Making Physical Objects FAIR - DiSSCo approach

Distributed System of Scientific Collections

Wouter Addink, wouter.addink@naturalis.nl
Deputy Director, DiSSCo Coordination and Support Office
https://orcid.org/0000-0002-3090-1761
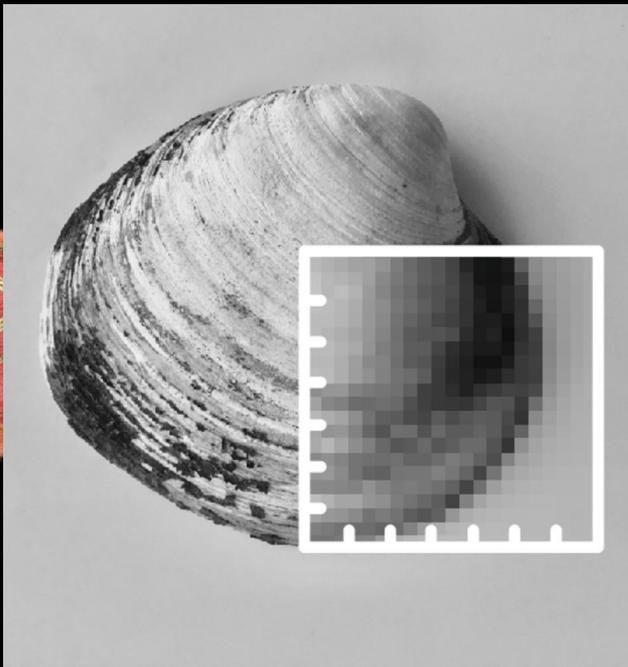
The goal: to <u>digitally</u> unify today's fragmented landscape of European NSCs into a single knowledge base under common curation, access policies and practices.

