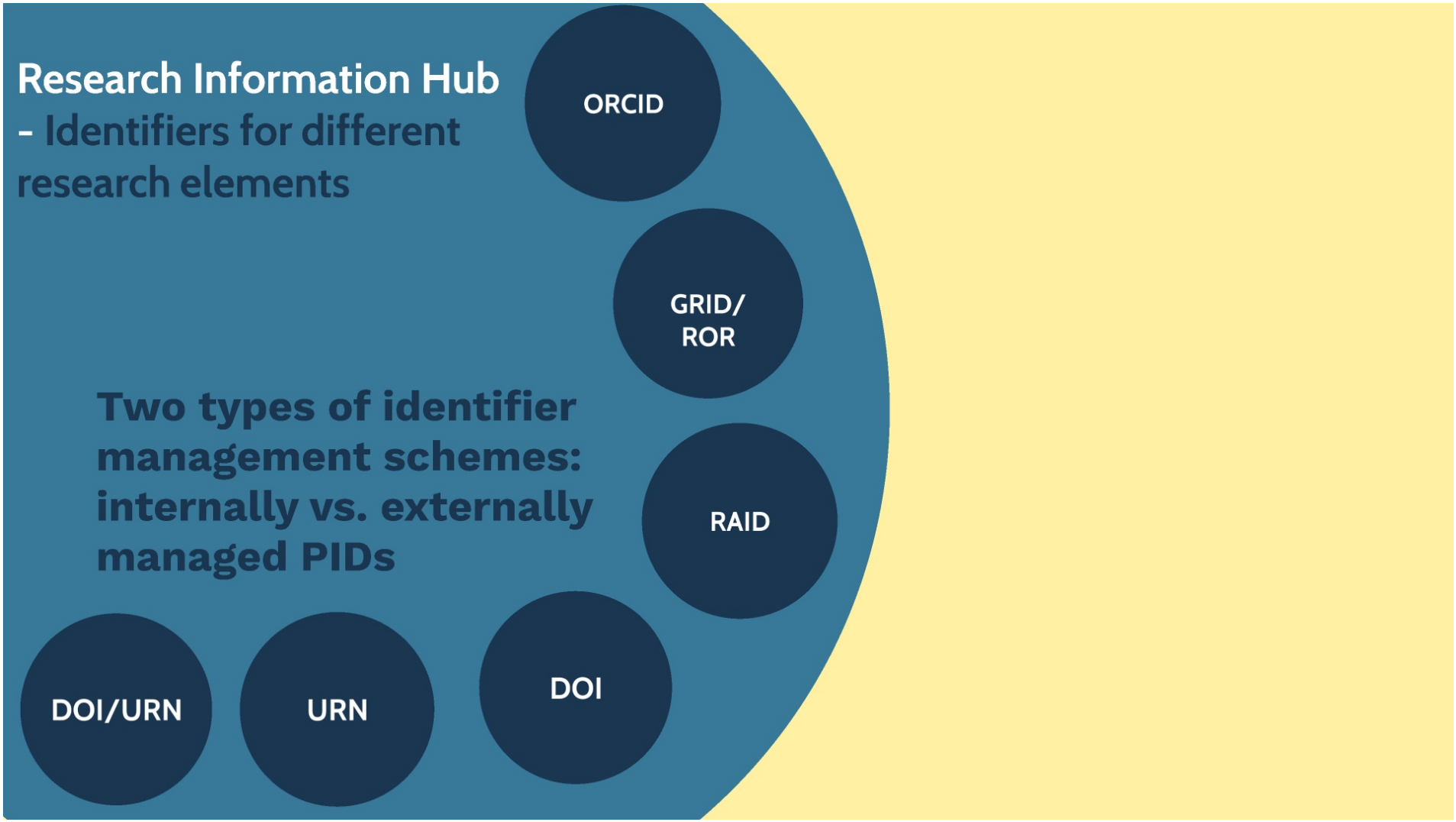
Unified Identifier Management and Resolution Services

Tommi Suominen, Jessica Parland-von Essen, Tiina Strengell
CSC - IT for Science
Pidapalooza 2020, Lisbon
30.01.2020

Research Information Hub
https://research.fi/

PID management service

Menti.com


Research Information Hub
- Identifiers for different research elements
ORCID
GRID/ ROR
Two types of identifier management schemes: internally vs. externally managed PIDs
RAID
DOI/URN
URN
DOI

