Exploring the FAIR requirements for federated infrastructures in the life sciences and beyond

Plenary session link:
https://www.rd-alliance.org/plenaries/exploring-fair-requirements-federated/
Group(s) name(s) organising the session: Life Science Data Infrastructures IG

Session summary (for Group co-chairs)

Please complete the table below by **7th June, close of business**. The information provided will be used to include your outcomes to the RDA community as a report organised by the Technical Advisory Board.

Summarise the session's key points and discussions in three sentences:

The session highlighted various approaches to enable seamless data discovery and analysis across regionally dispersed and heterogeneous resources. Presenters shared insights from building on technical standards such as those developed by GA4GH, broadly used platforms such as Galaxy and frameworks such as 1+MG Framework to enable FAIR and secure data management in compute intensive domains, such as genomics. Discussions emphasised future opportunities to involve a wider range of infrastructures and efforts to align general and discipline specific infrastructure solutions.

Key outcomes/actions/takeaways

1. *Showcased approaches to federating infrastructures using technologies from the global collaborations Galaxy, Gen3, and GA4GH with examples of implementations across Africa, Australia, and Europe.*
2. *Demonstrated that the FAIR principles can be interpreted and applied to a wide range of aspects related to federated infrastructure and research environments beyond data, such as tools, compute resources, data provisioning, workflow execution, resource entitlements etc.*
3. *Identified other RDA groups that IG members can engage with for broader or deeper discussions on the topics covered but also to develop potential topics for future joint sessions proposals.*
4. *There is a wide range of life science infrastructures that are unaware of the RDA or this IG today—can and should the IG do something to address this in the short term?*

Synergies and/or possible collaborations identified with RDA groups and other groups:

The following RDA groups were specifically mentioned as potential venues for continued discussions and drivers of joint activities/sessions on the topics covered:

- Sensitive Data IG
- Working with PIDs in tools IG
- Mapping the Landscape of Digital Research Tools WG

- Global Open Research Commons WG
- GORC Health Data Commons WG – RDA P22 BoF currently drafting its charter
- TRESD WG – proposed group focused on trusted research environments

Other groups were also mentioned, such as Canada's Alliance Cloud Connect, FDO forum, RO-Crate but there are many more existing and potential connections that could be created/strengthened.

Get involved in ***RDA Community***
Check out ***VP22 programme sessions***

This meeting will take place according to the **RDA Code of Conduct**

FIRST SESSION (15 May) 22:30 - 00:00 UTC

Attendee Check-in for the first session

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

Name	Affiliation	Location	Email/social media
Jeff Christiansen	Australian BioCommons	Brisbane, Australia	jeff@biocomons.org.au
Wolmar Nyberg Åkerström	ELIXIR / NBIS, SciLifeLab	Uppsala, Sweden	wolmar.n.akerstrom@uu.se
Reyna Jenkyns	World Data System	Victoria, BC, Canada	reyna@oceannetworks.ca
Gareth Price	Galaxy Australia, QCIF	Brisbane, Australia	g.price@qcif.edu.au
Andrea Budac	World Data System	Edmonton, AB, Canada	abudac@oceannetworks.ca
Romain David	ERINHA AISBL	BE / FR	romain.david@erinha.eu
Oliver Hofmann	University of Melbourne	Melbourne, Australia	oliver.hofmann@unimelb.edu.au
Anu Gururaj	NIAID	Rockville, MD	anupama.gururaj@nih.gov
Inès Autuoro	University of Strasbourg	France	iautuoro@unistra.fr
Christian Berger	Philipps-Universität Marburg	Marburg, Germany	berger@hrz.uni-marburg.de
CJ Woodford	Digital Research Alliance of Canada	Ontario, Canada	c.joseph.woodford@gmail.com
Michelle Barker	Research Software Alliance	Australia	michelle@researchsoft.org