170+ National Facilities,
23 Countries
1 European Collection

DiSSCo members

Austria - Naturhistorisches Museum Wien **Belgium** - Agentschap Plantentuin Meise - Royal Belgian Institute of Natural Sciences - Instituut voor Natuur-en Bosonderzoek (INBO) - Royal Museum for Central Africa, Tervuren - Flemish Research Institute for Nature and Forest - Royal Zoological Society of Antwerp - Université Libre de Bruxelles, Herbarium - University of Namur - Flanders Marine Institute **Bulgaria** - Institute of Biodiversity and Ecosystem Research – Bulgarian Academy of Sciences (IBER-BAS) - National Museum of Natural History, Sofia – Bulgarian Academy of Sciences (NMNHS-BAS) **Czech Republic** - Charles University, Faculty of Science - Institute of Botany, The Czech Academy of Sciences - Institute of Vertebrate Biology, The Czech Academy of Sciences - National Museum, Prague - Nature Conservation Agency of the Czech Republic - Masaryk University **Denmark** - Natural History Museum of Denmark, University of Copenhagen - Naturhistorisk Museum Aarhus - The Science Museums, Aarhus University **Estonia** - Estonian Museum of Natural History - Estonian University of Life Sciences - Tallinn University of Technology - University of Tartu **Finland** - Finnish Museum of Natural History - LUOMUS, University of Helsinki - Biodiversity Unit, University of Oulu - Biodiversity Unit, University of Turku - Digitarium, University of Eastern Finland - Kuopio Natural History Museum - Open Science Centre / Museum, University of Jyväskylä **France** - Muséum national d'histoire naturelle - Muséum d'histoire naturelle Philadelphe-Thomas de Gaillac - Centre de coopération internationale en recherche agronomique pour le développement - Centre Informatique National de l'Enseignement Supérieur - Nancy Museum-Aquarium- Conservatoire botanique national Alpin - Institut Recherche et Développement - Sorbonne University - La Société National des Sciences Naturelles et Mathématiques de Cherbourg - Le Jardin Botanique de la Ville de Lyon - Muséum d'histoire naturelle de La Rochelle - Université de Toulouse III-Paul Sabatier - Université Lille 1 – Sciences et technologies - Université Claude-Bernard Lyon - Université Clermont Auvergne - Université de Bourgogne - Université de Montpellier - Université de Rennes - Université de Strasbourg - Université Pierre et Marie Curie - Tela Botanica **Germany** - Bavarian Natural History Collections – SNSB - Berlin Natural History Museum - Botanischer Garten und Botanisches Museum Berlin – Freie Universität Berlin - Centrum für Naturkunde, Universität Hamburg - Senckenberg Gesellschaft für Naturforschung - Zoological Research Museum Alexander Koenig - Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science Berlin - Naturkundemuseum Stuttgart **Greece** - Natural History Museum of Crete, University of Crete - Mineralogy and Petrology Museum, National and Kapodistrian University of Athens - Museum of Geology and Palaeontology, National and Kapodistrian University of Athens - Museum of Zoology, National and Kapodistrian University of Athens - Botanical Museum, National and Kapodistrian University of Athens - Zoological Museum of the University of Patras - Department of Geology, University of Patras - Botanical Museum of the University of Patras - Museum of Zoology of the Aristotle University of Thessaloniki - Goulandris Natural History Museum - Benaki Phytopathological Institute - Botanical Garden "Ioulia & Alexandros N. Diomidis" - Cephalonia Botanica - Focas-Cosmetatos Foundation - Collection of Micro-organisms, Agricultural University of Athens - Hellenic Collection for Pathogenic Fungi - Herbarium, Agricultural University of Athens - Hippocrates Botanical Garden of Kos - Hippocrates Botanical Garden of Markopoulo - Mycetotheca, National and Kapodistrian University of Athens - Natural History Museum of the Municipality of Amaroussion - Phytotheca, National and Kapodistrian University of Athens - Sedbank, National and Kapodistrian University of Athens - The Apollon-Delphi Botanic Garden - Hippocrates Botanical Garden of Limnos, University of the Aegian **Hungary** - Hungarian Natural History Museum **Italy** - Natural History Museum, University of Florence - National Academy of Sciences Italian National Academy of Entomology - Italian Botanical Society - Italian Geological Society - Italian Palaeontological Society - Italian Society of Biogeography - National Association of Science Museums - National Research Council (DSSTTA) **Luxembourg** - Musée national d'histoire naturelle (MnhnL) **Netherlands** - Naturalis Biodiversity Center - Natural History Museum Rotterdam - Natuurmuseum Maastricht - NIOZ Royal Netherlands Institute for Sea Research - Stichting De Bastei (Natuurmuseum Nijmegen) - Stichting Museon - Stichting TwentseWelle - Teylers Museum, Haarlem – NLBIF - University of Amsterdam - Utrecht University Museum - Westerdijk Fungal Biodiversity Institute - Natuurmuseum Brabant - Natuurmuseum Fryslân **Norway** - Natural History Museum, University of Oslo - Tromsø University Museum - NTNU University Museum - University Museum of Bergen **Poland** - Museum and Institute of Zoology, Polish Academy of Sciences - The University of Warsaw **Portugal** - National Museum of Natural History and Science, University of Lisbon (MUHNAC – Ulisboa) - Science Museum of the University of Coimbra (MUC – UC) - Natural History and Science Museum, University of Porto (MHNC – UP) - Botanical Garden of the University of Coimbra (JBUC) **Slovakia** - Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Institute of Botany - Pavol Jozef Šafárik University in Košice - Comenius University **Spain** - National Museum of Natural Sciences (CSIC-MNCN) - Royal Botanic Garden of Madrid (CSIC-RJB) - Universidad de Navarra (UNAV) - Instituto Geológico y Minero de España (IGME) **Sweden** - Swedish Museum of Natural History - Bergius Foundation - Department of Biology, Lund University - Gothenburg Botanical Garden - Department of Biological and Environmental Sciences, University of Gothenburg - Department of Ecology and Environmental Science, Umeå University - Uppsala University, Museum of Evolution - Västarvet, The Gothenburg Museum of Natural History **United Kingdom** - Natural History Museum London - Royal Botanic Garden Edinburgh - Royal Botanic Gardens, Kew.

