**An online version of this document, *open for public comments*, can be found here:**

**https://docs.google.com/document/d/15mxia8Y_wsIlp1a_IDhseYz8ACrvaodGdw4L-VID9ww/edit?usp=sharing**

**For convenience, the following on this page are the RDA guidelines for writing a case statement for review and submission. The document below has been divided into the corresponding sections listed on this page and are marked in red text.**

**A Case Statement describes:**

1. What is the research case (will the WG produce something useful)?
2. What is the business case (will people use it)?
3. Is there capacity (are the right people involved to adopt and implement).

**Case Statement Content**

A Case Statement must contain the following components:

1. **WG Charter:** A concise articulation of what issues the WG will address within a 12-18 month time frame and what its "deliverables" or outcomes will be.
2. **Value Proposition:** A specific description of who will benefit from the adoption or implementation of the WG outcomes and what tangible impacts should result.
3. **Engagement with existing work in the area**: A brief review of related work and plan for engagement with any other activities in the area.
4. **Work Plan:** A specific and detailed description of how the WG will operate including:
    - The form and description of final deliverables of the WG,
    - The form and description of milestones and intermediate documents, code or other deliverables that will be developed during the course of the WG's work,
    - A description of the WG's mode and frequency of operation (e.g. on-line and/or on-site, how frequently will the group meet, etc.),
    - A description of how the WG plans to develop consensus, address conflicts, stay on track and within scope, and move forward during operation, and
    - A description of the WG's planned approach to broader community engagement and participation.
5. **Adoption Plan:** A specific plan for adoption or implementation of the WG outcomes within the organizations and institutions represented by WG members, as well as plans for adoption more broadly within the community. Such adoption or implementation should start within the 12-18 month timeframe before the WG is complete.
6. **Initial Membership:** A specific list of initial members of the WG and a description of initial leadership of the WG.

# *Raising FAIRness in health data and health research performing organisations (HRPOs)* WG Case statement (draft)

1. **WG Charter** *A concise articulation of what issues the WG will address within an 18-month time frame and what its "deliverables" or outcomes (including a Recommendation) will be.*

Reuse of health and clinical research data (including social care, and hereafter referred to as health research data) has major restrictions when compared to other research data in the biomedical domain. This primarily pertains to the concerns imposed by privacy, sensitivity and ethical issues raised by making data freely available. Meanwhile, research data management (RDM) practices such as the creation of data management plans (DMP), sharing datasets, the deposition of data in repositories, and the application of FAIR data principles to research outcomes are becoming increasingly common as they are required by funder mandates. Besides this requirement placed by funders there is also the wider need for researchers to share, find and access data to progress common goals as the primary value, and promote integrity and reproducibility.

The last few years have seen a rapid rise in uptake of the FAIR principles, which originated in the life sciences domain, but which have now been adopted to varying degrees across all research domains. Concomitant with the rise of FAIR datasets has been an increase in open research which urges researchers to make their data available for reuse, especially those that are publicly funded. However, an important caveat when thinking about FAIR when compared to open research is the phrase **"as open as possible, as closed as necessary"**.

The recent enforcement of the GDPR in Europe is a prime example of a legal framework that makes strict regulations around the processing and sharing of personal data and places the onus on the data controller to make sure provisions are in place to ensure this. Although the GDPR is the most far reaching data protection legislation currently in the world, there are other territories that have restrictions on secondary use of personal data and health data, e.g. USA (HIPAA), Ireland (Health Research Regulation), India (Personal Data Protection Bill) and S-Africa (Protection of Personal Information Act). As well as internationally enforced restrictions, there are those at national and local levels, and together they all require evidence that the sharing and reuse of health research data are carried out responsibly and in-line with stated aims. The legislation is not meant to impose barriers but to protect individuals' rights.

FAIR adoption in the health research domain is complicated by numerous factors including concerns regarding: ethical, moral, cultural, technical, and legal constraints of primary source data. We therefore propose this WG to address some of these issues to:

- Analyse and report legal and ethical issues surrounding data privacy of health research data at the national level.
- Identify commonalities across territories that can be a foundation for harmonisation of guidelines on FAIR adoption.
- Provide Health Research Performing Organizations (HRPOs) with a set of clear and simple guidelines for implementing FAIR Open Data policy in health research.

Therefore, the main purpose of this WG is to provide HRPOs (such as universities, public research institutes, hospitals, medical charities etc.) with a set of clear and simple guidelines, which will define, establish and enable implementation of an aligned FAIR data policy at the institutional level. At this

time, there is a lack of clear and practical instructions for HRPOs to implement the FAIR principles locally. By developing clear guidelines, HRPO researchers will be better equipped to share and drive reuse of their datasets alongside non-health research, having previously obtained approvals from Ethics Committees, when necessary. This last point is particularly important since the variables are hitherto unknown for innovation in unexpected fields of research outwith health research.

