Maintenance

After the data has been published, various data maintenance issues may arise. The type of data maintenance that is required depends on several factors:

- How **reliable** is the data that has been published? – Depending on the processes of data collection and verification preceding the publication of the data, the data may be expected to contain varying numbers and types of errors. Data holders most often can make an informed guess with regard to the quality of their data, although their awareness of possible errors is likely to increase after the data has been put to use in new contexts.

  Example: The dataset containing the names of all Swiss municipalities published by the Federal Statistical Office (FSO) can be expected to undergo a strict quality check before publication and is probably used often enough by the FSO itself to ensure that potential errors are corrected before the publication of the data. In contrast, the inventory of cultural properties of national and regional significance, assembled by the Federal Office for Civil Protection, was ridden with errors and imprecise coordinates when it was used as a basis for the organization of the Wiki Loves Monuments photo contest in Switzerland back in 2012 and 2013. In fact, the Federal Office for Civil Protection had simply taken over the data from the cantonal monuments protection services, which had varying quality assurance processes in place. Often, the coordinates for the various objects had been recorded at a time when GPS was not available and the data could not be automatically visualized on an interactive map. Hence the many errors in the coordinates that had not been detected before putting the data to use in the context of the photo contest. Also, monument-related data is most often used in the context of renovation activities, which do not happen that often for a given monument; the use of the data is therefore rather limited, and updates may take place only years after changes have occurred. Not to mention archeological objects for which wrong coordinates were recorded on purpose. The example illustrates different sources of inaccuracies in a dataset, which may have a varying significance depending on the usage context.

- To what extent is the **reality** which a dataset describes **likely to change**? – Even if a dataset is 100% accurate for a given point in time, the reality it describes may be subject to change. Depending on the dataset, such changes may occur with varying frequency.

  Example: While the population statistic of the city of Bern for the year 1990 based on a given census method is not subject to change, heritage institutions contained in an inventory may change their names, their addresses, etc. over time.

- To what extent can the data publisher be considered as the ultimate **authority** for the published data? – In some cases, the official role to publish authoritative data on a particular topic may be assigned to the data publisher by force of law, while in other cases, an organization may aspire to a similar role as part of its mission. In contrast, data publishers may also include data in their datasets for which they do not aspire to have any authority.

Example: The FSO is required by law to keep and publish a list of the official names of Swiss municipalities. Many datasets of other data publishers contain municipality names in order to specify their data (which municipality does the data refer to?) or to provide further context (which municipality is an organization based in?). Although they may maintain an internal database of municipalities, these other data publishers do not have the authority over the names of Swiss municipalities. In some cases, there may be conflicting authorities which need to be resolved, once the data enters the same data space. This is for example the case for population statistics of cities that may be calculated according to varying methods depending on whether the statistics are calculated by the FSO or by the city's own statistical office (for example weekly residents may be counted differently). In this case, an apparent conflict of authority can be resolved by specifying the method of calculation, i.e. by specifying more exactly the concept of reference used by the two statistical services.

In the linked data world, data maintenance needs to address the following aspects:

- **Establishing and documenting data quality:** Data publishers (or data users) may have to deal with incorrect or inconsistent data. When several data publishers ingest the same type of data into the same data space (which is typically the case when publishing linked open data), duplicates may arise. If statements are contradictory (e.g. different population statistics for the same city and the same year), definitions may be further refined, or erroneous entries may be removed. Duplicate statements that are not contradictory do not represent an issue per se. Also is it possible to have contradictory statements that are linked to varying sources (see the comments regarding authority and trust below). Depending on the case at hand, it may be useful to regularly monitor and document data quality and data completeness.
- **Building a network of trust:** In the long run, it is important to build a network of trust around any given dataset. First of all, the entity publishing the data should be identifiable by data users. The same goes for any entity that makes changes to the data (e.g. corrections etc.). In this way, data publishers can build up a reputation. Furthermore, statements can be referenced to reliable sources. If data from an authoritative source has been published as linked data, a data publisher may decide that it is enough to point to the authoritative source and stop publishing a given statement themselves (e.g. it would be enough for the publisher of pollution data to include the FSO identifier for a given data point in order to point to the municipality register published by the FSO from where the name of the municipality in the various languages as well as the district, the canton, and the country can be inferred).

Note: To enhance the possibilities of building a network of trust around data, the Wikidata community is presently considering the option of allowing users to sign statements as a form of endorsement[22]. This would give organizations the possibility to mark data as reliable and facilitate the collective identification of authoritative data in areas where it may not be clear from the outset which data publisher can generally be regarded as an authoritative source.

- **Ensuring the synchronization between the data published as linked data and the primary databases:** If relevant changes happen to the data either on the primary database (the source database) or in the data published as linked data, the synchronization between the two needs to be ensured and differences in the data need to be managed. Changes may take the form of corrections or updates, may be triggered by the data publisher or by the data users, and may be made directly on the primary database or made within a crowdsourcing environment like Wikidata.

3.5 Data Use

Data use plays an important role in the context of the publication process, as it provides valuable feedback regarding the usefulness of the data, the form of its publication, the correctness of the data, the adequacy of the data models employed, etc. Thus, data usage is an important driver of data quality and completeness. At the same time, data quality and completeness are important prerequisites of data use. As a result, data publishers are usually facing a chicken-and-egg problem, which calls for an iterative approach, in which various forms of data use are supported in a step-by-step process, starting with the low-hanging fruits by allowing simple uses of the data and progressively advancing towards more complex forms.

Data use may take place both by the data publishing organization and by the wider community of data users. Obviously, the possible uses of linked data are as manyfold as the data itself. Below a few examples of low-threshold uses of linked data. Whether they are possible or not in the case of a given dataset depends on the nature of the dataset:

- Link the data to authority files (or base registers) that have been published as linked data, and complement the source database with data from the authority files (or base registers). By this means, some of the data maintenance tasks related to data for which the data publisher does not act as the authority may be outsourced to the maintainer of the authority file (or the base register). While the systematic linking to authority files (or base registers) may be quite a burdensome task for some datasets, this step forms an important part of the data publication process anyways (see: 3.3.2 Matching Items /). Thus, only minimal extra effort is required to pull complementary data from the authority file (or base register). Integrating additional data fields into existing applications maintained by the data publisher would represent another logical step.

