



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

Version 1.0

IVOA Note 2025-12-2r6

Interest Group

Data Curation and Preservation

This version

<https://www.ivoa.net/documents/DOI4Archives/2025122r6>

Latest version

<https://www.ivoa.net/documents/DOI4Archives>

Previous versions

This is a draft

Author(s)

August Muench, Gilles Landais, Raffaele D'Abrusco, Anne Catherine Raugh, Edwin Henneken, Baptiste Cecconi, Tim Jeness

Editor(s)

G. 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

1	Concepts and definition	4
1.1	DOI landscape	4
1.2	Use cases	5
1.3	Issues	7
1.4	Resources treated in the document	8
1.4.1	Datasets	8
1.4.2	Collections	9
1.4.3	Service	9
1.4.4	Note about Knowledgebase	9
2	DOI Lifecycle	10
2.1	The DOI Ecosystem	10
2.2	How it Works, In Brief	11
2.3	Considering a DOI Solution	12
2.3.1	DOIs are <i>forever</i>	12
2.3.2	DOIs are not free.	12
2.3.3	Minimal metadata isn't (usually) good enough.	12
2.3.4	Landing pages aren't optional.	12
2.3.5	When object permanence matters, objects must be permanent.	13
2.3.6	Assigning a DOI implies "ownership".	13
2.4	Getting Into the DOI Business	13
2.5	Minting a DOI	14
2.5.1	Step 1: Collect and format the metadata	15
2.5.2	Step 2: Create a landing page	15
2.5.3	Step 3: Formulate a unique DOI	16
2.5.4	Step 4: Submit the DOI, the landing page URL, and metadata to the RA for publishing	17
2.6	Updated resources	17
2.6.1	Versioning	18
2.6.2	Evolving Datasets	19

2.7	Nuances and Gory Details	20
2.7.1	Draft DOIs	20
2.7.2	Prefixes	20
2.7.3	Metadata Maintenance	21
2.7.4	Tombstones	22
3	Core Metadata	22
3.1	Metadata consistency	23
3.2	Assigning resource type	23
3.3	Metadata list	25
3.3.1	Assigning the publisher	25
3.3.2	Title and description	26
3.3.3	Keyword is better with Web semantics	27
3.3.4	Dates	28
3.3.5	Creators and contributors	28
3.3.6	Licenses	31
3.3.7	Fundings	32
3.4	Citation extraction	32
3.4.1	Curate DOI is mandatory to extract citation	33
3.4.2	Using DOI for Evolving DataSet	34
3.4.3	Citing Collections	35
4	DOIs across networks	35
4.1	Dissemination of resources through complementary networks	35
4.2	Linked data with internal and external resources	36
4.3	A basic tour of relationship motivating by their usage	37
4.3.1	Express a Relationship	37
4.3.2	Term to link resources with their origin	38
4.3.3	Dealing with Collection	39
4.3.4	Managing version and obsolescence	40
4.4	Connecting DOI and the IVOA	41
4.4.1	Status of DOI in the Virtual Observatory	41
4.4.2	DOI and IVOA networks complete each other	41
4.4.3	How and why to connect DOI and IVOA networks?	42
4.5	DOI consuming in ADS/SciX	42
4.5.1	SciX curation requirements	43
4.5.2	Meta-data harvesting.	43

5 Appendix	44
5.1 Recommendation checklist	44
5.2 Building landing pages	45
5.3 Example of existing DOI	46
5.3.1 Chandra Data Archive	46
5.3.2 VAMDC & Zenodo workflow	46
5.3.3 ESA	47
5.3.4 VizieR	47
5.3.5 PDS	47
5.3.6 NSF-DOE Vera C Rubin Observatory	48
References	48

Acknowledgments

Alberto Accomazzi (ADS), Christophe Arviset (ESA), Arnaud Masson (ESA), Maria Henar Sarmiento (ESA), Markus Demleitner (GAVO), Andre Schaaff (CDS) and all participants of the DCP sessions.

1 Concepts and definition

1.1 DOI landscape

The DOI is widely used in the academic publishing process for articles, datasets or services. It is a persistent identifier created in 1996 for the long-term management of digital resources which includes both a URL and metadata that can be modified over time. These capabilities allow the DOI to be in phase with the data lifecycle and to be free of possible change of URL location.

The DOI is particularly well suited to citation. From its metadata, it is possible to extract the information needed to generate BibTeX. In addition, the DOI has a resolver that translates a DOI into a URL. The mechanism is supported by an architecture that allows to modify the URL over time. This feature makes the DOI easy to use for citing or linking resources in articles or from any web pages.

The architecture is robust and has been generalized to identify other types of data than those mentioned above. For example, DOIs are used for instruments ([Rubin Telescope](#)), and the [DataCite](#) organization which provides DOI for datasets, is a partner of [ARDC](#) to register identifiers for projects ([RAID](#)).

With the advent of open science, DOI has taken on a key role in the dissemination of information. It enforced the data traceability by linking

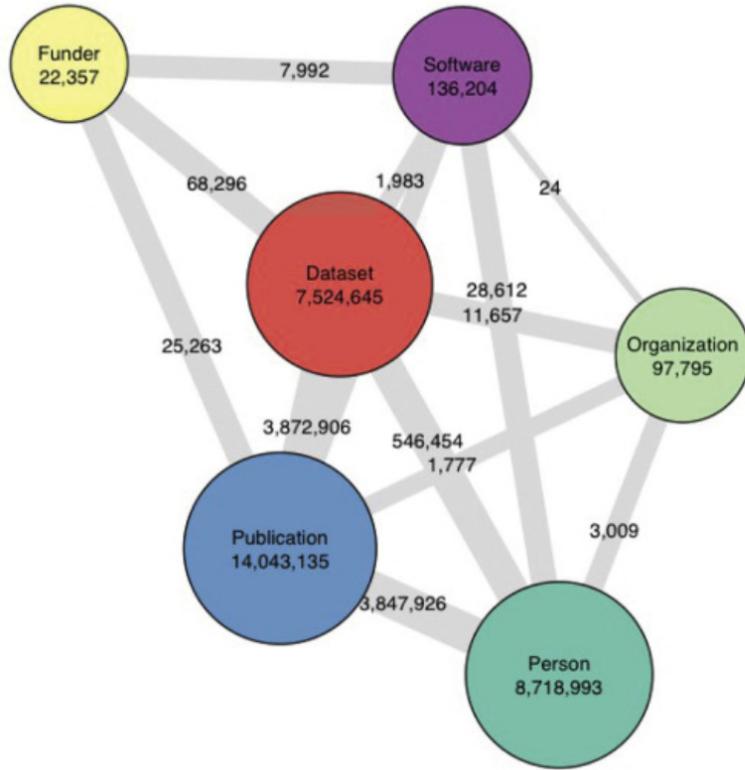


Figure 1: Connection between the Entities in the PID Graph, August 2020 (Cousijn et al.)

resources of different types according to relationships of dependency or authorship and includes other identification standards such as the author identification ([ORCID](#)), organizations ([ROR](#)) or Funders.

DOI enables the implementation of search engines and knowledge graphs, improving the discovery and traceability of digital resources.

1.2 Use cases

There are many reasons for creating a DOI. Below, we list the use cases that we found to be particularly relevant and which motivated to write this note.

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 articles in

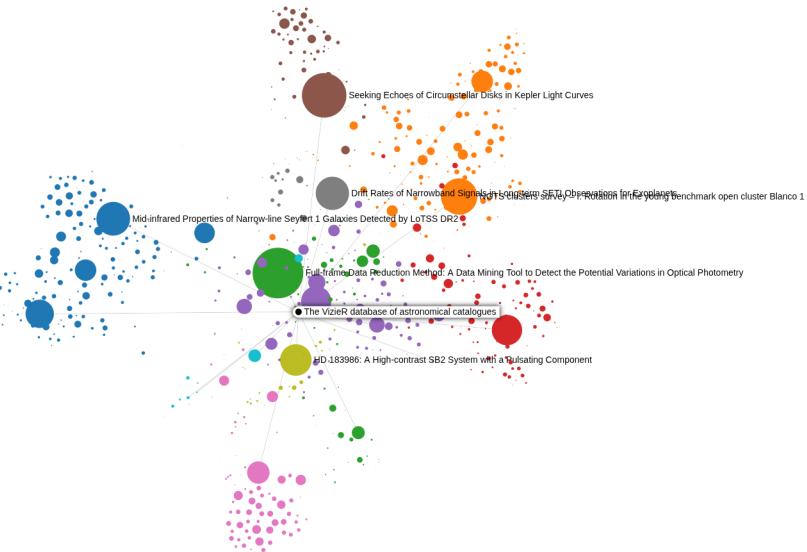


Figure 2: DOI Graph created by citationgraph.org from a vizier DOI

the same way and use the same tools to count citation and measure impact. Hence data producers and providers are expecting that data reuse is acknowledged through regular citation.

