RDA 25th Plenary Meeting [part of IDW 2025]
Collaborative Notes

Group(s) name(s) / *session title*	Life Science Data Infrastructures IG - *Linking Pixels, Proteins & Populations: Integrating Data Across Life Science Domains*
Session link for more information	https://www.rd-alliance.org/groups/life-science-data-infrastructures-ig/plenary-participation
Session programme time	Wednesday-Thursday, 15-16 October 23:00 – 00:30 UTC / Thursday, 16 October 09:00-10:30 AEST, Breakout 4
Presentation slides	RDA P25 Life Science Data Infrastructures IG.pptx

✂----------

Session summary (to be filled by group chairs):

We will use the content in the table below to highlight your work to the RDA community in a report by the Technical Advisory Board (RDA P24 report example). **Please complete ALL fields below by *Friday, 31 October*, to be included in the report** and other communication. We also ask you to update your group webpage with the key takeaways from your post-plenary session.

1.	**Describe what your session is about in three sentences maximum:** *This session explored emerging practices and standards for integrating and sharing multi-modal data (imaging, structural, genomic, and other life science data types). Presentations highlighted experiences from national efforts to support multi-modal studies such as BioFAIR UK and Melbourne "PX4 Health", regional efforts to connect services from across digital infrastructures focusing on overlapping and complementary modalities exemplified by the emerging European Life Science Connect EOSC Node, and global efforts to connect multi-modal data resources exemplified by the foundingGIDE (GIDE - Global Image Data Ecosystem) project.*
2.	**Highlight a maximum of five key outcomes/actions/takeaways of your session:** 1. *Tony Burdett, BioFAIR (UK) – introduced three conceptual layers of multimodal integration (assay-based, multi-experiment, and cross-domain) while emphasising the role of persistent identifiers and provenance to link data "from assay to insight."* 2. *Bernie Pope, PX4 Health (Australia) – provided examples from a multimodal phenotyping and multi-omics data platform addressing reproducibility and batch effect challenges through dedicated instrumentation and standardised workflows.* 3. *Maria Mirza, Founding GIDE (Global Image Data Ecosystem) – described efforts*

	to harmonise metadata across international bioimaging repositories to enable shared APIs and cross-repository querying for a wide range of imaging modalities. 4. *Aastha Mathur, EOSC Life Science Connect Node (Europe) – shared experiences from developing a federated model connecting ELIXIR, Euro-BioImaging, EMBL, and Instruct-ERIC within the European Open Science Cloud to support interoperable FAIR data and services.* 5. *Nils Hoffmann, MOMSI Working Group (RDA) – outlined the RDA Multi-Omics Metadata Standards Integration (MOMSI) WG's landscape review of existing omics standards, published in FAIRsharing, that can guide harmonisation and identify gaps and priorities for future interoperability work.*
3.	**What collaborations or synergies did you discover between your group and other RDA group(s) or external (outside of the RDA) organisations?** • *RDA Sample Type Classification WG* • *RDA Multi-Omics Metadata Standards Integration (MOMSI) WG* • *HUPO PSI MS Metabolomics Standards Initiative Lipidomics Standards Initiative, https://hupo.org/Proteomics-Standards-Initiative-(PSI)* • *Science for Life Laboratory (SciLifeLab) Sweden, www.scilifelab.se* • *foundingGIDE, https://founding-gide.eurobioimaging.eu/* • *BioFAIR UK, https://biofair.uk/* • *Life Sciences Connect EOSC Node, https://eosc.eu/building-the-eosc-federation/eosc-node-life-sciences-connect/*

✄- -

Attendee Check-in

New to the RDA? Check out where to start **here**
Get involved in the **RDA Community**
This meeting will take place in accordance with the **RDA Code of Conduct**

Please complete this table to indicate your attendance (add rows as needed):

