

CESSDA ERIC Persistent Identifier Policy Best Practice Guidelines

22 November 2017

Version 1.0

<https://doi.org/10.18448/16.0041>



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

TABLE OF CONTENTS

1	INTRODUCTION.....	3
2	BACKGROUNDS.....	3
3	PRINCIPLES.....	3
	Principle 1: Identifying.....	4
	Principle 2: Locating.....	5
	Principle 3: Resolving.....	6
	Principle 4: Referencing and Citation.....	7
	Principle 5: Visibility.....	8
	Principle 6: Flexibility.....	9
	Appendix A: Definitions and Acronyms.....	10
	Appendix B:.....	11
	PID - What are the benefits for the CESSDA community?.....	12
	Review of the PID Services provided by GESIS, DANS and SND.....	14

1 INTRODUCTION

1.1 Purpose of the document

This document contains guidelines for the implementation of the CESSDA Persistent Identifier Policy Principles. These guidelines address all CESSDA Service Providers (SP). This information is aimed at supporting CESSDA SP to implement the use of global Persistent Identifiers (PID) and to assess the consequences for their organisation.

We have chosen for short descriptions in the document to keep it readable. Suggestions for further reading and specialised documentation are given in the appendix.

1.2. Related Document

CESSDA ERIC Persistent Identifier Policy. Version 1.0. 22 November 2017. <https://doi.org/10.18448/16.0040>

2 BACKGROUND

As social sciences tend to be more and more data intensive, data repositories must facilitate several ways of identifying and locating data. This development poses complex technical and organisational challenges to data providers. Persistent identification is becoming a prerequisite for sustained and reliable resource discovery and reuse. The use of PID is an important feature of a certified and trustworthy data repository. PID support access to data as well as referencing and citing data. They are an advertisement for data integrity – ultimately PID are part of the proof that an object which a repository has responsibility has not changed. Additionally, the use of PID helps data repositories to be compliant with the FAIR principles (Findable - Accessible - Interoperable - Reusable)¹ set by FORCE 11² and provides a future-proof plan in case of the relocation of its holdings.³

The main task of CESSDA and its Service Providers is to provide their designated communities with well documented, verifiable, and understandable data for research. One way of doing this is to assign a PID to the data collection (and if desired even to other data collection related objects) and their later versions. The PID will accompany the specific version of the data collection: it is assigned inside the repository and displayed (cited) outside when the data collection is reused. This allows for keeping track of which dataset version is disseminated and gives the user a simple way of both citing the data creator and displaying which exact version was used.

To achieve this goal the assigned PID must be unique on a global scale and the PID service provider must be a trusted organisation with a clear policy on the long-term support of the service and a sustainable business model. The PID assigner (a CESSDA SP) must provide landing pages for each assigned PID with information about how to access the data, licensing rules, different versions and provenance. Some PID services make use of additional metadata alongside their PID which can contain information about related material such as publications or other data collections. If additional metadata is used, this should be made available on the landing page as well.

3 PRINCIPLES

The following pages contain a general description as well as recommendations and examples of the six CESSDA Persistent Identifiers Policy Principles.

¹ FAIR Data Publishing Group: <https://www.force11.org/group/fairgroup>

² Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>

³ See: Appendix B to this document: CESSDA PID Task Force: PID - What are the benefits for CESSDA community (users and providers of services)? Cologne, 26 September 2016.

Principle 1: Identifying

Each CESSDA SP shall use globally unique and persistent identifiers to identify their data holdings, which are of interest to CESSDA.

General information

A globally unique and persistent PID is a worldwide unique and persistent archival reference code. It is an identifier that is permanently assigned to an object and unique on a global scale. Persistence is a commitment by the issuing PID service provider ensuring that the system is administered comprehensively.

Attaching a unique identifier to a data collection helps the data provider to guarantee the origin, content and version of the data collection. This can be verified by resolving the PID. In order to work on a global scale the service must run 24/7 and must be supported and governed by a dedicated user community (see also principle 3).

The main PID systems for data collections are: Handle, DOI[®], URN:NBN and ARK.

DOI[®] and Handle are based on the same (Handle) system; URN:NBN on a network of services from National Libraries and ARK is a service developed by the California Digital Library (CDL).

DOI[®] and Handle are the ones most commonly used.

