Persistent Identifier (PID)

Options appraisal

Alex Hardisty

Wouter Addink

General Assembly (GA) 2, 16th October 2020

A PASSPORT CONTAINS DATA ABOUT A PERSON.
A DIGITAL SPECIMEN CONTAINS DATA ABOUT THE PHYSICAL SPECIMEN.

Each person has his own Personal Identification Number

Contains a reference to

Personal Identification Number

Passport with Passport Number

Each physical specimen has its own identifier (often not globally unique)

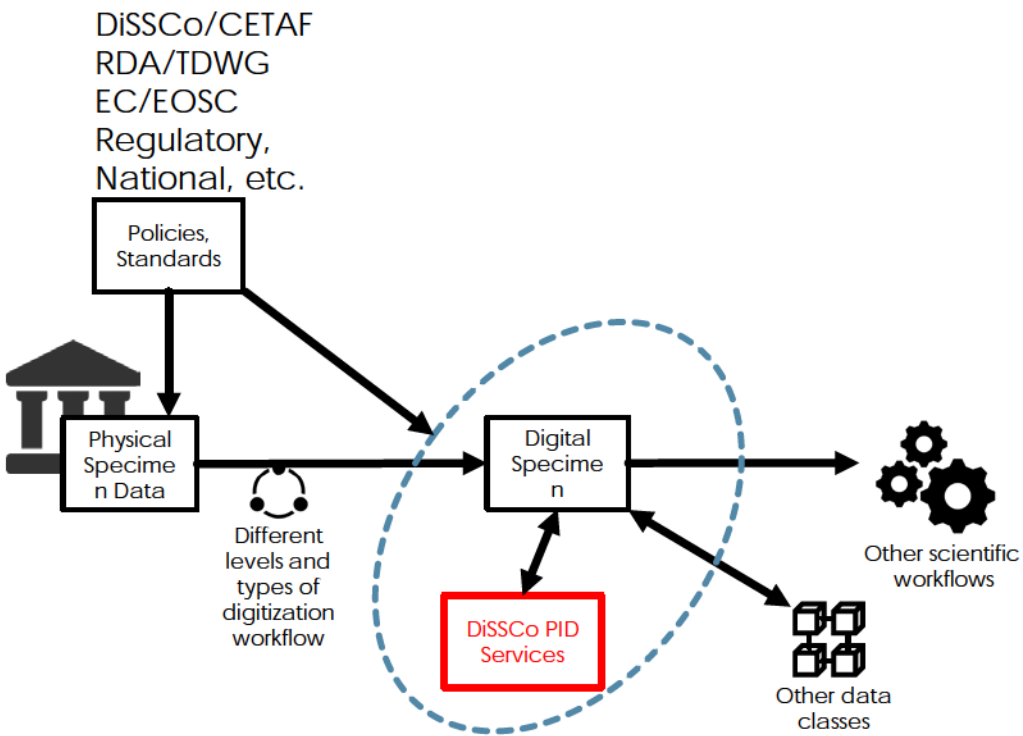
(Image by NHM London)

Contains a reference to

physical specimen identifier

Digital specimen with its unique persistent identifier (PID)

INTRODUCTION – SETTING THE SCENE



DiSSCo needs PIDs and PID services to support the ambition for Digital Specimens, virtual collections, workflows, etc. on the Internet; for ELViS loans and visits,for annotations, citations, attribution of work and microcredits; to pursue aims of common policies and procedures; and to transform work practices.

Example PIDs:

doi: 10.7299/X75Q4W7G
https://doi.org/10.7299/X75Q4W7G

hdl: 20.5000.1025/c2618387bb0932270617

https://hdl.handle.net/20.5000.1025/c2618387bb0932270617
https://doi.org/20.5000.1025/c2618387bb0932270617

REQUIREMENTS OF A PID SCHEME FOR DISSCO - SIX MAIN ONES

- **Scalability**: Scale for specimens, scale for machines, scale for global use.
- **Identifiers appropriate to the situation**: PIDs appropriate to the digital object type being persistently identified -----> Handle, DOI, ORCID Id, Wikidata item, GRID/ROR.
- **Trust**: User confidence; seeing the PID scheme as appropriate to their needs and trustworthy.
- **Persistence**: Heritage timescales – more than 100 years.
- **Governance**: By stakeholders themselves. Internationally accepted mechanism recognizing the foundational value of PIDs in delivering most of the open science aspirations in EU, US, and around the world.
- **Potential for global adoption**: Extensible towards a single PID scheme that could be adopted globally.