Christine Laney	US National Ecological Observatory Network (NEON)	USA	claney@battelleecology.org
Aline Grand	University of Strasbourg	Strasbourg, France	alinegrand@unistra.fr
Bernie Pope	Australian BioCommons	Melbourne, Australia	bernie@biocommons.org.au
Joshua Harris	Australian BioCommons	Melbourne, Australia	joshua@biocommons.org.au
David Medyckyj-Scott	Manaaki Whenua	New Zealand	medyckyj-scottd@landcareresearch.co.nz

Session Slides: LifeSci Data Inf IG VP22 Session Slides

Recording: Exploring the FAIR requirements for federated infrastructures in the…

Agenda *for the first Session*

Link to notes (this document):
https://docs.google.com/document/d/1_bTRIoAbxr0XDXC8STEw2Y3m-qhfXYMiDpsvY4Z6KWg/edit

00:00 - 00:20 (20 min)	**Welcome, icebreaker and introduction: Why this session?**	Jeff Christiansen Wolmar Nyberg Åkerström Mentimeter: menti.com/als95qn9dfr6 See also: Results view
00:20 - 00:60 (40 min)	**Federated infrastructures: State of the art in the biomolecular life sciences?**	Initiatives connecting and contributing to Global Alliance for Genomics and Health (GA4GH) standards, frameworks, and tools **Oliver Hofmann (Melbourne, Australia) and Bernie Pope (Melbourne, Australia) -** Exemplar projects that are using existing and upcoming GA4GH standards in the data infrastructure world - globally and in Australia (20 min)
		FAIR data analysis for the life science and beyond using the Galaxy platform **Gareth Price (Brisbane, Australia) -** Towards a global harmonised biodata analysis infrastructure - the use galaxy effort (20min)

00:60 - 01:10 (10 min)	**Flashtalks: Connecting FAIR data resources and services - What are promising approaches?**	Wolmar Nyberg Åkerström on behalf of **Melissa Konopko (Cambridge UK)** - European '1+ Million Genomes' Initiative and European Genomic Data Infrastructure (GDI) (5min) 1+MG Framework: The Global Reference Point to Connect to the European Genomic Data Infrastructure ● You can use it ● You can contribute ● GDI 18 months milestone ● GDI Starter Kit milestone
01:10 - 01:30 (20 min)	**Open discussion: Where do we go from here?**	All
01:30	**Session End**	

Collaborative Session Notes *from the first Session*

(To be used by participants and chairs during the session).

Welcome, icebreaker and introduction: Why this session?
Jeff: Welcome and introduction to the session.

- Focus on exploring FAIR requirements for federated infrastructures in life sciences and beyond.
- Aims to identify synergies and opportunities for further collaboration with other groups; and encourage communication, discussion, and coordination among individuals within and outside the RDA

Wolmar: Icebreaker poll results view:
https://www.mentimeter.com/app/presentation/alzomuowmc363xz3s9jdkthjzcntgzk1/embed

Federated infrastructures: State of the art in the biomolecular life sciences?
Bernie & Oliver: Exemplar projects that are using existing and upcoming GA4GH standards in the data infrastructure world - globally and in Australia
Bernie: Part 1

- Examples from genomic cohort data sharing, specifically the Data Commons for cardiovascular disease project.
- Motivation includes tackling atherosclerosis and coronary artery disease (CAD), leading causes of premature cardiovascular death in Australia, such as identifying new risk factors since many heart attack patients show no known risk factors.
- Assembling large, representative datasets of at-risk individuals and population-level data helps identify and validate biomarkers. Compiling various types of omics data (genomic, metabolomic, lipidomic).

- Process includes collecting samples, undergoing analyses, and aggregating large datasets. Machine learning and AI algorithms identify risk scores, feeding into clinical interfaces to improve outcomes.
- Collaboration began with CAD Frontiers to develop an aggregated data collection and analytics platform for biomarker identification, inspired by the NHLBI's BioData Catalyst platform, which uses Gen3 software for data exploration and analysis.
- Gen3 and similar systems across the NIH support diverse data commons, covering a range of data types and disease phenotypes, supporting multi-omic data, imaging data, and rich metadata across many samples.
- A data commons, as defined by Bob Grossman, is a platform and governance framework for data management, analysis, and sharing. Gen3 includes components like security, login landing, data exploration, a data dictionary, and data submission, all backed by cloud storage.
- The Australian Cardiovascular Disease Data Commons (ACDC) on the Gen3 platform currently hosts synthetic data and plans to include 18 participant cohorts, pending ethics approval, operating on Amazon's cloud service.
- The project emphasises data governance (legal, ethics, policy, trust) and aims to harmonise data from 18 historical and ongoing cohorts into a consistent data dictionary. Access management uses CILogon for institutional logins, with data access supported by REMS.

