Research data publishing

Role of the libraries

Mari Elisa Kuusniemi (a.k.a MEK), M.A, Science Information Specialist, Data Support, Helsinki University Library, ORCID: 0000-0002-7675-287X

Susanna Nykyri, PhD, Chief Specialist (Open Science), Tampere University Library, Publishing Manager @TUP_UniPress, ORCID: 0000-0002-5018-5176

Agenda

- What makes data research data and on LIBER Open Science Roadmap
- Data publishing services commonly already in place and what else is needed
- Different publishing models and conclusions

On research data

"Research data is a resource used by a researcher or a research group during a research process, i.e. the basic data of scientific or artistic research, in digital, analogue or physical form. Research data has been collected, observed, measured or created to confirm hypotheses and verify results."

Research data includes also research methods, software, code, infrastructures and alike. [DOI: https://doi.org/10.23847/isbn.9789525995480]

→ The role of the context!

Open data is not necessarily data publication. Data as a publication requires following commonly accepted practices, as citedness of data. [DOI: https://doi.org/10.23978/inf.109094]

Research data
as research publication
as meriting research output?

Great variety of national
and international actors involved, e.g.

DORA

SCIENCE EUROPE

ACADEMY OF FINLAND

Open Science

FINNISH NATIONAL BOARD ON RESEARCH INTEGRITY TENK

DMP Tuuli

LIBER Open Science Roadmap

ADOPT technology and provide services which support Open Science

- Provide a certified repository.
- Create a data catalogue.
- Publish content with a machine-readable license.
- Use open APIs to provide access to library services.
- Develop intelligent tools to automate metadata production and support FAIR data management during the entire data life-cycle.

Source: Ayris, et al. “LIBER Open Science Roadmap”. Zenodo, July 2, 2018. https://doi.org/10.5281/zenodo.1303002.

PLUS: Publish data produced by the library!

Data publishing services already in place

Data repository services

- Curated collections and/or not curated
- Dedicated data repositories are either self-hosted and outsourced, like
  - Dataverse
  - Figshare
  - EUDAT B2Share
- Or library can recommend to use national or international repositories

Long-term preservation of data sets

- Requires a data curation process
  - Preservation for long-term is expensive and some kind of selection of data is needed.
- Library can have an essential role on evaluation of documentation & metadata of the data set
  - Guidance, training and support for researchers
- Data curation process is a great opportunity to engage with researchers

Data catalogues

- help the discovery of datasets created or used by the organisation's researchers
- collect and highlight what is happening at the organisation
- promote interdisciplinary collaboration by identifying common research activity
- support the practice of data re-use

Can be established as part of CRIS or as a dedicated system for data sets

Source: [Research Data Discovery - Developing a Data Catalog at the University of Maryland, Baltimore]

NYU HEALTH SCIENCES LIBRARY

About How To Use the Catalog Frequently Asked Questions

NYU Data Catalog

Search

Filter by

NYU Datasets Only

Subject Domain

COVID-19	23
Risk Factors	12
Health Status	11
Population Characteristics	8
Health Care System	6
Health Care Utilization	4
Genomics	3
Chronic Disease	2

Keyword: covid

Page 1 of 3 Next >

Results Found: 27

Sort by: relevance Per page: 10

MIDAS Online Portal for COVID-19 Modeling Research

Description The MIDAS (Models of Infectious Disease Agent Study) Online Portal for COVID-19 Modeling Research is a collection of publicly-available COVID-19 resources to support dashboard monitoring, data processing, modeling, and visualization efforts. Collections listed in the portal include case counts and case line lists with documented metadata, peer-reviewed and non-peer-reviewed parameter estimates, and...

Subject COVID-19
Health Status

Access Rights Free to All

Other data publishing services

Data journals

- Data journals are **scholarly journals that publish datasets or data papers**.
- "a data paper describes a dataset, giving details of its collection, processing, software, file formats etc, without the requirement of novel analyses or ground breaking conclusions." *Geoscience Data Journal*
- Do you have a platform for research journals (university press)? Do you already have a data journal? Do you promote this option?