Current model
Slow
Fragmented
Expensive
Inefficient
Limited to physical access

Integrated RI model
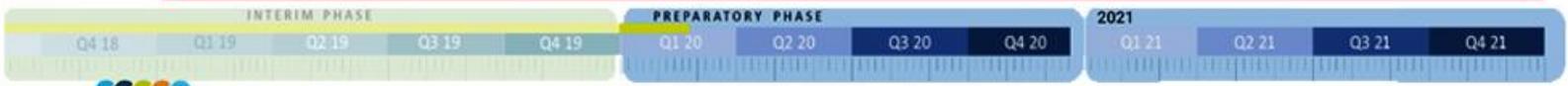Global access
Lower costs
Faster
New insights
Optimised
FAIR data

DISCO

ESFRI

- **DiSSCo links data: DIGITAL OBJECT architecture**
- **And not just any kind of data: actionable, FAIR data**

PHYSICAL SPECIMEN

DIGITAL SPECIMEN Actionable Knowledge Unit

Digitisation
Analysis
Computation

Genomic data
Biochemical data
Morphological data
Geographical data
Taxonomic information
Species interactions data
Ecological data

FAIRdata
Findable
Accessible
Interoperable
Reusable

Design considerations

9 Characteristics of DiSSCo data management essential to protect throughout the lifetime of the infrastructure

1. **Centrality of the digital specimen** - as digital surrogate of the physical object
2. **Accuracy and authenticity of the digital specimen** - best available digital representation
3. **FAIRness** - by adopting Digital Object Architecture
4. **Protection of data** - following legal regulations and community norms
5. **Preserving readability and retrievability** - including image data
6. **Traceability (provenance) of specimens** - store as part of the DS based on the W3C PROV framework
7. **Annotation history** - interpretations are part of the scientific and historical record
8. **Determinability (status and trends) of digitisation** - minimum information level calculated
9. **Securability** - authority and permissions to retrieve and/or modify sensitive information
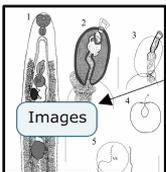
Digital Extended Specimen - concept
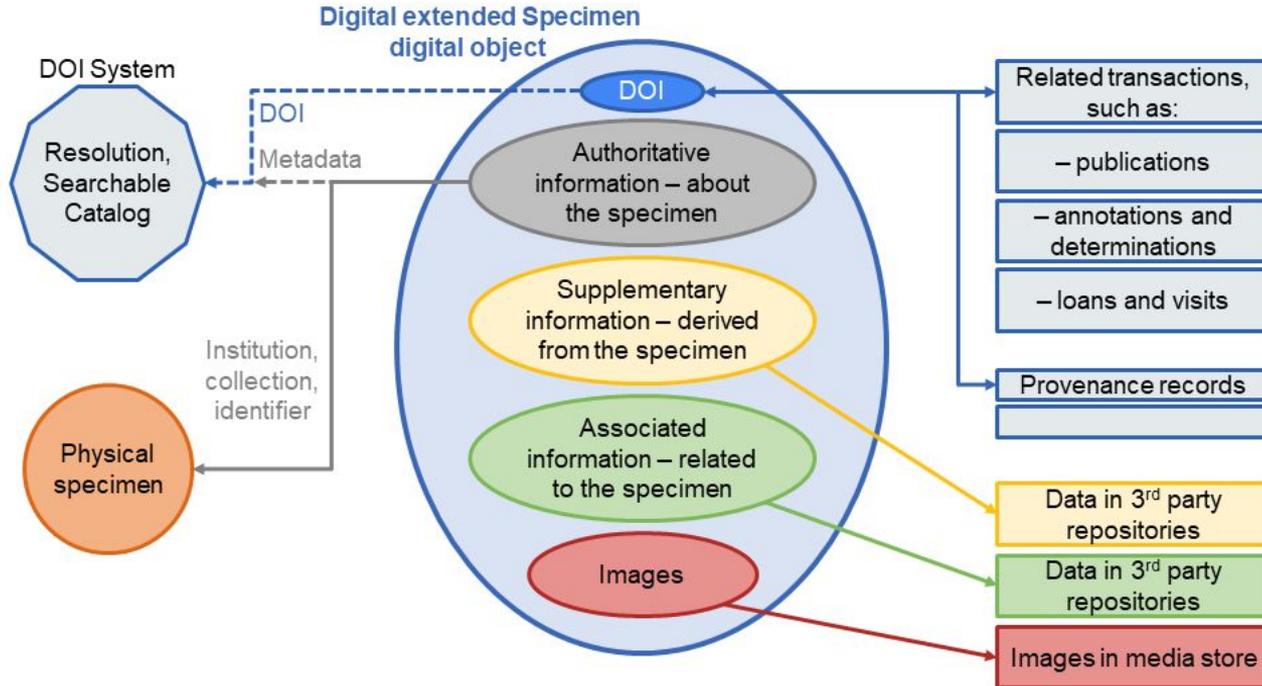
A linked Digital Specimen