2. **Reuse.** The Data consumer expects to get information about origin and rights. For access rights, existing license must be displayed but also readable by software.
3. **Findability.** Data published are disseminated in networks like the VO registry ([Dower and Demleitner et al., 2014](#)), google dataset search or the [EOSC EU](#) node. 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 are able to make reports and to generate graphs of citation.
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 knowledge base, he or she expects that the protocols stay compatible with the original implementation. For datasets, he or she should expect that data in the original study have not been altered.

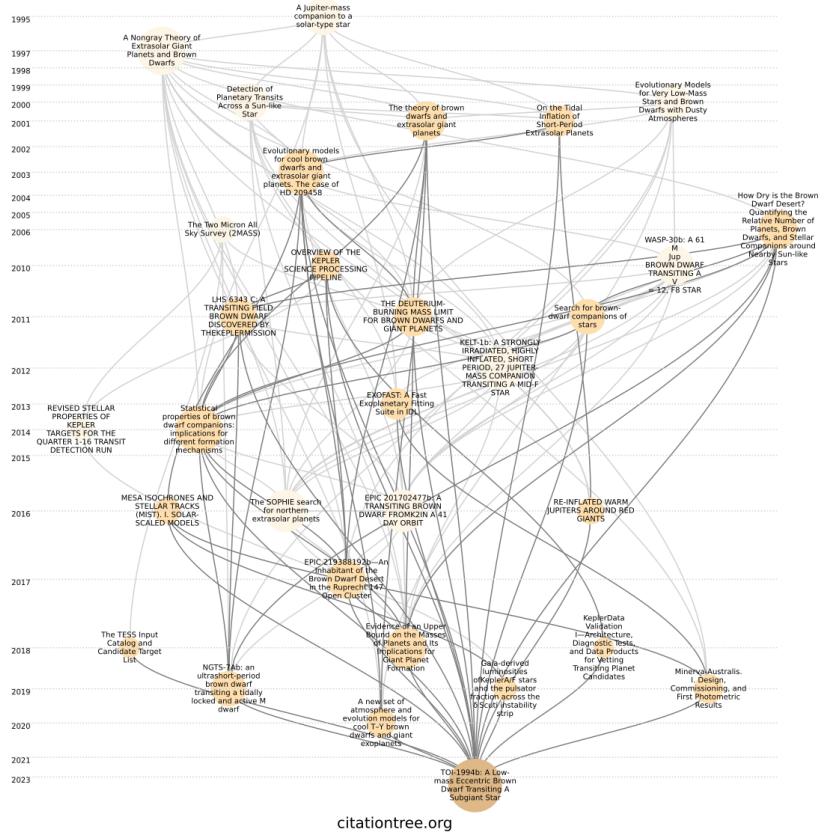


Figure 3: DOI Graph created by citationgraph.org from an article published in AJ (doi: 10.3847/1538-3881/ad1a18)

1.3 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 into DOI schema.

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.

1.4 Resources treated in the document

We observe 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, for knowledge base, etc.

1.4.1 Datasets

Dataset DOIs are the most common usage in astronomy-related repositories, and represent the most important part of all such DOIs. A dataset DOI is used for various resources that can be static, evolving and even multiple.

DataSet for individual Resource Individual resources are being assigned DOIs at various levels of granularity. At the most granular level, every observation 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. In a wider granularity, the European Space Agency ([ESA](#)) assigns a DOI per data release.

Dataset for multiple resources A DOI on multiple resources gathers together resources under the same entity. For instance a VizieR catalogue ([Ochsenbein, 1996](#)) possibly includes a collection of tables, spectra and images all attached to the same published article.

Evolving datasets Evolving dataset is a particular case of Dataset. They include resources such as logs of Observations which are updated incrementally, and they also include resources, such as a knowledge base or "High-Level Science Products" (HLSP) which have non-static data and can grow over time, accumulating both data revisions and additions.

Resultsets A Resultset is the result returned by a query. It is the result of a service, a user request and datasets executed at a precise moment, which may be not reproducible if it involves evolving data. Thus, certain services such as the [VAMDC](#) or [Canfar](#) provide a Query Store with a DOI (Resultset have not been treated in the scope of this Document).

1.4.2 Collections

The DOI collection directs 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 relation (see 4.3) defined in the DataCite schema (DataCite Metadata Working Group, 2024)

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

In any case, **Collection for DOI means a collection of resources accessible via sustainable URIs.**

1.4.3 Service

A service in DOI designates both a set of services accessible via the web and the whole organizational system, including staff, necessary for its operation.

An example of Service DOI is the VizieR Catalogue service including many access to a database ¹ or the Tübingen Oscillator Strengths Service (TOSS) in GAVO ²

1.4.4 Note about Knowledgebase

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

Knowledgebase is not a particular entity in the DOIs schemas ; it can be a dataset or a service depending on what has to be described (see [Assigning resource type](#)).

¹<https://api.datacite.org/dois/application/vnd.datacite.datacite+xml/10.26093/cds/vizier>

²<https://api.datacite.org/dois/application/vnd.datacite.datacite+xml/10.21938/3I01ISNUC0DNH1ZJBCVUWA>

2 DOI Lifecycle

DOIs are born, but they never die. This means they need long-term curation and management as well as contingency planning to ensure they are useful for generations to come.

2.1 The DOI Ecosystem

DOIs are supported by a hierarchy of agencies that is intended to insure that any DOI created will continue to resolve to something meaningful no matter what happens to the original creator of the DOI. This system includes:

- **Registration Authorities (RAs)**

The RAs are the key members of the [DOI Foundation](#) - the top-level of DOI support. The RAs allocate prefixes, register DOIs, and provide a schema to govern the metadata they collect for each DOI, which varies from RA to RA. Most RAs serve a specific type or population of DOI user.

[bsi.identify](#), for example, deals specifically with DOIs used to track construction products; [EIDR](#) provides DOI services for entertainment media; Japan, China, and Korea all have RAs providing DOI services for their agencies and citizens, etc. There are currently 12 RAs in the DOI Foundation. [DataCite](#) is one of these RAs, and was the first RA specializing in providing DOIs for data resources. The RAs together ensure that if one RA ceases operation, the DOIs that that RA managed will then be managed by one or more other RAs in the Foundation.

- **Users - DOI Creators**

RAs provide (or more often sell) DOI services to *users* - people who want to create DOIs for various reasons. Each RA will organize its users according to its own business model. DataCite does this through memberships and service fees. DataCite has two types of members: Consortia, in which one organization acts as a DOI agent for many smaller organizations (managing billing and often providing software for the DOI pipeline); and Direct Members, who are typically single organizations managing their own DOI billing/pipelines.

- **Users - DOI Consumers**

The RAs make the DOIs and their associated data available to the public. Typically the basic metadata is freely available for public reference and use. DataCite, for example, makes all metadata in their database accessible to the world at no charge.

2.2 How it Works, In Brief

Every DOI consists of a prefix and a suffix. In the DOI "10.26007/97r3-1e19", for example, the prefix is "10.26007" and the suffix is "97r3-1e19". To resolve this DOI you could enter <https://doi.org/> into a browser and you would be directed to the *landing page* (see 2.3.4) for that DOI. Here is a brief summary of what is going on in the background to reach this point:

- Every RA has a distinct domain of prefixes it can assign to its DOI creators.
- The RA assigns one or more prefixes to each creator. The same prefix is never given to two different DOI creators, but one creator may have more than one prefix assigned to them that they can use as desired.
- The DOI creator is responsible for making sure that every suffix assigned to a prefix is unique, so that the combination of *prefix/suffix* is guaranteed to be unique in the world. (The RA will check this before publishing the DOI.)
- The DOI creator is also responsible for curating the metadata for every DOI they create - in particular making sure there is a *landing page*, which is an absolute requirement for all published (that is, resolvable) DOIs.
- The DOI metadata is uploaded to the RA's metadata database. Most RAs allow DOI creators to reserve DOIs and upload metadata before the DOI is public. (more details in [Draft DOI](#))
- When a DOI creator wants to publish the DOI, the metadata is checked for minimal completeness by the RA and, if that is satisfied, the RA enables public access to the metadata.
- DOI consumers can then find the metadata directly in the RA's public database and can resolve the DOI through the <https://doi.org> URL demonstrated above.
- Once published, then DOI will *always* be resolvable through <https://doi.org>. If the object goes away, the landing page must be replaced with a tombstone page (see [Tombstones](#)) that tells users that the object is no longer available. If the DOI creator goes away, their prefixes can be transferred to another creator. If no other creator will take over curation, the RA will. If that RA goes away, the remaining RAs in the DOI Foundation will take over curation.

2.3 Considering a DOI Solution

Because a DOI is itself a permanent entity that is generally expected to lead to a permanent entity, there are a number of things a potential DOI creator should consider before opting for a DOI solution.