SOFTWARE TOOL ARTICLE

preCICE v2: A sustainable and user-friendly coupling library [version 1; peer review: awaiting peer review]

Gerasimos Chourdakis *, Kyle Davis *, Benjamin Rodenberg *, Miriam Schulte *, Frédéric Simonis *, Benjamin Uekermann *, Georg Abrams, Hans-Joachim Bungartz, Lucia Cheung Yau, Ishaan Desai, Konrad Eder, Richard Hertrich, Florian Lindner, Alexander Rusch, Dmytro Sashko, David Schneider, Amin Totounferoush, Dominik Volland, Peter Vollmer, Oguz Ziya Koseomur

This article is included in Excellent Science gateway

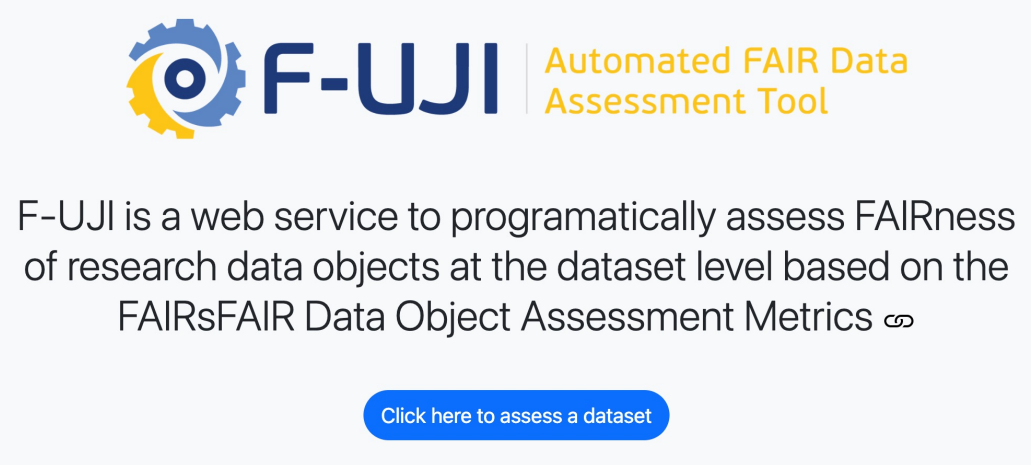
Article Authors Metrics

Abstract

preCICE is a free/open-source coupling library. It enables creating partitioned multi-physics simulations by gluing together separate software packages.
This paper summarizes the development efforts in preCICE of the past five years. During this time span, we have turned the software from a working prototype -- sophisticated numerical coupling methods and scalability on ten thousands of compute cores -- to a sustainable and user-friendly software project with a steadily-growing community. Today, we know through forum discussions, conferences, workshops, and publications of more than 100 research groups using preCICE. We cover the

Domain specific data repositories

- Do you know the domain specific data repositories or big data collections of your organisation?
- Can you offer any support for them?
  - PID service (DOI, URN…)
  - Ensure discovery (make sure they can be found from Re3data, FAIRsharing.org)
  - Promote Core Trust Seal serticifate & find help to apply certification
  - Help with FAIRification



https://www.f-uji.net/

FAIR advocacy

FAIR advocacy could be part of data publishing support services

- Integral part of data curation process
- Utilize FAIR metrics tools like F-UJI as a starting point of discussion

Motivation?

