

ELIXIR Report of the 21st RDA Plenary



ELIXIR RDA Focus Group

RDA Plenary 21 & SciDataCon 2023 report

Report Authors

- Fabio Liberante, ELIXIR Hub, [0000-0002-0192-5385](#)
- Fotis Psomopoulos, ELIXIR-GR, INAB|CERTH, [0000-0002-0222-4273](#)
- Gavin Farrell, ELIXIR Hub, [0000-0001-5166-8551](#)
- Marek Suchánek, ELIXIR-CZ, CTU in Prague, [0000-0001-7525-9218](#)
- Paulette Lieby, ELIXIR-FR, [0000-0002-9289-9652](#)
- Peter Maccallum, ELIXIR Hub, [0000-0001-5260-5915](#)
- Sveinung Gundersen, ELIXIR-NO, [0000-0001-9888-7954](#)
- Wolmar Nyberg Åkerström, ELIXIR-SE, NBIS, [0000-0002-3890-6620](#)

Introduction

This document has been prepared by [ELIXIR's RDA Activities Focus Group](#) to showcase the synergies and activities of the [Research Data Alliance \(RDA\)](#) which may be useful for ELIXIR members operating in the life sciences domain. The RDA was launched as a community-driven initiative in 2013 with the goal of building the social and technical infrastructure to enable open sharing and re-use of data. The ELIXIR RDA Activities Focus Group has prepared twenty reports of [RDA Plenary events](#) to date containing overviews of highlighted RDA recommendations and outputs. This report contains highlights from various sessions of the RDA 21st Plenary which took place during the International Data Week 2023: A Festival of Data as a hybrid event in Salzburg, 23-26 October, 2023. Additional links to collaborative note documents and/or abstracts provided by the session chairs are included.

About the 21st RDA Plenary

The 21st RDA Plenary was a fully hybrid event with virtual and onsite participants interacting and accessing recordings of all sessions using the [Whova](#) conferencing platform. The event co-located with International Data Week (IDW) 2023 was hosted by the University of Salzburg through its Data Science and Geoinformatics departments, supported by the Governor of Salzburg and with assistance from the Austrian Academy of Sciences - GIScience and the European Umbrella Organization for Geographic Information.

The organisers reported impressive engagement in this year's event:

- 834 participants (702 onsite and 132 online)
- Attendance from 48 countries
- 6 Plenary meetings
- 45 Breakout sessions

Following this year's successful event the upcoming [22nd RDA Plenary](#) will take place virtually 14-23 May, 2024.

RDA Group Acronyms

- Working Group - WG
- Interest Group - IG
- Birds of Feather - BoF
- Community of Practice - CoP

Summaries Section

Overview

This section contains shorter highlight recommendations for ELIXIR members from the event and other key links and information from the various RDA 21st Plenary sessions of interest to the life sciences community.

Highlight RDA sessions, progress & outputs for ELIXIR members

- **Secure Research Environments BoF**
 - Members likely to apply for RDA WG status. This could be a good group to build upon EOSC-ENTRUST project work, share TRE related outputs, and align to the work of ELIXIR on TREs through the work in the Compute Platform WP3, 2024-2026.
 - [Notes link](#), page 33
- **BoF Let's talk about FAIR mappings!**
 - Working towards common practices for sharing mappings and crosswalks was very well-attended and will result in a working group proposal federating activities across several RDA IGs/WGs.
 - FAIRCORE₄EOSC/FAIR Impact projects related session.
 - Several ELIXIR folk present at session: cross-linked via their EOSC projects.
 - [Notes link](#), page 30
- **FAIRification of Genomic Tracks: data-driven life science through granular discovery of biological sequence annotations with uniform metadata**
 - New [WG initiative](#) in extension of the FAIRtracks project (which is an ELIXIR RIR service), with support from the [RDA Tiger project](#).
 - Good to signpost this session to ELIXIR 's life science genomic data members.
 - Challenges getting folk in person due to plenary cost - but [recordings are available](#).
 - Good discussion and group engagement.
 - [Notes link](#), page 23
- **Pro₄rs RDA & ReSA: Policies in Research Organisations for Research Software (PRO₄RS)**
 - ELIXIR STEERS & EVERSE projects noted and linkages to this group.
 - Research software focused.
 - Convergence on policies to increase visibility and recognise research software.
 - 18 months to deliverable point via WG - good to follow progression of group's output recommendations.
- **Global open research commons (GORC): what it takes and the road to get there. Reflections and discussion on creating a commons attribute model and a commons integration roadmap**
 - [GORC international model v1.0](#) release and further details in notes below.
 - [Notes link](#), page 33
- **Building a semantic interoperability framework: towards FAIR mappings and crosswalks**
 - From FAIR Impact project, a first version of recommendations:
 - Use SSSOM [A Simple Standard for Sharing Ontology Mappings \(SSSOM\)](#) to structure, serialise with LD standards, use standard ontologies.

- For Metadata: which ones are required? Preliminary answer: use SSSOM as a guide, whether we use it or not.
 - [Notes link](#), page 7
- **[BoF - Valorisation of data and software in the open science world](#)**
 - Exploring the competing objectives between Open Science and private business models (relevant in public-private partnerships)
 - Valorisation vs openness
 - What about a global (as applying to the whole world) research law?
 - Increase engagement with industry, private sector, publishers..., as
 - It is a collective responsibility
 - Also, there is a need for funding: sustainability challenges
 - BUT in any case, preserve openness
 - [Notes link](#), page 11
- **[Researchers' engagement with data management: from cookbook to toolkit](#)**
 - This IG: want engagement from everyone: collect and share best practices:
 - [Cookbook](#)
 - Submit your case study: [template](#) or via [Google form](#).
 - Podcast - IDEA:
 - <https://www.idea.int/podcasts>
 - <https://podcasters.spotify.com/pod/show/datalibs>
 - [Notes link](#), page 28
- **[Semantics / Metadata Oriented Presentations of note](#)**
 - Interesting tools and ecosystems:
 - [NHLBI BioData Catalyst®](#)
 - Cloud-based ecosystem that facilitates analysis and access to data (at scale)
 - 4 Pb of data, incl. largest and most ethnically diverse set of whole-genome sequences from [Trans-Omics for Precision Medicine \(TopMed\) program \(GA4GH Driver Project\)](#)
 - [CEDAR workbench](#) (see also [Unleashing the potential of domain specific metadata](#))
 - End-to-end metadata management solution w/templates to support community standards
 - Autogenerates spreadsheets from CEDAR templates, incl. ontology support and validation
 - [Semantic Engine](#)
 - Suite of tools to manage *dataset* ↔ *schema* and *schema* ↔ *schema* relations
 - [Overlay Capture Architecture \(OCA\)](#)
 - Each feature is a layer of the schema in OCA (like a Docker image)
 - Mix-and-match scheme creation - different contributors to different parts, etc.

Notes section

Overview

This section contains more detailed notes by ELIXIR members at the event who reported back key info.

Note: not all sessions of the event had ELIXIR reporter coverage given the volume of sessions.

Table of Contents

Monday, 23 October

[Opening Plenary Session: Welcome to IDW 2023 'Spatial Data Science: Geographic Context Matters'](#)

[SDC-B1-3: Building a semantic interoperability framework: towards FAIR mapping and crosswalks](#)

[SDC-B1-7: Institutional / Strategy Oriented Presentations](#)

[RDA-B1-1: Structuring Semantic Information with Respect to Conventional Metadata \(Syntactic\) Structures](#)

[RDA-B1-6: Mastering the Art of Research Software Metadata and Metrics](#)

[RDA-B1-8: Valorisation of data and software in the open science world](#)

[SDC-B2-8: Data Science Lightning Talks](#)

Tuesday, 24 October

[RDA-B2-3: Ten recommendations for data repositories/catalogues to improve data discoverability](#)

[RDA-B2-6: Active Data Management Plans: how far have we come in the changing landscape](#)

[Plenary Session: Inclusivity in Open Science while advancing research assessment and career pathway impact](#)

[SDC-B3-3: Unleashing the potential of domain specific metadata](#)

[SDC-B3-4: WorldFAIR: Emerging Cross-Domain Perspectives - Session 2, Envisioning Solutions - the Cross Domain Interoperability Framework \(CDIF\)](#)

[SDC-B3-7: Data Stewards / Competency Centres / Training Presentations](#)

[RDA-B3-1: FAIRification of Genomic Tracks: data-driven life science through granular discovery of biological sequence annotations with uniform metadata](#)

[RDA-B3-2: PID Exchange - a curated information resource for PID adopters](#)

[Plenary Session: RDA Adoption & Impact](#)

Wednesday, 25 October

[SDC-B4-1: Interoperability and Interconnectivity in Developing Open Science Infrastructures](#)

[SDC-B4-5: Why Persistent Identifiers and How Their Connectivity Enhances Research](#)

[RDA-B4-3: Researchers' engagement with data management: from cookbook to toolkit](#)

[RDA-B4-6: Building towards FAIR for Machine Learning](#)

[RDA-B4-7: Revising the Versioning Principles: The Road to Actionable Recommendations](#)

[RDA-B4-8: Let's talk about FAIR mappings! Towards common practices for sharing mappings and crosswalks](#)

[SDC-B5-7: Data Management Plans Presentations](#)

Thursday, 26 October

[RDA-B5-3: Global open research commons: what it takes and the road to get there. Reflections and discussion on creating a commons attribute model and a commons integration roadmap](#)

[RDA-B5-7: 2nd BOF: Trusted / Secure Research Environments for Sensitive / Confidential Data: FAIRness for "Closed" Data and Processes](#)

[SDC-B6-3: Panel Discussion: Standards for a FAIR Data Future: How to Advance Standardisation in Data Management?](#)

[SDC-B6-5: Semantics / Metadata Oriented Presentations](#)

[SDC-B6-6: National / Institutional Approaches Presentations](#)


[RDA-B6-4: Key aspects of FAIR for infrastructure solutions in the biomolecular life sciences](#)


[RDA-B6-5: Equity and Inclusion: universal access to PIDs](#)

Session Reporting

Monday, 23 October

Opening Plenary Session: **Welcome to IDW 2023 'Spatial Data Science: Geographic Context Matters'**

 *When & where:* 23 Oct 9:00-11:00, Europa Hall, plenary

 *Reported by:* Marek Suchánek

The session was moderated by Univ.-Prof. Dr. Arne Bathke, dean of the Faculty of Digital and Analytical Sciences, who welcomed everyone and provided practical information. Dr. Wilfried Haslauer (Governor of Salzburg) opened the event by welcoming all participants, thanking the organisers and emphasised the importance of the event in context of growing needs related to data across all domains worldwide. Brief welcome came also from prof. Barend Mons (CODATA), Hilary Hanahoe (RDA), and Dr. David Castle (WDS) clarified the role of their organisations.

