The Old and New – working with existing and emerging PID types

FREYA workshop RDA-plenary 2019

Ketil Koop-Jakobsen and Tina Dohna

PANGAEA® Data Publisher

European Commission

Horizon 2020
European Union funding
for Research & Innovation

The Old and New – working with existing and emerging PID types

Overview

- The potential for expanding the PID graph with new and emerging PIDs
- The PID landscape
- Requirements for new PIDs
- PID graph functionality in practice

Expanding the PID graph with new and emerging PIDs

FREYA's role in the development of new and emerging PIDs

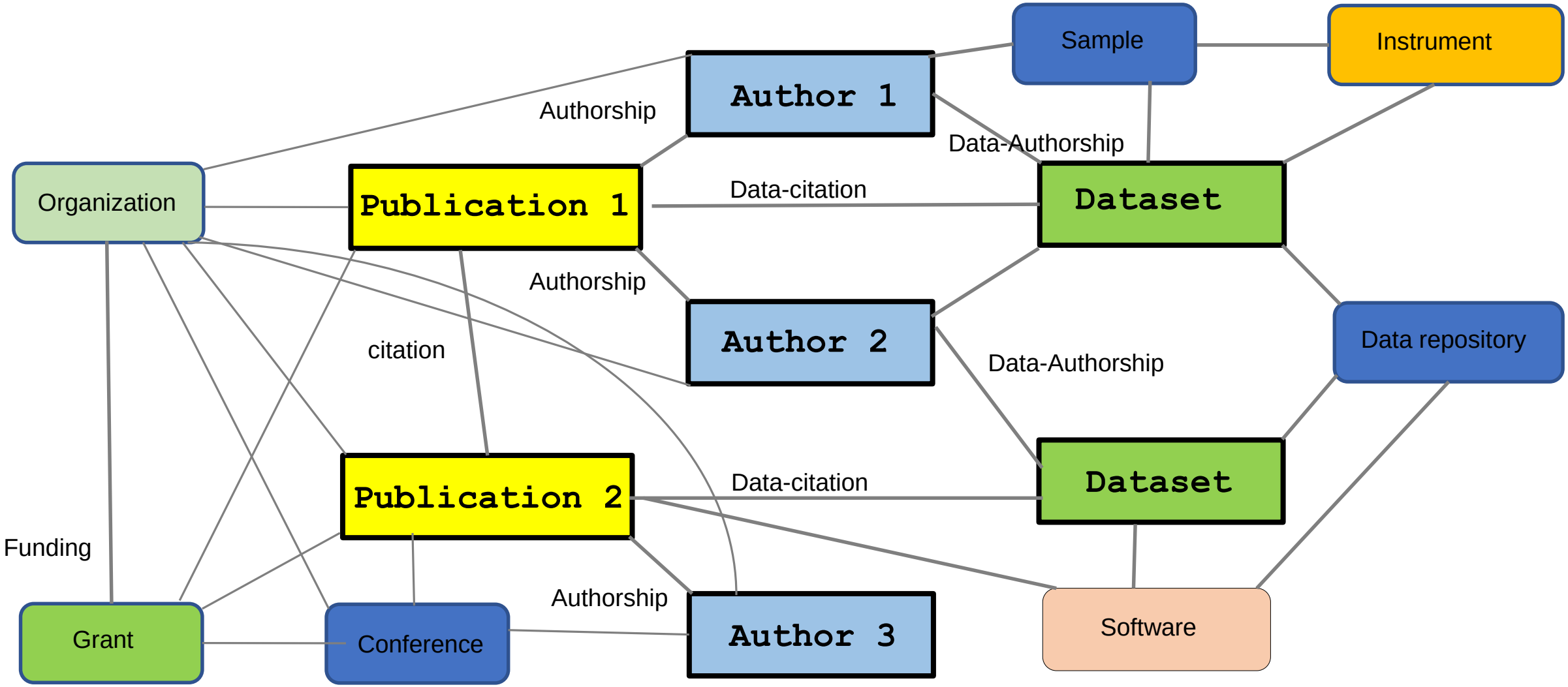
- FREYA is project – not a PID provider for new PIDs
- FREYA gathers information about new and emerging PIDs, their implementation and services.
- FREYA asks stakeholders about their current needs for news PIDs and the functionality of existing PIDs.
- FREYA is a testing ground for implementation and linking of emerging PIDs and the existing system of old PIDs (PID graph)

FREYA's role in the development of new and emerging PIDs

- FREYA …
  - small-scale implementation of emerging PIDs in existing repositories (e.g. PANGAEA)
  - elucidate benefits of an expansion of the PID landscape
  - improve/build services to exploit the PID graph functionality
  - gathers knowledge and experience on functional PID graph construction
  - identify best practices and key requirements for generation of sustainable PID services in a global perspective

New PID Types (WP3)

Extending the PID graph with New PIDs



The PID landscape

The PID landscape

Categories of PID maturity:

- MATURE PIDs: Existing PIDs with a mature PID Infrastructure and broad cross-community use and available PID services

- EMERGING PIDs: Existing PID with an emerging PID Infrastructure and upcoming services may be in pilot – early community adoption

- IMMATURE PIDs: PIDs in development/on drawing board - not yet widely used.

