Study in View of the Further Development of DCAT-AP CH

Final Report

Bern University of Applied Sciences
Stephan Haller, Beat Estermann, Angelina Dungga-Winterleitner

The Study has been mandated by the Swiss Federal Archives.

Version 1.0
20 December 2018

Management Summary

The Swiss application profile of DCAT (DCAT- AP CH) is the Swiss metadata standard used to facilitate cross-platform search and exchange of catalog data between different data portals in Switzerland. It is a national derivation of the European application profile of the Data Catalog Vocabulary (DCAT-AP) and the metadata standard currently in use on the Swiss Open Government Data Portal opendata.swiss.

On behalf of the Swiss Federal Archives, the Bern University of Applied Sciences has conducted a study concerning further development of the Swiss application profile of DCAT, the results of which are presented in this report. The study considers both the developments at the international level and the needs and requirements of the main stakeholder groups in Switzerland. It is based on desk research, expert interviews with representatives of key stakeholder groups, and input gathered during a meeting with the eCH Specialized Group "Open Government Data", who is responsible for the maintenance of DCAT-AP CH.

As the study has shown, the future development of DCAT-AP CH needs to be considered in the context of an interplay both between data portals and between the various derivations of DCAT that are in use on these portals. There is presently a cascade of data portals in place: While the European Data Portal harvests and aggregates data from national portals, they in turn aggregate data from lower-level administrative-territorial entities as well as from specialized portals, which may cater to more specific user needs. A similar situation exists with regard to the metadata standard: While the W3C Recommendation DCAT represents the smallest common denominator on a global scale, the application profiles at the various levels (European profile, national profiles) take into account the specificities of the respective level and serve as a laboratory for innovations. A permanent process of innovation and harmonization thereby ensures that the different derivations of DCAT remain in sync: Novelties that have been successfully introduced at lower levels are considered for adoption at higher levels, and lower level derivations are regularly adapted to the higher-level standards. In addition to the derivations at various levels, specialized application profiles of DCAT have been created, such as StatDCAT-AP, which standardizes the description of statistical data.

What also needs to be taken into account when deciding about the future development of the data portals and the corresponding metadata standards, is the inexorable trend towards the publication of linked open data: While the data portals, the metadata standard, and auxiliary documents should be designed to support the transition to linked open data, particular care should be taken to avoid addressing issues at the level of the data portals that are best resolved by publishing data as linked open data. This said, not all data will be published as linked open data at once; there will be a long transition period during which the provision of ordinary tabular data continues to be the norm.

The main challenges that should be tackled within the Swiss open data portal ecosystem comprise issues related to metadata quality, the lacking support of standard licenses on the Swiss Open Government Data Portal, issues related to data quality in view of better interoperability of datasets, multilingual support, as well as issues related to the representation of geographical data. While some of these issues can indeed be addressed by expanding or complementing the metadata standard, others need to be addressed at the level of the data portal or by the data providers themselves.

Consequently, recommended actions are formulated not only with regard to the future development of DCAT-AP CH, but also with regard to the future development of the Swiss Open Government Data Portal and the deployment of linked open data by public sector organizations, as they need to go hand in hand:

The main tasks in view of the further development of DCAT-AP CH by the eCH Specialized Group "Open Government Data" comprise the active involvement in international standardization activities in order to ensure the highest possible degree of interoperability at the international level; the agreement on controlled vocabularies and their publication as linked open data; the creation of a convention handbook to support the harmonization of practices when it comes to describing datasets; the creation of non-normative, low-threshold communication material about the standard to support the

sensitization of stakeholders in Switzerland; as well as the short- and long-term development of the Swiss application profile.

At the same time, the provider of the Swiss Open Government Data Portal should ensure full support of DCAT-AP CH and shift their focus on data quality instead of data quantity, which includes putting in place effective feedback loops with data users. All three, the eCH Specialized Group "Open Government Data" as well as the Swiss Confederation's Open Government Data and Linked Data projects, need to closely cooperate to support the progressive transition to linked open data. They also need to develop a shared vision of the intended pace of innovation and their role within the international context, as the further development of DCAT-AP CH should be closely aligned with the future development of the Swiss Open Government Data Portal.

Contents

1   Introduction

The Swiss Open Government Data Portal (opendata.swiss), which is managed by the Swiss Federal Archives, employs a data model that is derived from the European application profile of the Data Catalog Vocabulary (DCAT-AP). This metadata standard plays a crucial role in facilitating a cross-platform search and exchange of catalog data between different data portals – both within Switzerland and when aggregating data in the European Data Portal[1].

During the implementation of the Swiss Open Government Data Portal a few years ago, a Swiss derivation of the European application profile DCAT-AP was developed. This Swiss application profile of DCAT (DCAT- AP CH)[2] went through the eCH[3] standardization process and was officially recognized as a Swiss E-Government standard in June 2018.

Since the launch of opendata.swiss, there have been various innovations regarding the European application profile of the Data Catalog Vocabulary. This raises the question as to how the Swiss derivation of the metadata standard should be adapted in the future.

The following developments are relevant:

-    In autumn 2015, version 1.1 of the European application profile DCAT-AP was published[4], and in autumn 2018, version 1.2 was released[5].

-    In different EU member states, various derivations of DCAT-AP, similar to the Swiss application profile, have emerged. Some contain extensions to the original standard, such as additional data fields or controlled vocabularies[6].

-    In 2016, specialized application profiles were published in the fields of geodata and statistical data (GeoDCAT-AP[7], StatDCAT-AP[8]).

Against this backdrop, the Swiss Federal Archives have mandated the Bern University of Applied Sciences to carry out a study that provides an overview of the expectations and needs of Swiss stakeholders concerning further development of the metadata standard. In addition, the study is expected to identify trends at the international level regarding standardization activities around DCAT, and to suggest possible directions for the further development of the Swiss derivation of the standard.

---

[1]   https://www.europeandataportal.eu/
[2]   eCH-0200 DCAT-Anwendungsprofil für Datenportale in der Schweiz (DCAT-AP CH)
[3]   eCH is the standardization organization for e-government standards in Switzerland.
[4]   https://joinup.ec.europa.eu/release/dcat-ap-v11
[5]   https://joinup.ec.europa.eu/release/dcat-ap/12
[6]   See for example the German application profile: http://www.dcat-ap.de/def/dcatde/1_0/spec/specification.pdf
[7]   https://joinup.ec.europa.eu/release/geodcat-ap/v101
[8]   https://joinup.ec.europa.eu/release/statdcat-ap-v100

2 Research Questions and Methodology

The study addresses the following questions:

- What development trends have emerged over the past years regarding standardization activities around DCAT-AP?

- What elements of the new or revised application profiles (DCAT AP Version 1.1, GeoDCAT-AP 1.0, StatDCAT-AP 1.0) are of relevance when it comes to ensuring interoperability between the Swiss Open Government Data Portal and related data portals (e.g. geocat.ch, data portals of statistical services, European Data Portal)?

- Which enhancements to the metadata standard and/or its implementation on the Swiss Open Government Data Portal are required by portal users from various domains (geodata, statistical data, cultural data)?

- What are the experiences of portal operators in Switzerland regarding the use of the Swiss application profile of DCAT? – Are there any data fields that are used inconsistently by data suppliers? Is the standard applied in the same manner across different portals? Is there any need for improvement to the standard or its implementation?

- What synergies would result from closer cooperation with other countries? – To what extent could their standards (e.g. the German derivation of DCAT-AP) be directly adopted by the Swiss community? To what extent could Swiss players directly contribute to the further development of the European application profile and subject-specific application profiles? Would it be conceivable to rely on one of the existing standards and to abandon the maintenance of a separate Swiss application profile in the future?

- In reference to Tim Berners-Lee's five-star open data model[9], what contribution can be made using standardization activities to increase the maturity level of the data referenced on open data portals?

- What role should the Swiss application profile of DCAT play? What use case should be focused on in its further development?

A three-step process was used to address these questions:

1. The development trends relevant to the further development of DCAT-AP CH were analyzed employing desk research. Based on this first analysis, interview questionnaires were developed that were used in the second step.

2. Fifteen expert interviews were carried out with representatives of three different stakeholder[10] groups:

   a. Persons/organizations playing a leading role in the development of relevant standards (DCAT-AP, DCAT-AP DE, GeoDCAT, StatDCAT);

   b. Providers of data portals (e.g., opendata.swiss; geocat.ch; OGD[11] City of Zürich; Federal Statistical Office; cantonal statistical offices, providers of cultural data platforms);

   c. Representatives of specialist communities (e.g., geodata, statistical data, cultural data).

   The expert interviews were analyzed, and the results summarized in a preliminary study report.

---

[9] http://5stardata.info
[10] The interviewees comprised representatives of five data portal providers from Switzerland (most of which also act as data providers). They included two data portal providers from other countries, one portal developer (and user of open data), one linked data specialist (and user of open data), two stewards of specialized application profiles, three data providers, and one data user.
[11] "OGD" stands for "Open Government Data".

3. The preliminary results of the expert interviews were discussed among the eCH Specialized Group "Open Government Data", which provided feedback as to how to further develop DCAT-AP CH. The group is responsible for the maintenance of the Swiss application profile of DCAT.

This report includes the combined findings from all three steps. While sections 3 to 6 mainly draw on the results based on desk research, complemented by additional insights gained through the interviews, sections 7 to 9 mainly draw on the results of the expert interviews, complemented by the input received during the workshop with the eCH Specialized Group OGD. The final two sections provide a synthesis and a discussion of the results, along with the conclusions and recommended actions.

3   Data Portals

DCAT and its derivations, such as DCAT-AP, DCAT-AP CH, or GeoDCAT are used in the context of open data metadata portals, such as the open data portal of the city of Zurich, opendata.swiss, geo-cat.ch, or the European Data Portal. These portals provide a catalogue of datasets that are hosted on other platforms to facilitate searching and discovering reusable datasets across platforms. To assemble the catalogue, metadata about the various datasets are aggregated on the platform either manually or automatically through APIs, using either push or pull methods. Data catalogues at various levels harvest data from each other. For example, a subset of the datasets referenced in geocat.ch and the datasets referenced by the data portal of the city of Zurich are also referenced in the Swiss Open Government Data Portal, while the metadata of the Swiss Open Data Portal is in turn harvested and made available by the European Data Portal. Thus, some metadata portals have the function of aggregators in relation to other data portals (see figure 1). The DCAT standard and its derivations therefore not only facilitate the exchange of metadata between data providers and open data platforms but also between open data platforms at various levels.
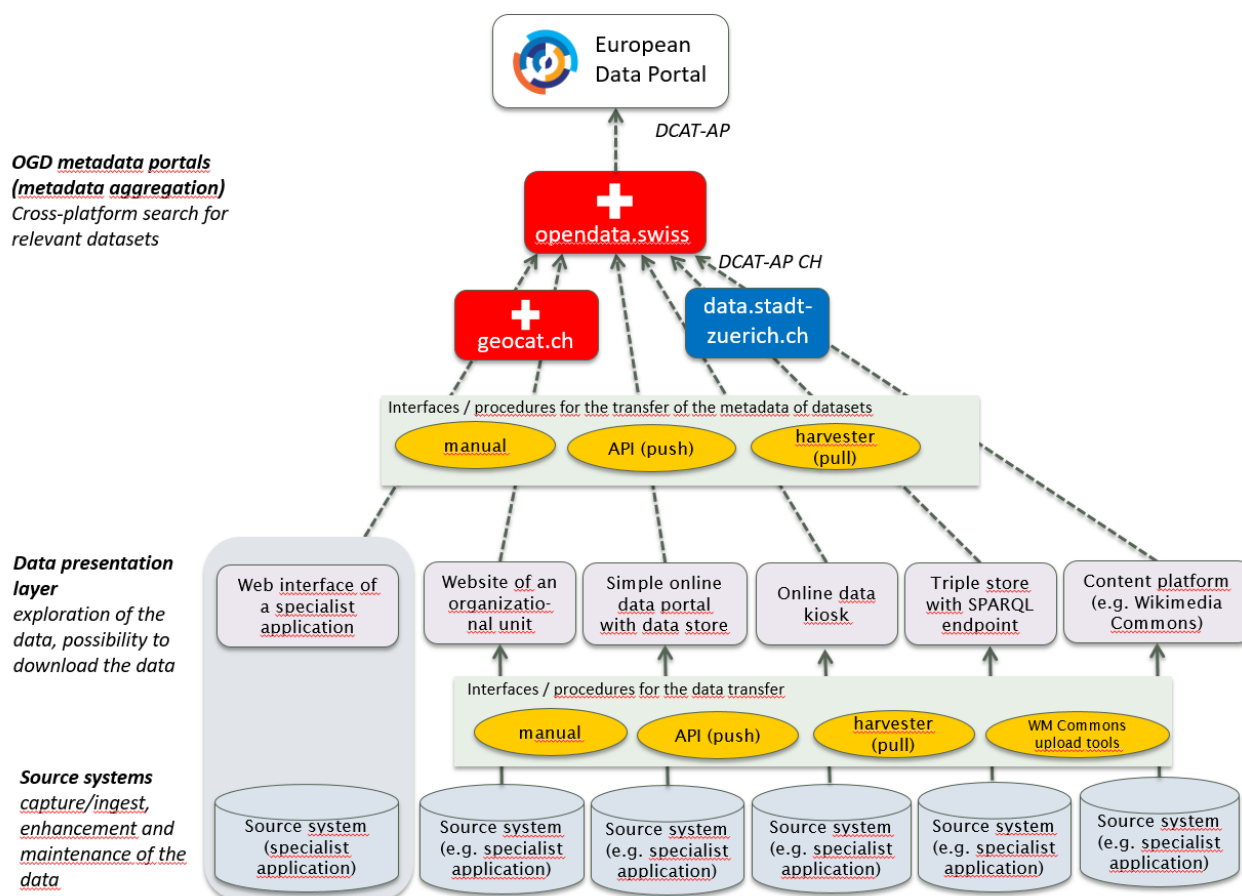
Figure 1: The Swiss Open Government Data Portal (opendata.swiss) in its context

The Swiss metadata portals most frequently mentioned during the expert interviews included opendata.swiss, opentransportdata.swiss and the respective portals of the Swiss Federal Railways, Swisscom, the Swiss Post, and the City of Zurich. Most interviewees responsible for publishing data and metadata on their organization's data portal reported that opendata.swiss automatically harvests their metadata. In one case, the metadata is entered directly on opendata.swiss, and in another case, the metadata is first harvested by geocat.ch and from there transferred to opendata.swiss.

Opendata.swiss was launched in February 2016 and has since seen a yearly 15 percent increase in usage. However, usage levels remain relatively low, and the use of open data portals is perceived as a niche activity, rather than an activity for the masses.

Interviewees use metadata portals for several reasons. Most often they browse the portals to get an overview of published data or to check the visibility of the data they have published themselves. Other usages include the building of an open data community, converting data to linked data, data retrieval, data publishing, or data-harvesting for further transmission to other portals. The portals are also used as references or practical examples for the development of other portals.

3.1 The Swiss metadata portals in their context

Figure 1 illustrates the Swiss open data metadata portals in their context. Note that data providers are responsible for the provision of the data on a website or platform where it can be explored and down-loaded. In some cases, the specialist applications used to manage the data come already equipped with a web interface. In many cases, however, data providers need to transfer data from a source system to a data portal or website to ensure that the data is made available to re-users. These data presentation layers vary in their level of sophistication, ranging from simple websites to online data kiosks[12] to triple stores with SPARQL endpoints.