The main part of the session was a keynote by Linda Beale (Esri) about the importance of spatial (meta)data including explanations on practical use cases:


- **Spatial is crucial and fundamental** to work that we do and can affect the results significantly.
- Data always needs (spatial) context otherwise it might be totally useless.
- Name does not matter, not even used software, **the meaning/intention matters**.
- **Maps are powerful:** putting data on a map is helpful, attractive for people and organisations, checking for missing regions, etc.

Another three short inspirational talks on spatial data followed:

- Getachew Gella (CDL-GEOHUM) presented work on integrating AI with Earth observation for building extraction in settlement areas. Various challenges related to recognition of buildings in settlement areas, mainly refugee camps, were demonstrated and clarified together with potential for AI-based solutions.
- Anita Graser (AIT) presented challenges in movement data analysis and its privacy challenges.
- Krzysztof Janowicz (UC Santa Barbara) talked about KnowWhereGraph (knowwheregraph.org) where geographical information is represented in a knowledge graph, FAIR by design, supporting various perspectives and enabling AI/ML processing (including a feedback loop to the graph for increasing density).

SDC-B1-3: **Building a semantic interoperability framework: towards FAIR mapping and crosswalks**

 *When & where:* 23 Oct 11:30-13:00, Mozart Hall 4+5, breakout (interactive workshop), [details](#)

 *Reported by:* Paulette Lieby, Wolmar Nyberg Åkerström


Paulette Lieby with later contributions from Wolmar:


Summary:

- From FAIR Impact project, a first version of recommendations
 - use SSSOM [A Simple Standard for Sharing Ontology Mappings \(SSSOM\)](#) to structure, serialise with LD standards, use standard ontologies
 - for Metadata: which one are required? Preliminary answer: use SSSOM as a guide, whether we use it or not
- From [FAIRCORE4EOSC](#), sister project:
 - Focus on service dev
 - Enhancing FAIRness in EOSC
 - A EOSC metadata schema and crosswalks registry
 - Collaborative session notes: <https://bit.ly/scidatacon-notes>
 - Slides: <https://doi.org/10.5281/zenodo.10033053>
 - Google Slides version: <https://docs.google.com/presentation/d/1ABSbo3NrZmlTOSnNMI3ifOeEn1gmtot6dJKUYVCG2do/edit>
 - Workshop on November 24th: <https://fair-impact.eu/events/fair-impact-events/documenting-mapping-community-practices>
- Intro Wolmar Nyberg Åkerström EOSC-A Semantic Interoperability Task Force
 - 13 Task Forces liaise with EOSC projects to offer feedback on developments
 - The Semantic Interoperability Task Force develops and implements the semantic interoperability recommendations of the [EOSC Interoperability Framework](#)
 - The Strategic Research & Innovation Agenda (SRIA) provides a roadmap and priorities for EOSC
 - governance of community-based metadata and data standards
 - semantic artefact catalogues in national infrastructures and guidelines
 - publishing semantic artefacts through institutional or vocabulary specific thematic repositories
 - 40+ TF members working to identify and address gaps across three themes: (Meta)data conventions; Semantic artefact catalogues; Use cases
 - [Call for input from the global data community](#) for semantic interoperability in EOSC
- FAIR Impact project Yann Le Franc, also part of EOSC-A Semantic Interoperability Task Force
 - Discussing engaging with all communities re FAIR
 - To converge on practices/policies
 - Metadata & ontologies
 - A semantic artefact for governance
 - In particular here, we talk about mappings: linking entities in the digital world
 - FAIRify: we talk about documented mappings → mapping sets/crosswalks, instead of just mappings
 - Pbs: 1-n or n-1 mappings across schemas
 - Pb: how to re-use already established mappings?
 - Need to be accessible
 - Interoperable
 - Reusable: no explanatory context
 - So, we need FAIR mappings
 - Automation of data retrieval
 - → establish guidelines for FAIRification
 - → ma common exchange model
 - → how? Engage with communities: co-construct
 - → governance framework
 - 1. Analysis of requirements
 - 2. Community mapping practices
 - Recommendations for FAIR mapping:

- Gathering topics around four groups
 - 1st version after collecting first community inputs; eg: use SSSOM [A Simple Standard for Sharing Ontology Mappings \(SSSOM\)](#) to structure, serialise with LD standards, use standard ontologies
 - Eg for Metadata: which one are required? Preliminary answer: use SSSOM as a guide, whether we use it or not
 - Analysing topology of mappings
 - A git repo will be set up with all docs
- [FAIRCORE₄EOSC](#): Future Tools for Mappings (40 minutes) - Joonas Kesäniemi, Tommi Suominen
 - FAIRCORE₄EOSC:
 - Sister project of FAIR impact (above)
 - Focus on service dev
 - Enhancing FAIRness in EOSC
 - Implementation challenges
 - RDGraph research discovery graph, from OPENAIRE
 - PIDGraph from Datacite
 - MSCR, EOSC metadata schema and crosswalks registry
 - DTR datatype registry
 - PIDMR EOSC PID Meta resolver
 - CAT EOSC Compliance assessment toolkit
 - Raid research Activity service identifier, from Oz
 - RSAC EOSC software heritage mirror
 - + EOSC software
 - Case studies
 - Release ... beta version soon
 - MSCR, EOSC metadata schema and crosswalks registry
 - Case studies
 - Built on top of Finnish dev tools
 - Terminologies tool
 - Reference tool
 - Data vocabularies tools
 - Add a crosswalk tool
 - Schemas can be submitted in a draft, discussed etc.. then published (versioned)
 - Schemas are versioned
 - Crosswalks always refer to versioned schemas
 - FAIR:
 - Registered content has PIDs, mappings are highly contextual
 - Publicly accessible
 - We have semantic, structural, value mappings

SDC-B1-7: Institutional / Strategy Oriented Presentations

 *When & where:* 23 Oct 11:30-13:00, Karajan-Hall 1, breakout (mixed paper session), [details](#)


 *Reported by:* Marek Suchánek


The session consisted of 4 talks related to different institutional strategies and lessons learned in different areas related to RDM:


- Robert Werth (Frankfurt University of Applied Sciences)
RDM practices at universities of applied sciences in Germany

- Focus on teaching, <1% research funds, many joint projects with business/government.
 - There is a lot of data to be “mined” (used).
 - Only 25% heard about FAIR principles, 16% about CARE principles.
 - Many work with data in research but do not realise that they perform RDM, thus for example also don't know about advantages/possibility to put data in repository/archive.
 - **They do not do RDM and make data FAIR because of extra workload.**
 - Researchers need permanent local RDM staff, raise awareness, less teaching burden, integrate RDM into curricula, and more reputation for data publication.
- Robin Rice (University of Edinburgh)
 - Why wasn't I told this service existed?**
 - They had a survey with 171 replies from 3 colleges during the summer. Later did interviews and focus groups.
 - **Limited awareness of services available**, people don't know about offered services and therefore not using them.
 - They now have a website <https://digitalresearchservices.ed.ac.uk> and plan to do roadshows, partnerships, digital curation interviews, conferences, and more.
- Matthew J. Chandler (Princeton University)
 - Building Sustainable RDM Infrastructure**
 - Sustainability through transiting from data management planning to preserving and publishing.
 - Significant **cultural change** is related to it (requirements, research integrity, etc.).
 - **TigerData** project = data management ecosystem at Princeton University
 - It strives to add value at multiple levels: Support Services, User Interface + API, Data Management System, Repositories and Storages.
 - It handles role hierarchy for projects from external collaborators up to data stewards and sponsors.
 - It should be sustainable and modular infrastructure at Princeton.
 - (note) Seems like very early stage of work in progress project: <https://github.com/pulibrary/tiger-data-app>
- Timothy B. Norris (University of Miami)
 - Geospatial Digital Special Collections**
 - They have a lot of geospatial datasets and are building a custom software for it.
 - **Geospatial data struggle to be FAIR** (difficult harmonisation, many standards, etc.).
 - Work in progress on their own GDSC solution (packaging, annotating, provenance, etc.).
 - It should result in a flexible and scalable metadata-rich repository for geospatial datasets (using PostGIS, Kubernetes, metadata store, ETL and Rook-Ceph PVC).

RDA-B1-1: Structuring Semantic Information with Respect to Conventional Metadata (Syntactic) Structures

 *When & where:* 23 Oct 14:00-15:30, Mozart Hall 4+5, breakout, [details](#)


 *RDA Groups:* [Metadata IG](#)

 *RDA Pathways:* FAIR, CARE, TRUST - Adoption, Implementation, and Deployment; Data Infrastructures and Environments - International; Semantics, Ontology, Standardisation


 *Reported by:* Sveinung Gundersen


- [Collaborative session notes](#)
- Metadata IG, one of the first RDA Interest Groups (2013)
 - Two successful WGs:
 - Metadata Standards Directory WG
 - Continued with [Metadata Standards Catalog WG](#)
- Metadata Element Set (Keith Jeffrey)
 - Data elements, more than single-valued attributes
- Vocabulary and Semantic Services IG (John Graybeal)
 - Re-envisioned in 2015
 - i-ADOPT WG
 - BoF: [Let's talk about FAIR mappings](#)
 - Semantic superpowers:
 - Referential precision and persistence
 - Accessible labels and descriptions
 - Standard representation
 - Computational logic
 - Connecting metadata and vocabularies
 - *Only interoperable if computable!*
- Metadata for Effective Scientific Knowledge Graphs (Andrea Mannocci)
 - [Scientific Knowledge Graphs Must Interoperate](#) (paper)
 - RDA WG Scientific Knowledge Graphs - Interoperability Framework
- Metadata Required for FAIR Digital Objects (Rainer Stotzka ++)
- RDA Endorsed recommendation:
 - Representation of FAIR Digital object should contain at minimum:
 - Identifier, e.g. PID
 - Information Record with:
 - Identifier
 - Profile (record structure definition)
 - Type of data
 - Link to data
 - Helmholtz Kernel Information profile
 - RDA Draft - 15 basic attributes

RDA-B1-6: Mastering the Art of Research Software Metadata and Metrics

 *When & where:* 23 Oct 14:00-15:30, Trakl-Hall, breakout, [details](#)

 *RDA Groups:* [Software Source Code IG](#)

 *RDA Pathways:* FAIR, CARE, TRUST - Principles; Semantics, Ontology, Standardisation; Research Software

 *Reported by:* Marek Suchánek

The interactive session was organised by Morane Gruenpeter (Software Heritage) and it focused best practices in Research Software, metadata, activities of FAIR-IMPACT, and building awareness of FAIR4RS. First, Morane presented various activities, the current state and plans in the interest group, FAIR-IMPACT, Software Heritage, and FAIR4RS.


Then, the more interactive part continued focusing on the value of software metadata (why it matters and how for different roles, what different roles provide, and who uses it).

