Datacite DOIs

Ari Lukkarinen 2019-11-15

CSC PID verkosto

Persisten identifiers

- Identifiers
  - Nametag (ID= something)
  - Black button for doors (ID= something)
  - Car's register plate, etc.
- These are needed to identify "things" in a standardized way in IT systems.

- Persistent identifiers
  - Promise that
    - the identifier will be maintained
    - The target will be maintained

- Resolver
  - System that maintain information about identifiers and their targets
  - Is able to inform user about the location of the target

CSC

PID services in Finland

- Handle based systems
  - DOI (http://www.doi.org)
    - Well defined metadata schema
    - Mandatory metadata attributes
    - Landing page as a target
    - Available from Datacite (CSC) and from Crossref
    - CSC is a member of Datacite Foundation
  - ePIC (http://www.pidconsortium.eu/)
    - No metadata requirements
    - Can also be used for individual files
    - Service provided by members. CSC is a member of ePIC.

- URN
  - Managed by Finnish national library
  - Identifier by itself does not contain any metadata information. It is a pointer.

Future work

- Metadata and PID systems
  - Metadata model for digital objects is needed anyway !
  - How much of this information should be harvested to PID systems ?
  - Perhaps only that part what can be expected to be usefull for end users !
  - Datacite 3.1 schema used widely
  - Openaire and B2Share use Datacite 3.1 schema
  - **Openaire has own requirements for metadata !**

- What "things" should have well defined persistent identifiers ?

- Dynamic data
  - Should the identifier point to a service providing up to date information ?

  - Or should all services be designed so that there is well defined way to create an identifier that can be later used to retrieve the answer from a system in a same form as it was when the identifier was created ?

Changes in Datacite service

- Datacite had a policy, that each country had one operator, allocating agent, that was a contact point in that country.

- CSC ís at the moment an allocating agent. Allocating agent is able to sell Datacite services.

- To manage Datacite services at Finland Datacite Finland consortium was formed

- Current members are
  - CSC fairdata services
  - CSC EUDAT services
  - Aalto University

Policies – DOI, Datacite, Datacite Finland

- DOI standard
  - Framework for DOI based identifiers
  - Metadata should support …

- Datacite
  - Detailed metadatamodel
  - Roles and responsibilities

- Datacite Finland
  - Detailed policy document

- Datacite Finland policy
  1. CSC's roles and responsibilites
  2. Customer roles and responsibilities
  3. Control of digital objects and metadata
  4. Ownership and intellectual property rights
  5. Proper use of service
  6. Data Protection and Information Security
  7. Change of allocating agency
  8. Exit procedures

CSC

Changes in Datacite service pricing

- Datacite had a policy, that each country had one operator, allocating agent, that was a contact point in that country.

- CSC ís at the moment an allocating agent. Allocating agent is able to sell Datacite services.

- In new operational model each organization can become a direct member of Datacite Foundation.

1. Fixed sum 8500€ annually.

2. Costs billed on allocating agent level Yearly membership fee + #prefix + #DOI

3. Membership fee (organization) + #repository + #DOI

Our old cost estimate is not sustainable any more

DOI Service Fees for Non-profit Organizations

1. Annual Membership Fee: 2000 euros
2. Annual Repository Fee
3. "New DOIs /year" Fee

Repositories		DOIs (Digital Object Identifiers)	
Ranges	Annual Fee	Ranges	Annual Fee
0-1	500€	0-10,000	500€
2-5	1,000€	10,001-100,000	2,000€
6-10	3,000€	100,001 and up	3,000€
11-50	6,000€		
50 and up	10,000€		

Datacite service pricing, direct member

Example pricing: Direct Member with 3 repositories and 10,000 DOIs

	Membership Fee	Number of Repositories	Repository Fee	Number of DOIs /year	DOI Fee	Total Membership Fee and Service Fee
Direct Member Organization	2000€	3	1000€	10,000	500€	3500€

Datacite service pricing, consortium member

	Membership Fee (split between organizations)	Number of Repositories	Repository Fee	Number of DOIs /year	DOI Fee	Total Membership Fee and Service Fee
Organization 1 (consortium lead)	666€	2	1000€	8,500	500€	2166€
Organization 2	666€	1	500€	600	500€	1666€
Organization 3	666€	1	500€	1,500	500€	1666€
Totals	2000€		2000€		1500€	5500€

Crossref and Datacite comparison

- Crossref yearly depends on revenue of publishing operations or expenses of an organization

- Datacite yearly depends on number of repositories.

	Datacite (€/early)	Crossref ($/quarterly)
0-10 000	0,05	0,06
10 000 – 100 000	0,02	0,03
100 000 – 1 000 000	0,003	0,02
1 000 000 – 10 000 000	0,0003	0,01
10 000 000 -	3000 / N	0,005

CSC

Datacite Finland

1. Fixed sum, about 2000 € annually, is not sustainable any more

2. Current consortium model with flat fee is difficult

Options:

1. All organizations will become direct members

2. Continue with consortium model with new price structure mimicing DataCite pricing model. Datacite membership fee would be shared among participants, but Datacite Finland yearly fee would need to be increased.

CSC

Datacite Finland service pricing, estimate

Example pricing: Direct Member with 3 repositories and 10,000 DOIs

	Membership Fee	Number of Repositories	Repository Fee	Number of DOIs /year	DOI Fee	Total Membership Fee and Service Fee
Direct Member Organization	2000€	3	1000€	10,000	500€	3500€

Datacite Finland Cost estimate

3000 €

4500 €

Datacite services, how to continue ?

- We would like to know your opinion !

- If we dismantle Datacite Finland, all organizations will become direct members of Datacite. Datacite will take care of all service issues.

- Informal PID forum will continue !

Datacite Finland and changes in Datacite organization

- Datacite had a policy, that each country had one operator, allocating agent, that was a contact point in that country.

- CSC ís at the moment an allocating agent

- Allocating agent is able to sell Datacite services.

- CSC provides first line of support and can help if needed.

New operational model

- In new operational model each organization can become a direct member of Datacite Foundation.

- Datacite provides first line of support

The FAIR principles for data

Findable

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

Accessible

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A2. Metadata are accessible, even when the data are no longer available

Interoperable

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

Reusable

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

CSC

PID microservice

- Service that would be used to create and modify "whatever" identifier in a standard way

- Would be used to
  - Track that data objects managed by CSC are accessible.
  - Keep change log (what, who, when…)
  - To avoid duplicate entries
  - To identify identical targets in different PID systems
  - To create standardized PIDs following CSCs logic
  - To act as a resolver

- PID microservice needs a database that knows identifiers and where data objects are

- Same database could also be used as one service component for data management at CSC services

CSC level dataset management