Full Name	Affiliation	Location	Email
Wolmar Nyberg Åkerström	ELIXIR Sweden / NBIS, SciLifeLab	Uppsala, Sweden	wolmar.n.akerstrom@uu.se
Jeff Christiansen	Australian BioCommons	Brisbane, Australia	jeff@biocommons.org.au
Tiff Nelson	Australian BioCommons	Newcastle, NSW, Australia	tiff@biocommons.org.au
Chuck Cook	Global Biodata Coalition	Cambridge, UK	ccook@globalbiodata.org
Nils Hoffmann	de.NBI & ELIXIR Germany, Research Center Jülich	Bielefeld, Germany	n.hoffmann@fz-juelich.de
Maria Mirza	Euro-BioImaging ERIC/foundingGIDE	Heidelberg, Germany	maria.mirza@eurobioimaging.eu

Tony Burdett	BioFAIR	Cambridge, United Kingdom	tony.burdett@biofair.uk
Julia Gehrmann	University Hospital Cologne	Cologne, Germany	julia.gehrmann1@uk-koeln.de
Marek Cebecauer	Heyrovsky Institute	Prague, Czech Republic	marek.cebecauer@jh-inst.cas.cas.cz
Aastha Mathur	Euro-BioImaging ERIC	Heidelberg, Germany	aastha.mathur@eurobioimaging.eu
Athina Papadopoulou	RDA Europe	Greece	athina.papadopoulou@rda-foundation.org
Steven Bentley	AGRF	Brisbane, Australia	Steven.bentley@agrf.org.au
Brian Collyer	AIHW	Brisbane, Australia	brian.collyer@aihw.gov.au
Aline Grand	Université de Strasbourg	Strasbourg, France	alinegrand@unistra.fr
Milan Ojsteršek	University of Maribor	Slovenia	milan.ojstersek@um.si
Yiming Bao	China National Center for Bioinformation	Beijing, China	baoym@cncb.ac.cn
Johan Gustafsson	Australian BioCommons	Adelaide, Australia	johan@biocommons.org.au

Meeting Aims

This meeting will explore emerging practices and standards for integrating and sharing multi-modal data—including imaging, structural, genomic, and other life science data types. The session will present early experiences from European efforts to bring together services from ELIXIR, EMBL, Euro-BioImaging, and Instruct-ERIC as part of a federated Life Science Research pilot node within the European Open Science Cloud (EOSC). It will also provide an opportunity to connect with global initiatives such as the Global Image Data Ecosystem (GIDE), to identify common challenges and promising approaches that could inform broader data integration efforts across biomolecular disciplines.

The examples presented will provide a starting point for discussion with the wider RDA community to identify opportunities for adoption of standards, future work, and new collaborations, emphasizing practical considerations for designing interoperable, cross-domain data services and supporting FAIR data sharing practices. The meeting outcome will be a short report summarising the approaches discussed, highlighting shared challenges and potential areas for collaboration, and outlining next steps for engaging interested participants in ongoing or new initiatives.

The meeting will focus on examples from the biomolecular life sciences but is open to and will provide context for anyone interested in disciplinary challenges related to combining different types of research data.

Agenda

Time	Topic	Format	Speakers / Facilitators

(AEST)			
09:00	Welcome & Introduction (10min)	Brief introduction to the RDA group(s) hosting the session followed by an interactive icebreaker to the topic where all participants are invited to engage.	Wolmar Nyberg Åkerström Jeff Christiansen
09:10	Setting the scene: Linking Pixels, Proteins & Populations (30min)	coordinated flash talks providing different perspectives on integration and sharing multi-modal data followed by Q&A	Tony Burdett Bernie Pope Maria Mirza Aastha Mathur
09:40	Discussion: Practices and standards for integrating and sharing multi-modal data (25min)	Guided discussions in smaller groups with reporting supported by Mentimeter and collaborative notes.	Wolmar Nyberg Åkerström
10:05	Synthesis: State of the art and opportunities for future collaborations (20min)	Reporting from the smaller groups and open plenary discussion moderated by the session chairs	Wolmar Nyberg Åkerström Nils Hoffmann
10:25	Wrap-Up & Next Steps (5min)	Summary of today and where next	Wolmar Nyberg Åkerström
10:30	END		

Collaborative Session Notes *(To be used by participants and chairs during the session)*

4 presentations:

Tony Burdett BioFAIR.