In terms of identifying, by using PID, CESSDA SP will be able to

- identify their holdings persistently on the Internet so that these holdings can be referenced
- keep track of disseminated data and their versions
- provide a simple way of citing and referencing both data collection and data creator
- add a building block towards the status of a Trusted Digital Repository⁴.
- be compliant with the CESSDA Data Access Policy⁵.

Recommendations

- CESSDA SP should assign globally unique PID to data that are published.
- CESSDA SP should commit to maintaining the PID permanently.⁶
- The assigned PID must always resolve to the same object.
- The PID should preferably be embedded within the data file (e.g. as a variable).
- Which PID service a CESSDA SP may use depends on various reasons (e.g. harmonisation with other national or already existing services etc.).
- When choosing a PID system and a PID service provider⁷, credibility and long-term viability of the PID system as well as the use of open standards should be considered. CESSDA SP should use PID service providers with clear and transparent policy and business models.

Examples

PID	Example PID name
Handle	hdl://hdl.handle.net/11022/0000-0000-0000-C
DOI [®]	https://doi.org/10.5878/002645
URN:NBN	urn:nbn:de:bvb:19-146642
ARK	http://bnf.fr/ark:/13030/tf5p30086k

⁴ For more information on Trusted Digital Repositories see: Edmunds, R; L'Hours, H.; Rickards, L.; Trilsbeek, P. and Vardigan, M. (2016): Core Trustworthy Data Repositories Requirements. <https://doi.org/10.5281/zenodo.168411>

⁵ <http://cessda.net/eng/About-us/Documents>

⁶ For further information on the provision of tombstone pages see Principle 5

⁷ See: Appendix B: CESSDA PID Task FORCE: Review of the PID Services provided by GESIS, DANS and SND. Cologne 29 August 2016.

Principle 2: Locating

All data holdings of each CESSDA SP shall be findable by their global PID via the Internet.

General information

PID are unique, alphanumeric, web compatible codes which do not only persist over time (see principle 1) but will also take the user to the identified archival object or to information about that object.

The retrievability is supported by associating PID with landing pages that are accessible via standard Internet browsers. A landing page gives essential information about the object identified by the PID, so that the user can be sure getting the object, in the exact version he/she searched for.

In terms of locating, by using PID, CESSDA SP will be able to

- make sure a data collection and its specific versions can be located

Recommendations

- A PID assigned by a CESSDA SP shall resolve to a web landing page at the website of the CESSDA SP.
- This landing page should include information about the data collection such as its origin, version, availability and accessibility. It shall indicate whether a newer version exists and preferably include information about the differences between the versions.
- The landing page shall also include information about related data collections and (if known) their PID.
- If further metadata on the data collection exists, this information shall also be provided on the landing page.

Examples

PID	Landing Page
doi:10.4232/1.0145	https://dbk.gesis.org/dbksearch/sdesc2.asp?no=0145&db=e&doi=10.4232/1.0145
doi:10.5878/002349	https://snd.gu.se/en/catalogue/study/SND0739/001/1.0
doi:10.17026/dans-x7r-395c	https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:56502

Principle 3: Resolving

CESSDA SP shall use global PID services that ensure 24/7 resolvability of PID.

General information

PID must be persistently actionable. This refers to both identifying (see Principle 1) and resolving. 'Resolving' is an automated service that translates the PID into a current web-address. This service must be available 24/7, on a global scale. This availability belongs to the responsibility of the PID service provider. The CESSDA SP, however, must ensure that the PID service provider they choose is committed to 24/7 resolvability. Most resolver services are interoperable and can resolve multiple PID systems.

The PID should resolve to a landing page (see Principle 2). Providing and maintaining the necessary metadata for the landing page and the object, as well as the link between the metadata/PID and the object is part of the responsibility of the data provider.

In terms of resolving, by using PID, CESSDA SP will be able to

- provide reliable access to their holdings
- make sure a data collection and its specific versions can be accessed irrespectively of a potential relocation of their holdings.

Recommendations

- CESSDA SP should use a trusted and well-known PID system running a 24/7 reliable global resolver.
- CESSDA SP shall use PID that resolve to a landing page on the Internet.
- CESSDA SP should guarantee resolvability of PID by maintaining up-to-date landing pages even if the data collection is not accessible any longer.

Examples

DOI[®], Handle, ARK and URN:NBN are examples of trusted global PID systems that comply with the CESSDA PID Principle 3 (Resolving).