The PID landscape

Contents

1 Introduction 5
2 Methodology 7
2.1 Assessment approach 7
2.2 Defining "maturity" 7
3 Results 9
3.1 Overview of PID types in usage and the maturity of respective PID infrastructure 9
3.2 PID usage around specific information entities 11
3.2.1 Publications 11
3.2.2 Conferences 11
3.2.3 Researchers 12
3.2.4 Organisations 13
3.2.5 Data 14
3.2.6 Data Repositories 15
3.2.7 Grants (Research Awards) 15
3.2.8 Projects 16
3.2.9 Experiments 17
3.2.10 Investigations 18
3.2.11 Analyses (studies of data) 18
3.2.12 Software 19
3.2.13 Computer simulations 20
3.2.14 License information for software 20
3.2.15 Equipment 21
3.2.16 Archival/Storage facilities 22
3.2.17 Research stations 23
3.2.18 Samples 24
3.2.19 Cultural artefacts 25
3.2.20 Historical and mythical personae 26
3.2.21 Temporal periods and historical places 26
3.2.22 Study registrations 27
3.2.23 Data Management Plans (DMPs) 28
3.2.24 Workflows 28
3.2.25 Protocols 29
4 Concluding observations and musings 30
Annex A: Abbreviations 32

- 25 entities identified - have or need a PID
- Significant overlap among disciplines
- Complicates determination of PID maturity

The PID landscape

Table 1 Entities, PID types and their maturity

Research entity	PID types used[5]	Maturity of PID Infrastructure
Publication	DOI, Accession number, Handle, URN, Scopus EID, Web of Science UID, PMID, PMC, arXiv Identifier, BibCode, ISSN, ISBN, PURL	Mature
Citation	OCI (secondary aggregation of information)	Emerging
Conference	DOI, Accession number	Emerging
Researcher (or Scholar)	ORCID iDs, ISNI (also DAIs, VIAFs, arxivIDs, OpenIDs, ResearcherIDs, ScopusIDs)	Mature
Organization	DOI; ISNI, GRID, Ringgold IDs	Emerging
Data	DOI, Accession number, Handle, PURL, URN, ARK	Mature
Data repository		Immature

Only PIDs for **researchers, publications, and data** have services that are deemed fully mature. The remaining are either emerging or immature

Deliverable 3.1 is available on the FREYA webpage

https://doi.org/10.5281/zenodo.1324296

Requirements for new PIDs

Use-cases

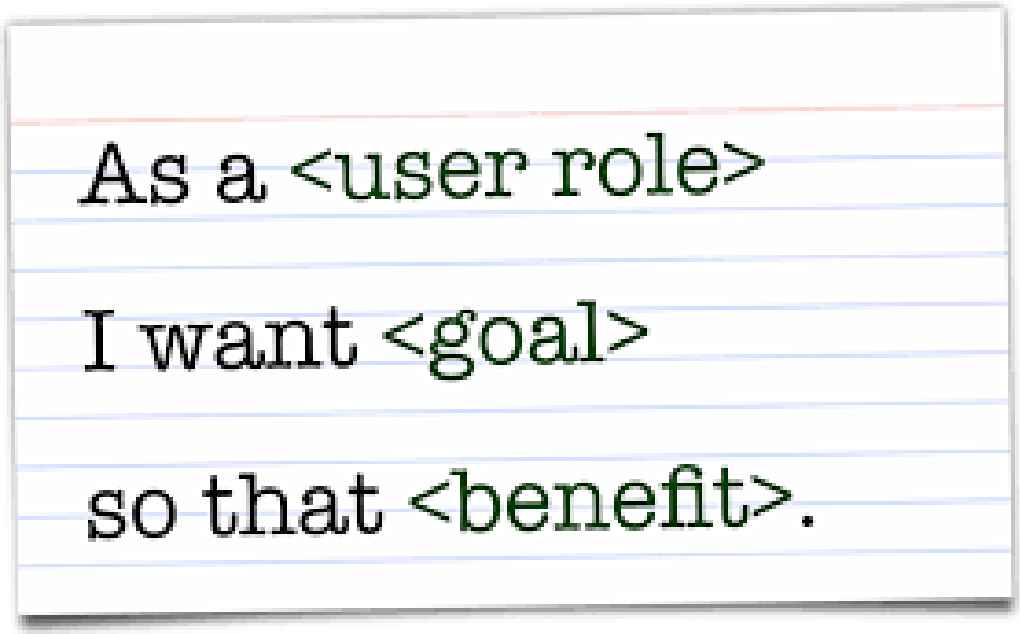
Requirements for new PIDs
Use-cases

Methods:

- Collection of Use-cases from the community, at conferences and from FREYA ambassadors (72 Use-cases collected in total)
- 30 Use-cases revolved around “New PIDs”
- Prioritize new PIDs in high demand
- Identify requirements for the progress of new PIDs
- Matching Need and Requirements with FREYA expertise

