Bridging policy and practice: advancing data visitation in health-related sciences

24th Plenary Meeting Collaborative Notes

Group(s) name(s) organising the session: BoF - Francis P. Crawley
Session link:
https://www.rd-alliance.org/members/francis-p-crawley/plenary-participation/?application_id=174660
Session scheduled date/time/breakout session: *08.04.2025, Breakout 3, 07:00 UTC*

Session summary (for Group co-chairs)

We will use the content in the table below to highlight your work to the RDA community as a report organised by the Technical Advisory Board.
Please complete ALL fields below by **Friday 25th April, close of business** to be included in the report & social media activities.

Summarise your session in three sentences:
The session explored how the data visitation model facilitates secure and ethical access to sensitive health data. This session emphasized the critical role of data visitation using the EOSC-Future/RDA Artificial Intelligence & Data Visitation (AIDV) Working Group outputs and the pilot testing of the FAIRlyz platform to ensure safe quality control (QC) during the AI-driven analysis of biomedical datasets. It highlighted the need for robust policy frameworks and secure platforms, helping to prepare the EOSC-Future/RDA AIDV-WG and AI4RDA RDA Plenary 25 BoF session proposal on Advancing secure and trusted data visitation in the RDA as well as the Case Statement for an RDA/CODATA Data Visitation for Health-related Sciences Working Group (DVHrS-WG).
Key outcomes/actions/takeaways
1. ***Convergence of Non-Data-Movement Approaches:*** *A significant outcome of the session is the clear recognition of parallel and convergent efforts across multiple initiatives aimed at enabling data analysis without necessitating physical data transfer. This shared objective is driven by the increasing sensitivity and sheer volume of health-related data, making traditional data sharing models less feasible and desirable. The session highlighted a growing community of practice exploring alternative methodologies.* 2. ***Identification of Key Implementing Groups:*** *Several prominent groups are actively developing and implementing methods aligned with the principles of data visitation.* 3. ***Action Item:*** *Foster Inclusivity and Collaboration: A primary action stemming from this session is the imperative to actively include the identified groups (FAIRlyz, ELIXIR/DOME, AHA!/ARDC, TRE Community, EOSC-ENTRUST) in future discussions, collaborations, and the development of best practices for data visitation in health-related sciences. This will prevent duplication of effort, encourage the*

sharing of knowledge and technical solutions, and accelerate the advancement of the field

4. ***Takeaway: Growing Momentum and Shared Vision:*** *The session underscored the significant and growing momentum behind approaches that prioritize data locality for analysis. The convergence of efforts across diverse international initiatives suggests a shared vision for a future where sensitive and large-scale health-related data can be effectively studied through secure and privacy-preserving data visitation methodologies. This collaborative spirit is crucial for overcoming the technical and policy hurdles in this domain*

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

These should be considered key stakeholders and potential collaborators in future advancements:

1. ***Lifetime Omics and the FAIRlyz Registry:*** *Lifetime Omics is actively developing the FAIRlyz platform focused on facilitating data visitation for scientific research and quality control of datasets, with specific applications in health-related sciences. RDA AIDV WG is guiding the ongoing work and infrastructure development for FAIRlyz' data visitation.*
2. ***The ELIXIR Organisation and their DOME Registry:*** *ELIXIR, a distributed infrastructure for life science information, is actively involved in non-data-movement approaches through initiatives like the DOME (Data-driven Outcome Modeling Ecosystem) registry. The DOME registry, a curated database of annotations for supervised machine learning models in biology (including both published and unpublished studies), facilitates analysis across datasets and is considering federated analysis vs centralizing the raw data before analysis.*
3. ***The AHA! Australian Harmonized Approach to Federated Learning for Research Sector (ARDC):*** *The ARDC's AHA! initiative specifically focuses on federated learning, a key technique for distributed analysis that avoids direct data sharing. Their work in harmonizing this approach within the Australian research sector provides valuable insights and potentially transferable frameworks for the broader health-related sciences community.*
4. ***The Trusted Research Environment (TRE) Community and the EOSC-ENTRUST Project:*** *The Trusted Research Environment (TRE) community, along with the EOSC-ENTRUST project within the European Open Science Cloud (EOSC), are actively engaged in establishing secure, controlled environments where researchers can access and compute on sensitive datasets co-located with the data source. Their focus on secure compute environments aligns directly with the goals of data visitation by bringing the analysis to the data, rather than the other way around.*

Get involved in **RDA Community**
This meeting will take place according to the **RDA Code of Conduct**

Attendee Check-in

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

Full Name	Affiliation	Location	Email
Paty Buendia	Lifetime Omics	Stuart, USA	paty@lifetimeomics.com
Francis P. Crawley	EOSC-Future/RDA AIDV-WG	Leuven, Belgium	fpc@gcpalliance.org
Gavin Michael Farrell	University of Padova	Padova, Italy	gavinmichael.farrell@studenti.unipd.it
Gnana Bharathy	Australian Research Data Commons (ARDC) and University of Technology Sydney	Sydney, Australia	gnana.bharathy@ardc.edu.au
Lars Eklund	Uppsala University/NBIS	Uppsala, Sweden	lars.eklund@it.uu.se
Aleksandra Lazić	University of Belgrade	Belgrade, Serbia	lazic.ale@gmail.com (@AleLazic)
Milica Tanasic	Clarivate	Belgrade, Serbia	milica.tanasic@clarivate.com
Fabien BORGET	Aix-Marseille University	Marseille, France	Fabien.Borget@univ-amu.fr
Nathanael Sheehan	University of Exeter	Munich, Germany/Exeter, UK	ns651@exeter.ac.uk
Hugh Shanahan	Royal Holloway, University of London	Egham, UK	hugh.shanahan@rhul.ac.uk
Su-Nee GOH	Nanyang Technological University	Singapore	sunee@ntu.edu.sg
Guillaume MANSARD	Université de Poitiers	Poitiers, France	guillaume.mansard@univ-poitiers.fr
Gabriel KARERANGABO	Sciences Po	Paris,France	gabriel.karerangabo@sciencespo.fr
Ronit Purian	Climatact and IVC	Tel Aviv	purian.ronit@gmail.com

Mireille Louys	Observatoire astronomique de Strasbourg , CDS	Strasbourg, F	mireille.louys@unistra.fr

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

It would be a very good idea to make sure that individuals from the Trusted Research Environment (TRE) community are involved here. They are working on something similar here, namely having access to compute which is co-located with data sets of interest. The idea is that data cannot be moved out of the TRE. RDA Norway posted something on this here about EOSC-ENTRUST: A European Network of TRUSTed research environments who I think would be a good point of contact.

This is the link to the security template: https://docs.google.com/document/d/1q-OjK8Z6_qsob4J2vjUtRS6hD8MjmA7oZZ2NbEtKaI4/edit?usp=sharing Access is limited