2.3.1 DOIs are *forever*.

Once a DOI is published it must always be resolvable. If a resource with a DOI is removed from circulation (a data set is offlined, for example, or a digital document is accidentally destroyed) then the DOI must resolve to a tombstone page. A tombstone page ([Tombstones](#)) indicates clearly that the resource is no longer available, and usually offers contact information or an option for users who are attempting to lay hands on the now-defunct resource. The tombstone page must remain as a legacy for as long as there is a DOI Foundation to resolve it.

2.3.2 DOIs are not free.

There is a nominal cost for creating a new DOI, but a much larger expense is tied up in the effort needed to create and curate the metadata associated with the DOI and its landing page. This might be done through manual editing of individual metadata files, or through coding of automated procedures to pull DOI metadata from an internal database.

2.3.3 Minimal metadata isn't (usually) good enough.

If DOIs are being created to make resources more findable and accessible, minimal metadata is not going to be good enough to check that box. Long, descriptive titles, subject keywords from standard vocabularies, abstracts containing significant and distinctive details, ORCIDs, ROR IDs, and in particular linking to the DOIs of related resources are all essential to improving findability. Explicit digital licensing is essential for accessibility, which also benefits from MIME types, file sizes, and similar descriptive metadata.

2.3.4 Landing pages aren't optional.

Metadata and landing pages are connected in the sense that the registered metadata in the very least should not contradict what is displayed on the landing page.

The DOI creator is generally expected to host the landing page or tombstone page for as long as the creator is curating any DOIs with the same prefix. Landing pages also typically require more curation than the DOI

metadata itself, particularly if the hosting organization is interested in making sure the pages are scraped for inclusion in databases like Google Scholar, or can be used to track downloads and citation.

2.3.5 When object permanence matters, objects must be permanent.

If DOIs are being created for resources that are considered part of the research literature and are intended to be used for citation to establish provenance and to support reproducibility, then it is essential that the DOIs resolve to the *same thing* every time. For example, if a data set is used as the basis of an analysis and is cited by DOI, then that DOI should always return the version of the data used in the analysis - not a later revision that might give a different result.

Note that the advice above is strictly limited to the case of DOIs that must support reproducibility. There are other reasons for creating DOIs that do not assume or require permanence (more detail in [Updated resource](#))

2.3.6 Assigning a DOI implies "ownership".

If the potential DOI creator has no control over the resource to which the DOI will be assigned, the DOI should not be assigned. If, for example, a DOI creator has no control over what URL a resource might be moved to, then the DOI creator cannot maintain the linkage between the landing page and the data. The DOI may suddenly be unable to resolve to a landing page that can lead a user to the resource.

2.4 Getting Into the DOI Business

If DOIs are the right solution, the following preparations are highly recommended prior to committing to a specific DOI solution:

Identify a DOI Service Provider. This might involve becoming a direct member of an RA, joining a consortium, or working through an existing office in a parent organization. If any part of the larger organization is already issuing DOIs, that office can likely point the way to their service provider. Direct membership in an RA might be preferable if the number of DOIs to be produced annually is large (several thousand or more), or the current provider does not provide a metadata schema with attributes appropriate to the resources to be tagged with DOIs (see the following). Cost may well be a significant factor in this decision, as might the desire/preference for either API or form-based access (or both) to DOI services.

Map the available local metadata to the RA schema. DOI creators must supply metadata according to the schema used by the RA that will be registering the DOIs. Each RA provides its own schema. The DOI creator should make sure that the metadata in hand can be mapped to the

RA schema, and that the RA schema contains sufficient attributes to support the reason(s) for assigning DOIs.

Draft the landing page design. Landing pages should minimally be able to take the user to the data in one or two clicks. Landing pages can also be designed to increase Search Engine Optimization, entice sites like Google Scholar to create listings, encourage citations, and help in tracking download metrics. Landing pages will need to be designed to meet the organizational goals, and then be hosted, maintained, and curated in a publicly accessible place. Specifically, the landing page cannot be behind a paywall or login. The data resource might be, but the landing page must be visible to all. More details about how to create a landing page are given in Appendix.

Define DOI policies and strategies. The bulk of this document is intended to assist in this effort. It is important, however, to think through and plan out how to manage DOIs as an institutional asset, as well as how to deal with the circumstances expected to arise. These include things like:

- Offlining old data
- Rapid drafting of DOIs to meet publication deadlines
- Transferring a DOI prefix (and its DOIs) to or from another organization
- DOIs for service portals
- DOIs for data that aren't static
- etc.

Consider curation. Depending on what an organization wants to achieve with DOIs, curation may involve: very little - fixing a typo here and there as it is noted, for example; a lot - tracking citations, enriching metadata as schemas evolve, and such; or the occasional crisis - having to change every URL because of a system failure. It is important to make sure that the effort available for routine curation, in particular, matches both the organizational goals for assigning DOIs to the resources and the available workforce.

2.5 Minting a DOI

The process of creating a new DOI, often called "minting", is conceptually very simple:

1. Collect and format the metadata
2. Create a landing page

3. Formulate a unique DOI
4. Submit the DOI, the landing page URL, and metadata to the RA for publishing

In actual practice, of course, there are many roads leading to Rome. The following sections elaborate on the various methods for accomplishing each step.

2.5.1 Step 1: Collect and format the metadata

This can range from an entirely manual activity to an entirely automated one. The manual effort to gather metadata can be significant, especially for heterogeneous repositories.

RA, such as DataCite or CrossRef, fixed Metadata schema, formats, and DOI submission modalities. The RA generally provides a form-based interface and APIs covering the registering protocols and metadata formats (such as JSON or XML). Metadata schema, formats, and DOI submission modalities are fixed by the RA. The RA generally provides a form-based interface and APIs covering the registering protocols and metadata formats (such as JSON or XML)

For DOI creators minting small numbers of DOIs (a few dozen per year, say), the RA may provide a form-based interface for creating DOIs that guides the user through the most commonly used metadata attributes. DataCite, for example, has its [Fabrica](#) interface for interactive DOI creation. (It can also be used to edit existing DOI metadata, one record at a time.)

Some RAs also provide APIs for creating, drafting, publishing, and updating DOI metadata. DOI creators can code their own custom DOI minting interfaces using these APIs. Note that these APIs will require credentials to make any change to the metadata in the RA database, which may not be available if DOI service is being provided by an intermediary organization.

Sometimes, RA provides a test and production environment. In DataCite, the “[test Fabrica](#)” is a production clone that emulates the DOI generation. It is common for DataCenter who develops a DOI workflow.

2.5.2 Step 2: Create a landing page

The landing page must exist and be publicly accessible (i.e., not behind a paywall or login) before the RA will publish the DOI.

The landing page itself can change, of course, and the DOI curator will always have permission to update the DOI metadata to change the URL if needed. If the resource is behind a paywall or login, the landing page should provide that notice as well as information about how to get permission to access the resource.

The document gives a good practice for landing page creation in the appendix. Note, that the landing page can evolve during all its life cycle. For instance a pre-publication landing page is modified when the resource is published and it may evolve again for a deaccessioned resource (see tombstone page at the end of this section).

Note: Metadata and landing pages are connected in the sense that the registered metadata in the very least should not contradict what is displayed on the landing page.

2.5.3 Step 3: Formulate a unique DOI

The DOI creator will be assigned one or more prefixes by the RA that will be unique to that creator. The DOI creator has a lot of freedom when it comes to creating suffixes to append to that prefix. The DOI Foundation allows "any printable characters from the legal graphic characters of Unicode", although it does stipulate that the string must be interpreted case-insensitively. So it is technically all right for the prefix '/' to be followed by numbers, letters, punctuation - including the slash ('/') character - and any other printable Unicode character. In practice, however, it is generally wise to avoid characters that will make life miserable for those who need to deal with DOI strings in forms, text, URLs, and APIs.

From an information science perspective, it is generally considered a good practice *not* to try to encode meaning into things like DOI suffixes, but sometimes a unique internal identifier is a very convenient way to ensure a unique DOI. There are utilities available to generate hashes that can be used when a unique identifier is required and not otherwise handy. DataCite, for example, will auto-generate a suffix that has a number of useful characteristics - like avoiding the letters i, l, and O that are so often confused with 1 and 0.

See also the DataCite blog [Cool DOI](#).

Example: <https://doi.org/10.3847/1538-4365/aab76a>

- the prefix 10.3847 is the prefix attributed by Crossref to the AAS journals
- the suffix 1538-4365/aab76a defined the article

Note Once created, the DOI is fixed. It is a persistent identifier dedicated to a single resource and can't be reused for another resource.

2.5.4 Step 4: Submit the DOI, the landing page URL, and metadata to the RA for publishing

Once the landing page is live and the metadata are complete (and validated, if possible), the information can all be uploaded to the RA with instructions to publish the DOI. Typically, the DOI will be available to the public from the RA's metadata database within minutes. It may take anything up to a day to propagate to the <https://doi.org> service, depending on load and maintenance schedules.