+ specific to natural sciences, alignment to EOSC PID policy, etc.

ALIGNING WITH THE EUROPEAN OPEN SCIENCE CLOUD (EOSC)

EOSCFAIR Executive Board Working Group

EOSCArchitecture Executive Board Working Group

Second draft Persistent Identifier (PID) Policy for EOSC
https://doi.org/10.5281/zenodo.3780423

- *A Persistent Identifier that supports and enables research that is FAIR is one that is globally unique, persistent, and resolvable*
- *To make it globally resolvable, the PID needs to be part of a namespace defined by a syntax that is controlled by an Authority*
- *The EU research community needs to be represented in the governance structure*

EOSCArchitecture Executive Board Working Group

PID Architecture for the EOSC v0.3
https://docs.google.com/document/d/1iJ0NP7Ec2o_P3_DkDsi_ngiiExNdRZDcvXhdAEIADx

- *The Handle technology provides most of the components described in the (EOSC) PID Architecture*
- *FAIR Digital Objects: A deep interconnection of both (FAIR and Digital Object) approaches can be extremely fruitful.*

20+ OPTIONS TO CHOOSE FROM

Scheme: / DiSSCo modes:		DOI (10.)	IGSN	ePIC (21.)	CNRI 5-digit prefix	New top-level prefix	Second level prefix	Three segment prefix	National-level services
		A	B	C	D	E	F	G	H
Ally with MPA	1	Possible	Not possible	Possible	Deprecated	Possible	Possible	Not possible	Not possible
Act as MPA	2	Not possible	Not possible	Not possible	Not possible	Possible	Not possible	Not possible	Not possible
Use existing RA	3	Possible	Possible	Possible	Possible	Not possible	Possible	Possible	Not desirable
Ally with RA	4	Possible	Possible	Possible	Possible	Not possible	Possible	Possible	Not desirable
Become an RA	5	Possible	Not possible (AA only)	Possible	Deprecated	Possible	Possible	Possible	Not desirable

MPA = Multi-Primary Administrator (e.g., International DOI Foundation). RA = Registration Agency (e.g., DataCite). AA = Allocating Agent.

EVALUATION

- First step: Reduce the number of alternatives to a sensible and practical subset. A coarse, three-level scoring against each major requirement (strong, weak, in-between).
- Second step: For 7-9 strongest and 'do nothing' scenarios, a more detailed assessment.
  - 'Do nothing' implies difficulty in the future of finding digitized specimens of interest and ad-hoc evolutions of existing local practices.

20+ OPTIONS TO CHOOSE FROM : 5 + 2 COMBO MAKE SENSE TO EVALUATE (against 10 dimensions, outcomes and impact, pros and cons)

Scheme: / DiSSCo modes:		DOI (10.)	IGSN	ePIC (21.)	CNRI 5-digit prefix	New top-level prefix	Second level prefix	Three segment prefix	National -level services
		A	B	C	D	E	F	G	H
Ally with MPA	1	Possible	Not possible	Possible	Deprecated	Possible	Possible	Not possible	Not possible
Act as MPA	2	Not possible	Not possible	Not possible	Not possible	Possible	Not possible	Not possible	Not possible
Use existing RA	3	Possible	Possible	Possible	Possible	Not possible	Possible	Possible	Not desirable
Ally with RA	4	Possible	Possible	Possible	Possible	Not possible	Possible	Possible	Not desirable
Become an RA	5	Possible	Not possible (AA only)	Possible	Deprecated	Possible	Possible	Possible	Not desirable