If such data portals contain many datasets that are referenced on an open data metadata portal, they are usually equipped with an API, so that the metadata can be transferred automatically to the metadata portal. In this case, the data portal needs to provide the metadata using a metadata-standard that is compatible with the metadata-standard used by the metadata portal.

The interviews conducted in this study have shown that the DCAT-AP CH-Standard is not locally used by any other portal than opendata.swiss. Data providers base their metadata descriptions on other standards, mostly on an individually adapted version of an international standard such as ISO-19139 or INSPIRE. Geocat.ch, for example, bases its metadata description on the model GM-03, a model based on ISO and INSPIRE, and developed by Swisstopo. The individual standards or models used for metadata description are then mapped to DCAT-AP CH to allow for data harvesting.

3.2 Various types of data referenced on open data metadata portals

The datasets currently referenced on Swiss open data metadata portals can roughly be assigned to one or several of the following types, which come with different requirements as to how they are presented and made available for re-use:

- **Statistical data** are quantitative data about various phenomena and topics.

- **Geospatial data** are data related to a geographical location or data defining such locations (e.g. shape files, etc.).

- **Base registers** are databases "that are legally controlled and maintained by public administrations and provide authentic sources of information on items such as persons, companies, buildings, or roads" (Estermann et al. 2018). They are often referenced in other datasets. In the heritage domain, **authority files** fulfill the same function, but as they relate to historical data, authority may be established through agreement by the heritage community and not through legislation.

- **(Quasi-)Real-time data** derive their use value from their availability immediately after their time of capture or generation.

- **Calendar data** are related to events taking place on a specific date and/or time. For planned or predicted events, they derive their main use value from their availability ahead of the event.

- **Object metadata, collection metadata, and archival finding aids** describe digital or analog artefacts and the record sets or collections they are part of. They are typically found in heritage institutions but may also be present in other organizations which manage text documents

---

[12] The term «data kiosk» in the context of open data refers to online data platforms which allow users to explore and visualize data before selecting a specific dataset or a subset thereof for download.

or photographic or audiovisual content. The metadata about datasets contained in open data catalogs also falls into this category.

- **Digital content** refers to digital-born text, photographic or audiovisual content or to digitized versions of content that has originally been produced in an analog format. It accounts for a large part of the data held by heritage institutions. Other organizations may also make digital content available for re-use. Digital content is typically accompanied by object metadata that provides further information about the content and its provenance. In many cases, digital content is covered by copyright.

3.3 Data maturity model

The data provided through open data portals can be of varying maturity. Data maturity can be rated with reference to Tim Berners-Lee's 5-star Open Data Maturity Model, which considers the publication of data in form of linked open data as the highest maturity level (see figure 2):

- **One-star data** is available on the Web (in whatever format) under an open license.

- **Two-star data** is additionally available as structured data (e.g. in tabular form instead of an image scan of a table).

- **Three-star data** is additionally available in a non-proprietary open format (e.g. CSV).

- **Four-star data also** contains 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 additionally linked to other data sources on the Web to provide context.
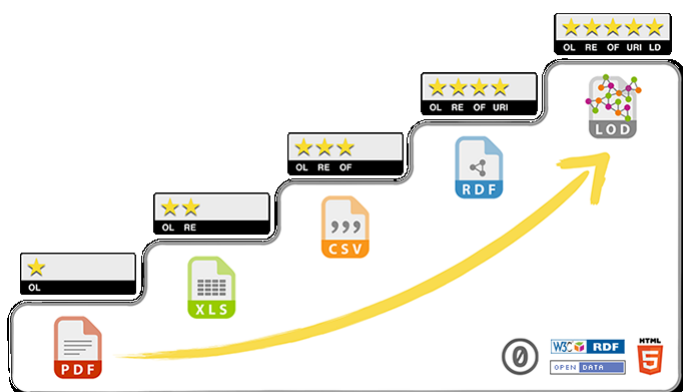
Figure 2: Tim Berners-Lee's 5-star Open Data Maturity Model

Independently of the maturity level of the data, the data may be made available in the form of downloadable data dumps or through an API, where third-party services can retrieve specific data entries. An intermediate approach consists in the provision of a data kiosk, where data can be explored and filtered before selecting data (e.g. a subset of a larger dataset) to download.

3.4 Frictionless data

Parallel to striving towards linked open data, there is also an initiative to simplify the shared use of tabular data (three-star data), called "Frictionless Data", which provides specifications and software to this effect[13]. "Frictionless Data" is expected to make it easier for data analysts, developers, or even automated agents to quickly notice whether datasets are compatible with each other[14]. The community around "Frictionless Data" also maintains a series of datasets with "core data"[15]. – Core data, according to their definition, are data which are often used in the context of other datasets (e.g. lists of countries, populations, geographic boundaries and more).

---

[13] https://frictionlessdata.io

[14] https://frictionlessdata.io/articles/oleg-lavrovsky/

[15] https://datahub.io/blog/core-data-essential-datasets-for-data-wranglers-and-data-scientists

3.5 International Image Interoperability Framework (IIIF)

The concept of linked data and the 5-star Open Data Maturity Model are not directly applicable to digital content. The International Image Interoperability Framework (IIIF) defines, however, application programming interfaces that provide a standardized method for describing and delivering images over the web. Similar to the linked data approach, this standard supports a decentralized platform architecture where content is made available by various servers from where it can be dynamically integrated into various online services provided by clients. While IIIF programming interfaces for image content have existed for several years, the IIIF standard for audiovisual content is currently under development; version 1.0 has been announced for the end of 2018[16].

3.6 Data platform maturity

When considering data platforms in a cross-organizational context, the following maturity model can be applied (figure 3; see Estermann et al. 2017):

1. **Data silos:** each organization makes their data available on their data platform.

2. **Federated search:** a central platform offers a federated search across several data platforms based on a shared data model that is limited to a few data fields.

3. **Central data aggregation:** metadata is aggregated on a central platform which provides a search option across all the aggregated metadata; usually, a more sophisticated data model is used in comparison to a federated search, and thanks to the aggregation of the data on one platform, more sophisticated search functionalities (like facetted search) can be offered.

4. **Automatized central data aggregation:** the same as central data aggregation; the metadata is however aggregated through automatized methods (e.g. data harvesting).

5. **Linked data:** data is kept in decentralized databases; thanks to a shared data model and inter-linking of data, the data can be queried and exploited across servers.

While the linked data approach supports a decentralized architecture, it requires centrally agreed upon data models that are shared among many data providers. Thus, some linked data services rely on central data aggregation to achieve the cross-platform harmonization of data models before exposing data as linked data (e.g. Wikidata, Europeana), while others take a more decentralized approach (e.g. the LINDAS linked data platform where each graph functions like a data silo until it is linked to other graphs, thanks to shared ontologies).
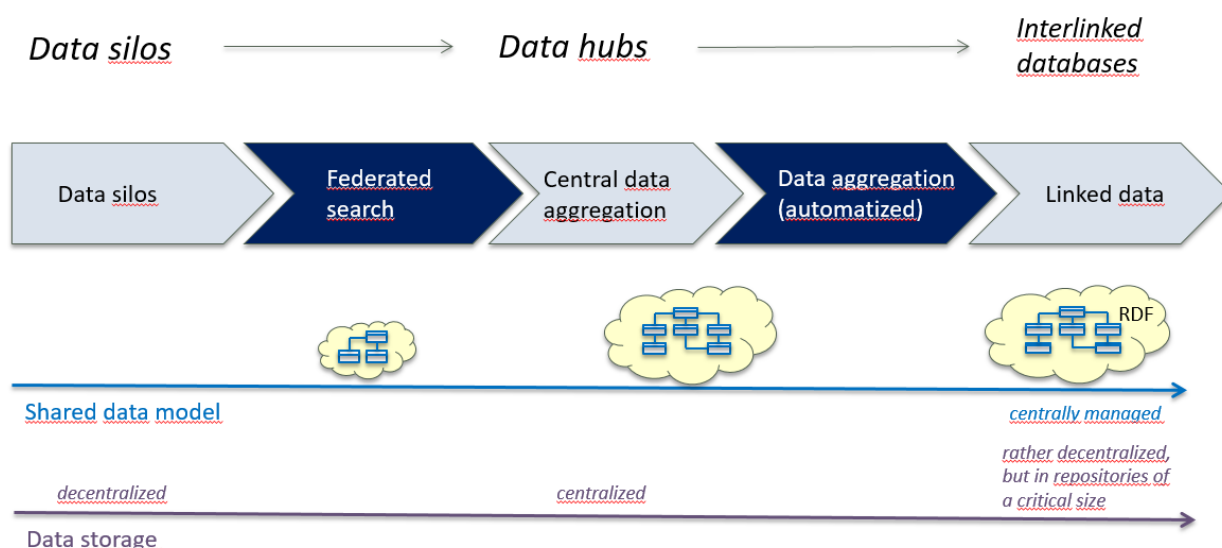
Figure 3: Maturity model of heritage data platforms holding object metadata (Estermann et al. 2017)

This data platform maturity model has been developed in the context of heritage data platforms. It can be applied *mutatis mutandis* to open data platforms. Hereby, it is interesting to note that the Swiss open data metadata portals employ the (automatized) central data aggregation approach, but data aggregation is limited to the metadata only. Therefore, open data metadata portals do not currently support data integration at the dataset level. Moreover, they do not seem to be designed for this. Data integration at the dataset level is however taking place on dedicated linked data platforms (e.g., LINDAS triple store, Wikidata; see figure 4).
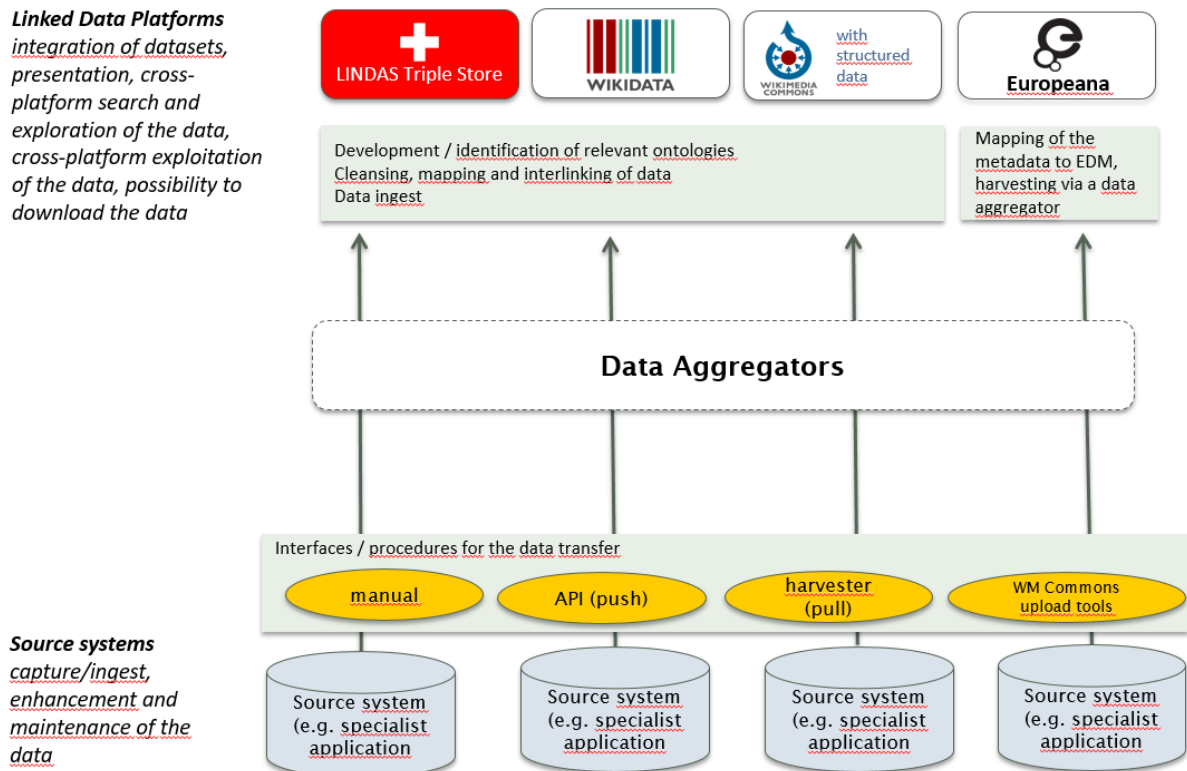
Figure 4: Provision of linked data

4 Overview of Development Trends in the Context of DCAT

This section identifies the development trends that should be considered when planning further development of DCAT-AP CH and examines the role the Swiss derivation of DCAT is to play in the context of related standards.

We first describe the changes in the European application profile DCAT-AP made in version 1.1. compared to version 1.01, and the minor change in version 1.2. The changes which are particularly relevant for the Swiss derivation are highlighted. We then look at the situation in various European countries where national derivations of DCAT-AP have been created and provide an overview of extensions in GeoDCAT-AP and StatDCAT-AP that might be of relevance. We conclude by relating how the interlocutors from our expert interviews conceive the role of the Swiss derivation of the standard and its interplay with the main standard and its derivations at various levels.

4.1 Novelties of version 1.1 of the European application profile DCAT-AP

Version 1.1 of the European application profile DCAT-AP was published in autumn 2015 (cf. PwC EU Services 2017). It includes several changes to its previous version, 1.01 (cf. PwC EU Services 2013). An overview of these changes is provided in Figure 2, while the tables in sections 4.1.1 and 4.1.2 list the changes in detail.

Note that the DCAT-AP standard defines the required level of all classes and properties and distinguishes between elements that are *mandatory*, *recommended* or *optional*. This requirement level relates primarily to the data supplier, e.g., elements marked as *mandatory* must be supplied by the data supplier. Data receivers, i.e., data portals, only have to be able to process such elements. The term "processing" means that "receivers must accept incoming data and transparently provide these data to applications and services. It does neither imply nor prescribe what applications and services finally do with the data (parse, convert, store, make searchable, display to users, etc.)".[17]

The Swiss extension also uses the term *conditional*: elements that are mandatory under certain conditions.

