International
Virtual
Observatory
Alliance

This is only an outline! Everything is still to be done.

Best practices for the creation of and metadata for digital object identifiers in astronomy archives

Version 1.0

IVOA Note 2024-12-29

Author(s)
>August Muench, Gilles Landais, Raffaele D'Abrusco, Anne Raugh, Edwin Henneken, Baptiste Cecconi, to complete: Alberto Accomazzi ?, ESA people ?

Editor(s)
>August Muench, Gilles Landais

Abstract

Many astronomy archives are minting digital object identifiers (DOI) for datasets and services. This document aims at summarizing current workflows for creating and using DOIs, diagnose issues in the metadata of existing DOIs, and develop best practices for workflows and metadata for future DOI deployment. This note is focused on archives in Astronomy, Planetary Science, and Heliophysics. Additional domains may be considered at a later time.

Status of this document

This is an IVOA Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

A list of current IVOA Recommendations and other technical documents can be found at https://www.ivoa.net/documents/.

Contents

Acknowledgments

Conformance-related definitions

The words "MUST", "SHALL", "SHOULD", "MAY", "RECOMMENDED", and "OPTIONAL" (in upper or lower case) used in this document are to be interpreted as described in IETF standard RFC2119 (**?**).

The *Virtual Observatory (VO)* is a general term for a collection of federated resources that can be used to conduct astronomical research, education, and outreach. The International Virtual Observatory Alliance (IVOA) is a global collaboration of separately funded projects to develop standards and infrastructure that enable VO applications.

1 Introduction

1.1 General observations

We focus on roles of digital object identifiers (DOIs) rather than the full landscape of FAIR data (Wilkinson and Dumontier et al., 2016). It is beyond the scope of this document to address all aspects of FAIRness. We consider that it is also unreasonable for DOIs to be laden with the problems of enabling all aspects of FAIRness. Instead we aim for this document to guide archives with achieving compliance for specific aspects of enabling FAIR data. In this document we are using DataCite Commons [1]. This document adopts DataCite Metadata Schema 4.6. We choose relevant examples from existing DOI taken in approximately 75000 dataset DOIs in astronomy (See Table 1).

1.2 DOI landscape

DOIs are a key element of the scholarly publication process. It is a persistent identifier for resources, which includes both a URL and metadata. The DOI is mostly used for sustainable resources such as scholarly journal articles, for datasets and softwares.

The DOI is a persistent identifier that emerges today among other identifiers such as the ORCID[2], for authors, the RoR[3] for organization or the

---

[1]https://commons.datacite.org
[2]https://orcid.org/
[3]https://ror.org/

RAID[4] for projects.

It enables a resource to be reliably identified, regardless of the professional network in which it is published. It improves traceability and citation with standard metadata that can be harvested and reused.

The DOI metadata can also include references and related products. This network of resources builds a so-called scholarly (or scientific) knowledge graph[5].

There are several DOI providers. CrossRef[6] is the main DOI provider for scholarly communications. Datacite[7] is the main DOI provider for data publication. Each provider has a specific metadata schema, extending Dublin-Core, and design to support the scope of the resources related to the DOI they provide. Data Centers minting DOIs for their datasets should carefully inform and curate the metadata registered in the DOI provider database.

DOIs are also used for grants (see, e.g., https://doi.org/10.3030/824064, which is redirecting to the ESCAPE project page on the CORDIS web site of European Commission), for instruments (see, e.g., the recent Datacite implementation of the RDA PID for instrument recommendation), for repositories (see, e.g., http://doi.org/10.17616/R31NJMS9, which is redirecting to the PADC record in the Re3data registry), or for FAIR concepts and resources (see, e.g., https://doi.org/10.25504/FAIRsharing.L4Q3H9, which is an identifier for the concept of "IVOA identifier" in FAIRsharing's registry). It is thus not always used for scholarly citations, but also as persistent identifiers to resources.

In this document, we propose good practices to provide efficient citation, interconnections with related resources, including IVOA resources.

1.3 Use case

1. **Citation**. Data producers (authors) and providers (data centers) expect that the datasets or collections they are publishing are cited in scholarly communication (journal articles, proceedings, posters, etc.) so that their contribution to the field is acknowledged and attributed. Using the regular citation method allows to treat data and article references equally and use the same tools to count citation and measure impact. Hance data producers and providers are expecting that data reuse is acknowledged through regular citation. Users also expect the Data Center website to provide a way to cite the resource.

---

[4]https://raid.org/
[5]https://doi.org/10.1162/qss_e_00160
[6]https://www.crossref.org/
[7]https://datacite.org/

2. **Reuse**. The Data consumer expects to get information about origin and rights. For access rights, existing licence must be displayed but also readable by software.

3. **Findability**. Data published are disseminated in networks like the VO registry, google dataset search[8] or the EOSC EU node (see 4.1.1). The users as well as the Data Center expect that the entries are consistent.

4. **Impact**. Data centers as well as authors need reports about the usage activity of their work. They query dedicated services (eg: OpenCitation) that make reports and are capable to graph the citation of a specific resources..

5. **Reproducibility**. A user expects sustainable access to a resource that he cited in an article or that he used in an experiment. He or she expects that the experiment is reproducible. If it uses a service or a knowledgebase (see 3.4), he or she expects that the protocols stay compatible with the original implementation. For datasets, he or she should expect that data present in the original study have not been altered.

1.4   Example of existing DOI

Table 1 lists a short selection of DOI covering different usage.

Archive	Prefix	DataCite Client	Yo	Count
Canadian Astronomy Data Centre	10.11570	nrc.cadc	-	104
Chandra Data Archive	10.25574	si.cda	-	57,724
ESO Science Archive Facility	10.18727	tib.esolib	-	447
European Space Agency	10.5270		-	28,858
IPAC (26131 — 26135)	10.2613X	caltech.ipacdoi	-	680
Mikulski Archive for Space Telescopes	10.17909	stsci.mast	-	3,034
NASA Planetary Data System	10.17189	nasapds.nasapds	-	2,287
++Small Bodies Node of PDS	10.26007	sbn.archive	-	4,837
Strasbourg Astronomical Data Center	10.26093	inist.cds	-	22,181

Table 1:   Example of organizations generating DOIs (statistics generated 2025/03/28).

TODO: Please, add ObsParis !

2   Role within the VO Architecture

---

[8]https://datasetsearch.research.google.com/

2.1 Status of DOI in the Virtual Observatory

The IVOA registry uses IVOA identifiers (IVOID) (Demleitner and Plante
et al., 2016) which are only usable within the IVOA registry ecosystem. At
the time of writing, DOIs are not fully integrated in VO standards. They
can be used in the IVOA Registry since VOResource 1.2 (Dower and Dem-
leitner et al., 2014) for any ResourceName such as alternate identifier or
related resources. They are also used in the DataOrigin Note (Landais and
Muench et al., 2024) as metadata that a service can provide in the result of
a query and in the BibVO note (Demleitner and Accomazzi, 2024) which is
an architecture based on registry to link Data and articles.

2.2 Comparison with the VO Registry schemas

The IVOID   is a unique identifier for resource (dataset or service) indexed
in the Virtual Observatory registry. Registry records has metadata which
includes entry points to access the resource via WEB services. The access
information needs a machine interpretation and its usage is generally hidden
for users. The IVOID and the IVOA Registry are meant to ensure findability
and technical interoperability (how do I find a service and what protocol can
I use?). They are mostly used for services. When they represent datasets,
the services to access the data are described in details. For instance the
vocabulary of the registry extends the relations with a term "isServedBy"
to link Data and Service. The absence of an equivalent term in DataCite
makes it illusory to try and transcribe the same level of interoperability in
DOI.