The DOI[®] system is best known among researchers as it is the most common PID for publications. It is also rapidly becoming the global standard for data citation⁸. It has by now a sustainable and affordable business model, and a well-developed infrastructure and network of trusted service providers. The DOI[®] system is based on the Handle technology. The International DOI Foundation (IDF) operates a central Handle server to resolve the DOI names.

⁸ <http://bit.ly/2ugiTrs> (retrieved on 07 December 2017)

Principle 4: Referencing and Citation

PID need to be used to ensure referencing and citation of the data holdings of each CESSDA SP.

General information

PID facilitate referencing and citation of (digital) objects because they provide unique and robust links to digital objects (see Principles 1 to 3). They ensure that the original referenced/cited data collection can be (re)located at any time, under any circumstances, even if it has been moved.

The Data Citation Principles drawn up by FORCE¹¹ stresses the necessity of unique and persistent identifiers. The inclusion of PID in (rich) metadata is also part of the FAIR principles.

Referencing and citation require more metadata than just the PID. Apart from specific citation styles in the publishing domain within scientific communities there is usually a common standard for a minimum set of metadata, which should include the PID.

Granularity is an important aspect when implementing the use of PID. Granularity concerns the degree of aggregation of an object. Different levels of granularity can be useful depending on discipline, type of data or purpose. In some cases it might be most useful to assign the PID to a data collection, or in other cases the persistent identification of a dataset or even a variable are necessary.

When deciding which level of granularity is used the aspects of citation is crucial. What are the current citation and research practices among the data user community? What is likely to be cited?

In terms of referencing and citation, by using PID, CESSDA SP will be able to

- provide citable permanent links to data collections used in publications, for reference and impact measurements
- improve scholarly communication
- be more effective in their compliance to the FAIR principles.

Recommendations

- CESSDA SP should provide information on how to cite the data collections.
- CESSDA SP should include the PID in their recommended citation standard.
- CESSDA SP should assign PID (at least) on data collection level.

Examples

Recommended citation in the GESIS Data Catalogue: (see also: <http://bit.ly/2u9JRiu>)

de Winter, Lieven; Bryder, Tom; Linch, Philip; Norris, Pippa; Chavel, Cecile; Weßels, Bernhard; Kielhorn, R.; Marsh, Michael; Sola, Giorgio; Gardella, Louisa; Thomassen, Jacques; et.al. (1999): European Candidates Study 1994. GESIS Data Archive, Cologne. ZA3077 Data file Version 1.0.0, doi:10.4232/1.3077

Recommended citation in the UK Data Archive Catalogue: (see also: <http://bit.ly/2vdiCAP>)

University of Essex. Institute for Social and Economic Research, NatCen Social Research. (2014). *Understanding Society: Waves 1-4, 2009-2013*. [data collection]. 6th Edition. UK Data Service. SN: 6614, <http://dx.doi.org/10.5255/UKDA-SN-6614-6>

Principle 5: Visibility

PID must be included in the resource-discovery metadata provided by CESSDA SP for the Product and Service Catalogue (PaSC).

General information

The inclusion of the PID on the landing page and in the metadata describing the data collection ensures visibility of the PID itself (for easy citation and reference), as well as of the data collection it refers to, and its current status (version, duplicates, accessibility).

It also provides the possibility to create relations with other 'objects' (e.g. published articles, funder information, related data collections), thus providing context to the data collection itself.

Providing the PID in a richer metadata-context means the visibility and discoverability of identified datasets increase, e.g. because it facilitates meaningful inclusion in aggregated services no matter where the data collection is stored, maintained or moved to. This increase in visibility benefits both the individual dataset and the aggregated data collection.

In terms of visibility, by using PID, the CESSDA SP will be able to

- provide unique metadata for the CESSDA Product and Service Catalogue (PaSC).
- identify relations and duplicates of data collections

Recommendations

- CESSDA SP should provide the PID both in their resource-discovery and in their harvestable metadata.
- CESSDA SP should implement a version control and assign a new PID to each new version of a data collection.
- If a data collection is moved to a different location, the PID should be maintained.
- If a data collection is taken offline or is no longer accessible, a tombstone landing page should be maintained by the CESSDA SP, including the citation metadata and information about the data provider.
- In order to avoid the publication of duplicates of the same data collection (at different archives) in the CESSDA PaSC appropriate (technical) solutions should be developed.

