

Piloting the use of RAiD in research.fi

16 | 05 | 2024 by Tommi Suominen, Sonja Sipponen, Laura Himanen **CSC – IT Center for Science** & Clifford Tatum **SURF**, Shawn Ross **ARDC**



Funded by
the European Union



Structure of the presentation :

- Introduction of research.fi
- Introduction of RAiD
- Introduction of FC4E (context)
- Piloting the adoption of RaiD in research.fi
- Next steps
- It's all about governance...



Research Information Hub & Research.fi service

- The Ministry of Education and Culture (MINEDU) has authorized the implementation of the National Research Information Hub to CSC – IT Center for Science Ltd.
 - The project for Research Information Hub was launched in 2017
- The Research.fi service, launched in 2020, is the most visible part of the Research Information Hub
 - Collects information on research
 - Contains eg. statistics on the development of research resources and impact
 - Information on the Finnish research system



Search for information on research in Finland

Search target ▼

For example, publication, field of science, keyword

Q SEARCH

Search help ▼

 Publications 249 634

 Funding calls 244

 Granted funding 8 212

 People 9

 Research data 10 331

 Infrastructures 214

 Organizations 85

 Projects

Science and research in Finland



Researcher's Profile Tool

Create a public profile to Research.fi service or log in to your existing profile

Latest science and research news

Europe's flagship supercomputer LUMI top-listed once again – powers scientific breakthroughs in AI and more

13.05.2024

CSC - Tieteen tietotekniikan keskus Oy

Kutsu: Helsingin yliopiston EU-vaalipaneeli 16.5. klo. 16

Open funding calls

Apurahahaku taloustieteen 1. ja 2. vuoden jatko-opiskelijoille

Hakuaika päättyy 17.5.2024

Yrjö Jahnssonin säätiö sr

RCF - NIH joint call 2024: thematic areas of the Finnish Research Flag-ships

Aims of Research.fi



Provide landscape of Finnish research ecosystem and basis for science policy

- R&D analytics
- Visualizations
- Business Intelligence (BI)
- Landscape



Ability to search for Finnish research and researchers

- Research.fi portal
- Semantic search
- Linked data



Support the reuse of research information in other services and systems

- APIs
- Shared data models
- Vocabularies
- Reference data



Merit researchers and make their research output and activities more visible

- Researcher's Profiles
- Traditional research outputs
- Other research activities

What is RAiD?

RAiD is designed to address a key challenge faced by researchers, research administrators, funders, publishers, and others in the research ecosystem—maintaining consistent and up-to-date information on projects throughout the research lifecycle.

Affordances:

- enter once use multiple
- reporting
- reproducibility
- assessment, impact analysis



WHAT IS A RAID?

A RAiD is a persistent identifier for...

- Research projects and activities, linking organisations, people, inputs, and outputs to a project and providing key project information found nowhere else
- Governed by [ISO Standard 23527:2022](#)

A RAiD is not for...

- Grants
- Researchers
- Durable organisations or organisational units (teams, centres, groups, departments, etc.)
- Documents, papers, articles, books, recordings, or other digital objects
- Software or datasets
- Instruments
- Samples or specimens

HOW DOES RAID WORK?

A RAiD has two parts: The [RAiD identifier](#) and the [RAiD metadata record](#).

RAiD uses the [DOI system](#) to create [globally unique, persistent identifiers](#).

The [metadata record](#) includes other PIDs (where possible) for various project inputs and outputs like:

- Collaborators (people)
- Organisations (institutions)
- Grants, awards, and investments
- Infrastructure, tools, instruments, and services
- Data and software
- Publications, reports, and events

Where necessary, the [metadata record](#) captures project information found nowhere else, such as a project's title, description, and subject. Many RAiD elements can be time-bound.



What a RAiD metadata record looks like



FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

<https://raid.org/10.25.10.1000/j6k28>



Title (primary) Lorem Ipsum Project
Title (acronym) LIP
Start date 2023-01-30
Description (short) Lorem ipsum dolor sit amet, consectetur adipiscing elit.