The DOI   and their metadata are persistent identifiers for resources. The
DOI is visible for users and it handles a landing page human-readable.
The metadata include attribution, references, relations, licence, but noth-
ing about programmatic access or interoperability.

   We propose in section 4.4.4 a method to facilitate link between DOI and
registry record.

   VO Registry and DataCite schemas are based on DublinCore terms ex-
tended with their own specifications. Despite a common basis, the vocabu-
laries differ. The DataOrigin note provides in Appendix a mapping between
DataCite and VOResource for a sub selection of terms. Note that the en-
dorsement process of DataOrigin has not been initiated.

3 Extant use of Digital Object Identifiers in Archives

We observe four different uses for DOIs in astronomy-related archives. Identifiers are being assigned for individual datasets (at various levels of granularity); for collections of datasets; for services; and for what we will call "knowledgebases" or curated meta-collections of data and calculated results. These uses are described in more detail below.

3.1 Datasets

Dataset DOIs are the most common usage in astronomy-related repositories, and represent about »XX% of all such DOIs. Individual datasets are being assigned DOIs at various levels of granularity. At the most granular level, every observation [9] in the Chandra Data Archive (CDA) is assigned an individual DOI. Similarly, the European Space Agency assigned DOIs to every individual Herschel, ISO, Planck and XMM-Newton observation. This list also includes every Hubble Space Telescope (HST) observation, which are also archived at the Mikulski Archive for Space Telescopes (MAST) but are not assigned DOIs individually by MAST. Individual dataset DOIs are implemented mostly consistent with the usage of DOIs by institutional repositories, generalist repositories (e.g., Zenodo), etc. That is, the dataset DOI resolves to a landing page (see 4.3) describing the dataset. In some cases the landing pages are rich summaries of the dataset, often including much more information than could be gleaned from the DOI metadata alone. In other cases the landing pages pop the user into a web user interface with very little (or almost no) contextual information about the DOI.

Dataset DOIs are also being minted for "High-Level Science Products" (HLSP). Such HLSP datasets are not static and can grow over time, accumulating both data revisions and additions. HLSP landing pages for DOIs have no set structure or content and change regularly as the linked data evolve. None of these dataset DOIs are versioned. The data discovered by following a dataset DOI to an individual observation will change based upon the version of the archive pipeline used to create it, up real-time or on-demand reprocessing for some archives (see also 4.5.2).

The typical Resource type in DataCite is DataCite:dataset

TODO: (see also) explain difference between knowledgebase and evolving dataset.

Question: what is the difference between knowledgebases and HLSP ?

---

[9]https://cxc.cfa.harvard.edu/csc/organization.html

3.2 Collections

Collection-type DOIs direct users to a collection of individual data records in a particular archive. Examples of collection-type DOIs include the DOI services provided by the Mikulski Archive for Space Telescopes (MAST)[10] (Novacescu and Peek et al., 2018), Chandra Data Archive (CDA) (Novacescu and Peek et al., 2018), and the VAMDC Consortium[11] Query Store service (Moreau and Zwolf et al., 2018). There is significant variation in the expected use cases for these Collection DOIs. There are also strong variations in the metadata of Collection-type DOIs.

Considering use-cases, both MAST and Chandra Collection DOIs collect dataset identifiers within their respective databases. In both cases there is no expectation that these Collection DOIs would ever themselves collect attribution (i.e., be cited in the reference list of a corresponding Journal article). However, VAMDC Collection DOIs, are intended to collect and distribute credit to the collection of database identifiers they contain. They attempt to distribute this citation/credit to the collection of resources by "citing" all the related resources in the saved VAMDC query record. Because this depends upon the capabilities of both the chosen DOI minting service (Zenodo) and the DataCite Schema, additional notes on the details and the outcomes of this effort is provided in one of our case studies in Appendix B.

Similarly the metadata of Collection DOIs vary between MAST and CDA examples. As described in D'Abrusco and DPAOps Team (2023), Chandra Data Collection DOIs create complete records of individual using `relatedIdentifier` tags and predicates. The metadata of MAST Collection DOIs do not provide detailed information on the individual MAST records collated in the collection.

See 4.2, 4.3, 4.4, 5.1, 5.2, 5.4

The typical Resource type in DataCite is   DataCite:**Collection**

Alternative text (remove VAMDC and the citation part which will be in section 4.2.3)

Collection-type DOIs direct users to a collection of individual data records that are related to each other (e.g. : individual observations from the same survey or archive, etc.). The collection provides access to resources (identified with a DOI or with another identifier) via relationships defined in the DataCite schema.

Examples of collection-type DOIs include the DOI services provided by the Mikulski Archive for Space Telescopes (MAST) [12]

---

[10]https://archive.stsci.edu
[11]https://vamdc.org
[12]https://archive.stsci.edu

(Novacescu and Peek et al., 2018), Chandra Data Archive (CDA) (Rots and D'Abrusco et al., 2018). There is significant variation in the expected use cases for these Collection DOIs. There are also strong variations in the metadata of Collection-type DOIs.

Considering use-cases, both MAST and Chandra Collection DOIs collect dataset identifiers within their respective databases. For instance, Chandra Collection DOI is made of DOIs of different types. Its metadata and related identifiers are defined to make a Chandra collection DOI the most versatile and general aggregation that can be obtained using Chandra data "entities".

Similarly the metadata of Collection DOIs vary between MAST and CDA examples. As described in D'Abrusco and DPAOps Team (2023), Chandra Data Collection DOIs create complete records of individuals using `relatedIdentifier` tags and predicates. The metadata of MAST Collection DOIs do not provide detailed information on the individual MAST records collated in the collection.

(TODO: add MASER example (Baptiste example) The collection https://doi.org/10.25935/1SXH-AQ56 is a grouping of https://doi.org/10.25935/2F8-C505, https://doi.org/10.25935/XS9J-ND90, https://doi.org/10.25935/F8NS-0911, https://doi.org/10.25935/5JFX-DH49, https://doi.org/10.25935/9ZAB-FP47, https://doi.org/10.25935/ZKXB-6C84, https://doi.org/10.25935/6D16-BM97 and https://doi.org/10.25935/B9NF-9558.)

3.3 Services

Examples of DOIs for services include IRSA (DUST). Service DOIs lead to query tools. They may lead to query results. Sometimes Collection DOIs act like Service DOIs but they are not. Collection DOIs may result from queries performed at a Service.

Examples:

- the VizieR Service, doi:
  https://api.datacite.org/dois/application/vnd.datacite.datacite+xml/10.26093/cds/vizier

- TOSS (Tübingen Oscillator Strengths Service, GAVO):
  https://api.datacite.org/dois/application/vnd.datacite.datacite+xml/10.21938/3I01ISNUCODNH1ZJBCVUWA

See 4.3, 4.4, 5.1, 5.2, 5.4
The typical Resource type in DataCite is DataCite:**Service**

3.4 Knowledgebases

A knowledgebase is a collection of material collated from many discrete sources. All of the values contained in a knowledgebase have a provenance traced to other resources and have been curated into a single database for reuse. Examples include: Simbad (as originally described in, Wenger and Ochsenbein et al., 2000), NASA Exoplanet Archive (NEA) (NASA Exoplanet Science Institute, 2020) (as originally described in, Akeson and Chen et al., 2013), NASA Extragalactic Database (NED) (NASA/IPAC Extragalactic Database (NED), 2019) (as originally described in, Helou and Madore et al., 1991).