[22] https://phabricator.wikimedia.org/T138708

- As a further step, complementary data may be pulled not only from the authority file (or base register) itself, but also from other datasets that link to the same authority file (or base register). This step requires some reflection about which organizations can be considered as authorities in a given field as well as some investigation into the maturity of the form of data publication (data quality, reliability, regular updates, etc.). But most interestingly, this is a two-way process, as our data publisher may not only be on the receiving, but also on the giving end, as other data publishers may start thinking about complementing their data with the data from our data publisher. Thus, it becomes possible to discuss aspects of data quality, form and reliability of data publication, data models used, etc. on a mutual basis.
- A valid alternative to undertaking the above-mentioned steps consists of ingesting some of the data into Wikidata and in turn complementing the source dataset with Wikidata identifiers[23] and any other data from Wikidata that is deemed sufficiently reliable. In many fields, Wikidata acts as a hub linking to various existing authority files, or can be used as an authority file itself. Due to its collaborative nature, the threshold for creating new authority records is lower than for other authority files or base registers. This may be considered a problem if one primarily values stability, reliability and due process, but it may also be an advantage due to the flexibility and the responsiveness of the approach: New records that are created today by anybody in the world can tomorrow be enriched and linked to by any other entity. This greatly facilitates multilateral collaboration. Furthermore, data that is ingested into Wikidata is automatically available for use in Wikipedia, one of the most visited websites in the world. Data from Wikidata is for example used for the infoboxes on the top-right side of Wikipedia articles. Catalan Wikipedia has been a pioneer in this area, creating infobox templates with interactive maps, detailed career information for biographies and various other information directly fetched from Wikidata. Thus, as of fall 2017, over 50% of the 550'000 articles on Catalan Wikipedia are using data from Wikidata. Other language versions of Wikipedia are however still at the beginning of this integration process, among them the large English, French and German Wikipedias.[24]

3.6 Embeddedness in the Wider Data Landscape

As established above, data publishers engaging in linked open data publication are not filling a void or starting from scratch, but contributing to an existing ecosystem of data, data models, and community members. Data publication and maintenance takes place in a broader context. An important part of the data publication process therefore consists in finding one's

[23] See the examples listed at "Wikidata for authority control"

[24] See: "Wikidata, a rapidly growing global hub, turns five" (blog post by Andrew Lih and Robert Fernandez, 30 October 2017). Switzerland-related data that is directly integrated into Catalan Wikipedia comprises for example the names of current city mayors, population statistics, prizes received by artists, etc. On Portuguese Wikipedia, some list articles are generated dynamically from Wikidata, e.g. https://pt.wikipedia.org/wiki/Lista_de_pinturas_de_Alfredo_Norfini.

place within the linked data ecosystem, identifying important partners and allies and collaborating with them. Important questions that should be asked are:

- In which field does my organization aspire to publish authoritative data?
- Who publishes authoritative data in adjacent fields that may have overlaps with our own data?
- Who are the publishers of complementary data? What other data sources should my organization link to or synchronize data with?
- What are viable avenues for data improvement? (e.g. comparison of the data to other data sources, gathering feedback from data users, crowdsourcing, etc.)
- How to map between existing data models? How to integrate them and to further develop shared ontologies?
- How best to engage in community building and know-how exchange? How to engage in a dialogue with (potential) data users?

3.7 Interactions and Feedback Loops Between the Different Aspects

As has already been noted, for most data publishers, the publication of linked open data will take the form of an iterative process, with many interactions and feedback loops between the different steps of the process (see figure 2). Thus, data provision and data maintenance may trigger work in the area of ontology development and vice versa. And especially data use may be expected to lead to valuable feedback that will not only affect data maintenance and ontology development, but also trigger the provision of further data.

4 The Role of Community and Collaboration

The publication of linked data is presently on the rise. More and more organizations are starting to make their data available as linked open data, which opens up new opportunities for its use. As a consequence, an increasing number of professionals involved in the development and the exploitation of data infrastructures need to get acquainted with the technology and its implications in their respective fields. At the same time, more widespread use of linked data brings to the fore various challenges that still need to be tackled (e.g. how to document data provenance and track this information throughout the query pipeline? how best to approach the historicization of data?). The increasing use of linked data also acts as a driver for the development of user-friendly tools and software components that provide the typical functionalities of the data lifecycle to model, create, store, secure, publish, edit, search, retrieve, integrate, process, and visualize linked data. In many areas, mature and integrated solutions are still lacking, and initiatives to create new or improved tools abound. This has resulted in a tools landscape that is scattered and in constant evolution. As more and more institutions from various fields are publishing their data as linked data, an increasing number of data models are ported to the RDF format. Furthermore, the prospect of being able to query data across domains and across institutional boundaries provides additional impetus with regard to the integration and the alignment of various ontologies.

Quite naturally, community and collaboration play an important role in this context, most notably with regard to the following aspects:

- A community of practice is instrumental in documenting and exchanging know-how among experienced members and in facilitating the integration of new members to the linked open data community. Community gatherings are also occasions in which training can take place.
- The community serves as a place where information about new tools is exchanged and where requirements with regard to tools that still are to be developed can be articulated. It therefore also provides a forum for the exchange between developers and users of software tools that can be used in the context of linked data.
- The community serves as a forum where data modelling practices are agreed upon and where shared ontologies are developed.

Community and collaboration manifest themselves at concrete events, such as at conferences, hackathons, working group meetings, etc. as well as in the documentation and tools the various groups share among each other. Below is a short, and certainly not exhaustive list of recurrent events and documentation hubs consecrated to linked open data. The focus is on events that are relevant to publishers and users of linked data from the Swiss government and heritage sectors.

4.1 Community Events in Switzerland

The following events typically serve as a get-together for members of the Swiss linked data community in the government and/or heritage sectors:

- **LDSW – Linked Data Switzerland Workshop**
  The annual Linked Data Switzerland Workshop is organized by the HES-SO Valais/Wallis (Data Semantics Lab), in cooperation with the Swiss Federal Archives and the eCH Specialized Group OGD. It brings together LOD practitioners related to the public and the heritage sectors with the purpose of fostering the exchange among them.