Principle investigator orcid.org/0000-0002-3843-0000 **Date:**
Role <https://credit.niso.org/contributor-roles/conceptualization/>
Co-investigator orcid.org/0000-0002-3639-2080 **Date:**
Role <https://credit.niso.org/contributor-roles/data-curation/>



Lead organisation ror.org/00rqy9422
Partner organisation ror.org/03b94tp07

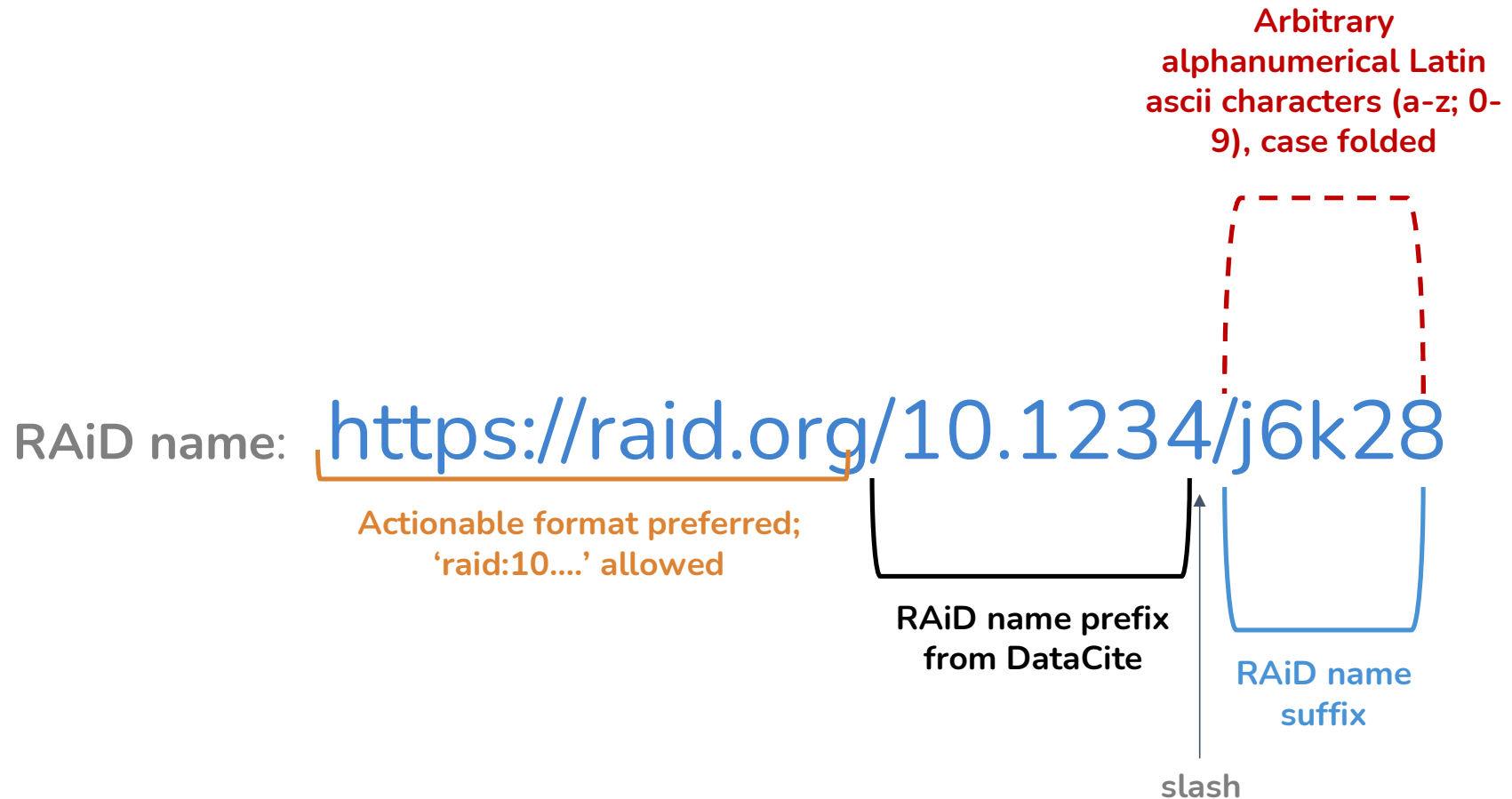


Grant doi.org/10.8948/908234D93EAF
Dataset doi.org/10.1594/PANGAEA.726855
Article doi.org/10.1038/nphys1170
Instrument doi.org/10.1337/jdlc-tima
Sample XXAB000IH