ORCID: Researchers

- **ORCID** is the default identifier for researchers, created by researchers themselves, leading to inadequate coverage (approx 24% of new publications reported through research.fi have at least one ORCID ID)
- **ISNI** is also used by e.g. national libraries, generated based on publication activity, not administered by researchers themselves, leading to inadequate quality

GRID/ROR: Organisations

- research organisations (ROR 95.000, annual growth 5000)
- All organisations ISNI
- funding organisations (Funder ID)
- publishers (ISBN-publisher part)
- PIC for organisations that have received EU funding

RAID: Funding decisions/ projects

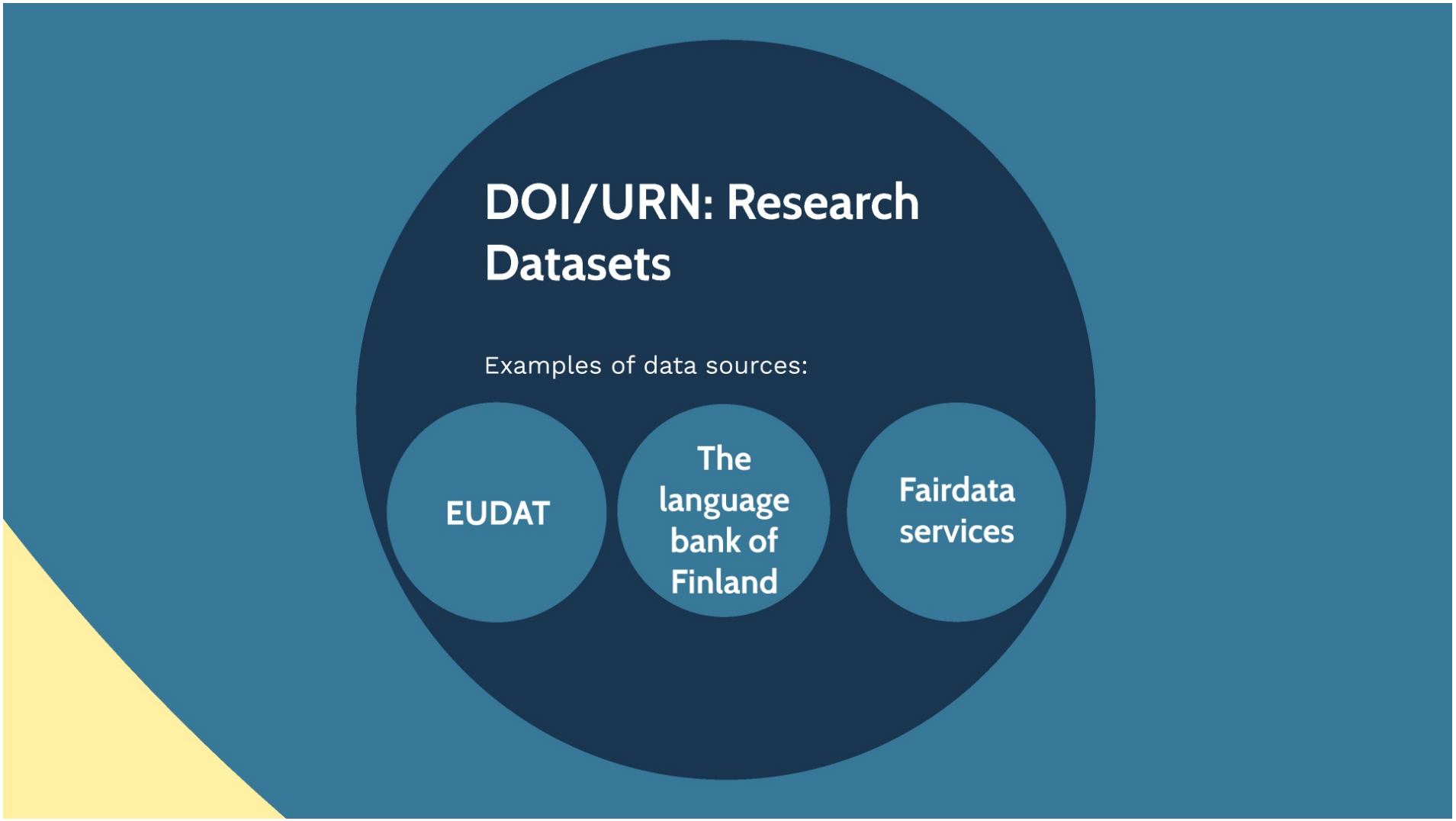
Several funding decisions may link to a the same project

