

cessda eric

Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

PID services, what why how?

A short dip in the PID pond

Wednesday 14 November 2018
Belgrade, Serbia

CESSDA Widening Task

Birger Jerlehag

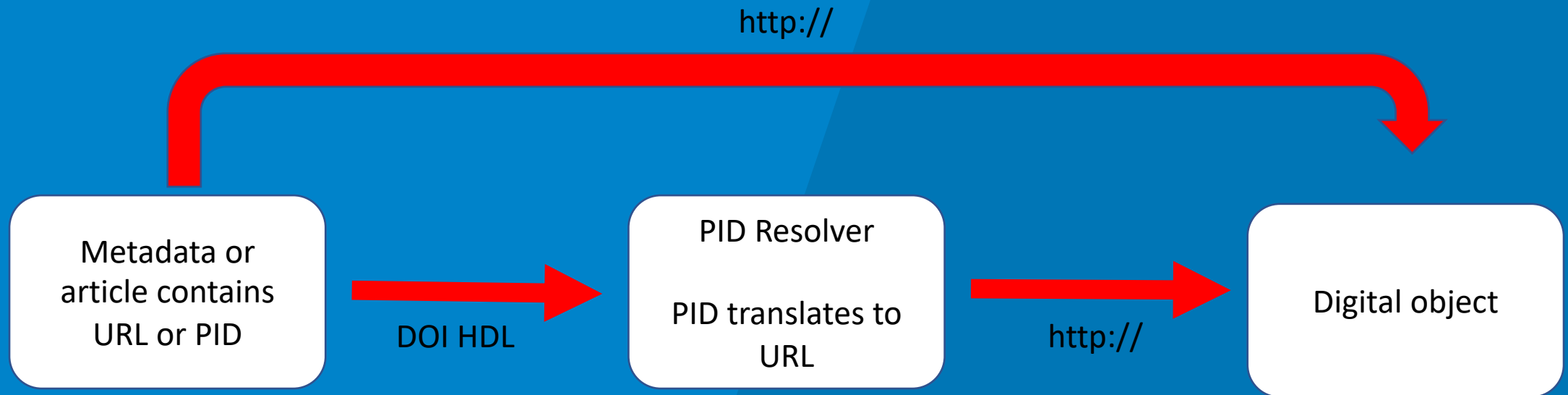
Senior Adviser

Swedish National Data Service, SND

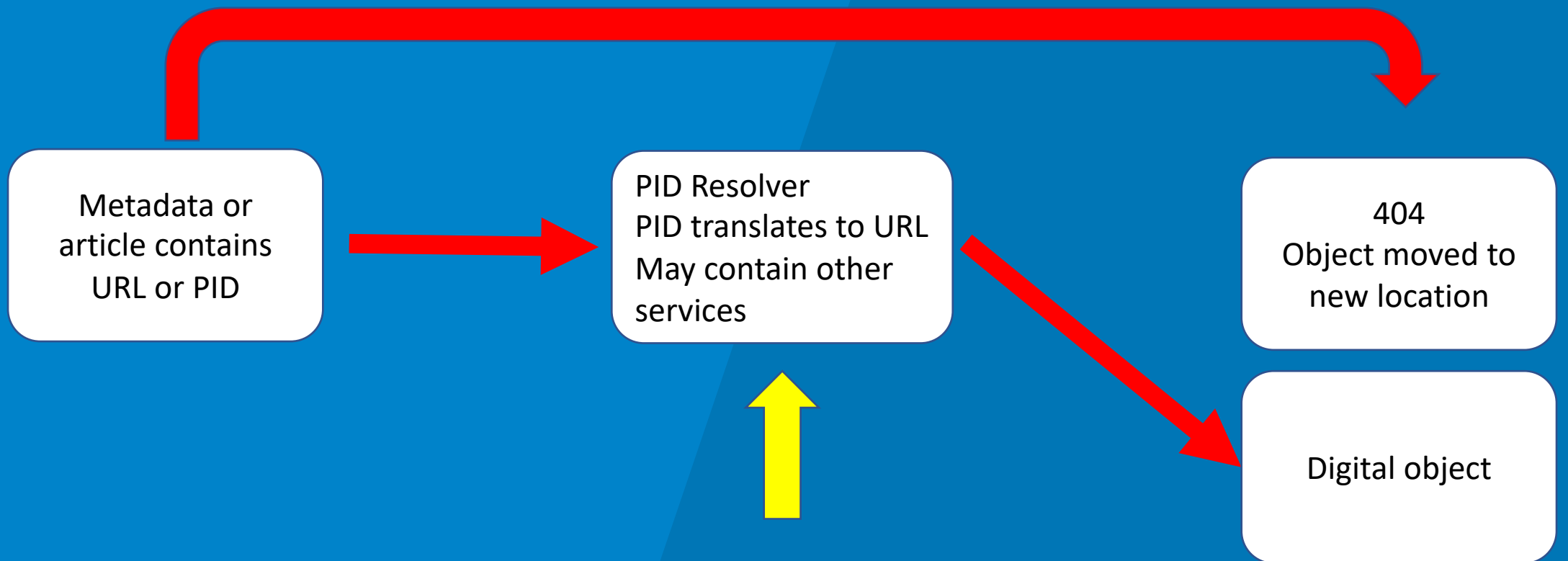
Widening Workshop



What is a Persistent Identifier



What is a PID Service



Most common PID services

Handle system

The DONA Foundation

DOI

International DOI Foundation

Digital Object Identifier

DataCite is a member of IDF

ARK

The California Digital Library

Archival Resource Key

URN:NBN

National Libraries

Uniform Resource Name : National Bibliographic Number

Why are PIDs important

Can I trust you with the data I have created or compiled

Can I rely on that data from you is trustworthy

Can data I use and cite be located and accessed in the future

The use of a trustworthy PID service is an important step to take towards FAIR data and becoming a Trusted Digital Repository

DataCite is



a global membership-based not for profit organisation

member of IDF (International DOI Foundation)

a Registration Agency for DOI prefixes

cessda eric

DataCite

90+ members/Allocation Agencies serving 1700+ Data Centers/Clients with PID prefixes

More than 15 million DOIs minted through DataCite

Financed by membership fees and projects funding (Freya ...)

Uses its own metadata schema with 19 properties

DataCite

Runs an array of services publicly available

Runs re3data (Registry of Research Data Repositories)

Works closely with other PID organisations such as CrossRef and ORCID

CESSDA Service Providers GESIS and SND are members

Costs, where to get a DOI prefix,

DataCite members can assign DOI prefixes to clients

There might be a DCmember from your country you can ask

Each DCmember can set their own cost model so costs vary

GESIS da|ra offers DOIs without cost to CESSDA SPs

GESIS da|ra

The use of the da|ra service is open to institutions interested in registering their stored research data with DOI names.

It is free of charge for clients coming from the academic field.

To become a da|ra client the respective institution has to sign a Service Level Agreement with da|ra.

CESSDA SPs GESIS, CSDA, FORS and AUSSDA are all clients at da|ra

cessda eric

CESSDA and PIDs

CESSDA/SND lead DASISH Task 5.2 – Robust PID Services – 2012/2013

The PID Services Report (D5.1) from 2014 gave a good insight to what demands SPs in all 5 SSH infrastructures had on PID Services.

The PID Task Force was formatted in 2015/2016 and could after a series of interviews, workshops and webinars present a suggestion to a PID Policy which was accepted by the General Assembly in 2017.

CESSDA PID Policy, background and considerations