Oliver: Part 2

- Australian Genomics was established in 2016 with about 70 different entities to show the benefit of genomics for patients, evaluate cost-effectiveness in a clinical setting, and advise policymakers on what a national rollout should look like.
- Clinical flagship projects cover many different diseases and require a data model suitable for various, sometimes challenging, datasets.
- Clear mandate from about 90% of the participants in those studies that they wanted their datasets to be available for future research. A system was developed to capture this data and make it accessible.
- The data store uses S3 buckets on Amazon, split into staging and storage. Validation is done on the staging bucket based on a manifest provided by the clinical labs before moving it into storage. Data announcements are made on the website and through the community, but also discovered via publications.
- The data access control involves researchers submitting proposals, reviewed by a committee. In practice, it's more manual, involving emails and processes for staging, retrieving, and sending data, which is cumbersome and inefficient.
- To improve this, the Data Release Coordinator was developed with Australian Genomics and Australian BioCommons. It automates many processes, integrates with REMS, and facilitates dynamic consent management, creating a self-service system for researchers.
- Based on standards, many of them are GA4GH, including findability through the Beacon Network. User authentication through the Australian Access Federation, to use home institution accounts for identification.
- Envisioning having a Gen3 front end for more targeted views for some communities, currently an environment on AWS where researchers can explore the data as is
- Workflow Execution System (GA4GH) can bring workflows or code to the datasets but most researchers prefer to download the data and work in their established environment.
- Aim to have agreement on what a research environment looks like for different use cases and be able to move data on demand. If data is needed in a different trusted environment, move it over. UK Biobank and others have moved data between countries for joint analysis.

- National genomics infrastructure with a mandate to align with international data sharing initiatives, focusing on standards-based interoperability across systems, especially between different states with different setups.

Gareth: Towards a global harmonised biodata analysis infrastructure - the use galaxy effort

- Focusing on FAIR principles and infrastructure, including local and global job resourcing, tools, and workflow harmonisation for researchers.
- Galaxy is a web interface for uploading research data, launching tools, accessing reference data, as well as workflow creation and sharing.
- Galaxy's large public servers, called useGalaxy servers, update software and tools regularly. The Galaxy ToolShed hosts nearly 10,000 tools, with some hosted via CERN VM file systems
- Job distribution uses the Pulsar system to access infrastructure beyond the main node, allowing high CPU, memory, GPU, and rapid I/O VMs for optimal analysis. The system enables global data sharing and job resourcing.
- The Total Perspective Vortex (TPV) is an abstraction layer aiding job scheduling. Now used by all Galaxy servers, helping to manage HPC complexity, scale resources, and analyze job profiles to improve efficiency.
- The Intergalactic Utilities Commission (IUC) ensures tools are well-wrapped, curated, and validated. Automation processes for Docker and Singularity containers support the tools' development and global distribution.
- The Intergalactic Workflow Commission (IWC) produces high-quality workflows, regularly reviewed and tested, with detailed metadata. They are deposited at Dockstore and WorkflowHub for discoverability and immediate use in Galaxy instances.

Flashtalks: Connecting FAIR data resources and services - What are promising approaches?