  URL: Webpage of the last edition of the workshop
- **Swiss Open Cultural Data Hackathon**
  The Swiss Open Cultural Data Hackathon is organized by the OpenGLAM Working Group of the opendata.ch association. The annual hackathon features a side programme that regularly addresses issues related to linked open data and Wikidata and offers introductory sessions. Furthermore, linked data and Wikidata have been the object of various hackathon projects.

  URL: Hackathon overview page

4.2 Selected Community Events at the International Level

Members of the Swiss linked data community in the government and heritage sectors also regularly attend the following international conferences and gatherings:

- **International Semantic Web Conference**

  The annual International Semantic Web Conference (ISWC) is the leading conference for research on Semantic Web topics.

  URL: Overview page
- **Wikimania**

  The annual conference of the Wikipedia/Wikimedia community regularly features sessions about Wikidata.

  URL: Wikimania
- **WikidataCon**

  The conference dedicated to the Wikidata community, first held in October 2017.

  URL: WikidataCon 2017
- **WikiCite**

  Conference on citation metadata in Wikidata, first held in 2016.

URL: WikiCite

- **LODLAM Summit**

  The biannual unconference brings together practitioners from the field of linked open data in libraries, archives, and museums.

  URL: LODLAM Summit 2017

4.3 Community Hubs

The following community hubs are used to share best practices and to coordinate collective action in their respective fields:

- **Linked Data @ World Wide Web Consortium (W3C)**

  The W3C maintains various working groups and lists of resources:

  URL: Recent Press, Current Status of Specifications, Working Groups

  URL: List of various resources about linked data

- **Wikidata + Heritage Data**

  The Wikidata + GLAM community is quite active and maintains both a Facebook group and an overview page on Wikidata. Its members also contribute to the monthly newsletter of the GLAM-Wiki community:

  URL: Facebook Group

  URL: Overview Page on Wikidata

  URL: This month in GLAM Newsletter (includes a Wikidata report)

- **Wikidata in general**

  There are various forums where Wikidata-related issues are addressed, for example:

  - Project chat (in English) / Bistro (in French)
  - Various thematic Wikiprojects
  - Weekly status updates (on-wiki or by email)

5 Platforms and Tools

As mentioned above, the landscape of tools, services, and technical functionalities available for the use in the context of linked data is in constant evolution. Many tools have been developed in the course of research projects, as part of a volunteer effort, or have been donated to the community by the company that has initially developed them. It is therefore not uncommon to see that tools are not further maintained and that services are discontinued; in other cases, the business model and conditions of use may evolve, and tools or services once offered for free are further developed and offered on a commercial basis. Sometimes, the source code of the original tools remains available for reuse under an open source license. In the meanwhile, new tools and services pop up, potentially going through a similar cycle as the previous ones. For the use with Wikidata alone there are over sixty tools.[25] It is therefore impossible to provide a comprehensive list of platforms and tools in the context of this Whitepaper. In order to give a flavor of the kind of tools and services that are used in the context of linked data publication, this section provides an overview of various types of tools with some examples.

5.1 Tools for Data Extraction

Tools for data extraction serve the generation of structured data from non-structured data by distilling specific information from free-flowing textual or from audio or visual sources. For example, for structured information to be extracted from unstructured texts, the following main subtasks are typically involved[26]:

- **Preprocessing:** the text is prepared for processing with the help of computational linguistics methods such as tokenization, sentence splitting, morphological analysis, etc.
- **Identifying and classifying concepts:** mentions of people, things, locations, events and other pre-specified concepts are detected and classified.
- **Connecting the concepts:** relationships between the extracted concepts are identified.
- **Unifying the data:** the extracted data are brought into a standard form.
- **Removing duplicates:** duplicate data is eliminated.

Information extraction can be entirely automated or performed with the help of human input. Good data extraction solutions effectively combine automated processes with human processing.

Examples: DBpedia Spotlight, Babelfy

[25] See Wikidata:Tools

[26] https://ontotext.com/knowledgehub/fundamentals/information-extraction/

5.2 Tools for Converting non-RDF Data Sources into RDF

So-called RDFizers or converters are software or services for converting non-RDF data sources into one or more of the RDF data model serializations. Similarly to the tools for data extraction, they often come with some degree of ontology matching. They may come as standalone tools or may be included in larger data publishing tools (see below).

Example: Tarql: SPARQL for Tables

5.3 Tools for Data Cleansing and Enhancement

Often, the data to be published as linked data has shortcomings and needs to be cleaned up before its publication: data may be untidy[27] and contain errors or inconsistencies. Also, the source data may be available in a format that requires some reformatting before its transformation into RDF, and data structures may need to be aligned with target ontologies. Usually, these issues have to be resolved either before the start or during the data publication process. Furthermore, data may be enhanced by completing or extending it with the help of web services and external data.

Example: OpenRefine

5.4 Data Publishing Tools

Data publishing tools shield publishers from dealing with technical details such as content negotiation and ensure that data is published according to the linked data community best practices. Some of the tools also provide SPARQL query access to the served data sets and support the publication of RDF dumps. Some tools allow for the transformation of data held in non-RDF data formats (e.g. relational databases, databases conforming to the OAI-PMH standard, or data formats of popular CMS) to RDF data.

Examples: Virtuoso, Trifid

5.5 Data Ingestion Tools

Data ingestion tools facilitate data ingestion into a specific linked data platform, such as Wikidata.

Example: Quick Statements

[27] See: Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10), 1-23.

5.6 Tools for the Interlinking of Data

During or after its publication, data may be reconciled and matched against authority files (base registers) or against any other dataset that has overlaps with the published data.

Examples: SILK, LIMES, Mix'n'match

5.7 Tools for the Creation and the Management of Ontologies

During the data publication process, one may face the need to create a new ontology or to extend existing ones. There are various tools supporting the ontology development process, from the initial creation of the ontology to its maintenance. While some tools cover the entire lifecycle of an ontology, others focus on specific aspects like the initial creation of an ontology, its editing, its mapping to other ontologies, or its visualization in order to facilitate its inspection.

