

DATA PAPER

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

Riccardo Albertoni^{1†}, David Browning², Simon Cox³, Alejandra N. Gonzalez-Beltran⁴,
Andrea Perego⁵, Peter Winstanley⁶

¹Istituto di Matematica Applicata e Tecnologie Informatiche "Enrico Magenes",
Consiglio Nazionale delle Ricerche (IMATI-CNR), Genoa, Italy

²Independent Consultant, United Kingdom

³Land and Water, CSIRO, Melbourne, Australia

⁴Scientific Computing, Rutherford Appleton Laboratory, Science and Technology
Facilities Council, Harwell Cam- pus, OX11 0QX, United Kingdom

⁵European Parliament, Luxembourg

⁶Semantic Arts, Fort Collins, CO 80524, USA

Keywords: Data Catalogs; Interoperability; Discoverability; FAIR data

Citation: Albertoni R., Browning D., Cox S., et al.: The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake. *Data Intelligence* 6(2), 457–487 (2024). doi: https://doi.org/10.1162/dint_a_00241

Submitted: November 21, 2023; Revised: December 10, 2023; Accepted: December 15, 2023

ABSTRACT

DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. Since its first release in 2014 as a W3C Recommendation, DCAT has seen a wide adoption across communities and domains, particularly in conjunction with implementing the FAIR data principles (for findable, accessible, interoperable and reusable data). These implementation experiences, besides demonstrating the fitness of DCAT to meet its intended purpose, helped identify existing issues and gaps. Moreover, over the last few years, additional requirements emerged in data catalogs, given the increasing practice of documenting not only datasets but also data services and APIs. This paper illustrates the new version of DCAT, explaining the rationale behind its main revisions and extensions, based on the collected use cases and requirements, and outlines the issues yet to be addressed in future versions of DCAT.

[†] Corresponding Author: Riccardo Albertoni (E-mail: riccardo.albertoni@cnr.it; ORCID: 0000-0001-5648-2713).

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

1. INTRODUCTION

Data has become the most important asset that enables addressing issues ranging from societal challenges, such as pandemics and climate change, to everyday business insights. Thus, data descriptions and data cataloging are fundamental for supporting these data-driven approaches. The last few years have seen an increase in the trend towards Open Data, originally related primarily to *public sector information*, and then with increasing emphasis on facilitating the sharing and re-use of *research data*—for example, the Research Data Alliance (RDA)^① and funder policies—, as well as an understanding of the importance of metadata—for example, with the uptake of FAIR data principles [1] for Findable, Accessible, Interoperable and Reusable data. Besides enabling data discovery and re-use, metadata is now also considered crucial to providing all the information necessary to reproduce an experiment—not only in order to verify the research results in scientific studies, but also in cases where data are used in support to policy making and impact assessment in the public sector. In addition, the qualitative and quantitative costs of not providing FAIR data and metadata have been estimated to be really high: an estimated impact of €10.2 bn for the European economy [2].

The Data Catalog Vocabulary, or DCAT, is a notable contribution to this picture. DCAT is a metadata vocabulary designed to facilitate interoperability between data catalogs published on the Web, irrespective of the domain, community, or platform. Consequently, by using DCAT, data published on the web can be exchanged between systems in an unambiguous manner and with a shared meaning. It was developed following the World Wide Web Consortium (W3C) standardization processes.

Originally developed and hosted at the Digital Enterprise Research Institute (DERI), DCAT was considered by the W3C e-Government Interest Group, and further refined by the Government Linked Data (GLD) Working Group, which published it as a W3C Recommendation in 2014 [3]. Since then, it has been adopted and adapted by different parties—a notable example being DCAT-AP [4], the profile of DCAT being used across Europe as metadata interchange format.

In this paper, we describe the revision of DCAT, referred to as DCAT 2, which was developed by the W3C Dataset Exchange Working Group (DXWG)^② in response to a new set of use cases and requirements gathered from implementation experiences with the original version (2014) of the W3C DCAT vocabulary, and new applications that were not considered at that time. These encompass the ability to categorize datasets and other resource types, such as data services, and elucidate connections between datasets and among datasets and other cataloged resources. Moreover, it involves representing information about quality and representing various types of identifiers. These requirements play a pivotal role in facilitating the exchange of datasets across diverse communities, establishing their origin, suitability for specific purposes, and deduplication in the case distinct catalogs harvest the same data. The use cases and the requirements stemming from the use cases are collected in the report [5].

^① <https://www.rd-alliance.org/> (accessed 10 February 2023)

^② <https://www.w3.org/2017/dxwg/> (accessed 10 February 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

Overall, DCAT 2 harmonizes approaches emerging from different communities of usage, extending the core on which profiles can ensure the uniformity of semantics required for a lossless interoperability.

DCAT 2 was published as a W3C Recommendation in February 2020 [6]. This paper complements the formal recommendation, offering insights into the requirements and the process considered in the new version of DCAT.

The paper is organized as follows. Section 2 explains the methodology, detailing the design principles adopted for the development of DCAT 2. Section 3 gives a brief summary of the requirements that drove the revision. Section 4 presents the DCAT model and highlights the features and guidelines introduced in DCAT 2. Section 5 reviews and discusses contributions in relation to other well-known metadata vocabularies. Section 6 discusses the implementation evidence and the uptake of DCAT. Finally, Section 7 summarizes the contributions and outlines future activities.

2. METHODOLOGY AND DESIGN PRINCIPLES

The revision of DCAT has been developed by the W3C Data Exchange Working Group (DXWG), which was chartered to maximize interoperability between services such as data catalogs, e-infrastructures, and virtual research environments.^③ The revision of DCAT was one of the planned deliverables, together with two other specifications concerning guidelines for the publication of application profiles and profile-based content negotiation.

DCAT 2 is released as a W3C recommendation. W3C recommendations are recognized as a web standard. Gaining and maintaining the group's consensus on technical issues, ensuring the group members' engagement, ensuring a wide public review, and demonstrating implementation for each recommended feature are the main challenges in developing a W3C recommendation. The DXWG leveraged the W3C process to tackle these challenges. The W3C Process [7] is designed to foster consensus, uphold quality, garner endorsement, and promote adoption within both W3C members and the broader community. Moreover, W3C equips working groups with a suite of tools, including IRC channels for minute-taking, proposal voting, and member comment queuing during group calls, a GitHub repository to collaborate on recommendation writing and track the discussion of technical discussion and issues, a wiki to collect meeting agendas and public and member-restricted mailing lists. The group discussions took place in circa 130 teleconferences and four face-to-face meetings, as well as via the DXWG mailing list, issue tracker and GitHub repository. Following the formal W3C process, all these resources are publicly available, including the agenda and minutes of each meeting.^④

The efforts of DXWG have focused on fulfilling requirements expressed in a W3C Working Group Note, the *Dataset Exchange Use Cases and Requirements* [5], which documents 51 use cases collected by the working group, and from which the requirements for the revision were identified. Beside the use

^③ See the DXWG charter: <https://www.w3.org/2017/dxwg/charter> (accessed 10 February 2023)

^④ All these resources are publicly available from the DXWG wiki: <https://www.w3.org/2017/dxwg/> (accessed 06 March 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

cases and requirements documented in [5], the working group took into account the feedback received in response to four intermediate versions of the specification, consisting of three public Working Drafts and a Candidate Recommendation, each publicized within relevant communities.

This paper explicitly refers to requirements and technical design issues to guide interested readers into interlinked working group resources, which deepen the discussion and elucidate the design choices made.

The paper references to working group resources as follows:

Issues All the DCAT issues are documented in the GitHub space of the DXWG. The paper cites them in the text by number, e.g., Issue 1009 for <https://github.com/w3c/dxwg/issues/1009>.

Requirements Requirements are documented in [5] and replicated as separated GitHub issues to track discussion and changes triggered by the requirements. The paper refers to them by their handles, also pointing to the related issues when specific discussions need to be referenced. For example, the paper refers to “Dereferenceable identifiers [RDID]” by [RDID], and to its related issue available at <https://github.com/w3c/dxwg/issues/53> as Issue 53.

Use Cases Use Cases are documented in [5]. The paper refers to them by their identifiers. For example, it refers to “Modeling service-based data access [ID18]” as ID18, available at <https://www.w3.org/TR/dcat-ucr/#ID18>.

The working group adhered to the following guiding principles designing DCAT 2.

Preservation of the backward compatibility with existing implementations. In designing DCAT 2, the working group strove to minimize the impact on existing implementations. Governmental agencies have already deployed broadly the DCAT standard, and the working group aimed to preserve current implementations by avoiding the need to enforce changes unless strictly necessary. DCAT 2 does not make obsolete any pre-existing terms, and introduces new practices by complementing those already in place. New implementations of, e.g., application profiles are expected to adopt DCAT 2, while the existing implementations will not need to be upgraded unless owners want to use the new features. In particular, current DCAT deployments that do not overlap with the DCAT 2 new features (e.g., data services, time and space properties, qualified relations, packaging) do not need to change anything to remain conformant with DCAT 2.

Reuse of terms from consolidated metadata vocabularies. DCAT 2 incorporates terms from pre-existing vocabularies where stable terms with appropriate semantics could be found. This is consistent with the Data on the Web Best Practice (DWBP) #15 “Use terms from shared vocabularies, preferably standardized ones, to encode data and metadata.”[8]. DCAT reuses terms from Dublin Core [9], FOAF [10], and PROV-O [11], and defines a minimal set of classes and properties of its own. Informal summary definitions of the externally-defined terms are included in the DCAT vocabulary for convenience, while authoritative definitions are available from the normative references. Changes to definitions in the references, if any, will be expected to take precedence over the summaries given in DCAT.