The RA intake systems are smart enough not to overwrite an existing DOI by an attempt to create a new DOI with a DOI string that is already in use. As expected, they will also refuse to create a DOI without authenticating credentials, or create a DOI with a prefix that is not assigned to the credentials used.

Once published, the DOI cannot be removed, even if it was created in error. In the event of error, the metadata should be nulled out and the URL directed to a landing page that explains that the DOI was created in error and corresponds to no resource.

2.6 Updated resources

The resources subjected to updates require the DOI creator to adopt a management plan. The plan depends both on the nature of the updates and the nature of the resource.

In this section, we treat different use cases:

- Evolving datasets (see the definition and examples in section [Evolving datasets](#))
- Corrections on datasets contents
- Versioning dataset: for instance a version per data release (eg: SDSS DR1, ...)

Foremost, it is important to distinguish dataset versioning which is managed by the DOI creator, and metadata versioning which is operated by the RA for each metadata update. In this section we will talk only on dataset versioning.

DataCite recommends "**Register a new identifier for a major version change**".

It is out of the scope of this document to give a definition of what is a major change. However, we give 4 concretes examples:

- A log of observation is a dataset updated regularly (may be daily). In that case, the Data Center doesn't update the metadata (not even the update dates) because it is the nature of the resource to evolve.

Example: VizieR Occultation lights curves

<https://doi.org/10.26093/cds/vizier.102033>

- Fixing typo in a dataset (for instance a column containing authors names) could be managed as a minor correction. It doesn't need a new DOI (ie no versioning), but it is recommended to update the **DataCite:dateType** "updated" attributes and possibly update the version number (**DataCite:version**)
- Any correction that impacts the scientific result should be subject to versioning (see [Versioning](#))
- A new data release must be subject to versioning

Example ESA GAIA release DR3: <https://doi.org/10.5270/esa-qa4lep3>
 ESA GAIA release DR2: <https://doi.org/10.5270/esa-ycsawu7>

2.6.1 Versioning

DataCite (version 4.6) recommends creating a new DOI for each major version and stipulating the version number with **DataCite:version**. The versioning mechanism in DataCite schema is based on relationships terms **DataCite:relatedIdentifier**. Versioning allows access to the landing page to each versions. Each version has its own DOI and is linked to the other versions.

Different mechanisms exist:

- Zenodo method: makes a DOI collection of versions. Each version has its own DOI
 Version and collection DOI are linked together using **DataCite:relatedIdentifier** (HasVersion and isVersionOf)
- Make 1 DOI for version and link the DOI version together using related identifiers and attributes **DataCite:relationType** "isNewVersionOf" and "isPreviousVersionOf".

Examples:

- Data for Figures in Unexpected Long-Term Variability in Jupiter's Tropospheric Temperatures
 Zenodo creates a DOI which contains all version (<https://doi.org/10.5281/zenodo.5768656>)

```

"relatedIdentifiers": [
  {
    "relationType": "HasVersion",
    "relatedIdentifier": "10.5281/zenodo.5768657",
    "relatedIdentifierType": "DOI"
  },
  {
    "relationType": "HasVersion",
    "relatedIdentifier": "10.5281/zenodo.7336240",
    "relatedIdentifierType": "DOI"
  },
  {
    "relationType": "HasVersion",
    "relatedIdentifier": "10.5281/zenodo.7583188",
    "relatedIdentifierType": "DOI"
  }
]

```

Each of this versioned DOI use a relation **DataCite:relationType** "IsVersionOf" to the main collection.

Ex: <https://api.datacite.org/dois/10.5281/zenodo.7336240>

- Example of Dataset, published by ASTROMAT using **DataCite:relationType** "isNewVersionOf"
<https://doi.org/10.26022/IEDA/1129401129>

Semantic versioning. Version includes usually different levels. The first level refers to major releases, the second level to minor releases, and a third level is added for patches (corrections, etc.). For example, "semantic versioning" (Preston-Werner, 2023) focuses on software. Others, such as "Calendar versioning" (Hashemi, 2016), suggest adding a date to the version as a reference point.

2.6.2 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.

There is no dedicated "evolving" metadata in DataCite yet. The recommendation given below remains informative for the end users.

For those types of datasets, we suggest adding in the DOI **DataCite:Title** or in **DataCite:Description** the evolving nature of the datasets and, when applicable, to update the **DataCite:dateType** "updated" attribute. We also recommend specifying the evolving nature in the **DataCite:ResourceType** which is a free text. However, these recommendations are just descriptive, they are not used for instance in citation generation tools such as those

offered by DataCite. They remain informative and can be used for a home-made citation creation.

Example: VizieR Occultation lights curves
<https://doi.org/10.26093/cds/vizier.102033>

2.7 Nuances and Gory Details

Following are additional details and items of note for DOI creators and curators.

2.7.1 Draft DOIs

Because publishing a DOI is an irrevocable act, most RAs make it possible to reserve a DOI in a non-public state, typically called "Draft" or "Reserved", so that the metadata can be created, examined, and tested before going public. A DOI in a draft state can be deleted entirely, but as long as it exists in the RA's database the DOI string is protected from being duplicated. Publishing a draft DOI typically only involves changing a flag in the RA database.

Draft metadata is not visible to users other than the DOI creator (identified by credentials) who created the draft.

Draft DOIs can be used to eliminate race conditions with data citations. A draft DOI can be supplied for citation of the data in a paper in preparation, as long as the journal editor is confident that the DOI will be live by publication of the paper. In the event of a very skeptical journal editor or a very careful referee, a draft DOI can be published with a pointer to a landing page explaining that the data are, for example, in hand and undergoing review, and that those wishing to see the data before it is released can contact some named party for access. Once the data are published, the already-published DOI and landing page can be updated accordingly.

Some repositories provide draft DOIs as a matter of course to facilitate timely publication.

A good example of a repository providing DOI registration before publication is the [China-VO Paper Data](#). Frameworks such as [Invenio](#), [Dataverse](#) or the [U.S. Department of Energy Office of Scientific and Technical Information \(OSTI\)](#) also provide this possibility.

2.7.2 Prefixes

Prefixes are assigned by the RAs to specific DOI creators. Typically, a creator can have as many prefixes as they need.

There are two reasons why a DOI creator might want more than one prefix:

1. **Accounting.** Separate prefixes make it possible to allocate costs to specific DOI applications, and also make it easier to generate statistics by prefix for DOIs that have been created.
2. **Planning.** It is generally pretty easy to transfer all the DOIs associated with a single prefix to a new curator. It is all but impossible to transfer a subset of DOIs from a single prefix to a new curator. If it is possible that a set of DOIs might need to be passed off to another curator, it is wise to plan for that by giving them a unique prefix.

As a practical example, the Planetary Data System Small Bodies Node ([SBN](#)) is a Direct Member of DataCite. The SBN has three parts: The Asteroid Subnode in Arizona; the Comet Subnode in Maryland; and the Minor Planet Center (MPC) in Massachusetts. Each of these subnodes has its own DOI prefix. The MPC generates thousands of DOIs a year for its publications. It has its own prefix primarily for accounting purposes. The Asteroid and Comet subnodes create a few dozen DOIs each per year. They have their own prefixes primarily for planning, so that if the organization of the node changes during a recompete, the DOIs of one (or - let's hope not - both) of these subnodes can easily be transferred to a new institution.

2.7.3 Metadata Maintenance

It is important to note that the only thing that is truly permanent in the DOI metadata is the DOI value itself. Everything else can change. Some of it *shouldn't* change, but it all *can* change. So it is possible, sometimes even advisable, to update a DOI metadata record with a new, more explicit title and abstract when the existing content is deemed too vague or too likely to produce unsatisfying search results. And it is certainly possible to fix any typos, correct mistakes, and upload "missing" metadata as these issues are discovered. This is part of the value of DOIs - the metadata can be actively curated to serve the resource it is describing. The immutable DOI provides the permanent link.

Care should be taken to ensure that the metadata used for citing a resource is complete and correct at the time the DOI is published. Many systems generate citation and reference strings automatically from DOI metadata. Changing an author list or publication date after a citation has been printed, for example, can have a large impact on individuals.

Resources will usually dictate what can be curated after publication. The only thing that absolutely must be curated is the link to the landing page. Moving or reorganizing the site that hosts the landing pages can result in a major effort to update all affected URLs. Not doing so reflects badly on the RA, and thus is usually noted at a high level.

RAs are also concerned about metadata completeness. Incomplete metadata leads to bad search results from the RA metadata databases. This also reflects badly on the RAs, and there is increasing pressure to "do something about it". In general, all applicable fields of the RA metadata schema should be filled wherever possible.

