23rd Plenary Meeting Collaborative Notes

Group(s) name(s) organising the session: *Health Data Commons GORC Profile WG*
Session link: *Health Data Commons GORC profile: Methodology, Means, and Moving forward*
Session scheduled date/time/breakout session: *Tuesday, 12 November, Breakout 1, 10:00 – 11:30 CST*
Venue: *Plaza de La Autonomía, Aula Magna*

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 and to a wider audience through English & Spanish social media mentions.
Please complete ALL fields below by **Friday, 29th November, close of business** to be included in the report & social media activities.

Summarise the session in three sentences:
In this session, the HDC GORC WG presented it's main objectives, timelines, and workplan which focus on a gap analysis (literature review and questionnaire), hearing from the health data community (speaker series, community of practise), and exposing metadata on choices made by HDCs on implementing their infrastructures, specifically regarding FAIR, CARE, and TRUST, via a GORC model profiles. The draft questionnaire was discussed, with identification of differences in service needs between the GORC model and HDCs already identified. Presentations on the EU Health Data Space and complimentary work done for HDCs in the US were presented and discussed.
Key outcomes/actions/takeaways
1. *The questionnaire is a good starting tool to gather data, but should not be the only tool used to gather information - holding space for conversations will be paramount.* 2. *Explicit consideration for federated organizations should be included in any tools we use and communications.* 3. *The questionnaire needs to be further refined to ensure adequate guidance to HDC participants and that the options presented reflect what we already know of the health data community.*
Synergies and/or possible collaborations identified with RDA groups and other groups:
GORC IG, WG RDARI WG
Highlight text that will be used in social media mentions (please make sure the text is clear and appropriate for public consumption & comprehension)

Direct link to Group home page
https://www.rd-alliance.org/groups/health-data-commons-gorc-profile-wg/

*Get involved in **RDA Community***
*Check out **P23 programme sessions***

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
Matt Chandler	Princeton University	USA	mchandler@princeton.edu
Susanna Sansone	University of Oxford; ELIXIR	UK	susanna-assunta.sansone@oerc.ox.ac.uk
Athina Papadopoulou	RDA Europe	Greece	athina.papadopoulou@rda-foundation.org
Anne-Lotte Masson	Erasmus Medical Center	The Netherlands	a.masson@erasmusmc.nl
Hilary Hanahoe	Research Data Alliance	United Kingdom	Hilary.hanahoe@rda-foundation.org
Michael Witt	Purdue University, re3data.org	USA	mwitt@purdue.edu
Victoria Yorke-Edwards	University College London	United Kingdom	v.yorke-edwards@ucl.ac.uk
Wolmar Nyberg Åkerström	NBIS / ELIXIR Sweden	Sweden	wolmar.n.akerstrom@uu.se
Rory Macneil happy to contribute generally and in particular on interoperability in particular	Research Space	Scotland	rmacneil@researchspace.com
Mijke Jetten	Health-RI	The Netherlands	mijke.jetten@health-ri.nl

Alyssa Foote	World Data System	Canada	afoote@oceannetworks.ca
Anu Gururaj	NIH	USA	anupama.gururaj@nih.gov
Lauren Maxwell	Universitatsklinikum Heidelberg	Germany	lauren.maxwell@uni-heidelberg.de
Natalie Harrower	Canadian Research Data Centre Network (CRDCN)	Canada	natalie.harrower@crdcn.ca
Lars Eklund	NBIS/SND/UPPMAX Uppsala University	Sweden	lars.eklund@it.uu.se
Robert Grossman	University of Chicago	USA	rgrossman1@uchicago.edu

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

Session slides (with links!) HDCGORCWG_P23 Presentation.pptx
(Sciensano) slides on Health DCAT-AP for Health Data Commons
Recording: https://www.youtube.com/watch?v=uUmjKCmLE3Q

Questionnaire:
https://docs.google.com/document/d/1bpVKmJbGLE68w33O1sWHiy3RRcg6C17aV1eZtB9bZm0/edit?tab=t.0

Mijke Jetten - I would like to participate to bring in The Netherlands (Health-RI as a HDC) perspective, but I have only very limited time available

— Mijke we are happy to have you I any time you have will be greatly appreciated /Lars

Wolmar Nyberg Åkerström – happy to bridge to the EOSC Association's Health Data Task Force.

— Thank you Wolmar that would be great /Lars

Notes taken from recording, after session:

- This is the first WG meeting, the case statement has been submitted after council review and we're awaiting formal endorsement.
- Menti:
- In this WG, not looking at how HDCs should address the FAIR, CARE, or TRUST principles - not interested in promoting or prescribing best practise. We want to identify what HDCs are doing to address these in their own context an making that metadata more available and described (exposed) so that we can think about interoperability across

them. Start with the GORC model and seeing what considerations are applicable to HDCs, what needs to be added, etc .