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

Minimization of the ontological commitment. The group strives to minimize the ontological commitment of DCAT 2. From a practical point of view, that implies avoiding over-axiomatization of DCAT, e.g., by introducing restrictions that might limit the re-usability of DCAT. Moreover, following the DWBP #16 “Choose the right formalization level” [8], DCAT 2 has removed or relaxed domain and range restrictions for properties (such as those concerning the specification of data themes, keywords, and landing pages). As a rule of thumb, DCAT delegates to application profiles the burden of setting restrictions or providing guidelines for specific applications and communities.

Balancing normative specification and Open-World Assumption. The specification of DCAT 2 is influenced by common assumptions made in contexts of the Semantic Web and linked data. In particular, DCAT is a metadata schema based upon the “Open-World Assumption” (OWA), and it is defined by using the Resource Description Framework (RDF) data model [12]. The OWA implies that the metadata schema is not closed, and it can be extended using types and relationships borrowed from other schemas. RDF promotes an inherently machine-actionable approach, where each term in a metadata schema has its own identifier, which can be used to retrieve the term’s semantics, and terms from distinct vocabularies can be jointly used. These assumptions have proven to scale on uncoordinated open environments such as the Web, but the flexibility offered by the OWA must be taken into account when dealing with the notion of conformance. DCAT-compliant catalogs may include additional non-DCAT metadata fields and additional RDF data in the catalog’s RDF description. The contents of all metadata fields that are held in the catalog (and that contain data about the catalog itself), as well as the corresponding cataloged resources and distributions, are included in this RDF description, and are expressed using the appropriate classes and properties from DCAT. All classes and properties defined in DCAT are used consistently with the semantics declared in the DCAT Recommendation. Constraints on instances can be provided using shape languages such as ShEx and SHACL [13, 14, 15].

3. REQUIREMENTS FOR DCAT 2

Table 1 summarizes the requirements addressed by DCAT 2. The following sections present the modeling solution introduced in DCAT 2, which refer to the requirements in the table.

Table 1. Requirements addressed in DCAT 2 identified by their IDs. The loosely-structured catalog requirement (Issue 253) emerged from the community in form of GitHub issue.

Requirement	Description
Dataset access [RDSA]	Provide a way to specify access restrictions for both a dataset and a distribution.
Distribution schema [RDIS]	Define a way to include identification of the schema the described data conforms to.
Spatial coverage [RSC]	Provide means to specify spatial coverage with geometries.

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

Table 1. *Continued*

Requirement	Description
Temporal coverage [RTC]	Allow for specification of the start and/or end date of temporal coverage.
Funding source [RFS]	Provide means to describe the funding (amount and source) of a Dataset (or entire Catalog).
Related datasets [RRDS]	Ability to represent the different relationships between datasets.
Project relation [RPR]	Provide a means to indicate the relation of Datasets to a project.
Dataset publications [RDSP]	Provide a way to link publications about a dataset to the dataset.
Dataset type [RDST]	Provide a mechanism to indicate the type of data being described and recommend vocabularies to use given the dataset type indicated.
Qualified forms [RQF]	Define qualified forms to specify additional attributes of appropriate binary relations (e.g. temporal context).
Loosely-structured catalog [Issue 253]	Provide a best practice for a loosely-structured catalog.
Distribution definition [RDIDF]	Revise definition of Distribution. Provide better guidance for data publishers.
Distribution package [RDIP]	Define way to specify content of packaged files in a Distribution.
Distribution service [RDISV]	Provide a mean to describe that a distribution is provided by a service.
Dereferenceable id [RDID]	Encode identifiers as dereferenceable HTTP URIs.
Primary & alternative id [RIDALT]	Provide means to distinguish the primary and alternative (legacy) identifiers.
Identifier type [RIDT]	Indicate type of identifier (e.g. prism:doi, bibo:doi, ISBN).
Quality-related info [RDQIF]	Define a way to associate quality-related information with Datasets.
Data quality model [RDQM]	Identify common modeling patterns for different aspects of data quality based on frequently referenced data quality attributes found in existing standards and practices.
Dataset citation [RDSC]	Provide a way to specify information required for data citation (e.g., dataset authors, title, publication year, publisher, persistent identifier).
Entailment of Schema.org [RES]	Define schema.org equivalents for DCAT properties to support entailment of Schema.org compliant profiles of DCAT records.

4. DCAT METADATA SCHEMA

The backbone of DCAT 2 [6] consists of three main classes: `dcat:Catalog`, `dcat:Resource`, `dcat:Distribution`. Figure 1 provides an overview of DCAT 2 model, showing the classes of resources that can be members of a Catalog, and the relationships between them. The diagram uses UML-style class notation, but it should be interpreted following the usual RDF Open-World Assumption around the presence/absence of properties, relationships, and cardinalities. To assist in understanding the full scope of each class, the inherited properties are copied down from each super-class. Cardinalities are shown in a few places to reinforce expectations, but these are not axiomatized or enforced in any way by the normative recommendation.

`dcat:Catalog` represents a catalog, which can be seen as a kind of dataset in which each individual item is a metadata record describing a DCAT resource. `dcat:Resource` represents any resource that may be described by a metadata record in a catalog. It is the parent class of `dcat:Dataset` and `dcat:DataService` — the most typical resources types documented in a DCAT catalog. DCAT profiles or applications can define other kinds of resources to be cataloged as sub-classes of `dcat:Dataset`, `dcat:DataService` or `dcat:Resource`. It is worth noting that `dcat:Resource` and its subclasses can be used also for datasets and services which are not included in any catalog. `dcat:Distribution` represents a specific representation of a dataset. A dataset might be available in multiple serializations that may differ in various ways, including natural language, media-type or format, schematic organization, temporal and spatial resolution, level of detail or profiles (which might specify any or all of the above).

DCAT 2 borrows from the Dublin Core Metadata Terms (DCTERMS) vocabulary [9] a set of properties that are transversely applicable to different items, including datasets, data services, catalogs, and distributions. In particular, `dcterms:title` and `dcterms:description` to title and describe items; `dcterms:issued` and `dcterms:modified` to indicate the date of formal issuance and the most recent modification date of an item; `dcterms:license` and `dcterms:rights` to indicate a legal document under which the item is made available and its copyright statements.

To enhance comprehension of DCAT and the newly introduced features in DCAT 2, Appendix A incorporates snippets of serialized RDF in Turtle format. In an effort to showcase a diverse range of DCAT's capabilities while maintaining concise examples, we have opted to illustrate a fictional catalog. This catalog is designed to embody plausible features, offering a comprehensive overview of the DCAT features discussed in this article. Further, even non-fictitious examples can be found in the recommendation of DCAT 2 [6], or in real catalogs adopting DCAT 3 (for example, the DCAT description[©] for the LusTRE framework [16]). Listing 1 defines the namespaces of the vocabularies reused in the following examples. In Listing 2, the entity `ex:catalog` represents a fictional catalog published by the transparency office `ex:transparency-office` (see lines 6 and 17-19), which includes two datasets `ex:dataset-001` and `ex:dataset-002` (see line 11) and a data service `ex:figure-service-001` (see line 13). In Listing 3 - line 29, the dataset `ex:dataset-001` is distributed as a CSV and as its packaged and compressed counterpart represented by the entities `ex:dataset-001-csv` and `ex:dataset-001-targz`, respectively.

© <https://raw.githubusercontent.com/riccardoAlbertoni/LUSTRE-DCAT2/master/LusTRE-DCAT2.ttl> (accessed 06 March 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

411). For example, Listing 3 - line 15 shows the use of `dcterms:conformsTo` to state the conformance of `ex:dataset-001` with the Commission Regulation (EU) No 1089/2010 (represented in lines 47-50).

DCAT 2 elaborates the guidelines to handle licenses and rights (see Issue 114). Different best practices recommend providing data license and right information (e.g. DWBP [8]). However, multiple use cases fall under the umbrella of license and right information. DCAT 2 provides guidelines distinguishing three main cases: one to associate a resource that represents “license”; a second, to associate a resource denoting only access rights (e.g., whether data can be accessed by anyone or just by authorized parties (Req.RDSA, Issue 59)); a third, to cover all the other cases - i.e., statements not concerning licensing conditions and/or access rights (e.g. copyright statements).

For the first case, DCAT 2 recommends the property `dcterms:license` to refer to canonical URLs of well-known licenses such as those defined by Creative Commons usually applied to the datasets’ distributions (see Listing 3 - line 38 applying `cc-by` license to the distribution `ex:dataset-001-csv`). For the second, it recommends the property `dcterms:accessRights` to express statements specify access rights by referring to code lists/taxonomies, such as the access rights code list MDR-AR[®] used in DCAT-AP [4] or the Eprints Access Rights Vocabulary Encoding Scheme[®] (see Listing 3 - lines 16-17 declaring public access for the dataset `ex:dataset-001`). For the third, all the other types of rights statements such as copyright statements, which are not covered by `dcterms:license` and `dcterms:accessRights`, DCAT 2 recommends the property `dcterms:rights` (see Listing 3 - lines 18-20 indicating the copyright of the dataset `ex:dataset-001`). Finally, in the particular case when rights are expressed via Open Digital Right Language (ODRL) policies, DCAT 2 recommends to use the `odrl:hasPolicy` property as the link from the description of the cataloged resource or distribution to the ODRL policy according to the W3C ODRL model [18] and vocabulary [19], in addition to the corresponding DCTERMS property that matches the same ODRL policy type.

The following subsections provide more detailed descriptions of the specific components of DCAT 2.