Once these baselines are met, there are other curatorial chores to perform as resources allow. Some organizations, for example, actively track citations of their products and add the information to their own DOI metadata. Also, many archives are now thinking seriously about "metadata enrichment" - adding metadata to both standardize and improve the quality and success of searching, accessing, and (re)using their data holdings.

2.7.4 Tombstones

A tombstone page is used to indicate that the associated DOI does not lead to a resource. This could be for a number of reasons:

- The resource has been near-lined (moved to cold storage but still available on request).
- The resource has been off-lined (not available online without special arrangement).
- The resource has been deleted/purged/lost (not available, period).
- The resource was never received.
- The DOI was published in error.

The point of a tombstone page is to prevent a user from wasting time continuing to seek a resource that is no longer available. When a resource is available but requires special procedures for access (like retrieval from deep storage), the tombstone should provide instructions or contact information to make the request. Some tombstones may contain or point to the available metadata, or perhaps some documentation, to help a disappointed user.

Except in the case of a DOI published in error, the DOI metadata related to a tombstone page should be left intact.

In general, tombstones should be rare, but they are an effective way to maintain a link to legacy data cited in the literature once improved versions are available and getting the bulk of the contemporary attention.

3 Core Metadata

3.1 Metadata consistency

A DOI describes a published resource resulting from an activity.

Sometimes, the final product is the result of activities, often operated by different institutes. For instance a dataset attached to an article has been enriched with added values. A compilation derived from several original resources, etc.

Each of those activities can be published with a DOI and has its own metadata. For instance, each activity has a publisher - idem, the creation and update dates deal with the resource resulting from the activity, they are distinct to the date of the original data.

Compilations product is a well representative use-case of metadata consistency activity. The author of a compiled resource uses original resources, but by cross-referencing them, the author modifies the scope of the new resource and adds its own intellectual touch. In this case, the creator of the data is the author of the compilation and not the authors of the original resources. This use-case could be applied to any derived resource that modified the content. For those, the provenance information and the citation of the original resource is managed by relations (see 4.2)

What about metadata inheritance? Rather than inheritance, it is often the provenance which has to be specified.

Inheriting metadata from the original product to the derived product can hinder understanding of the product's activity. However, we observe some exceptions:

- License is a DOI metadata which can be subject to inheritance due to the contaminants property of some licenses.
- Creators of resources resulting from an activity that doesn't modify the scientific content.
Authors of the original resource remain the intellectual owners of the (inherited and curated) resource. Curator(s) responsible for the added value could be added as a contributor (see [Creators and contributors](#)). In any case, the relation to the original resource should be assigned (see 4.2)
- Resource copied (isCopyOf term) keep the original metadata.

3.2 Assigning resource type

DataCite provides two items to manage the resource type.

The first is [DataCite:ResourceTypeGeneral](#) which is **mandatory** in the DataCite schema. It consists of a controlled vocabulary that allows the term to be machine-readable. The "ResourceTypeGeneral" is used to make

categories. The second is **DataCite:ResourceType** and it completes "ResourceTypeGeneral" with free text. "ResourceType" is an option which is not generally used by citation extractors (see [Citation extraction](#)). It remains, however, a common way to accurately determine the resource type.

The choice of the "ResourceType" (see also definition in [Resources treated in the document](#)) affects the understanding of the resource. For instance if you want to publish a DOI for your Knowledge base, you can expose its content (so the database) with the type **DataCite:ResourceTypeGeneral** "Dataset", or the Service or organization that maintains it with the type **DataCite:ResourceTypeGeneralType** "Service".

A DataSet should not be confused with a Collection. A collection is an entity which doesn't point to the data directly, but gathers resources described elsewhere having each their own PID. The collection provides a list of relations to resources and has metadata that applies to all of these external resources. The Dataset contains one or multiple resources under the same DOI. For instance, an author who publishes a set of Documents made of spectrums under the same DOI is a **DataCite:ResourceTypeGeneral** "DataSet".

"ResourceType" is a description of everything that is associated with the DOI, taken as a whole. DataCite provides the term "formats" to describe the files comprising the resource. The term can list multiple specific MIME types for users who want to know whether they will be able to do anything with what they download.

A non exhaustive proposal or ResourceType

	ResourceTypeGeneral	(suggestion) ResourceType
Table or set of tables	Dataset	observation datasets compilation evolved datasets See "product-type" IVOA vocabulary
Resource such as a table available in VO	Dataset	IVOA resource ObsCore table EpnCore table
Individual image	Image	Image 2D, Image 3D Sky survey
Collection of images	Dataset	Images, curated images, ...
Collection of Spectra	DataSet	Spectra
Logs of observation	Dataset	Evolving Datasets
Collection of resources having their own PID	Collection	(specify the resource type) Collection of images Collection of observations ...
Set/collection of resources	DataSet	
VO service	Service	IVOA service IVOA TAP service
Service Web	Service	Web service
Incremental dataset (HLSP)	Dataset	Evolving Dataset
Instruments	Instrument	(*) see also PID for instrument Howto

Use web semantic for ResourceType Even if the term is free, we encourage the usage of external vocabulary (rdf-semantic) such as

- "product-type" IVOA vocabulary
- "Content-type" IVOA vocabulary

3.3 Metadata list

The DataCite schema groups items in 20 categories. All items have impacts (discoverability, reusability, citation, provenance, etc.) and all have to be considered by the DOI producer.

In this section we focus on items highly used for reusability and citation.

3.3.1 Assigning the publisher

Publisher is **mandatory** in DataCite schema and consumed for provenance and citation. The publisher is the organization responsible for hosting, and

distributing the resource. Organization has to be named and when it exists, it can also be identified with a persistent identifier such a RoR.

Example: Datacite extract (JSON)

- simple publisher in free text

```
"publisher": { "name": "Centre de Donnees Strasbourg (CDS)" }
```

- publisher with RoR:

```
"publisher": {  
    "name": "Observatoire astronomique de Strasbourg",  
    "schemeUri": "https://ror.org",  
    "publisherIdentifier": "https://ror.org/04xsj2p07",  
    "publisherIdentifierScheme": "ROR" }
```

3.3.2 Title and description

Titles is a DataCite **Mandatory** metadata, it describes the resource and is exploited by search engines such as ADS (see **Scix curation requirements**) or **EOSC**.

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 of a title used in an article and its additional dataset:

Reference article: ApJ (Draper Z.H, 2000), "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"

How to feed titles ? We give a non exhaustive list of thought to be considered to create a useful title:

- Object or type studied
- Facility used
- Release version
- Measurement method (spectroscopy, photometry)
- etc.

Description completes the title. 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 controlled vocabulary for description qualifiers (Abstract, Methods, ...).

Description for data derived from an 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 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.

When the data is derived or completes an existing publication, It can be useful to specify the Origin in the text.

Example: VizieR DOI description example

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

3.3.3 Keyword is better with Web semantics

We encourage the usage of recognized keywords such as UAT ([Unified Astronomy Thesaurus](#)), IVOA-UAT, FOS ([Field of science](#)) or any keywords driven with a Web semantic.

```
"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"  
  }  
]
```

3.3.4 Dates

Dates are Recommended by DataCite. Date is not mandatory, but it is an important metadata to generate BibTeX (see 3.4.1).

A DOI can have multiple dates related to the different steps in a publishing workflow. Each date is qualified by a controlled vocabulary ([DataCite:dateType](#)). For instance "Created" and "Updated" are the most common dates.

Examples of dates, sorted according to their role in a workflow, and applies to the same DOI resource

1. **Submitted:** the date when the data were submitted by the data producer in a DataCenter
2. **Created:** the date when the data was created in the DataCenter. This date is particularly interesting for BibTeX generation
3. **Issued:** the Date when the data has been published the first time
4. **Updated:** the last modification (concern modification on the data). The date is important for evolving Datasets.

Some resources, such as Services, Knowledgebase or Collection are by nature resources that may evolve. The frequency of updates may make it inappropriate to use the update date. For changes in architecture or data model, DataCite:versions should be used.

5. **Withdrawn:** is a particular status that has to be specified when a resource is removed data (see [Tombstones](#) section)

Again, the Date(s) involve(s) the publisher activity only (see [Metadata consistency](#))

3.3.5 Creators and contributors

DataCite makes a difference between creator and contributor.

- The creator list is **mandatory** in the DataCite schema. A creator is any person (or organization) directly related to the creation of a resource. Creators mean to be credited in citation (they are generally in citations such as APA or BibTeX).

- The contributor list is optional. Contributors complete creators with any person (or organization) who took a part in the resource creation. The contributor is qualified with a controlled vocabulary in the DataCite schema (eg: **DataCite:contributorType** "ContactPerson", "ProjectLeader", "Sponsor", etc.). Contributors are generally not in citations (such as those generated by ADS or DataCite)

"Creator" is the Dublin Core generalization of the "author" concept, it is a fundamental metadata which allows to credit the authors. The creator list is used by citation generators who extract information from DOI (see **Citation extraction**). The good usage consists of adding all authors (and not only the first author) and to provide ORCID and affiliations each time they are known.