Digital Specimen - benefits

- Enables **community curation** through annotations
- Researchers can build on efforts by other researchers to validate and enhance the data, **saving time in data wrangling** and speeding up scientific discoveries
- FAIR metadata that provides provenance and **attribution**
- Enables **linking** with derived and related data, usage tracking, better citation
- Enables AI services that **speed up digitization** and extract new information from specimens.
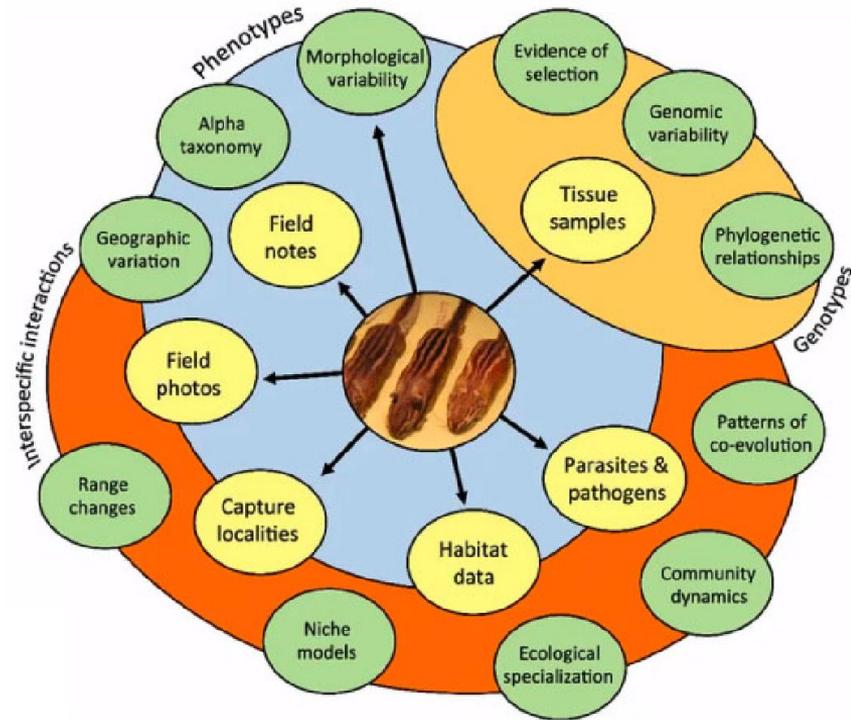
FAIR design concepts

- "**Data deluge**" → Digitization produces too much data for curation by humans

- Vision: A **global Integrated Virtual Data Collection** that can be autonomously navigated and appropriately processed by machines ("**Machine Actionability**")

- Enable machines to **autonomously detect patterns in data**

DS

PID:20.5000.1025/f9dd382c670e64cccf21

PID Layer

#0120 return License
#0123 return gathering
#0124 return DNAseq
#0121 extractROI
#0122 updateTraitInfo
#0125 return MIDS info
#0126 update MIDS level

Typed Operations

Data Model Layer

ods:type:<https://terms.dissco.tech/opends/DigitalSpecimen>
ods:name: "Hieracium kuekenthalianum"
ods:midsLevel: 1
ods:institution: <https://ror.org/01wz97s39>,
ods:institutionCode: "FR",
ods:materialType: "Dry – pinned"
dct:license:<http://creativecommons.org/licenses/by/4.0/>

rdf:type
ods:2DImageObject

Resource Layer

FDO Conceptual Model (FDOF)

FDO for Semantic Alignment (BioDT)

FAIR **DIGITAL OBJECTS** FORUM

A new PID for the Digital Specimen

- The objects in the collections have an ID attached that is often only locally unique, non-resolvable and not always persistent
- Changing the physical object IDs would be very expensive and time consuming
- Institutions make individual decisions on curation including the assignment of identifiers

Solution:

- Create a new online digital object representing the physical object with a digital object identifier
- Maintain a 1:1 relationship with the physical identifier attached to the object - one DS per object
- Use a digital object object identifier that is resolvable, persistent and machine actionable

Similar approach: BSI.Identify

Results for "track"

🔍 track

We found 3 results related to your search.