Local storage uq.edu.au/114/32
Cloud storage 79.152.127.243
Sub-project <https://raid.org/10.3010/401XQPOI>

RAiDs are now DOIs (via DataCite)



RAiD & PIDgraph

RAiD Core metadata is registered at DataCite, facilitating connections within the PID Graph

- **Near term solution:** PIDgraph will only include 'core' project metadata submitted to DataCite (including related PIDs)
- **Longer term:** project metadata is available via RAiD API and can be incorporated into the PID graph or can be retrieved via API for custom analysis / visualisation
- RAiD full metadata scheme: <https://metadata.raid.org/en/latest/> (will be updated shortly to reflect recent - minor - changes)

RAiD-Datacite core metadata mapping (draft)

RAiD Core Metadata	DataCite Metadata 4.4	DataCite Occurrence	Details
1 Identifier	1 Identifier	1-n	DOIs with prefix per RAiD structure 10.25.DD.DDDD/suffix
	2 Creator	1-n	identifier.owner.id and Organisation name (resolved from ROR)
2 Date	8 Date	0-n	date.startDate
3 Title	3 Title	1-n	title.text (Primary)
	4 Publisher	1-n	identifier.registrationAgency.id Organisation name (resolved from ROR)
	5 PublicationYear	1-n	date RAiD created
	6 ResourceType	1-n	type 'Other' and 'RAiD Project'
4 Description	17 Description	0-n	description.text (Primary)
5 Contributor	7 Contributor	0-n	contributor.id and Contributor name (resolved from ORCID)
6 Organisation	7 Contributor	0-n	organisation.id and Organisation name (resolved from ROR)
7 RelatedObject	12 RelatedIdentifier	0-n	relatedObject.id with type 'HasPart' if it is an 'output'.
8 AlternateIdentifier	11 AlternateIdentifier	0-n	alternateIdentifier.id
9 AlternateURL	12 RelatedIdentifier	0-n	alternateUri.uri with type 'IsSourceOf'
10 RelatedRAiD	12 RelatedIdentifier	0-n	relatedRaid.id with type 'HasPart'
11 Access	16 Rights	0-n	identifier.license and access.type.id and

Identity management & log-in (<https://app.demo.raid.org.au>)

ARDC Research Activity Identifier (RAiD)

DEMO

This is the Oceania region implementation of the [RAiD](#) ISO standard.

Maintained by the [ARDC](#).

[PRIVACY STATEMENT](#) [USAGE TERMS](#)