4.1.1 Changes in V1.1 on Class level

Class name	URI	Change
Category	`skos:Concept`	*mandatory → recommended*
Category scheme	`skos:ConceptScheme`	*mandatory → recommended*
Licence document	`dct:LicenseDocument`	*optional → recommended*
Checksum	`spdx:Checksum`	New *optional* class
Identifier	`adms:Identifier`	New *optional* class
Kind	`vcard:Kind`	Replaces `v:VCard`
Provenance Statement	`dct:ProvenanceStatement`	New *optional* class

4.1.2 Changes in V1.1 on Property level

`dcat:Catalog`			
Property	URI	Range	Change
has part	`dct:hasPart`	`dcat:Catalog`	New *optional* property
is part of	`dct:isPartOf`	`dcat:Catalog`	New *optional* property

---

[17] cf. DCAT-AP V1.1 (PwC EU Services 2017), p. 7
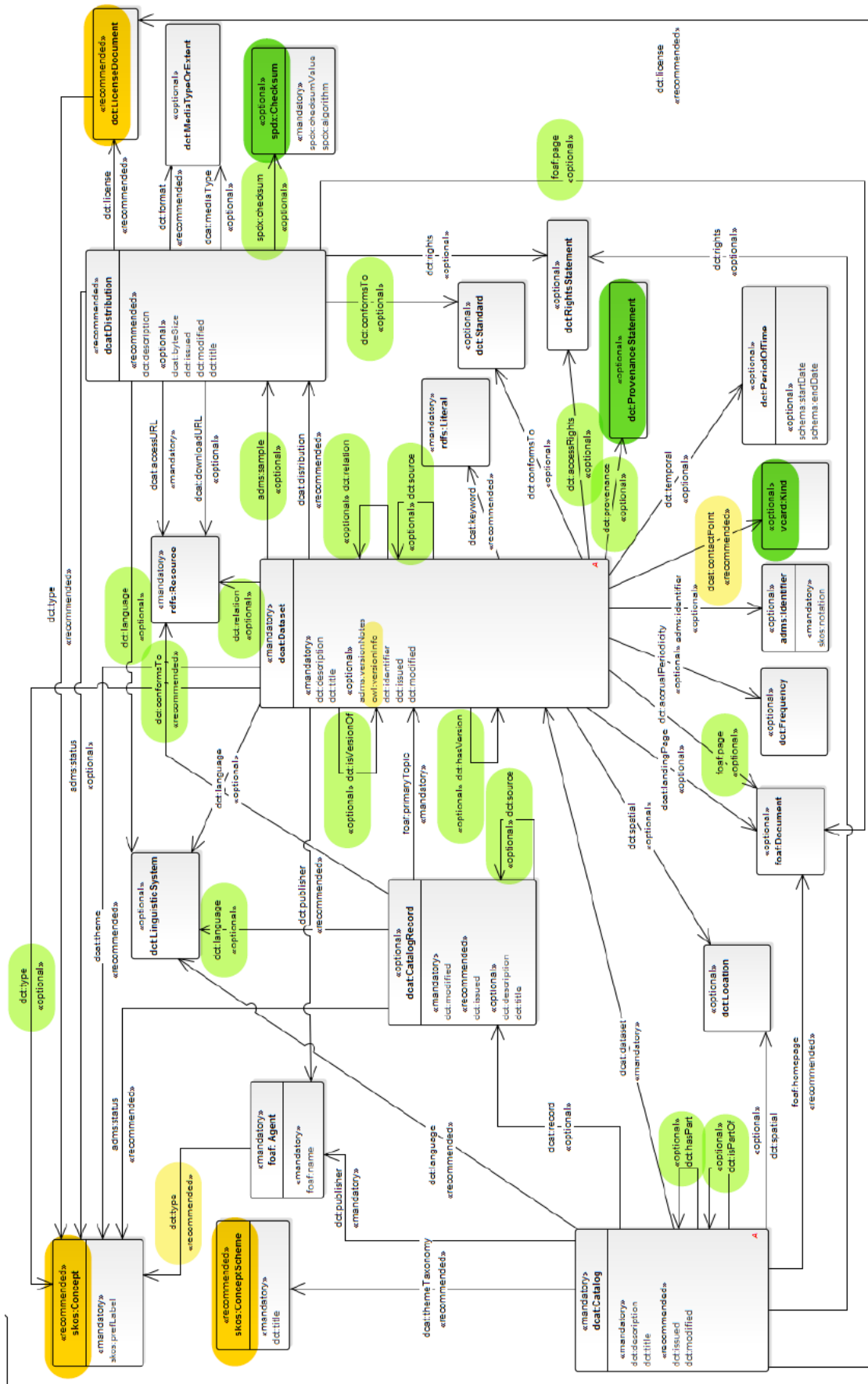
Figure 2: Overview DCAT-AP Changes in V1.1

`dcat:CatalogRe-cord`			
Property	URI	Range	Change
application profile	`dct:conformsTo`	`rdfs:Resource`	New *optional* property
language	`dct:language`	`dct:LinguisticSystem`	New *optional* property
source metadata	`dct:source`	`dcat:CatalogRecord`	New *optional* property

`dcat:Dataset`			
Property	URI	Range	Change
contact point	`dcat:contactPoint`	`vcard:Kind`	Changed URI and Range (`adms:contactPoint`, `v:vCard`)
access rights	`dct:accessRights`	`dct:RightsStatement`	New *optional* property
documentation	`foaf:page`	`foaf:Document`	New *optional* property
has version	`dct:hasVersion`	`dcat:Dataset`	New *optional* property
is version of	`dct:isVersionOf`	`dcat:Dataset`	New *optional* property
Provenance	`dct:provenance`	`dct:ProvenanceStatement`	New *optional* property
related resource	`dct:relation`	`rdfs:Resource`	New *optional* property
Sample	`adms:sample`	`dcat:Distribution`	New *optional* property
Source	`dct:source`	`dcat:Dataset`	New *optional* property
Type	`dct:type`	`skos:Concept`	New *optional* property
Version	`owl:versionInfo`	`rdfs:Literal`	New URI (previously `adms:version`)

`dcat:Distribution`			
Property	URI	Range	Change
checksum	`spdx:checksum`	`spdx:Checksum`	New *optional* property
documentation	`foaf:page`	`foaf:Document`	New *optional* property
language	`dct:language`	`dct:LinguisticSystem`	New *optional* property
linked schemas	`dct:conformsTo`	`dct:Standard`	New *optional* property

`dcat:Agent`			
Property	URI	Range	Change
type	`dct:type`	`skos:Concept`	Name change (previously "publisher type")

`spdx:Checksum` *(NEW CLASS)*			
Property	URI	Range	Cardinality
algorithm	`spdx:algorithm`	`spdx:checksumAlgorithm_sha1`	*mandatory* 1..1
checksum value	`spdx:checksumValue`	`rdfs:Literal` typed as `xsd:hexBinary`	*mandatory* 1..1

`adms:Identifier` *(NEW CLASS)*			
Property	URI	Range	Cardinality
notation	`skos:notation`	`rdfs:Literal` typed with the URI of one of the members of the DataCite Resource Identifier Scheme	*mandatory* 0..1

`dct:LicenceDocument`			
Property	URI	Range	Change
licence type	`dct:type`	`skos:Con-cept`	different Range (previously `rdfs:Class`)

4.2 Novelties of version 1.2 of the European application profile DCAT-AP

Version 1.2 of the European application profile was published in November 2018. Compared to version 1.1 only one further change was made:

`dct:Licence-Document`		
Property	URI	Change
licence type	`dct:type`	Cardinality: 0..1 → 0..n This property can be repeated for multiple licences.

This is only a minor change, but it should be reflected in the Swiss application profile as well. According to current information, the next major release of DCAT-AP is scheduled for November 2019.

4.3 Changes relevant for the Swiss derivation DCAT-AP CH

The table on the following pages lists all classes and attributes of DCAT-AP V1.1. Changes to V1.01 are shown in color and bold-faced:

- Green: New class / New attribute
- Yellow: Changed class / Changed attribute.

Additional properties (in green) and changed properties of the Swiss derivation DCAT-AP CH (cf. eCH OGD 2018b) are listed on a light gray background. Changes usually concern the requirement level *mandatory/recommended/optional*. According to DCAT-AP CH, attributes marked with "†" should not be used.

The last two columns assess the relevance for the Swiss standard and prioritize the potentially necessary adjustments:

A - Conflict between DCAT-AP V1.1 and DCAT-AP CH, should be resolved immediately.

B - Recommended adjustment

C - Possible adjustment, but not urgent.

Class	Class URI	Mandatory / Conditional[18]	Recommended	Optional	Relevance of the Change	Prio.
Agent	`foaf:Agent`	`foaf:name`	**dct:type**[19]		No adjustment necessary.	-
Category	`skos:Concept`	`skos:prefLabel`			Class requirement level lowered (*mandatory → recommended*). Change can be adopted.	C
Category Scheme	`skos:ConceptScheme`	`dct:title`			Class requirement level lowered (*mandatory → recommended*). Change can be adopted.	C
Catalogue	`dcat:Catalog`	`dcat:dataset` `dct:description` `dct:publisher` `dct:title`	`foaf:homepage` `dct:language` `dct:license` `dct:issued` `dcat:themeTaxonomy` `dct:modified`	**dct:hasPart** **dct:isPartOf** `dcat:record`† `dct:rights` `dct:spatial`†	New attributes can be adopted.	C
		`dct:issued` `foaf:homepage` `dct:modified*`		`dct:language` `dct:license` `dcat:themeTaxonomy`		
Catalogue Record	`dcat:CatalogRecord`	`dct:modified` `foaf:primaryTopic`	**dct:conformsTo** `adms:status` `dct:issued`	`dct:description` **dct:language** **dct:source** `dct:title`	New attributes can be adopted.	C
Checksum	`spdx:Checksum`	`spdx:algorithm` `spdx:checksumValue`			New *optional* class. Change can be adopted.	C

[18] Conditional properties are marked with an asterisk (*)

[19] Property name change (previously "publisher type"), but no change of URI

Class	Class URI	Mandatory / Conditional[18]	Recommended	Optional	Relevance of the Change	Prio.
Dataset	dcat:Dataset	dct:description dct:title	dcat:contactPoint[20] dcat:distribution dcat:keyword dct:publisher dcat:theme	adms:identifier† adms:sample adms:versionNotes† dcat:landingPage dct:accessRights dct:accrualPeriodicity dct:conformsTo† dct:hasVersion dct:isVersionOf dct:identifier dct:issued dct:language dct:modified dct:provenance dct:relation dct:source dct:spatial dct:temporal dct:type foaf:page owl:versionInfo[21]	owl:versionInfo replaces adms:version, which should not be used according to the Swiss standard. It is unclear whether the usage of the new owl:versionInfo should be discouraged as well. All other changes can be adopted as well.	B
		dct:publisher dcat:contactPoint dct:identifier dct:issued* dct:modified* dct:language* dcat:theme* dcat:distribution* dcat:landingPage*		dcat:keyword dct:coverage dct:relation rdfs:seeAlso schema:image		

---

[20] Changed URI and Range (adms:contactPoint, v:vCard)

[21] New URI, previously adms:version†

Class	Class URI	*Mandatory / Conditional*[18]	*Recommended*	*Optional*	Relevance of the Change	Prio.
Distribution	dcat:Distribution	dcat:accessURL	dct:description dct:format dct:license	adms:status† dcat:byteSize dcat:downloadURL dcat:mediaType **dct:conformsTo** dct:issued **dct:language** dct:modified dct:rights dct:title **foaf:page** **spdx:checksum**	dct:language is now part of DCAT-AP V1.1. All other changes can be adopted as well.	C
		dct:issued dct:rights dct:title* dct:description* dct:language* dct:modified* dcat:byteSize* dct:format*		**dct:coverage** **dct:identifier** **schema:image** dct:license		
Document	foaf:Document				No adjustment necessary.	-
Frequency	dct:Frequency				No adjustment necessary.	-
Identifier	adms:Identifier	skos:notation			No adjustment necessary.	-
Kind	vcard:Kind				Replaces v:VCard, as provided already in DCAT-AP CH.	-

Class	Class URI	Mandatory / Conditional[18]	Recommended	Optional	Relevance of the Change	Prio.
Licence Document	`dct:LicenseDocument`	**`dct:type`**[22]			Class requirement level upgraded (*optional → recommended*). Change should be adopted. See also the additional change of cardinality (to 0..n) implemented in V1.2 of DCAT-AP.	B
License Type	`skos:Concept`				No adjustment necessary.	-
Linguistic System	`dct:LinguisticSystem`				No adjustment necessary.	-
Literal	`rdfs:Literal`				No adjustment necessary.	-
Location	`dct:Location`				No adjustment necessary.	-
Media Type or Extent	`dct:MediaTypeOrExtent`				No adjustment necessary.	-
Period Of Time	`dct:PeriodOfTime`				No adjustment necessary.	-
Provenance Statement	`dct:ProvenanceStatement`				New optional class. Change can be adopted.	C
Publisher Type	`skos:Concept`				No adjustment necessary	-
Resource	`rdfs:Resource`				No adjustment necessary	-

[22] Different range, `rdfs:Class → skos:Concept`

Class	Class URI	Mandatory / Conditional[18]	Recommended	Optional	Relevance of the Change	Prio.
Rights Statement	dct:RightsStatement				No adjustment necessary	-
Standard	dct:Standard				No adjustment necessary	-
Status	skos:Concept				No adjustment necessary	-

4.4 National Derivations of DCAT-AP

Several other countries have extended DCAT-AP at the national level. A broad overview of extensions from Belgium, Germany, Ireland, Italy, The Netherlands, Norway, Spain, Sweden and Switzerland can be found in the report "Analysis of the DCAT-AP extensions" (Cochez et al. 2017), which was prepared by PwC EU Services on behalf of the ISA[2] programme:

> *"In this report, the analysis of national profiles implementing DCAT-AP v1.1 has been presented. In the process, we indicated several properties which could be discussed for inclusion in the next iteration of DCAT-AP or the W3C DCAT recommendation. Examples of already existing properties which have been modified frequently include, dct:identifier, dct:publisher, dcat:theme, and the way to use the vCard class. Furthermore, we identified a need to standardise more clearly how license and mediaTypes/formats are specified. New properties to be considered for future revisions of DCAT-AP include those related to spatial properties and relationships between the class Dataset and Distribution.*
> *We also indicated several changes made by national profiles which limit interoperability or which only help implementations capable of dealing with these specific requirements, while other implementations ignore the information as they are unable to interpret it. In the future, the ISA² Programme could help DCAT-AP implementers overcome these interoperability challenges by, for example, creating additional guidelines that ensure the compatibility of extensions with DCAT-AP and the interoperability of extensions among each other, or by checking the compliance of national extensions with DCAT-AP."*

While the details of all national extensions are documented in the report, the most important modifications with regard to DCAT-AP are listed in the following table, as these are candidates for further standardization or amendments of the current Swiss derivation:

Type of Change	
Changes to *mandatory* properties	The changes include the addition of restrictions, the increase of the requirement level from *optional*/*recommended* to *mandatory* and the introduction of new *mandatory* properties (in brackets the number of national extensions where this is proposed): - `dct:identifier` in the classes `Dataset` (4), `Agent` (2) and `Standard` (1) - `dct:publisher` in `Dataset` (4) - `dct:licence` in `Distribution` (3) and `Catalog` (1) - `dcat:theme` in `Dataset` (3) - `dcat:modified` in `Dataset` (2) and `Catalog` (1) - `dct:issued` (2), `dcat:mediaType` (2), `rdf:type` (2), `dct:format` (2) and `dcat:accessURL` (2). However, the changes in these five properties are not consistent across the different derivations.
vCard	Several restrictions have been defined in the various national derivations.
License	The provision of licensing information is a controversial topic, as illustrated by the declaration in DCAT-AP CH of `dct:rights` as mandatory and `dct:license` as optional, as well as related discussions in the eCH Specialized Group. Quote from the report (Cochez et al. 2017): *"Germany, Italy and Switzerland have made `dct:license` a mandatory property[23] for the class `Distribution` while the Netherlands has excluded it for the same class. Instead the Dutch extension has added a recommended licence attribute to the `Dataset`*

---

[23] The report may be based on a preliminary version of DCAT-AP CH. In any case, DCAT-AP CH 1.0 defines dct:license as an optional property.