Distributed infrastructure

Local/National considerations

Different levels of readiness and knowledge

No mention of PIDs in statutes or other guidelines

CESSDA PID Policy

Principle 1 – Identifying

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

Principle 2 – Locating

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

Principle 3 – Resolving

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

Principle 4 – Referencing and Citation

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

Principle 5 – Visibility

PID must be included in the resource-discovery metadata provided by CESSDA Service Providers for the CESSDA Data Catalogue.

Principle 6 – Flexibility

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

*CESSDA stands for Consortium of European Social Science Data Archives and ERIC stands for European Research Infrastructure Consortium.



Is scheduled for revision 2019

Shall be fully implemented by all SPs in 2020

cessda eric

CESSDA PID Policy Best Practice Guide

A living document where recommendations from the PID community (RDA-groups, Freya, DataCite ...) will be adopted

Identifying



- Assign globally unique PID to data that are published
- Commit to maintain PID permanently
- Do not change assigned PID
- Embed PID in data file (e.g. as variable)
- Use PID service with clear and transparent policy & business models

Locating



- Resolve assigned PID to a landing page on the website of the CESSDA SP
- The landing page should include information on the data collection (e.g. origin, version, availability, accessibility etc.)
- The landing page should include information about relations to other data collections (& their PID)

Resolving



- Use trusted, reliable & well-known PID system
- Use PID that resolve to landing page on Internet
- Guarantee resolvability of PID by maintaining landing pages (even if data collection is not accessible any longer)

Referencing and Citation



- Provide information on how to cite data collections
- Include PID in recommended citation standard
- Assign PID (at least) on data collection level

Visibility



- Provide PID in resource-discovery & harvestable metadata
- Implement version control and assign a new PID to each new version of a data collection
- If a data collection is moved to 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, including citation metadata & information about the data provider

Flexibility



- Introduction of PID for other entities than data collections (researchers creating/collecting the data and institutions)
- Work on interoperability between different PID systems and on connecting PID with LOD



CESSDA PID Project 2018

Implementing the CESSDA Persistent Identifier Policy

Extended Best Practice Guidelines on specific topics related to the FAIR principles was created.