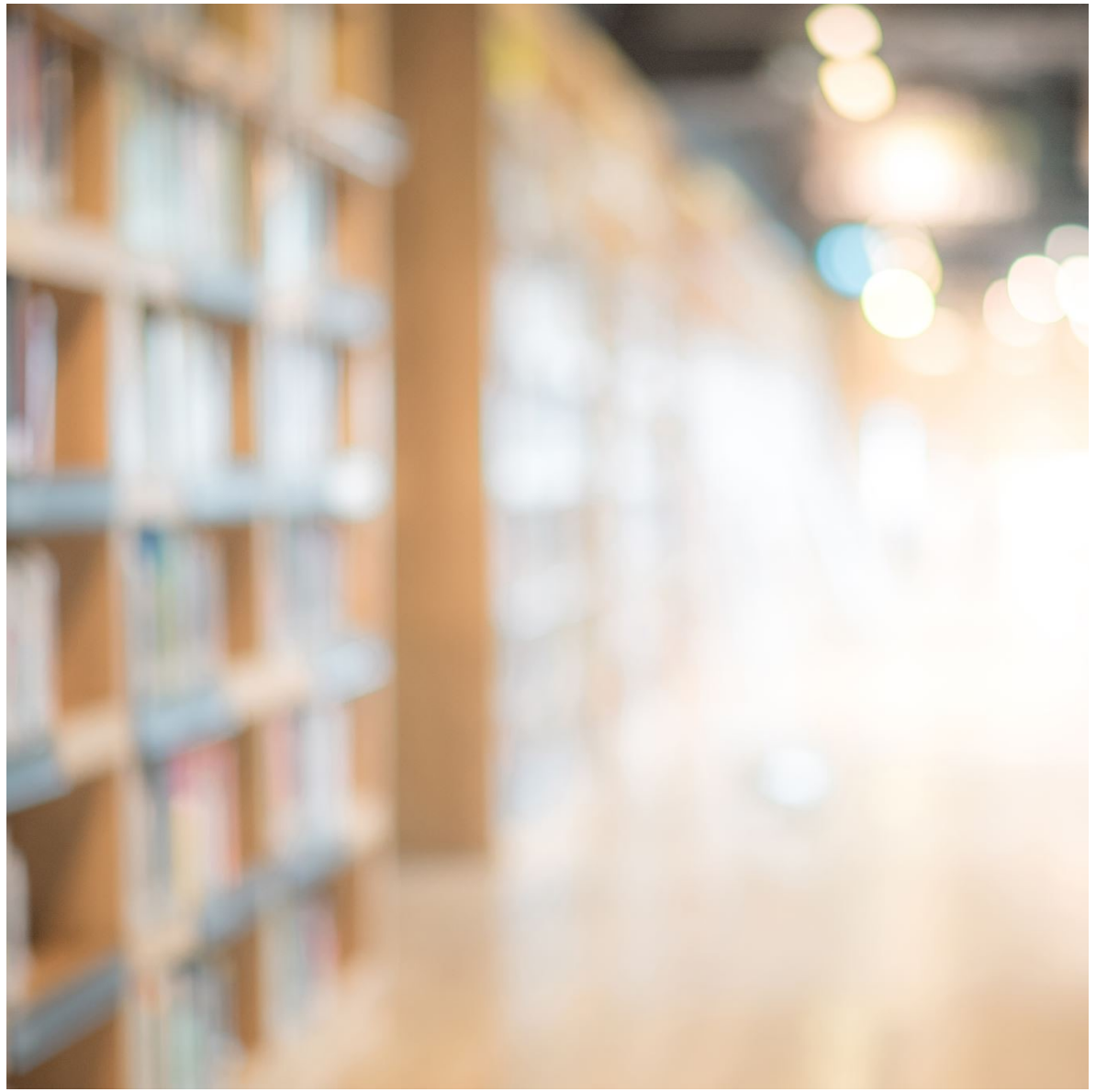
- Quality and impact of research & Research assessment systems & possibilities to enable responsible OS
- Metrics? (In Finland: research.fi, plans to include more quantitative indicators related to open data in the next national OS monitoring) - What should be measured and awarded?
- We're not there yet!
- Libraries: as experts in (biblio)metrics and in relation to developing of meriting practices, there is need for active input in the development of altmetrics and datametrics (tools, data, methods, skills)

Hot topic: New skills and specialised roles needed

- How skills needed related to metadata of research data differ from traditional metadata skills?
  - **Coaching** data documentation, **creating** new practices vs. following the cataloguing rules
  - Providing **inspiration** to see the effort
- What kind of skills you need for FAIRification support services?
  - Motivator, interpreter, expertise on selection of relevant & easy to use tools, etc.
- Skills as research output publisher?
  - Discipline & data type specific needs
- What kind of skills you need for long-term preservation services?
  - Data curation, data documentation & metadata
  - Data management (GDPR, agreements, transfer of rights, etc.)
  - Data archiving

Conclusions

Aims and possible service models for the research libraries as research data publishers



Service	Role of the library	Service models	Service providers
Data repository	Make possible to publish data in a data repository.	Create a institutional repository OR recommend to use a spesific data repository	National services, international repositories, domain specific repositories
Long-term preservation archive	Make possible to preserve valuable data sets in a long-term preservation archive. Provide curation of metadata.	Create an institutional service OR use national or international data archive	National services, international data archives, domain specific data archives
Data catalogue	Collect metadata of published data sets	Create a institutional data catalogue OR use national or international catalogue	National services, PID services collect metadata, domain specific data catalogues
Data journals	Promote data journals and data papers	Part of university press OR use outsourced OA journal platforms	Institutional and international journal publishing platforms
Support for domain specific repositories	Ensure support for domain specific repositories or data collections in your organisation	Provide support by your self OR find experts from outside the organisation	EOSC, FAIRsFAIR, Digital Competence Centers

1) Centralised model

A) Own repository

B) Peer-reviewed data journal

2) Distributed model

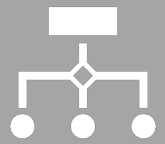
- Utilise services outside of the organisation
- Critical role in ensuring all the needed services and tools are available – and used (awareness, skills)

Both models require substantial support and services!

- Data management skills and roles (guidance, education, training etc)
- FAIR data
- Own data catalogue (metadata)

Final aims of a research library are

A) to enable research data publishing for all the disciplines covered within the university



B) to take care of the needed integrations, processes and workflows.

Main challenges

Centralised model

- Completely centralised not possible
- The awareness and activity of researchers more critical, may end-up to isolation and the "best" data not published and/or not documented on university data catalogue

Distributed model

- Know-how on "data pearls"
- Relies more on the skills and know-how on the work (being) done on the field – more active model for the support services

Thank You!

Tampere University

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

mari-elisa.kuusniemi@helsinki.fi

susanna.nykyri@tuni.fi