	class with a limited number of possible values. The rationale seems to be that multiple distributions of the same dataset will have the same licence anyway."
	Germany wants a future extension of DCAT AP to include a possibility to add a license attribution text. Currently, Germany uses the property `dcatde:licenseAttributionByText` defined in the German national extension.
`dcat:mediaType` vs. `dct:format`	Sweden and Norway have replaced `dcat:mediaType` with `dct:format`, while Switzerland on the other hand, has elevated `dcat:mediaType` to *recommended* (instead of *optional*).
Geodata	This is also an important issue on which there is currently no consensus. To quote from the report: "*Some of the extensions have also added geospatial metadata elements. Ireland has included some geospatial elements for data interoperability across the nation. The extended properties are defined for the class* `Dataset`, *namely* `GeographicBoundingBox`, `SpatialReference System`, *and* `Spatial` Resolution. *The German extension added* `politicalGeocodingLevelURI`, `politicalGeocodingURI`, *and* `geocodingText`. *Italy included* `geographicalIdentifier`, `geographicalName`, *and* `Geometry` *(geometry). If* `loc:geometry` *is specified, then the three mandatory items CRS, coordinates, and geometry Type must also be provided. The coordinates represent coordinates of the geographic area covered by the dataset, CRS is the spatial reference system in which the data are represented, and geometry Type is the type of geometry that characterizes the spatial object used for the location of the dataset (e.g. point).* *Spatial (*`dct:spatial`*) is a DCAT-AP property that represents the geographical area coordinates where the dataset applies. Norway, Spain, Sweden, Switzerland, and the Netherlands have added range restrictions on the property spatial for the class* `Dataset`*. Each extension has introduced its own type of restriction. Sweden has also added restrictions for the class* `Catalogue`*.*"

4.5 GeoDCAT-AP

In 2016, a specialized application profile was published in the field of geodata: GeoDCAT-AP (cf. PwC EU Services 2016a). According to PwC EU Services (2016b), "*GeoDCAT-AP is an extension of DCAT-AP for describing geospatial datasets, dataset series, and services. It provides an RDF syntax binding for the union of metadata elements defined in the core profile of ISO 19115:2003 (Geographic information – Metadata) and those defined in the framework of the INSPIRE Directive. Its basic use case is to make spatial datasets, data series, and services searchable on general data portals, thereby making geospatial information better searchable across borders and sectors. This can be achieved by the exchange of descriptions of data sets among data portals.*"

GeoDCAT-AP does not define any extensions regarding classes or properties of DCAT-AP, but instead describes how to map other existing geodata standards into GeoDCAT-AP, as well as best practices regarding the use of the defined elements.

To stay compatible with INSPIRE and ISO 19115, GeoDCAT-AP restricts the modeling of spatial coverage to a bounding box based on 4 coordinates or a geographic identifier, even though the Core Location Vocabulary would allow more complex geometries. Furthermore, GeoDCAT-AP states that "cur-

rently there is no agreement on a preferred format to be used in RDF for the representation of geometries. In GeoDCAT-AP, geometries can be provided in any, and possibly multiple, encodings, but at least one of the following must be made available: WKT or GML" (cf. PwC EU Services 2016a, p. 57).

4.6 StatDCAT-AP

In 2016, a specialized application profile was published in the field of statistical data: StatDCAT-AP (cf. PwC EU Services 2016b). It contains several extensions.

4.6.1 New optional classes

Class	URI	Description
Annotation	`oa:Annotation`	A statement providing explanatory information about a resource. This property is used for statements related to the quality of the Dataset, including rating, quality certificate, feedback that can be associated to datasets or distributions.
Attribute Property	`qb:AttributeProperty`	A component property which represents an attribute of observations in the Dataset, e.g. unit of measurement.
Dimension Property	`qb:DimensionProperty`	A component property which represents a dimension in the Dataset.
Size or duration	`dct:SizeOrDuration`	A dimension or extent, e.g. the number of data series in a Dataset.

4.6.2 New optional properties

`dcat:Dataset`			
Property	URI	Range	Änderung
Attribute	`stat:attribute`	`qb:AttributeProperty`	Neues Attribut, optional
Dimension	`stat:dimension`	`qb:DimensionProperty`	Neues Attribut, optional
number of data series	`stat:numSeries`	`rdfs:Literal` typed as `xsd:integer`	Neues Attribut, optional
quality annotation	`dqv:hasQualityAnnotation`	`oa:Annotation`	Neues Attribut, optional
unit of measurement	`stat:statUnitMeasure`	`skos:Concept`	

`dcat:Distribution`			
Property	URI	Range	Änderung
Type	`dct:type`	`rdfs:Resource`	Neues Attribut, optional

4.7 Vocabularies

As shown in the table below, DCAT-AP V1.1 mandates a number of controlled vocabularies and named authority lists that must be used:

Vocabulary Name	URI	Classes	Property
ADMS change type vocabulary	http://purl.org/adms/changetype/	Catalogue, Record	`adms:status`
ADMS licence type vocabulary	http://purl.org/adms/licence-type/	Licence Document	`dct:type`
ADMS publisher type vocabulary	http://purl.org/adms/publishertype/	Agent	`dct:type`
ADMS status vocabulary	http://purl.org/adms/status/	Distribution	`adms:status`
Dataset Theme Vocabulary	http://publications.europa.eu/resource/authority/data-theme	Dataset	`dcat:theme`
		Catalogue	`dcat:themeTaxonomy`
Geonames	http://sws.geonames.org/	Catalogue, Dataset	`dct:spatial`
IANA Media Types	http://www.iana.org/assignments/media-types/media-types.xhtml	Distribution	`dcat:mediaType`
MDR Continents Named Authority List	http://publications.europa.eu/mdr/authority/continent/	Catalogue, Dataset	`dct:spatial`
MDR Continents Named Authority List	http://publications.europa.eu/mdr/authority/continent/	Catalogue, Dataset	`dct:spatial`
MDR Countries Named Authority List	http://publications.europa.eu/mdr/authority/country/	Catalogue, Dataset	`dct:spatial`
MDR File Type Named Authority List	http://publications.europa.eu/mdr/authority/file-type/	Distribution	`dct:format`
MDR Frequency Named Authority List	http://publications.europa.eu/mdr/authority/frequency	Dataset	`dct:accrualPeriodicity`
MDR Languages Named Authority List	http://publications.europa.eu/mdr/authority/language/	Catalogue, Dataset	`dct:language`
MDR Places Named Authority List	http://publications.europa.eu/mdr/authority/place/	Catalogue	`dct:spatial`
MDR Places Named Authority List	http://publications.europa.eu/mdr/authority/place/	Catalogue, Dataset	`dct:spatial`

Note that the time span covered by the authority lists is limited, e.g. the "Countries Named Authority List" only references countries that have been in existence since 1950.

In the report about national derivations of DCAT-AP, nothing is said about how controlled vocabularies are used.

StatDCAT-AP makes use of the Data Cube Vocabulary[24] to describe what attributes with which dimensions are used in a data set, and the Data Quality Vocabulary[25] for information about data quality.

For its extended profile, GeoDCAT-AP uses a number of controlled vocabularies, mainly to ensure interoperability with INSPIRE and ISO 19115:

- http://inspire.ec.europa.eu/media-types
- http://inspire.ec.europa.eu/metadata-codelist/ResourceType
- http://inspire.ec.europa.eu/metadata-codelist/ResponsiblePartyRole
- http://inspire.ec.europa.eu/metadata-codelist/SpatialDataServiceCategory
- http://inspire.ec.europa.eu/metadata-codelist/SpatialDataServiceType
- http://inspire.ec.europa.eu/metadata-codelist/TopicCategory
- http://inspire.ec.europa.eu/theme
- http://www.iana.org/assignments/character-sets/
- http://www.epsg-registry.org/
- http://www.opengis.net/def/crs/EPSG/

DCAT-AP CH contains references to several controlled vocabularies and named authority lists:

- the Swiss Confederation's terminology database TERMDAT;
- the multilingual thesaurus of the European Union (http://eurovoc.europa.eu);
- the official Directory of towns and cities (maintained by Swisstopo);
- vocabularies that have been standardized as part of eCH standards (e.g. eCH-0049, eCH-0070 and eCH-0145 for the description of functions and services of public authorities).

The European Commission has defined a series of "core vocabularies"[26], which are defined as *"simplified, re-usable and extensible data models that capture the fundamental characteristics of a data entity in a context-neutral and syntax-neutral fashion"*. So far, the following classes of entities have been covered:

- legal entities ("Core Business Vocabulary");
- public organizations ("Core Public Organization Vocabulary");
- locations represented as an address, a geographic name, or a geometry ("Core Location Vocabulary");
- persons ("Core Person Vocabulary");
- public services related to business and life events ("Core Public Service Vocabulary Application Profile");
- criteria, based on which judgements or decisions are made, and the evidence used to prove that such criteria are met ("Core Criterion and Core Evidence Vocabulary").

These core vocabularies have not been specifically referenced in DCAT-AP V1.1. Compatibility between the two would need to be investigated separately.

4.7.1 License vocabularies

As discussed above, how to define access rights and licenses is a controversial issue. DCAT-AP V1.1 says the following on this topic: "*Concerning licence vocabularies, implementers are encouraged to use widely recognised licences such as Creative Commons licences, and in particular the CC Zero Public Domain Dedication, the Open Data Commons Public Domain Dedication and License (PDDL), the ISA*

---

[24] W3C. The RDF Data Cube Vocabulary. W3C Recommendation 16 January 2014. https://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/

[25] Data on the Web Best Practices: Data Quality Vocabulary. W3C Working Draft 19 May 2016. https://www.w3.org/TR/vocab-dqv/

[26] https://joinup.ec.europa.eu/release/core-vocabularies-v20

Open Metadata Licence, the European Union Public Licence (EUPL) or an open government licence such as the UK Open Government Licence. Further activities in this area are undertaken by the Open Data Institute with the Open Data Rights Statement Vocabulary (ODRS) and by the Open Digital Rights Language (ODRL) Initiative" (PwC EU Services 2017, p. 19).

Europeana, the European platform for cultural heritage, currently foresees the use of 14 standardized licenses and rights statements[27], which comprise Creative Commons licenses (http://creativecommons.org), as well as rights statements from RightsStatements.org.

4.7.2 The use of dct:license and dct:rights statements in Switzerland

According to DCAT-AP CH, every distribution needs to have a dct:rights statement, while a dct:license statement is optional. Quite curiously, opendata.swiss does not support dct:license statements for distributions, but enforces a non-standard-compliant use of the ODRS license tag within dct:rights statements that is based on its own controlled vocabulary for expressing the terms of use. This leads to the incompatibility of the metadata at the international level.

The use of licenses in the context of open data has been a contentious issue in Switzerland since the inception of the Confederation's open government data project. A report by the legal team concluded that the legal provisions governing the publication of government data were sufficient to regulate the re-use of open government data and that no additional "licenses" were required (Wiedmer & Seiberth 2015). Citing Creative Commons licenses as examples, they portrayed "licenses" as contracts and as instruments belonging to the private law sphere while the publication of open government data belonged to the realm of public law. Interestingly, they ignored the fact that standard licenses do not necessarily need to be designed as contracts (e.g. Creative Commons Zero Public Domain Dedication or Datenlizenz Deutschland). They also overlooked the fact that, in the context of open government data, there is a need for standardized, machine-readable licenses alongside human-readable standard formulations and the legal provisions contained in various federal laws. As a consequence, opendata.swiss currently enforces human-readable standard formulations regarding the terms of use applying to distributions. It also allows to reference the legal basis by including a dct:relation statement in the dct:rights field. At the same time, it fails to support the use of standard licenses, although various Swiss cantons, cities, and even federal agencies are using such licenses in the context of open government data.

By overlooking the need for standard licenses, Wiedmer and Seiberth (2015) missed the opportunity to explain why several of the standard licenses commonly used in countries of the European Union cannot be applied to data from Switzerland. In fact, unlike the European union, Switzerland does not recognize any *sui generis* database rights. Thus, most of the data published as open government data in Switzerland is not protected by intellectual property law, which means that there is no legal basis to apply Creative Commons Attribution, Creative Commons Share-Alike or similar licenses to that data (cf. Lee 2017). There are however two possibilities how standard licensing terms can be applied to Swiss data: Firstly, the Creative Commons Zero Waiver or a similar license can be used, which is the case for various public-sector organizations in Switzerland. Secondly, licensing terms could be expressed with the help of a standard vocabulary; these terms of use would however rest not on copyright protection but on the legal terms enshrined in public law, which may have the same effect as Creative Commons Attribution, Creative Commons Share-Alike, or Creative Commons Non-Commercial licenses (Wiedmer & Seiberth 2015). Unfortunately, existing rights vocabularies, such as the Open Data Rights Statement Vocabulary (ODRS), currently focus on copyright and do not provide a standard way of linking to the legal basis of specific terms of use. When opting for licensing terms that are substantially different from the Creative Commons Zero Waiver, it should be kept in mind that incompatibilities may be introduced regarding the use of the data[28], which would be counter to the spirit of open data.

---

[27] https://pro.europeana.eu/page/available-rights-statements

[28] Wikidata, for example, requires data to be licensed under the Creative Commons Zero Waiver or equivalent, in order to facilitate data integration and re-use.

4.8   DCAT-AP CH and its interplay with the main standard and its various derivations

Based on the expert interviews and the workshop with the eCH Specialized Group, the interplay between the different derivations of DCAT can be described as follows (figure 3):

While the W3C Recommendation DCAT represents the smallest common denominator on a global scale, the application profiles at the various levels (European profile, national profiles) consider the specificities of the various levels and serve as a laboratory for innovations. A permanent process of innovation and harmonization thereby ensures that the different derivations of DCAT remain in sync: Novelties that have been successfully introduced at lower levels are considered for adoption at higher levels, and lower level derivations are regularly adapted to the higher-level standards. The communication across the different levels thus plays a key role and has been somewhat neglected so far during the development of the Swiss application profile.
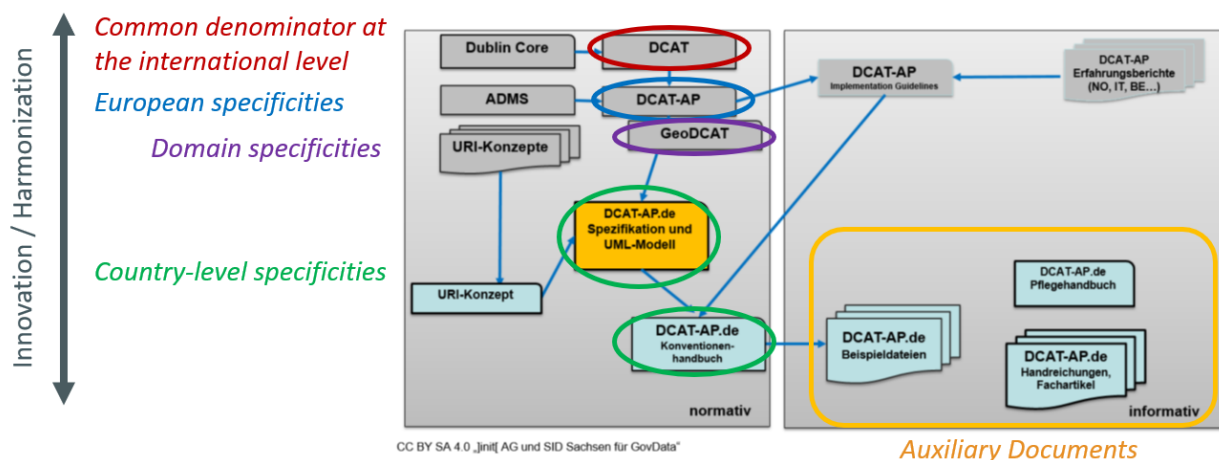

Figure 3: Different derivations of DCAT and auxiliary documents (example: Germany)

To conclude, the Swiss application profile enables the harmonization at the Swiss level (given the mandatory character of eCH standards), helps ensure compatibility with the European standard, facilitates communication about the standard by providing a simplified version of the European standard, and can be used to impose more restrictive requirements regarding metadata quality and completeness at the national level. It can also be used to cover Switzerland-specific particularities where necessary. According to the interviewees, such particularities include the multilingual situation, the definition of administrative levels, specific legal systems, the Swiss coordinate system, specific terminology, and Switzerland-specific ontologies.