Important links to resources mentioned during the session:


- [CodeMeta](#)
- [Software Heritage](#)
- [FAIR4RS Principles](#)
- [ReSA](#)
- [RSMD Checklist \(from FAIR-IMPACT\)](#)


Collaborative notes: <https://bit.ly/3PRBUX5>

RDA-B1-8: Valorisation of data and software in the open science world

 *When & where:* 23 Oct 14:00-15:30, Karajan-Hall 1, breakout, [details](#)

 *RDA Groups:* BoF 

 *RDA Pathways:* Other

 *Reported by:* Paulette Lieby

Summary/highlights

- Exploring the competing objectives between Open Science and private business models (relevant in public-private partnerships)
 - Valorisation vs openness
- What about a global (as applying to the whole world) research law?
- There is interest, and practice from all(?) continents, mostly EU though for the time being
- Increase engagement with industry, private sector, publishers..., as
 - It is a collective responsibility
 - Also, there is a need for funding: sustainability challenges
 - BUT in any case, preserve openness
- Collaborative session notes:
<https://docs.google.com/document/d/1d-MW3ZjQQ4rpRMgwKoittkUoiNd7BvbDdidXDjtoS4/edit#heading=h.rhwgt4e87dyq>
- Stéphanie Kuss, Hilary Hananoe, Agnès Robin (UMontpellier)
- BoF: around how to answer possibly competing mission objectives
- Stéphanie Kuss
 - C.U.R.I.E network
 - French association for academic research valorisation
 - What is valorisation
 - “Make research results.. Accessible/usable or marketable”
 - Through knowledge and technology transfer
 - And partnership
 - Valorisation officers
 - Examples of valorisation
 - In many fields
 - Let’s focus on IP (patents, software, know-how)
 - IP: do not publish before IP protection
 - What about OS?
 - HOW?


- Agnes Robin
 - "Droits des données de la recherche" pub. 2022
 - Open science vs knowledge valorisation?
 - Ex: LPRN 2016 "Loi pour la République Numérique" French Law that exploits the how of research data 'openness'
 - If funding > 50% open data
 - Cases for closing:
 - Personal data
 - Biodiversity (Nagoya)
 -
 - What are the good practices?
 - CURIE: how to integrate OS clauses in public-private partnerships
 - What about a global an uniform research law?

- Use case: Hillary Hananoe
 - Ups and downs of public-private partnerships
 - Bottom-up, starting from communities: working groups
 - A framework for openness
 - RDA: multi-stakeholder
 - Main: public research + academia
 - Increase engagement with industry, private sector, publishers...
 - A collective responsibility
 - Also, need funding: sustainability challenges
 - BUT preserve openness
 - Thus, a framework for sponsorship
 - Process maintained
 - WG facilitated on of working groups, increase visibility of work
 - 1st: ORACLE:
 - 2 WGs
 - Ups
 - Professional support
 - Financial support
 - Downs
 - Perception & culture
 - Volunteer work for sale
 - Giving knowledge to commercial organisation
 - Handle with care, but handle
 - Cost, responsibility and risk sharing exercise
 - OS is everyone's concern but also everyone's benefits

- Discussion: Is this a topic of interest/concern in your country?
 - We need an intermediate to help work on the ground
 - As a lawyer: OS is a data management method; didn't catch
 - EU project: FAIRPlus, closed 2022, consortia between public research and big pharma
 - Fact: Most of the data never reused
 - But with most of the data could be opened
 - So in need of FAIRification
 - DE (Germany, one specific large research institute): we have a OS policy
 - Provoked a discussion, many technology transfer departments
 - Since then, set up a policy of sorts; but we do not believe there exists one answer for everyone
 - We should not overemphasise legal issues
 - We should stress IP to allow openness (so that to prevent stealing by other interests)

- NL
 - Issues around partnerships
 - Q: where do we step in this process (decision making wrt to the issues discussed here)?
- Issue of decision at the individual level? Or collective (which? Organisation, state, EU, other?) level
- One of the aims: produce education material to connect OS issues with valorisation issues
 - But we are maybe not at this stage yet
 - There are many ways of valorisation, not necessarily money
 - Also, a lot of "OS missions" created in FR are not even talking about valorisation
- Different types of valorisation: for example, valorisation for publishers (closed) and start-up (open)
 - Some element of answers (eg Plan S)
- Finland
 - HPC centre
 - RDM solutions for academic, and partnerships
 - Pb: understanding the business realities, this is where the gaps are
 - We should have the business side represented here
- There is a IG "sharing rewards and credit", mainly oriented towards academics, but is not restricted
 - Currently writing up recommendations
- Do a landscape analysis to evaluate the issue
- Chile:
 - Part of public-partnership
 - Pushing the conversation towards open science
 - We could benefit a lot from this conversation
- Important to reach beyond Europe
- Q: do we talk about public-private partnerships or valorisation?
- UK
 - A public-private partnership
 - Had to have this conversation
- Good examples in FR for public-private valorisation for software (did I understand this properly?)
- FR - Humanum
 - What about metadata? Since typically metadata is supposed to be open, always
 - For (French) legal aspects: no difference between data and metadata, but the law is very vague on this topic, so it could be worthwhile
- What is happening at the EU level? Not necessarily in research
 - EU rules are in the process of changing
 - New rules/regulations requiring private sector to open, under conditions
- Another issue: duration of embargo

SDC-B2-8: Data Science Lightning Talks

 *When & where:* 23 Oct 16:00-17:30, Mozart Hall 2+3, breakout (lightning talks), [details](#)

 *Reported by:* Sveinung Gundersen


Seven misc. scientific talks. Of particular interest:

- [Comparing statistical methods for analysing longitudinally measured ordinal outcomes in rare diseases](#) (Martin Geroldinger et al.)
 - Cross-over study of rare diseases
 - Usually not standard parametric methods
 - Comparison of nonparametric methods
- [Analysis-Ready AI-Datacubes: Easier Access \(Not Only\) for Geo Data](#) (Peter Bauman)
 - Earth data:
 - Move away from sensor data oriented file-based binary graveyards


- Instead: actionable, queryable spatio-temporal services
- Concept of datacubes as a unifying paradigm:
 - Standards like ISO SQL/MDA, OGC WCPS, and the ISO 19123 family
- Field of “Array Databases: based on semantic-rich array data model
- Declarative query language allows to state the question, rather than devise an algorithm
- Perfect match for AI/ML

Tuesday, 24 October

RDA-B2-3: Ten recommendations for data repositories/catalogues to improve data discoverability

 *When & where:* 24 Oct 9:00-10:30, Mozart Hall 4+5, breakout, [details](#)

 *RDA Groups:* [Data Discovery Paradigms IG](#)

 *RDA Pathways:* Data Infrastructures and Environments - Regional or Disciplinary; Data Infrastructures and Environments - International; Data Lifecycles - Versioning, Provenance, Citation, and Reward


 *Reported by:* Fotis Psomopoulos

Collaborative notes: [RDA P21 Collaborative Notes - Ten recommendations for data repositories/catalogues to improve data discoverability](#)


Summary

Overall the session was attended by ~100 people onsite and ~20 people online. The main goal of the session was to review and finalise the drafted recommendations for data repositories to improve data discoverability.


- Rule 1: Know your user (Choose an appropriate user study method)
- Rule 2: Improve presence in Web search
- Rule 3: Be present in literature search
- Rule 4: Support both disciplinary and interdisciplinary search
- Rule 5: Support broad terms to narrow terms
- Rule 6: Handle duplication
- Rule 7: Search data within data
- Rule 8: Support user's metadata needs (from user's perspective and machine-to-machine perspective)
- Rule 9: Improve metadata quality / linked data
- Rule 10: Include identifiers (PID)

As part of the discussion, the case study of PANGEA ([Data Publisher for Earth & Environmental Science](#)) was presented as they have already implemented a number of the rules, with different levels of maturity. The next step identified for the IG is to finalise the  [Ten-Things for repositories to improve data discovery](#) document before the end of the year, so that it can be effectively used by the wider community.

RDA-B2-6: Active Data Management Plans: how far have we come in the changing landscape

 *When & where:* 24 Oct 9:00-10:30, Europa Hall, breakout, [details](#)

 *RDA Groups:* [Active Data Management Plans IG](#)

 *RDA Pathways:* Data Infrastructures and Environments - Institutional; Data Infrastructures and Environments - International; Training, Stewardship, and Data Management Planning

 *Reported by:* Paulette Lieby

Summary:

- A world tour of the major maDMP tools and infrastructures : DMPRoadmap, DSW, ARGOS, Japan, China; and RDA-DMP-CS
 - Full blast towards automated data flows between CRIS and DMPs
- Result of breakout rooms: challenges
 - Institutions
 - Need for incentives
 - Use a minimum DMP template
 - Plan for cost recovery
 - Support visibility
 - Funders
 - Interaction with which tools?
 - Interoperability then critical
 - maDMP: when do the funders actually consume the maDMP?
 - Data stewards and librarians
 - Focus on support: awareness, sense of urgency, import FAIR principles,
 - DMP as a chore
 - Researchers
 - Also focus on support
 - How to manage big data
 - Metadata standards come from the communities: bottom up

- Collaborative session notes:

<https://docs.google.com/document/d/1UBdXDM5vpsu4JtyjdFbP74MOesjMfbWPrFeOujpp99A/edit#heading=.8p8cwrkl8tqg>


- The only session about DMPs
- Update from RDA DMP CS (Marek)
 - RDA open call project
 - Refactoring of CS
 - Meta-what? Metamodel for the specification
 - Still the same specification, the form will itself change towards machine-actionability
 - Later: extension, made easier using metamodel
 - See poster #20

- DMPs platforms across regions (Elli GR)
 - NII in Japan
 - Compliance re funders
 - Towards automation of data management process based on DMP
 - le, practice RDM along the DMP
 - And RDM compliance with DMP
 - And RDM monitoring with DMP
 - CASDMP China
 - 2019: open data and DMP required "big data driven discovery"
 - DMP as the core of RDM towards researchers, projects, funders...,
 - For publication too
 - Using RDA DMP CS, with modification wrt dataset and publication (indicators)
 - Focus on interoperability: vocabulary
 - Automated Evaluation, tracking, statistics
 - UK - DCC
 - DMPonline (DCC) and DMPRoadmap
 - DCC as a service used by academia around the world
 - DMPRoadmap (open source) as a common base to various services provided by national bodies
 - How to we connect to policies?