4.2 Resources

The class `dcatalog:Resource` represents a cataloged resource. In previous versions of DCAT, datasets were the only kind of entities in DCAT catalogs. DCAT 2 newly introduces the `dcatalog:Resource` class, which is an extension point for defining a catalog of any resource. The original `dcatalog:Dataset` is a sub-class of `dcatalog:Resource`. Besides properties transversely applicable, the class `dcatalog:Resource` includes all the properties that were made available in the previous version of DCAT for datasets and might serve for other kinds of resources in DCAT 2. In particular, `dcatalog:landingPage` indicates a Web page that can be navigated in a Web browser to gain access to the resources, the catalog, a dataset, its distributions and/or additional information. `dcatalog:contactPoint`, `dcterms:creator` and `dcterms:publisher` indicate respectively the contact information for the cataloged resource (expressed in `vCard` [20]), the entity responsible for creating the resource and the entity for making the resource available, both expressed as `foaf:Agent`. `dcterms:language`

[®] <https://publications.europa.eu/en/web/eu-vocabularies/at-dataset/-/resource/dataset/access-right> (accessed 10 February 2023)

[®] http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_AccessRights_Vocabulary_Encoding_Scheme (accessed 06 March 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

refers to the natural language used for textual metadata (i.e. titles, descriptions, etc) of a cataloged resource. `dcat:keyword` classifies the resources using free-text keywords, while `dcat:theme` classifies resources with concepts taken from Knowledge Organization Systems (KOS) and possibly available as Linked Data. Listing 3 provides examples for some of the above properties, see lines 1-10. In particular, line 5 states `ex:finance-employee-001` as creator and line 8 declares `ex:finance-ministry` as publisher. Lines 62-73 define a fictional organization, (i.e., Finance Ministry `ex:finance-ministry`), and declare the agent `ex:finance-employee-001` as part of it.

`dcat:Dataset` is a subclass of `dcat:Resource` which represents a collection of data, published or curated by a single agent, and available for access or download in one or more representations, schematic layouts and formats or serializations. The property `dcat:distribution` relates a dataset to its distributions (`dcat:Distribution`), see Listing 3 - line 29.

`dcat:DataService` is a subclass of `dcat:Resource` which represents a Web API or service that provides access to data, specifically to download distributions of a dataset, see the entity `ex:figure-service-001` in Listing 3 - lines 52-58.

Other subclasses of `dcat:Resource` can be defined to support applications that catalog other kinds of resource, for example, “specimens”.

4.2.1 DCAT 2 new features in Resource

DCAT 2 provides flexible mechanisms to indicate the type of cataloged resources (Req. RDST and Issue 64). DCAT can be used to model a variety of resources - including documents, software, images and audio-visual content. To ensure the flexibility potentially required by catalogs serving different communities and application cases, DCAT 2 provides two mechanisms for typing resources. First, a cataloged resource description has an RDF type to denote a sub-class of `dcat:Resource` - initially `dcat:Dataset` and `dcat:DataService`. Second, the property `dcterms:type` may be used to indicate a sub-type. It is strongly recommended that the value of this property is taken from a well-governed and broadly recognized set of resource types (e.g., the DCMI Type vocabulary [9], the DataCite resource types [21], the ISO-19115-1 scope codes [22], the MARC intellectual resource types). Using `dcterms:type` is particularly appropriated for referring to classifications provided by other standards, and to enable interoperability with existing catalogs (see use cases ID8 and ID20). For example, Listing 3 - line 55 uses the softtyping via `dcterms:type` to indicate that the data service `ex:figure-service-001` is a View Service according to the INSPIRE directive. When describing a resource which is not a `dcat:Dataset` or `dcat:DataService`, it is recommended to create a suitable sub-class of `dcat:Resource`, or use `dcat:Resource` with the `dcterms:type` property to indicate the specific type.

DCAT 2 provides information required for data citation (see Req. RDSC and Issue 61). DCAT 2 provides equivalents to all the mandatory elements in DataCite [21]. The original DCAT already supported title, publisher, publication year, resource type, DCAT 2 has specifically considered `dcterms:creator` to indicate creator and it provides guidelines for dealing with different types of identifiers (see section 4.5).

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

DCAT 2 provides a way to deal with a wide set of relations. Resources might be related in many different ways and complex relations might characterize the context in which resources have been created, for example, to track its input data, the software used, the agents and founders involved (e.g., see use cases ID9, ID12, ID31, ID32). The property `dcterms:relation` is recommended for use in the context of a cataloged resource to capture general relationships, including related datasets (Req. RRDS) and the case where the package of resources associated with a cataloged item includes a mixture of representations, parts, documentations and other elements which are not strictly ‘distributions’ of a dataset (see Issue 253 expressing the requirement on loosely-structured catalogs). Listing 4 shows dataset `ex:d33937` that is just a bag of files, in the example, the `dcterms:relation` specifies the files that are contained in the bag. The property `dcterms:relation` is a super-property of a number of more specific properties which express more precise relationships, such as `dcatalog:distribution`, `dcterms:hasPart`, (and its sub-properties `dcatalog:catalog`, `dcatalog:dataset`, `dcatalog:service`), `dcterms:isPartOf`, `dcterms:conformsTo`, `dcterms:isFormatOf`, `dcterms:hasFormat`, `dcterms:isVersionOf`, `dcterms:hasVersion`, `dcterms:replaces`, `dcterms:isReplacedBy`, `dcterms:references`, `dcterms:isReferencedBy`, `dcterms:requires`, `dcterms:isRequiredBy`. The `dcterms:relation` is not inconsistent with a subsequent reclassification with more specific semantics, though the more specialized sub-properties should be used to link a dataset to component and supplementary resources if possible. For example, DCAT 2 uses the property `dcterms:isReferencedBy` to associate the resource described in the catalog with an external resource that references, cites, or points to the cataloged resource. See Listing 3 -line 28 that links the dataset `dataset-001` to the DOI of a (fictional) publication that cites the dataset. By applying this property, DCAT 2 tracks publications that reuse or describe a specific dataset (see Req. RDSP and Issue 63). DCAT 2 tracks the project that has generated a resource: `prov:wasGeneratedBy` links datasets to the projects that have generated them (Req. RPR and Issue 77).

DCAT 2 supports complex non-binary relations. It uses qualified relations to deal with relations not covered by the above or other known properties (e.g., PROV-O properties such as `prov:wasDerivedFrom`, `prov:hadPrimarySource`) and to overcome the limitation related to binary relations (see the requirement “qualified forms” [Req. RQF] discussed in Issue 79). Even when the relations are represented in known properties, there may be the need of providing additional information concerning, e.g., the temporal context of a relationship, which requires the use of a more sophisticated representation, for example, to specify the temporal dimension of a role—i.e., the time frame during which an individual/organization played a given role - and, maybe, also other information e.g., the organization where the individual held a given position while playing that role (see use cases ID19 and ID13, and Issue 66). DCAT 2 models relationships between resources and agents with property `prov:qualifiedAttribution` (for example, the funding source Req. RFS) and relationships between resources with `dcatalog:qualifiedRelation`. Property `prov:qualifiedAttribution` links the resource to instances of the class `prov:Attribution`, which ascribes the resource to an agent indicated by the property `prov:agent`. For example, Listing 5 - lines 1-7 specifies that the dataset `ex:DS987` has been funded by the Department of Education - Australian Government, where the Department of Education is represented via its URI, and the role of funder is borrowed by a controlled list of roles provided by CSIRO. Property `dcatalog:qualifiedRelation` links the resource to a relation `dcatalog:Relationship` involving another resource pointed by the property `dcterms:relation`. The property

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

`dcat:hadRole` is used in `prov:qualifiedAttribution` to denote the relation the resources have and in `dcat:qualifiedRelation` to indicate the roles an agent plays. For example, Listing 5 - lines 9-15 specifies that the dataset `ex:Test987` was originated by `ex:DS987`.

DCAT 2 supports a rich set of temporal and spatial properties to characterize datasets. The previous version of DCAT offered `dcterms:issued`, `dcterms:modified` and `dcterms:accrualPeriodicity` to indicate when a dataset is issued, modified and its update schedule (see Listing 3 - lines 12-14). DCAT 2 adopts new properties specifically dealing with the temporal coverage (Req. RTC). It introduces the property `dcat:temporalResolution` to specify the minimum temporal separation of items in a dataset encoded as `xsd:duration` and adopts `dcterms:temporal` to indicate the temporal extent of a dataset. The extent is expressed as instances of the class `dcterms:PeriodOfTime`, indicating the start and end of the interval by using properties `dcat:startDate` or `time:hasBeginning`, and `dcat:endDate` or `time:hasEnd`, respectively. The interval can also be open - i.e., it can have just a start or just an end (see Issue 85 for further discussions). For example, Listing 3 specifies that the dataset `dataset-001` covers the temporal intervals from July to September with a measure for day (i.e., P1D), see lines 21-25. Similarly, DCAT 2 introduces two new properties to express spatial coverage (Req. RSC, see Issue 83 for the detailed discussion). `dcat:spatialResolutionInMeters` specifies the minimum spatial separation of items in a dataset, expressing it as a decimal values in meters. `dcterms:spatial` expresses the spatial extent of a dataset. Its values are a spatial region or named place `dcterms:Location`, in which, the property `locn:geometry` specifies an extensive geometry (i.e., a set of coordinates denoting the vertices of the relevant geographic area), `dcat:bbox` specifies a geographic bounding box delimiting a spatial area, `dcat:centroid` indicates a geographic center of a spatial area, or another characteristic point. For example, Listing 3 specifies that the dataset `dataset-001` covers the European Union with a resolution of 30 meters, see lines 26-27.

DCAT 2 adds mechanisms for including data services. Data is often served via web services. A service may provide access to more than one dataset, and it is necessary to know how to query the service API to get the data (see use cases ID18 and ID6). DCAT 2 specializes `dcat:Resource` with a new class `dcat:DataService` to model data services (see Issue 180). A data service is a collection of operations that provides access to one or more datasets or to data processing. The `dcat:servesDataset` property links a service to data that it can distribute. The kind of service can be indicated using the `dcterms:type` property; its value may be taken from a controlled vocabulary such as the INSPIRE spatial data service type code list[®]. `dcat:endpointURL` provides the root location or primary endpoint of the service (a Web-resolvable IRI). Property `dcat:endpointDescription` provides a description of the services available via the endpoints, including their operations, parameters, etc. The endpoint description gives specific details of the actual endpoint instances, using `dcterms:conformsTo` to indicate the general standard or specification that the endpoints implement. An endpoint description may be expressed in a machine-readable form, such as an Open API [23] description, an OGC GetCapabilities response WFS [24, 25], WMS [26, 27], a SPARQL Service Description [28], an OpenSearch [29] or WSDL [30] document, a Hydra API description HYDRA [31]. For example, Listing 3 specifies that `ex:figure-service-001` serves the dataset `dataset-001` (line 58) and it is a view service

[®] <http://inspire.ec.europa.eu/metadata-codelist/SpatialDataServiceType/> (accessed 06 March 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

conforming to the OGC Web Map Service 1.3 specification (line 54), available at the indicated endpoint URL (line 57) and described as in the endpoint description (line 56).