Service from British Standards Institution for Manufacturers for quick, unambiguous identification of any product at the unit level

Insulfix Track - Double R130D
Insulfix Track

UPIN 10.21/NND2F2Q9 ⎘

Insulfix Track - Double R130D. For use with Double rafters of minimum depth 130mm. Provides a quick, easy and reliable way to correctly install PIR insulation boards between rafters. Holds the boards firmly in place, minimises gaps and reduces the need for messy spray foam. By making it easy to install offcuts it can reduce waste per board by up to 90%.

Height: 130mm   Width: 97mm   Length: 2400mm   Material: uPVC

Filter results

Brand name ⌄

Classifications name ⌄

Classifications value ⌄

Manufacturer ⌄

Apply filters

Insulfix Track - Single R130S
Insulfix Track

UPIN 10.21/PQ9TMZGR ⎘

Insulfix Track - Single R130S. For use with Single rafters of minimum

UPIN = Universal Persistent Identification Number (10.21/XXXXXXXX)

Persistent Identifier Infrastructure

Every object gets a PID that needs to be …

Resolvable	Handle System
Persistent	DOIs
Informative	FDO Profile
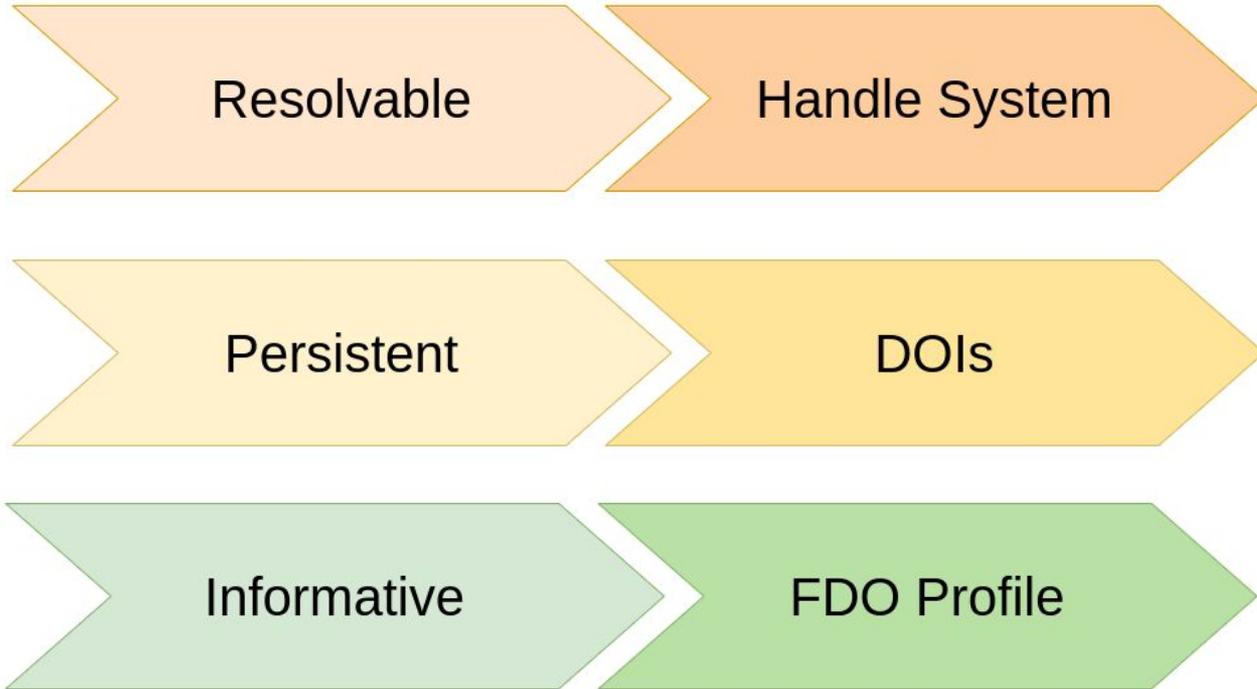
Core values and principles for scientific data
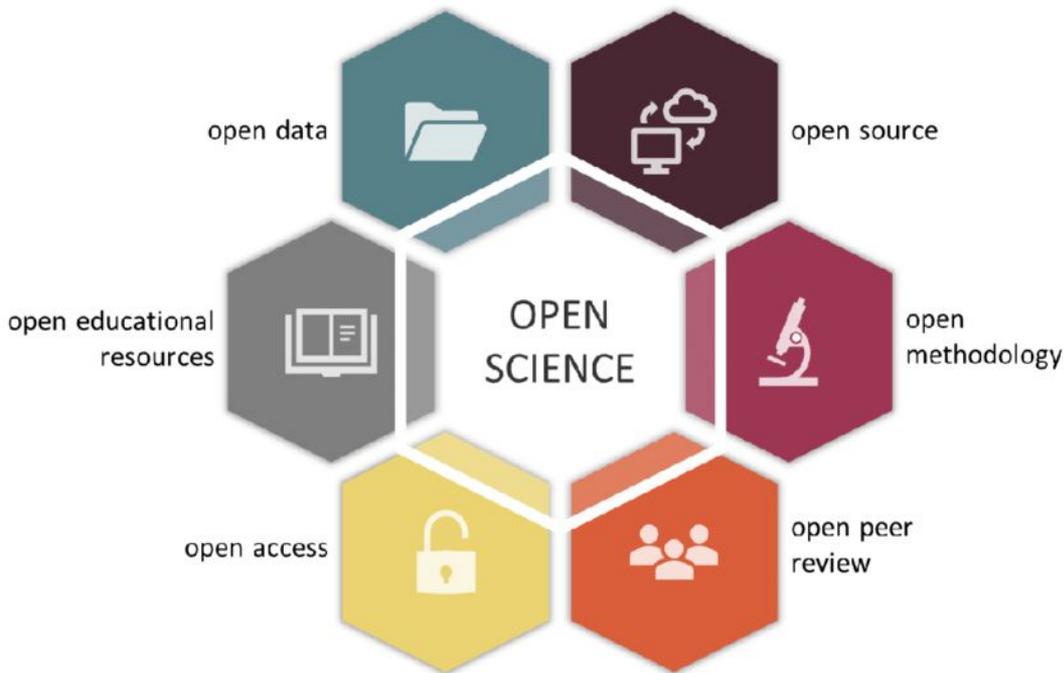
Free for researchers no pay for resolving or using the PID.