DOI: Publications

- Most often DOI, but coverage is still incomplete, even for new publications (approx 63% of new publications we receive through VIRTA have a DOI)
- Also ISBN and ISSN, which are possible to convert into URN

URN: Infrastructures

Catalog for research infrastructures,
currently one URN per infrastructure


DOI/URN: Research Datasets
Examples of data sources:
EUDAT
The language bank of Finland
Fairdata services

EUDAT B2Share

- handle - for files
- DOI - for datasets

Language bank

- URN (metadata and landing pages)
- Handle (as above)

KIELIPANKKI
The Language Bank of Finland

fairdata.fi

- research datasets in storage and preservation (DOI)
- for metadata and external datasets (URN), but an external dataset can have an externally managed DOI as an alternate identifier

Unified Identifier Management and Resolution Services

Tommi Suominen, Jessica Parland-von Essen, Tiina Strengell
CSC - IT for Science
Pidapalooza 2020, Lisbon
30.01.2020

Research Information Hub
https://research.fi/

PID management service

Menti.com

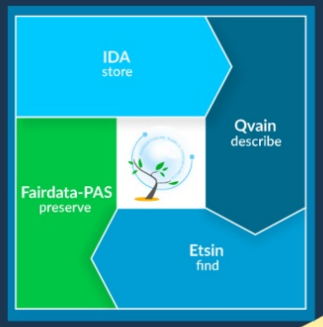

Registering services to the PID service
Harmonised suffix creation
Creating selected PID types
PID management service (under construction)
• Research datasets can have multiple (different types of) identifiers, and they move around between storage systems
• A dataset can first in a working phase have a URN, and once finalised get a DOI.
• The same dataset might be stored in different services using different PIDs, independently managed and assigned
• FAIR principles require a more managed approach to PID management
• CSC has a PID policy that requires the creation of the appropriate PID in the appropriate context
Validating that PIDS resolve
Interlinking PIDs pointing to the same item

CSC hosts several research and education services

- Short-term data storage (object storage)
- IDA for sharing and publication of research datasets
- International services (EUDAT B2Share)
- Long-term preservation (Fairdata PAS)
- FIN-CLARIN (Language bank)
- open educational resources (aoe.fi)
- Research Information Hub (research.fi)

KIELIPANKKI
The Language Bank of Finland

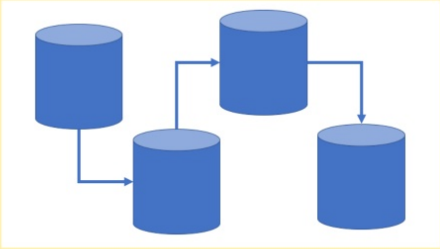
EUDAT



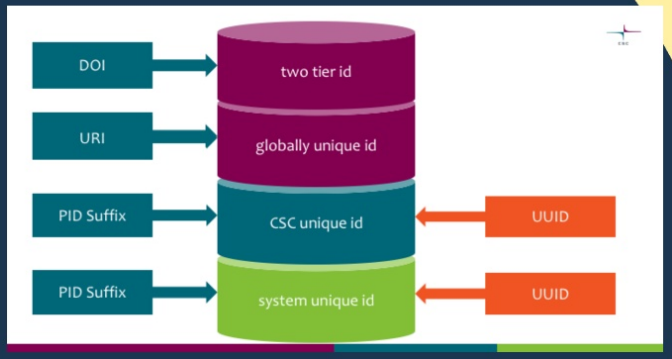
Harmonising suffixes across services

- suffix can function as an internal ID)
- ensuring ids are correctly formed
- enabling recognition of CSC created suffixes
- possibility to use the same suffix across different PIDs
- control digit for correctness validation
- supporting data life cycle management (tracking across services)

- E.g. suffix NRD-123457 -->
http://urn.fi/urn:nbn:csc:nrd123457
https://doi.org/10.34577/nrd123457



PID creation



- holds rules for PID syntax per registered service
- creates identifiers and then registers them
- starting off with URNs
- adding new PID types as we go
- the PID MS acts as the centralised requestor for all new ids of all types