Examples

Versioning:

GESIS (three-digit version number: major.minor.revision)

Data collection: German General Social Survey - ALLBUS 2010

Version 1.0.1 <https://doi.org/10.4232/1.11782>

Version 1.0.0 <https://doi.org/10.4232/1.11692>

<https://dbk.gesis.org/dbksearch/sdesc2.asp?no=4612&db=e&doi=10.4232/1.11782>

Tombstone page:

Scientific Use File German Ageing Survey (SUF DEAS) 1996, Version 2.0

<https://doi.org/10.5156/DEAS.1996.M.002>

Migrants to Stockholm 1880-1925

<https://snd.gu.se/en/catalogue/study/SNDo236> (<https://doi.org/10.5878/000143>)

Principle 6: Flexibility

The PID policy shall be reviewed at least every two years and adjusted according to the latest strategic and technological developments within and outside of CESSDA.

General information

The developments around PID are fast and cover a wide range of technical and organisational details. A lot is happening, especially in the communities around the individual PID systems. Much of it is important research and development as well as proof-of-concept work that indicate the directions in which PID may develop over time. It is important for CESSDA to remain aware of these changes and if possible and necessary adjust the CESSDA PID policy accordingly.

Within the CESSDA community, exists variation in the understanding and use of (global) PID-systems. A major task is the harmonisation of this understanding to a level that will enable further developments within CESSDA (e.g. the PaSC, the question bank). It will require the development of an (organisational) structure within the CESSDA community that enables partners who are well underway in their PID implementation to assist partners who are just starting to use PID, by providing guidance.

It is therefore important that the CESSDA PID policy is treated as a 'living document': the present version is to provide a framework for the first harmonisation step necessary for current CESSDA SP.

Regular review of the policy at two-year intervals will enable tracking whether its implementation is proceeding according to plan. It will also ensure that the CESSDA PID policy can incorporate further developments in the field at a speed that it can handle.

Examples are:

- The introduction of PID for other entities than data collections, such as for researchers creating/collecting the data and their institutions;
- The work on interoperability between different PID-systems;
- The work on connecting PID and Linked Open Data.

Appendix A: Definitions and Acronyms

Acronyms	Definition
CESSDA SP	CESSDA Service Providers
Dataset	A file that contains data, published or curated by a single agent, and available for access or download in one or more formats. ⁹
Data collection	A collection of files, some of which may contain data, and some other types of information.
DOI	Digital Object Identifier
FAIR	Findable – Accessible – Interoperable – Reusable. Principles set by FORCE 11
IDF	International DOI Foundation
PA	Publication Agents
PaSC	Product and Service Catalogue
PID	Persistent Identifiers
Landing page	A Web page that can be navigated to in a Web browser to gain access to the dataset, its distributions and/or additional information. ¹⁰
Resource-discovery metadata	Metadata that are provided for information retrieval systems that are based on a comprehensive central search index covering several metadata sources.
URN	Uniform Resource Name
URN:NBN	National Bibliography Number

⁹ <https://www.w3.org/TR/vocab-dcat/#class-dataset>

¹⁰ https://www.w3.org/TR/vocab-dcat/#Property:dataset_landingpage

Appendix B:

The documents in appendix B clarify the content of the Best Practice Guidelines.

CESSDA PID Task Force:

PID - What are the benefits for CESSDA community (users and providers of services)?

Cologne 26 September 2016

CESSDA PID Task Force:

Review of the PID Services provided by GESIS, DANS and SND.

Cologne 29 August 2016

CESSDA PID Task Force:

PID - What are the benefits for the CESSDA community?

Cologne 26 September 2016

Implementing and maintaining the use of persistent identifiers in an organisation is a classic change-management process. It requires both investment, in time and other resources, and commitment from staff. The return on this investment, however, can be significant.

When implementing persistent identifiers a data provider should consider the direct and indirect benefits¹¹.