An assessment of the results of the Research Data Alliance (RDA)-EU adoption project on “Persistent Identifier Types for the Social Sciences” was carried out.

A collocated RDA-event ‘The use of Persistent Identifiers in the CESSDA ERIC’ was arranged and a Poster was presented at the RDA Plenary.

GESIS, DANS, SND

cessda eric

CESSDA PID TO DO?

Locate
Identify
Cite

Longitudinal data
Granularity
Relations
Fragments (EQB)
Multinational datasets
PIDs for people and Orgs
Alignment with other ERICs
Other types of data, databases

cessda eric

Tip of the Iceberg — Image by © Ralph A. Clevenger/CORBIS

PID projects and future enhancements

Freya

PID Graphs

[GESIS Graph](#)

PIDappaloza

Research Data Alliance (RDA)

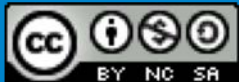
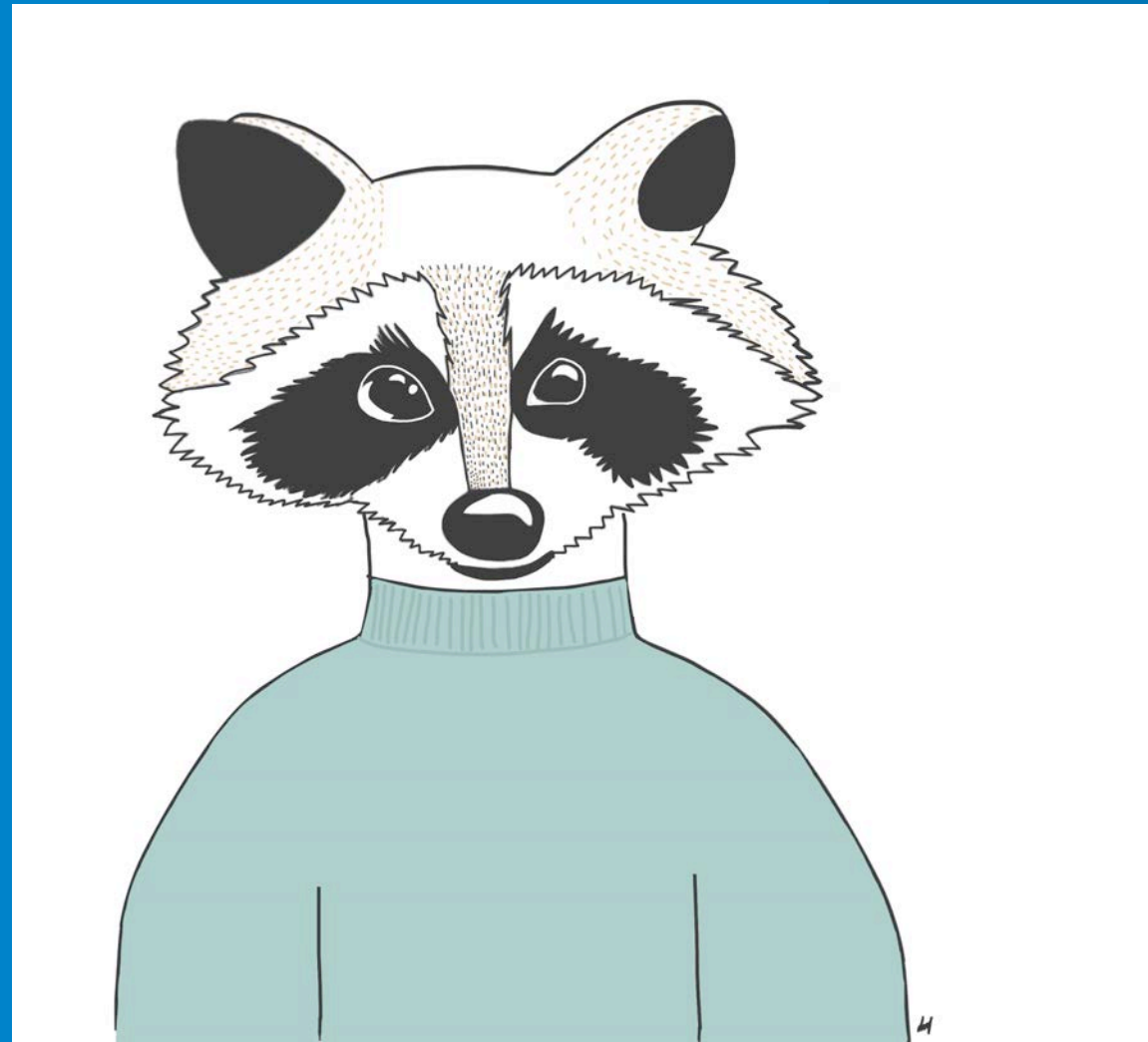
Scholix

Make Data Count

[Cite data, an ICPSR Movie](#)

cessda eric

OVER AND OUT.
???



cessda eric