Current observations about DOIs for knowledgebases include:

- DOIs for knowledgebases lead users to interstitial landing pages rather than directly to the collated, curated resources.

- DOIs for knowledgebases do not lead to individual values, e.g., the results of a query against that knowledgebase.

- DOIs for knowledgebases never provide information in their *metadata* about the *state* of a knowledgebase: its current version; last update; etc. Nor is this information on the interstitial landing pages of knowledgebases.

- Services that return DOIs for queries against knowledgebase are considered Service DOIs (See Section 3.3).

See 4.3, 4.5.2, 5.1, 5.2, 5.4

The typical Resource type in DataCite is DataCite:**Dataset** or DataCite:**Service**

3.5 Issues

The primary issue evident in DOI creation today is a mismatch between the metadata created by an archive and the use case (intended to be) implemented by that DOI. Succinctly, the metadata supplied by astronomy archives is often insufficient to ensure the accurate citation of the datasets.

By detailing these pathologies and triaging their less-than-desirable outcomes we can aim to develop empirically-defined best practices to guide repositories forward with the use of identifiers. Here is a topical list of pathologies:

1. Incomplete metadata: missing or incorrectly formatted author data, generic or misstated titles, misunderstood dates.

2. Inconsistent metadata: transmutations of metadata between systems lead to inconsistent metadata deposits. Example: transformation from schema.org to Crossref left ESA DOIs in a nasty state;

3. (Un)versioned data: versioning is mostly nonexistent and when provided it is ill defined and often opaquely transmitted;

4. Misconceptions: DOIs do not lead to data, or what they do lead to is widely varied.

4 Best Practices for DOI Workflows

4.1 Pre-requisite

Ten Registration agencies are providing DOIs, Crossref[13] usually used for articles and DataCite[14] for data are the best known. We strongly recommend the use of DataCite which offers the most appropriate metadata for recording data.

To create a DOI (DataCite), you must be a member of DataCite. Data publishers who are not members must contact a member organisation (in general, countries and universities have a contact point).

4.1.1 Providing metadata using the appropriate schema

DOI workflow requires metadata curation based on the DataCite schema[15] which allows to improve the FAIRness of the datasets. Other workflows exist with their own semantic and metadata serialization. These are multiple, they overlap and are also specific.

For instance, the IVOA framework provides a registry with a high level of interoperability and point directly to the resources, whereas DataCite is specialized in data citation and link a human web page called "landing page".

> **Note** DOI implies to maintain a sustainable mechanism to provide a URL. This requirement, even if also in usage in the IVOA registry, is mandatory for DOI.

Datasets distributed in diverse frameworks complement each other and are likely harvested by platforms or search engine such as ADS, EOSC, Google Dataset. These indexing services group together the information relating to the same resource, they merge the metadata which are both redundant and specific to each harvested network. This cross operation process (for instance OpenAire) is often a black box and depends on the indexing service's strategy. Note that DOI, as a unique identifier, facilitates the cross operations.

---

[13]https://www.crossref.org/
[14]https://datacite.org/
[15]https://datacite-metadata-schema.readthedocs.io/

The list of metadata tends to increase, but the most popular are DataCite schema, Dublin Core[16], VOResource [17] (the registry of the Virtual Observatory), schema.org [18] (extends Dublin Core and is used by Google), DCAT[19] (linked catalogues, datasets and services. DCAT is a concurrent of the VO registry), OpenCitation[20] (a schema of linked citation), etc.

> **Note** All are specific, and we highlight the importance for Data providers to disseminate consistent metadata (for instance list the whole authors in all output).

The importance, but also the specifity of the metadata have been discussed in IVOA DCP session (see the presention of H.Enke, Interop 2023, Bologna [21]). In practice, it is better for implementers to think since the beginning about the different output in order to provide consistent workflow.

Note DataCite provides several serialization of the metadata, in particular "schema.org".

> **Note** Maintain the disseminations workflows togethers

4.1.2 DOI syntax

The DOI architecture includes a URL resolver to link the DOI identifier and the resource for the long term. Only the DOI identifier is fixed, the URL as well as the metadata are maintained by the DOI creator.

The DOI syntax is composed with a prefix followed by a suffix. The prefix is assigned by DataCite or Crossref and is used for all resources provided by the DOI producer. The suffix is created by the producer to identify the resource.

For instance:

https://doi.org/10.3847/1538-4365/aab76a

- the prefix 10.3847 is the prefix attributed by Crossref to the AAS journals

---

[16]https://www.dublincore.org/
[17]https://ivoa.net/documents/VOResource/
[18]https://schema.org/
[19]https://www.w3.org/TR/vocab-dcat-3/
[20]https://opencitations.net/
[21]https://wiki.ivoa.net/internal/IVOA/InterOpMay2023RegistryDCP/DOI-AIP-20230510.pdf

- the suffix 1538-4365/aab76a defined the article

DataCite recommends using an opaque syntax. To remove any significance in a name avoids bad interpretation and is more sustainable. For example, the data center where the data are deposited may change and therefore should not be used in the syntax.

Note also the DOI usage in web pages for which it is recommended to provide clickable links. It is therefore up to the web maintainer whether or not to hide the DOI with appropriate text

4.2 Citation requirements

> **Note** Semantic forewords. Citation means citation in the reference list of an article. A citation is not a footnote URL, an acknowledgement or any DOI string in the text. It ONLY refers to listing the reference in a reference list of an article.

4.2.1 Generating BibTeX

DataCite schema includes the required metadata to generate citation in various formats: BibTeX, APA, etc. The citation quality is generally fixed by journals, ADS or Data publishers. It is possible to make a citation from DOI metadata when these are properly curated (BibTeX generated by ADS is richer than those provided by DataCite). We will describe the mandatory items that cover the known citations and we will complete the metadata with added relevant information.

Note: Example of citation provided by DataCite in BibTeX format:

```
curl -LH "Accept: application/x-bibtex" https://doi.org/10.5270/esa-1ugzkg7
```

Full BibTeX template : the following example maps BibTeX with DataCite schema (see table 2).

We use the ADS BibTex type "@dataset". This term is specific to ADS and is not used by service such as DataCite which uses the generic term "@misc".

```
@dataset{{localref},
        author = {authors},
        title = "{title}",
        year = {year},
        month = {month},
        eid = {usualName},
```

14

```
            url = {url},
            keywords = {keywords},
            publisher = {publisherName},
            copyright = {rights},
            DOI = {DOI}
    }
```

BibTeX	DataCite	Relevance for citation
authors	Authors	MUST, see 5.2
title	Title	MUST, see 5.5
year	Date:creation	MUST, see 5.4
month	Date: creation	
eid	alternateIdentifier	MAY, see 4.4
url		SHOULD, use https://doi.org/{DOI}
keywords	Sujects	SHOULD, Please, see 5.3
publisher	Publisher	MUST
copyright	Rights	RECOMMENDED, see 5.6
DOI		MUST
resource_type	resourceType	MUST, see 5.1

Table 2: linking BibTeX and DataCite schema

Citation in APA style:

```
{authors} ({year}). {title} [resource_type]. {publisher}. {DOI}
```

VizieR example: (catalogue J/MNRAS/320/451)

```
Beers, T. C., Rossi, S., O'Donoghue, D., Kilkenny, D., Stobie, R. S.,
Koen, C., & Wilhelm, R. (2006). A-G star metallicity [Data set].
Centre de Donnees Strasbourg (CDS). https://doi.org/10.26093/CDS/VIZIER.73200451
```

4.2.2   Evolving datasets

See DataCite evolving dataset guidance[22]

PLEASE RE-VIEW!