The target audience for the proposed WG guidelines (beyond HRPOs themselves) are primarily researchers. In health research, researchers (and physicians) will be the responsible individuals tasked with implementing established policies of their data. Moreover, senior management will also be involved in shaping and driving these policies. In both cases, there is evidence that these individuals have little or no prior knowledge of FAIR and other technical issues, which is still the case in other research domains too[1]. However, anecdotal evidence suggests that these same individuals are the most willing to implement FAIR practices in their research institutions where possible and practicable. To achieve our stated objectives, the following tasks will be undertaken:

- Perform a landscape analysis to inform the guidelines.
- Invite specific stakeholders to help draft the guidelines, aiming for diversity in regional representation.
- Disseminate the draft guidelines to a wider community and agree a final version.
- Devise an outreach strategy so that the guidelines can be adopted by the intended community.

To underpin the drafting of the guidelines, the WG will draw upon a comprehensive analysis of current barriers (technical, ethical, security, legal, cultural, behavioural and economic), factors that reinforce them and potential overcoming mechanisms, which was and is currently investigated further by the EU H2020 funded FAIR4Health (F4H) project[2]. This WG will also build upon previous BoFs, held in RDA Plenary 13 and Plenary 14, initiated by F4H, about the necessity for creating useful and straightforward FAIR guidelines for HRPOs. Attendance at RDA 13 was the first time where this specific issue was discussed and the need for a dedicated WG was raised, and we were able to bring together a diverse group of interested individuals as well as establish contact with founders and members of the Health Research Data IG. The RDA 14 BoF built upon this.

Together, the proposed aims will be used to formulate an RDA approved output. The WG will take as its starting point 5 principles and 10 steps derived from a recently completed deliverable[3] in F4H, and further refine them through the course of its work to ensure they are fit for purpose for HRPOs to implement a FAIR data policy globally. They take into account all the legal, ethical, cultural and technical challenges that need to be overcome.

**2. Value Proposition** *A specific description of who will benefit from the adoption or implementation of the WG outcomes and what tangible impacts should result.*

The main beneficiaries of these results are those that are responsible and have the authority to establish policies in HRPOs globally, the researchers themselves, and societal outcomes. For the latter, adoption of these guidelines will aid transparency since health research can be opaque in many circumstances. This will require engagement of patients' groups, medical charities and public research institutions, as well as drawing from the growing field of citizen science. Moreover, there will be beneficiaries outwith the health research domain that will be able to access such data where

[1] Brock, J. "A love letter to your future self": What scientists need to know about FAIR data *Nature Index* **11 Feb 2019**
[2] https://www.fair4health.eu/
[3] https://osf.io/3u7dt/

allowable to foster innovation in, for example, technology. The results of the implementation of the FAIR data policy gathered in this WG will allow facilitation and encouragement of the HRPO community to share and reuse their datasets derived from publicly funded research initiatives.

The potential impact of the proposed FAIR data policy, in terms of health outcomes and health research, will be:

- adoption policy is expected to have a beneficial impact on science (sharing and reuse of health research results), the economy (optimisation of resources) and society (alignment of health outcomes with the values, needs and societal outcomes).
- increase visibility of international FAIR data policies and the applicable research data which has the potential to be reused.
- increased awareness by outreach to HRPOs to use FAIR and to be as open as possible through public engagement and deliberation for the health research community, data scientists, industry, and general public. This will be based on existing and ongoing work being conducted by the F4H project.

**3. Engagement with existing work in the area** *A brief review of related work and plan for engagement with any other activities in the area.*

Better definition of the FAIR principles to different research domains and resource types, in a practical sense and including applying metrics to FAIRness levels, is currently being examined by a number of international consortia. These will be leveraged to inform the guidelines that will be delivered by this WG. Some consortia and work that are known so far include:

- FAIR4Health
- GO FAIR
- EOSC (-Life) and the Turning FAIR into Reality WG related to EOSC-Life project
- FAIRplus
- FAIRsFAIR
- Horizon 2020 and Horizon Europe, provides legislative drivers such as the newly revised EU Directive on *Public Sector Information (PSI).*
- *Open Science agenda for Europe*, different communities, stakeholders and institutions have been producing their own guidelines, recommendations or other kinds of documents tackling RDM (including FAIR data and/or Open Data approaches).
- *Practical Guide to the International alignment of Research Data Management*[4] addressed to Research Funding Organizations (RFOs), launched by Science Europe in January 2019.
- Recommendations on managing research data addressed to Researchers[5], by *Maredata Spanish Research network of Open Research Data*[6].
- *Guidelines on FAIR Data Management in Horizon 2020*[7] by the EC addressed to EU funding applicants and beneficiaries to help them to make the data produced by their research projects Findable, Accessible, Interoperable and Reusable.