4.3 Distributions

dcat:Distribution is a specific class for representation of a dataset. A dataset might be available in multiple serializations that may differ in various ways, including natural language, media-type or format, schematic organization, temporal and spatial resolution, level of detail or profiles (which might specify any or all of the above). Distributions represent a general availability of a dataset, whose access can include different access methods (e.g., direct download, API, or through a Web Page). For the distributions, dcat:downloadURL provides the URL for a downloadable file in a given format. The “format” of a distribution should be specified through the property dcat:mediaType when a correspondent IANA Media Types [32] exists, or dcterms:format otherwise. dcat:byteSize specifies the size of distribution in bytes. When a direct link to the downloadable file is not available, dcat:accessURL indicates a URL of the resource that gives access to a distribution of the dataset. It should be used for the URL of a service or location that can provide access to this distribution, typically through a Web form, query or API call. See Listing 3 lines 31-37 for examples of the above properties.

4.3.1 DCAT 2 new features in Distributions

DCAT 2 introduces distribution service to support use cases where the distribution of a dataset is made by Web services (ID6 and Req. RDISV). DCAT 2 adds the property dcat:accessService which relates distributions to their dcat:DataService detailed information about how users can interact with distribution services (Issue 267).

DCAT 2 revises and clarifies the definition of distribution (Req. RDIDF). The previous definition of dcat:Distribution allowed a number of alternative interpretations. The definition has been rephrased to clarify that distributions are primarily representations of datasets. DCAT 2 clarifies that lossless transformations between representations are not always possible. In some cases, distributions of the same dataset might have different levels of fidelity to the underlying data (see discussion in Issue 52). Moreover, the question of whether different representations can be understood to be distributions of the same dataset, or distributions of different datasets, is application-specific. Judgment about how to describe them is the responsibility of the provider, taking into account their understanding of the expectations of users, and practices in the relevant community.

DCAT 2 supports packaged and compressed distributions (Req. RDIP see Issue 54). Distributions can include multiple files made available in compressed archives. DCAT 2 introduces the property dcat:packageFormat and dcat:compressFormat to indicate the package and compression formats of the distribution. Both formats should be expressed using a media type as defined by IANA [32], if available, see Listing 3 lines 43-45 that represents ex:dataset-001-csv packaged and compressed. *DCAT*

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

2 recommends to indicate distribution schema. It uses the property `dcterms:conformsTo` to indicate the model or schema used for the representation of dataset (Req. RDIS and Issue 55), see Listing 3 line 34.

4.4 Catalog and Catalog Record

A `dcatalog:Catalog` is a curated collection of metadata about resources such as datasets and data services. `dcatalog:Catalog` is characterized by further properties besides those transversely applicable: `foaf:homepage` indicates the homepage of the catalog which usually is a public Web document available in HTML; `dcatalog:themeTaxonomy` refers to the Knowledge Organization System (KOS) providing concepts to classify the cataloged resources; `dcatalog:record` links a catalog to a `dcatalog:CatalogRecord` describing the registration of a single cataloged resource that is part of the catalog. Using `dcatalog:record` and `dcatalog:CatalogRecord` is possible to distinguish between the metadata of a cataloged resource (i.e., instances of `dcatalog:Resource`) and the metadata of the metadata of the cataloged resource (i.e., instances of `dcatalog:CatalogRecord`). This is required in specific cases, for example, to express the date when a resource has been registered or modified in the catalog (`dcterms:issued` and `dcterms:modified` attributed to instances of `dcatalog:CatalogRecord`), which may differ from the publication or modification of the concrete resources (aka `dcterms:issued` or `dcterms:modified` attributed to instances of `dcatalog:Resource`), see Listing 2 - line 24: `ex:record-002` specifies via `dcterms:issued` when `ex:dataset-002` has been included in the catalog.

4.4.1 DCAT 2 new features in Catalog and Catalog Record

DCAT 2 clarifies the scope of DCAT catalogs. DCAT was originally conceived to model data catalogs. DCAT 2 opens to novel first-class cataloged resources providing `dcatalog:Resource` as an extension point for community-specified cataloged resources (see Issue 172 and section 4.2). It adds `dcatalog:DataService` for representing data services and subsumes `dcatalog:Dataset` and `dcatalog:DataService` with `dcatalog:Resource`. It provides properties to deal with the new kinds of cataloged resources (see Issue 116): `dcterms:hasPart`, to specify a cataloged resource irrespective of its type; `dcatalog:service`, to specify a cataloged data service, see Listing 2 - lines 15 and 13, respectively.

DCAT 2 enables provision for catalogs to be composed of other catalogs, in particular, `dcatalog:Catalog` has been made a sub-class of `dcatalog:Dataset`, and the property `dcatalog:catalog` is provided to specify sub-catalogs (see Issue 182), see Listing 2 - line 14.

DCAT 2 extends the type of thematic resources which can be considered to classify datasets. It relaxes the global range of the property `dcatalog:themeTaxonomy` allowing the linking to a KOS that is not formalized as a `skos:ConceptScheme` (See Issue 119). Beside SKOS concept schemes, SKOS collections [33, 34] or OWL ontologies [35] are recommended advising that each member of the KOS can be denoted by an IRI and published as linked data, see Listing 2 - line 34 that implies the sub-catalog `dataset-002` uses DBPEDIA categories as theme taxonomy.

DCAT 2 includes specific mechanisms to state the conformance of metadata to standards. It adopts the property `dcterms:conformsTo` for `dcatalog:CatalogRecord` to represent the conformance of a record metadata

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

with a metadata standard (see Issue 502), see Listing 2 ex:record-002 - line 25 implies the dataset-002 metadata conforms to DCAT 2.

4.5 Guidelines

In addition to the feature discussed above, DCAT 2 elaborates guidelines to meet specific requirements posed by the community. Guidelines systematize emerging solutions based on W3C vocabularies such as DQV [17] and ADMS [36] which are stable enough to be adopted even if they have not reached the status of W3C recommendation.

DCAT 2 provides guidelines to deal with different kinds of identifiers. As pointed out in the use case ID11, a number of different (possibly persistent) identifiers are widely used in the scientific community, especially for publications, but now increasingly for authors and data. Different approaches are used for representing them, best practices are needed to enable their effective use across platforms. But more importantly, they need to be made actionable, irrespective of the platforms they are used in (see Req. RIDID). Encoding identifiers as HTTP URIs seems to be the most effective way of making them actionable. Notably, quite a few identifier schemes can be encoded as dereferenceable HTTP URIs, and some of them are also returning machine-readable metadata (e.g., DOIs, ORCIDs). Moreover, they can still be encoded as literals, especially if there is the need of knowing the identifier type (Req. RIDT). In such a case, a common identifier type registry would ensure interoperability. DCAT 2 reuses terms provided by DCTERMS [9] and VOCAB-ADMS [36]. Data providers can apply `dcterms:identifier` to any kind of resources binding their HTTP dereferenceable proxy IDs with legacy identifiers, non-HTTP dereferenceable identifiers, locally minted or third-party-provided identifiers (Issue 53). Another issue concerns the ability to specify primary and secondary identifiers. This may be a requirement when resources are associated with multiple identifiers (Req. RIDALT). The property `adms:identifier` can express other locally minted identifiers or external identifiers, like DOI, ELI, arXiv for creative works, and ORCID, VIAF, ISNI for actors such as authors and publishers, as long as the identifiers are globally unique and stable. The property `adms:identifier` ranges in instances of the class `adms:Identifier`, for which `skos:notation` indicate the identifier as a literal with datatype IRI (e.g., "PA 1-060-815"^^`ex:type`), `adms:schemaAgency` and `dcterms:creator` represent the authority that defines the identifier scheme (e.g., the `ex:type` in the example). `adms:schemaAgency` is used when the authority has no URI associated (see Issue 67). The type of identifiers can be provided as RDF datatypes [12] or custom OWL datatypes [37] if not already registered as URI type. Examples of common types for identifier scheme (arXiv, etc.) are defined in DataCite schema[®] and FAIRsharing Registry[®] (see Issue 68). Specific examples dealing with identifiers are available in the DCAT recommendation, section "Dereferenceable identifiers"[®].

DCAT 2 provides guidelines for documenting the quality of resources and distributions. Consistently with the recommendations from the Data on the Web Best Practices (DWBP) [8], the use cases ID45

[®] <https://schema.datacite.org/meta/kernel-4.1/include/datacite-relatedIdentifierType-v4.xsd> (accessed 6 March 2023)

[®] <https://fairsharing.org/search?q=identifier> (accessed 6 March 2023)

[®] <https://www.w3.org/TR/vocab-dcat-2/#dereferenceable-identifiers> (accessed 11 November 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

and ID14 stress the need for a uniform representation of data quality so that consumers understand the possibilities and risks of using and reusing the data. DCAT 2 reuses the Data Quality Vocabulary (DQV) [38] [17] to associate quality-related information to datasets (Req. RDQIF) and offer common modeling patterns for different aspects of Data Quality (see, Req. RDQM, Issue 57, Issue 58). The property `dqv:hasQualityAnnotation` relates datasets and distributions with reviews, users' feedback and quality certificates (modeled as `dqv:QualityAnnotation`). The property `dqv:hasQualityMeasurement` relates resources and distributions to quality measurements (`dqv:QualityMeasurement`) evaluated by community-defined domain-specific metrics (`dqv:Metric`) which provide quantitative or qualitative information about the dataset or distribution. `dqv:QualityPolicy` models policies or agreements that are chiefly governed by data quality concerns. As previously discussed, `dcterms:conformTo` can state the compliance with standards, specifications. DCAT 2 includes examples of how DQV can express the degree of conformance to best practices (e.g. the DWBP [8] or the FAIR Principles [1]) and combines DQV with the Evaluation and Report Language (EARL) [39] and PROV ontology [11] to express details about the results of conformance and quality tests. Examples of documenting the quality of resources and distributions are available in the DCAT recommendation, section "Quality information"[Ⓞ].