- At different levels
 - Through acquiring mandatory information
 - Constraining how some issues are dealt with
 - Through contextual guidance
 - And workflows to other tools
 - Use cases for technical capacity
 - Interrogating repositories, metadata standards, policies, grant PIDs
- DAMAP Tomasz
 - TU Wien & Graz, from RDA DMP CS
 - Aiming at high automation
 - So closely integrated with a specific CRIS
 - And closely automated
 - Future: automated feedback and feedback
 - With ARGOS & DSW
 - Paper <https://datascience.codata.org/articles/10.5334/dsj-2023-028>
- DSW Marek
 - DSW highly customisable: export & import
 - Based for Chalmers University
 - Building API adapters for questionnaires/projects
 - To follow up
 - Future
 - KM to cover all entities & properties of CS: here we talk about the Common DSW KM, for example not covering costs
 - Extension of CS
- ARGOS OpenAIRE Elli
 - DMP CS compliant
 - Create blueprints for software... combine different templates
 - IDs assigned through Datacite
 - Automated prefilling, API calls, publication
 - Integration use-cases
 - RoHub, Rspace
- Q&A
 - Q: which feature gets the most positive feedback?
 - DCC: Targeted guidance
 - DAMAP: ?
 - CAS: help in organising data
 - There was mention of including software and other outputs in the planning tools. I'm interested in which of the tools presented can incorporate information on outputs other than data?
 - DSW: is a general expert system, so can include anything
 - DCC: consider lifecycle of software and data? Have to think about this issue when designing tools that answer more than data
 - Automated publication of DMPs
 - DCC: wherever the most useful, Zenodo for example
- Breakout rooms
 - Institutions
 - Need for incentives
 - Use a minimum DMP template

- Plan for cost recovery
 - Support visibility
- Funders
 - Interaction with which tools?
 - Interoperability then critical
 - maDMP: when do the funders actually consume the maDMP?
- Data stewards and librarians
 - Focus on support: awareness, sense of urgency, import FAIR principles,
 - DMP as a chore
- Researchers
 - Also focus on support
 - How to manage big data
 - Metadata standards come from the communities: bottom up

Plenary Session: Inclusivity in Open Science while advancing research assessment and career pathway impact

 *When & where:* 24 Oct 11:30-13:00, Europa Hall, plenary, [details](#)

 *Reported by:* Marek Suchánek

Plenary session consisted of 4 presentations on topic of inclusivity from various perspectives and contexts:

- Prof. Ginny Barbour ([OA Australasia](#)) presented an approach for **inclusivity and equity in assessment** in Australasia. The main issues are that we only value what can be measured and often take a one size fits all approach thus cannot easily value what is important. In Australasian context, they have rich indigenous cultures and currently indigenous researchers may be excluded. Different solutions are appearing: statements of intent by institutions (e.g. declaring support for ECR), locally led initiatives (e.g. [TUWHERA](#)), or embedding appropriate practices ([gida-global.org/care](#)).
- Dr. Ana Ortigoza ([PAHO](#)) talked about the case from Latin America, mainly about the needs of indigenous and other groups, what needs to be addressed in this aspect, and who should support the change. They have census data, health survey data and vital statistics where there is a growing trend to include ethnic identification. There is also legal gender recognition (4 countries). **Data is both a technical and political tool** to promote access to services, expand citizenship, and guarantee human rights for Afro-descendants, Indigenous people, and sexual minorities. **Window of opportunity for FAIR and CARE data implementation** being addressed by PAHO as well as in WorldFAIR WP8.
- Abel L Packer ([SciELO](#)) presented impact, diversity, equity, inclusion and accessibility aspects in SciELO Network (Open Science infrastructure in three continents). Mainly observations in evaluation % of languages of publications and various topics has been presented. Research assessment as an integral function of the research cycle as Open Science practice. Research communicated via nationally edited journals are of highest relevance in addressing the SDGs.
- Prof. Maui Hudson ([Uni Waikato](#)) also provided a view on indigenous needs for inclusivity in Open Science. The presentation was mainly based on a publication "Indigenous Peoples' Rights in Data: a contribution toward Indigenous Research Sovereignty" ([10.3389/frma.2023.1173805](#)) - mainly clarification of 5 As/levels: **acknowledgement, attribution, authorship, access, and authority**.

SDC-B3-3: Unleashing the potential of domain specific metadata

 *When & where:* 24 Oct 14:00-15:30, Doppler Hall, breakout (mixed paper session), [details](#)

 *Reported by:* Sveinung Gundersen

- Examples of domain-specific metadata templates connected to the CEDAR infrastructure
 - GENEMEDE – generic metadata descriptors to comprehensively describe a neuroimaging experiment (Niccolo Bonacchi & Praveen Sripad)
 - Building cognitive neuroscience metadata template (Zefan Zheng)
- Representing community standards as metadata templates (Mark A. Musen)
 - On the CEDAR workbench & infrastructure
 - Enthusiastic presentation, highly recommended:
 - [Video](#) (starts 28 mins in)
 - Examples from NCBI BioSample:
 - 73% of “Boolean” metadata are not actually *true* or *false*
 - 26% of “integer” metadata values cannot be parsed into integers
 - 68% of metadata that are supposed to follow ontologies do not actually do so
 - MIAPPE and family as examples of metadata standards
 - CEDAR approach:
 - Community-endorsed reporting guidelines as *templates*
 - Use ontologies whenever possible
 - User interface:
 - Ontology values are sorted according to popularity (based on context)
 - GOFAIR:
 - Metadata for Machines workshops
 - Intensive 1-3 day workshops with invites experts to agree on recommendations
 - Ultimately result in CEDAR templates
 - CEDAR templates:
 - Plug-and-play with a variety of software systems
- Implementing domain-specific metadata into Dryad
 - Dryad is data publishing platform
 - Partnership with CEDAR allows standardised metadata for particular domains
 - CEDAR’s embedded editor integrated into Dryad submission process
 - New features including auto-save (!)

SDC-B3-4: WorldFAIR: Emerging Cross-Domain Perspectives - Session 2, Envisioning Solutions - the Cross Domain Interoperability Framework (CDIF)

 *When & where:* 24 Oct 14:00-15:30, Europa Hall, breakout (practice papers), [details](#)

 *Reported by:* Paulette Lieby, X Y

Summary:

The project WorldFAIR, the Cross Domain Interoperability Framework (CDIF)

- Collaboration with authoritative organisations
- Turning FAIR into reality: an interop framework
- Lots of case studies, across heaps of disciplines
- NOT a new standard, BUT an interoperability framework
 - Aim for implementation, not modelling
- Recommendations published in May 2024
 - Along SSSOM, DDI-CDI


- Organised by CODATA
 - <https://www.scidatacon.org/IDW-2023-Salzburg/sessions/555/>
 - The WorldFAIR project: the CDIF
 - What we do and where do we go?
 - EU funded project (from ERA) : CODATA originally around
 - Around FAIR
 - Address global issues
 - Partnerships
 - GoFAIR, DDI, GBIF, OneGeochemistry
 - RDA
 - With authoritative international entities!
 - Turning FAIR into reality: an interop framework
 - To support inter-disciplinary research, adopt global standards where relevant
 - Lots of case studies, across heaps of disciplines
 - Chemistry, social sciences, life sciences (biodiversity), ...
 - Methodology
 - An exploratory FIP
 - CDIF workshop: engage in depth
 - → Recommendations, revised FIP
 - FIPs:
 - See FIP from DSW
 - The conversation
 - Encoded as a nanopublication
 - What is CDIF?
 - NOT a new standard
 - An interop framework
 - World FAIR: important international reach
 - 11 rich case studies
 - 11 completed and prospective FIP
 - CDIF
 - Pilot of modelling research methodology


- CDIF: goals and status update
 - IS
 - NOT a new standard
 - Concrete guidance for cross-domain implementers: at the machine level
 - Coordinated set of metadata profiles and services, on existing spec and technologies
 - *ma* interop FAIR implementation; step-wise implementation
 - Lingua franca: restrict the scope: a many-to-one framework
 - A common language for most domains
 - Status
 - Spring 2024: set a recommendation
 - Discovery, access, integration, context, universal
 - Schema.org + dcat, ODRL, DDI-CDI (data structure), SKOS/XKOS, OWL, SSSOM (semantics), PROV-O, DDI-CDI, I-ADOPT/O&M (context)
 - More advanced profiles later; 2nd round of WorldFAIR possibly
 - Discoverability
 - Draft for level one recommendation
 - <https://github.com/Cross-Domain-Interoperability-Framework/Discovery/blob/main/discoverability.md>
 - JSON-LD
 - Level two under development
 - DCAT CDI

- Level three: queryable services...
 - Access, integration, context, universals...
 - Level one recommendations for some (access I believe)
- Here is some background on the DDI Variable Cascade, from earlier this year:
 - <https://codata.org/initiatives/data-skills/ddi-training-webinars/the-variable-cascade/>
- CDIF Working Group Presentations
 - Data integration and semantic mappings Yann Le Franc
 - See
 - <https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/2946531350/2023+Defining+a+Core+Meta+data+Framework+for+Cross-Domain+Data+Sharing+and+Reuse>
 - Recommendations for 3 questions
 - Identify real use-cases (RO-crates, lots..., DCAT and schema.org...)
 - Starting with dummy use-case:
 - Two csv files two different measurements in rivers
 - Q: can we establish correlations?
 - Start with semantics
 - First merge on date
 - Use DDI-CDI framework to squeeze URL ontologies into column headers (use variable cascade from DDI)
 - → to define the level of mappings
 - Recommendation 1: use SSSOM for simple semantic mappings
 - Simple Enumeration mapping: use constrained classification mapping different controlled lists
 - 2: use SSSOM for simple enumeration mappings
 - N-m enumeration mapping:
 - what about XKOS?, is SSSOM enough?
 - Where do we put the mappings
 - 3. Part of PROV-O of the resulting files
- Data access
 - The poor cousin of interoperability
 - Tricky concept: can cover lots of concepts
 - Eg usage different from access
 - Needs *ma* not only machine readable: mediating access to data
 - Mostly done by prose
 - → lack of automation
 - Not much harmonisation, heterogenous terminology
 - Access vs usage
 - Access: mediating initial access
 - Usage: subsequent access
 - Use ODRL (open digital right language) to model access
 - 4 use-cases
 - Consistent representation
 - Policy Metadata aggregation
 - Shared access services
 - Access for federated analysis
 - ODRL will need to be extended for more complex use-cases:
 - Sits in process interoperability
 - Look at workflows and orchestration
 - Build a CDIF profile

- Looking forward: expanding and sustaining WorldFAIR
 - A shift from bibliography to a technology approach
 - Project ends May 24
 - Next step: WorldFAIR+
 - Q about sustainability


SDC-B3-7: Data Stewards / Competency Centres / Training Presentations


 *When & where:* 24 Oct 14:00-15:30, Trakl-Hall, breakout (mixed paper session), [details](#)


 *Reported by:* Marek Suchánek


- UniVienna (Tereza Karlová, Monika Bergmann, ...)
 - RDM and Data Stewardship training portfolio**
 - Generic and domain-specific courses, courses on specific tools
 - Courses mainly for PhD students
 - For example, RDM for the Life Sciences (for PhD candidates) = 2-3h online course + interactive session (2x3h in-person) with practical exercise(s) + snacks
 - Data Steward Certificate Course
 - 2 semester / 15 ETCS, 2950 EUR, certificate, English
 - Next actions
 - Legal helpdesk
 - Growing commitment by PIs and depts
 - Erasmus+ project
- NFDI4Health (Jens Dierkes, Julia Fürst, Birte Lindstädt, ...)
 - Data Stewards and FAIR Data Training in Biomedical Research**
 - Focusing on different roles and responsibilities in complex RDM.
 - Cross-domain graduate programme in RDM and data science based on curriculum of Uni Bremen.
 - Data Train concept, syllabus, implementation report (published in Oct 2023: [10.4126/FRL01-006441348](https://doi.org/10.4126/FRL01-006441348)).
 - Data Steward pilot projects in academic libraries (ZB MED and USB Cologne).
- Paula Oset Garcia (UGent), Veerle Van den Eynden (KULeuven), Tessa Acar (FWO), Ziad Choueiki (UGent):
 - Building a community of data stewards in Flanders**
 - [FRDN Knowledge Hub](#) started in February 2021.
 - Community canvas to analyse the community – **identity, experience and structure**.
 - They host “Network Day” to get together people involved in Open Science.
- Lyve Kvale (UiO) Norwegian competency framework
 - Aims to define a set of skills and competencies, applied in development of education for a profession and used for recruiting.
 - Project funded by National Library (Norway).
 - Full-day online workshops with interviews, 40 participants.
 - Libraries are the central point in terms of skills for data management.
 - Challenges: role and expectations, accreditation, career path(s).