TODO: recall the difference between knowledgebase and Evolving

We consider different citation for evolving datasets. The method depends on the Data Center implementation.

---

[22]https://datacite-metadata-schema.readthedocs.io/en/4.6/guidance/dynamic-datasets/

1. Cite a snapshot of the Dataset. In this approach, the Data Center makes snapshots and adopts a DOI versioning mechanism (see 4.5).

2. Cite the dataset as to be an evolving Dataset. Example in APA style:

   ```
   {authors} ({year}). {title} [evolving Data set]. {publisher}.
   {DOI}. Accessed {date_of_access}
   ```

   In the example date_of_access could be the 'update' date of the DOI record (assuming that the dataset and its DOI are well synchronized).

3. Cite a DataSet extraction resulting of a query.
   The solution may become complex if we take or not the reproducibility aspect. A simple solution without reproducibility constraint consists to cite the access protocol (for example IVOA SCS, TAP), optionally completed with query details.

   Example in APA style: (note that "protocol" is independent of the DOI record!

   ```
   {authors} ({year}). {title} [evolving Data set]. {publisher}.
   {DOI}. Accesses {date_of_access}, via {protocol}
   ```

   An option taking into account the reproducibility has been adopted by the VAMDC Query Store[23] where query and result are hosted with a DOI (giving details are not in the scope of this note).

Note   see also the acknowledgment proposed in DataOrigin[24] Appendix Citation, Template.

4.2.3 Collections

Collections improve discoverability thanks to relationships linking individual resources and allow global metadata factoring. However, the usage of Collection for citation is ambiguous.

Practice shows that the quality required to extract a good citation is often not obtained from the DOI of Collections. This lack of information is often due to the concept which is of a different nature.

There are different views that depend on the Data Center strategy but which could be ambiguous for article editions. We report two usages, without wanting to endorse either.

May be add a reference to the RDA term "Dynamic"?

PLEASE REVIEW!

---

[23]https://cite.vamdc.eu/
[24]https://www.ivoa.net/documents/DataOrigin/

- Collection is not intended to be cited in articles. The role of Collection (potentially having poor metadata) seems to be essentially to establish relation (with the advantage of the flexibility of the URL resolution that the DOI offers). This approach is chosen by MAST

- The preferred citation is the Collection. Collections are intended to collect and distribute credit to and collection of entities they contain. Sometimes, individual entities have poor metadata in favor of global metadata (in Collections). In this architecture, because heritage is not trivial for harvesting, it results that individual resources have no metadata allowing citation. This approach is chosen by Chandra

4.2.4 Service

We encourage Data publishers to provide a way to cite or acknowledge services used by authors in articles. The recommendation is in the Publisher discretion and should be visible in the landing page of the Service (see 4.3).

It is rather a common practice to acknowledge than cite a service in an article. For both, the Service DOI which includes the URL resolver has to be considered.

4.3 Landing pages

4.3.1 Building a landing-page

Landing page is a human readable WEB page attached to the DOI. Datacite provides a documentation about good usage [25].

> **Note** The landing page is primarily dedicated for Human. The DOI metadata should be visible in the web page. In particular, the DOI, title, authors, licenses should be highlighted.

Additional information is often added. In particular links to access the data, but also any other information, specific to the data center or included in other workflows provided by the Data Center (BibTeX, schema.org [26], DCAT [27], etc)

Landing page for service of Knowledgebase The landing page of a service or a Knowledgebase is generally its web portal. It includes both service access and information (eg: contact, about, etc.)

---

[25]https://support.datacite.org/docs/landing-pages
[26]https://schema.org/
[27]https://www.w3.org/TR/vocab-dcat-3/

Landing page for collection This page contains information that is common to every datasets in the collection. Each Dataset specificity is not shown on this page. The content depends on the nature of the collection: for instance the Organization, an abstract, the date of creation or the materials used to generate the individual datasets.

The page should provide the way to access individual datasets.

4.3.2 Generate a machine-readable landing page

Search engines harvest landing page for indexation and expect some metadata serialized with a standard semantic. For instance "schema.org" or DCAT serialized in JSON-LD are needed by Google Datasets [28].

The FAIRness of the landing page can be evaluated with validators:

- FAIRchecker [29]

- Google Search central [30]

4.3.3 Pathologies

- https://doi.org/10.17189/1519607 (NASA/PDS): there is a title, authors, DOI, summary... but there is no direct link to the data. The "Search/Access Data" link goes to a search interface, where the user should search again with some parameter (not specified, to be guessed)

- https://doi.org/10.48322/rgf7-3h67 (HPDE): there is everything about the citation, title, doi, summary, content... but the formatting is very close to the SPASE XML record. It's ok for user fluent with SPASE, not so much for outsiders. The SPASE record lists a number of related identifiers, not present in the DataCite metadata.

- https://doi.org/10.57780/esa-3xcjd4w (ESA): All required/recommended information is present, but the granularity (a single DOI for the full experience, not split with version ro processing levels...). And the data access goes to a search interface, where the user should search again for the products.

- https://doi.org/10.3334/ornldaac/2351. Besides having semantic information in the DOI, the biggest problem is the fact that metadata says that this is a collection DOI. However, the metadata is for a dataset that is in the collection. The collection itself is specified in the

---

[28]https://datasetsearch.research.google.com/
[29]https://fair-checker.france-bioinformatique.fr/
[30]https://developers.google.com/search/docs/appearance/structured-data

relatedIdentifiers section with attribute IsPartOf, but the collection itself does not have a DOI.

4.4 Linked Data

4.4.1 Connect resources together

Linked Data allows to connect resources together, it helps discovery and allows to generate a map of interconnected resources. Linked Data are expressed with DataCite:**RelatedIdentifier** in DataCite schema. They have been the subject of a supplementary section in the DataCite guidance.

We recommend to add at least one relation (for instance to link a dataset with a reference article).

DataCite provides a list of controlled vocabulary that allows to specify the relationship between resources. These relations may use DOI or any other sustainable identifier (eg: bibcode).

> **Note** The relationship vocabulary is precise. Too often, providers give their own interpretation of the terms. If there's any doubt about the meaning of the vocabulary, it's better to put no relations than a misunderstood relationship.

See relation vocabulary and their definitions in the DataCite Schema[31].

> **Note** Privilege DOI linking to any relations

DOI has to be privileged than any other identifier or URI. DOI facilitates the harvesting/merging jobs. For instance the EOSC EU Node portal displays related resources having DOI.

4.4.2 Link resources within a DataCenter

DataCenter may provide resources in different granularity. For instance Chandra provides catalogues of observations. Each observation is a Product having his own DOI. Then the products are linked to the catalogue which has a DOI too.

---

[31]https://datacite-metadata-schema.readthedocs.io/_/downloads/en/4.5/pdf/, section 12

Particular interest for Collections In the case of Collection, DataCite provides terms to link a resource to its collection. `DataCite:`**isPartOf** seems to be appropriate to link an individual Dataset to a Collection. The `DataCite:`**relatedIdentifiers** picked to express the relationship between elements of a collection and the collection are defined by the relationship in the hierarchy of the data holdings.