5. RELATED WORK

This section reviews metadata models that readers might perceive as overlapping with DCAT in terms of coverage or goals. The discussion points out the distinct metadata models' peculiarities and their mapping into DCAT. Overall, the discussion clarifies that DCAT is not redundant with the existing metadata models. Instead, a joint of the discussed metadata models with DCAT brings advantages in the overall metadata expressivity and cross-sector, cross-platform sharing, and reuse.

CERIF. The Common European Research Information Format (CERIF) models Research Environment, including research outputs, persons, organizations, projects, funding programs, facilities as first-class citizens and capturing the semantic relationships of entities with each other as well as entity classifications (i.e. roles). The European Commission mandated euroCRIS to maintain, develop and promote CERIF as an EU recommendation to Member States. euroCRIS now has more than 100 institutional members in approximately 40 countries and there are hundreds of implementations of CERIF, including by several commercial ICT suppliers. CERIF is currently being used in numerous systems in production across Europe (e.g., national or institutional research information systems), as well as in European FP7 e-infrastructure projects, such as OpenAIREplus, EuroRIs-Net+ and ENGAGE [40]. CERIF and DCAT differ in terms of goals and specificity. CERIF specifically focuses on research environments, while DCAT focuses on Data Catalogs. Partial mapping of DCAT into CERIF exists [41]. For example, DCAT Datasets can be modeled as ResultProduct, but CERIF does not natively provide distinctions between catalogs, datasets, distributions, nor other details such as access details.

DataCite. The DataCite metadata schema [21] is a list of core metadata properties chosen for accurate and consistent identification of a resource for citation and retrieval purposes, along with recommended

[Ⓞ] <https://www.w3.org/TR/vocab-dcat-2/#quality-information> (accessed 11 November 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

use instructions. It is managed by the DataCite consortium, founded in late 2009 with the goal of easing the access to scientific research data on the Internet, increasing acceptance of research data as legitimate, citable contributions to the scientific record, and supporting data archiving that will permit results to be verified and re-purposed for future study. DataCite infrastructure is responsible for issuing persistent identifiers (in particular, DOIs) for datasets, and for registering dataset metadata. Such metadata is to be provided according to the DataCite metadata schema. While DataCite's Metadata Schema has been expanded with each new version, it is, nevertheless, intended to be generic to the broadest range of research datasets, rather than customized to the needs of any particular discipline. DataCite metadata primarily supports citation and discovery of data; It does not include specific terms for Catalogs and Distributions, it is not intended to supplant or replace community-specific metadata. DataCite enables providing other metadata schemas via DOI content negotiation. In particular, it supports JSON-LD [42] to serve metadata according to Schema.org. A mapping from DataCite to DCAT is defined in CiteDCAT-AP [43], a metadata profile used in Zenodo[®], the most popular European research data repository.

ISO 19115. ISO 19115-1:2014 [22] defines a metadata schema for describing geographic information and services by means of metadata. It provides information about the identification, the extent, the quality, the spatial and temporal aspects, the content, the spatial reference, the portrayal, distribution, and other properties of digital geographic data and services. Mapping of ISO 19115 to DCAT has been developed, in particular, GeoDCAT-AP [44] is an extension to the DCAT application profile for European data portals (DCAT-AP) for the representation of geographic metadata. GeoDCAT-AP was designed to enable the cross-sector and cross-platform sharing and re-use of INSPIRE and, more in general, metadata following the ISO 19115/19119 standards and the corresponding XML-based implementation (ISO 19139).

Schema.org. In 2011, the major search engines Bing, Google, and Yahoo (later joined by Yandex) created Schema.org to provide a single schema across a wide range of topics that included people, places, events, products, offers, and so on [45]. Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on Web pages, in email messages, and beyond. Schema.org includes a number of types and properties based on the original DCAT work (see `sdo:Dataset` as a starting point), and the index for Google's Dataset Search service relies on structured description in Web pages about datasets using both Schema.org and DCAT [46]. This class is modeled starting from W3C DCAT work, and benefits from collaboration around the DCAT, ADMS and VoID vocabularies[®]. In particular, Schema.org mimics the DCAT backbone, the (abstract) `sdo:Dataset` and (concrete) `sdo:DataDownload` matches `dc:Dataset` / `dc:Distribution`, as for the relationship of Datasets to DataCatalogs. Contrary to DCAT, Schema.org is not a W3C standard, the project is not governed by W3C, the W3C advisory group or the W3C Process; rather, it stems from an informal collaboration. In terms of workflow, the primary difference between Schema.org and W3C's recommendation track process is an emphasis on incremental publication of releases (several releases per year) approved by a small steering group whose role is to evaluate and approve release candidates prepared by the project webmaster on the basis of wider discussion which takes place in a dedicated W3C community group and

[®] <https://zenodo.org/> (accessed 06 March 2023)

[®] See <http://www.w3.org/wiki/WebSchemas/Datasets> (accessed 06 March 2023) for full details and mappings.

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

related GitHub project. DCAT 2 [6] provides a mapping between DCAT and Schema.org to clarify the relation between DCAT and Schema.org and promote the discoverability by mainstream search engines (see Req. RES).

VoID. VoID [47] is an RDF vocabulary for expressing metadata about RDF datasets. It covers (i) general metadata following the Dublin Core model; (ii) access metadata describing how RDF data can be accessed using various protocols; (iii) structural metadata describing the structure and schema of datasets for tasks such as querying and data integration; (iv) description of links between datasets for understanding how multiple datasets are related and can be used together. VoID is quite popular in the context of Linked data and extended by other vocabularies such as DataID [48]. However, being specifically suited for RDF dataset and linked data practices, it does not cover all the types of data required by the open and research data community (e.g., CSV, JSON). Fruitfully jointly use of DCAT and VoID have been shown (e.g., by DataID [48]).

6. DCAT IMPLEMENTATIONS AND UPTAKE

The W3C recommendation process requires the collection of implementation experiences to show that a specification is sufficiently clear, complete, relevant to market needs, and to ensure that independent, interoperable implementations of each feature of the specification are realized. In view of that, the editors of DCAT 2 prepared a DCAT 2 implementation report [49]. The report also shows preliminary evidences of DCAT 2 uptake. It focuses on two types of evidence: i) DCAT-based vocabularies; ii) data catalogs, data services, and datasets.

As for DCAT-based vocabularies, different profiles are based on DCAT 2 [6] or extend the original version of DCAT [3] with properties and classes included in DCAT 2, showing implementation evidences of the reviews included. Due to the large number of DCAT-based vocabularies and data catalogs supporting DCAT, this section includes only a representative subset, providing nonetheless enough implementation evidence of the revisions proposed in DCAT 2.

A noteworthy example of DCAT profiling is the development of DCAT-AP. Developed as part of Interoperable Europe[®] mission to foster data interoperability in Europe, DCAT-AP is a DCAT profile for sharing information about catalogues containing datasets and data services descriptions in Europe, under maintenance by the SEMIC action. The specification was elaborated by a multi-disciplinary Working Group with representatives from 16 European Member States, some European Institutions and the US. DCAT-AP [4] is a profile of DCAT used across Europe since 2014 as a metadata interchange format, primarily for catalogs of government data, and, to some extent, for scientific data. As such, it has a broad geographic coverage, and it is supported in data catalogs (e.g., the European Data Portal[®]) and catalog platforms (e.g., CKAN[®]). Over the course of the last decade, DCAT-AP has evolved into a comprehensive ecosystem that includes numerous interconnected specifications.

[®] <https://joinup.ec.europa.eu/interoperable-europe> accessed 16 November 2023

[®] <https://data.europa.eu/> (accessed 06 March 2023)

[®] <https://ckan.org/> (accessed 06 March 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

GeoDCAT-AP [44] and StatDCAT-AP [50] are domain-specific extensions of DCAT-AP for geospatial and statistical data, respectively, and they share the same geographic coverage of DCAT.

CiteDCAT-AP [43] and DCAT-AP-JRC [51] are extensions of DCAT-AP specifically designed for multidisciplinary research data, and they are implemented in the corporate catalog of the European Commission's Joint Research Centre[®]. Moreover, CiteDCAT-AP is supported in Zenodo[®], the research data catalog and repository most widely used in Europe.

DCAT-AP has also been used as a basis for the development of country-specific extensions (see [52]). Such extensions have not been included in this review, but they provide additional support to the implementation evidence for the revisions proposed in DCAT 2 already included in DCAT-AP.

DCAT-AP aligns with DCAT 2 since version 2.0, and such alignment will eventually be reflected in the DCAT-AP extensions. For example, GeoDCATAP 2.0 [44] (released in December 2020) is aligned with DCAT 2.

Moreover, in the context of scientific data, projects and initiatives such as EOSC-pillar [53], FAIRsFAIR [54] and ExPaNDS encourage data repository owners to publish their datasets by mapping their metadata with the DCAT standard when following the FAIR principles.

DCAT 2 is adopted in FAIRification of Citizen Science platform [55], and open source platforms such as SEEK [56] to improve interoperability between digital assets on the Web and enable cross-domain markup. It is a core building block for developing REST API aiming at creating, storing, and serving FAIR metadata (see FAIR Data Point (FDP) [57]).