RDA-B3-1: FAIRification of Genomic Tracks: data-driven life science through granular discovery of biological sequence annotations with uniform metadata

 *When & where:* 24 Oct 16:00-17:30, Doppler Hall, breakout, [details](#)

 *RDA Groups:* BoF 

 *RDA Pathways:* FAIR, CARE, TRUST - Adoption, Implementation, and Deployment; Semantics, Ontology, Standardisation; Discipline Focused Data Issues

 *Reported by:* Paulette Lieby, Sveinung Gundersen

Summary:


- Initiative for a new RDA WG: [FAIRification of Genomic Annotations](#)
- Previous work: FAIRtracks
 - Unified model for data analysis: mental model of coordinate system along a reference genome
 - ELIXIR Implementation study + RIR
- BoF; three use-cases
 - Three use-cases
 - Collect biomedical datasets to build AI models
 - Annotations in biodiversity as FAIR objects
 - Enhance Track Hub infrastructure with metadata
 - How (recommendations)
 - Harmonised metadata schemas
 - Executable metadata transformation flows (ie mappings): maintainable and scalable data flows
 - Persistent storage
 - PIDs
 - Standardised search APIs


Collaborative session notes: <https://bit.ly/48Mdlm7>


- Supported by RDA Tiger
 - Service provision for WG
 - Lower barrier to WG activity
 - Till end 2025, 4 calls per year
- Background: FAIRtracks
 - Unified model for data analysis: mental model of coordinate system along a reference genome
 - → Genome browser
 - But also for non-visual analysis
 - Different data types → genomic tracks
 - Genomic tracks
 - Summaries of raw data
 - Enables discovery of genomic datasets and their relationships
 - Lots of big consortia involved, and small projects
 - Each with its own data portal, data model....!!
 - So: get some alignment in there; started ELIXIR-NO and SP
 - Outcomes
 - Recommendations from the FAIRtrack project
 - Services (RIR)
- Building an integrating repository for genomic data Anna Bernasconi
 - 2016-2021
 - 3 axes

- Genomic data model
 - Genomic query language
 - Conceptual model (enriching semantics)
 - → discover relationships etc... across a broad range of heterogeneous data
 - Semantic search...
- Genome browser EMBL-EBI Peter Harrison
 - e!Ensembl, the Track Hub Registry
 - ELIXIR CDR
 - Focus on standardisation across different species
- Why two genome browsers? Two registries?
 - A number of browsers
 - Different software, various context
 - Different genomes: a browser per particular coordinate system defined from a version of a reference genome assembly.
- Case statement (draft)
 - Naming? "Genome annotations"
 - Three use-cases
 - Collect biomedical datasets to build AI models
 - Annotations in biodiversity as FAIR objects
 - Enhance Track Hub infrastructure with metadata
 - How (recommendations)
 - Harmonised metadata schemas
 - Executable metadata transformation flows (ie mappings): maintainable and scalable data flows
 - Persistent storage
 - PIDs
 - Standardised search APIs

RDA-B3-2: PID Exchange - a curated information resource for PID adopters

 *When & where:* 24 Oct 16:00-17:30, Europa Hall, breakout, [details](#)


 *RDA Groups:* [Working with PIDs in Tools IG](#), [National PID Strategies IG](#)


 *RDA Pathways:* FAIR, CARE, TRUST - Adoption, Implementation, and Deployment; Data Infrastructures and Environments - International; Data Lifecycles - Versioning, Provenance, Citation, and Reward

 *Reported by:* Marek Suchánek

- Introduction to the [National PID Strategies WG](#) + acks RDA TIGER + IG goals
- [PIDfest.org](#) (June 2024, Prague, CZ)
- Practical issues when implementing PIDs presented by Rory Macneil and Vaida Plankyte (Research Space), e.g., IGSN integration and other promoting interoperability.
- Xiaoli Chen (DataCite) presented support of FAIR Workflows with PID integrations mainly in the context of PID graphs and PID metadata in DataCite.
- Session notes: https://docs.google.com/document/d/110C2ZWqYLL4FkfgDZL_Ed-hFzgtDTRgR-fyE-YGhNoM/edit


Plenary Session: *RDA Adoption & Impact*

 *When & where:* 24 Oct 17:45-19:15, Europa Hall, plenary, details

 *Reported by:* Marek Suchánek

[RDA Strategic Plan 2024-2028](#) has been presented with focus on sustainability, empowering, and innovation. There is also an implementation plan. Then, RDA Outputs (and where to find them) presentation covered how the outputs can be found through a new web interface including some impact metrics. There were also short presentations followed by panel discussion. The topics were about adoption of FAIR₄RS (ELIXIR, FAIR-IMPACT, Netherlands eScience Center), Core Trust Seal / Metadata Standards, and Brazilian perspective on outputs.

SDC-B4-1: Interoperability and Interconnectivity in Developing Open Science Infrastructures

 *When & where:* 25 Oct 9:00-10:30, Doppler Hall, breakout (mixed paper session), [details](#)

 *Reported by:* Paulette Lieby

Summary:

- Presentation of GOSC: Global Open Science Cloud Initiative
- Several infrastructures presented, of which the most interesting are
 - the African Open Science Platform (presented by NICIS Anwar Vahed)
 - Federated infrastructures
 - UbuntuNet alliance: provide connectivity across continent <https://ubuntunet.net/>
 - → services, tools, data science and analysis, education, priorities
 - Five regional open science nodes
 - Two accepted: Egypt (remote sensing and space sciences)
 - East Africa: capacity development (Kenya)
 - Challenges: Language barriers, Low visibility of scholarly output, Restricted access to research funding, Internet connectivity (mostly lack of infrastructures)
 - And Solutions: Pan-African Open Access portal, Existing digital infrastructure for research dissemination, With physical ownership of research data by African stakeholders
 - Australian Research Data Commons (Rosie Hicks, ARDC)
 - → Provide competitive advantages to researchers
 - Next
 - Construction of RDC research data commons
 - People, planet, HASS (SS and indigenous)
 - Fully funded for the next few years (5 years)
 - Q: what is secure research environment
 - Privacy concerns: a proliferation of different solutions
 - But that's not good enough: siloed systems
 - So ARDC: describe specs and technical requirements → integration

Notes:

- Intro to GOSC infrastructures: Global Open Science Cloud Initiative
 - UNESCO 2021
 - Federated services
 - New solutions to difficult data transfer issues: move processes to data
 - WGs
 - Technical infrastructures: AAI Federation
 - IAAS Cloud Federation model (open source, based on block chain)
 - Services operability
 - → co-develop testbed and deliver services
 - → build a toolkit
 - → refine framework


- African Open Science Platform (NICIS Anwar Vahed)
 - Federated infrastructures
 - → services, tools, data science and analysis, education, priorities
 - Five regional open science nodes
 - Two accepted: Egypt (remote sensing and space sciences)
 - East Africa: capacity development (Kenya)

- UbuntuNet alliance: provide connectivity across continent <https://ubuntunet.net/>
 - Challenges
 - Language barriers
 - Low visibility of scholarly output
 - Restricted access to research funding
 - Internet connectivity (mostly lack of infrastructures)
 - Solutions
 - Pan-African Open Access portal
 - Existing digital infrastructure for research dissemination
 - With physical ownership of research data by African stakeholders
 - Policy dev and dialogue
 - Training (UbuntuNet to participate)
 - Open repositories for publications and data
 - South African Open Science Cloud
 - NICIS: **National Integrated Cyber Infrastructure System**
 - Computing
 - Networking
 - Data services
 - Discovery, federated authentication, repository (focused on metadata; among others space data SKA, CERN), DMP tool, DOI minting
 - Q about sustainability
 - More node establishment
 - Support dialogue across continent
 - Infrastructure expansion across continent
- Australian Research Data Commons (Rosie Hicks, ARDC)
 - Funded by federal Gov from NCRIS
 - Provide competitive advantages to researchers
 - Half of 42 universities members, also CSIRO
 - Services
 - OpenStack Nectar
 - PIDs, communities of practice,
 - Next
 - Construction of RDC research data commons
 - People, planet, HASS (SS and indigenous)
 - People
 - Data strategy & discovery, secure data access, data integration, advanced analytics
 - Planet
 - Again, identify challenges
 - Trusted data supply chain
 - ...
 - Now, we have a clear pathway
 - Fully funded for the next few years (5 years)
 - Q: what is secure research environment
 - Privacy concerns: a proliferation of different solutions
 - But that's not good enough: siloed systems
 - So ARDC: describe specs and technical requirements → integration
 - Advancing technical interoperability and interconnectivity in China: Showcase of CSTCloud (Haiming ZHANG and Yuepeng LI)
 - Global resources
 - And local resources
 - → resource sharing is urgently needed

- CST cloud: manage CAS' resources uniformly
 - And provide integrated services (integration etc...)
- Research Center for Open Science and Data Platform, National Institute of Informatics (Mikiko, RCOS, NII)


SDC-B4-5: Why Persistent Identifiers and How Their Connectivity Enhances Research

 *When & where:* 25 Oct 9:00-10:30, Europa Hall, breakout (panel discussion), [details](#)

 *Reported by:* Marek Suchánek


The panel session focused on the significance of Persistent Identifiers (PIDs) in the global research ecosystem and the importance of connecting various PID systems. PIDs, which provide globally unique names and rich metadata for research elements, have become **vital components of research data infrastructure**. The session aimed to **showcase successful PID implementations**, emphasise the need for connectivity among PID systems, and explain why coordination in the research community was crucial. The session began with a presentation by GFZ German Research Centre for Geosciences on PID connections, followed by lightning talks from international organisations and stakeholder groups, including Crossref, DataCite, ORCID, RAiD, Jisc, GFZ Potsdam, IGSN e.V., RDA Complex Citations Working Group, and RDA/WDS Scholarly Link Exchange (Scholix) WG. The structured panel discussion explored approaches to **building interrelations among PID systems**, with dedicated time for participants to pose questions and discuss challenges and needs related to PIDs. Attendees gained insights into current PID connectivity efforts, future aspirations, and the rewards for the research community through these synergies.