It should be noted however that many interviewees are not entirely happy with the way the Swiss derivation of the standard is used today. Five out of fifteen interviewees even question the need for a Swiss application profile altogether. According to the interviewees, one of the international standards would do the job just as well. Others call for a strict limitation of "suissitudes", as they obstruct international interoperability in the longer run. Moreover, one interlocutor observed that it might be worthwhile testing some real innovations at the Swiss level in cooperation with the Swiss portal providers, but that this is not what was happening in practice. Instead, the Swiss community had been debating about insignificant deviations from the international standard. The view that a Swiss setting could be a place where innovations are tested in practice before being proposed at a higher level, was also shared by one of the members of the eCH Specialized Group.

4.9 Domain-specific extensions

Several interviewees recognized the usefulness of domain-specific extensions, as they allow catering to the domain-specific needs of specialized communities and help to ensure compatibility with existing domain-specific standards while keeping the main standard relatively "slim". The domains mentioned, where domain-specific extensions should be used, were statistical and geographical data. One interviewee also mentioned heritage data as a possible area where a domain-specific extension might be useful.

5   Evolving Requirements from a User Perspective

As open data portals mature, and the number of referenced datasets increases, requirements from a user perspective may also evolve. As Lourenço (2015) notes, user requirements need to be assessed with regard to the two main purposes of open data provision, namely facilitating data re-use to allow for the creation of new products and services (economic value of the data) and improving accountability of public agents (democratic value of the data).

Lourenço (2015) identifies a set of requirements open data portals should fulfill which are relevant to users intending to use government data from a transparency and accountability point of view:

- Portals should foresee mechanisms for external and independent quality assurance processes and publish the results of such revision processes alongside each published dataset.

- Portals should provide a mechanism to clearly identify and distinguish accountability-related datasets, i.e. datasets that are of primary interest from a transparency point of view (as opposed to datasets mainly intended for re-use).

- Portals should provide a list of all relevant governmental data sources, including information about the 'target' entities covered (i.e. those entities which might be accountable to citizens), the data provider in charge, as well as a list of all accountability-related informational items expected to be disclosed with an indication of all the time periods that ought to be covered. Data portals should make it easy for the user to see which of this information has been provided, and to assess the degree of openness of each data provider.

- The metadata provided for each dataset should include the identification of the 'authoring' and 'target' entities, informational items, and time periods covered.

- Portals should provide free search, oriented search and browsing mechanisms to help users find the required and related/complementary datasets.

- Portals should provide clear and simple descriptions of the concepts associated with the data being disclosed. Such descriptions or definitions should provide a common language to help describe the datasets, to categorize them, and to relate their metadata to the list of all datasets, allowing for a complete assessment.

- Portals should provide an indication of its granularity level (unit of analysis) for each dataset.

- Portals should provide information that allows for assessing delay in information provision (e.g. release date and/or periodicity of publication of data, and the actual release and update date of each dataset).

- Portals should provide mechanisms to allow users to express some measure of value or usefulness of the data disclosed as well as mechanisms to suggest missing valuable data.

Other user requirements have been expressed in form of OGD data publication standards, such as the open data principles of the Sunlight Foundation[29]. The table below contains a (slightly adapted) list of the ten principles and concrete requirements from a data provider's point of view that were identified during the open government data project of a Swiss canton.

---

[29]   https://sunlightfoundation.com/policy/documents/ten-open-data-principles/

Open data principle	Challenges / aspects to consider
1. Completeness and understandability of the data	Provide high-quality metadata.
2. Release of primary source data	It should be aimed for the release of primary source data; in some cases, this requires prior rights clearance and clarifying who is responsible for data publication.
3. Timeliness	Provide information on the update frequency of the data. State the level of data quality. When providing real-time data, use a disclaimer if plausibility checks have not yet been applied. Provide quality-checked data later. Provide version information.
4. Ease of access	Ensure findability of the data. Ensure that the data can be retrieved. Ensure a high level of usability of data platforms. Provide access to the data without prior registration; do not use captchas (the requirement of access without registration applies to basic services; extended services, like the unconstrained use of an API, may require registration). Provision of the data in common formats (in case of doubt, the data should be provided in different formats).
5. Machine readability	Publish data as 3-star data at least (see also point 7).
6. Non-discrimination	For all data that may be published: Ensure the «open by default» principle within the entire public administration. Do not require prior registration to access the data.
7. Commonly owned or open standards	Use W3C-Standards. Use open and well-documented APIs.
8. Licensing	Avoid license terms with liability clauses that hold data users responsible for third party claims at the address of public administration. If texts, images, or audio-visual content are out of copyright or have not been protected by copyright in the first place: Clearly indicate that no copyright restrictions apply and that the content is in the «public domain». Adapt the legal provisions that are contrary to the open data principles (e.g. restrictions with regard to commercial use).
9. Permanence	Datasets that are released must be versioned. Use permanent URIs. Provide an archive of earlier releases of the data.
10. No Usage Fees	Adapt the legal provisions that are contrary to the open data principles (e.g. fees regulations).

Currently, a systematic assessment of the Swiss open data metadata portals from a user perspective does not exist. Anecdotal user feedback suggests that users have encountered issues related to unclear or misleading metadata, the inaccessibility of referenced data, the publication of partial datasets[30], as well as lacking machine-readability of some data. In the participants surveys of the 2016 and 2017 editions of the Swiss Open Cultural Data Hackathon[31] (N = 51;43), when asked about how to improve the sustainability of the hackathon's impact, hackathon participants rated the item «improve the quality and/or completeness of open datasets» higher than the item «increase the number of open datasets». Furthermore, the difference between the two ratings increased from 2016 to 2017. Thus, as the number of open datasets referenced on open data metadata portals increases, we may see an increase in user expectations concerning data quality and completeness.

Regarding quality assurance, metadata portal providers can either try to enforce basic quality requirements up-front or provide users with filtering mechanisms that allow them to filter out data that do not meet certain standards.

---

[30] e.g. https://opendata.swiss/de/dataset/medikamentenkonsum (consisting of only two data points extracted from a larger dataset)

[31] https://glam.opendata.ch/hackathons/

6 Specific Needs of the Heritage Community

In 2017, the OpenGLAM working group of the Opendata.ch association began to systematically reference open data from Swiss heritage institutions in the Swiss Open Government Data Portal; in this context, several GLAM-specific requirements regarding the portal's functions have been brought forward. To put the different challenges and demands into perspective, this section provides an overview of the various types of heritage data, the heritage data platform situation in Switzerland, and the specific needs that have been expressed by the OpenGLAM community.

6.1 Various types of heritage data

The final report of the 2017 edition of the Open Cultural Data Hackathon (Estermann 2017) highlights the need of engaging data providers in a dialogue on how to handle the heterogeneity of datasets from the heritage sector. The author suggests that best practices should be developed regarding the publication of various types of datasets.

According to the report, datasets in the heritage sector are rather diverse; they typically contain one of the following types of data or a combination of them. Most of them can be allocated to one of the types of data identified in section 3.2 *(indicated in brackets)*:

- structured data describing or documenting some cultural or historical phenomena *(calendar data or statistical data)*;
- collections of digital artefacts (photographs, scans of prints, paintings, etc.) *(digital content)*;
- metadata pertaining to collections of digital or physical artefacts (i.e. object metadata) *(object metadata, collection metadata, or archival finding aids)*;
- geodata services (e.g. historical maps available through a web map service) *(geospatial data)*;
- inventories or registers of heritage institutions or collections *(base registers)*;
- administrative data from cultural institutions *(calendar data or statistical data)*;
- *ontologies, thesauri, or vocabularies;*
- *registers of ontologies, thesauri, or vocabularies.*

The last two types of heritage data are currently not referenced on the Swiss Open Data Portal.

According to the report, another source of heterogeneity lies in the varying granularity of datasets:

- "While some datasets comprise entire catalogues, covering the collections of many institutions (e.g. catalogues of platforms such as Swissbib or e-codices), others pertain to the collection of a single institution (sometimes also available through one of the platforms).
- While some digital collections comprise many thousands of artefacts, others only comprise a few dozen of them. Whether or not to break one collection up into many smaller ones is usually at the discretion of the data owner" (Estermann 2017, p. 5).

Moreover, the report notes that a further source of heterogeneity of datasets "lies in the fact that some collections may comprise both artefacts that are in the public domain or freely licensed and artefacts whose copyright status is unknown or the use of which is restricted. Here again, it is up to the data owner to decide whether or not to split up existing collections into smaller ones based on the licensing criterion. From a thematic point of view, splitting collections up based on this criterion may often not make much sense" (Estermann 2017, p. 5).

6.2 Heritage data platform situation in Switzerland

There are currently about 15 heritage data platforms with a national scope in Switzerland. Figure 4 provides an overview of the various platforms. Some platforms are supposed to cover the entire catalogue or finding aid of an institution (e.g. library catalogues in the case of Swissbib or archival finding aids in the case of Archives Online), while others focus on specific fonds or collections (e.g. e-codices

focusing on old manuscripts or Memobase focusing on photo and audio-visual collections). While some provide only search functions based on metadata (often including thumbnails), others also serve as presentation platforms where content can be viewed. As a method of data integration, most platforms either use federated search or (automatized) central data aggregation. Presently, only swissbib offers a linked data service. In addition, some heritage data is made available through the Wikidata platform (Estermann 2018), and several Swiss heritage institutions make some of their digital content available through Europeana. Both platforms provide a linked data service which allows querying the data through a SPARQL endpoint. e-codices is the only Swiss heritage platform that currently supports IIIF.
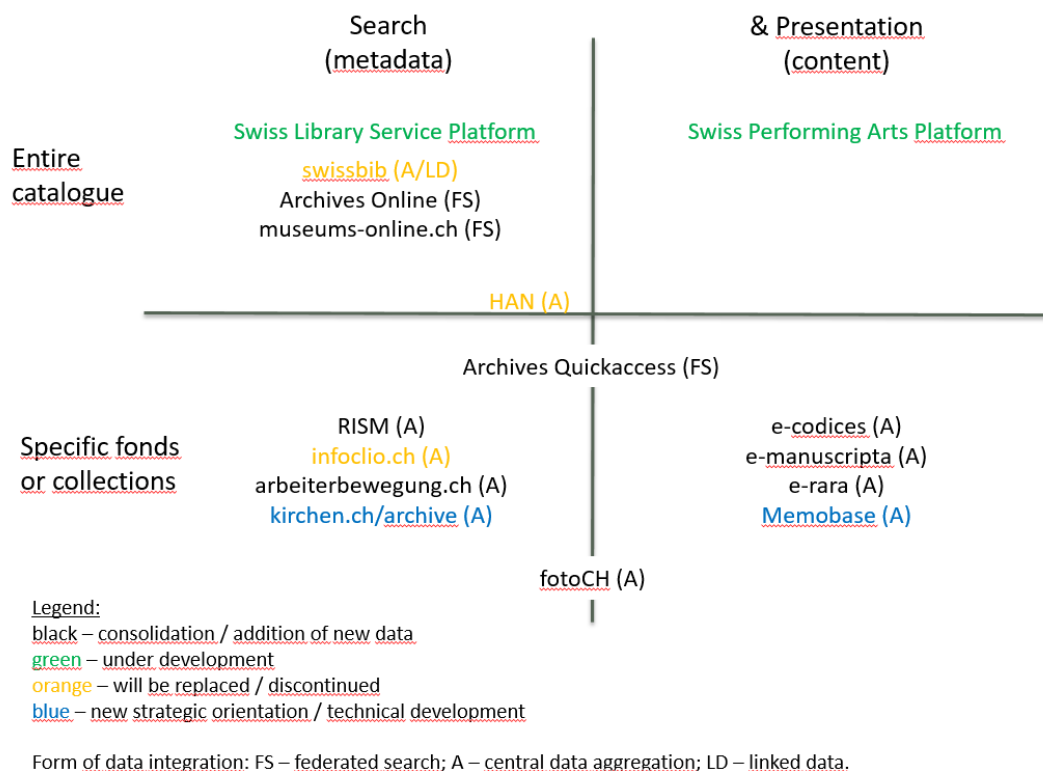

Figure 4: Overview of heritage data platforms in Switzerland with a national scope

It is presently unclear how the various heritage data platforms are supposed to interact with the Swiss Open Government Data Platform:

- To what extent should the heritage data platforms be configured to systematically provide metadata about specific fonds or collections to opendata.swiss? – Note that this is presently not the case, but some heritage data platform providers make their entire metadata catalogue available as open data (e.g. swissbib, e-codices). Furthermore, some individual fonds or collections available through heritage data platforms have also been manually referenced on opendata.swiss.

- When data (and content) from various sources is aggregated, it is sometimes unclear what constitutes a dataset in terms of DCAT. When data is aggregated on a linked data platform, such as Wikidata, the boundaries between different datasets may become blurred. – This raises the question as to what extent and in which granularity such data should be referenced on opendata.swiss.

- Should digital content and metadata pertaining to content be referenced on opendata.swiss or should only specialized platforms be used for this purpose? – If specialized platforms are to be used: what happens to the content that does not fit into one of the existing platforms in terms of its scope?

6.3 Specific Needs that have been expressed by the Swiss OpenGLAM community

According to Estermann (2017), the hackathon team would like to see the following improvements to the Swiss Open Data Portal:

- "The ability to add references to standard licenses (e.g. Creative Commons licenses) in the metadata.

- The ability to add a picture to the metadata record (which is useful especially in the case of collections of content, like photographs, prints, scans of paintings, etc.).

- Adaptations to the functionality allowing the integration of data from opendata.swiss on third party websites (such as widgets showing up-to-date metadata from the portal).

- The ability to tag datasets at the level of the catalogue. At present, tags for datasets are managed exclusively by the organizations, which makes it difficult to tag a specific subset of datasets for a specific event (e.g. a hackathon). Furthermore, the automatic harvesting of some organizations is set up in a way that prevents the attribution of tags in language versions not covered by the harvesting routine" (Estermann 2017, p. 4).