DCAT is recommended by the ExPaNDS project as part of its “Final Recommendations for FAIR Photon and Neutron Data Management”[®].

7. CONCLUSION AND FUTURE WORK

DCAT 2 is a metadata schema that facilitates data catalogs' interoperability on the Web. DCAT gives people and machines a specific and domain-independent approach to create catalogs that express the core elements of a dataset description in a standardized way that is suitable for publication on the Web, and enables cross-domain interoperability by being used either on its own or alongside, as a complement to other data catalog standards. Thanks to this, DCAT facilitates effective search and retrieval and permits easy scaling up of the query process either through “frictionless” aggregation of dataset descriptions and catalog records from many different sources and domains, or by applying the same query across multiple catalogs and aggregating the results. These patterns can also be varied slightly so as to provide communities with tailored approaches to the dataset catalog that respect the specific nuances of a particular type of data.

[®] <https://data.jrc.ec.europa.eu/> (accessed 06 March 2023)

[®] <https://zenodo.org/> (accessed 06 March 2023)

[®] <https://doi.org/10.5281/zenodo.6821676> (accessed 17th February 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

DCAT 2 is designed as a community effort by DXWG, adheres to design principles specifically suited to establish it as a lingua franca for exchanging data coming from different catalogs. In particular, the back compatibility with the previous version aims at preserving existing implementations; the reuse of terms from consolidated metadata vocabularies eases the interoperability promoting the adoption of cross-vocabulary modeling patterns; the minimization of ontology commitment opens to its reuse and specialization from the different domain communities; the Open-World Assumption unlocks DCAT complementation with other existing metadata vocabularies.

Building upon the foundational work initially published in 2014, Version 2 significantly influences real-world scenarios, particularly within the realms of public sector information (PSI directive) and the sharing and reuse of research data according to the FAIR principles. DCAT 2 introduces enhancements that enable the comprehensive representation of spatial and temporal coverage, descriptions of data services, and the establishment of complex attribution and relations among cataloged resources. By skillfully leveraging and extending existing vocabularies, DCAT 2 puts forth guidelines for the representation of various identifier types and the quality attributes of cataloged resources. DCAT 2 introduces qualified forms that facilitate the representation of funders and delineate data production chains, referencing specific, community-based sets of roles that are already in use. DCAT serves as a robust backbone onto which a multitude of community-developed controlled vocabularies seamlessly integrate. This adaptability empowers DCAT to accommodate similar requirements arising from different user groups, fostering a common, more inclusive, and comprehensive representation of data across a spectrum of domains.

DCAT editors and DXWG support DCAT 2 adopters by assisting the specific doubts and issues via the DXWG public mailing list[Ⓜ] and related GitHub space[Ⓜ]. Further DCAT releases are planned, DXWG is discussing including a more explicit notion of data series and versioning in DCAT. Going forward, the WG expects the incorporation of classes to describe data services into the model will make DCAT an increasingly useful tool in data science and provide a well-trodden path for those implementing the FAIR principles to follow. Exploring DCAT anti-patterns is another promising avenue that deserves future consideration. This exploration might require broadening the collection of implementations and delving into a representative corpus of published DCAT fragments. Examining a representative set of published DCAT fragments is particularly intriguing, as it would offer an opportunity to identify and document potential misuse, which can be addressed in subsequent rounds of DCAT standardization.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the contributions made to DCAT version 2 by all members of the working group, especially Annette Greiner, Antoine Isaac, Armin Haller, Dan Brickley, Ine de Visser, Jaroslav Pullmann, Lars G. Svensson, Linda van den Brink, Makx Dekkers, Nicholas Car, Rob Atkinson, Tom Baker.

[Ⓜ] <https://lists.w3.org/Archives/Public/public-dxwg-wg/> (accessed 06 March 2023)

[Ⓜ] <https://github.com/w3c/dxwg> (accessed 06 March 2023)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

The authors also gratefully acknowledge the chairs of the Data eXchange Working Group: Karen Coyle, Caroline Burle and Peter Winstanley and W3C staff contacts Philippe Le Hgaret, Phil Archer and Dave Raggett.

Riccardo Albertoni was partially supported by TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No 952215.

David Browning's work on this was funded by refinitiv.com (previously Thomson Reuters).

AUTHOR CONTRIBUTIONS

Riccardo Albertoni: Conceptualization, Methodology, Writing-Original draft preparation, Project administration;

David Browning: Conceptualization, Methodology, Project administration;

Simon Cox: Conceptualization, Methodology, Writing- Reviewing and Editing, Project administration;

Alejandra Gonzalez Beltran: Conceptualization, Methodology, Writing-Reviewing and Editing, Project administration;

Andrea Perego: Conceptualization, Methodology, Writing-Reviewing and Editing, Investigation, Project administration;

Peter Winstanley: Conceptualization, Methodology, Writing-Reviewing and Editing, Project administration.

REFERENCES

- [1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I.J., et al: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), (2016)
- [2] European Commission. Directorate General for Research and Innovation., PwC EU Services., Cost-benefit analysis for FAIR research data: cost of not having FAIR research data., Publications Office, LU. Available at: <https://data.europa.eu/doi/10.2777/02999> (2018). Accessed 11 April 2024
- [3] Maali, F., Erickson, J.: Data Catalog Vocabulary (DCAT), W3C Recommendation, W3C. Available at: <http://www.w3.org/TR/2014/REC-vocab-dcat-20140116/> (2014). Accessed 11 April 2024
- [4] Nuffelen B. V.: DCAT Application Profile for data portals in Europe. Version 2.1., Technical Specification, Euro-pean Commission. Available at: <https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe> (2020). Accessed 11 April 2024
- [5] Pullmann, J., Atkinson, R., Isaac, A., Faniel, I.: Dataset Exchange Use Cases and Requirements, W3C Working Group Note, W3C. Available at: <https://www.w3.org/TR/2019/NOTE-dcat-ucr-20190117/> (2019). Accessed 11 April 2024
- [6] Albertoni, R., Browning, D., Cox, S., Beltrn, A. G., Perego, A., Winstanley, P.: Data Catalog Vocabulary (DCAT)-Version 2, W3C Recommendation, W3C. Available at: <https://www.w3.org/TR/2020/REC-vocab-dc-at-2-20200204/> (2020). Accessed 11 April 2024

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

- [7] Rooney, N., Rivoal, F.: World wide web consortium process document, Available at: <https://www.w3.org/2019/Process-20190301/> (2019). Accessed 11 April 2024
- [8] Burle, C., Loscio, B. F., Calegari, N.: Data on the Web Best Practices, W3C Recommendation, W3C. Available at: <https://www.w3.org/TR/2017/REC-dwbp-20170131/> (2017). Accessed 11 April 2024
- [9] DCMI Usage Board, DCMI Metadata Terms, DCMI Recommendation, DCMI, Available at: <http://dublincore.org/documents/dcmi-terms/> (2012). Accessed 11 April 2024
- [10] Brickley, D., Miller, L.: FOAF Vocabulary Specification 0.99, Namespace document. Available at: <http://xmlns.com/foaf/spec/> (2014). Accessed 11 April 2024
- [11] McGuinness, D., Lebo, T., Sahoo, S.: PROV-O: The PROV Ontology, W3C Recommendation, W3C. Available at: <http://www.w3.org/TR/2013/REC-prov-o-20130430/> (2013). Accessed 11 April 2024
- [12] Wood, D., Cyganiak, R., Lanthaler, M.: RDF 1.1 Concepts and Abstract Syntax, W3C Recommendation, W3C. Available at: <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/> (2014). Accessed 11 April 2024
- [13] Labra Gayo, J. E., Prud'hommeaux, E., Boneva, I., Kontokostas, D.: Validating RDF Data, volume 7 of Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool Publishers LLC (2017). <http://dx.doi.org/10.2200/s00786ed1v01y201707wbe016>
- [14] Prud'hommeaux, E., Boneva, I., Labra Gayo, J. E., Kellogg, G.: Shape Expressions Language 2.1, W3C Community Group Report, Shape Expressions Community Group. Available at: <http://shex.io/shex-semantics/> (2019). Accessed 11 April 2024
- [15] Knublauch, H., Kontokostas, D.: Shapes Constraint Language (SHACL), W3C Recommendation, W3C. Available at: <https://www.w3.org/TR/2017/REC-shacl-20170720/> (2017). Accessed 11 April 2024
- [16] Albertoni, R., Martino, M. D., Podestà, P., Abecker, A., Wössner, R., Schnitter, K.: Lustre: A framework of linked environmental thesauri for metadata management. *Earth Science Informatics*, 11, 525–544 (2018)
- [17] Albertoni, R., Isaac, A.: Data on the Web Best Practices: Data Quality Vocabulary, W3C Note, W3C. Available at: <https://www.w3.org/TR/2016/NOTE-vocab-dqv-20161215/> (2016). Accessed 11 April 2024
- [18] Iannella, R., Villata, S.: ODRL Information Model 2.2, W3C Recommendation, W3C. Available at: <https://www.w3.org/TR/2018/REC-odrl-model-20180215/> (2018). Accessed 11 April 2024
- [19] Myles, S., Rodríguez-Doncel, V., Iannella, R., Steidl, M.: ODRL Vocabulary & Expression 2.2, W3C Recommendation, W3C. Available at: <https://www.w3.org/TR/2018/REC-odrl-vocab-20180215/> (2018). Accessed 11 April 2024
- [20] McKinney, J., Iannella, R.: vCard Ontology-for describing People and Organizations, W3C Note, W3C. Available at: <http://www.w3.org/TR/2014/NOTE-vcards-20140522/> (2014). Accessed 11 April 2024
- [21] DataCite Metadata Working Group, DataCite Metadata Schema, Technical Specification, DataCite e.V.. Available at: <https://schema.datacite.org/> (2021). Accessed 11 April 2024
- [22] ISO/TC 211, ISO-19115-1:2014: Geographic information-Metadata-Part 1: Fundamentals, International Standard, ISO. Available at: <https://www.iso.org/standard/53798.html> (2014). Accessed 11 April 2024
- [23] Miller, D., Whitlock, J., Gardiner, M., Ralphson, M., Ratovsky, R., Sarid, U.: OpenAPI Specification v3.1.0, Technical Specification, OpenAPI Initiative. Available at: <https://spec.openapis.org/oas/v3.1.0> (2021). Accessed 11 April 2024
- [24] Vretanos, P. A.: Web Feature Service 2.0 Interface Standard, OGC Interface Standard, OGC. Available at: <http://www.opengeospatial.org/standards/wfs> (2014). Accessed 11 April 2024
- [25] ISO/TC 211, ISO 19142:2010: Geographic information Web Feature Service, International Standard, ISO. Available at: <https://www.iso.org/standard/42136.html> (2010). Accessed 11 April 2024
- [26] de la Beaujardiere, J.: Web Map Service Implementation Specification, OpenGIS Implementation Standard, OGC. Available at: <http://www.opengeospatial.org/standards/wms> (2016). Accessed 11 April 2024