RDA-B4-3: Researchers' engagement with data management: from cookbook to toolkit

 *When & where:* 25 Oct 11:30-13:00, Karajan-Hall 2+3, breakout, [details](#)

 *RDA Groups:* [Engaging Researchers with Data IG](#)

 *RDA Pathways:* Training, Stewardship, and Data Management Planning

 *Reported by:* Paulette Lieby

Summary:

- This IG: want engagement from everyone: collect and share best practices
- Cookbook <https://www.openbookpublishers.com/books/10.11647/obp.0185>
 - 24 case studies, started in Botswana
 - A Standardised collection
 - **Submit your case study: template**
https://docs.google.com/document/d/1NqcX_dKtVSpIraWqUqPUnsrubRmEdp5v29iyehpCUY/edit?usp=sharing or google form <https://forms.gle/WUaoEny7oRurdKDL8>
- Podcast - IDEA <https://www.idea.int/podcasts> , <https://podcasters.spotify.com/pod/show/datalibs>
 - 20-30 min episodes
 - Mostly from US for now, looking for diversification

Notes:


Collaborative session notes:

https://docs.google.com/document/d/1V93mqCCIAUKS75RMhHh1RZRajtZr9CUW8KvNz_ENI8/edit


- This IG: want engagement from everyone: collect and share best practices
- Cookbook <https://www.openbookpublishers.com/books/10.11647/obp.0185>

- 24 case studies, started in Botswana
- Standardised collection
- New calls: Template to download and fill in: https://docs.google.com/document/d/1NqcX_dKtVSplraWqUqPUnsrubRmEdp5v29iyehpCUY/edit?usp=sharing
- Podcast - IDEA <https://www.idea.int/podcasts>, <https://podcasters.spotify.com/pod/show/datalibs>
 - Because not all efforts in the RDM is published (indeed!)
 - So → easily digestible way of presenting material
 - 20-30 min episodes
 - An interview process
 - Mostly from US for now, looking for diversification
- Toolkit for online engagement activities....
- Create case studies; template https://docs.google.com/document/d/1NqcX_dKtVSplraWqUqPUnsrubRmEdp5v29iyehpCUY/edit?usp=sharing or google form <https://forms.gle/WUaoEny7oRurdKDL8>

RDA-B₄-6: Building towards FAIR for Machine Learning

 *When & where:* 25 Oct 11:30-13:00, Trakl-Hall, breakout, [details](#)

 *RDA Groups:* [FAIR for Machine Learning \(FAIR₄ML\) IG](#)


 *RDA Pathways:* FAIR, CARE, TRUST - Principles; Semantics, Ontology, Standardisation; Research Software

 *Reported by:* Fotis Psomopoulos, A B, X Y


// report here (text, bullets, etc.)

The session provided a quick update on the efforts of the FAIR₄ML IG since the last plenary, particularly in the context of the two Task Forces (“FAIR₄ML white paper and the ML lifecycle”, and “Metadata schemas for ML”). Following that, there was a discussion around potential interactions both at the organisational level and around standards on metadata and registries. A focus was around elements of the ML life cycle and the metadata required to better describe them.

Both elements of the session (ML lifecycle and metadata schema for ML) have a direct connection to the ELIXIR ML Focus Group effort (further emphasised through the presentation of the DOME recommendations). Within the context of the session there was a discussion around potential interactions both at the organisational level and around standards on metadata and registries. Particular focus was around the ML life cycle and the metadata required – this effort will be also continued in the context of a follow-up mini hackathon that was organised by NFDI₄DS in December 2023, with two key outcomes also available in Zenodo: [Lifecycle for FAIR Machine Learning](#) and [Towards metadata for machine learning - Crosswalk tables](#).


Collaborative notes:  [RDA P21 Collaborative Notes - Building towards FAIR for Machine Learning](#)

RDA-B₄-7: Revising the Versioning Principles: The Road to Actionable Recommendations

 *When & where:* 25 Oct 11:30-13:00, Mozart Hall 2+3, breakout, [details](#)

 *RDA Groups:* [Data Versioning IG](#)

 *RDA Pathways:* Data Lifecycles - Versioning, Provenance, Citation, and Reward

 *Reported by:* Marek Suchánek


In the presentation, the current status of the Data Versioning Interest Group (IG) was initially introduced, providing a concise overview of its ongoing activities and focus areas. The central portion of the discussion delved into the **practical application of the BIBFRAME 2.0 model within the context of data versioning**. The BIBFRAME 2.0 model,

designed to support the representation and management of bibliographic data in the digital environment, was explored for its relevance and potential implications in the realm of data versioning.


A significant part of the presentation was dedicated to a thorough examination of **Persistent Identifiers (PIDs) and their role in versioning**. This involved a detailed discussion on the distinction between release/version Digital Object Identifiers (DOIs) and collection DOIs within platforms such as Zenodo. Specifically, the conversation likely delved into the challenges and considerations surrounding the use of PIDs in the versioning process, addressing issues related to identifying and citing specific versions of datasets. The **comparison between release/version DOIs and collection DOIs in platforms like Zenodo** would have been elucidated to underscore the differences and implications for ensuring accurate and reliable versioning practices.


Collaborative notes: <https://bit.ly/P21-data-versioning>

RDA-B4-8: Let's talk about FAIR mappings! Towards common practices for sharing mappings and crosswalks

 *When & where:* 25 Oct 11:30-13:00, Mozart Hall 4+5, breakout, [details](#)

 *RDA Groups:* BoF 

 *RDA Pathways:* FAIR, CARE, TRUST - Principles; FAIR, CARE, TRUST - Adoption, Implementation, and Deployment; Semantics, Ontology, Standardisation

 *Reported by:* Wolmar Nyberg Åkerström

This BoF session proposed the formation of a new RDA group to discuss common practices for sharing mappings and crosswalks. Many RDA IGs and WGs have independently worked on or are working (indirectly) on mappings and while it's a waste of effort not to reuse mappings there is a lack of common practices and representations. The session aimed to showcase various types of mappings and models to share them and to start a discussion on how to make them FAIR and to align practices.

The FAIR Impact project is working on guidelines on how to make mappings FAIR, a machine-actionable common exchange model, and a governance framework for mappings. It includes practical aspects of mappings from creation to maintenance and has been guided by workshops with stakeholder from a wide range of communities such as the [Why Mappings Matter and how to make them FAIR?](#) and the upcoming workshop [Documenting mapping community practices](#). The preliminary model covers models and exchange formats, metadata, PIDs, and related services / APIs.

The FAIRCORE4EOSC project is working on building a repository for mappings called the [Metadata schema and crosswalk registry \(MSCR\)](#). It supports registering schema in a wide variety of formats ranging from JSON Schema and SHACL to CSV and PDF. It accepts mappings ranging from XSLT to SSSOM and also has a crosswalk editor with an internal format for mappings. Mappings converted to the internal format can also be executed using a separate data transformation engine.

A series of [13 flash talks](#) presented further examples of mappings and solutions for sharing mappings:


- Wim Hugo [[link to position in the slide deck](#)] – conceptual model, identifying vocabulary needs, defining a metadata schema, and specifying service methods and payloads for a Brokering Registry
- Kurt Baumann [[link to position in the slide deck](#)] – RML-based mappings from different Sources to schema.org
- Alexandra Kokkinaki [[link to position in the slide deck](#)] – examples of Unit Conversions in the marine domain
- Javier Millan Acosta [[link to position in the slide deck](#)] – BridgeDb for omics identifier mappings and export to SSSOM

- Gerhard Goldbeck [[link to position in the slide deck](#)] – [OntoCommons Bridge](#) approach applied to mappings in chemistry and materials
- Iseult Lynch [[link to position in the slide deck](#)] – mapping shape representations in (nano)materials
- Sveinung Gundersen [[link to position in the slide deck](#)] – omniipy for scalable and maintainable executable metadata crosswalks for harmonisation of genomic annotations
- Nicolas Matentzoglu (Presented by Yann) [[link to position in the slide deck](#)] – SSSOM to define mappings across terminologies used in rare disease resources
- Rorie Edmunds [[link to position in the slide deck](#)] – Crosswalk Recommendations for translation of IGSN metadata into the DataCite Schema V4.4
- Keith G Jeffery [[link to position in the slide deck](#)] – A proposed RDA metadata element set and RDA Metadata Standards Catalog to inform Metadata Mappings, Crosswalks, Conversion
- Andrea Mannocci [[link to position in the slide deck](#)] – Interoperability Framework for Scientific Knowledge Graphs
- Milan Ojsteršek [[link to position in the slide deck](#)] – Metadata mappings between different metadata schemas

The session concluded with a discussion on how to approach forming an RDA group on FAIR mappings supported by Sli.do [[live results](#)].

Collaborative notes: https://docs.google.com/document/d/164Gg3j_a4ZKPIPYDFIlogtX6o-PLiVVNCqh_VpcUVlw/edit

SDC-B5-7: Data Management Plans Presentations

 *When & where:* 25 Oct 16:00-17:30, Mozart Hall 4+5, breakout (mixed paper session), [details](#)

 *Reported by:* Marek Suchánek


There were X short presentations on DMP-related work:

- *ChatDMP* (Karl Benedict, UNM) – It was about the use of generative AI tools, such as ChatGPT, in the development of research project proposals and data management plans (DMPs). It acknowledges the opportunities and risks associated with using AI to generate DMPs and highlights the common desire among researchers for streamlined DMP development based on the reuse of existing templates. The University of New Mexico's Research Data Services team is conducting an applied research project to experiment with tools and methods for collecting structured data management plan components **aligned with the RDA DMP Common Standard**. The goal is to produce machine-actionable DMP data, which will be used to **generate narrative DMPs through the ChatGPT API**. The paper presents early results and incremental improvements achieved through experimentation with DMP prompt engineering. It also addresses the technical and ethical considerations of using generative AI in the research proposal development process, emphasising both the risks and opportunities in leveraging AI as an assistive technology for DMP creation.
- *Leveraging DMPs to improve Data Discoverability* (Soo-Yon Kim, RWTH Aachen University) – The Cluster of Excellence "Internet of Production" (IoP) aims to facilitate collaboration and knowledge exchange across the entire production chain. Large-scale interdisciplinary research projects, such as the IoP, face challenges in setting up effective strategies and infrastructure for data-driven collaboration among researchers. The paper identifies four major challenges: interdisciplinary nature, collaboration among various institutes, varying levels of experience and openness toward data management practices, and limited support capacities. To address these challenges, the authors propose a **solution leveraging data management plans (DMPs)** within the IoP. Researchers create DMPs using the RDMO tool, allowing data stewards to **automatically list and manage datasets in a registry for enhanced discoverability**. The approach aims to be **discipline-agnostic, non-disruptive to existing institutional solutions, low-barrier for researchers**, and supported by RDMO's API


for **automated data extraction and analysis**. This effort seeks to improve data discoverability and support data-driven collaborative research within large-scale projects.