Examples[28]: Protégé, TopBraid Composer

5.8 Tools for Data Maintenance

Data that has been published as linked data needs to be maintained: Data may be checked for internal inconsistencies or for inconsistencies between various datasets, or the data may be complemented based on reasoning. For this purpose, a semantic reasoner is used – a piece of software that is able to infer logical consequences from a set of asserted facts or axioms. Thus, new statements can be created based on existing data; if such logical reasoning results in contradictory statements, this is an indication that there is something wrong with the data, and inconsistencies can subsequently be dealt with.

Furthermore, the data can be validated against a set of predefined constraints as defined by the data model. To do so, the constraints themselves are described in RDF format using the Shapes Constraint Language (SHACL)[29], and the actual data is subsequently validated against the conditions that have been defined. SHACL includes features to express conditions that constrain the number and types of properties an item of a particular class may have, the number of values that a property may have, the type of such values, numeric ranges, string matching patterns, as well as logical combinations of such constraints.

And finally, the original data source from where the linked data was generated as well as other datasets a given dataset links to may be in constant evolution, which requires continuous monitoring and updating of the data. Tools are therefore needed to monitor differences

[28] For an overview, see: http://wiki.opensemanticframework.org/index.php/Ontology_Tools

[29] For the SHACL specification, see: https://www.w3.org/TR/shacl/; for a tutorial, see: https://www.topquadrant.com/technology/shacl/tutorial/

between datasets, data completeness or the evolution of datasets over time.

Example[30]: Pellet

5.9 Tools for the Discovery of Data

Tools facilitating the discovery of linked open data comprise linked data browsers and linked data search engines. Linked data browsers provide the mechanisms for navigating the information space, allowing end-users to live-browse the Web of Data. Data search engines crawl linked data from the Web by following RDF links, and provide query capabilities over aggregated data, thus allowing end-users to search the Web of Data. While some search engines are human-oriented, others serve the needs of applications built on top of distributed linked data by providing APIs through which linked data applications can discover RDF documents on the Web that reference a certain URI or contain certain keywords[31].

Example: LOD Live Browser

5.10 Tools for Data Analysis and Visualization

There are various tools for data analysis and visualization. While some are integrated in the query interfaces of triple stores, others come as standalone tools. Another distinction can be made between generic visualization systems for linked data, and domain specific visualization systems.

Examples[32]: Wikidata Query Service, Histropedia, Rhizomer

5.11 Programming Frameworks

Programming frameworks provide a universal, reusable software environment that provides particular functionality as part of a larger software platform to facilitate development of software applications, products and solutions. There are several programming frameworks which support the use of data in RDF format.

Examples: Apache Jena, Eclipse RDF4J (formerly known as Sesame)

---

[30] For a list of reasoners, see: http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/

[31] Christian Bizer, Tom Heath, and Tim Berners-Lee (2009). Linked data-the story so far. Semantic services, interoperability and web applications: emerging concepts (2009): 205-227.

[32] For an overview and discussion of existing linked data visualization tools, see: Nikos Bikakis and Timos Sellis (2016). Exploration and Visualization in the Web of Big Linked Data: A Survey of the State of the Art, Workshop Proceedings of the EDBT/ICDT 2016 Joint Conference (March 15, 2016, Bordeaux, France).

6 Present State of Linked Data Publication by Swiss Public-Sector Organizations

This chapter provides an overview of the present state of linked data publication by Swiss public-sector organizations, including heritage institutions and international organizations located in Switzerland. It not only lists the datasets and Switzerland-specific ontologies and vocabularies that have already been published as linked data, but also points to various use cases and case reports.

6.1 Public Sector Information Available as Linked Data

This section contains an overview of datasets that have already been published as linked data. In the future, these datasets shall be listed on the opendata.swiss platform and the search function shall be enhanced to enable filtering for data published as linked data.

Dataset	Publisher / Maintainer	Links
aLOD Selected archival finding aids of the Swiss Federal Archives and several cantonal and municipal archives.	aLOD Consortium	- Description - SPARQL endpont (aLOD)
Art in the Urban Space The dataset contains information about objects of art that are in the possession of the city of Zurich and accessible to the public.	City of Zurich, Office for Art in the Public Space Data ingested into Wikidata by a volunteer.	- Description - Sample query - SPARQL endpoint (Wikidata Query Service)
DiDok (location documentation) The dataset contains the Swiss Federal Railways' operating points, including all the stops in Switzerland.	Swiss Federal Railways	- Description with sample queries - SPARQL endpoint (LINDAS)
Historicized Municipality Inventory	Federal Statistical Office	- Description with sample queries - SPARQL endpoint (LINDAS)

Dataset	Publisher / Maintainer	Links
PCP Inventory Inventory of cultural properties of national or regional significance in Switzerland.	Federal Office for Civil Protection, PCP Section Data ingested into Wikidata by a student at Bern University of Applied Sciences.	- Description - Sample query - SPARQL endpoint (Wikidata Query Service)
linked.swissbib.ch Catalog of all Swiss university libraries, the Swiss national library, several cantonal libraries and of other institutions, which release their data under CC0.	Maintained by Swissbib (swissuniversities and University of Basel)	- Data model - API documentation with sample queries - API
Professional Performing Arts Companies The dataset contains information about professional performing arts companies in Switzerland.	Swiss Archives of the Performing Arts (formerly: Swiss Theatre Collection) Data ingested into Wikidata by a student at Bern University of Applied Sciences.	- Description - Sample query - SPARQL endpoint (Wikidata Query Service)
Selected environmental data So far, datasets about air, soil, and water quality have been published.	Federal Office for the Environment	- Description with sample queries - SPARQL endpoint (LINDAS)
swissBOUNDARIES 3D The dataset contains the administrative territorial units and borders of Switzerland and Fürstentum Liechtenstein.	Swisstopo	- Description with sample queries - SPARQL endpoint (Geo Admin)
Swiss GLAM Inventory Inventory of Swiss heritage institutions.	OpenGLAM CH	- Description - Sample query (all GLAMs and heritage institutions in Switzerland and Fürstentum Liechtenstein) - SPARQL endpoint (Wikidata Query Service)
Taxonomic names of species Database containing currently accepted scientific names of species and their synonyms.	Plazi	- Description - SPARQL endpoint (LINDAS Test Server)
The Stapfer-Enquête	Swiss Federal Archives	- Description

Dataset	Publisher / Maintainer	Links
Database containing the results of a unique survey of the school situation in the Helvetic Republic dating from 1799.	SFA	- SPARQL endpoint (LINDAS)

6.2 Switzerland-specific Ontologies and Vocabularies

This section contains an overview of ontologies and vocabularies that are Switzerland-specific and have been published in RDF format. Ontologies and vocabularies widely used by Swiss government agencies typically have the form of eCH standards.