Some of these requirements were filed in form of change requests during the standardization process of DCAT-AP CH. Thus, the possibility to add references to standard licenses and to add a picture to the metadata record has been included in the standard. It should be noted, however, that the latter change was disputed among the members of the specialized group as was noted in the plenary meeting's minutes held on 7 May 2018:

"The new element "schema:image" was adopted by simple majority. The decision was preceded by a longer discussion where the following points were raised:

- The use of thumbnail pictures is quite common among cultural heritage platforms. Examples of data catalogs using thumbnail pictures include data.stadt-zuerich.ch and http://open-collections.okfn.org. The OpenGLAM CH Working Group also used preview pictures on its list of open datasets and collections […]. After adopting opendata.swiss as the main data catalog, this feature is missing.

- Having preview pictures for some catalog entries, but not for others, introduces some unwanted heterogeneity among the datasets. As an alternative, it is possible to make preview pictures available on a separate document that can be added to the dataset (e.g. make a few pictures available in a PDF document). Furthermore, a ZIP-file with all the thumbnails can be made available in addition to the high-resolution images. Heritage institutions should agree on a best practice in this area.

- It is debatable whether thumbnails qualify as metadata: If one limits oneself to machine-readable metadata, they would not qualify as metadata. If human readable metadata are included, they clearly qualify as metadata, as they are able to confer a lot of information at a glance that are not readily available as structured data or are tedious to look up for humans.

- It is also debatable whether thumbnails should be made available in a linked data environment, as they may require a different type of infrastructure. Counter-examples are Zenodo or Europeana, which make their data available as linked open data and heavily rely on thumbnail pictures to support the (human) search process.

- The copyright status of thumbnail pictures should be clarified: Do the same rules apply as for the full-size pictures or are there any limitations to copyright that are applicable when it comes to providing them in catalogues or finding aids?

- Pictures could also be used to illustrate other datasets (not only open collections); the entries in the data catalog would become more appealing" (eCH OGD 2018a, p. 2).

7 Perceived Challenges Today

The experts interviewed for this study mentioned a number of challenges in the open data portal eco-system. According to the interviewees, the most significant challenges currently concern the quality of the metadata and the quality of the data itself – all but one interviewee mentioned challenges in these two areas.

7.1 Metadata quality

Several interviewees stated that currently the metadata quality on opendata.swiss was "bad" or at least inconsistent. Insufficient metadata quality can result in outcomes that are detrimental to the extended use of open data:

- **Reduced interoperability**, as it is unclear to which other datasets a specific dataset is related or can be linked to. This affects all actors in the eco-system: data providers, open data portal providers, harvesters, and users.

- **Limited discoverability**, because it is currently difficult for a user to find relevant data sets just by querying the metadata portal without looking at the actual data sets.

- **Lowered customer satisfaction**: Due to the difficulty in finding data, users may have a wrong impression and assume that very little relevant data is available, thus leading the user to question the usefulness of open data and open data metadata portals in general.

Concretely, the following issues regarding metadata quality were identified:

- 10 out of 15 interviewees stated that often the **descriptions of the datasets were insufficient**, making it difficult for a user to correctly understand what data is included in a given dataset and how to judge the relevancy for their purposes. Five interviewees mentioned explicitly that **data comparability** is currently an issue.

- There is often **confusion about date properties**: Is it the date of publication of the data, or rather the publication of the metadata? Is it the date of the original publication, or the date of the last update? Are there any validity dates?

- A related issue is the **management of datasets over time**. For a user of the data, it is currently difficult to understand the relationship of different versions of the same dataset and how they are interlinked.

- To better identify and link relevant data sets, **agreed and understandable globally unique identifiers** are necessary, ideally without any central agency assigning such IDs. The current definition in DCAT-AP-CH is not seen as being sufficient in this respect.

- The **usage of keywords** in the dataset description is currently incoherent, making it difficult to find related datasets.

- There is also a perceived **lack of homogeneity of metadata** from different providers and across datasets, as well as across the different levels of government.

- Aggregating metadata portals often have **less metadata than what is available on the underlying harvested sites**. This is an issue both for opendata.swiss as well as for the European Data Portal.

- In DCAT-AP, there is currently **no standard way to describe the quality** of a dataset; thus, users need to look at the data itself to decide if a given dataset fits their intended use.

7.2 License and usage rights

Five of the eleven Swiss interviewees raised the issue of licenses and terms of use. The current Switzerland-specific use of `dct:rights` on opendata.swiss is seen as hard to understand and unnecessarily complicates the use of the datasets, especially in the international context. According to the interviewees, well-accepted standard Creative Commons licenses should be used.

7.3 Data quality in view of better interoperability of datasets

Not only the quality of the metadata, but also the quality of the data itself is sometimes an issue according to the interviewees. The following three issues were highlighted:

- There is a need for **globally unique identifiers** in the datasets to facilitate discovery and interlinking of related datasets.

- Generally, **consistent semantics of individual data fields** across datasets are needed. Example: The field "population" (for a city) can contain different numbers, as different counting methods are used: In the dataset from the city of Zurich the *economic population* is used, as this is relevant regarding the usage of the urban infrastructure, while in the dataset from the canton of Zurich the number of *registered inhabitants* is used, which is relevant regarding the number of taxpayers etc.

- For the reuse of data, it is important that the data actually conforms to the specified schema, i.e., that the data has been **validated against the schema**.

7.4 Multi-language support

Four interviewees pointed out concerns regarding multi-language support, a topic which is especially important for Switzerland. Obviously, the European Data Portal (EDP), the EU Open Data Portal (EU ODP), as well as some other European countries are faced with this challenge as well. For example, while the portal website may be multilingual, the actual datasets are not, or the dataset is not available in the desired language. Around two years ago, the EU ODP started to use machine translation to solve this problem. The results of machine translation are not always good, but the portal providers still believe it better than not having any translation at all.

7.5 Geographic data

For geographic data, three issues were mentioned:

- The metadata on opendata.swiss **lacks information about scale and perimeter** (bounding box, polygon etc.). As highlighted already in section 4, how to define the geometry of a geographic area is also a contested issue elsewhere in Europe.

- **Nomenclature for spatial structure** (in Switzerland typically cantons and communes) is different from country to country, which makes it hard to query the data and to make comparisons.

- There is a difficulty comparing historical and current data as **geographical perimeters change over time**: A commune may grow, e.g., through municipal mergers, and thereby may even change the canton it belongs to, as could be the case with Clavaleyres BE if it merges with Murten FR, as planned.

8   Suggestions for Improvement

In the course of the expert interviews, many suggestions for improvement were made. Some apply to the open data portal itself, while others were made in view of the further development of DCAT-AP CH. The preliminary results were also discussed during the workshop with the eCH Specialized Group "Open Government Data".

8.1 Suggested Improvements at the level of the open data portal / metadata catalogue

Given the improvement of the open data portal (the focus was mainly on opendata.swiss) and its metadata catalogue, the following suggestions were made:

- The interviewees widely agreed that **transparent license and rights information** is very important. Thereby, standard licenses (e.g. CreativeCommons licenses) should be used wherever possible. As several interviewees pointed out, it is not in the spirit of open data, if users need to read through national legislation before they know what rules apply with regard to data use. As it appears, there are two opposing views with regard to using standard licenses within public administration: Some argue that the present legal framework does not allow the public administration to apply licenses on their data, while others have chosen a pragmatic approach by applying the corresponding licenses from the available set of international standard licenses and have fared well with this.

- The **metadata catalogue** should also be provided **in the form of linked open data**. This will allow for interlinking and facilitate its re-use. Furthermore, as one of the platform providers pointed out, this is in his own experience a good and relatively easy way to raise data providers' awareness to linked open data.

- The **quality of metadata** should be improved: A certain level of quality and completeness of the data should be ensured at the time of data ingestion. The SHACL shape graph[32] that is presently being developed might be instrumental in highlighting some of the quality issues, but testing against the shape file may not be sufficient. How the same type of data (e.g. statistical time series) are represented on the portal, should be harmonized.

- Most interviewees found **information about data completeness** at the level of individual datasets important. There was however little reflection during the interviews about representing data completeness at an inter-organizational level (e.g. the availability of the same type of data from all cantons for the same time periods).

- Most interviewees found **information about data compatibility** important: information about the ontologies and vocabularies used in the datasets should be provided, along with detailed descriptions of the different data fields (schema information).

- Several interviewees also pointed out the importance of providing data providers with **incentives to improve data quality**: While some interviewees stressed the importance of encouraging the publication of linked open data, one pointed to the fact that, so far, metrics at the portal level have mainly focused on quantity (e.g. the number of datasets provided) instead of quality. He thinks that the time has come to move from a quantity-focused approach, which made sense at the beginning when only few datasets were available, to a quality-focused one. As the discussion among the eCH Specialized Group showed, quality issues are tricky and sometimes hard to resolve. A broader discussion will be needed about what quality means in a given context. As one participant observed, many of the quality issues at hand are not specific to open data but occur whenever data from different sources are combined.

- Several interviewees pointed to the importance of ensuring a **feedback loop with users** in the form of a comment function or a discussion forum. One interviewee observed that data quality was not a state, but a process, and that the interaction between data providers and data users plays a crucial role in this process. Today, no appropriate feedback channels are made available. One of the interviewees mentioned datathons or hackathons as a good method to get people to use the data and to provide feedback about the quality of the data. They pointed out that it is only through usage that possible inconsistencies of the data appear.

---

[32]   The Shapes Constraint Language (SHACL) is a World Wide Web Consortium (W3C) specification for validating graph-based data against a set of conditions. A SHACL shape graph for DCAT-AP CH is currently being developed to facilitate the validation of entries in open data catalogues against the DCAT-AP CH standard.

- Several interviewees suggested that **examples of data use** be provided on the portal. In addition to pointers to secondary data uses by third parties, a thorough description of the primary use of the data within the public administration would be appreciated, as knowledge about the purpose and circumstances of the original data collection or generation would greatly enhance the understanding of the data.

- Moreover, many interviewees requested an **improvement of the usability of the data portal**. Aspects mentioned in this context included:

    - Minimizing the number of clicks to access the actual data.

    - Providing an attractive preview of the data (several interviewees mentioned that this was not the case today on opendata.swiss).

    - Providing a data kiosk for explorative data visualization and data selection.

    - Providing direct access to APIs (e.g. by referencing data services in addition to datasets).

    - Providing pointers to high-quality documentation of APIs.

    - Harmonizing the presentation of datasets (aggregation, versioning, structure, granularity of description). Guidance concerning these aspects could be provided by means of a "convention handbook" ("Konventionenhandbuch") as it exists in Germany.

    - Providing multilingual metadata: One interviewee referred to the automatic translation of metadata entries on the European portal as a positive example; while the metadata thus generated was certainly not perfect, it was still useful to get an idea of the content of the dataset. The maintainer of the portal indicated that the quality of the automatic translation still needs improvement.

    - Ensuring that the datasets referenced on opendata.swiss are findable through Google dataset search. – *Note that this issue has already been resolved at the time of writing.*

    - Allowing full-text search on datasets.

    - Providing images to illustrate datasets where appropriate.

    - Improving links and references between datasets. This includes facilitating data citation by making the different versions of datasets and distributions citable.

    - Providing statistical information, such as the number of views or the number of downloads.

    Obviously, not all these improvements can be implemented by the provider alone, as many of them require action by, and close cooperation with, data providers. Minimizing the number of clicks, for example, also relies on the metadata quality.

As was pointed out by several interviewees, the implementation of linked data might be instrumental in achieving some of these improvements (see section 9). At the same time, one of the portal providers observed that the CKAN platform software was inadequate when it comes to supporting linked data, as it is neither semantic nor multilingual by default.

8.2 Further points to be considered at the level of the data portal

Several interviewees raised further questions that should be considered at the level of the data portal. Most of the concerns are likely to require further debate within the community, as there is presently no consensus concerning the preferable course of action:

- Should opendata.swiss **reference only datasets that are fully open** according to the Sunlight Foundation's open data principles (not the case today) or should a **catalogue of all existing public sector datasets** be provided, be they open or not (also not the case today)? – Maybe, both could be done on different platforms. The interviewees with a geocat background tend to ask for the referencing of all available datasets (as the geocat.ch platform aims to do), while representatives of the open data community ask for stricter enforcement of the open data

principles on opendata.swiss so that the portal would live up to the promise implied by its domain name.

- Should the scope of opendata.swiss be **limited to public sector data only** or should **data from the private sector** also be accepted? – While the platform provider is presently imposing a limitation to what they define as public sector data, referring to potential issues of data quality and objectiveness, other interviewees stress that a strict limitation to public sector data does not make much sense where private sector organizations provide similar or complementary datasets. It is presently unclear where such data should be referenced, as no private sector open data portal currently exists in Switzerland.

- Should **online collaboration** be possible on the platform? – Some interviewees are calling for online collaboration features on the opendata.swiss platform that go beyond a mere comment function or discussion forum. One interviewee asks the Confederation to "provide a collaboration platform instead of a data dump". Another interviewee suggested the possibility of using collaborative tagging to improve the findability of datasets. During the discussion with the eCH Specialized Group it was also mentioned that tackling data quality issues would require community interaction and therefore some sort of community platform.

- One interviewee mentioned that the metadata portal should increasingly **cater to the needs of software agents** instead of focusing solely on improving the usability for human users. In contrast, others seem to favor a development path where the needs of software agents are increasingly accommodated by systematically publishing data as linked open data. This would make the opendata.swiss portal redundant from the point of view of software agents. At the same time, the portal could retain its role as a central access point for humans.

- One interviewee suggested that it might be worthwhile to not only **aggregate metadata on the platform, but also the actual data itself**. This suggestion should also be assessed in view of linked data publication.

- Two interviewees suggested that the **portal provider take over additional tasks**, such as fostering data use, organizing community events, improving the OGD Handbook, or supporting data providers with opening-up data. While it is not obligatory that all these tasks be carried out by the portal provider itself, they may play a coordinating role in ensuring that these aspects be effectively addressed within the Swiss open data community.

8.3 Further development of DCAT-AP CH

The interviewees made a series of suggestions about further development of the Swiss derivation of DCAT, the most important one being that it **should be aligned with the European profile**. The principle of subsidiarity should be applied: Wherever harmonization at the next higher level makes sense, standardization at the higher level should be pursued. As one interviewee observed, in an ideal world, there would be no need of DCAT-AP CH. One interviewee also questioned the usefulness of maintaining a German and a French translation of the standard at the Swiss level. Several interviewees would prefer that more energy be put into the transformation of data into linked data instead of spending time on further elaboration of the Swiss standard.

Further suggestions for improvement include the following:

- **Controlled vocabularies** should be agreed upon and provided to enhance the interoperability of datasets. Thereby it is important not only to align data modelling practices at the national level, but also to ensure data compatibility at the international level.

- **Additional metadata fields** were suggested in the following areas:
  - **Data quality** – the exact data which should be provided requires further analysis and discussion.
  - **Completeness of the data** – here again, how exactly to approach this issue requires further analysis and discussion.

- **Ontologies and vocabularies** used in the dataset (i.e. schema information, including the geo standard used, as well as vocabularies or thesauri used in the dataset).

- **The contact point of the data aggregator** – as opposed to the contact point of the data provider. This is particularly useful in situations when data is aggregated at several levels.

- Provide a specific field for references to the **legal foundations**; at present, a workaround is being used for this on opendata.swiss.

- A **field for general comments in a free-text form** should be provided, as is the case of the OGD portal of the City of Zurich whose data model is based on Dublin Core. Such a free-text field helps to avoid the misuse of other data fields when relevant information is added when no dedicated field currently exists.