Requirements for new PIDs
Use-cases

General Use-case template



Requirements for new PIDs
Use-cases

*Use-Case Example: **Software***

As a software author, I want to be able to see the citations of my software aggregated across all versions. so that I see a complete picture of reuse

Image from:
https://opentextbc.ca/selfpublishguide/chapter/screenshots-of-software/

Requirements for new PIDs
Use-cases

Use-Case Example: Software

*As a **software author**, I want to be able to **see the citations of my software** aggregated **across all versions**, so that I see a complete picture of **reuse***

Image from:
https://opentextbc.ca/selfpublishguide/chapter/screenshots-of-software/

Requirements for new PIDs
Use-cases

Use case frequency to gage popularity – bottom up approach

Entity	Popularity
Instrument	10
data	8
article	6
person	5
Repository	5
Organisation	4
Sample	4
software	4
Grants	3
project	1
study	1
conference	1

Requirements for new PIDs
Use-cases

Selected 9 PIDs for further work:

- Instruments (PANGAEA)
- Facilities (STFC)
- Grants (EMBL-EBI)
- Organisations (DataCite)
- Software (DataCite)
- Research campaigns (PANGAEA)
- Data Management Plans (DataCite)
- Physical samples and cultural artefacts (Brit. Library)
- Conferences (CERN)

Deliverable D3.2 will provide more detailed information - https://zenodo.org/record/2649230

Requirements for new PIDs
Use-cases

1. **Identify needs based on user-stories**: Why do people want this PID?
2. **Validate current status**: What does it take to expand this PID type
3. **Validate the extent of existing PIDs**: Cross-disciplinary approach
4. **Identifying expertise:** Who can move this PID-type forward

Report is available in Zenodo at https://zenodo.org/record/2649230

PID graph Functionality in Practice

PID graph functionality in practice

CASE-STUDY: Expansion of the PANGAEA PID graph through implementation of PIDs for physical samples (IGSNs):

How implementation of new PIDs provides new service opportunities

I am a Geologist interested in sediment cores.

It has come to my knowledge that interesting research is going on in lakes of the French Alps.

I search the PANGAEA data repository for data and find the work of Bajard et al 2015

What kind of information does PANGAEAs use of PIDs provide me?

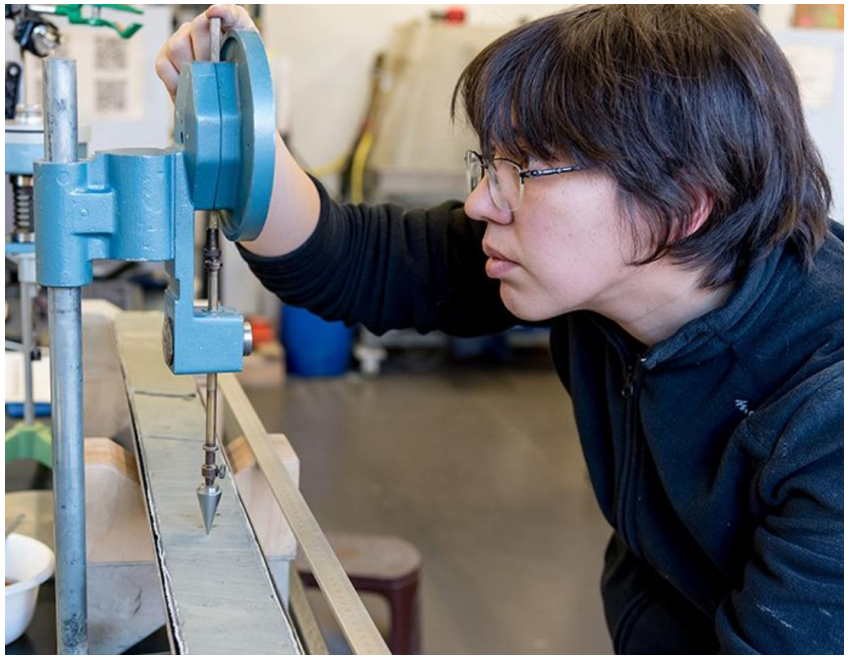


Photo: MARUM – Center for Marine Environmental Sciences, University of Bremen; V. Diekamp

PID graph functionality in practice

CASE-STUDY: Expansion of the PANGAEA PID graph through implementation of PIDs for physical samples (IGSNs):

PANGAEA.

Data Publisher for Earth & Environmental Science

SEARCH SUBMIT ABOUT CONTACT

Citation:

Bajard, Manon; Sabatier, Pierre; David, Fernand; Develle, Anne-Lise; Reyss, Jean-Louis; F… anuel; Arnaud, Daniel; Augustin, Laurent; Crouzet… ne; Arnaud, Fabien (2015): Chemical compos… Mastercore. *PANGAEA*, https://doi.org/10.159…