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

- [27] ISO/TC 211, ISO 19128:2005: Geographic information Web map server interface, International Standard, ISO. Available at: <https://www.iso.org/standard/32546.html> (2005). Accessed 11 April 2024
- [28] Williams, G.: SPARQL 1.1 Service Description, W3C Recommendation, W3C. Available at: <https://www.w3.org/TR/2013/REC-sparql11-service-description-20130321/> (2013). Accessed 11 April 2024
- [29] Clinton, D.: OpenSearch, Technical Specification. Available at: <https://github.com/dewitt/opensearch/blob/master/opensearch-1-1-draft-6.md> (2018). Accessed 11 April 2024
- [30] Chinnici, R., Moreau, J. J., Ryman, A., Weerawarana, S.: Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language, W3C Recommendation, W3C. Available at: <http://www.w3.org/TR/2007/REC-wsdl20-20070626> (2007). Accessed 11 April 2024
- [31] Lanthaler, M.: Hydra Core Vocabulary, Unofficial Draft, Hydra Community Group. Available at: <https://www.hydra-cg.com/spec/latest/core/> (2018). Accessed 11 April 2024
- [32] IANA, IANA Media Types. Available at: <https://www.iana.org/assignments/media-types/> (2021). Accessed 11 April 2024
- [33] Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., Summers, E.: Key choices in the design of simple knowledge organization system (SKOS), *J. Web Semant.* 20, 35–49 (2013)
- [34] Bechhofer, S., Miles, A.: SKOS Simple Knowledge Organization System Reference, W3C Recommendation, W3C. Available at: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/> (2009). Accessed 11 April 2024
- [35] McGuinness, D., Bao, J., Kendall, E., Patel-Schneider P.: OWL 2 Web Ontology Language Quick Reference Guide (Second Edition), W3C Recommendation, W3C. Available at: <http://www.w3.org/TR/2012/REC-owl2-quick-reference-20121211/> (2012). Accessed 11 April 2024
- [36] Archer, P., Shukai, G., Dekkers, M.: Asset Description Metadata Schema (ADMS), W3C Note, W3C. Available at: <https://www.w3.org/TR/vocab-adms/> (2013) Accessed 11 April 2024
- [37] Patel-Schneider, P., Motik, B., Parsia, B.: OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax, W3C Recommendation, W3C. Available at: <http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/> (2009). Accessed 11 April 2024
- [38] Albertoni, R., Isaac, A.: Introducing the Data Quality Vocabulary (DQV), *Semantic Web* 12, 81–97 (2021)
- [39] Abou-Zahra, S.: Evaluation and Report Language (EARL) 1.0 Schema, W3C Note, W3C. Available at: <https://www.w3.org/TR/2017/NOTE-EARL10-Schema-20170202/> (2017). Accessed 11 April 2024
- [40] Jeffery, K. G., Houssos, N., Jörg, B.: A. Asserson, Research information management: the CERIF approach, *IJMSO* 9, 5–14 (2014) . <http://dx.doi.org/10.1504/IJMSO.2014.059142>
- [41] Theodoridou, M., Ivanovic, D., Martin, P., Remy, L., Muckensturm, M.: X3ML mappings from common metadata schemes to CERIF RDF, (2019)
- [42] Kellogg, G., Champin, P. A., Longley, D.: JSON-LD 1.1: A JSON-based Serialization for Linked Data, W3C Recommendation, W3C. Available at: <https://www.w3.org/TR/2020/REC-json-ld11-20200716/> (2020). Accessed 11 April 2024
- [43] Perego, A., Austin, T.: DataCite to DCAT-AP Mapping, Technical Specification, European Commission, Joint Re-search Centre (JRC). Available at: <https://ec-jrc.github.io/datacite-to-dcat-ap/> (2021). Accessed 11 April 2024
- [44] Perego, A., Nuffelen, B. V.: GeoDCAT-AP-Version 2.0.0: A geospatial extension for the DCAT application profile for data portals in Europe, SEMIC Recommendation, European Commission. Available at: <https://semic.eu.github.io/GeoDCAT-AP/releases/2.0.0/> (2020). Accessed 11 April 2024
- [45] Guha, R. V., Brickley, D., Macbeth, S.: Schema.org: evolution of structured data on the web, *Commun. ACM* 59, 44–51 (2016)

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

- [46] Brickley, D., Burgess, M., Noy, N. F.: Google dataset search: Building a search engine for datasets in an open web ecosystem, in: L. Liu, R. W. White, A. Mantrach, F. Silvestri, J. J. McAuley, R. Baeza-Yates, L. Zia (Eds.), *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, ACM, pp. 1365–1375, (2019)
- [47] Cyganiak, R., Alexander, K., Hausenblas, M., Zhao, J.: Describing Linked Datasets with the VoID Vocabulary, W3C Interest Group Note, W3C. Available at: <https://www.w3.org/TR/2011/NOTE-void-20110303/> (2011). Accessed 11 April 2024
- [48] Freudenberg, M. et al.: The Metadata Ecosystem of DataID. In: Garoufallou, E., Subirats Coll, I., Stellato, A., Greenberg, J. (eds) *Metadata and Semantics Research. MTSR 2016. Communications in Computer and Information Science 672*, Springer, 317–332 (2016)
- [49] R. Albertoni, D. Browning, S. Cox, P. Winstanley, A. Gonzalez Beltrn, A. Perego, DCAT Version 2 Implementation Report, W3C Document, W3C. Available at: <https://w3c.github.io/dxwg/dcat-implementation-report/> (2020). Accessed 11 April 2024
- [50] Dragan, A., Sofou, N.: StatDCAT-AP DCAT Application Profile for description of statistical datasets Version 1.0.1., Technical Specification, European Commission. Available at: <https://joinup.ec.europa.eu/release/stat-dcat-ap/101> (2019). Accessed 11 April 2024
- [51] Perego, A.: DCAT-AP-JRC: An extension to DCAT-AP for multidisciplinary research data., Technical Specification, European Commission, Joint Research Centre (JRC). Available at: <https://ec-jrc.github.io/dcat-ap-jrc/> (2019). Accessed 11 April 2024
- [52] Cochez, M., Karim, N., Dimitriadis, I.: Analysis of the DCAT-AP extensions, Technical Report, European Commission. Available at: <https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/document/national-extensions-analysis-dcat-ap> (2017). Accessed 11 April 2024
- [53] Galeazzi, F., Franc, Y.L., Wezel, J.V., Hönegger, L., Beckmann, V., Carrillo, R., Nieuwerburgh, I.V., Tanlongo, F., Geistberger, J.S., Cazenave, N., Huynh, F., & Hashibon, A.: EOSC-Pillar first annual report (2020)
- [54] Devaraju, A., Huber, R., Mokrane, M., & Herterich, P.: FAIRsFAIR data object assessment metrics (2020)
- [55] Luna, R. A., Zubcoff, J., Garrigós, I., Gonz, H.: Fairification of citizen science data, in: T. Di Noia, I.-Y. Ko, M. Schedl, C. Ardito (Eds.), *Web Engineering*, Springer International Publishing, Cham, pp. 450–454 (2022)
- [56] Löbe, M., Ulrich, H., Beger, C., Bender, T., Bauer, C. R., Sax, U., Ingenerf, J., Winter, A.: Improving findability of digital assets in research data repositories using the W3C DCAT vocabulary, in: *MedInfo*, volume 290 of *Studies in Health Technology and Informatics*, IOS Press, pp. 61–65 (2021)
- [57] da Silva Santos, L. O. B., Burger, K., Kaliyaperumal, R., Wilkinson, M. D.: FAIR Data Point: A FAIR-oriented approach for metadata publication, *Data Intelligence* 5, 163–183 (2023)

APPENDIX A. EXAMPLES

- 1 @ prefix xsd : <http :// www .w3.org /2001 / XMLSchema # > .
- 2 @ prefix rdfs: <http :// www .w3.org / 2000 / 01 / rdf - schema # > .
- 3 @ prefix dcterms: <http :// purl. org/ dc/ terms/ > .
- 4 @ prefix foaf: <http :// xmlns. com / foaf /0.1/ > .
- 5 @ prefix skos: <http :// www .w3.org / 2004 / 02 / skos/ core # > .
- 6 @ prefix dcat: <http :// www .w3.org/ns/ dcat # > .
- 7 @ prefix prov : <http :// www .w3.org/ns/ prov # > .
- 8 @ prefix ex: <http :// example . org/ > .