RAiD Sign-in



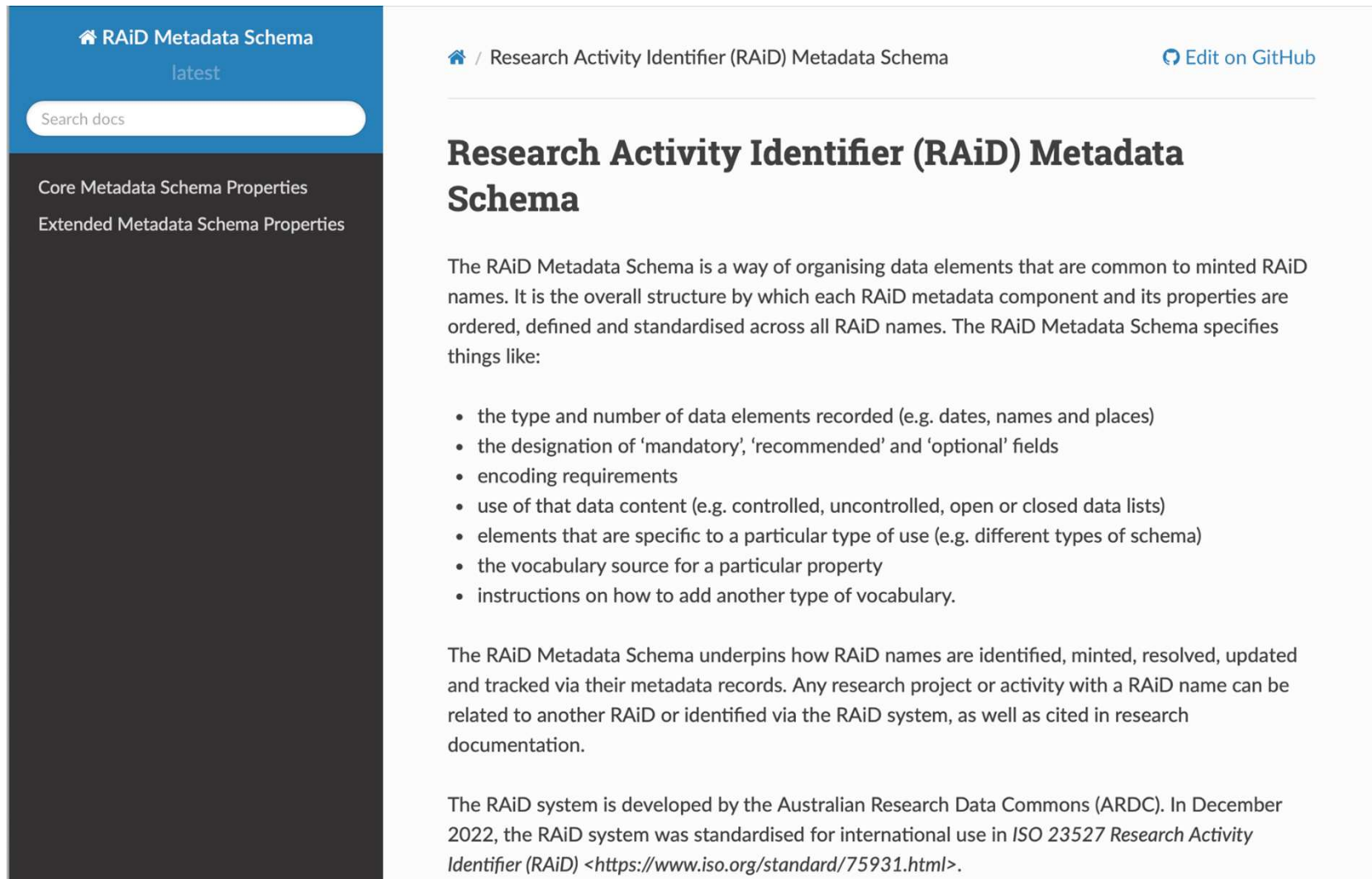
Please select your preferred sign-in method

 GOOGLE

 AAF

 EDUGAIN

 ORCID



The screenshot shows the RAID Metadata Schema website. The left sidebar is dark blue with a search bar and navigation links. The main content area is white with a header bar and a detailed introduction to the schema.

RAiD Metadata Schema
latest

Search docs

Core Metadata Schema Properties
Extended Metadata Schema Properties

[/ Research Activity Identifier \(RAiD\) Metadata Schema](#) [Edit on GitHub](#)

Research Activity Identifier (RAiD) Metadata Schema

The RAiD Metadata Schema is a way of organising data elements that are common to minted RAiD names. It is the overall structure by which each RAiD metadata component and its properties are ordered, defined and standardised across all RAiD names. The RAiD Metadata Schema specifies things like:

- the type and number of data elements recorded (e.g. dates, names and places)
- the designation of 'mandatory', 'recommended' and 'optional' fields
- encoding requirements
- use of that data content (e.g. controlled, uncontrolled, open or closed data lists)
- elements that are specific to a particular type of use (e.g. different types of schema)
- the vocabulary source for a particular property
- instructions on how to add another type of vocabulary.

The RAiD Metadata Schema underpins how RAiD names are identified, minted, resolved, updated and tracked via their metadata records. Any research project or activity with a RAiD name can be related to another RAiD or identified via the RAiD system, as well as cited in research documentation.

The RAiD system is developed by the Australian Research Data Commons (ARDC). In December 2022, the RAiD system was standardised for international use in *ISO 23527 Research Activity Identifier (RAiD)* <<https://www.iso.org/standard/75931.html>>.

RAiD API documentation (<https://api.demo.raid.org.au/swagger-ui/index.html#/raido-stable-v1>)

 **Swagger**
Supported by SMARTBEAR

/docs

Explore

Servers

- Default Server URL

Authorize

service-point the service-point API

GET

/service-point/{id}

✓

🔒

PUT

/service-point/{id}

✓

🔒

GET

/service-point/

✓

🔒

POST

/service-point/

✓

🔒

raid the raid API

GET

/raid/{prefix}/{suffix}

Read a raid

✓

🔒

PUT

/raid/{prefix}/{suffix}

Update a raid

✓

🔒

GET

/raid/

List raids

✓

🔒

POST

/raid/

Mint a raid

✓

🔒

GET

/raid/{prefix}/{suffix}/history

A list of base64 encoded changes to the raid in JSON Patch (RFC 6902) format.