CESSDA Service Providers

- The data are identified in a unique and persistent way so that they can be referred to
- Reliable access to digital objects
- Cost-savings in staff time spent handling objects or re-identifying information (versioning)
- Potential for the development of further services (e.g. linking of objects)
- Greater confidence in managing information and objects
- Enhanced ability to publish information and to make the CESSDA SP's collection better visible
- Ability to share CESSDA SP's collection through other portals: local, regional, national, thematic and international
- Ensuring that information and knowledge is used effectively in the future even if local staff changes
- Linking of objects
- Compliance with certification frameworks (such as the DSA, guideline #13)
- Compliance with CESSDA Access Policy (2016)
- Compliance with the FAIR Principles

CESSDA Portal

- Enhanced ability to publish information and to make the CESSDA collection better visible
- Greater confidence in managing information and objects
- Cost-savings in staff time spent handling objects or re-identifying information (versioning)
- Potential for the developments of further services (e.g. linking of objects)
- Lower threshold for multidisciplinary research

Primary researchers

- Get a permanent, citable link to the dataset usable in publications, for reference and impact measurements.
- Get credits for the creation of datasets
- Increase his/her visibility

Secondary researchers

- Get a clear link and citation to the dataset used for research.

Funders:

- Can associate both direct and indirect (secondary use) outcome of funding.
- Greater clarity to funders about the extent and content of your collections

Further Links:¹²

<http://www.ncdd.nl/en/pid/>

<http://www.ncdd.nl/en/pid-wijzer/>

¹¹ See: www.athenaeurope.org/getFile.php?id=779

¹² All links are retrieved on 10 December 2016.

<http://www.ands.org.au/guides/persistent-identifiers-working>
<https://www.clarin.eu/content/goals-and-requirements-pid-systems>
<http://www.athenaeurope.org/getFile.php?id=779>
<http://www.datasealofapproval.org/en/assessment/>
https://www.lshtm.ac.uk/research/researchdataman/depositdata/data_compass_doi_policy.pdf
<http://www.da-ra.de/en/for-data-centers/register-data/why-do-i-need-a-doi/>
http://dasish.eu/publications/projectreports/DASISH-D5_1_BS_version141106.pdf
<http://www.persid.org/initiative.html>
https://www.doi.org/driven_by_doi/DOI_Marketing_Brochure.pdf
<http://www.persistent-identifier.de/?lang=en>
<http://www.langzeitarchivierung.de/Subsites/nesstor/DE/Standardisierung/PI.html>
<http://www.ariadne.ac.uk/issue56/tonkin>
<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/persistent-identifiers>
<http://www.pidconsortium.eu/>
<http://handbook.dpconline.org/technical-solutions-and-tools/persistent-identifiers>
<https://www2.usgs.gov/datamanagement/preserve/persistentIDs.php>
<http://www.gbif.org/resource/80575>
<https://rd-alliance.org/groups/pid-information-types-wg.html>
<http://biorxiv.org/content/early/2016/12/28/097196>

CESSDA PID Task Force:

Review of the PID Services provided by GESIS, DANS and SND

Cologne, 29 August 2016.

1. DataCite - PID Services based on the DOI[®] system¹³

Introduction

DataCite is an international not-for-profit organisation formed in 2009. The organisation consists of a managing agent (currently the German National Library of Science and Technology) and members, and is represented by a board.

DataCite operates globally, but is nationally represented to reach out to research groups in different countries. Member institutions interact directly with clients, and offer services to the research community. Member institutions that provide identifiers for clients are called Allocation Agencies.

DataCite's mission is to establish easier access to research data on the Internet, increase acceptance of research data as legitimate, citable contributions to the scholarly record and to support data sharing that will permit results to be verified and re-purposed for future study.

The global uniqueness of the centralized allocation of DOI names is secured by using prefixes to create unique namespaces that can be used within different organisations. Suffixes are assigned by the registrant/organisation, and are unique for each object within that organisation. Prefixes are distributed by Allocation Agencies that are Members of Registration Agencies, such as DataCite (see the section about Governance Structure).

Additional metadata, one of the features that distinguish DOI names from "original" Handles, are stored at the Registration Agency DataCite Metadata Store but also at some Allocation Agencies (CDL-EZID, ETH DOI Desk, GESIS-da|ra). The metadata are used for providing additional services.

Governance Structure

The International DOI Foundation (IDF) is the Registration authority of the ISO standard (ISO 26324) of the DOI system. IDF is also the governance and management body for a federation of Registration Agencies.