Sabatier, Pierre

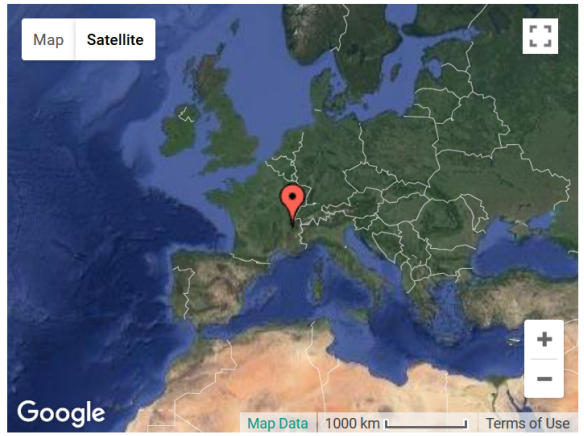
https://orcid.org/0000-0002-9620-1514

pierre.sabatier@univ-savoie.fr

In supplement to: Bajard, M et al. (2015): Erosion record in Lake La Thuile sediments (Prealps, France): Evidence of montane landscape dynamics throughout the Holocene. *The Holocene*, **26(3)**, 350-364, https://doi.org/10.1177/0959683615609750

Always quote above citation when using data! You can download the citation in several formats below.

RIS Citation | BibTeX Citation | Text Citation | Facebook | Twitter | Google+ | Show Map | Google Earth



Coverage:

Latitude: 45.530000 * *Longitude:* 6.056700

Date/Time Start: 2012-07-13T08:28:42 * *Date/Time End:* 2013-05-22T12:25:30

Minimum DEPTH, sediment/rock: 0.02000 m * *Maximum DEPTH, sediment/rock:* 6.23000 m

Event(s):

THU10-Mastercore * *Latitude:* 45.530000 * *Longitude:* 6.056700 * *Date/Time:* 2010-04-25T00:00:00 * *Elevation:* 874.0 m * *Device:* Piston corer (PC) * *Comment:* IGSN THU10-P1: IEFRA00BA; IGSN THU10-I: IEFRA00B

Mature and actionable PIDs available in PANGAEA

PIDs:

Author-PID:

Data-PID:

Article PID:

https://doi.pangaea.de/10.1594/PANGAEA.855427

Bajard, Manon; Sabatier, Pierre; David, Fernand; Develle, Anne-Lise; Reyss, Jean-Louis; Fanget, Bernard; Malet, Emmanuel; Arnaud, Daniel; Augustin, Laurent; Crouzet, Christian; Poulenard, Jérôme; Arnaud, Fabien (2015): Chemical composition of sediment core THU10-Mastercore. *PANGAEA*, https://doi.org/10.1594/PANGAEA.855427,

In supplement to: Bajard, M et al. (2015): Erosion record in Lake La Thuile sediments (Prealps, France): Evidence of montane landscape dynamics throughout the Holocene. *The Holocene*, **26(3)**, 350-364, https://doi.org/10.1177/0959683615609750

Sabatier, Pierre
https://orcid.org/0000-0002-9620-1514
pierre.sabatier@univ-savoie.fr

The Holocene
Journal Home | Browse Journal | Submit Paper | About | Subscribe
Erosion record in Lake La Thuile sediments (Prealps, France): Evidence of montane landscape dynamics throughout the Holocene
Manon Bajard, Pierre Sabatier, Fernand David, more...
First Published December 31, 2015 | Research Article
https://doi.org/10.1177/0959683615609750
Abstract
Lake La Thuile, in the Northern French Prealps (874 m a.s.l.), provides an 18-m long sedimentary sequence spanning the entire Lateglacial/Holocene period. The high-resolution

ORCID
Manon Bajard
ORCID iD https://orcid.org/0000-0003-4622-0482
Employment (1)
CNRS en Alpes: Le Bourget du Lac, Savoie
2014-10-01 to present | PhD (EDYTEM)
Works (1 of 1)
Legacy of early anthropogenic effects on recent lake eutrophication (Lake Bénit, northern French Alps)
Anthropocene
2018-12 | journal-article
DOI: 10.1016/j.ancene.2018.11.005

PANGAEA.
Data Publisher for Earth & Environmental Science
Citation: **Bajard, Manon; Sabatier, Pierre; David, Fernand; Develle, Anne-Lise; Reyss, Jean-Louis; Fanget, Bernard; Malet, Emmanuel; Arnaud, Daniel; Augustin, Laurent; Crouzet, Christian; Poulenard, Jérôme; Arnaud, Fabien (2015):** Chemical composition of sediment core THU10-Mastercore. *PANGAEA*, https://doi.org/10.1594/PANGAEA.855427,
In supplement to: Bajard, M et al. (2015): Erosion record in Lake La Thuile sediments (Prealps, France): Evidence of montane landscape dynamics throughout the Holocene. *The Holocene*, **26(3)**, 350-364, https://doi.org/10.1177/0959683615609750
Always quote above citation when using data! You can download the citation in several formats b
RIS Citation | BibTeX Citation | Text Citation | Facebook | Twitter | Google+ | Show Map | Google

874.0 m * *Device:* Piston corer (PC)

Mature a… …GAEA

PIDs:

Author-PID:

Data-PID:

Article PID:

Sample-PID:

https://app.geosamples.org/sample/igsn/IEFRA00BA

Go Back

IGSN: IEFRA00BA

IGSN: IEFRA00BA
Sample Name: THU10_P1
Other Name(s):
Sample Type: Core
Parent IGSN: Not Provided