Include related identifiers linking the collection DOI to internal and if possible external identifiers for the dataset(s). This will make the DOIs broadly more useful and interoperable.

4.4.3   Link data to external resources

There are many reasons to link a resource with external products, we give just few examples:

- link the datasets that have been used to create a product

- link other copy

- for a derived product, link the original data

- cite a resource

- etc.

DataCite schema allows to qualify the relation. Although these qualifications are precise and must not be subjected to bad interpretation. We list a short list of terms often used for this purpose (Note that DataCite relationships have no scope restriction, they can be applied to link external or internal archives.)

- `DataCite:`**isIdenticalTo** (exact copy),

- `DataCite:`**isVariantFormOf** , `DataCite:`**isDerivedFrom**

- `DataCite:`**Cite**

- etc.

Linking Dataset to a publication   DataCites provides terms to link Datasets with articles. For instance `DataCite:`**Cites** , `DataCite:`**isSupplementTo** , `DataCite:`**isDescriebdedBy** , `DataCite:`**References** , etc. Choosing the good term is specific to the Datasets.

Note the `DataCite:`**Cites** relations which makes an automated citation (the article is cited when the DOI of the dataset is created). Other terms like `DataCite:`**References** generates statistics in DataCite too.

Citation between Articles and dataset is possible in either direction (term DataCite:**Cites** or DataCite:**isCitedBy**). The choice of citation depends on the workflow. Sometimes the article may cite a dataset that has been published previously. At other times, the dataset is published in conjunction with the article or has been reproduced in its identical form or with added value at a later date. It should be noted, however, that the notion of DataCite:**isCitedBy** requires a workflow synchronization that is not always easy to implement.

Example: link a dataset with its reference article.
The above example provides a relations using bibcode
(extract from https://doi.org/10.26093/cds/vizier.22640008).

```
"relatedIdentifiers": [
  {
    "relationType": "IsSupplementTo",
    "relatedIdentifier": "2023ApJS..264....8H",
    "relatedIdentifierType": "bibcode"
  }
]
```

Linking Dataset with its original resource   We distinguish a mirror copy and a resource derived from an original resource. Both can have their own DOI. We encourage derived resources as well as copy mirrors to link the original resource using DataCite relations.

There are lots of reasons why a DataCenter that provides a derived resource wants to provide a DOI. Often, the derived product is the result of a curation that provides added values.

Before creating a DOI, we encourage the DataCenter to tell their DOI plan to the original archive. Some metadata of the original resources should be copied in the derived product, and others will be dedicated to the derived product.

- Authors: the full authors (creators and contributors) list should be replicated in the derived metadata. The list can be completed with contributors involved in the derived product curation.

- Dates: the creation and the modification dates are the dates of the DataCenter that provide the derived product.

- Abstract: it is not needed to duplicate the original abstract, but the added curation should be explained.

- relations: add a link to the original product. You can Cite (relation type DataCite:**Cites**) the original resource that you can complete

with a relation explaining the relationship between the original and
the derived product.

Please : refer to DataCite:**relationType** vocabulary and do not extrapolate the semantic.

```
"relatedIdentifiers": [
  {
    "relationType": "IsVariantFormOf",
    "relatedIdentifier": "10.5270/esa-qa4lep3",
    "relatedIdentifierType": "DOI"
  },
  {
    "relationType": "Cites",
    "relatedIdentifier": "10.5270/esa-qa4lep3",
    "relatedIdentifierType": "DOI"
  }
]
```

4.4.4   Link other networks

Link an other network (see 4.1.1) is not a capability provided by DataCite.
However, networks using a specific identifier can be exploited to link the DOI
record.
DataCite:**alternateIdentifier** is a term to specify another identifier of the
same resource. For instance the Usual Name used by the Data Center or the
IVOID of the Virtual Observatory.

> **Note** When it exists, we suggest to add the IVOID of the resource
> in DOI using DataCite:**alternateIdentifier** and to add the DOI
> in the VO Registry records using ALTIDENTIFIER.

```
"identifiers": [
  {
      "identifier": "ivo://CDS.VizieR/j/mnras/320/451",
      "identifierType": "ivoid"
  }
]
```

4.4.5   Linked Data for Services and knowledgebase

It is again the role of the Data Center to choose relevant links. However, we suggest to limit links with a reference article.

For Query tool Service (or web Service) we haven't found any equivalent in DataCite of the SERVEDBY term of the VO registry to link datasets with services.

For Organized System Service, then terms like DataCite:**isSourceOf** , DataCite:**isCompiledBy** can be envisaged.

4.5   Versioning and evolving datasets

4.5.1   Manage version

Distinguish dataset versioning (the Data Center responsibility) and metadata versioning (automated versioning done by DataCite at each metadata update).

A good practice consists to follow a versioning mechanism for each data update. DataCite (version 4.6) recommends to create a new DOI for each major version and stipulating the version number with DataCite:**Version** .

It is recommended to adopt a version number following the pattern:

```
major_version.minor_version[.patch_version]
```

Different mechanisms exist:

- Zenodo [32] method: makes a DOI collection of versions. Each version has its own DOI which is linked (related identifier) with the DOI collection.

- Make 1 DOI for version and link the DOI version together using related identifier DataCite:**isNewVersionOf** , DataCite:**isPreviousVersionOf**

> Provide example of what a "data update" or "major version ?

---

[32]https://zenodo.org/

4.5.2 Evolving datasets

Versioning is well adapted for data subject to planned update, such as survey releases. Versioning implying a DOI per version is preferable for reproducibility. However, there are datasets that evolve regularly and for which versioning is inappropriate. For instance logs of observations evolve regularly.

For those types of datasets, we suggest adding in the DOI DataCite:**Title** or in DataCite:**Description** the evolving nature of the datasets and to modify the "Update" date at each modification.

We also recommend specifying the evolving nature in the DataCite:**ResourceType** (see examples in 5.1).

4.6 Tombstone page

In principle, a resource having a DOI should be sustainable (For instance, a DOI in an article links a resource which is relevant for its content).

If for any unfortunate reason, the resource needs to be discontinued, the landing page must be replaced by a "tombstone page". It is the responsibility of the data provider to maintain the URLs of their landing pages through their DOI registration agencies.

The tombstone page must include the following pieces of information:

- DOI, title, creators, citation (as for a regular landing page)

- Clear statement that the resource is discontinued

- Reason for the removal of the resource

- Link to a new version or other replacing resources (if applicable)

5 Metadata list

We focused on a sub list of terms needing a special attention in curation.

5.1 Choosing a resource type

The Resource type is specified with DataCite:**ResourceTypeGeneral** which uses a controlled vocabulary, completed with a DataCite:**ResourceType** which is a free text.

Examples: Choosing the good Resource type leads to the Data publisher. The following list is indicative only and does not constitute a rule of good practice:

	ResourceTypeGeneral	ResourceType
table	Dataset	Dataset
spectrum	Dataset	Spectrum
image	Image	Image
logs of observations	Dataset	Evolving Dataset
Collection	Collection	Data files, Documents...
Service VO	Service	IVOA Service
Service Web	Service	Web Service
Service/Project	Service or Project	
knowledgebase	Dataset	Evolving Dataset

Table 3

Example:

```
"types": {
  "resourceType": "evolving Dataset",
  "resourceTypeGeneral": "Dataset"
},
"titles": [
  {
        "lang": "en",
        "title": "The Gemini Observation Log (evolving Dataset)"
  }
]
```

5.2 Authors

Note for datasets. The good usage consists of adding all authors and not only the first author.