There are several Registration Agencies for registering DOI names, since different communities require different services. They all provide services that allocate DOI prefixes, register and resolve DOI names, as well as store metadata corresponding to the registered objects. DataCite is one of the Registration Agencies in the IDF. DataCite Members, in turn, can be Allocating Agencies that allocate DOI names on behalf of the DOI Registration Agency of DataCite. Allocating Agencies work with data centres the so called Publication Agents (PA)

Services

Assign DOI names

Through the web interface or the API of the DataCite Metadata Store¹⁴ the PA are able to mint DOI names.

¹³ A DOI (Digital Object Identifier) is a persistent identifier. The DOI name is defined as a digital identifier of an object, rather than an identifier of a digital object. The DOI[®] system is managed by the International DOI Foundation (IDF), and is an extension of the Handle System architecture.

Once created, information about a DOI name is available through different services such as search¹⁵, event data¹⁶, OAI-PMH¹⁷, and others (see below).

Metadata Schema¹⁸

DataCite gathers metadata for each DOI name assigned to an object. DataCite operates its own metadata schema. Current version is 3.1.

Cite your data¹⁹

Data citation is fundamental as it enables easy reuse and verification of data, making it possible to track and quantify the impact of data. Data citation creates a scholarly structure that recognises and rewards the producers of data. DataCite recommends a specific format for data citation which includes the DOI name of the object as a linkable, permanent URL.

Format your citation²⁰

The DOI Citation Formatter²¹, a service created in collaboration with CrossRef²², allows for formatting of citations according to the users' needs.

Find a dataset²³

DataCite Metadata Search²⁴ is a service that allows people to search for datasets registered with DataCite, via the metadata associated with the datasets.

Get your DOI statistics²⁵

DataCite provides open statistics on DOI registrations and DOI resolutions, filtered by Allocating Agency, data centre or Prefix.

Content Negotiation²⁶

DataCite's Content Resolver exposes the metadata stored in the DataCite Metadata Store using multiple formats. It can also redirect to content hosted by DataCite PA, making it possible to access data directly using a DOI name.

OAI-PMH Provider²⁷

The DataCite OAI-PMH Provider²⁸ is a service that exposes DataCite Metadata using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)²⁹.

Metadata Store³⁰

The DataCite Metadata Store³¹ is a service allowing PA to mint DOI names and register associated metadata.

¹⁴ <https://mds.datacite.org/>

¹⁵ <https://www.labs.datacite.org/search.html>

¹⁶ <https://www.labs.datacite.org/eventdata.html>

¹⁷ <https://www.labs.datacite.org/oaipmh.html>

¹⁸ <http://schema.datacite.org/>

¹⁹ <https://www.datacite.org/cite-your-data.html>

²⁰ <https://www.datacite.org/citation.html>

²¹ <http://citation.crosscite.org/>

²² <http://crossref.org/>

²³ <https://www.datacite.org/search.html>

²⁴ <http://search.datacite.org/ui>

²⁵ <https://www.datacite.org/stats.html>

²⁶ <https://www.datacite.org/content.html>

²⁷ <https://www.datacite.org/oaipmh.html>

²⁸ <http://oai.datacite.org/>

²⁹ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

³⁰ <https://www.datacite.org/doiis.html>

³¹ <https://mds.datacite.org/>

The service requires data centres to first register for an account with a DataCite Member (Allocating Agency)³².

Test Environment³³

Additionally to the Metadata Store³⁴, DataCite has set up an environment for testing all services including DOI registration. The test environment is a closed system; i.e. DOI names registered here are not resolvable with any resolvers other than the one in the test system. Access for PA or interested data centres is granted via the Allocation Agencies.

Labs Search³⁵

This service is a labs version of the Metadata Search. It includes all functionality of the DataCite/ORCID claiming tool.

Profile³⁶

DataCite Profiles integrates DataCite services from a user's perspective and provides tools for personal use. In particular, it is a key piece of integration with ORCID³⁷, where researchers can connect their profiles and get automatic updates of their ORCID record when any of their works obtains a DOI name.

Status³⁸

DataCite's status page provides information about scheduled maintenance, service outages and statistics. It is accessible via a web interface, an API, a mailing list and Twitter.

Event Data³⁹

DataCite Event Data service (prototype) retrieves and exposes the activity that occurs around research data objects. In particular, Event Data brings to light links between publications and data, citations, software, reuse, documentation, etc. This service is accessible via a web interface and an API.