Description	
Material:	Sediment
Classification:	Not Provided
Field Name:	Not Provided
Description:	Not Provided
Age (min):	Not Provided
Age (max):	150 years
Collection Method:	Coring>GravityCorer>Pilot
Collection Method Description:	Not Provided
Size:	1900
Geological Age:	Holocene
Geological Unit:	Not Provided
Comment:	Not Provided
Purpose:	Not Provided

Geolocation	
Latitude (WGS84):	45.530083
Longitude (WGS84):	6.056717

THU10-Mastercore * *Latitude:* 45.530000 * *Longitude:* 6.056700 * *Date/Time:* 2010-04-25T00:00:00 * *Elevation:* 874.0 m * *Device:* Piston corer (PC) * *Comment:* IGSN of cores: THU10-P1: IEFRA00BA; THU10-I: IEFRA00BB

Core in Bremen IODP Core repository - each has an IGSN

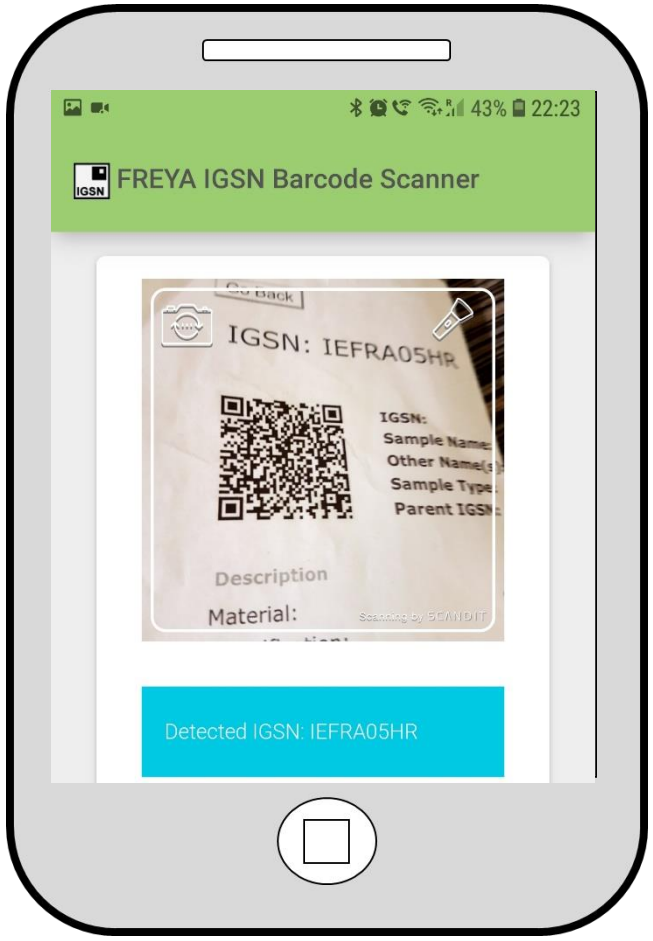
Photos: MARUM Bremen Univerisity - https://www.marum.de/en/Research/IODP-Bremen-Core-Repository.html

PID graph Functionality in Practice

FREYA IGSN Barcode scanner

GOAL: Build an app to collect metadata about a specific physical sample by using PID graph functionality

1. Automated compilation of metadata for a specific IGSN (sample PID)
2. Expanded metadata collection to include sources outside PANGAEA
3. Facilitate access to related IGSNs

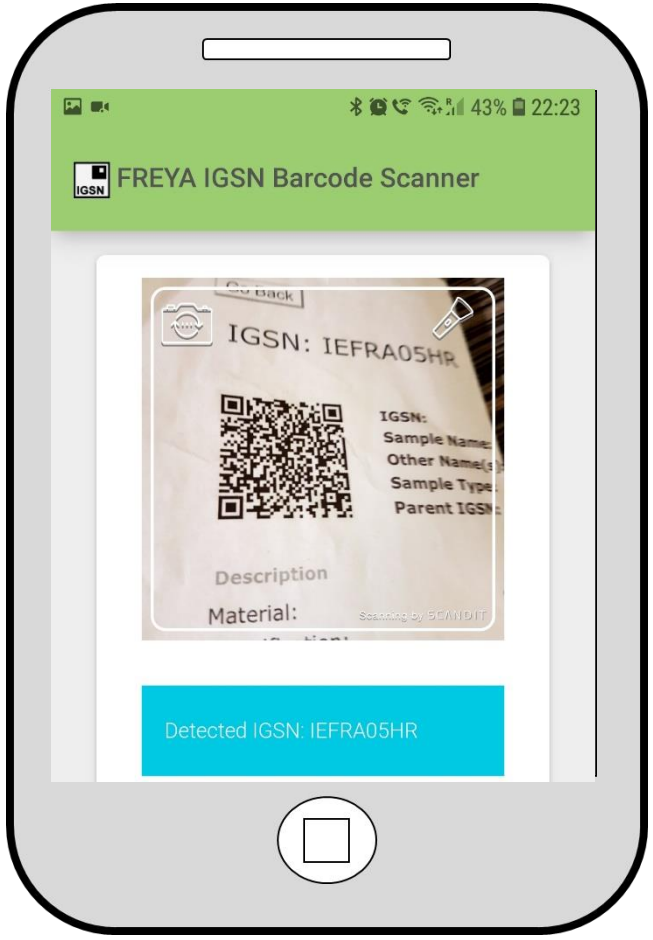


PID graph Functionality in Practice

FREYA IGSN Barcode scanner

In a nutshell:

- Provide the researcher with an easy-to-use tool
- Scan the barcode and get access to the following metadata:
  1. Data originating from the sample
  2. Researcher that has worked on the sample
  3. Funder information
  4. Related IGSNs
  5. Related publications



https://dataportals.pangaea.de/freya/igsn/

Conclusions

- Our use-case-oriented method gives a practical orientation about the USERs demand for new PIDs

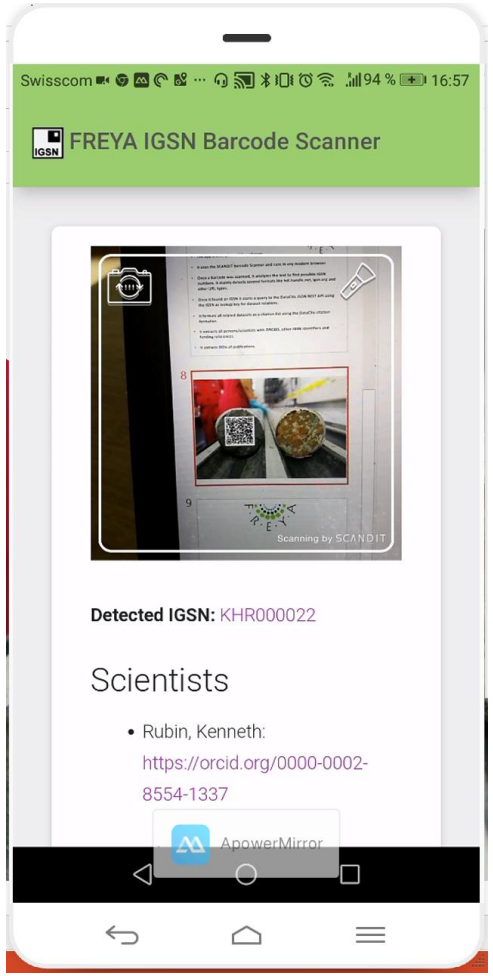
- In particular, PIDs for **Instruments, Organizations, Physical samples, Grants, Software** are sought for by the community

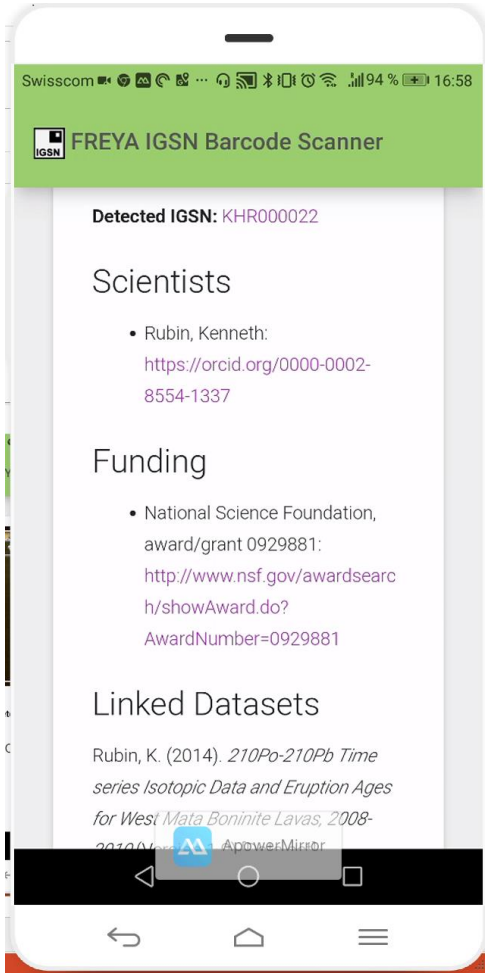
- Implementation of some of new PIDs will improve the USERs access to additional information