Wolmar on behalf of Melissa: European '1+ Million Genomes' Initiative and European Genomic Data Infrastructure (GDI)

- Involves multiple countries at the policy level, with incentives from governments and European engagement to facilitate cross-border data movement and regulatory changes. Aim to make over a million genomes available for research and health applications.
- Several projects feed into this. The 1+ Million Genomes initiative involves policymakers and country representatives, followed by research and healthcare projects. The Beyond 1 Million Genomes project focused on specifications and design. The Genomic Data Infrastructure (GDI) project, which recently hit an 18-month milestone, continues with further projects planned, including one on population genomics.
- An overview of the framework provides a user-friendly entry point to the specifications, guidelines, and outputs from the Beyond 1 Million Genomes project, with updates during the GDI period.
- The framework page structure includes problem contexts and recommendations, categorised as required, recommended, or informational.
- The ambition is for the framework to be a global reference for genomic data infrastructures in Europe. It's available on GitHub, accepting contributions under specific conditions, encouraging engagement and future collaboration.
- 1+MG Framework: The Global Reference Point to Connect to the European Genomic Data Infrastructure
  - You can use it
  - You can contribute
  - GDI 18 months milestone
  - GDI Starter Kit milestone

Discussion:

- The session aims to reach other RDA groups on these topics.
- Consider joining the group or reaching out for joint activities.
- There is a large number of research infrastructures (RIs) within life sciences and inviting them on a broader scale of could be considered, being mindful of scope
- The group is open to everyone

SECOND SESSION (22 May 07:00 - 08:30 UTC)

Attendee Check-in for second session

1. **Indicate your attendance in the table below**
2. **Ice breaker poll** menti.com/alkwifo789e4

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

Name	Affiliation	Location	Email/social media
Jeff Christiansen	Australian BioCommons	Brisbane, Australia	jeff@biocommons.org.au
Wolmar Nyberg Åkerström	ELIXIR / NBIS, SciLifeLab	Uppsala, Sweden	wolmar.n.akerstrom@uu.se
Ryan O'Connor	RDA Europe	Cork, Ireland	ryan.oconnor@rda-foundation.org
Ville Tenhunen	EGI Foundation	Savonlinna, Finland	ville.tenhunen@egi.eu
Julia Gehrmann	University Hospital Cologne	Cologne, Germany	julia.gehrmann1@uk-koeln.de
Tilo Mathes	Research Space	Berlin, Germany	tilo.mathes@researchspace.com / LI
Miguel Ángel López González	IRTA (Institute of Agrifood Research and Technology)	Spain	miguelangel.lopez@irta.cat
Carme Reverté Reverté	IRTA (Institute of Agrifood Research and Technology)	Spain	carme.reverte@irta.cat
Federico Bianchini	University of Oslo / ELIXIR Norway	Oslo, Norway	fredebi@uio.no
Rory Macneil	Research Space	Edinburgh, Scotland	rmacneil@researchspace.com
Sveinung Gundersen	University of Oslo / ELIXIR Norway	Elverum, Norway	sveinugu@uio.no
Rob Hooft	Health-RI	Hoek van Holland, The	rob.hooft@health-ri.nl

		Netherlands	
Lars Eklund	SND/NBIS/Uppsala university	Sweden	lars.eklund@it.uu.se
Leo Chiloane	SAEON	South Africa	pl.chiloane@saeon.nrf.ac.za
Giulia Caldoni	University of Bologna	Italy	giulia.caldoni2@unibo.it
Andreas Czerniak	Bielefeld University	Germany	andreas.czerniak@uni-bielefeld.de
Takudzwa N Musarurwa	University of Cape Town	South Africa	tnmusarurwa@gmail.com
Nina Grau	INRAE, Montpellier	France	nina.grau@inrae.fr
Anne-Sophie Bage	INRAE, Rennes	France	anne-sophie.bage@inrae.fr
Björn Grüning	Uni-Freiburg	Germany	bjoern.gruening@gmail.com