3 conceptual layers of multimodal integration:

- Multimodal assays (one expt, several different techniques) - e.g. spatial omics (webAtlas from HCA, wheat atlas)
- Multimodal Experiments (several coordinated experiments to probe one phenomenon) - e.g. conducting molecular (multi-omics) and cellular phenotyping to create a holistic feature of a single perturbation (MorPhiC)
- Cross-domain analysis - combining and resign data of multiple modalities for exploration or hypothesis testing - e.g. HoloFood (gut microbiome of framed animals + genomics from hosts). Farm Scale Digital twinning from Rothamsted - drone, climate

Combine / linking data at Sample level (e.g. BioSample)

BioFAIR BioCOmmons - Data COmmons + Methods COmmons + People COmmons.

Qs for Tony:

- Any reflections on what are the common focal points for integration? E.g. molecule, cell, tissue, organism, population? UniProt, BioSamples, … where is there global consensus?
    - A: Not yet a consensus. Can model things explicitly with a detailed knowledge graph, but can be very difficult to tease apart later. Very use case dependent. Think about anticipated reuse of the data. E.g. plant cross breeding - very complex breeding program (20 generations) - consider the last generation + passport info on how the last generation was produced could be sufficient.
- Where do you see the greatest practical challenge in making multi-modal data integration possible? Are formats and repositories often too specialised for one modality?
- What strategies have proven effective in motivating people to invest the effort required to add comprehensive metadata on platforms such as BioSamples?

Bernie Pope, Australian BioCommons
PX4 Heath initiative UoM.
Assay based multi-modal approach
Phenotyping clinic - variety of tests (BMI, clinical tests, wearables etc)
Integrated multimodal omics (genomics, epigenetics, metabolomics, lipidomics, etc)
Dedicated equipment to assist in consistency.

Multimodal data = combo of data of multiple modalities about the same individual and groups of individuals.

Challenges:

- Batch effects - especially relevant for longitudinal studies
- Highly dimensional data, so dimension reduction will be necessary
- Interoperability - with external studies (e.g. UK BioBanks, Precise SIngapore)
-

Qs for Bernie:

- Is there a difference in what you do to enable effective primary use of the data and what is later asked for to enable reuse of it? How did you approach harmonising data/metadata?
- using dedicated mass spec instrument does not sound very scalable. Do you consider implementing some calibration approach?
    -
- Who should drive the work on methodology for multi-modal data integration?

Maria Mirza - EuroBioImaging
FoundingGIDE - increasing coordination among global bioimage data resources. Image file sizes are ever increasing, >160 file formats. Aiming to bring together DBs across Europe, Japan, Australia. Can 3 repositories identify shared info, and arise at an API to query across. Metadata was harmonised.
Once the foundations are built, aiming to link bioimaging data to other data types - e.g. Same study - imaging data (in BioImag Archive) + Proteomics data (in PRIDE), Transcriptomics data (ArrayExpress), and Data Analysis code (GitHub)

Qs for Maria:

- Are the descriptive models for the Global Image Data Ecosystem aligned with those of Omics and other modalities? E.g. for samples, molecules, cells …

- The approach is to describe the images using sufficient metadata and to be able to iteratively add more metadata as needed. The metadata model is developed iteratively using feedback from the community and it is aligned with the imaging community needs.

Aastha Mathur - EuroBioImaging

A life Science Node in EOSC
EuroBioImaging + ELIXIR, Instruct ERIC, EMBL
The vision of a web of FAIR Data and Services to enable horizontal analysis
EOSC - sharing infrastructure, data software practices in an interoperable way.
The life science node - a dedicated