Sometimes, maintaining a creator list becomes difficult and specifying a full organization is privileged. It is particularly important when a lot of resources are generated by an institute. For instance Chandra pipeline specifies the full organization for their DOI observations.

Example: <https://api.datacite.org/dois?query=10.25574/29770>

Again, creators as well as contributors must be involved in the activity that generates the resource (see **Metadata consistency**). The derived resources for which the scientific content has been modified don't copy generally creators of the original resource.

For resource coming from a third party, it is advisable to:

1. the DOI producer has to think about "Who took part in the activity of the resource that I publish? (creators/contributors) "
2. The DOI producer should contact the third party: "Who has to be credited (citation) in the resource that I publish? (creators)"

Examples:

- A service like VizieR, which adds value to the data published in articles by making the data VO-compliant or by adding any other information without modifying the original content, keeps the list of authors..
- A compilation product using external resources does not keep the original authors
- The sustainability of a Knowledgebase that has evolved data and a mobile team, will put the organization only.

- Example of Resource (PDS) using creator and contributors:
"Lucy Multispectral Visible Imaging Camera (MVIC) Dinkinesh Calibrated Data Collection" <https://api.datacite.org/dois/10.26007/f61e-0f52>
- An heterogeneous collection of datasets having each their own DOI and their own creators specifies only the publisher organization

Who is a creator, a contributor? Datacite provides a controlled vocabulary for contributors ([detailed in the DataCite schema](#)) that allows to include persons or organizations that took part in the creation or the publication of a resource. For instance, an experience that required funding can add the Funder organisation in the contributor list.

Making the difference between creator/contributor has been debated (see url in the end of the section) because it is often difficult to classify the role of each one (for instance, what about persons in charge of the Data such as Data curator or Data manager?)

To help the choice, we propose 2 criteria:

1. The creator takes a substantial part in the data creation (imagine, observe, calibrate, analyse, etc.)
2. The creator approves the decision to publish the final result

Defining creator/contributor has been debated:

<https://library.unt.edu/metadata/creator-contributor-definitions.html>

<https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>

Example: extract of DataCite (JSON)

```
"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"
            }
        ]
    }
]

```

Examples:

- DOI having Data curator (ESO):
<https://doi.eso.org/10.18727/archive/65>
- DOI with diverse contributors (NSF):
<https://api.datacite.org/dois/application/vnd.datacite.datacite+json/10.26024/g8p7-wy42>

3.3.6 Licenses

The "Rights" in DataCite is optional but it is a pillar for the reusability of the resources (see [FAIR principles](#))

Many licenses are used today for data or softwares. The most popular are listed in [SPDX](#). SPDX includes [Creative Commons licenses](#) which are often used for Datasets.

The choice of license is often guided or imposed by the governance (institute, country law) or may be subject to a legacy license in case of derived product. The DOI creator must be aware of the license imposed by the publisher's governance. For instance:

- NASA encourages CC0 license (see [SPD-41](#))
- French government imposes LO/OL or CC-by licenses (see [Etalab](#))

To make a license interoperable means to be machine readable. It is advisable to use standard license acronyms and add the URL definition. The Creative Common and SPDX websites offer such URLs and acronyms.

Example: CC-by-4.0 URL

- SPDX: <https://spdx.org/licenses/CC-BY-4.0.html>
- Creative Common: <https://creativecommons.org/licenses/by/4.0/>

Example: extract of DataCite (JSON)

```
"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"
  }
]
```

3.3.7 Fundings

Financial support can be attached to a DOI using [DataCite:fundingReference](#). Funds information is expressed with a reference to the organism, generally with an organism identifier (eg.: ISNI, Grid, Cross Funder, ROR) and an award reference (see DataCite guidance)

Examples:

- DOI generated by NASA Planetary Data System:
<https://api.datacite.org/dois/application/vnd.datacite.datacite+xml/10.26007/97r3-1e19>
- DOI generated by Chandra:
<https://api.datacite.org/dois?query=10.25574/29770>
- Rubin Example: (using RoR)
<https://commons.datacite.org/doi.org/10.71929/rubin/2583847>

Fundings is a metadata consumed by scientific platforms such as EOSC. They are used by research funding agencies who wish to establish usage statistics related to their investment.

3.4 Citation extraction

Now that the DOIs have been properly curated, we will be able to extract metadata to make citations and to credit resources.

DOI is well suitable to link data in articles. It can be inserted everywhere (in footnote, in the text or in acknowledgment) with hyperlinks using the DOI or hidden under a human readable text using metadata such as title, authors or content type.

Citations in an article are the references list which credits the authors. A citation is not a footnote URL, an acknowledgment or any DOI string in the text. It ONLY refers to listing the reference in a reference list of an article. Citations use formats such as BibTeX or APA which include the basic DOI metadata. Many indexing services harvesting DOIs (from DataCite, Cross-Ref or other) extract metadata that they then format into web pages with sometimes in addition a citation service for the resources.

For instance, DataCite provides a citation output in BibTeX format:

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

Some agencies make statistics on citation extracted from articles which include references such as articles or datasets. DataCite provides its own public statistics on any resources which have been cited in DOI metadata. Each time a DOI is generated with a "Cites" relation (see linked data in section 4.3) the counter is incremented. The statistics are available through the DataCite API.

Note: Publishers that provide DOI for resources deriving from other products should consider how to cite the original data. Citing is possible in the DataCite schema that provides relationships to cite resources (see linked data in section 4.3).

3.4.1 Curate DOI is mandatory to extract citation

The citation that can be extracted from a DOI is only useful if the DOI's metadata is properly curated with the quality fixed by journals, ADS or Data publishers.

In this section, we summarize the metadata expected to obtain a clean citation in an annotated BibTeX Format. The BibTeX is based on ADS.

BibTeX template using DOI metadata

Note: the BibTeX type "@dataset" which is specific to ADS and is not used by services such as DataCite which uses the generic term "@misc".

```
@dataset { { localref },
    author = { authors },
    title = "{ title }",
    year = { year },
    month = { month },
    eid = { usualName },
    url = { url },
    keywords = { keywords },
    publisher = { publisherName },
    copyright = { rights },
```

DOI = {DOI}
}

BibTeX	DataCite	Relevance for citation
authors	Authors	MUST, see Creators and contributors
title	Title	MUST, see Title and description
year	Date:creation	MUST, see Dates
month		
eid	alternateIdentifier	MAY, see Linked data in section 4.2
url		SHOULD, use DOI URL
keywords	Subjects	SHOULD, see Keyword is better with Web semantics
publisher	Publisher	MUST, see Assigning the publisher
copyright	Rights	RECOMMENDED, see Licenses
DOI		MUST
resource_type	resourceTypeGeneral	MUST, see Assigning resource type

Template Citation in APA style

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

VizieR example in APA style: (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>

3.4.2 Using DOI for Evolving DataSet

There are different approaches for [evolving Datasets](#) citation which depends on the Data Center implementation.

- Cite a snapshot of the Dataset.

In this approach, the Data Center makes snapshots and adopts a DOI versioning mechanism.

- The Data center provides a template for citation.

- Cite a DataSet extraction resulting from a query.

The solution may become complex if we take or not the reproducibility aspect. A simple solution without reproducibility constraint consists, in an article, to cite the evolving Dataset and to specify the query outside the citation, somewhere in the text.

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

3.4.3 Citing Collections

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.

The usage of Collections can be ambiguous for article editions. We report two usages, without wanting to endorse either.

- 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 & 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.

4 DOIs across networks

4.1 Dissemination of resources through complementary networks

Open data emerged in the academic landscape through online publication via various networks and web services. Some are historical , developed by institutes such as ADS, while others have been added more recently such as FairSharing, <https://www.academia.edu/>, or the European Open Science Cloud (EOSC).

A multitude of standards have been created, covering roles such as identification (DOI, RoR, ORCID, ISSN, ...) or data modeling. Each Network implements one or more standards (e.g., the IVOA includes DOI, ORCID in VOResource) depending on their objective. Even if a common basis can be extracted, which allows to map serialization together, these are complementary and specific (see Connecting DOI and the IVOA).

Datasets distributed in diverse frameworks are harvested by platforms that combine the different sources. 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.

The list of metadata tends to increase, but the most popular are DataCite schema, Dublin Core, the registry of the Virtual Observatory (Dower and Demleitner et al., 2014), schema.org (extends Dublin Core and is used by Google), DCAT (linked catalogues, datasets and services. DCAT is a concurrent of the VO registry), OpenCitation (Peroni and Shotton, 2020) (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 specificity of the metadata have been discussed in IVOA DCP session (see the presentation of H.Enke, Interop 2023, Bologna). In practice, it is better for implementers to think from the beginning about the different output in order to provide a consistent workflow. Note DataCite provides several serializations of the metadata, in particular "schema.org".