- **Parent-child relationships** should be represented between datasets when datasets contain other datasets (e.g. in the case of data aggregation at the cantonal or the federal level). Users should be directed to the dataset that is most useful to them, which will often be the data that has already been aggregated. A similar problem arises when it comes to expressing the relationship between data extracts and their source databases (e.g. one-time snapshot from a larger library database).

- One interviewee suggested that the **linked data specific elements** from the VoID vocabulary[33] be integrated into the DCAT application profile.

- And one interviewee mentioned that the **specialized extensions of DCAT** should also consider the level of the Swiss profile.

Regarding the definition of additional metadata elements, one interviewee observed that more metadata elements lead to greater complexity of the data model and makes it harder for data providers to provide and maintain the metadata.

And finally, as was pointed out during the discussion with the eCH Specialized Group, the Swiss standardization group should not only focus on the elaboration of the Swiss application profile but should get involved in the DCAT-AP working group at the European level as well as in the international DCAT working group. Suggestions to these working groups can be provided on github, which is straightforward and not very time-intensive. Active participation in the working group deliberations, in contrast, is time-intensive; thus, if full participation is to be ensured on behalf of the Swiss community, resources need to be allocated accordingly.

---

[33] https://www.w3.org/TR/void/

9 Transition to Linked Data

The interviewees are convinced that the future belongs to linked data. The transformation of data into linked open data is an investment by data providers to improve the usability of the data. One of the primary purposes of linked data publication is the interlinking of datasets between various institutions It is therefore a collective effort that requires coordination – between organizations, across domains, and at the international level.

9.1 Advantages of linked open data

Linked open data comes with a series of advantages in comparison to data of lower maturity levels and provides a valid response to some of the challenges identified by our interviewees:

- With linked data, it becomes unnecessary to provide **detailed ontology/vocabulary information** at the metadata level. This information is already contained in the data itself and could be extracted at the level of the portal if required.

- Linked data provides **great richness of expression** when it comes to providing information of various sorts about the datasets at hand.

- Linked data makes it easier to **spot similar, related and compatible datasets**. Publishing data as linked open data thus provides extra value to the downstream data user, as more information can be extracted from interlinked data. For example, linked data makes it easier to compare data from various sources (e.g. population statistics from different cities).

- Linked data makes the **sharing, exchange, and adaptation of applications** easier as the data model information is provided in a standardized form and can easily be compared between one dataset and the other.

- In a similar vein, linked data makes it easier to **integrate new data in existing applications**, by facilitating automatic updates and synchronization tasks.

9.2 Current challenges regarding linked data publication

As some of the interviewees noted, when it comes to publishing public sector data as linked open data, Switzerland is well on track compared to other countries. Having a central data catalogue of open data available on opendata.swiss is a great advantage compared to the situation a few years ago when open datasets had to be searched for on various government websites. On the other hand, there is no doubt that we are still at the beginning of a longer process, and that there is a chicken-or-egg problem that needs to be overcome: If there is limited linked open data, there is little incentive to create cutting-edge applications leveraging the advantages of the semantic web, and if there are no applications, there is little incentive for data providers to put in the extra effort required to improve the maturity level of their data.

During the interviews, the following challenges related to linked data publication were mentioned:

- The most widely noted challenge was the **lack of awareness and know-how among data providers**.

- Several interviewees also mentioned that many fields were still **missing ontologies** that are ready for the semantic web. In some cases, new standards need to be created, while in other cases the main challenge lies in transiting from legacy standards to standards supporting the linked data approach.

- Several interviewees also observed that there was an evident **lack of awareness and know-how among data users**. Linked data is complex at first sight and requires different programming techniques than other data formats.

- When it comes to interlinking data across different institutions over time, properly historicizing entity definitions and **interlinking historicized data** constitutes a significant challenge. At the time being, most institutions publish snapshot data and do not properly document the historical dimension of their datasets (e.g. municipalities are constantly being merged or separated). Datasets, therefore, undergo changes over time, but there is no single agreed-upon approach to deal with this issue; thus, every institution chooses its own approach and these

approaches are usually not or only poorly documented). If datasets are not properly historicized, the whole idea behind linked data fails regarding longitudinal data analyses.

- Several interviewees mentioned that there is still a **lack of tools** that allow people to work with linked data on a regular basis; currently, many specialized information systems do not support linked data.

- As was pointed out by several interviewees, **producing linked data is resource-intensive**. Organizations therefore think twice before putting in the extra effort to publish their data as linked data.

- One interviewee expressed concern regarding **data protection issues** due to possible de-anonymization of data. In fact, according to legal provisions, there are certain datasets that must not be linked to certain other datasets. Therefore, publishing such data as linked open data is highly problematic, even if it seems harmless at the level of each individual dataset. By embracing the linked data approach, data publishers would virtually interlink such datasets and invite users to run analyses on the two datasets simultaneously.

- One interviewee also mentioned that **linked data cannot be applied to all types of data**, mentioning radar data as an example.

- Two further issues were mentioned that are not specifically related to linked data publication but concern open data in general: Certain data holders are still **reluctant to release their data as open data**, which results in **varying levels of data provision** in different (geographic, thematic, etc.) areas.

9.3 How to facilitate and promote the provision of linked open data

The interviewees made various suggestions how the publication of linked open data can be facilitated by portal providers, standardization bodies, data holders, and/or the open data community at large. Three key measures were often mentioned and should receive careful attention:

- **Shared domain ontologies and vocabularies** need to be developed and made available. Portal providers may play the role of facilitators or instigators to initiate this process, but in the end the data holders of a given domain need to agree on common standards.

- **Existing base registers and taxonomies should be provided as linked open data**. They play a crucial role in interlinking datasets from various sources. Thus, data providers holding data that could potentially be re-used in many other datasets should prioritize their transformation into linked open data.

- Portal providers should **provide the metadata catalogue** as linked open data. In other countries, this has played an important role in raising the awareness among data providers with regard to linked open data.

In addition, the following suggestions were made by individual interviewees:

- A **topic ontology** should be provided on the portal that goes beyond the present category system. Maybe, a crowdsourcing approach could be used to achieve this.

- **Wikidata** could be used to develop data modelling approaches and shared vocabularies at an international level.

- Tools could be made available on the metadata portal that **facilitate the exploration of RDF data**.

- **Examples of linked data use** and its advantages should be provided.

- Linked open data should **be published at the highest aggregation level** to facilitate harmonization (*e.g. centralized data publication by the Federal Statistical Office, instead of cantonal statistical offices publishing their data separately*). This would greatly enhance the usability and the usefulness of the data.

- **Training** should be provided for data publishers and for data users.

- One interviewee pointed out that data holders do not necessarily need to build up all linked-data-related competencies in-house. Instead, they should mandate **specialized service providers** to accelerate the data transformation process.

- Another interviewee raised the question whether **data conversion services** could be made available at the level of open data portals, e.g. by providing a toolset or some sort of data wizard at the level of the metadata portal that would lead data providers through the data transformation process and assist them with this task.

9.4 How can open data metadata platforms support the transition towards linked data?

There are different ways open data metadata platforms can support the transition toward linked open data. The following measures could be envisaged in a step-by-step process:

- **Measure 1:** While open data metadata platforms serve as central catalogues of datasets independently of their data maturity level, they increase the findability of datasets of higher maturity levels by indicating the data maturity level for each referenced dataset (in terms of the 5-star model).

- **Measure 2:** In addition, for each dataset that has been published as linked open data, open data metadata platforms provide information about the data models, ontologies, and controlled vocabularies used in the dataset and list the authority files/base registers that are linked to by the referenced dataset.

- **Measure 3:** Open data metadata platforms support the transition toward linked open data by also serving as catalogues of RDF ontologies and authority files/base registers and by providing information about the frequency of their use.

In addition, open data metadata platforms may provide the metadata of the referenced datasets in the form of linked open data through a SPARQL endpoint.

10 Discussion

In this section, we discuss the insights gathered throughout the study regarding the research questions formulated in section 2.

10.1 Development trends regarding DCAT and its derivations

While the W3C Recommendation DCAT represents the smallest common denominator on a global scale, the application profiles at the various levels (European profile, national profiles) take into account the specificities of the respective level and serve as a laboratory for innovations. A permanent process of innovation and harmonization thereby ensures that the different derivations of DCAT remain in sync: Novelties that have been successfully introduced at lower levels are considered for adoption at higher levels, and lower level derivations are regularly adapted to the higher-level standards. The communication across different levels thereby plays a key role and has so far been neglected by the Swiss open data community.

At the international level, four tendencies can be observed:

- There is a multitude of application profiles that provide **extensions to DCAT**; these application profiles can take the form of national or domain-specific profiles (e.g., DCAT-AP DE; StatDCAT-AP).

- In some countries, **convention handbooks** have been published in addition to the national application profile, describing how to use a standard based on best practices. Similarly, GeoDCAT-AP can be considered as a convention handbook in the area of geo-spatial data; it does not extend the standard *per se*.

- Some application profiles contain lists of **controlled vocabularies** that are to be used in the context of the metadata. Controlled vocabularies have also been made available by other players, such as the European Commission.

- In some countries, further **auxiliary documents** (example data, information material, etc.) have been provided.

While auxiliary documents certainly benefit from translation and localization, it is unclear which of the other country and domain specializations are meant to remain in place, and which ones are meant to be consolidated at the international level. Ideally, for the sake of data interoperability, a maximum of consolidation and harmonization should occur over time.

10.2 Ensuring the interoperability of DCAT-AP CH with other application profiles

To ensure the interoperability of DCAT-AP CH with other application profiles, several adjustments have been recommended in cases where it seems advisable to follow the development of DCAT-AP. In addition, reference should be made to the extensions of StatDCAT as well as to the conventions stipulated in GeoDCAT.

Note that the ISA[2] report (Cochez et al. 2017) calls for corrective actions regarding the changes made by national profiles that limit data interoperability at the international level. The further development at the international level therefore needs to be monitored to make the necessary adjustments to the Swiss application profile in the future. Similarly, innovations introduced in the Swiss application profile should be propagated to other countries. If there is no uptake in other countries, it might be advisable to abandon them in future releases of DCAT-AP CH.

10.3 Enhancements to the metadata standard required by portal users

The primary requirement by data portal users regarding the Swiss application profile is to align the application profile with the European standard. Most Swiss stakeholders would rather put more energy into publishing data as linked open data instead of spending time on the further elaboration of DCAT-AP CH. And finally, as was pointed out during the discussion with the eCH Specialized Group, the

Swiss standardization group should not primarily focus on the further elaboration of the Swiss application profile, but rather get involved in the DCAT-AP working group at the European level as well as in the international DCAT working group to push for the improvement of the standard at these levels.

Enhancements of the DCAT standard that should be aimed for include:

- Allow for direct pointers to APIs; this would be particularly useful in the case of (quasi-)real-time data, in the case of current calendar data, as well as in the case of "living" databases.
- Provide additional data fields for data quality, completeness of the data, ontologies, and vocabularies used in the dataset, the contact point of the data aggregator, as well as references to the legal foundations governing the generation and the publication of the data.
- Allow preview images; this would be particularly useful in the case of digital content (already implemented in DCAT-AP CH v. 1.0).
- Ensure that license information is unequivocal (in the case of digital content, does the license apply to the object metadata or to the digital objects themselves?).
- Provide a field for general comments in free-text form.
- Provide additional possibilities to describe relationships between datasets.
- Add linked data specific elements from the VoID vocabulary.

At the same time, when extending the standard, an eye should be kept on maintenance cost. There is no reason to have a myriad of specialized data fields if no one uses them.

As important as the additional data fields are the **controlled vocabularies**, which should be agreed upon and included in the standard. But here again, it is crucial to ensure compatibility at the international level.

Interestingly, no mention was made during the expert interviews of improvements to the metadata that would enhance the democratic value of data, e.g. by tracking the regularity of the publication of specific datasets by public agents (cf. Lourenço 2015). There may be little demand for this type of data in Switzerland or the wrong people may have been interviewed.

10.4 Need for harmonization in the presentation and description of datasets

Apart from the general improvement of metadata quality (in some cases important information just seems to be lacking), many interviewees called for an effort to harmonize practices concerning the description of datasets. Aspects to focus on include: insufficient descriptions of datasets; diverging/unclear use of date properties; incoherent interlinking of datasets; unclear attribution of unique identifiers; incoherent use of keywords; lack of homogeneity of the metadata; lack of standardization when describing the quality of datasets.

In some areas, additions to the standard may be needed. In others, it would be useful to agree on shared practices that could then be described in a **convention handbook**. Here again, efforts should be coordinated at the international level. It should be examined to what extent the German convention handbook (cf. Sklarß et al. 2018) could serve as a model.

10.5 International cooperation

Increased attention should be given to international cooperation. As was pointed out during the discussion with the eCH Specialized Group, the Swiss standardization group should not primarily focus on the elaboration of a Swiss application profile but should get involved in the DCAT-AP working group at the European level, as well as in the international DCAT working group. **Contributions to the international standards should be the default** and the further elaboration of the Swiss application profile the exception. Resources should be allocated accordingly.

10.6 Supporting the move towards 5-star linked open data

Publishing linked open data should be a priority. The progression towards linked open data can be supported at various levels:

- The publication of **base registers, authority files, and taxonomies** as linked open data should be prioritized. This should be done by the data holders in question; however, a shortlist of these datasets, along with the information regarding their publication status, could be kept at the level of the national open data portal.

- Ideally, base registers and authority files would be **historicized**: they should take into account the historical development of the entities they describe, e.g. mergers and acquisitions between companies, mergers and separations of municipalities, etc. Some coordination will probably be needed when it comes to deciding how exactly to historicize certain elements.

- In the individual datasets, **globally unique identifiers** should be used that link to base registers and authority files. To facilitate longitudinal comparisons, links should be provided to **historicized data** (e.g., a pointer to a given municipality at a given point in time).

- Information about the **ontologies and vocabularies** used in the datasets should be provided. For the transition period, defining dedicated data fields for this purpose may be considered. Once the data are published as linked open data, this information can be generated automatically by the data portal.

- In many fields, **relevant ontologies** are still missing. Community efforts to develop and agree on such ontologies should be supported. There should be a **community platform** where data modelling and data quality issues can be addressed.

- **Schema information** should be provided, and datasets should be validated against the schema. Here again, it could be examined to what extent this could be done before actually moving to linked open data. Providing "frictionless data" might be a first step on the road towards publishing linked open data.

- One of the main challenges identified by the interviewees is the relative **lack of awareness and know-how** among data providers and data users. Metadata platform providers could raise awareness levels by publishing the data catalog as linked open data. Furthermore, community events, such as hackathons, focusing on linked open data, could be organized to foster the exchange among data providers and users and to facilitate collective action.

- In the longer run, open data metadata platforms should support the transition towards linked open data not only by providing **information about the maturity level** of each dataset, but also by automatically extracting and displaying **information about the data models, ontologies, and controlled vocabularies** used in the various datasets and the **authority files/base registers** that are linked to by the referenced datasets. This information could be complemented by a central, dynamically created catalog of RDF ontologies and authority files/base registers linked to by the various datasets, along with information about the frequency of their use.