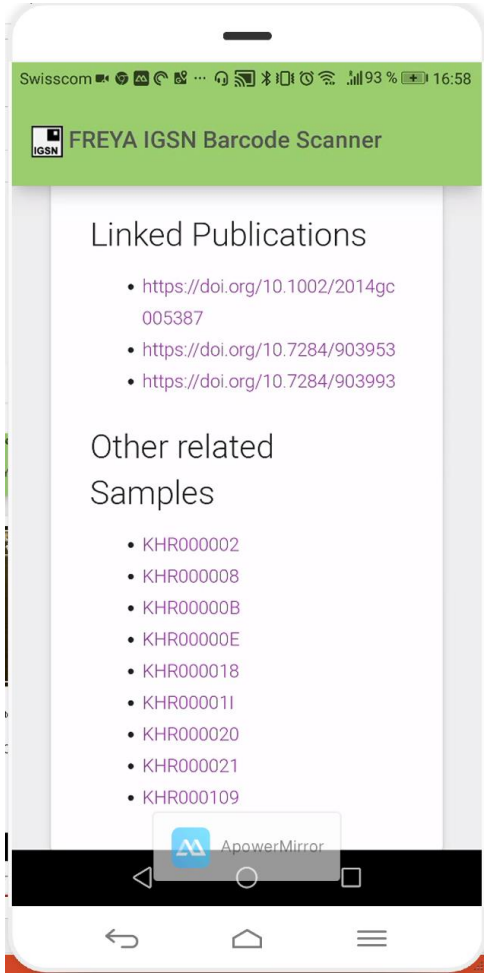
Thank you

www.project-freya.eu | twitter: @freya_eu

European Commission
Horizon 2020
European Union funding
for Research & Innovation

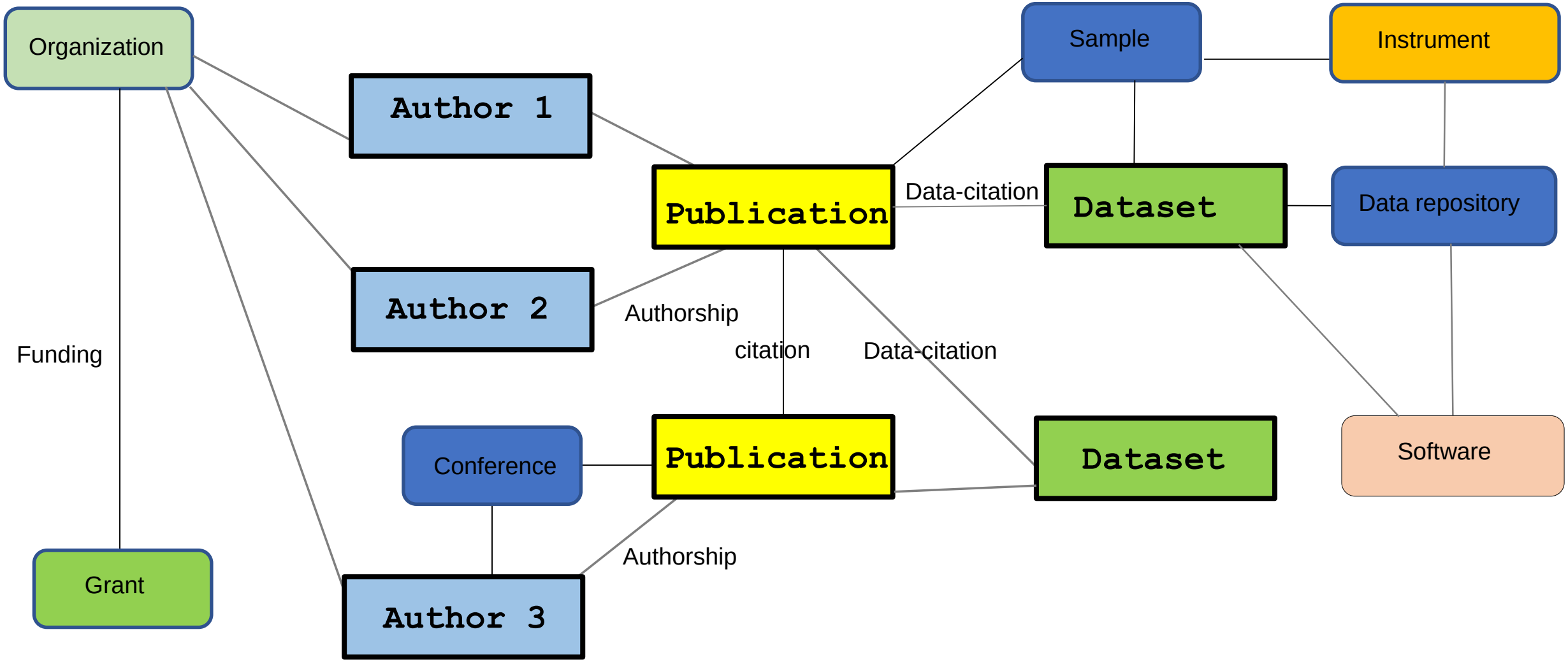

Swisscom 94 % 16:57
FREYA IGSN Barcode Scanner
Scanning by SCANDIT
Detected IGSN: KHR000022
Scientists
Rubin, Kenneth: https://orcid.org/0000-0002-8554-1337
ApowerMirror


Swisscom 94 % 16:58
FREYA IGSN Barcode Scanner
Detected IGSN: KHR000022
Scientists
Rubin, Kenneth: https://orcid.org/0000-0002-8554-1337
Funding
National Science Foundation, award/grant 0929881: http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0929881
Linked Datasets
Rubin, K. (2014). 210Po-210Pb Time series Isotopic Data and Eruption Ages for West Mata Boninite Lavas, 2008-
ApowerMirror