- GORC is a future vision of having research commons interoperable on a global scale such that researchers have access to the research objects and services they need to conduct research seamlessly, with the appropriate protocols and policies in place. Doesn't necessarily mean open, but that systems can work together for authorized parties and objects. GORC IG is creating a roadmap (how to get to GORC), GORC IM WG created a model with a set of considerations for what research commons can look to 1. Have a common language to describe commons components and 2. Work towards global interoperability of their infrastructure. Model says "what do you need to consider"., roadmap says "how do you do it".
    - GORC IG typology and definitions: what is a commons, essential elements of a commons
    - GORC IM V1.1 endorsed as a recommendation and available, built on the essential elements identified by the IG. ~ 400 considerations, so lots there!
- This WG will look at how do we use the model, specifically for HDCs.
- Current workplan and timeline:
    - Systematic review of literature and landscape for HDCs
    - Questionnaire to send to HDCs to gather information to put into profiles, build on the model, and determine who might be interested in being a speaker and/or being a part of a CoP
    - Both approaches will result of HDC GORC profile
    - Doing a gap analysis (review+questionnaire), hearing from HDCs directly (speaker series, CoP, case studies [application of the interim profiles]), resulting in the profile itself.
    - Want it to be easy to use and understand, profile (metadata) should be available in multiple formats and containers including a knowledge graph
    - Review and questionnaire are first, apply to GORC model revisement iteratively, then last part is case studies and formalization of the profile
- Questionnaire is drafted, let's discuss!
    - Essential identification questions of course
    - Can a HDC only provide one service, or do they always provide more than one?
        - Lars: I know of some in Sweden that are focused on 1 service, so it would be good to provide the choice instead of assuming.
    - Services that HDCs provide, options are high level items in the GORC model. Are we missing any?
        - Training on data is technically included in "services for direct research management", but could be brought out/made more prevalent. How to make data FAIR, etc.

- How the questionnaire could work for distributed systems, where there's multiple organizations that work together and together could be considered a HDC?
    - Federated systems for sure we want to capture. So for an organization like that, you would be captured as "providing access to services" since your organization specifically might not be providing the service itself with your own staff or infrastructure. This also captures folks that have 3rd party service providers too, really about what your researchers see when they log into your system.
- Is a questionnaire the best way to capture this information, it might be better to have it through conversation? Could take the questions in the questionnaire as a topic list. Also, how to capture folks who are in between offering services, maybe don't have things right now but will have them in the near future (e.g. in the mission and vision)?
    - There are questions about mission and vision, etc., but more importantly is that we'd like to use the questionnaire as a first step to determine who's interested in participating in those conversations (speaker series and/or CoP) - question at the bottom to opt in.
    - That should be higher in the questionnaire, maybe one of the first questions!
- Guidance should be given on how to answer the question "describe your HDC in 1-2 sentences".
    - Did want to capture how HDCs think of themselves in this open question, but we can definitely revisit and provide more guidance.
- Take a look at it and leave feedback and ideas!

- Pascal DERYCKE, Data discovery in the European Health Data Space: HealthDCAT-AP for findable health data
    - Policy context: In EU we have a new health data regulation to enable connection of data - EHDS. Want to create a federated infrastructure with interconnected data catalogs.
    - Data in scope: not one specific domain, everything that has an impact on health. Trying to create and provide metadata on all types of data.
    - Focusing here on data discovery, but the EHDS has many other components.
    - Why HealthDCAT-AP? Core metadata vocabulary for all HDCs, finetuning over the last several years. Not better than another one, but it has wide use in EU.
    - Looked at policies in place and what enables the use of data.
    - When you extend DCAT to a new domain, you try and find what you can use that is already there before introducing new fields. Also created new controlled vocabularies. Discussing with communities on best use.
    - Data have different access rights, so will give different fields based on access rights: open, protected, sensitive (personal level)

- Bob Grossman, Introduction to some interoperability frameworks for HDCs
    - There's lot of HDCs in US and being able to interoperate between them, so wanted to expose to the metadata between them. Core issue of this WG.
    - NIH centres and communities, looking at SAFE environments (papers published, Secure and authorized FAIR Environments)
    - Contain health data or could be used to analyze it.
    - Trusting users has known protocols, wanted to introduce a principle for platforms and environments. Allows platforms to be marked in the right way for health data use and storage.
    - Includes an authorized platform IDs.
    - Looking also at meshes, where commons are working together and have interoperable components. Paper on 10 pillars of data meshes available on arxiv
    - Importantly, data remains in the original environment until it's moved to an analysis environment
    - Different kinds of meshes depending on how open they are and what principles they're emphasizing.
    - Pillars:
        - Data should be identified with PIDs
        - The repo should have FAIR APIs so that the metadata associated with a PID can be accessed
        - The repo should have FAIR APIs so that the data associated with a PID can be accessed
        - There should be an API for authentication and authorization
        - A data platform ina data mesh should provide an API so that authorized analysis environments can access data on behalf of authenticated usyers who are authorized to access data in the platform
        - A data mesh should specify the shared governance model covering data governance, platform governance, and sustainability.
        - A data mesh should agree on the types of data objects supported, and the minimum metadata required for each type of data object
        - The data mesh metadata service (DMMS) should assign a PID to each dataset, study, or other entity in the mesh
        - Return of results, each data hub in the data mesh should return usage stats, usage information, and other relevant results
        - The data mesh operator should provide a public API interface to the DMMS for the data mesh metadata
- Questions:
    - There's a proliferation of commons and meshes, so want to highlight the RDARI session today that will talk about these infrastructures for institutions as well and how they fit into all this.
    - Should also think about reuse of data and provenance should have it's own PID

        - Bob: Issue when you work with derived data for sure. Addressed with other frameworks, so we didn't add it as a pillar here.
- Call to participate!