Metadata served as part of the PID record will always need to be public domain (Creative Commons 0), to support **Open Data**. Note that the data may have a different license.

Only one PID per object Solved by central authority (DiSSCo) minting the PID but requires a locally unique physical identifier

Handle system advantages

- Mature, scalable system, for resolvable identifiers; DOIs and other Handles
- Decoupling between URL and identifier, allowing the URL to change over time
- Central resolver
- Provenance metadata + optional other metadata **in the PID record**

But also:

urlappend

The value of this parameter is appended to the end of the URL used for redirection, e.g. for versioning.

e.g.: 10.22/ABC-4D5-RX7 points to latest version of the digital specimen

    10.22/ABC-4D5-RX7?urlappend=/3 points to version 3 of the digital specimen

locatt=key:value

For multiple redirection (e.g. to different landing pages or serialisations); specifies a key:value pair to determine the choice of redirection from 10320/loc values.

Handle.Net®

Handle Values for: 20.5000.1025/a6f8b178fab07cee8d9b

Index	Type	Timestamp	Data
100	HS_ADMIN	2021-06-22 09:08:57Z	handle=20.5000.1025/repo; index=0; [create hdl,delete hdl,create derived prefix,delete derived prefix,read v...
1	10320/loc	2021-06-22 09:08:57Z	\<locations\> \<location href="http://nsidr.org/objects/20.5000.1025/a6f8b178fab07cee8d9b" weight="0" view="json" /\> \<location href="http://nsidr.org/#objects/20.5000.1025/a6f8b178fab07cee8d9b" weight="1" view="ui" /\> \</locations\>
2	0.TYPE/DOIPService	2021-06-22 09:08:57Z	20.5000.1025/service