Swisscom 93 % 16:58
FREYA IGSN Barcode Scanner
Linked Publications
https://doi.org/10.1002/2014gc005387
https://doi.org/10.7284/903953
https://doi.org/10.7284/903993
Other related Samples
KHR000002
KHR000008
KHR00000B
KHR00000E
KHR000018
KHR00001I
KHR000020
KHR000021
KHR000109
ApowerMirror

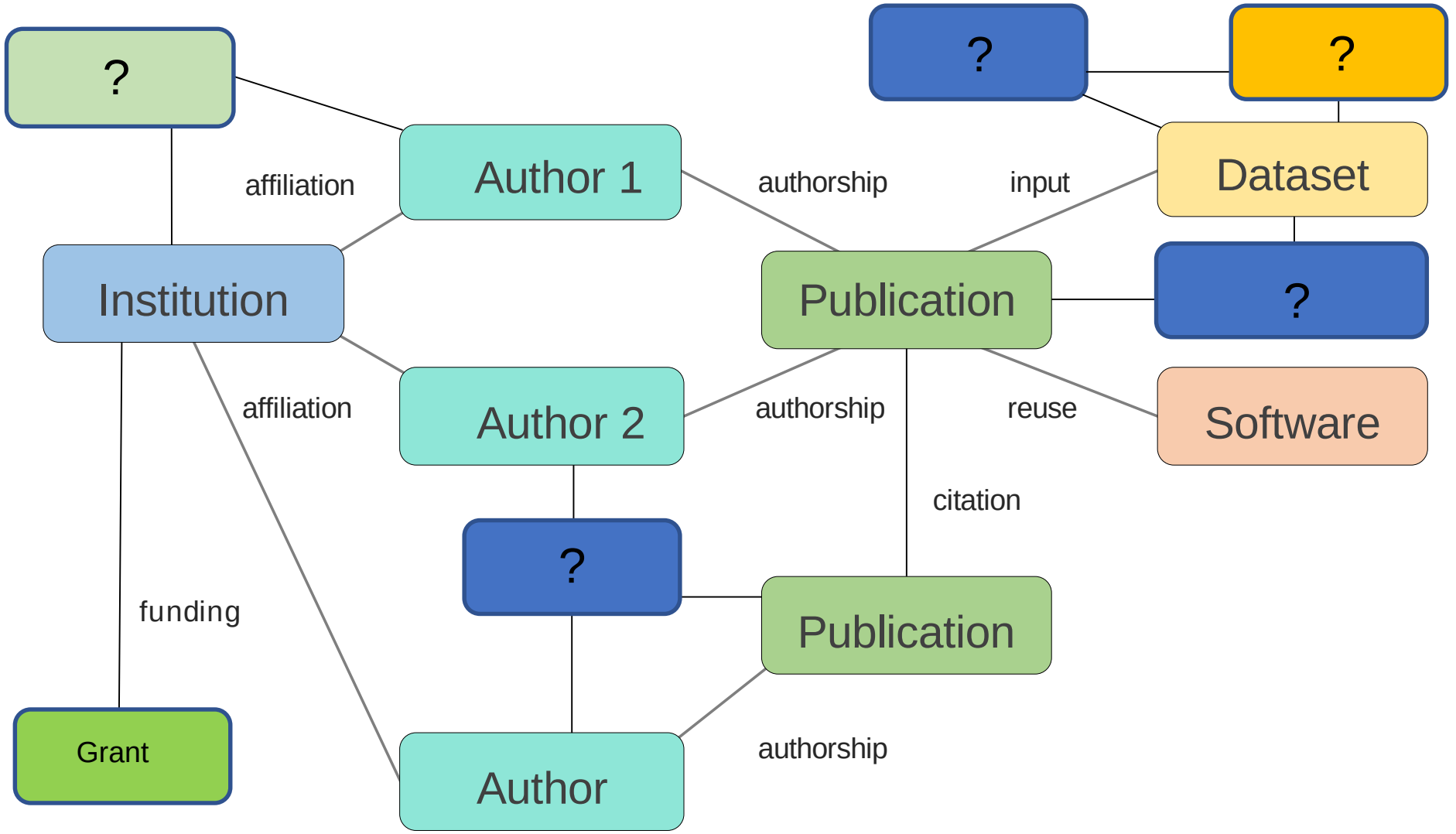
New PID Types (WP3)

Extending the PID graph with New PIDs



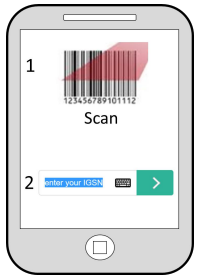
New PID Types (WP3)

Extending the PID graph with New PIDs



PID graph Functionality in Practice

FREYA IGSN Barcode scanner



- The app is a simple mobile webpage
- It uses the SCANDIT barcode Scanner and runs in any modern browser
- Once a barcode was scanned, it analyzes the text to find possible IGSN numbers. It mainly detects several formats like hdl.handle.net, igsn.org and other URL types.
- Once it finds an IGSN it starts a query to the DataCite JSON REST API using the IGSN as lookup key for dataset relations
- It formats all related datasets as a citation list using the DataCite citation formatter.
- It extracts all persons/scientists with ORCIDS, other IGSN identifiers and funding references.
- It extracts DOIs of publications