Ontology / Vocabulary	Publisher / Maintainer	Links
Government Ontology Data Standards Concerning Administrative Territorial Entities (eCH-0008 / eCH-0071)	eCH / Zazuko GmbH	gont.ch

6.3 Use Cases

This section contains a selection of use cases that serve as test and study cases to tackle current challenges and demonstrate the usefulness of linked open data in practice.

6.3.1 Worldwide Museums Database

Description	A worldwide museums database is to be realized by ingesting existing national museums databases into Wikidata and by drawing on the Wikipedia/Wikidata community to enhance and complement the data. The project idea was raised by the International Council of Museums (ICOM), who has expressed its interest in having a worldwide museums database at its disposal for disaster management purposes, by the Wikimedia community who is interested in improving the Wikidata database and in enhancing the coverage of museums within Wikipedia, as well as by representatives of the OpenGLAM and GLAM-Wiki communities who are interested in promoting open data and crowdsourcing among heritage institutions. Due to obvious synergies, data on archives and libraries may be ingested alongside the museum data.
Aspects covered / challenges raised	- Ingesting data into Wikidata; breaking down the ingestion process into sizeable chunks and writing guidelines to allow many contributors to get involved in the ingestion process; - Data modelling issues, e.g. with regard to modelling postal addresses in Wikidata (link to existing eCH standards); - Managing multilingual thesauri within Wikidata; - Integration of the historicised inventory of Swiss municipalities along with the museums data (today, the GLAM inventory needs to be updated manually every year in order to reflect mergers of municipalities); - Crowdsourcing data enhancement and completion of the database; - Automatically adding coordinates based on postal addresses; - (Semi-)automatic update of data in Wikidata from source databases (e.g. from VMS museums inventory and from the inventory of Swiss municipalities); - Interaction (mutual update) between data on LINDAS and data on Wikidata; - Using data from Wikidata within Wikipedia; - Tackling the question of complementarity between LINDAS and Wikidata: to what extent does it make sense to mirror data that has been ingested into Wikidata also on LINDAS? (What is the scope, what are the notability criteria on LINDAS? What is the added value of LINDAS compared to Wikidata and vice versa?).
Parties involved / implementation status	The project was kicked-off on 1 December 2016. The parties involved were ICOM, Wikimedia Foundation, Unesco, Europeana, Wikimedia Sweden, and OpenGLAM CH. Due to changes of personnel at ICOM, the project was paused in 2017. Data for all heritage institutions in Switzerland have been ingested (Swiss GLAM Inventory published by OpenGLAM CH). Based on this experience

	and based on draft instructions established by other community members, instructions for the ingestion of data from further countries will be established. In the meanwhile, the Swiss National Libarary has published another inventory of Swiss heritage institutions: ISplus.
Relation to other datasets	- Historicized inventories of municipalities and higher-level administrative-territorial units (in Switzerland and in other countries) - Inventories of private enterprises (in Switzerland and in other countries) - Inventories of public sector organizations (in Switzerland and in other countries) - Inventories of physical addresses with coordinate location information (in Switzerland and in other countries) - GND, LoC, Rameau, etc. - Library Authority Files (for the typology of heritage institutions)
Case reports / documentation	- Ingesting Swiss Heritage Institutions (Beat Estermann, November 2016) - Towards Ingesting GLAM Inventories on a Global Scale (Beat Estermann, January 2017) - Wikidata Guideline - Editing Data in Spreadsheet Mode (Beat Estermann, January 2017)
Applications making use of the data	The museum template of the Italian Wikipedia is making use of data from Wikidata in the absence of data entries for selected fields on Wikipedia itself (example).
Contact	Beat Estermann, (Bern University of Applied Sciences / OpenGLAM CH) beat.estermann@bfh.ch

6.3.2 Worldwide Database of Historical Monuments

Description	Wikipedia's monuments database contains data about historical buildings in about 70 countries and territories. It is from this database that the world's largest photo contest has been run since 2010. In 2017, the entire database that originally was established by harvesting monuments lists from Wikipedia is expected to be ingested into Wikidata as all the new tools that are to be used in the context of the contest (apps, visualizations, maintenance bots, etc.) will run on the basis of Wikidata in the future.
Aspects covered / challenges raised	- Ingesting data into Wikidata; breaking down the ingestion process into sizeable chunks and writing guidelines to allow many contributors to get involved in the ingestion process; - Managing multilingual thesauri within Wikidata; - Integration of the historicised inventory of Swiss municipalities along with the monuments data (today, monuments lists on Wikipedia need to be updated manually every year in order to reflect mergers of municipalities); - (Semi-)automatic update of data in Wikidata from source databases (e.g. from the Protection of Cultural Properties (PCP) inventory and from the inventory of Swiss municipalities); - Feedback of errors to the maintainer of the source database; - Using data from Wikidata within Wikipedia; - Tackling the question of complementarity between LINDAS and Wikidata: to what extent does it make sense to mirror data that has been ingested into Wikidata also on LINDAS? (What is the scope, what are the notability criteria on LINDAS? What is the added value of LINDAS compared to Wikidata and vice versa?).
Parties involved / implementation status	The project is coordinated by the Wiki Loves Monuments International Team and Wikimedia Sweden. Volunteers in various countries are involved. Data providers for Switzerland are the Federal Office of Civil Protection, PCP service, as well as cantonal monuments protection offices (not all cantons make data available as open data though). The data provided by the Federal Office of Civil Protection has been ingested into Wikidata by a student from the Bern University of Applied Sciences.
Relation to other datasets	- Historicized inventories of municipalities and higher-level administrative-territorial units (in Switzerland and in other countries)
Case reports / documentation	- Ingesting Data about Italian heritage from WLM lists (Nicola Vitucci, Summer 2016) - Ingesting the Swiss Inventory of Cultural Properties (User:Affom, May 2017) - Video of conference presentation: Wikidata as the future of cultural heritage data – for Wiki Loves Monuments and beyond (Jean-Frédéric Berthelot, October 2017)