Session Slides: LifeSci Data Inf IG VP22 Session Slides

Recording: Exploring the FAIR requirements for federated infrastructures in the…

Agenda *for the second Session*

00:00 - 00:10 (10 min)	**Welcome, icebreaker and introduction: Why this session?**	Jeff Christiansen Wolmar Nyberg Åkerström Ice breaker poll: menti.com/alkwifo789e4 See also: Results view
00:10 - 00:50 (40 min)	**Federated infrastructures: State of the art in the biomolecular life sciences?**	1. **Takudzwa (Taku) Nyasha Musarurwa (Cape Town, South Africa)** - Implementing GA4GH standards for federated data analysis in Africa (20 min)
		2. **Björn Grüning (Freiburg, Germany) [slides]** - FAIR data analysis for the life science and beyond using the Galaxy platform (20min)
00:50 - 01:10 (20 min)	Flashtalks: Connecting FAIR data resources and services - What are promising approaches?	**Rob Hooft (Rotterdam, Netherlands) -** European '1+ Million Genomes' Initiative and European Genomic Data Infrastructure (GDI) (5 min)

		Sveinung Gundersen (Oslo, Norway) - Empowering Users: Orchestrating Sensitive Data Access for Interactive Federated Analysis in Virtual Research Environments (5 min)
		Rory MacNeil - The New England Research Cloud NERC: a starting point for a series of federated research clouds around the US (5 min)
01:10-01:30 (20 min)	**Open discussion: Where do we go from here?**	**All** - Identifying synergies and opportunities for collaboration with other groups leading up to future RDA plenaries Exit poll: menti.com/al3npzbk99o8 See also: Results view
01:30	**Session End**	

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

Welcome, icebreaker and introduction: Why this session?
Jeff: Welcome and introduction to the session.

- Focus on exploring FAIR requirements for federated infrastructures in life sciences and beyond.
- Aims to identify synergies and opportunities for further collaboration with other groups; and encourage communication, discussion, and coordination among individuals within and outside the RDA

Wolmar: Icebreaker poll results view:
https://www.mentimeter.com/app/presentation/al21sidyv8kgtwgviq73o9mtusvdhh31/embed

Federated infrastructures: State of the art in the biomolecular life sciences?
Takudzwa: Implementing GA4GH standards for federated data analysis in Africa

- The eLwazi platform is designed as an open data science resource for researchers across Africa, ensuring accessible and findable data, with tools and workflows for analyses on various computing environments all centralised through a user-friendly interface.
- The platform's main components are infrastructure, data, and tools, leveraging Terra, Gen3 implementations, Data Biosphere concepts, and GA4GH standards to integrate diverse data types and access policies.
- Goals include developing an African data science platform to support health discovery, addressing challenges where researchers currently use different platforms independently, thus complicating data sharing and collaboration.

- Infrastructure details include a gateway and workspace for data access and tool selection, with data catalogues adhering to various policies; partners span the US and Africa, supporting diverse research hubs within the DS-I Africa initiative.
- A pilot platform demonstrates data sharing and compute capabilities across Africa using GA4GH standards; future plans involve refining the platform, integrating additional GA4GH standards, enhancing user experience, and providing flexible access to various computing services.
- Current advancements include replacing WESkit, adding features like a DRS resolver and passport tokens for secure access, implementing a reference panel imputation service asynchronously across locations, exploring Crypt4GH for security, and considering Gen3 and DNA Stack for APIs, with a demo link provided for feedback and an invitation for questions.

Björn: FAIR data analysis for the life science and beyond using the Galaxy

- Galaxy is an open-source platform started in 2005, used for accessing and sharing data, tools, workflows, and infrastructure with over 130 public instances and more than 3,000 tools available on major continental servers.
- Serves a broad user base with over 300,000 registered users across its European, American, and Australian servers, catering differently to scientists, service providers, and educators by offering varied functionalities and abstraction layers for distributed and federated storage.
- Galaxy includes a front end for workflow creation and a robust backend with over 500 API endpoints, enabling fully automated tool execution, workflow management, and user administration programmatically, highlighting its flexibility for large-scale user and data management.
- The platform uses metadata and annotation standards like schema.org and SPDX, supports TRS, DRS, and TES GA4GH standards for data exchange, and configures to utilise various compute and storage resources, making it adaptable to diverse data import, export, and authentication plugins.
- As an agnostic framework, Galaxy allows different communities to customise the platform with specific tools and visualisations, demonstrated by European collaborations with ELIXIR to create Galaxy flavours for various scientific domains including life sciences, climate science, and astrophysics.
- Galaxy addresses the entire research data management lifecycle by integrating with data providers and supporting GA4GH standards, offering tools for data transformation, and facilitating the export and preservation of research artefacts, emphasising the importance of data sharing and reusability.
- Galaxy enhances storage management with solutions for federated infrastructures, user-based object storage, and export options to permanent archives, supported by a storage dashboard and notification system, aiming to efficiently manage and scale storage needs while ensuring data transparency and reproducibility.