Listing 1: Prefixes used

- 1 # *catalog*
- 2 ex: catalog a dcat: Catalog ;
- 3 dcterms: title " Imaginary Catalog "@en ;
- 4 rdfs: label " Imaginary Catalog "@en ;
- 5 foaf: homepage <http :// example . org/ catalog > ;
- 6 dcterms: publisher ex: transparency - office ;
- 7 dcterms: language <http :// id. loc. gov / vocabulary / iso639 -1/ en > ;
- 8 dcat: theme Taxonomy <http :// aims. fao . org/ aos/ agrovoc >; # *AGROVOC Thesaurus*
- 9 dcterms: issued " 2020 -11 -30 " ^^ xsd : date ;
- 10 dcterms: modified " 2023 -11 -11 " ^^ xsd : date ;
- 11 dcat: dataset ex: dataset -001, ex: dataset -002 ;
- 12 dcat: record ex: record -002 ;
- 13 dcat: service ex: figure - service -001 ;
- 14 dcat: catalog ex: dataset -002 ;
- 15 dcterms: hasPart ex: dataset -001, ex: dataset -002, ex: record -001, ex: figure - service -001.

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

- 16 # *superfluous statement as all resources are typed as service, datasets, etc*
- 17 # *The transparency office publishing the catalog*
- 18 ex: transparency - office a foaf: Organization ;
- 19 rdfs: label " Transparency Office "@en .
- 20
- 21 # *Record of dataset -002 in the catalog*
- 22 ex: record -002 a dcat: Catalog Record ;
- 23 foaf: primary Topic ex: dataset -002 ;
- 24 dcterms: issued " 2023 -11 -11 " ^^ xsd : date ; # *added in the catalog in date*
- 25 dcterms: conformsTo <https :// www .w3.org/TR/vocab - dcat -2/ > .
- 26 # *DCAT 2 Rec URI*
- 27 # *Sub - catalog*
- 28 ex: dataset -002 a dcat: Catalog, dcat: Dataset ;
- 29 dcat: landing Page <http :// example . org/ Catalog2 > ;
- 30 dcterms: title " A second fictional example "@en ;
- 31 dcterms: accessRights <http :// publications . europa . eu/ resource / authority / access - right/ PUBLIC > ;
- 32 dcat: keyword " thesauri"@en , " environment "@en, " framework "@en ;
- 33 dcat: theme <http :// dbpedia . org/ resource / Category : Thesauri >, <http :// dbpedia . org/ page / Category : Environmental_science >;
- 34 dcat: theme Taxonomy ex: DBPEDIACategories ;
- 35 dcat: distribution ex: dump ;
- 36 dcterms: issued " 2013 -11 -30 " ^^ xsd : date ;
- 37 dcterms: modified " 2015 -09 -06 " ^^ xsd : date ;
- 38 dcat: contactPoint ex: team ;

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

```
39 dcterms: creator ex: team ;
40 dcterms: publisher ex: team ;
41 dcat: dataset ex: d1, ex: d2 ;
42 dcat: service ex: sparqlEnd Point .
43
44 # DBPEDIA categories
45 ex: DBPEDIACategories a skos: ConceptScheme ;
46 skos: prefLabel " The set of categories provided by DBPEDIA "@en ;
47 foaf: homepage <https :// dbpedia . org/ resource / Category : Main_topic_classifications > .
```

Listing 2: A fictional DCAT 2 Catalog

```
1 # Fictional dataset included in the catalog
2 ex: dataset -001 a dcat: Dataset ;
3 dcterms: title " Imaginary dataset"@en ;
4 dcat: keyword " air quality "@en, " health "@en ;
5 dcterms: creator ex: finance - employee -001 ;
6 dcat: contactPoint <http :// dcat. example . org/ transparency - office / contact > ;
7 dcat: landing Page <http :// example . org/ dataset -001. html > ;
8 dcterms: publisher ex: finance - ministry ;
9 dcterms: language <http :// id. loc. gov / vocabulary / iso639 -1/ en > ;
10 dcat: theme <http :// aimdatas . fao . org/ aos/ agrovoc/ c_a2ef545f >, # AGROVOC concept for air
quality
11 <http :// aims. fao . org/ aos/ agrovoc/ c_3511 > ; # AGROVOC concept for health
12 dcterms: issued " 2011 -12 -05 " ^^ xsd : date ;
13 dcterms: modified " 2011 -12 -15 " ^^ xsd : date ;
14 dcterms: accrualPeriodicity <http :// purl. org/ linked - data / sdmx /2009 / code # freq -W> ;
```

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

```
15 dcterms: conformsTo <http://data.europa.eu/eli/reg/2014/1312/oj>;
16 dcterms: accessRights
17 <http://publications.europa.eu/resource/authority/access-right/PUBLIC>;
18 dcterms: rights [ a dcterms: RightsStatement ;
19 rdfs: label " Copyright 2021 ACME Inc."@en
20 ];
21 dcterms: temporal [ a dcterms: Period OfTime ;
22 dcat: startDate " 2011-07-01 " ^^ xsd : date ;
23 dcat: end Date " 2011-09-30 " ^^ xsd : date ;
24 ];
25 dcat: temporalResolution " P1D " ^^ xsd : duration ;
26 dcterms: spatial <http://sws.geonames.org/6695072/>; # geonames URI for European Union
27 dcat: spatialResolution In Meters " 30.0 " ^^ xsd : decimal ;
28 dcterms: isReferenced By <https://doi.org/xx.yyyy/fictionalpaper>; # Doi of a fictional paper
29 dcat: distribution ex: dataset-001-csv, ex: dataset-001-targz .
30
31 # a CVS distribution of the dataset-001
32 ex: dataset-001-csv a dcat: Distribution ;
33 dcat: download URL <http://dcat.example.org/files/001.csv>;
34 dcterms: conformsTo <http://dcat.example.org/files/001.csv-metadata.json>; # a CSV
Schema
35 dcterms: title " CSV distribution of imaginary dataset 001 "@en ;
36 dcat: media Type <http://www.iana.org/assignments/media-types/text/csv>;
37 dcat: byte Size " 5120 " ^^ xsd : non Negative Integer ;
38 dcterms: license <https://creativecommons.org/licenses/by/4.0/> .
```

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

39

40 # ex: dataset -001 - csv packaged and compressed

41 ex: dataset -001 - targz a dcat: Distribution ;

42 dcat: download URL <http :// dcat. example . org/ files /001. tar. gz > ;

43 dcat: media Type <http :// www .iana .org/ assignments / media - types/ text/ csv > ;

44 dcat: package Format <http :// publications . europa . eu/ resource / authority / file - type / TAR > ;

45 dcat: compressFormat <http :// www .iana .org/ assignments / media - types/ application / gzip > .

46

47 # Reference standard / specification

48 <http :// data . europa . eu/ eli/ reg / 2014 / 1312 / oj > a dcterms: Standard ;

49 dcterms: title “ Commission Regulation (EU) No 1089 / 2010 of 23 November 2010 implementing Directive 2007 / 2 / EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services”@en ;

50 dcterms: issued “ 2010 -11 -23 ” ^^ xsd : date .

51

52 # Data Service

53 ex: figure - service -001 a dcat: Data Service ;

54 dcterms: conformsTo <http :// www . opengis. net/ def/ service Type / ogc/ wms /1.3 > ; # OGC Web Map Service 1.3

55 dcterms: type <https :// inspire . ec. europa . eu/ metadata - codelist/ SpatialData Service Type / view > ;
View Service

56 dcat: endpointDescription <http :// example . org/ api/ figure -006/ params > ;

57 dcat: endpointURL <http :// example . org/ api/ figure -001 > ;

58 dcat: servesDataset ex: dataset -001 .

59

60 # fictional Finance Ministry

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

```
61 ex: finance - ministry a vcard : Organization, foaf: Group ;
62 vcard : hasEmail <mailto : finance - ministry@ example - finance - ministry . org > ;
63 foaf: mbox <mailto : finance - ministry@ example - finance - ministry . org > ;
64 vcard : hasMember ex: finance - employee -001;
65 foaf: member ex: finance - employee -001;
66 vcard : title " Finance Ministry "@en .
67
68 # fictional Finance Ministry employee
69 ex: finance - employee -001 a foaf: Person, card : Agent ;
70 foaf: family Name " Rossi";
71 foaf: firstName " Mario ";
72 foaf: homepage <https :// example - finance - ministry . org/ mariorossi > .
```

Listing 3: Fictional DCAT 2 datasets and service

```
1 # Dataset as bag of files
2 ex: d 33937 a dcat: Dataset;
3 dcterms: description " A set of RDF graphs representing the International [ Chrono ] stratigraphic Chart,
..."@en ;
4 dcterms: identifier " https :// doi. org / 10 .25919 / 5 b 4 d 2 b 83 cbf2 d " ^^ xsd : any URI ;
5 dcterms: creator <https :// orcid . org /0000 -0002 -3884 -3420 > ;
6 dcterms: relation ex: ChronostratChart2017 -02. pdf ;
7 dcterms: relation ex: ChronostratChart2017 -02. jpg ;
8 dcterms: relation ex: timescale . zip ;
9 dcterms: relation ex: d33937 - jsonld ;
10 dcterms: relation ex: d33937 - nt ;
11 dcterms: relation ex: d33937 - rdf ;
```

The W3C Data Catalog Vocabulary, Version 2: Rationale, Design Principles, and Uptake

12 dcterms: relation ex: d33937 - ttl .

Listing 4: A legacy datasets that are just a bag of files

1 # *Qualified attribution*

2 ex: DS987 a dcat: Dataset ;

3 prov : qualified Attribution [

4 a prov : Attribution ;

5 prov : agent <https://www.education.gov.au> ; # *link to the Department of Education - Australian Government*

6 dcat: had Role <http://registry.it.csiro.au/def/isotc211/CI_Role Code/funder> # *role of funder*

7] .

8

9 # *Qualified relation*

10 ex: Test987 a dcat: Dataset ;

11 dcat: qualified Relation [

12 a dcat: Relationship ;

13 dcterms: relation ex: DS987 ;

14 dcat: had Role <http://www.iana.org/assignments/relation/original>

15] .

Listing 5: Qualified attributions and relations