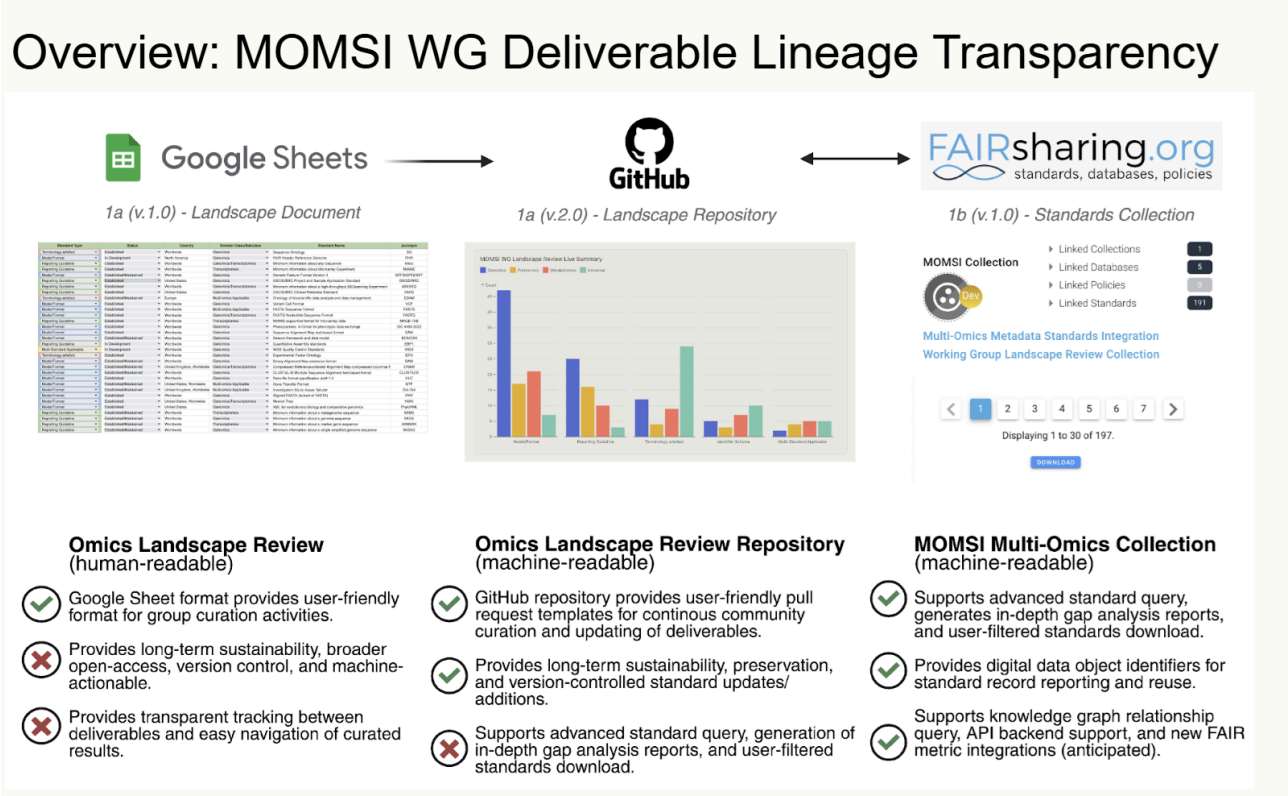
Qs for Aastha:

- Who decides which (meta)data models to use for the joint services? Are you already aligned where the partners have services for the same modalities?
- What obstacles do you expect concerning semantic data interoperability within and between EOSC nodes? Will EOSC favor a few centralized repositories to reduce complexities of Interop?
  - EOSC wants to preserve richness of existing repositories and resources - aim for linking across these. EOSC is not just for life science - the discipline nodes are aiming for a more meaningful inroad to the services though for people what are associated with that discipline
- are there examples of practice to integrate multimodal data in other disciplines in EOSC (eg physics etc) that we in life science might learn from?

GENERAL DISCUSSION:

Tony - BioFAIR is thinking about data journeys as opposed to only data repositories (which are specialist, siloed) - a set of transformations from assay to insight, with bridges to describe transformations (using PIDs etc) in a robust way.

MOMSI WG



- Omics Standard Landscape Review Curation Workflow & Interactive Web-based Dashboard Tool (10.15497/RDA00133)
- https://rda-momsi.github.io/Dashboard/
-

INTERNATIONAL DATA WEEK
13-16 October 2025
Brisbane, Australia

RDA
RESEARCH DATA ALLIANCE