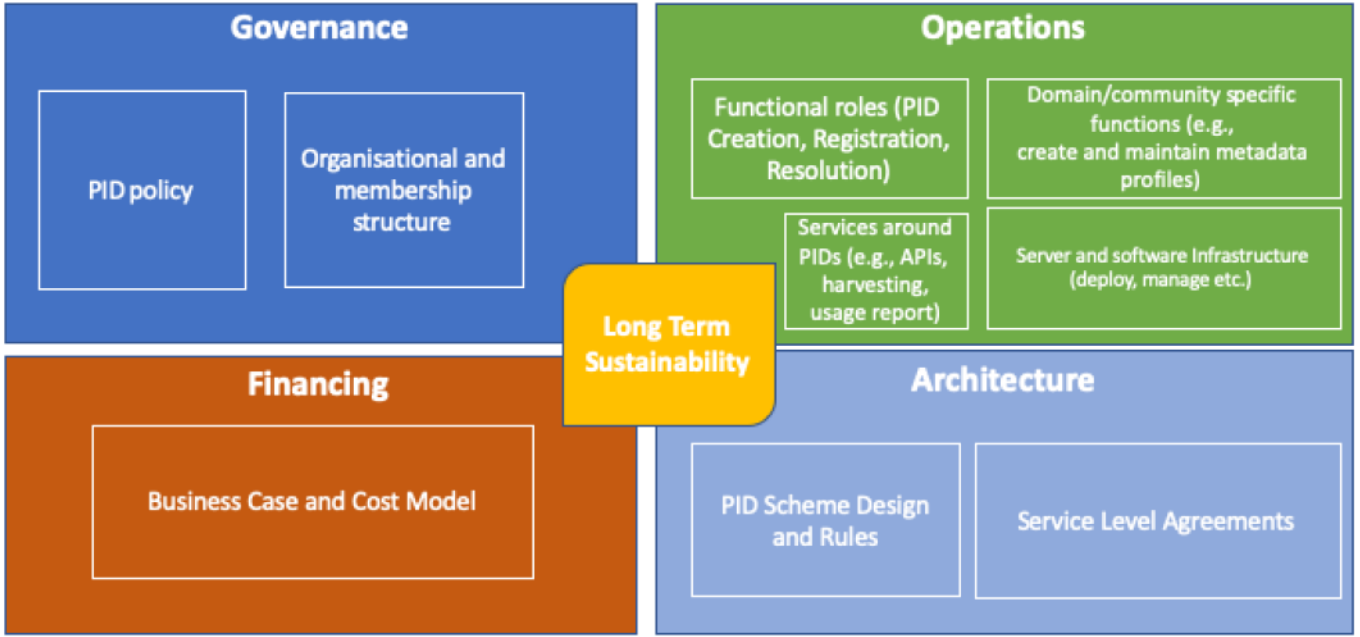
MPA = Multi-Primary Administrator (e.g., International DOI Foundation). RA = Registration Agency (e.g., DataCite). AA = Allocating Agent.

EVALUATION AND PREFERENCE

'powered by DOI®'

- Rule out option to act as MPA/New top-level prefix. Is a step too far. Has rigorous obligations (and associated costs) that community probably would find hard to accept.
- **Preferred option**: Choosing Digital Object Identifiers (DOI) and allying with the International DOI Foundation (IDF) comes out more strongly than other options.
  - Mainly for reasons related to the substantial achievements, operational experience and reputation of DOI/IDF to date over multiple industry sectors.
  - Also: Good uptake and familiarity of DOIs in the NSC community already, especially around journal articles, supplementary materials and datasets publishing. Also, already being used by GBIF. And compatible with EOSC PID Policy and FAIR. Financially viable. Acceptable globally.

STILL, SOME REQUIREMENTS NOT FULLY MET
(SCALE, GOVERNANCE, PERSISTENCE, SPECIFICITY TO SECTOR - METADATA).
HOW TO PRACTICALLY IMPLEMENT THE CHOICE HAS DIFFERENT POSSIBILITIES - DISCUSS IN DEPTH WITH DOI FOUNDATION AS NEXT STEP.



- Brand differentiation
  Natural Science Identifiers, (NSId), 10.22 as recognisable prefix.
- Achieving operational autonomy
  Can lead to a new Registration Agency alongside DataCite, Crossref and others.

NEXT STEPS

Customise with natural sciences community characteristics

- With DiSSCo Prepare Task 6.2 partners + optionally, a few additional DiSSCo members (expressions of interest/in-kind contribution needed now):
  - Make robust, committed plan for operations, governance and sustainability over the long-term.
  - Establish a Local Handle Service to gain operational experience in an experimental context as basis for developing the scheme architecture
  - Scheme design and rules, metadata elements, plan for necessary software development.
- Continued consultation in the DiSSCo/CETAF/International community.
  - CETAF WGs, National Nodes, international stakeholders
- Intensify discussions with DOI Foundation.
  - Aim to become general member in early 2021.

QUESTIONS AND COMMENTS