[4] https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf
[5] https://digital.csic.es/handle/10261/173801
[6] One of the Maredata (https://maredata.net) members (UC3M) is also a member of FAIR4Health.
[7] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

- European Data Protection Supervisor's recently published: *A Preliminary Opinion on data protection and scientific research*[8].
- National Institutes of Health (NIH) *Policy for Data Management and Sharing*, that is NIH data management and sharing activities related to public access and Open Science[9], launched in November 2019.
- USA - HIPAA
- Sage Bionetworks
- Atul Butte at UCSF
- South Africa -Protection of Personal Information Act
- ELIXIR has published guiding principles on FAIR Data Management in life science, 7 of the nodes are involved in developing a FAIR Service Architecture.
- NHMRC National Statement on Ethical Conduct in Human Research (Australia)
- NHMRC Management of Data and Information in Research (Australia)

Some of the above groups have already been approached by F4H and there is ongoing engagement to address concerns. These have included teleconferences and attendance at fora organised by respective groups. Further engagement with these and other initiatives will be to facilitate the inclusion and participation of HRPOs to ensure that research outcomes are shared, and FAIR data policy strategies are useful for the diverse needs of these institutions.

**4. Work Plan** *A specific and detailed description of how the WG will operate including:*

- *The form and description of the final Recommendation of the WG,*

The RDA Recommendation from the WG will be titled *Guidelines for implementing FAIR data within Health Research Producing Organisations.* The Recommendation will be provided in a non-proprietary, open file format. It will include the following components:

- Guidelines for HRPO senior managers on how to write a FAIR aligned data policy for their organisation, to enable the reuse of health research data.
- Guidelines for health researchers working within HRPOs on publicly funded research on how to make their data FAIR in a practical sense. This will be a longer document with links to supporting resources and references to best practice examples of FAIR health data.
- A checklist and/or rubric that can assist data creators within HRPOs to make their data FAIR. This will be a short and concise document that will allow researchers to quickly check their data against a specific FAIR checklist.
- Training documents aimed primarily at early career health researchers. This will include an introduction to the FAIR principles and how these can be implemented without risking privacy of human research participants.

In line with RDA guidance, the recommendation will also include a maintenance and retirement plan.

- *The form and description of milestones and intermediate documents, code or other deliverables that will be developed during the course of the WG's work,*

---

8 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

9 https://osp.od.nih.gov/scientific-sharing/nih-data-management-and-sharing-activities-related-to-public-access-and-open-science/

Milestones and intermediate documents:

The main method for formulating and disseminating the guidelines will be through written documents. Draft documents and other deliverables will be shared for review with the members of the WG through online channels such as email while the RDA website will provide an open forum for community review before finalisation.

The tasks for the WG include:

1. **Review** of the current landscape for FAIR adoption of health research data taking into consideration levels of privacy and ethical considerations granted to such data in regions around the world. This will be based on existing and recently completed work done by F4H in the EU context and the EOSC Landscaping Work currently being carried out. [M1-3]
2. **Assessing the impact** of such restrictions in terms of research, as well as financial and societal impacts, drawing on previously mentioned work by F4H, EOSC and others.[M2-6]
3. **Analysis** of national and international guidelines and policies, and their enforcement to compare and contrast and provide a foundation for the WG's own recommendations. [M3-9]
4. **Implementation** of draft guidelines at the 4-5 HRPOs to test feasibility. [M9-12]
5. Community surveys to gauge requirements and assess proposed deliverables of the WG. [M1-12]
6. Deliver **final output** as a written set of principles to be adopted by the HRPO community, with care to be taken that these should not be seen as rules that need to be strictly followed depending on the particular regional and local contexts. [M12]

| **GANTT Chart** | | **Month** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| **Tasks (numbered above)** | 1 | | | | | | | | | | | | |
| | 2 | | | | | | | | | | | | |
| | 3 | | | | | | | | | | | | |
| | 4 | | | | | | | | | | | | |
| | 5 | | | | | | | | | | | | |
| | 6 | | | | | | | | | | | | |

- *A description of the WG's mode and frequency of operation (e.g. on-line and/or on-site, how frequently will the group meet, etc.),*

WG's mode and frequency of operation:

The guidelines will be reviewed, and different versions will be generated, at the same time that the objectives and milestones of this WG are achieved.

1. Plenary 15 - Melbourne: Presentation of the objectives of the WG and establishing the initial approach to the principles and steps to be included in a global HRPO policy for FAIR data.
2. Plenary 16 - Costa Rica: Review in depth of the principles and steps to guide HRPOs to create a policy (internationally valid) for FAIR data.
3. Plenary 17 - Edinburgh: Approval of the guidelines and preparation of their proposal to HRPOs.
4. Monthly teleconferences (Skype/MS Teams/GTM/Zoom), F2F meetings at RDA P16 and P17 and other relevant conferences and workshops, mailing lists, open discussion through Twitter polls, or other simple mechanisms, will be used to develop and review the guidelines and

written intermediate documents, landscape analyses and survey results. Sharing of materials will be done through OSF and/or Google Drive.