All authors are indeed asked by journals for citations. We also encourage the author's ORCID with their affiliations. When possible, inform the RoR.

Example of a unique author.

```
"creators": [
  {
    "name": "Ochsenbein, Francois",
    "nameType": "Personal",
    "givenName": "Francois",
    "familyName": "Ochsenbein",
    "affiliation": [
      {
        "name": "Observatoire astronomique de Strasbourg",
        "schemeUri": "https://ror.org",
        "affiliationIdentifier": "https://ror.org/04xsj2p07",
        "affiliationIdentifierScheme": "ROR"
      }
    ],
    "nameIdentifiers": [
      {
        "schemeUri": "https://orcid.org/",
        "nameIdentifier": "0000-0003-4667-015X",
        "nameIdentifierScheme": "ORCID"
      }
    ]
  }
]
```

Creators and contributors   DataCite distinguishes Creators and Contributors. The authors are conceptually the same as the Datacite term Creators. [PLEASE RE-]

The DataCite:**creator** can be a DataCite:**Person** or an DataCite:**Organization**. Contributors are persons or organizations that contributed to the development of the resources (sic DataCite Schema 4.6). A contributor has a role (eg: Editor, Supervisor, etc.) taken in a controlled vocabulary (see DataCite:**contributorType** term and the complete list in DataCite Schema).

In a way DataCite:**Contributors** is an alternative of DataCite:**Creators**. For instance, Zenodo provides a way to add a role to creators in its upload form. However, DOI created by Zeonodo used only Creator's term.

The usage of Contributor in citation has not been evaluated today.

Authors for Collection, Services, Knowledgebases Staff implied in this type of resource evolves. We recommend specifying the Organization (and the RoR if it exists) rather than contributors. Adding humans (for instance the main Creator of the service) is in the leaves to the Data publisher discretion.

5.3 Keywords

We encourage the usage of recognized keywords such as UAT[33], IVOA-UAT[34] or any keywords driven with a Web semantic.

The term in the DataCite schema is `DataCite:subject`.

Example:

```
"subjects": [
  {
    "subject": "Sky surveys",
    "valueUri": "https://astrothesaurus.org/uat/1464",
    "schemeUri": "http://astrothesaurus.org",
    "subjectScheme": "UAT"
  },
  {
    "subject": "Earth (planet)",
    "valueUri": "https://astrothesaurus.org/uat/439",
    "schemeUri": "http://astrothesaurus.org",
    "subjectScheme": "UAT"
  }
]
```

5.4 Dates

DataCite provides a list of terms to qualify dates such as `DataCite:Created`, `DataCite:Updated`, `DataCite:Validated`, etc.

The dates involve the publication in the data center only. For example, a dataset provided by a Data Center A may have been created months or even years ago and already published in a reference article or another Data Center B. The `DataCite:Creation` date of the resource published by A is the date of creation in A.

---

[33]https://astrothesaurus.org/
[34]https://www.ivoa.net/documents/uat-as-upstream/

27

Note Dates have not to be used as versioning.

Dates for Services, Knowledgebase and Collection By nature, these types of resources may evolve. The frequency of updates may make it inappropriate to use the update date. For changes in architecture or data model, versions should be used.

5.5   Title and description

Titles is an important metadata, it describes the resource and is often exploited by search engines (such as ADS or VO-registry) in a text search process.

Assigning a title is specific to each dataset. It is a short sentence that contains the most relevant aspects of the dataset but descriptive enough to be understandable/interpretable by non-experts. This would be in the light of supporting Open Science. For datasets derived or attached to a reference article, it is better to create a new description that describes the dataset.

Example : Reference article: ApJ (Draper Z.H, 2000), [35]
"Disk-loss and disk-renewal phases in classical Be stars. II. Contrasting with stable and variable disks"
Dataset Title : VizieR, https://doi.org/10.26093/cds/vizier.17860120
"Spectropolarimetric survey of classical Be stars"

Making a good title is of course specific to the dataset. For instance, we can name the object or type studied, the facility used, the release version, or the measurement methods if it is spectroscopy or photometry, etc.

Description completes the title. The ADS would index this as the abstract for the dataset record. Just like the title, ADS hopes this is a descriptive text, understandable for non-experts. Note that, like title, description describes the data and not its reference, even when the data comes from an original resource.

Datacite provides a description qualifier (Abstract, Methods, ...) which depends on the resource.

Description for data derived from external resource When the data derived from an external resource is not a simple copy, it is recommended to adapt the original description. For example, in the case of data attached to a published article, rather than reproducing the abstract of the reference

---

[35]https://doi.org/10.1088/0004-637X/786/2/120

article (which may also be subject to the same license as the article), the description may focus on the dataset content with added information such as those useful to produce the datasets.

Another good practice consists of adding in the description a reference to the data origin.

Example : VizieR DOI description example -

"VizieR online Data Catalogue associated with article published in journal Monthly Notices of the Royal Astronomical Society with title ...

5.6 Licenses

License is one of the metadata required by FAIR principles[36]

Institutes as well as countries encourage the usage of licenses for the datasets. For instance:

- NASA encourages CC0 licence (see SPD-41)

- French government imposes LO/OL or CC-by licenses (see Etalab)

Note Use machine readable licences

It is highly recommended to use machine-readable licences using the official term or the URL link:

- SPDX licenses https://spdx.org/licenses/

- Creative Common https://creativecommons.org

Datacite metadata (see "Rights" in the DOI metadata schema)

```
"rightsList": [
  {
    "rights": "Creative Commons Attribution 4.0 International",
    "rightsUri": "https://spdx.org/licenses/CC-BY-4.0.html",
    "schemeUri": "https://spdx.org/licenses/",
    "rightsIdentifier": "CC-BY-4.0",
    "rightsIdentifierScheme": "SPDX"
  }
]
```

---

[36]https://www.go-fair.org/fair-principles/

5.7 Funding references

Financial support can be attached to a DOI using DataCite:**fundingReferences** .

Funds information is expressed with a reference to the organism, generally using a Crossref Funder ID (OFR, Open Funder Registry)[37] or a RoR, and an award reference (see DataCite guidance)

Example: DOI generated by NASA Planetary Data System:
https://api.datacite.org/dois/application/vnd.datacite.datacite+xml/10.26007/97r3-1e19

Example: DOI generated by Chandra:
https://api.datacite.org/dois?query=10.25574/29770

5.8 Identifier for project

At the time of writing, RAID (Research Activity Identifier) is new in the open DATA scene. We do not have feedback to provide any best practice. However, an agreement between the RAID maintenance organisation (Australian Research Data Commons[38]) and DataCite makes possible to generate such identifiers from DOIs (using DataCite:**resourceTypeGeneral** "Project")

6 Recommendation checklist

1. Choose adequate granularity. The granularity depends on parameters such as citation, discoverability, traceability. Consider also that DOI need maintenance for the long term (see 4.1.1)

2. Select the **Resource type** and adopt the usage specific to the chosen category (see 5.1).

3. **Landing page** (see 4.3)

   - Check if DOI metadata are displayed in the landing page
   - Complete the landing page with added information such as links to datasets and any other metadata used in other dissemination workflows
   - Check if the landing-page is machine-readable

4. **Provide curated metadata**, in particular:

   - List all authors with ORCID and affiliation when they are known (see 5.2)

---

[37]https://www.crossref.org/services/funder-registry/
[38]AustralianResearchDataCommons

- Add machine readable licence (see 5.6)
- Add the publication date in the Data Center (see 5.4)
- Choose title matching with the Content and the type of the resource (see 5.5)

5. Check if **metadata allows to generate citations** (see 4.2).

   For ADS indexing, it would be desirable to have a description (abstract) and keywords in addition to the BibTex metadata.