Note Maintain broadcast workflows together.

4.2 Linked data with internal and external resources

Open data provides reusable and interoperable data. The term interoperability taken in a broad context, describes an interconnected network made of a multitude of documents shared through specific sub networks. The DOI metadata allows linking documents together in order to improve the provenance with additional documents, to cite their origin or to link versioned resources.

Linked data are currently exploited by OpenCitation that makes citation graphs, by EOSC for discoverability by providing links in the results of search queries. This capability interests ADS who could integrate the information in the future.

Relationships allow to make linked data in the DataCite schema; they have been the subject of a supplementary section in the DataCite guidance. RelatedIdentifier is an option in DataCite that recommends adding at least one relation (for instance to link a dataset with a reference article).

4.3 A basic tour of relationship motivating by their usage

In this document, we focus on a short list of relationType responding to the most important use case and dedicated to Provenance, Citation and Versioning.

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.

4.3.1 Express a Relationship

DataCite schema allows to build relationships, including the nature of the relation, with the term **DataCite:RelateIdentifier**.

A relationship can be summarized in a RDF triple :

<current resource> <predicate> <external resource>

Example: linking a dataset with its article

<<https://doi.org/10.26093/cds/vizier.22640008>>
isSupplementTo <<https://doi.org/10.3847/1538-4365/ac9af8>>

The predicate is a term taken into a controlled **relationship vocabulary** which expresses the nature of those relationships (more details in the next sections).

The identifiers must be unique and, preferably persistent (PID) and resolvable. The identifier type must be specified from a controlled vocabulary using the term **DataCite:relatedIdentifierType**. The list includes common identifiers (eg: DOI, bibcode, ISSN), URL or resolvable identifier such as PURL or W3id³.

Choosing the identifier is important and depends on their usage in the Open Data landscape. For instance, the bibcode which is available in Datacite schema but not in OpenAire⁴ have a limited use for each service that uses the OpenAire process.

Note To facilitate the harvesting/merging jobs, the DOI has to be privileged than any other identifier or URI.

³https://guidelines.openaire.eu/en/latest/data/field_identifier.html

⁴https://guidelines.openaire.eu/en/latest/data/field_identifier.html

4.3.2 Term to link resources with their origin

It is commonly used to link a dataset with an article or with its original dataset. These references (in bold) are currently exploited by EOSC.

- **Cites**

If you applied a specific, published technique to generate your data set, reference the article that describes that technique with a referenceType of "Cites".

Authors motivated by credits can query DataCite that make (public) statistics with the term "Cites" . Statistics are incremented each time a DOI is created with the "Cites" term. However, these statistics are independent of the statistics generated by classical indexing services such as ADS.

- **isCitedBy**

Is the "Cites" reciprocal term that could be used, for instance; when the dataset is cited by the article. 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 "isCitedBy" requires a workflow synchronisation that is not always easy to implement.

- **IsSupplementTo**

if the dataset is supplementary (e.g., it contains the data used to create a plot or table accompanying the article), use a relationType of "IsSupplementTo"

- **IsDerivedFrom**

If your data were derived from a previous dataset, include the DOI of that data set with a referenceType of "IsDerivedFrom". IsDerivedFrom has to be used for any data that has been made from an original resource. DataCite accepts the case that a dataset is derived from a larger dataset or dataset with values that have been manipulated from their original state.

- **IsVariantFormOf**

If your data is a copy of an original Dataset, the term is adapted for you. In a more general case, the term includes a dataset in which the original content has not been modified. It includes a dataset with added values or with any "decoration" that does not modify the scientific content.

- **IsCollectedBy**

Specify an instrument used to collect the dataset.

Other terms

- What about "IsIdenticalTo" ?
IsIdenticalTo means datasets byte-by-byte equivalent and disapproved of any additional data or fantasy. The intellectual or administrative property of the document is also not a sufficient reason to use this term.
Frequently, even copies are variants, **so privilege other terms such as isVariantFormOf**.
- "**References**" may be used to link another resource which is particularly important for the understanding of the data. For instance an instrument calibration document which was used to perform observation.

Examples:

- Linking Data with article in VizieR:
(extract from <https://doi.org/10.26093/cds/vizier.22640008>).

```
< relatedIdentifiers >
  relatedIdentifier relatedIdentifierType="bibcode" relationType="IsSupplementTo">
    2023ApJS..264.8H
  </ relatedIdentifier >
</ relatedIdentifiers >
```
- Linking Dataset with original resource
(extract from <https://doi.org/10.26093/cds/vizier.1355>).

```
< relatedIdentifiers >
  < relatedIdentifier relatedIdentifierType="DOI" relationType="IsVariantFormOf">
    10.5270/esa-qa4lep3
  </ relatedIdentifier >
</ relatedIdentifiers >
```

See also the **GREI Metadata Recommendations** from DataCite schema version 4.6

4.3.3 Dealing with Collection

The following terms allow to report the hierarchy in an archive. They should be emphasized with type Collection.

- **isPartOf**
Any resource which is a part of a collection. For instance a Document resulting from a set of observations in a mission context.

- **hasPart**

Collections having a reasonable list of children can use this term.

Example

- Collection in Chandra archive:

Dataset observation: <https://doi.org/10.25574/23336>

Which isPartOf the chandra Collection: https://doi.org/10.25574/csc2.stk.acisfj2008299p002118_001 (defined as ResourceTypeGeneral=Dataset)

- Data releases from the NSF-DOE Vera C. Rubin Observatory also defines a parent DOI for the entire data release and then link each distinct dataset type with hasPart/isPartOf relationships. This approach is described in <https://doi.org/10.71929/rubin/2583847>.
- The deep coadd images <https://doi.org/10.71929/RUBIN/2570313> is-PartOf Rubin LSST Data Preview 1: <https://doi.org/10.71929/rubin/2570308>

4.3.4 Managing version and obsolescence

See also the section [Versioning](#).

- **IsNewVersionOf / IsPreviousVersionOf**

The terms allow linking different versions of a resource together as a graph view. It can be used without using a Collection hierarchy.

- **HasVersion**

This term is used by Zenodo which creates a Collection that includes all versions (see

- **ObsoletedBy / Obsoletes**

When a dataset is obsoleted by a new product. Obsolescence is neither versioning, nor reflecting that a resource disappeared (see tombstone page).

The state of obsolescence of a resource always relates to an object (cd may be a method, software that no longer adapts to a new format, data cannot be used in a recent context, etc.)

The obsolete stuff continues to be accessible, but can't be used for current or future operations.

4.4 Connecting DOI and the IVOA

4.4.1 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 ([Plante and Demleitner et al., 2018](#)) for any "ResourceName" such as alternate identifier or related resources. They are also used in the DataOrigin([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.

4.4.2 DOI and IVOA networks complete each other

The IVOID is a unique identifier for a resource (dataset or service) in the Virtual Observatory Registry. Registry record 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 resource and what protocol can I use?). When they represent datasets, the services to access the data are described in detail. 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 is a persistent identifier for resource(s) that can be used for citation. The DOI is visible for users and it handles a landing page human-readable. The metadata include attribution, references, relations, license, but nothing about programmatic access or interoperability.

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

DOI and IVOA registry have their own scope. IVOID is used by the Virtual Observatory to query data whereas the DOI is widely used in the science context in literature, scientific search engines (ADS, Eudat, scholar, researchgate, etc), or platforms such as EOSC. In order to connect the networks, we propose a solution, illustrated by usage scenarios, based on the concepts of 'alternative identifiers' existing in both environments.

4.4.3 How and why to connect DOI and IVOA networks?

In this section we discuss scenarios that motivate to make the bridge between IVOA and DOI networks.

The proposed solution, for each use-case, is based on alternative identifiers existing in each schemas.

- A user finds a relevant dataset (from EOSC or any other platform). The DataCite metadata indicates an IVOID as an alternate identifier. Then information is used by the user who queries the registry included in VO tools or API such as pyvo. Finally, he/she finds the IVOA description that allows to extract and manipulate data.

Proposal: Specify the ivoid as AlternateIdentifier in DOI metadata

Example of an alternate identifier in a DataCite record using IVOID.

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

- The user finds an interesting dataset using VO tools, thanks to the metadata in the registry, the user finds the bibliographic links (DOI).

Proposal: Specify the DOI in the VO registry record

Example of DOI in a registry record:

```
< altIdentifier >doi:10.26093/cds/vizier.73200451</altIdentifier>
```

- A user finds a relevant article in ADS/EOSC and he/she would like to query the datasets related to the articles. (today, you can use the VO registry and make a search based on criteria such as authors, title description, etc). ADS or EOSC uses the relations in the Datasets to link the article with the dataset, then the alternate identifier specifies if the data are registered in VO.