- *Streamlining FAIR Data Stewardship with DMPs* (Marek Suchánek) – The presentation introduced the FAIR Wizard platform, **addressing challenges from the manual implementation of FAIR principles in Data Management Plans (DMPs)**. The FAIR Wizard streamlined FAIR data stewardship by integrating and automating DMP creation, customization, and execution. Key strengths included **seamless integration within organisational environments, ensuring consistency and accuracy in data stewardship**. The platform automated metadata extraction, data description enrichment, and FAIR guideline compliance validation, reducing the burden on researchers and minimising human error. The presentation highlighted the FAIR Wizard's features, benefits, and transformative impact on data management practices across diverse research domains through real-world examples and success stories. Overall, the **FAIR Wizard was showcased as a revolutionary tool** that accelerated scientific discovery, fostered an open research culture, and ensured long-term accessibility and usability of data for the wider research community.
- *Data Management Planning in South Africa* (Nobubele Shozi) – The work presented the state of research data management planning in South Africa, emphasising the need for effective planning in collecting, organising, storing, and caring for data during its lifecycle. Currently, **data management planning in South Africa operates in silos**, lacking integration and machine-actionable capabilities. Research funders in South Africa now require data management plans for fund allocation. The South African Data Management Planning tool, developed for national use by researchers, **aims to provide interoperability with DIRISA services**. Version 2 of the tool **facilitates automated DMP submission, streamlining the process for requesting DIRISA services**. The paper highlights the adoption of institutional DMPs on the SA DMP Tool by various South African research entities and underscores DIRISA's efforts towards machine-actionable DMPs.

RDA-B5-3: Global open research commons: what it takes and the road to get there. Reflections and discussion on creating a commons attribute model and a commons integration roadmap

 *When & where:* 26 Oct 9:00-10:30, Karajan-Hall 2+3, breakout, [details](#)

 *RDA Groups:* [GORC International Model WG](#), [Global Open Research Commons IG](#)


 *RDA Pathways:* Data Infrastructures and Environments - Institutional; Data Infrastructures and Environments - Regional or Disciplinary; Data Infrastructures and Environments - International

 *Reported by:* Marek Suchánek


During this session, the GORC IG and GORC WG introduced their objectives, recent developments, and supporting outputs, leading to a **lively discussion on the GORC-WG International Model**. Various potential future projects and areas of interest were explored, and additional adopters of the model were identified. Key outcomes included the decision to move forward with the adoption of the model and a plan to resubmit it as a recommendation. The GORC vocabulary, comprising essential elements and a model glossary, will transition into a persistently identifiable space using the ARDC vocabulary service. Furthermore, the **GORC model will be transformed into RDF/JSON-LD format**, and its application to the health data space will result in a discipline-specific profile and an implementation map. This approach will also facilitate the creation of a more general implementation map for the model, incorporating scaling heuristics, which can be integrated into the IG roadmap for direct translation to application rules. The session clarified the goal of creating a roadmap and a common model, highlighting the typology of essential elements for integrating different commons. The collaborative efforts involved numerous meetings, interviews, reviews, and task groups, initially captured in a spreadsheet.


Collaborative notes: https://docs.google.com/document/d/1GJQFAy92hgyTYKdYByRlrsI_3i8ah2hh2qasulUU7Ok/edit

RDA-B5-7: 2nd BOF: Trusted / Secure Research Environments for Sensitive / Confidential Data: FAIRness for "Closed" Data and Processes

 *When & where:* 26 Oct 9:00-10:30, Europa Hall, breakout, [details](#)

 *RDA Groups:* BoF 

 *RDA Pathways:* FAIR, CARE, TRUST - Principles

 *Reported by:* Paulette Lieby

Summary:

- Hoping to submit WG proposal soonish
- Big Q:
 - HOW to run TRE (trusted research environments)
 - Data visiting, bring compute to data, etc...
 - Also need to collaborate across national, disciplinary, ... boundaries
 - New technologies emerging
 - Bring together infra providers, data stewards, data owners, researchers
- Four problems, 4 work packages
 - Blueprints
 - What are the components?
 - Interoperability between TREs


- What are the risks / risk levels?
- FAIRness for closed data: what metadata can be made available?
- Comment: lots of discussion around the difficulty of the task, but a consensus such a WG must be set up; but one or four WGs?
- Most important outcome expected: clear definition of the problem


Notes:

- Collaborative notes :
 - <https://docs.google.com/document/d/1Fv9TOGlXMIg99OIGotlzo4l8leS-fkYrwtAGpsb5p6M/edit#heading=h.2aqsh8u3zaqw>
 - <https://www.rd-alliance.org/2nd-bof-trusted-secure-research-environments-sensitive-confidential-data-fairness-closed-data-and>
- Introduction
 - Hoping to submit WG proposal soonish
 - 4 work packages
 - Big Q:
 - HOW to run TRE (trusted research environments)
 - Data visiting, bring compute to data, etc...
 - Also need to collaborate across national, disciplinary, ... boundaries
 - New technologies emerging
 - Bring together infra providers, data stewards, data owners, researchers
 - Four problems
 - Blueprints
 - What are the components?
 - Interoperability between TREs
 - What are the risks / risk levels?
 - FAIRness for closed data: what metadata can be made available?
- P1 blueprints
 - A shared understanding of components, processes, roles, functionalities, vocabularies
 - A lit review
 - Data access
- P2 Interoperability; several levels
 - Social arch between people: roles, ...
 - Machine-actionable interfaces
 - Trusted interfaces
 - Compliant in federating users in TREs: legal interoperability
 - Q: do we have the right people in this WG? Esp. legal? .. good question
 - Q: how can one trust software?
 - →different risk profiles
 - Comment: people seem to think it too ambitious, and narrow the focus
- P3 Risks
 - Identify a risk catalogue
 - Get an understanding of the most likely risks
 - → ie understand what we are doing
 - "Identifying structure mistrust"
 - Q: document best practices

- P4: metadata, open, closed?, rich?

SDC-B6-3: Panel Discussion: Standards for a FAIR Data Future: How to Advance Standardisation in Data Management?

 *When & where:* 26 Oct 11:30-13:00, Europa Hall, breakout (panel discussion), [details](#)

 *Reported by:* Paulette Lieby

Summary:

Key questions about standardisation efforts in RDM will be addressed:

- What makes a good standard? What are necessary and supporting conditions?
- Which successful standards do we already have? And which do we need to develop?
- How can we advance adoption and compliance with existing standards?
- How can we connect standardisation efforts internationally and cross-disciplinary?

Notes:

<https://www.scidatacon.org/IDW-2023-Salzburg/sessions/522/>

Key questions about standardisation efforts in RDM will be addressed:

- What makes a good standard? What are necessary and supporting conditions?
- Which successful standards do we already have? And which do we need to develop?
- How can we advance adoption and compliance with existing standards?
- How can we connect standardisation efforts internationally and cross-disciplinary?

- Petra Ritter, prof. In brain simulation
 - Legal framework on health data: re TREs (digital simulated brains); within EOSC, ie sensitive and personal data
 - Standards: metadata
 - Semantic, technological, and legal interop
 - Building use-cases

- Stefan Weisgerber from DIN , the German Institute for Standardization
 - Mostly involved in industry data
 - Interlinked with research data?
 - Successful standardisation requires cosmopolitan standards


- Brigitte Mathiak
 - Metadata and knowledge graphs
 - Hard vs soft standards: well defined vs not very well defined, adopted...
 - So: 1. Define, then 2. Operationalise so that if errors, there will be feedback

- Andreas Kosmider
 - regulation/policy level
 - Huge pb of unused data
 - Keep it simple

- Christian Grimm
 - Selecting and applying standards in network infrastructure
 - Because interop is critical!!

- Qs:
 - What is standardisation?
 - It depends.... ;) what is the context? Etc...
 - What do I want the standards for?
 - Rather a lower-level focus: make the data accessible as this would allow research questions arise, and then ask what the right standards should be
 - Having a clear use-case or not? What about unknown data? So, do we need a strategy or not?
 - Needs politically enforced incentives
 - Rather a push-pull approach
 - It needs investment, but there an obvious benefit
 - How can we improve adoption?
 - Important to adopt outside your own community
 - But needs to be 'easy' to adopt, ie documented etc...
 - GDPR is not the problem but poor knowledge of GDPR
 - Give a proper example of how to do things
 - But: what happens when it doesn't work: should we say "it cannot be done"?
 - Take into account possibly vastly interpretation/transcription of GDPR in different countries

SDC-B6-5: Semantics / Metadata Oriented Presentations


 *When & where:* 26 Oct 11:30-13:00, Mozart Hall 2+3, breakout (mixed paper session), [details](#)

 *Reported by:* Sveinung Gundersen

- Six papers. Highlights of particular relevance:
 - ***NHLBI BioData Catalyst®: Democratizing Data Discovery and Computational Resources to Harness Data Science for Precision Medicine and Beyond (Sweta Ladwa , Regina Bures)***
 - Cloud-based ecosystem that facilitates analysis and access to data (at scale)
 - 4 Petabytes of data, including:
 - Largest and most ethnically diverse set of whole-genome sequences from Trans-Omics for Precision Medicine (TopMed) program
 - Longitudinal cohort epidemiology studies, clinical trials, Covid-19 data, imaging data...
 - July 2021 - gone public
 - Multiple Data/Metadata models - per domain
 - Analytical workspaces, pipelines, etc.
 - GA4GH Driver Project
 - ***From Lowly Spreadsheets to High-Quality Metadata: An End-to-End Approach for Easing Adherence to Community Standards for Scientific Metadata (Martin O'Connor et. al)***
 - Extension to CEDAR workbench (see also [SDC-B3-3: Unleashing the potential of domain specific metadata](#))
 - End-to-end metadata management solution
 - Ensures compliance with community standards
 - User groups did not like web form for entry
 - Autogenerates spreadsheets from CEDAR templates
 - Ontology/vocabulary terms as selections

- Also tools for validation and ingestion of CEDAR spreadsheets
- **Decentralised Semantics: A Semantic Engine user perspective (Carly Huitema et. al)**
 - Suite of tools called "Semantic Engine"
 - Connecting datasets with schemas
 - Overlay Capture Architecture (OCA)
 - Each feature is a layer of the schema in OCA (like a Docker image)
 - Different tasks and different contributors
 - JSON based
 - Content-based identifiers to link and track content
 - Multi-language support
 - <https://semanticengine.org>
 - UI tools to help create schemas

SDC-B6-6: National / Institutional Approaches Presentations

 *When & where:* 26 Oct 11:30-13:00, Karajan-Hall 1, breakout (mixed paper session), [details](#)

 *Reported by:* Marek Suchánek


There were 6 papers presented in this session:

- *Wageningen Common Data Solutions, developing a Data Mesh* – The presentation outlined Wageningen University & Research's (WUR) past approach, Wageningen Common Data Solutions (WCDS), which **adopted the iRODS software environment** to address multidisciplinary data challenges. WCDS combined technology development with education, training, policies, planning, community building, and external partnerships to integrate tools into WUR's research environment. The Wageningen Data Competence Center (WDCC), established in 2017, played a central role in strengthening the combination of environmental, plant, animal, food production, and social sciences with interoperable data. The presentation explored the alignment of WCDS with the data mesh concept, aiming to transform past data management into a practical and rewarding activity contributing to societal value in research and education.
- *Charting multidimensional academic careers: how can repositories contribute?* – The presentation explained the Dutch efforts since 2019 to broaden recognition in academia, focusing on Open Science values and breaking from traditional reward practices. **The Dutch program prioritises quality assessments** but acknowledges the potential increase in academic experts' workload. It introduced the myresearch.institute prototype service, which automatically traced and collected diverse contributions made by researchers to various web portals, providing insights into their Open Science activities. The presentation suggested a **contributor-centric approach** akin to a Current Research Information System (CRIS) for profiling researchers' activities, highlighting challenges, including the lack of support in scholarly platforms for retrieving contributions based on researcher identifiers like ORCID. It proposed initiatives for scholarly **repositories to become more contributor-centric** by supporting automatic collection of information about uniquely identified researchers' contributions.
- *The Australian Research Data Culture Conversation* – The presentation discussed the challenges in measuring research data volume management in Australian institutions due to the lack of alignment on definitions for national research data assets. It highlighted the increasing curation load resulting from a 40% compound annual growth rate (CAGR) in central research storage infrastructure between 2017 and 2020, emphasising the need for sustainable practices under financial constraints. The Research Data Culture Conversation (RDCC) was introduced as a collaborative effort by five major Australian universities to **address the impact of the data**


deluge and develop effective research data culture practices. The RDCC identified the importance of addressing beliefs, values, and practices in research data culture and emphasised the need for machine-actionable research data management plans (RDMPs) tied closely to university services. Additionally, the **concept of "Green and Pink Space"** was introduced to illustrate the distinction between the objectives and obligations of institutional digital corpus and research community data, highlighting the need for institutions to align infrastructure design with community objectives for effective research data management.

- *Data Stewardship: New challenges and opportunities in the Swiss University environment* – The Swiss National Strategy for Open Research Data (ORD) focuses on **maximising the openness of research data** in line with FAIR principles. Good Research Data Management (RDM) practices are crucial for ORD implementation, particularly considering the disciplinary diversity in RDM and data publication. The emergence of "Data Stewards" internationally, such as in the Netherlands, Belgium, Austria, and Germany, addresses domain-specific support in RDM. In response to the ORD Action Plan, the University Library of Bern, University of Bern (UniBE), Switzerland, aims to **strengthen research quality, provide cross-domain support on critical topics, and develop sustainable infrastructures and services.** The newly established domain-specific RDM support, including Data Stewardship, covers sensitive data management, data science, and reproducibility, offering subject-specific assistance to researchers throughout the project lifecycle. This initiative is supported by the delegation of Open Science "swissuniversities" ORD Action Plan B5.2 Data Stewardship.
- *Linking publications and datasets: connecting KU Leuven's institutional literature and data repository* – LIBIS, the digital service provider of KU Leuven Libraries, enhances the **connection between its data repository (RDR) and literature repository (Lirias).** Through a nightly export, metadata of published datasets is registered in Lirias, creating a bidirectional link using DOIs or Lirias IDs. The Dataverse installation at KU Leuven incorporates custom look-ups for author fields and related publications, simplifying input and strengthening connections with Lirias. To **automate metadata retrieval for externally published datasets,** a tool utilises DOIs, reducing manual entry for researchers and improving the university's dataset overview. These initiatives serve as inspiration for other institutions seeking to enhance connections, bolstering findability, reusability, and reporting capabilities.
- *A Dataverse integration dashboard: pulling data from data management systems* – A year ago, the LIBIS team at KU Leuven developed a versatile dashboard for their Dataverse-based institutional data repository, RDR, enabling users to pull data from external tools like **iRods, OSF, Gitlab, and GitHub into a dataset.** The decision to reverse the typical workflow, pulling data from external tools, not only offered flexibility and ease of integration but also led to a more creative and user-friendly interface. Beginning with GitHub as a test case, the team extended the dashboard to GitLab and realised its **potential to easily accommodate other tools, making data source additions straightforward with basic API calls for each tool.**

RDA-B6-4: Key aspects of FAIR for infrastructure solutions in the biomolecular life sciences

 *When & where:* 26 Oct 14:00-15:30, Trakl-Hall, breakout, [details](#)

 *RDA Groups:* [Life Science Data Infrastructures IG](#)

 *RDA Pathways:* FAIR, CARE, TRUST - Adoption, Implementation, and Deployment; Data Infrastructures and Environments - Regional or Disciplinary; Discipline Focused Data Issues

 *Reported by:* Paulette Lieby, Wolmar Nyberg Åkerström

Summary:

- What happened at P20
 - A reflection on the landscape; a few comments
 - Different infrastructures do not support the same stage of the data lifecycle, and not at the same maturity level
 - Not all functionalities need to be built for FAIRness
- Post-plenary reflections
 - Could access to data and services be made more seamless?
 - Research infrastructure in context
- Gathering a few points of critical importance

Notes:

Collaborative notes:

<https://docs.google.com/document/d/1ihlqsDjSd5o9PDOmwRbDaDpxYlqV1mSwOvNnYAnSN5w/edit>

This session was a continuation on the topic "[Infrastructure supporting the FAIR data principles in life science research practice](#)" [[recording](#)] introduced during the P20 plenary session in Gothenburg. The session aimed to host a discussion on key aspects/recommendations that could help life science infrastructures better support FAIR digital assets framed by landscape overview based on the presentations and input to the P20 session as well as some post-plenary reflection on data life cycles and how access to data and services across infrastructures could be made more seamless.

Reflections based on the 7 presentation from the P20 plenary session:

- Different infrastructures support different stages of the research data lifecycle in research projects
- Standards and benchmarks can support data repositories, tool repositories, workflow repositories, workspaces
- User-friendly interfaces, automated validation and curation support can improve user-supplied metadata

Poste plenary reflections:

- Often a complex patchwork of infrastructure services where parts of data and metadata are produced and handed over across different providers
- An overview of digital assets and metadata in the context of a data life cycle can guide discovery of relevant standards and hand-over points across infrastructures
- Adopting a federation's perspective on infrastructures could support convergence on shared shared standards and practices

RDA Digital assets and metadata in context (II)

Enabling reproducible, transparent research.

scientific hypothesis

PUBLICATIONS SLIDES
DATA METADATA
RESULTS LOGS
WORKFLOWS

researchobject.org
<https://www.researchobject.org/>

576+ Formats
788+ Terminologies
245+ Guidelines

Linked Open Data Executable Discoverable Reproducible

RDA Digital assets and metadata in context (II)

Study & data design Sampling & specimen collection Sample preparation Sample analysis & data generation Data processing to prepare inputs for analysis Data analysis Communicating results

Procedures: data protection, ethics permits, practical processing steps, infrastructure standards, sample quality assessment, data back-ups, labo access, ...
Biosamples and instruments: procedures (biological) and inclusion criteria, infrastructure standards, long term storage facilities, sample quality assessment, sample identification, magnets, instruments, kits, ...
Data and computational workflows: digital processing steps, working storage conditions, long term storage facilities, data quality assessment, sample/data metadata, reference data, analysis method, ...
Outputs: publications, data, workflows, reports, identifiers, ...


Key aspects of FAIR for infrastructure solutions in the biomolecular life sciences | RDA 21st Plenary Meeting 41

The session concludes with a discussion on Key aspects of FAIR for infrastructure summarised in the [collaborative session notes](#). The discussions were guided by a poll and covered challenges and desirable features from a research infrastructure users perspective and solutions that infrastructure managers provide to support FAIR and areas where they are looking for further advice. Other topics that were discussed include engaging instrument manufacturers in discussions around FAIR and how to approach capturing audit trails across infrastructures.


- Today: The key aspects for a FAIR infrastructure
- P20 summary
 - Reflection on the landscape
 - Mostly Generalist resources
 - How FAIR are they (self-assessment)?
 - F: better metadata
 - A: greater automation and integration
 - I/R: idem
 - General: better guidelines for certain type of infrastructure
 - Presentations P20
 - Different infra do not support the same stage of the data lifecycle, and not at the same maturity level
 - Some stages are overrepresented (eg analysis)
 - Not all functionalities need to be built for FAIRness
 - Broad themes
 - Repositories and workspaces: observing standards
 - Tools to do FAIR assessment (benchmarking)
 - User-supplied metadata if often lacking
 - Post-plenary reflections
 - Could access to data and services be made more seamless?
 - Research infrastructure in context
 - Sharing data and metadata in context (bundling eg RO-crate)
 - Pb: there are so many standards...


- Answer?: Federated infras and standards: then aggregation etc...
- Key aspects of FAIR for infrastructure
 - Researchers
 - Using several infrastructures in research: lots of handover points
 - Challenges in this...

RDA-B6-5: Equity and Inclusion: universal access to PIDs

 *When & where:* 26 Oct 14:00-15:30, Mozart Hall 4+5, breakout, [details](#)

 *RDA Groups:* [PID IG](#)

 *RDA Pathways:* FAIR, CARE, TRUST - Principles; FAIR, CARE, TRUST - Adoption, Implementation, and Deployment; Data Infrastructures and Environments - International

 *Reported by:* Marek Suchánek

During the collaborative session, **Nobuko Miyairi was introduced as the new co-chair for PID IG**, bringing her expertise and enthusiasm to the group. Tshiamo Motshegwa, representing Prof. Muliaro Wafula of the African Open Science Platform (AOSP), discussed the need for infrastructures to showcase African research data. This involves creating metadata and computing structures for interoperability, securing authority for sharing, and obtaining people and funding for sustainability. The session covered various projects, including H3ABioNet, South African National Biodiversity Institute (SANBI), DataFirst - CTS, African Academy of Science Open Research, The Square Kilometre Array, Indigenous Knowledge, and initiatives such as the Kenya Open Data Instruments.

Joy Owango presented the **Africa PID Alliance's approach** to using Persistent Identifiers (PIDs) for equity, focusing on Indigenous knowledge, cultural heritage, and patents. Challenges were highlighted, including the lack of national information on patents at a continental level and the goal to increase the **visibility of African research and knowledge**. Helena Cousijn discussed the DataCite Global Access Program, emphasising outreach, tech infrastructure development, and funding to support PID infrastructure awareness in lower-income countries. Paloma Narin-Arraiza presented the **ORCID Global Participation Program**, comprising a membership equity program and a global participation fund, supporting community outreach, technical implementation, and integration.

During the Q&A session, the community discussed ways to support global equity in data infrastructure, emphasising the importance of capacity building, policy instruments, open science frameworks, and local engagement. The need for robust engagement to develop coordinated policies and awareness of PIDs and research infrastructure was highlighted, stressing out the significance of sharing technical and implementation knowledge for effective PID adoption in Africa.

Collaborative notes:

https://docs.google.com/document/d/12GWSN86oWmIF_AxxAALwVmoVyFXpgrz2oVtGE-DXPJw/edit