Find a repository⁴⁰

DataCite supports all researchers intending to deposit and/or find data, in collaboration with re3data.org.

1.1 da|ra - a DOI-Based PID Service

Introduction

In cooperation with DataCite, GESIS - Leibniz Institute for the Social Sciences and the ZBW - Leibniz Information Centre for Economics offer a DOI registration service in Germany for social science and economic data.⁴¹

The service began in 2010 as a GESIS pilot project for developing and testing the technical and organisational concept of DOI registration. This service has been available online since 2010 and by the end of that year the

³² <https://www.datacite.org/members.html>

³³ <https://www.datacite.org/test.html>

³⁴ <https://mds.datacite.org/>

³⁵ <http://search.datacite.org/>

³⁶ <https://profiles.datacite.org/>

³⁷ <http://orcid.org/>

³⁸ <http://status.datacite.org/>

³⁹ <https://eventdata.datacite.org/>

⁴⁰ <https://www.datacite.org/re3data.html>

⁴¹ <http://www.da-ra.de/en>

entire GESIS data archive holdings had been registered. Since 2011 the registration service has been available for research data centres and data producers in Germany and beyond. In cooperation with the ZBW the service was extended to include economic data as well. Meanwhile da|ra is the central registration agency for social and economic data in Germany.

Governance

da|ra is jointly run and administered by GESIS and ZBW. It is publicly funded and the DOI service is provided free of charge to the academic community.

Clients

da|ra serves a broad variety of users.⁴²

Services

Metadata schema⁴³

da|ra provide a specific metadata schema that is adapted to the needs of social and economic research areas.

Minting of DOI names⁴⁴

da|ra allows for minting DOI names via the da|ra web entry mask and xml-upload as well as the da|ra web service (API). The API enables to work automated with the DOI registration service without being forced to log in to an existing da|ra user account at the web page.

Provision of a test environment⁴⁵

da|ra offers a special interface for prospective PA. Additionally to the live system da|ra has set up an environment for testing the DOI registration. It is connected to the DataCite test system and uses the same test features.

Website⁴⁶

The da|ra website provides all administrative and technical information needed for the DOI registration as well as access to a comprehensive search functionality, a DOI resolver plugin, a citation output (beta), linking data and full text papers (beta).

User support⁴⁷

da|ra offers a Help desk and Best Practice Recommendations.

Search & Access API⁴⁸

This is a service for checking the availability of the da|ra system (isAlive) and searching the metadata via a web service.⁴⁹

⁴² <http://www.da-ra.de/en/about-us/our-users/>

⁴³ <http://www.da-ra.de/en/technical-information/doi-registration/#c880>

⁴⁴ <http://www.da-ra.de/en/for-data-centers/obtain-a-user-account>

⁴⁵ <http://dara-test.gesis.org:8084/dara/mydara?lang=en>

⁴⁶ <http://www.da-ra.de/en/home/>

⁴⁷ <http://www.da-ra.de/en/about-us/da-ra-policy/best-practice/>

⁴⁸ <http://www.da-ra.de/dara/api/isAlive>

⁴⁹ <http://www.da-ra.de/fileadmin/media/da-ra.de/PDFs/daraSearchAccessApi.pdf>

OAI-PMH Data Provider⁵⁰

It offers access to the metadata of the registered research data using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The da|ra OAI-PMH Data Provider is able to disseminate records in the following formats: DDI-Lifecycle 3.1 and OAI DC (Dublin Core).

The da|ra PA have access to all DataCite services as well, see chapter 1.

1.2 SND - a DOI-Based PID Service

Introduction

SND became a member of DataCite in 2010 to get access to a DOI prefix for its own data collections. SND does not run any special services in the PID area, but as the membership puts SND in the role of allocation agent of DOI prefixes nationally there are some activities such as promoting the usage of PID in general and advising clients and others interested of how to use them. These activities are not restricted to any specific disciplines but cover all research areas.

Clients

SND serves a broad variety of users.⁵¹

Services

The SND PA have access to all DataCite services, see chapter 1.

1.3 DANS - a DOI-Based PID Service

Introduction

DANS is Publication Agent (PA) and assigns a new DOI, within the DANS-prefix, to each dataset deposited in the EASY system. EASY is the system that manages the ingest and archival storage of research datasets.⁵² A dataset consist of data files and metadata. An individual file does not get its own identifier. DOIs are assigned when a dataset is submitted in EASY; registration of the DOI with DataCite takes place when the dataset is published in EASY. DANS is also able to import existing DOIs when a dataset is already deposited in another place.