✓

🔒

Video Demo



Dissemination Materials

- Component Demo Video
- Watch the Demo Video below!



RAiD

FAIRCORE4EOSC in a nutshell

- today at month 24/36

Full name: Developing EOSC-Core components to enable a FAIR EOSC ecosystem

Research and Innovation Action

Budget: 10 million EUR

Duration: June 2022 – May 2025

Consortium: 22 partners, coordinated by CSC – IT Center for Science

Coordinator Tommi Suominen (CSC), **Project Manager** Anu Märkälä (CSC) and
Technical Coordinator Mark van de Sanden (SURF)

Website: faircore4eosc.eu

Key results: In response to the gaps identified in the SRIA, the project will develop nine new EOSC-Core components aimed to improve the discoverability and interoperability of an increased amount of research outputs.



Context

Enhancing FAIRness in the EOSC ecosystem

The European Open Science Cloud (EOSC) is an ecosystem of research data and related services that will enable and enhance seamless access to and reliable re-use of FAIR research objects (including data, publications, software, etc.).

The Strategic Research and Innovation Agenda (SRIA) for EOSC was created in 2021, as a roadmap for future development. Priorities highlighted in the SRIA are the establishment of the Web of FAIR data and a Minimum Viable EOSC (MVE) by 2027, that is the core components and functions to enable EOSC to operate (the EOSC-Core).



2021



Minimum Viable 

Web of FAIR Data

Findable Accessible Interoperable Reusable



2027

The 9 FAIRCORE4EOSC components



EOSC Research Discovery Graph (RDGraph) to deliver advanced discovery tools across EOSC resources and communities.



EOSC PID Graph (PIDGraph) to improve the way of interlinking research entities across domains and data sources on the basis of PIDs.



EOSC Metadata Schema and Crosswalk Registry (MSCR) to support publishing, discovery and access of metadata schemas and provide functions to operationalise metadata conversions by combining crosswalks.



EOSC Data Type Registry (DTR) to provide user friendly APIs for metadata imports and access to different data types and metadata mappings.



EOSC PID Meta Resolver (PIDMR) to offer users a single PID resolving API in which any kind of PID can be resolved through a single, scalable PID resolving infrastructure.



EOSC Compliance Assessment Toolkit (CAT) to support the EOSC PID policy compliance and implementation.



EOSC Research Activity Identifier Service (RAiD) to mint PIDs for research projects, allowing to manage and track project related activities.



EOSC Research Software APIs and Connectors (RSAC) to ensure the long-term preservation of research software in different disciplines.



EOSC Software Heritage Mirror (SWHM) to equip EOSC with a mirror of the Software Heritage universal source code archive.

Aims of piloting RAiD in Research.fi

- The main goal is to gain insights on how this type of identifier could work in the context of Research.fi and how to adopt it to the Finnish landscape
 - Data available of research projects in Finland
 - Project activities as a form of supporting funding
 - Impact analysis
- RAiD datamodel as a basis for designing our model for research projects in Finland
- As a part of the piloting Research.fi takes over the content of a legacy repository of over 12 000 descriptions of research projects from 'Projectnet' as a case example
 - Projectnet was a service that collected information about research projects focusing on natural resources from six Finnish research organizations between 2008-2023