Applications making use of the data	- Monumental is a web-application for the display of information about cultural heritage monuments based on data gathered on Wikidata, Wikipedia, and Wikimedia Commons. It also provides an interface to edit data about monuments on Wikidata.
Contact person	Beat Estermann, (Bern University of Applied Sciences / OpenGLAM CH) beat.estermann@bfh.ch

6.3.3 Integrating Biodiversity-related Data: Red list of Endangered Species

Description	Data on an organism is always linked to a taxonomic name. This global, universal reference system has the potential to link the many isolated databases towards a seamless knowledge base. Taxonomic names are published by scientists including diagnostic data for their identification. With increasing knowledge, names can change. It is thus important to have access to a synonymic list allowing to follow the history of a name as well as the scientific taxonomic treatments delineating the case for change. This pilot demonstrates the use of LOD for a group of species listed in the Red List of threatened plants in Switzerland allowing querying the list of threatened species based on the scientific and vernacular names (German, Italian, French), their data in published Swiss Floras, and distribution records held in Swiss Natural History Museums and or Infoflora or the Centre Swiss de la Cartographie de la Faune (CSCF). The use case brings together three elements: government held data on species (red listed species; list of plant species in Switzerland), the taxonomic reference system, and content describing the species (taxonomic treatments). The various datasets are transformed into LOD and uploaded to the LINDAS Triple Store. Adding new data is an ongoing manual to fully automated process. New species described by Swiss scientists are added at the moment the publications are available. Among the main target users of the database are scientists, conservationists, administrators, journalists, and citizen scientists who need to be able to look up data about species under different name variants. As names do not stop in front of borders, the pilot study can easily be used beyond Switzerland.
Aspects covered / challenges raised	- Ingesting of data from the the published flora or other pertinent publications - Import TreatmentBank data into Triple Store

	- Triple store - Application - Modeling red list - Adopt treatmentOntology - Demonstrate value of time series / synonymies
Parties involved / implementation status	The parties involved in the implementation of the use case are Plazi, Factsmis-sion, and Zazuko. Data providers are BAFU (Redlist) and Infoflora (*List of plants of Switzerland*). TreatmentBank is an active producer of extracted data on species from the published literature. The data is already being served as LOD and held in a tri-ple store together with Zazuko.
Case reports / documentation	- Scientific Species Names in eCH/LINDAS
Contact person	Donat Agosti (Plazi), Reto Gmür (Factsmission)

6.3.4 linked.swissbib.ch

Description	The project linked.swissbib.ch aims to integrate the meta catalogue swissbib into Linked (Open) Data, by transforming, interlinking and enriching the data. Once the data deduplicated, transformed and interlinked, it is made available for computer clients and human users. Firstly, massive data reuse is possible for machines through a RESTful API. Since the library networks providing swissbib have various terms of use, this could lead to difficulty of attribution for the re-user. To address this issue, a mechanism filters the data in such a way that only CC0-compatible records are made accessible via the API. Secondly, an experimental interface is being developed for the end user, whose goal is to offer an improved search and exploration experience based upon the new in-terconnected data. Expected benefits of that approach include better data in-teroperability, easier data reuse and a more enriching user experience. The project linked.swissbib.ch (2014-2017) was realized within the CUS pro-gramme 2013-2016 P-2 "Scientific Information : access, processing and preservation".
Aspects covered / challenges raised	- Semantic integration of bibliographic metadata from a meta catalogue - Data modelling and transformation (MARC to RDF format) - Interlinking and enrichment with external corpora (VIAF, DBpedia, GND) - Developing a RESTful API with CC0-compatible (record) data - Enriching the user interface with new search and browsing functionali-ties

<table>
<tr><td>Parties involved / implementation status</td><td>
- Basel University Library (swissbib)
- University of Applied Sciences HEG Geneva
- University of Applied Sciences HTW Chur
- GESIS – Leibniz Institute for the Social Sciences

In April 2017 the beta versions have been released:

- GUI: linked.swissbib.ch
- RESTful API: data.swissbib.ch
</td></tr>
<tr><td>Case reports / documentation</td><td>
Documentation

Documentation of the above-mentioned services: https://linked-swissbib.github.io/datamodel/

Blog Series

Different aspects of the project have been described in four blog posts (available both in German and French):

- „Metadatentransformation, Modellierung, Indexierung“
- „Verlinkung und Anreicherung“
- „Präsentation der angereicherten Daten“
- „Hydra Web API for smarter clients“.

Publications

PRONGUÉ, Nicolas, SCHNEIDER, René, 2017. Data streams in linked.swissbib.ch: the Swiss metacatalog in the linked Open Data Cloud. In: Everything changes, everything stays the same? understanding information spaces: proceedings of the 15th International Symposium of Information Science (ISI 2017), Berlin, Germany, 13th-15th March 2017. Glückstadt: Verlag Werner Hülsbusch, 2017. S. 359-361.

BENSMANN, Felix, ZAPILKO, Benjamin and MAYR, Philipp (2017) Interlinking Large-scale Library Data with Authority Records. Front. Digit. Humanit. 4:5. doi: https://doi.org/10.3389/fdigh.2017.00005.

PRONGUÉ, Nicolas, SCHNEIDER, René, 2015. Modelling Library Linked Data in practice: three Swiss case studies. In: Re:inventing information science in the networked society: proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19.-21. Mai 2015. Glückstadt: Verlag Werner Hülsbusch, 2015. S. 118-128.

Posters

PRONGUÉ, Nicolas, SCHNEIDER, René, 2017. Data streams in linked.swissbib.ch: the Swiss metacatalog in the linked Open Data Cloud. International Symposium of Information Science (ISI 2017). Berlin. 13th-15th March 2017. Glückstadt: Verlag Werner Hülsbusch, 2017. S. 359-361.