Identifier mapping

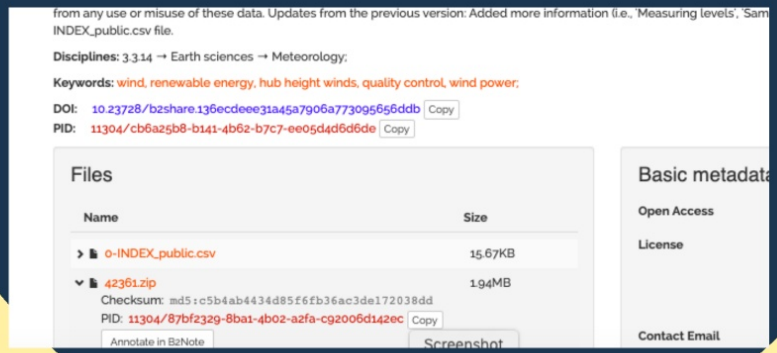
- Linking together identifiers that point to the same resource
- "Does this DOI have a URN?"
- Master metadata for datasets is always elsewhere

ID

URN

Quality checking

- Measuring adherence to FAIR criteria
- Checking that identifiers resolve
- Can warn if a service is about to create a PID for a resource for which a PID already exists



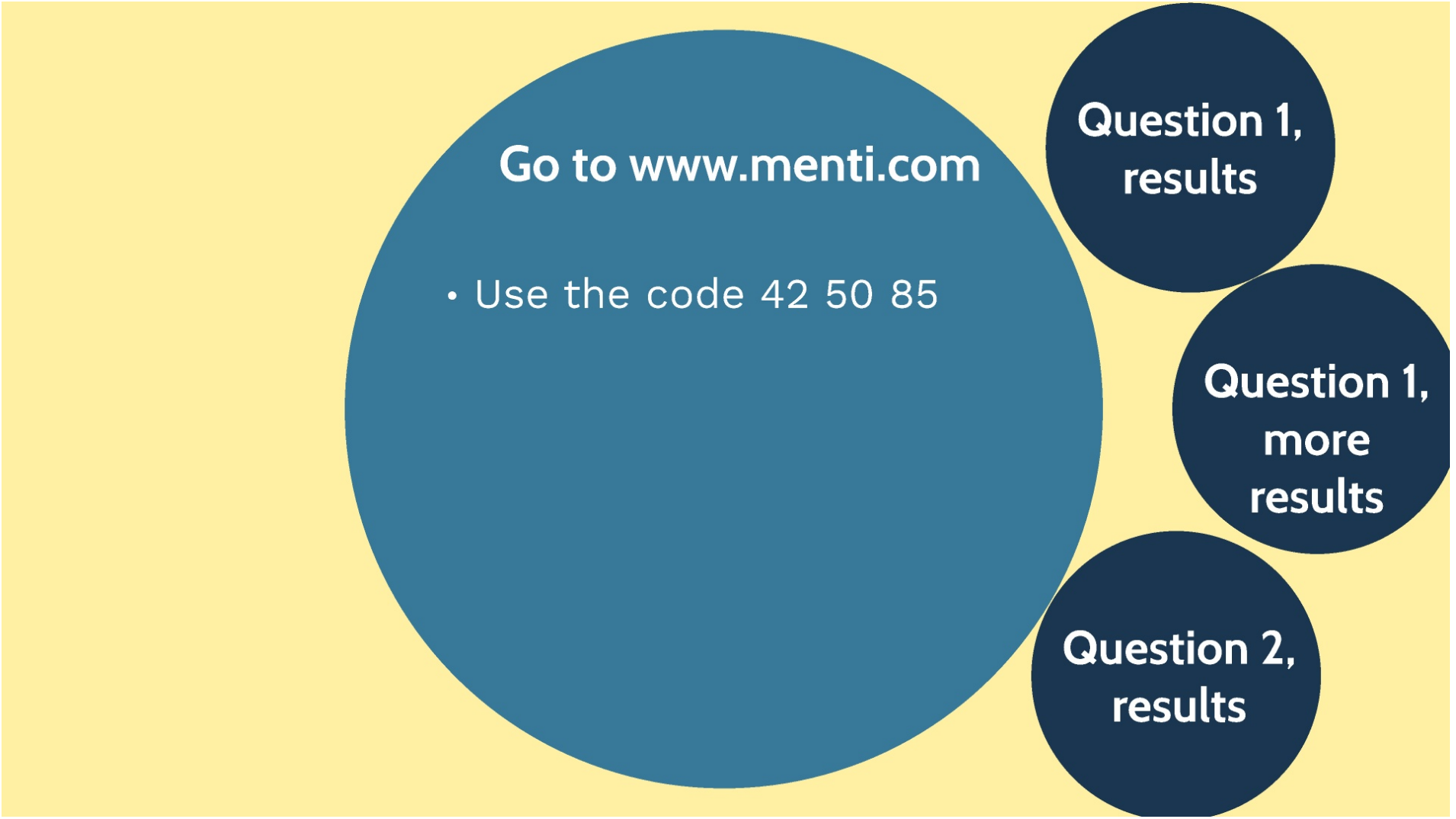
Unified Identifier Management and Resolution Services