Flashtalks: Connecting FAIR data resources and services - What are promising approaches?

Rob Hooft: European '1+ Million Genomes' Initiative and European Genomic Data Infrastructure (GDI)

- European infrastructure project for sharing human genetic data, aiming to collect and utilise over a million European genomes for research, healthcare, and health policy, inspired by efforts starting in 2016 and gaining momentum with signatures from 27 countries.
- European governments meet regularly to discuss realisation, leading to the B1MG project and the ongoing European Genomic Data Infrastructure (GDI) project launched at the end

of 2022, moving from design and testing to implementation and sustainability.

- Varied progress across countries, with examples like Denmark's established national genome centre and the Netherlands' developing health data infrastructure, aims to unify efforts under European rules and regulations to facilitate data sharing and infrastructural cohesion.
- European privacy laws (GDPR) intended to enable data sharing within the EU, support initiatives like an anticipated Population Genomics Project aimed at gathering representative genomic data across the European population, not just hospital patients.
- Various European projects and initiatives need to share information, captured in the 1+ Genomes framework documenting standards, explorations, and infrastructure-building efforts while adhering to laws and practical aspects such as data localization.
- Technical infrastructure defined by interfaces, allowing countries to use different tools while interoperating through minimal specifications and standards like TES and WES, with flexibility for countries to choose between commercial or on-premise solutions.
- Central catalogue concept classifying datasets and using DCAT ontology to describe and locate data, centralising requests while specifying all interfaces for data use, documented as a global reference point to enhance interoperability and reusability, aiming for adoption beyond Europe.

Sveinung: Empowering Users: Orchestrating Sensitive Data Access for Interactive Federated Analysis in Virtual Research Environments

- Issues in sensitive data analysis include aligning with policies and regulations, establishing a chain of trust, delegating data access efficiently, maintaining data within geographic boundaries, ensuring data security, orchestrating seamless distributed data access, and providing an intuitive user-controlled analysis experience.
- The implementation of GA4GH standards addresses many of these issues: AAI and Passports for trust, Passports and DUO for delegation, various standards for geolocation, Crypt4GH for data security, TES and WES for orchestration, and focuses on user experience.
- Crypt4GH is a versatile solution supporting multiple aspects of data security and access, enabling secure data management “in situ, in statu, and in motu”
- Crypt4GH works through header re-encryption ("recryption") using a public-private key pair mechanism allowing secure user access without dataset alteration, demonstrated within the Federated EGA for sensitive genomic data.
- The Strengthened Data Management in Galaxy study employed Crypt4GH for encrypted analysis and data security in a prototype without sharing private keys.
- Future plans for integration into Galaxy as a production implementation, prioritizing user control and experience, eliminating central storage by allowing direct data provisioning into trusted compute nodes, with appropriate key management and sharing capabilities.

Rory: The New England Research Cloud NERC: a starting point for a series of federated research clouds around the US

- The presentation introduces a different perspective, contrasting with the life sciences-specific infrastructures discussed earlier, by focusing on two emerging forms of generalist research infrastructure: Research Commons and Research Clouds, highlighting the need for effective interfacing with domain-specific infrastructure.
- The discussion revolves around interoperability in research infrastructures, citing the Global Open Research Commons (GORC) model as an inspiration for emerging infrastructure. It emphasises two types of interoperability: interaction between storage/compute and research tools, and interoperability between research tools.
- Three examples illustrate the concept: the proposed Norwegian Research Commons,

which follows the GORC model in detail, focusing on tool interoperability throughout the research life cycle to ensure smooth data passage and consistent metadata association.