Implementing DataOrigin in VOTable The DataOrigin note (Landais and Muench et al., 2024) defines a framework to describe the data that were used to produce a resultset. It includes the IVOID and DOI identifiers.

4.5 DOI consuming in ADS/SciX

The Science Explorer (SciX) (Bartlett and Polimera et al., 2025) is a digital library portal for researchers in Astronomy, Earth science, Heliophysics,

Physics, and Planetary Science. SciX is a multidisciplinary library based on the successful [Astrophysics Data System](#) (ADS) that has served astronomers and astrophysicists for more than a quarter century. While currently still represented by two websites, soon the ADS will [transition](#) to SciX. The system provides links to articles and associated data with a base of information such as authors, dates, journal, titles, descriptions , keywords. The project curates the metadata (such as author identification with ORCID) that allows it to provide clean citations in different formats. This citation, commonly used by authors that need to cite their sources, allows the ADS to generate statistics on the indexed resources.

4.5.1 SciX curation requirements

SciX works with Editors, Publishers and Organizations to harvest resources. DOIs, created in Crossref or in DataCite, are used in input (for harvesting) and in output (for citation). One of the biggest challenges ADS encounters falls basically in three categories: bad metadata, incomplete metadata and inconsistent metadata.

- Bad metadata occurs when the metadata supplied does not follow the metadata model.
It includes both, misunderstanding of the schema (for instance incorrect resourceType that happens in particular with Collections) and uncurated data (for instance when the information does not match the schema definition. Eg: eg: <creator>Henneken et al. (2024)</creator>).
- Incomplete metadata means less useful records (or meaningless records in the worst case). The [SciX blog](#) on data indexed in SciX (to capture citations) specifies a minimal set of metadata desired for indexing data
- Inconsistent metadata occurs when the registered metadata differs from what is listed on the landing page (like e.g. a different list of creators). These differences in metadata impact the quality of (BibTeX) citations of data which becomes dependent on the place where they are generated or harvested.

4.5.2 Meta-data harvesting.

SciX has consumed Crossref metadata for decades, allowing a highly complete service of published articles in the community. Now, SciX has started to also harvest DataCite metadata for indexing and linking records. However, the practices used by many disciplinary archives fail to provide consistent rich metadata across all networks.

The challenge consists in building an information system with metadata coming from different networks which use each of their schemas. For instance the DOI schema and schema.org which is sometimes used in landing pages. The effort is hard with inconsistent metadata. SciX therefore expects that data centers adopt workflows that provide consistent metadata (for instance with extracting the schema.org from the DataCite metadata)

In December 2025, relations are not yet extracted from the DOI metadata by SciX. However, extracting this information is on the SciX roadmap to be implemented in 2026.

5 Appendix

5.1 Recommendation checklist

1. Take in consideration that creating a DOI is forever ([DOI Lifecycle](#)) and implies ownership and maintenance for the long term ([Considering a DOI Solution](#))
2. Target the resources and choose the adequate granularity. The granularity depends on parameters such as citation, discoverability, traceability.
3. Map the metadata with the DOI schema (the schema depends on the RA). Take a particular attention with:
 - The choice of the resources ([Assigning resource type](#))
 - List all authors with ORCID and affiliation when they are known ([Creators and contributors](#))
 - Add machine readable license ([Licenses](#))
 - Make metadata consistent with the activity (dates, title, ... must describe your activity) ([Metadata consistency](#))
4. Make landing page
 - 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
5. Check if metadata allows to generate citations ([Citation extraction](#)).
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 ([A basic tour of relationship motivating by their usage](#))
 - Do not misinterpret the meaning of a semantic. The relations semantics are precise. The IVOA DCP working group can also help you.
 - If existing, make the links between resources that you provide (eg: collection, ...)
 - 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 ([How and why to connect DOI and IVOA networks?](#))
7. Ensure the maintenance of the DOI
Ensure that all your dissemination workflows are consistent, in particular, with IVOA registry and with "schema.org"
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) ([Versioning](#))
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 ([Tombstones](#))

5.2 Building landing pages

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

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, DCAT,, 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.)

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.

Generate a machine-readable landing page Search engines harvest landing pages 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.

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

- FAIRchecker
- Google Search central

5.3 Example of existing DOI

5.3.1 Chandra Data Archive

The CXC created DOIs for each single Chandra observation (D’Abrusco and DPAsOps Team, 2023), 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”.CDA is doing cool things with related identifiers.

See (D’Abrusco and DPAsOps Team, 2023), (Rots and D’Abrusco et al., 2018).

Example : <https://api.datacite.org/dois/10.25574/29770>

5.3.2 VAMDC & Zenodo workflow

The good and the bad of transitory data citation.

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 (Moreau and Zwolf et al., 2018).

Example: VAMDC query Store record:

<https://cite.vamdc.eu/references.html?uuid=a819e879-760d-4ede-831a-38d359b5864c>

Landing page: <https://zenodo.org/records/11243136>

5.3.3 ESA

add ESA ?

Example: Landing page: <https://doi.org/10.5270/esa-qa4lep3>

(JSON) DOI: <https://api.datacite.org/dois/10.5270/esa-qa4lep3>

5.3.4 VizieR

VizieR ([Ochsenbein, 1996](#)) 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>

5.3.5 PDS

The Planetary Data System (PDS) is a long-term archive of digital data products returned from NASA's planetary missions, and from other kinds of flight and ground-based data acquisitions, including laboratory experiments. But it is more than just a facility - the archive is actively managed by planetary scientists to help ensure its usefulness and usability by the world wide planetary science community.

Examples:

- (Dataset in NASA PDS: Small Bodies Node) "Lucy MVIC Dinkinesh Calibrated Data Collection" <https://api.datacite.org/dois/10.26007/f61e-0f52>
- (Supplement to article in NASA PDS: Small Bodies Node) "Taxonomy, colors, and slope parameters for asteroids from the Dark Energy Survey Bundle" <https://api.datacite.org/dois/10.26033/m95p-bn08>

5.3.6 NSF-DOE Vera C Rubin Observatory

Rubin Observatory ([NSF-DOE Vera C. Rubin Observatory, 2025](#)) is partially funded by the US Department of Energy and that allows for DOIs to be published through the [DOE Office of Scientific and Technical Information \(OSTI\)](#). Rubin has published instrument DOIs for the three instruments and as described in [DMTN-318](#) we publish collections of DOIs for every formal data release. Additionally, a subset of our technical documentation is also published. Each DOI entry is harvested by ADS; a process made easier by the use of a "rubin" infix in each DOI to make them easy to locate in DataCite.

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>.
- Bartlett, J., Polimera, M., Lockhart, K., Accomazzi, A., Kurtz, M. and Science Explorer Team (2025), ADS and SciX: Pioneering the Next Generation of Interdisciplinary Research Discovery, in 'American Astronomical Society Meeting Abstracts #245', Vol. 245 of *American Astronomical Society Meeting Abstracts*, p. 442.04. <https://ui.adsabs.harvard.edu/abs/2025AAS...24544204B>.
- D'Abrusco, R. and DPAOps Team (2023), 'A New Way to Cite Chandra Data', *Chandra News* **34**, 5. <https://ui.adsabs.harvard.edu/abs/2023ChNew..34....5D>.
- DataCite Metadata Working Group (2024), 'Datacite metadata schema documentation for the publication and citation of research data and other research outputs v4.6'. doi:[10.14454/MZV1-5B55](#), <https://datacite-metadata-schema.readthedocs.io/en/4.6/>.
- 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/>.
- Hashemi, M. (2016), ‘Calendar Versioning — CalVer’. <https://calver.org/>.
- 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 Duvernay, 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>.
- 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>.
- NSF-DOE Vera C. Rubin Observatory (2025), ‘Legacy Survey of Space and Time Data Preview 1 [Data set]’. doi:10.71929/RUBIN/2570308, <https://www.osti.gov//servlets/purl/2570308>.
- Ochsenbein, F. (1996), ‘The vizier database of astronomical catalogues’. doi:10.26093/CDS/VIZIER, <https://vizier.cds.unistra.fr>.
- Peroni, S. and Shotton, D. (2020), ‘Opencitations, an infrastructure organization for open scholarship’, *Quantitative Science Studies* **1**(1), 428–444. doi:10.1162/qss_a_00023, http://dx.doi.org/10.1162/qss_a_00023.
- Plante, R., Demleitner, M., Benson, K., Graham, M., Greene, G., Harrison, P., Lemson, G., Linde, T. and Rixon, G. (2018), ‘VOResource: an XML Encoding Schema for Resource Metadata Version 1.1’, IVOA Recommendation 25 June 2018. doi:10.5479/ADS/bib/2018ivoa.spec.0625P, <https://ui.adsabs.harvard.edu/abs/2018ivoa.spec.0625P>.
- Preston-Werner, T. (2023), ‘Semantic Versioning 2.0.0’. <https://semver.org/>.

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>.