Tommi Suominen, Jessica Parland-von Essen, Tiina Strengell
CSC - IT for Science
Pidapalooza 2020, Lisbon
30.01.2020

Research Information Hub
https://research.fi/

PID management service

Menti.com


Go to www.menti.com
• Use the code 42 50 85
Question 1, results
Question 1, more results
Question 2, results

Menti Results
Q1, Part 1

What kind of issues have you encountered having several PIDs for the same resource?

- Citation count misleading.
- Metadata reconciliation
- Maintaining multiple systems.
- Two version DOIs for the same dataset
- Reliability of the PID providers
- 2 DOIs for the same research object
- I've reported (to ORCID) a number of cases of authors with multiple ORCID's - these seem to have occurred either because the author created two entries several years apart, or their institution created one record and the author another.
- Implement about a global resolver that understands all the PID rules... Why not using Name-To-Thing... :)
- Confusion which one to use for citation

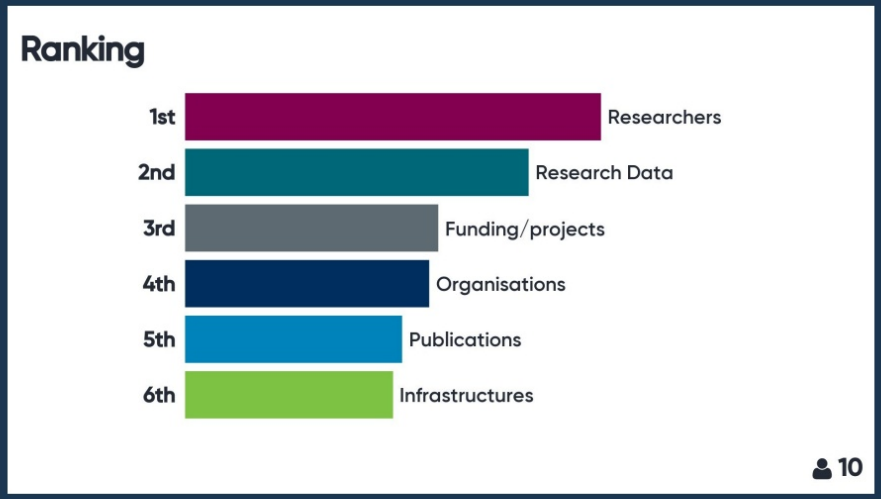
Menti Results
Q1, Part 2

What kind of issues have you encountered having several PIDs for the same resource?

- With updates, the situation gets messy quickly. Much harder to update many interlinked PIDs than just one.
- Different affiliations, tat should have been the same
- overlaps, mismatches, all sorts of data entry errors
- One of the PIDs stops resolving because a second one was minted as a replacement.
- none so far
- how to mange different versions of a paper?
- One of the PIDs is to the wrong item
- Each PID may provide different set of services, such as linking to different URLs. If e.g. a DOI is linked to a URL behind a paywall, users cannot access the resource if the DOI is the only PID in the metadata record is the DOI.

Menti Results, Q2

How would you prioritize the different identifier types, for the presented PID service?



Unified Identifier Management and Resolution Services

Tommi Suominen, Jessica Parland-von Essen, Tiina Strengell
CSC - IT for Science
Pidapalooza 2020, Lisbon
30.01.2020

Research Information Hub
https://research.fi/

PID management service

Menti.com