- Interoperability challenges include technical integration issues and human or organisational willingness to collaborate, using iRODS as a potential mechanism. The Norwegian commons aim to enhance tool interoperability but face uncertainties in collaboration between storage providers and tool developers.
- The New England Research Cloud, a bottom-up initiative by five major New England universities, exemplifies a computing and HPC-based approach, suggesting a set of generalist research services atop cloud infrastructure, indicating a blend of data management and compute perspectives.
- Canada's Alliance Cloud Connect, another example of a compute and store infrastructure, calls for research-facing services, already focusing on tools like Jupyter Notebooks and Galaxy, and expressing the need to incorporate more generalist tools such as Dataverse and DSpace.
- An open question of interoperability between generalist infrastructures and domain-specific resources like Galaxy is emphasised, noting the importance of effective collaboration for the future, despite challenges, particularly around storage and other integration aspects, presenting both challenges and opportunities.

Discussion:

- Continuing collaborations and discussions within and outside the RDA.
- Exit poll for ideas on groups to reach out to and possible actions:
  - Exit poll: menti.com/al3npzbk99o8
  - See also: Results view
- Two relevant RDA groups: GORC and Mapping the Landscape of Digital Research Tools. The latter has produced a map of research tools and seeks engagement with the life sciences community, which would benefit from interaction with this group.
- Add suggestions to collaborative notes or use the Menti poll.
- Will organize a post-plenary discussion inviting those who have added their email addresses to the notes.

From the Zoom chat: “Jeff Christiansen (22. May 2024, 10:01)
Re. the handover points between different infrastructures, it's previously been identified that integrating **sidecar files (like RO-Crates)**, with the use of standard APIs would better enable exchange from/to other infrastructures.”

What are some synergies and opportunities for collaborations that we should look into?

Which groups should we contact for future work and collaborations? (within and beyond the RDA)

- GORC Global
- FDO forum
- Ofr RDA Mapping the Landscape of Digital Tools WG
- Canada’s Alliance Cloud Connect

Give one or a few examples of your key takeaway messages and how you will follow up on this topic?

Rob: Very nice to see all the different tools people are using, I do notice that some of the tools that are proposed are not using semantic interoperability. E.g. the metadata that iRods maintains is not semantic. I think the components of an infrastructure that needs interdisciplinary or inter-national interoperability should develop semantics.

Ryan O'Connor: There is a spin-out WG from the GORC International Model WG in planning - the GORC Health Data Commons WG is in its set-up phase at present (i.e., drafting its Case Statement) and would have some connections with the initiatives described today. (disclosure: The WG is being supported by RDA TIGER project, which I am working on)

THIRD (FOLLOW-UP) SESSION (27 May)

Due to a lack of time in either session, a 1 hour follow-up session was held post-Plenary on 27th May to identify synergies for collaboration moving forward

Invitation *distributed to those who added their email in this documents*

> Dear participant,
>
> We would like to follow up on the two sessions with a discussion focusing on synergies and opportunities for collaboration with other groups leading up to future RDA plenaries and propose a meeting for those who are available next week on Monday (27th May) at 8:00 UTC.
>
> Register below to receive a calendar invitation or to get direct access to the meeting (60 min before the meeting starts):
> https://uu-se.zoom.us/meeting/register/u5Erd-qurDMtHNVIuvi_LGBzN-KIHozT49yI
>
> Poll on opportunities for future work and collaborations:
> https://www.menti.com/al3npzbk99o8
>
> See also: Results view
>
> Kind regards,
> Wolmar
>
> --
> Wolmar Nyberg Åkerström
> Co-chair of RDA Life Science Data Infrastructures IG
>
> NBIS - National Bioinformatics Infrastructure Sweden
> Uppsala University
> www.nbis.se, www.uu.se

Collaborative meeting notes *from the post-plenary meeting*

2024-05-27 Discussion - Exploring the FAIR requirements for federated infrastructures in the lif…