The transition to linked open data will not make metadata catalogues obsolete, as metadata portals will still be used by humans to search for datasets and to explore them. The current hierarchical cascade of data portals may however give way to a more distributed data portal landscape in the future. DCAT itself will not become redundant, but data catalogues should be kept simple and efforts should focus on the deployment of linked data instead of making the catalogues more sophisticated. In the linked data world, good solutions are needed more than ever to facilitate user interaction with large living databases: Users need to be given the opportunity to explore the data in easy ways and to interactively define their own "datasets" for download (e.g. through a data kiosk) and to gain direct access to the respective API.

10.7 Scenarios for the further development of DCAT-AP CH

To date, the Swiss application profile serves the harmonization at the level of Switzerland (given the mandatory character of eCH standards). It helps ensure compatibility with the European standard, facilitates communication about the standard by providing a simplified version of the European standard and is used to impose more restrictive requirements regarding metadata quality and completeness at the national level. It may also be used to cover Switzerland-specific particularities where necessary. According to the interviewees, such particularities comprise the multilingual situation, the definition of administrative levels, the Swiss coordinate system, Switzerland-specific ontologies and terminologies, as well as specificities of the legal system.

At the same time, maintaining a national application profile of DCAT does not come without costs and always bears the risk of defining exceptions and deviations from the main standard that result in interoperability issues in the international context. For this reason, a number of stakeholders are clearly concerned as to whether a Swiss application profile is actually needed. An alternative approach would involve filing relevant change requests directly with the international standardization bodies. This concerns both the core standard itself, as well as controlled vocabularies and a convention handbook that would need to be developed at an international level.

A further alternative scenario compared to the current approach would involve using the Swiss setting as an experimentation field to test innovations at a national level before they are implemented in other countries. This would, however, require close cooperation between Swiss platform providers and the Swiss standardization body, as well as a clear commitment to innovation in a specific field. In this case, novelties could be implemented on the relevant platforms and simultaneously be reflected in DCAT-AP CH. One interviewee suggested that heritage data could be such an experimentation field where some real innovation could be provided. Whether it would make sense to create a specialized application profile for heritage data is up for discussion. After all, some of the specifics of the heritage sector may not be exclusive to data from heritage institutions but may also appear in the context of classical public administration. Such special requirements concern the need for preview images, the problem related to providing access both to large "living" databases and to extracts from them, "blurred" boundaries between datasets, object collections with various licenses, as well as complex rights situations that cannot be expressed by applying one license per data object. Note that in this scenario, coordination at the international level would be as crucial as in the other scenarios. There is also little doubt that open data platforms, such as opendata.swiss, will be unable to replace existing heritage platforms, as there will be a continued need for dedicated presentation platforms that are optimized for the centralized viewing of digital content (such as Memobase, Europeana, or Wikimedia Commons), or the exploration of other heritage-specific data, such as archival finding aids or library catalogues.

Apart from the further development of the DCAT application profile, providing controlled vocabularies, and adopting a convention handbook, consideration should be given to the development of information material that helps raise awareness of the standard among various stakeholder groups in Switzerland. In contrast to the first set of documents that are primarily targeting experts and technical staff, it might be useful to provide low-threshold information material in the local languages.

10.8 Improvements at the level of opendata.swiss

In addition to the answers to the primary research questions of this study, many suggestions for improvements at the level of the opendata.swiss portal have been made. Many of these improvements fall under the responsibility of the portal provider; others require the active cooperation on the part of data providers. The suggestions include (in approximative order of importance and urgency):

- The possibility to add **references to standard licenses** in the metadata is imperative. In this point, opendata.swiss must fully implement DCAT-AP CH. Over the past five years, this requirement has been reiterated by various stakeholders, and it is time to take these calls seriously.

- Full implementation of DCAT-AP CH should also be the goal regarding the ability to **add preview images to metadata records**. In general, DCAT-AP CH should be implemented in a way

that allows for the meaningful display of a maximum of data that is harvested from lower-level aggregators.

- The **portal's usability** should be continuously improved. The current report contains many suggestions that should be examined in more detail.

- The portal's **metadata catalog** is to be published **in form of linked open data**.

- There needs to be a shift away from a focus on the number of datasets referenced on the portal to a **focus on the quality of the data in terms of data usability**. This includes the quality of metadata, data quality, data completeness, data maturity, and licensing. Apart from the licensing issue (several core datasets referenced on the portal are not actually provided as open data), ensuring high quality data and metadata constitute the greatest challenge perceived by the experts interviewed. In this context, it is important to conceive of data quality as a process, and not as a state. It is therefore crucial to **ensure an effective feedback loop** between data users and data providers.

- More **examples of actual data use** should be provided on the portal; ideally, these examples would be linked from the metadata entries concerning the datasets used. In addition, it might be advisable to proactively communicate such examples to potential target groups. Concrete examples of data use are also a good starting point for reflections on data quality and data completeness.

- The functionality allowing the **integration of data from opendata.swiss on third party websites** should be improved.

- Clarification as to how to reference **digital content** on opendata.swiss should be provided. Furthermore, guidance is needed regarding open data and open content that presently fits no platform's scope. To clarify these issues, the portal provider should closely cooperate with the heritage community and other interested parties.

Furthermore, as appears from the discussion regarding the possible scenarios for the further development of DCAT-AP CH, it is important that the provider of the opendata.swiss portal clarify what their **intended pace of innovation** is.

11 Conclusions and Recommended Actions

This study has examined the current state of development of the Swiss application profile of DCAT in the context of international developments and its usage among the Swiss open data community. Thereby it has also shed light on how the Swiss open data metadata portals, and specifically the national portal opendata.swiss, are currently perceived by the various stakeholders. Most interviewees view the existence of the metadata portal and the metadata standards in a very positive light. At the same time, they identified many areas requiring improvement. Concrete recommendations for improvement have been detailed in the previous two chapters. From these we can deduct a number of actions that we recommend should be taken to further propagate the use of open data in Switzerland. In this final section, recommended actions are detailed for the eCH Specialized Group OGD, the provider of opendata.swiss as well as for the Confederation's OGD and Linked Data projects. As has been pointed out in various parts of this report, some of the proposed measures will only be successful if they are also actively supported/embraced by the data providers.

11.1 Recommended actions for the eCH Specialized Group OGD

The following table summarizes recommended actions to be taken by the eCH Specialized Group OGD, as this group has the required expertise and legitimation to tackle these issues. However, it must be stressed that for this group to implement the recommendations, appropriate **funding needs to be provided**.

ID	Recommendation
eCH-1	**Actively participate in international developments**.
	Take an active role in the deliberations around the next major release of DCAT and DCAT-AP. Systematically foster community discussion in Switzerland regarding the issues at hand, with the goal of providing well-founded input and submitting change requests at the international level. This also minimizes the risk of incompatible extensions being made at different levels. The yet unpublished DCAT-AP OP extension used by the EU Open Data Portal should be examined in this context as well.
eCH-2	**Adopt controlled vocabularies and have them published as linked open data.**
	Provide lists of controlled vocabularies and base registers (authority files) that are to be used in the context of the open data catalogue and within the datasets themselves. To facilitate the transition to linked open data, these vocabularies and base registers should be published as linked open data.
	This will require coordination at the international level and may also involve cooperation with other eCH Specialized Groups and standardization bodies.
eCH-3	**Create a convention handbook**.
	As in Germany, normative conventions on how to apply DCAT-AP and DCAT-AP CH should be defined and collated in a handbook. This would lead to better metadata quality as well as to a reduced need of changes in DCAT-AP CH. In the intermediate term, the consolidation of national convention handbooks at the European level should be considered.
eCH-4	**Create non-normative communication material about the standard**.
	There are several types of stakeholders that need to be addressed in the context of the standard, which can roughly be divided into two target groups:
	- On one hand, there are the experts involved in the actual development of the standard as well as technical staff responsible for its implementation. This target group is best served by standard documents in English, as the development of the standard should happen with the international setting in mind.

ID	Recommendation
	- On the other hand, there are open data coordinators and domain specialists who need to be informed about the purpose and the content of the standard, along with best practices for its implementation. These stakeholder groups should have the possibility to provide feedback on the standard at an approachable level that is not too technical. Different communication material should be created, and different communication channels should be used for the two target groups. Note that this is an issue that has been encountered in other eCH specialized groups as well. It has also been raised during the last Linked Data Switzerland Workshop regarding the creation of material that helps the propagation of linked open data.
eCH-5	**Make minimal adaptations to DCAT-AP CH.** The modifications suggested in this report (section 4.3), along with further issues identified during the creation of the SHACL shape graph for DCAT-AP CH in fall 2018[34], should be addressed in a minor release of DCAT-AP CH in 2019.
eCH-6	**Decide about the future of DCAT-AP CH.** After the next major release of DCAT-AP (expected for November 2019), the decision to further maintain or to discontinue DCAT-AP CH needs to be made.

11.2 Recommended actions for the provider of opendata.swiss

The following table summarizes the recommended actions to be taken by the provider of the opendata.swiss metadata portal.

ID	Recommendation
ODS-1	**Ensure full support of DCAT-AP CH.** Full support of DCAT-AP CH should be ensured. This concerns mainly the display of license information and ideally also the display of preview images.
ODS-2	**Focus on data quality instead of quantity.** While increasing the number of available datasets was a well-taken goal during the first years after the portal's launch, data quality has become more and more the focus of the users. Mediocre data quality bears a significant risk that the portal is regarded as useless by potential users. Decisive action is needed to gradually improve the quality of the metadata and the actual datasets provided through the open data portal.
ODS-3	**Clarify the intended pace of innovation.** Concerning the future activities of the eCH Specialized Group "Open Government Data", the provider of opendata.swiss should clarify, after consultation with the Swiss open data community, what their intended pace of innovation at the level of the metadata portal is. The Swiss standardization body can then accordingly choose the best scenario regarding the further development of DCAT-AP CH.

---

[34] Several issues that call for minor corrections of DCAT-AP CH have been identified during the last meeting of the eCH Specialized Group "Open Government Data" on 2 November 2018.

ID	Recommendation
ODS-4	**Support the transition towards linked open data.**
	The transition towards linked open data should be supported by publishing the data catalogue as linked open data and by automatically extracting and displaying information about the data models, ontologies, and controlled vocabularies used in the various datasets that have been published as linked open data and by providing information about the authority files/base registers that are linked to by the referenced datasets (cf. section 9.4).

11.3 Recommended actions for the Confederation's OGD and Linked Data projects

Starting from 2019, the Confederation's OGD Project will be coordinated by the Federal Statistical Office, while the Linked Data Project continues to be led by the Swiss Federal Archives. The following table summarizes the actions that should be taken in close cooperation between the two projects.

ID	Recommendation
LD-1	**Ensure close cooperation between different federal agencies**.
	The operator of opendata.swiss and the provider of the Confederation's Linked Data Service (in continuation of the LINDAS project, under the lead of the Federal Archives) need to work hand in hand to advance the provision of linked open data by the public sector in Switzerland. As has been shown in the present report, most stakeholders strongly advocate for the publication of linked open data.
LD-2	**Ensure that base registers be published as linked open data**.
	The operator of opendata.swiss and the provider of the Confederation's Linked Data Service should take the necessary actions to ensure that important base registers that do not contain personal data are published as linked open data.
	Note that data protection issues may arise when publishing and linking other datasets to such base registers. These issues need to be addressed in the context of the publication of such datasets; respective guidelines for data holders need to be provided.
LD-3	**Foster linked data publication and use.**
	Linked data publication and its use should be fostered through adequate measures, which comprise:
	- systematically including on open data portals metadata about datasets that have been published as linked open data and providing direct pointers to the respective SPARQL endpoint and example queries;
	- using linked open data hackathons to raise awareness and skills and to foster coordination;
	- fostering the coordinated publication of linked open data in a given area by all players in the area and/or by the top-level aggregator of such data in Switzerland;
	- encouraging specialized communities to develop relevant ontologies in their fields.

12 Bibliography

Arnold, J., Gubler, J., Huber, C. H., Lang, J., & Noyer, F. (2017). Whitepaper Archivportale, Version 1.0, VSA-AAS. Retrieved from: http://vsa-aas.ch/wp-content/uploads/2017/09/Whitepaper_Portale.pdf (06.03.2018).

Cochez, M., Karim, N., & Dimitriadis, I. (2017). Analysis of the DCAT-AP extensions. https://joinup.ec.europa.eu/sites/default/files/document/2017-10/DCAT-AP%20extensions%20analysis_v1.00.pdf (19.06.2018).

eCH OGD (2018a). Minutes of the plenary meeting of 7 May 2018. eCH Specialized Group "Open Government Data". Internal document.

eCH OGD (2018b). eCH-0200 DCAT-Anwendungsprofil für Datenportale in der Schweiz (DCAT-AP CH). Version 1.0. https://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0200&documentVersion=1.0 (19.06.2018).

Estermann, B. (2017). Swiss Open Cultural Data Hackathon 2017: Final Report. Open GLAM CH. Retrieved from: https://glam.opendata.ch/wordpress/files/2017/10/GLAMHack2017_Final-Report_20171026.pdf (05.06.2018).

Estermann, B., Dungga Winterleitner, A., & Fraefel, M. (2017). Zukunft Memobase: Ergebnisse einer Studie im Auftrag des Vereins Memoriav zur künftigen Ausrichtung der Memobase. Berner Fachhochschule, E-Government-Institut.

Estermann, B. (2018). eCH-0205: Linked Open Data. Accessory Document. eCH association, Zurich.

Estermann, B., Fraefel, M., Neuroni, A. C., & Vogel, J. (2018). Conceptualizing a National Data Infrastructure for Switzerland, Information Polity, 23(1), 43-65.

Lee, J. (2017). Licensing Open Government Data. The Chinese University of Hong Kong. Faculty of Law. Research Paper No. 2017-07. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2964704 (12.12.2018).

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. Government Information Quarterly, 32(3), 323-332.

PwC EU Services (2013). DCAT Application Profile for data portals in Europe. Version 1.01. https://joinup.ec.europa.eu/sites/default/files/distribution/2014-05/DCAT-AP_Final_v1.01.pdf (19.06.2018).

PwC EU Services (2016a). GeoDCAT-AP: A geospatial extension for the DCAT application profile for data portals in Europe. Version 1.0.1. https://joinup.ec.europa.eu/sites/default/files/distribution/2016-08/geodcat-ap_v1.0.1.pdf (19.06.2018).

PwC EU Services (2016b). StatDCAT-AP – DCAT Application Profile for description of statistical datasets. Version 1.0.0. https://joinup.ec.europa.eu/sites/default/files/distribution/2016-12/statdcat-ap_version_1.0.0_1.pdf (19.06.2018).

PwC EU Services (2017). DCAT Application Profile for data portals in Europe. Version 1.1. https://github.com/SEMICeu/DCAT-AP/raw/master/releases/1.1/dcat-ap_1.1.pdf (19.06.2018).

Sklarß, S., Gattwinkel, D., Göldner, A., Horn, C., Wittig, C. & Koroyin, O. (2018). DCAT-AP.de Konventionenhandbuch, V1.0.1. https://www.dcat-ap.de/def/dcatde/1.0.1/implRules.pdf (22.11.2018).

Wiedmer, A. & Seiberth, C. (2015). Konzept: Rechtliche Rahmenbedingungen zur Publikation von Daten als Open Government Data (OGD). OGD Schweiz. https://www.egovernment.ch/index.php/download_file/force/473/3631/ (12.12.2018).