6. **Link your Datasets** with resources (see 4.4)

   - Do not misinterpret the meaning of a semantic. The relations semantics are precise, in case of doubt it is better to put nothing. The IVOA DCP working group can also help you[39]
   - Link the DOI with resources which are used to generate the dataset. For instance datasets attached to an article, or when they come from an original archive.
   - add Alternate identifiers. In particular add IVOID if it exists (see 4.4.4)

7. Ensure the maintenance of the DOI
   Ensure that all your dissemination workflows are consistent, in particular, with IVOA registry and with "schema.org" (see 4.1.1)

8. Use a **versioning mechanism** each time your datasets evolved.
   Distinguish dataset versioning (the Data Center responsibility) and metadata versioning (automated versioning done by DataCite at each update) (see 4.5)

9. Ensure the sustainability of the landing pages.
   If for any unfortunate reason, the datasets is no more available, provide a Tombstone page explaining the reasons (see 4.6)

A Catalog of Repositories and Archives issuing DOIs

This is not an update of the Chen and Schmitz et al. (2022) appendix. This is meant to also provide additional information about what kinds of DOIs are minted and what types of deposits the archive will accept.

---

[39]mailto:datacp@ivoa.net

A.1 Domain Specific repositories

TODO add brief descriptions

1. CADC

2. CDA

3. China-VO

4. ESA

5. ESO

6. IPAC

7. MAST

8. NEA

9. PDS

A.2 Generalist repositories

1. Dataverse: Open source research data repository software.
   Dataverse includes data management from input to the output, provides DOI, OAI-PMH.

2. Dryad: Open data publishing platform for authors.

3. Zenodo: Zenodo is a popular platform for authors. Zenodo uses the DataCite and generates DOI. The curation leads to authors.
   Zenodo developed by the CERN provides Invenio (Open Software) which is an alternative to Dataverse

4. FAIRsharing FAIRsharing Team (2018): A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies. The Standards and endorsed notes provided by the IVOA have a DOI created by FAIRsharing

B Important DataCite Metadata Case Studies

This section describes case studies around the metadata created by three archives. These case studies let us break out specific DataCite metadata recommendations and JSON snippets

B.1 Chandra Data Archive

The CXC created DOIs for each single Chandra observation, data entities of the Chandra Source at the single observation and stacked observation level, and publication-based, user-requested Chandra Data Collections. CXC mints DOIs with its own DataCite prefix and different DOI names for each of the classes of DOIs created (see below). For each class of Chandra DOIs, a comprehensive set of Chandra-related DataCite metadata are automatically specified and updated upon specific trigger actions; multiple types of DataCite related identifiers are used to connect DOIs of related data entities in the Chandra archive".

See (D'Abrusco and DPAOps Team, 2023), (Rots and D'Abrusco et al., 2018).

B.2 VAMDC & Zenodo workflow

The good and the bad of transitory data citation. See (Moreau and Zwolf et al., 2018).

The VAMDC Consortium is a consortium of Institutes and Research Institutions that share a common technical and political framework for the distribution and curation of atomic and molecular data. VAMDC provides a Query Store including data storage for query result with possible DOI generation based on Zenodo

Example:
VAMDC query Store record: https://cite.vamdc.eu/references.html?uuid=a819e879-760d-4ede-831a-38d359b5864c
Landing page: https://zenodo.org/records/11243136

B.3 ESA

add ESA ?
Example:
Landing page: https://doi.org/10.5270/esa-qa4lep3
(JSON) DOI: https://api.datacite.org/dois/10.5270/esa-qa4lep3

B.4 VizieR

VizieR produces some awesome DataCite metadata.

The VizieR catalogue service is a CDS service for data published in journals or by space / base-ground agencies. The data are curated by CDS with added-values such as cross-correlations, links and metadata required for the

Virtual Observatory. It is a POST publication data flow that generates DOI, landing page, and which disseminate metadata, especially in the Registry of the Virtual Observatory.

Example:
Landing page: https://cdsarc.cds.unistra.fr/viz-bin/cat/J/AJ/167/89
IVOA record: https://cds.unistra.fr/registry/?verb=GetRecord&metadataPrefix=ivo_vor&identifier=ivo://cds.vizier/J/AJ/167/89
(JSON) DOI: https://api.datacite.org/dois/10.26093/cds/vizier.51670089

B.5   TODO B.5 MASER (Baptiste?)

C   Workflows

added by Gilles - TODO: discuss the relevance

Figure 1 represents a simplified view of Data dissemination. In particular, we see the importance of the IVOA registry harvested by EUDAT[40] and then EOSC[41] which cross records with DataCite.

Figure 2 proposes relations ships. This diagram had been proposed in a report feedback of the IVOA (A.Schaaff et al.)
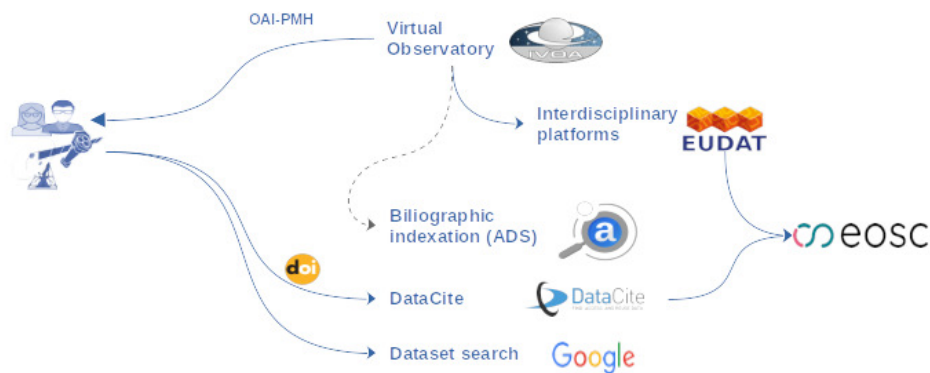

Figure 1: Dissemination workflow

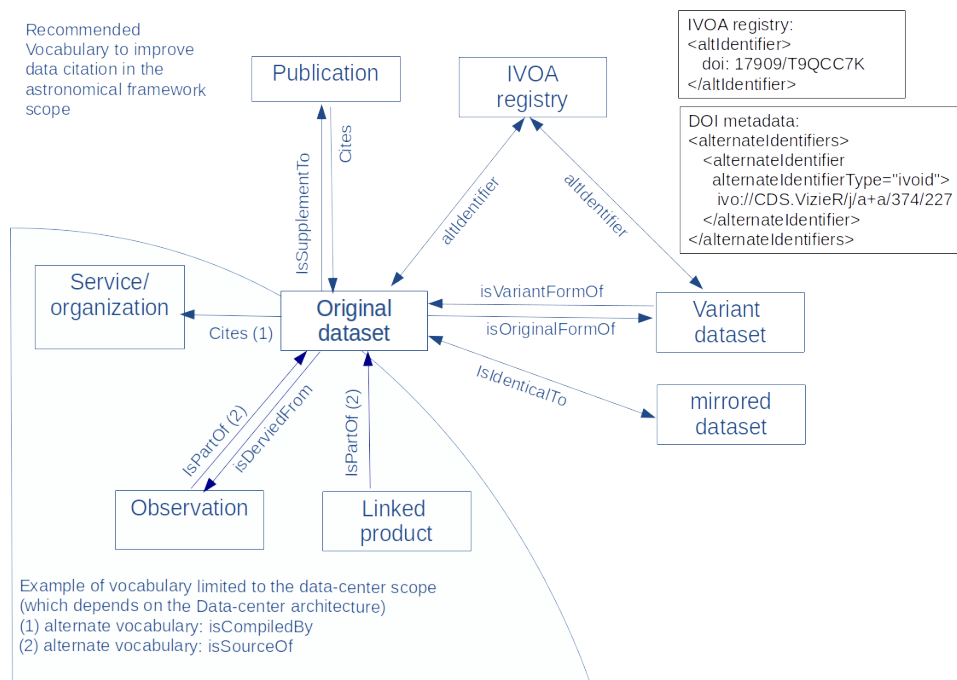## D   Changes from Previous Versions

No previous versions yet.