Information on the use of PID at DANS is available for the public.⁵³ This document explains what a persistent identifier is and how DANS implements the PID (both URN:NBN and DOI) in practice.

Clients

Both individual researchers and institutes are able to deposit in EASY, and have a DOI minted to their dataset.

Governance

DANS registers the metadata with the Dutch DataCite Allocating Agency at TU Delft.⁵⁴

⁵⁰ <http://www.da-ra.de/oaip/>

⁵¹ <https://snd.gu.se/en/about-us/services/pid-service>

⁵² <http://easy.knaw.nl>

⁵³ https://dans.knaw.nl/en/deposit/information-about-depositing-data/persistent-identifiers/persistent-identifiers?set_language=en

⁵⁴ <http://datacite.tudelft.nl/info/home/>

2. An European URN:NBN-based PID Service – the URN:NBN Cluster Service

Introduction

In contrast to the DOI services which came into existence on an international and European level the URN:NBN based PID services were developed on the national scale independent of each other. They are provided by the National Libraries and other national organisations like DANS in the Netherlands.

The first URN:NBN PID service has been introduced in Germany by the German National Library (DNB) in 2001. Since then, other European countries – such as the Netherlands, Sweden, Norway, Finland, Hungary, Italy, the Czech Republic, and Austria – have set up their own URN:NBN services. Currently not all EU countries are covered.

Organisation of the URN:NBN Cluster

In 2012, the DNB started the implementation of the URN:NBN Cluster which resolves the German and Swiss URN subdomains, and supports redirection of the URN subdomains Austria, Czech Republic, Finland, Hungary, Italy, the Netherlands, Norway, Sweden and the PID-Systems DOI, Handle and ARK. The goal of the DNB and its partners is to create a common infrastructure for URN:NBN resolution in Europe, with one resolution service as a single point of entry with high availability for the multiple (national) URN namespaces. Administration and resolving of URN:NBNs will be separated in this infrastructure.

DANS an URN:NBN-based PID Service

Introduction

DANS assigns a new URN:NBN, with the DANS-prefix, to each dataset deposited in the EASY system. DANS hosts the national resolver for the URN:NBN:NL domain.⁵⁵ In collaboration with university libraries, research institutions, the Royal Library of the Netherlands and EduStandaard, DANS works to maintain the national URN:NBN infrastructure. The Royal Library is the national registration agency for the URN:NBN.⁵⁶ DANS also participates in several persistent identifier projects. One is the project Persistent Identifiers⁵⁷ of the Digital Heritage Network⁵⁸. This project has developed a PID-guide⁵⁹, online available via the website of the National Coalition for Digital Preservation (NCDD).

In EASY (see above) URN:NBNs are assigned when a dataset is submitted; registration of the URN:NBN with the national resolver infrastructure takes place when the dataset is published in EASY. Information on the use of PID at DANS is available for the public.⁶⁰ This document explains what a persistent identifier is and how DANS implements the PID (both URN:NBN and DOI) in practice.

Clients

All institutions assigning URN:NBN:NL identifiers in the Netherlands⁶¹ make use of the resolver infrastructure for resolving the PID to the current URL. Both individual researcher and institutes are able to deposit in EASY, and have a URN:NBN minted to their dataset.

⁵⁵ <http://persistent-identifier.org/>

⁵⁶ <https://www.kb.nl/organisatie/onderzoek-expertise/informatie-infrastructuur-diensten-voor-bibliotheken/registration-agency-nbn>

⁵⁷ <http://www.ncdd.nl/en/ncdd-projects/digital-heritage-network/project-persistent-identifiers/>

⁵⁸ <http://www.ncdd.nl/en/ncdd-projects/digital-heritage-network/>

⁵⁹ <http://www.ncdd.nl/en/pid-wijzer/>

⁶⁰ https://dans.knaw.nl/en/deposit/information-about-depositing-data/persistent-identifiers/persistent-identifiers?set_language=en

⁶¹ <https://www.kb.nl/organisatie/onderzoek-expertise/informatie-infrastructuur-diensten-voor-bibliotheken/registration-agency-nbn/catalogus>