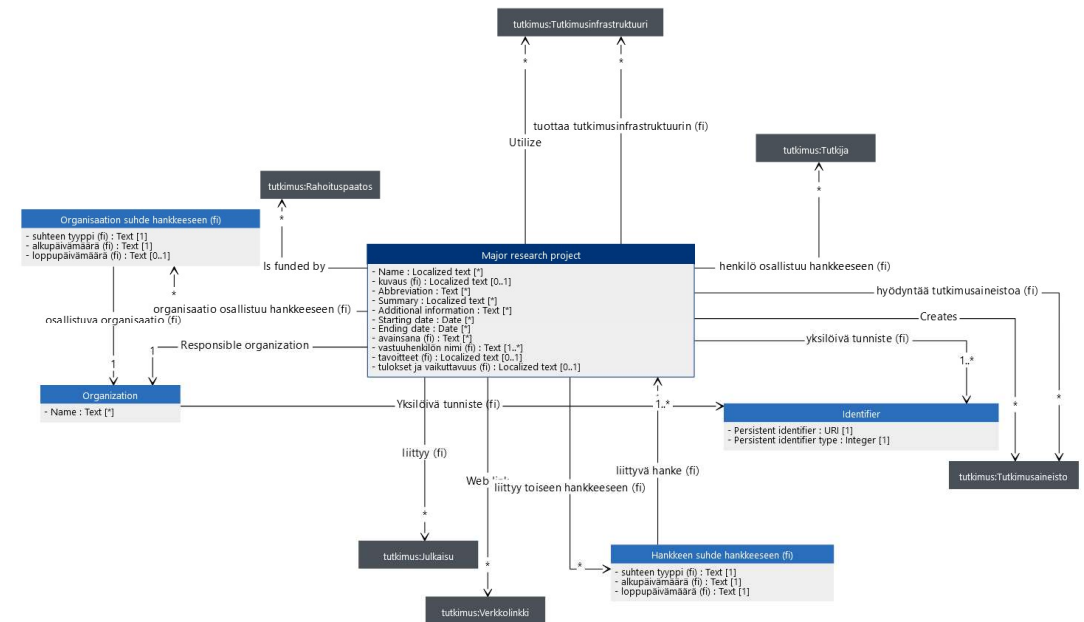
Challenges faced in piloting

- In Finland projects are mainly thought to be owned by research organizations as in RAiD model the people as contributors have more weight
 - Research.fi datamodel has previously always collected information on people affiliated to organizations as contributors
- Main challenges faced are concerning legacy data from 'Projectnet' which was not created this use in mind
 - Legacy data lacks persistent identifiers necessary to create needed links between project outputs and activities and for example organizations and people involved in the project
 - For example the lack of ORCIDs for people involved in the Projectnet-data required ARDC to make an exception to their requirements
- Concerns regarding organisations ability to upkeep project information with all the variables in the future
 - Current legacy data will, for the most part, not be updated due to it's nature



Results of piloting so far

- First draft of research project datamodel created for Research.fi based on RAiD
(<https://tietomallit.suomi.fi/model/mrp2024x1/>)
 - Mapping out previous questionnaires about information collected and Projectnet-data to RAiD datamodel
 - Key mandatory information identified to fit RAiD datamodel
 - Key related entities already collected in Research.fi
- Missing persistent identifiers
 - Manual labour to identify organizations involved in projects
 - ARDC making an exception to contributors on missing person identifiers
- First projects have been given their RAiD
 - Not yet adopted to all Projectnet-data



Benefits of RAiD so far

- Based on the piloting, we believe that the very versatile RAiD datamodel with a few modifications could function well in all Finnish research organizations
 - Modifications needed for the Finnish research project model:
 - Enabling organisations to provide abbreviations to projects
 - (possibly contributor affiliation)
 - Enabling detailed funding information to be linked to projects without persistent identifiers
 - Mapping ROR identifiers to other commonly used identifiers in Finnish landscape
 - Creating links between projects and outputs in Research.fi thus enriching the project entity
 - Description of results and impact
- RAiD datamodel has also provided important input to defining different types of connections between research entities in Research.fi



Next steps of RAiD implementation in Finnish landscape

- Finishing manual enrichment of legacy Projectnet-data, minting RAiDs to rest of legacy data and publishing the data on Research.fi portal
- Setting up a working group of key stakeholders to finalize the Research.fi research project model and pilot data transfers on based on the finalized model
 - Finalized model to be based on the work done based on a joint international standard
 - Output: the widely accepted data model for research projects in Finland that proactively promotes interoperability nationally and internationally
- Other key task for working group in collaboration with the Finnish PID working group to set up a governance model to RAiDs in the Finnish landscape
 - Research.fi can mint persistent identifiers for project entities transferred but can't maintain the actual content
 - Materdata is owned by research organizations
 - Research organizations must be willing to update project information



Conclusions: all (Sonja, Clifford, Laura, Shawn – Kyle?)

When the technical parts are set in the pipeline, its time to think about the practicalities

ORCID as mandatory information a challenge, for the legacy data (=old data) but also in the context of universities of applied sciences where project leaders might not identify themselves as someone who should have an ORCID.





We are FAIRCORE4EOSC !