PRONGUÉ, Nicolas, SCHNEIDER, René, 2016. Data flows in linked.swissbib.ch: from MARC to Linked Data. Kongress BIS. Luzern. 31. August - 3. September 2016.
</td></tr>
</table>

	Presentations SCHNEIDER René. Linked.swissbib.ch. Vortrag im Rahmen des Masterstudiengangs Management und Leadership im ABD-Bereich. Lausanne, 30.09.2017. SCHNEIDER, René. Linked.swissbib.ch. Vortrag mit Diskussion im Rahmen des SemanticWeb MeetUp Treffens. Zürich. 23. März 2017. https://www.meetup.com/de-DE/Zurich-Semantic-Web-Meetup-com/events/237607759/ SCHÜPBACH, Sebastian, 2016 Elasticsearch as Hub for Linked Bibliographic Metadata. Vortrag im Rahmen der Meetups von Elastic Switzerland Zürich. 31. August 2016. http://files.meetup.com/7646592/20160831%20Elasticsearch%20as%20Hub%20for%20Linked%20Bibliographic%20Metadata.pdf BENSMANN, Felix, HELLSTERN, Mara, KUNTSCHIK, Philipp, PRONGUÉ, Nicolas, 2016. Swissbib goes Linked Data. *Semantic web in libraries*. Bonn. 30. November 2016. SCHNEIDER, René, 2016. Teilnahme an der Panel-Diskussion "Experiences and challenges of Linked Data in the realm of libraries". In: VERBORGH, Ruben. Library science talk: "Linked happily ever after", Genf/Zürich, 5.-6. Dezember 2016.
Applications making use of the data	In April 2017, the test versions have been released: - GUI with new search and browse functionalities due to linked and enriched data: linked.swissbib.ch (beta) - RESTful API: data.swissbib.ch (beta)
Contact	- René Schneider, HEG Genève (project leader) - Günter Hipler, swissbib (technical project lead)

6.3.5 ld.geo.admin.ch

Description	ld.geo.admin.ch is the Linked Data service of the Swiss Federal Spatial Infrastructure. Most of the data produced today has a geographic component and geospatial data is an important part of the LOD Cloud. With the project ld.geo.admin.ch, swisstopo, as maintainer of the Federal Spatial Data Infrastructure, intends to unlock the potential of the geospatial data by publishing it on the web and according to the Linked Data principles. The service so far provides the dataset of the administrative units (Country, Cantons, Districts, and Municipalities); more datasets will be published in the future.
Aspects covered / challenges raised	- Data modelling and transformation (geospatial data to RDF format, GeoSPARQL) - Interlinking (Wikidata, data.admin.ch) - Versioning
Parties involved / implementation status	swisstopo The system is in production since March 2017
Case reports / documentation	**Documentation** https://www.geo.admin.ch/linkeddata **Presentations** Linked Data @ swisstopo. Linked Data Seminar. Amsterdam. 2.12.2016 FSDI and Linked Data. Linked Data Switzerland Workshop. Bern. 15.09.2016 Federal Spatial Data Infrastructure – Linked Data Service. Linked Data Switzerland Workshop. Bern. 20.09.2017
Contact	- Pasquale Di Donato, swisstopo - David Oesch, swisstopo

6.4 Further Case Reports

This section contains a list of further case reports documenting selected aspects of the publication of Swiss public-sector data as linked data:

- Municipality Index and Wikidata (Reto Gmür, 2017)
- Ingesting Data about Professional Theatre Troupes in Switzerland (User:Affom, May 2017)

7 Disclaimer / Reference to Third Party Rights

eCH-standards which the registered association **eCH** provides the user free of charge or which make reference to **eCH**, shall only have the status of recommendations. The registered association **eCH** will not be liable in any event for any decisions made or measures taken by the user based on these documents. The user will be responsible for verifying the documents himself prior to their use and to seek advice if required. **eCH**-standards can and shall not replace the technical, organizational or legal advice in the individual case.

Documents, procedures, methods, products and standards which are made reference to in **eCH**-standards are possibly protected by trademarks, copyrights or patents. It is the exclusive responsibility of the user to obtain the necessary licenses from the entitled persons and/or organizations.

Although the registered association **eCH** has taken adequate care to prepare the **eCH**-standards with due diligence, it cannot grant any warranty or guarantee that the information and documents provided are up-to-date, complete, true or without any errors. **eCH** reserves the right to change the contents of the **eCH**-standards at any time and without prior announcement.

Any liability for damage caused by the use of the **eCH**-standards by the user shall be excluded to the extent legally admissible.

8 Copyright

Persons preparing **eCH**-standards shall remain the owners of their intellectual property rights. These persons, however, obligate themselves to provide their intellectual property rights or other rights in third party intellectual property rights, to the extent possible, to the relevant technical units and the registered association **eCH** for free and for unlimited use and further development as part of the purpose of the association.

The standards prepared by the technical units can be used, distributed and developed further for free and to an unlimited extent by stating the name of the respective author(s) and of **eCH.**

eCH-standards are fully documented and free of any restrictions of licence and/or patent law. The associated documentation can be requested for free.

These provisions shall apply to the standards prepared by **eCH** only, however, not to any standards or products of third parties which include reference to **eCH**-standards. The standards include the relevant references to third party rights.

Annex A – References & Bibliography

Auer, S., Bühmann, L., Dirschl, C., Erling, O., Hausenblas, M., Isele, R., ... & Stadler, C. (2012). Managing the life-cycle of linked data with the LOD2 stack. In International semantic Web conference (pp. 1-16). Springer, Berlin, Heidelberg.

Berners-Lee, T. (2006-2009). Linked data-design issues.
Retrieved from: http://www.w3.org/DesignIssues/LinkedData.html

Bikakis, N., & Sellis, T. (2016). Exploration and visualization in the web of big linked data: A survey of the state of the art. arXiv preprint arXiv:1601.08059.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. Semantic services, interoperability and web applications: emerging concepts, 205-227.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an Ontology?. In Handbook on ontologies (pp. 1-17). Springer Berlin Heidelberg.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology, 1(1), 1-136.