example PID record

```
id: "20.5000.1025/763-X5F-LQP",
midsLevel: 1,
version: 1,
created: "2022-11-14T12:40:41.994203Z",
type: "BotanySpecimen",
physicalSpecimenId: "http://data.rbge.org.uk/herb/E0011363",
physicalSpecimenIdType: "cetaf",
specimenName: "Prunella",
organizationId: "https://ror.org/0349vqz63",
datasetId: "Royal Botanic Garden Edinburgh Herbarium",
physicalSpecimenCollection: "http://biocol.org/urn:lsid:biocol.org:col:15670",
sourceSystemId: "20.5000.1025/3XA-8PT-SAY",
data:
{
        dwc:sex: "not recorded",
        dwc:genus: "Prunella",
        dwc:county: "Nujiang Lisu Aut.
        dwc:family: "Labiatae",
        dwc:country: "CN",
        dwc:locality: "Gongshan Suburbs
        dwc:eventDate: "1997-09-08",
```

https://hdl.handle.net/20.5000.1025/763-X5F-LQP?noredirect

Handle.Net®

Handle Values for: 20.5000.1025/763-X5F-LQP

Index	Type	Timestamp	Data
1	pid	2022-11-14 12:40:41Z	https://hdl.handle.net/20.5000.1025/763-X5F-LQP
2	pidIssuer	2022-11-14 12:40:41Z	{"id":"https://doi.org/10.22/10.22/2AA-GAA-E29","pidType":"DOI","primaryNameFromP
3	digitalObjectType	2022-11-14 12:40:41Z	{"id":"http://hdl.handle.net/21...","pidType":"Handle","primaryNameFromPid":"Digital Sp
4	digitalObjectSubtype	2022-11-14 12:40:41Z	{"id":"https://hdl.handle.net/21...","pidType":"Handle","primaryNameFromPid":"BotanySp
5	10320/loc	2022-11-14 12:40:41Z	\<locations>\<location href="https://sandbox.dissco.tech/api/v1/specimens/20.5000.1025/76
6	issueDate	2022-11-14 12:40:41Z	2022-11-14
7	issueNumber	2022-11-14 12:40:41Z	1
8	pidStatus	2022-11-14 12:40:41Z	DRAFT
11	pidKernelMetadataLicense	2022-11-14 12:40:41Z	https://creativecommons.org/publicdomain/zero/1.0/
14	digitalOrPhysical	2022-11-14 12:40:41Z	physical
15	specimenHost	2022-11-14 12:40:41Z	{"id":"https://ror.org/0349vqz63","pidType":"ROR","primaryNameFromPid":"Needs to be
100	HS_ADMIN	2022-11-14 12:40:41Z	handle=300:0.NA/20.5000.1025; index=200; [create hdl,delete hdl,create derived prefix,de admin,list]

Handling of the PID lifecycle
(related to the object lifecycle)

PID State

- RESERVED
- DRAFT
- ACTIVE

Lifecycle State

- ARCHIVED
- OBSOLETE
- FAILED
- MERGED
- SPLIT

powered by DOI®

- Substantial achievements, operational experience and reputation of DOI/IDF
- Good uptake and familiarity of DOIs in the NSC community: journal articles, supplementary materials and datasets publishing
- Already being used by GBIF
- Compatible with EOSC PID Policy and FAIR
- Financially viable
- Acceptable globally

Hardisty AR, Addink W, Glöckler F, Güntsch A, Islam S, Weiland C (2021) A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). Research Ideas and Outcomes 7: e67379.
https://doi.org/10.3897/rio.7.e67379

20+ OPTIONS TO CHOOSE FROM (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

DOI advantages

DOIs are in use for over 20 years, it is a mature system with no signs of aging
– a good candidate for PIDs for objects that have a lifespan of more than 100 years

Mature infrastructure:
- > 66,500,000,000 resolutions
- DOI proxy service spread between 4 regions, autoscaled
- All requests are logged and there is a public uptime status page.

Mature organization (organised to guarantee persistence):
- Founded in 1997 and self-sustainable
- ISO 26324 Registration Authority
- 12 established Registration Authorities

Mature adoption:
- DOIs: > 238,000,000
- Widespread usage in science
- Over 5000 assigners