---

[40]https://www.eudat.eu/
[41]https://eosc.eu/

Figure 2: DOI schema

References

Akeson, R. L., Chen, X., Ciardi, D., Crane, M., Good, J., Harbut, M.,
Jackson, E., Kane, S. R., Laity, A. C., Leifer, S., Lynn, M., McElroy,
D. L., Papin, M., Plavchan, P., Ramírez, S. V., Rey, R., von Braun, K.,
Wittman, M., Abajian, M., Ali, B., Beichman, C., Beekley, A., Berriman,
G. B., Berukoff, S., Bryden, G., Chan, B., Groom, S., Lau, C., Payne,
A. N., Regelson, M., Saucedo, M., Schmitz, M., Stauffer, J., Wyatt, P.
and Zhang, A. (2013), 'The NASA Exoplanet Archive: Data and Tools for
Exoplanet Research', PASP **125**(930), 989, arXiv:1307.2944. doi:10.1086/
672273, https://ui.adsabs.harvard.edu/abs/2013PASP..125..989A.

Chen, T. X., Schmitz, M., Mazzarella, J. M., Wu, X., van Eyken, J. C.,
Accomazzi, A., Akeson, R. L., Allen, M., Beaton, R., Berriman, G. B.,
Boyle, A. W., Brouty, M., Chan, B. H. P., Christiansen, J. L., Cia-
rdi, D. R., Cook, D., D'Abrusco, R., Ebert, R., Frayer, C., Fulton,
B. J., Gelino, C., Helou, G., Henderson, C. B., Howell, J., Kim, J.,
Landais, G., Lo, T., Loup, C., Madore, B., Monari, G., Muench,
A., Oberto, A., Ocvirk, P., Peek, J. E. G., Perret, E., Pevunova,
O., Ramirez, S. V., Rebull, L., Shemmer, O., Smale, A., Tam, R.,
Terek, S., Van Orsow, D., Vannier, P. and Wang, S.-Y. (2022), 'Best
Practices for Data Publication in the Astronomical Literature', ApJS
260(1), 5, arXiv:2106.01477. doi:10.3847/1538-4365/ac6268, https://ui.
adsabs.harvard.edu/abs/2022ApJS..260....5C.

D'Abrusco, R. and DPAOps Team (2023), 'A New Way to Cite Chan-
dra Data', *Chandra News* **34**, 5. https://ui.adsabs.harvard.edu/abs/
2023ChNew..34....5D.

Demleitner, M. and Accomazzi, A. (2024), 'Bibliographic interfaces in the
virtual observatory'. https://www.ivoa.net/documents/Notes/BibVO/.

Demleitner, M., Plante, R., Linde, T., Williams, R., Noddle, K. and the IVOA Registry Working Group (2016), 'Ivoa identifiers'. https://www.ivoa.net/documents/IVOAIdentifiers/.

Dower, T., Demleitner, M., Benson, K., Plante, R., Auden, E., Graham, M., Greene, G., Hill, M., Linde, T., Morris, D., O'Mullane, W., Rixon, G., Stébé, A. and Andrews, K. (2014), 'Registry interfaces'. https://www.ivoa.net/documents/RegistryInterface/.

FAIRsharing Team (2018), 'Fairsharing record for: Simbad astronomical database'. doi:10.25504/FAIRSHARING.RD6GXR, https://fairsharing.org/FAIRsharing.rd6gxr.

Helou, G., Madore, B. F., Schmitz, M., Bicay, M. D., Wu, X. and Bennett, J. (1991), The NASA/IPAC extragalactic database., *in* M. A. Albrecht and D. Egret, eds, 'Databases and On-line Data in Astronomy', Vol. 171 of *Astrophysics and Space Science Library*, Springer, Dordrecht, pp. 89–106. doi:10.1007/978-94-011-3250-3_10, https://ui.adsabs.harvard.edu/abs/1991ASSL..171...89H.

Landais, G., Muench, A., Demleitner, M. and Savalle, R. (2024), 'Data origin in the vo'. https://www.ivoa.net/documents/DataOrigin/.

Moreau, N., Zwolf, C.-M., Ba, Y.-A., Richard, C., Boudon, V. and Dubernet, M.-L. (2018), 'The VAMDC Portal as a Major Enabler of Atomic and Molecular Data Citation', *Galaxies* **6**(4), 105. doi:10.3390/galaxies6040105, https://ui.adsabs.harvard.edu/abs/2018Galax...6..105M.

NASA Exoplanet Science Institute (2020), 'Planetary systems table'. doi:10.26133/NEA12, https://catcopy.ipac.caltech.edu/dois/doi.php?id=10.26133/NEA12.

NASA/IPAC Extragalactic Database (NED) (2019), 'Nasa/ipac extragalactic database (ned)'. doi:10.26132/NED1, https://catcopy.ipac.caltech.edu/dois/doi.php?id=10.26132/NED1.

Novacescu, J., Peek, J. E. G., Weissman, S., Fleming, S. W., Levay, K. and Fraser, E. (2018), 'A Model for Data Citation in Astronomical Research Using Digital Object Identifiers (DOIs)', ApJS **236**(1), 20, arXiv:1801.00004. doi:10.3847/1538-4365/aab76a, https://ui.adsabs.harvard.edu/abs/2018ApJS..236...20N.

Rots, A., D'Abrusco, R. and Winkelman, S. (2018), A Model for Using DataCite DOIs in Observatory Bibliographies, *in* 'European Physical Journal Web of Conferences', Vol. 186 of *European Physical Journal Web of Conferences*, p. 12011. doi:10.1051/epjconf/201818612011, https://ui.adsabs.harvard.edu/abs/2018EPJWC.18612011R.

Wenger, M., Ochsenbein, F., Egret, D., Dubois, P., Bonnarel, F., Borde, S., Genova, F., Jasniewicz, G., Laloë, S., Lesteven, S. and Monier, R. (2000), 'The SIMBAD astronomical database. The CDS reference database for astronomical objects', A&AS **143**, 9–22, arXiv:astro-ph/0002110. doi:10.1051/aas:2000332, https://ui.adsabs.harvard.edu/abs/2000A%26AS..143...9W.

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., da Silva Santos, L., Bourne, P., Bouwman, J., Brookes, A., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C., Finkers, R., Gonzalez-Beltran, A., Gray, A., Groth, P., Goble, C., Grethe, J., Heringa, J., t Hoen, P., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S., Martone, M., Mons, A., Packer, A., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016), 'The fair guiding principles for scientific data management and stewardship'. doi:10.1038/sdata.2016.18.