Hyland, B., Atemezing, G., & Villazón-Terrazas, B. (2014). Best Practices for Publishing Linked Data. W3C Working Group Note, 09 January 2014. *2014-09-04.*
Retrieved from: https://www.w3.org/TR/ld-bp/

Raymond, E. S. (2001). *The Cathedral & the Bazaar: Musings on linux and open source by an accidental revolutionary*. " O'Reilly Media, Inc.".

Vrandecic, D. (2010). Ontology Evaluation. Doctoral Thesis.
Retrieved from: https://pdfs.semanticscholar.org/7d97/25cd4367979e56a2f89539d550071eb25f3e.pdf

Wickham, H. (2014). Tidy data. Journal of Statistical Software, 59(10), 1-23.

Annex B – Cooperation & Verification

Main author:

- Beat Estermann (Bern University of Applied Sciences)
  https://orcid.org/0000-0003-3050-793X

Further contributors and reviewers (in alphabetical order):

- Donat Agosti (Plazi, Bern)
- Fabian Cretton (HES-SO Valais/Wallis, Data Semantics Lab)
- Jean-Luc Cochard (Swiss Federal Archives)
- Reto Gmür (Factmission)
- Mara Hellstern (HTW Chur)
- Daniel Hladky
- Andy Mabbett
  https://orcid.org/0000-0001-5882-6823
- Elena Mastrandrea (HTW Chur)
- Markus Pilzecker (Die Wissens-Ingenieure)
- Nicolas Prongué (HEG Genève)
- Alex Stinson (Wikimedia Foundation)
- Birk Weiberg (Swiss Archive for the Performing Arts)

Annex C – Abbreviations and Glossary

Below is a list of abbreviations used in this document:

Abbreviation	In full (with explanation)
API	“Application Programming Interface”, interface between the application software and the application platform, across which all services are provided.
CC0	“Creative Commons Public Domain Dedication”. A standard copyright waiver provided by Creative Commons[33].
CMS	“Content Management System”, a computer application that supports the creation and modification of digital content; it typically supports multiple users in a collaborative environment.
CSV	“Comma Separated Values”, a data interchange format which stores tabular data in plain text.
DBpedia	DBpedia ("DB" for "database") is a crowdsourced community effort to extract structured information from Wikipedia and make this information available on the Web. DBpedia allows users to semantically query relationships and properties of Wikipedia resources, including links to other related datasets. DBpedia was initiated in 2007 by people at the Free University of Berlin and Leipzig University, in collaboration with OpenLink Software. It thus predates Wikidata, the structured data project of the Wikimedia community, by five years, and has had an important role in the development of linked open data.
DCAT	“Data Catalog Vocabulary” is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.
GLAM	“Galleries, Libraries, Archives and Museums”. The acronym is used to refer to heritage institutions in general.
GND	“Gemeinsame Normdatei” (Integrated Authority File), an international authority file for personal names, subject headings and corporate bodies. It is used mainly for documentation in libraries and increasingly also by other types of heritage institutions. The GND is managed by the German National Library in cooperation with various regional library networks in German-speaking Europe and other partners.

[33] https://creativecommons.org/publicdomain/zero/1.0/

GUI	“Graphical User Interface”, a type of user interface that allows users to interact with electronic devices through graphical icons and visual indicators, instead of text-based user interfaces, typed command labels or text navigation.
HTML	“Hypertext Markup Language”, the standard markup language for creating web pages and web applications.
HTTP	“Hypertext Transfer Protocol”, an application protocol for distributed, collaborative, and hypermedia information systems; it provides the foundation of data communication for the World Wide Web.
HTTPS	A secure from of HTTP
JSON-LD	“JavaScript Object Notation for Linked Data”, a method of encoding Linked Data
LOD	“Linked Open Data”, linked data that is “open” according to the Open Definition[34], meaning that it may be freely used by anyone for any purpose.
LODLAM	“Linked Open Data for Libraries, Archives, and Museums”
MARC	“MAchine-Readable Cataloguing”, a set of standards for the description of items catalogued by libraries, such as books. The standards are published by the Library of Congress.
OAI-PMH	“Open Archives Initiative” – “Protocol for Metadata Harvesting”, a protocol used for the harvesting of descriptions of records in an archive so that services can be built using metadata from many archives.
OpenGLAM	OpenGLAM is an initiative run by the Open Knowledge Foundation that promotes free and open access to digital cultural heritage held by heritage institutions.
OWL	“Web Ontology Language”, a family of knowledge representation languages for authoring ontologies.
RDF	“Resource Description Framework”, a family of W3C specifications used for the formal representation and exchange of linked data.
RDFa	“Resource Description Framework in Attributes”, a W3C Recommendation that adds a set of attribute-level extensions to HTML, XHTML and various XML-based document types for embedding rich metadata within Web documents.

[34] http://opendefinition.org/

RDFS	"RDF Schema", a set of classes with certain properties using the RDF extensible knowledge representation data model, providing basic elements for the description of ontologies.
RESTful	"Representation state transfer" (REST). RESTful APIs are a way of providing interoperability between computer systems on the Internet. They allow requesting systems to access and manipulate textual representations of Web resources using a uniform and predefined set of stateless operations.
SHACL	"Shapes Constraint Language", a language for validating RDF graphs against a set of conditions. T conditions are provided as shapes and other constructs expressed in the form of an RDF graph.
SPARQL	"Simple Protocol and RDF Query Language", the language used to define queries to a datastore of linked data.
URI	"Uniform Resource Identifier", a string of characters used to identify names or resources on the Internet.
VIAF	"Virtual International Authority File", an international authority file resulting from a joint project of several national libraries to interlink their national authority files. Since 2012, the service is operated by the Online Computer Library Center (OCLC).
W3C	"World Wide Web Consortium", the main international standards organization for the World Wide Web.
XHTML	"Extensible Hypertext Markup Language", an XML markup language adapting the widely used Hypertext Markup Language to XML.
XML	"Extensible Markup Language", a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable; it is widely used for the representation of arbitrary data structures.

Annex D – Changes in Comparison to the Previous Version

This is the first version.

Annex E – Table of Figures