5. Ad-hoc information sharing with relevant initiatives whenever possible (e.g., EOSC Life Turning FAIR into Reality WG) to ensure alignment with parallel efforts and to help improve uptake of outputs.

- *A description of how the WG plans to develop consensus, address conflicts, stay on track and within scope, and move forward during operation, and*

<u>To develop consensus, address conflicts, stay on track and within scope, and move forward during operation:</u>

By its own stated goals, this WG aims to resolve conflicts of interest across different regions around the world to harmonise existing FAIR data policies. This will require addressing sensitivities in relation to each region and which can only be overcome by inclusion of all relevant stakeholders. RDA will provide the forum to increase the visibility of this WG and its members form a truly global community which is perfect for the underlying issues to be tackled by this WG. Moreover, engagement with those parties that are also doing similar work outlined above will have a significant impact and will help resolve conflicts. The foundation for creating a set of guidelines will hinge on finding commonalities across all regions.

Sub-groups will be formed and tasked with each of the stated objectives and these will report back to the rest of the WG and the wider community with results to meet each milestone and drive uptake.

- *A description of the WG's planned approach to broader community engagement and participation.*

<u>Community engagement and participation:</u>

- Health Data IG
- FAIR Data Maturity Model WG
- Working Group for Data Security and Trust (WGDST)
- WDS/RDA Assessment of Data fitness for Use WG
- RDA/CODATA Legal Interoperability IG
- RDA/NISO Privacy Implications of Research Data Sets IG
- Ethics and Social Aspects of Data IG
- PID Kernel Information WG
- Reproducibility IG
- Metadata Standards WG
- Metadata IG
- PID IG
- PID Kernel WG
- Data Foundations Terminology IG
- Research Data Provenance IG
- Go FAIR IG
- RDA FAIRsharing WG

Further to these, there will also be engagement with consortia, organisations and communities outlined previously.

**5. Adoption Plan** *A specific plan for adoption or implementation of the WG Recommendation and other outcomes within the organizations and institutions represented by WG members, as well as plans for adoption more broadly within the community. Such adoption or implementation should start within the 18 month time frame before the WG is complete.*

Adoption of the guidelines that will be generated by this WG will be undertaken by members of the F4H consortium in the first instance for validation. Since there are known jurisdictional variances in health research data from within the project and its members, this can be used as a sandbox for the wider global context. Nevertheless, we aim to test the guidelines beyond the EU context, with North America and Australia in particular as specific targets due to their already advanced state of data protection legislation.

Training documents will be generated that will allow interpretation of the guidelines into tangible results, and which will be tailored for local contexts where necessary. These will be aimed primarily at early career researchers but will also require buy in from senior management to encourage their use. To this end, all stakeholders will be engaged where possible to ensure that this can happen.

A checklist and/or rubric will be made that can be used by those generating data and/or responsible for their curation to easily manage data objects and their workflows to comply with a unified FAIR data policy.

**6. Initial Membership** *A specific list of initial members of the WG and a description of initial leadership of the WG.*

| | Name | Affiliation / FAIR Project |
|---|---|---|
| **1** | Faisal M. Fadlelmola | Centre for Biomedical Informatics and Systems Biology, University of Khartoum, SD |
| **2** | Oya Beyan | RWTH Aachen University, DE |
| **3** | Chris Duncan | NIEHS, USA |
| **4** | Janice Masud-Paul | Drexel, USA |
| **5** | Martine Gagnon | Universite Laval, CA |
| **6** | Philippe Despres | Universite Laval, CA |
| **7** | Eve Paquette-Bigras | Universite de Montreal, CA |
| **8** | Anne Cambon-Thomsen | SHARC group RDA IG co-leader; in charge of ethical aspects in the IMI FAIRplus project (EU funded), FR |
| **9** | Jeremy Geelen | CIHR, CA |
| **10** | Alyssa Grimshaw | Yale, USA |
| **11** | Shannon Sheridan | Drexel, USA |
| **12** | Mark Leggott | RDC, CA |
| **13** | Luiz Bonino | GO FAIR, Europe |
| **14** | Jennie Larkin | NIH NIDDK, USA |
| **15** | Amy Nurnberger | MIT, USA |
| **16** | Eliane Fankhauser | DANS, NL |
| **17** | Hannah Calkins | Children's Hospital of Philadelphia, USA |
| **18** | David Carr | Wellcome Trust, UK |
| **19** | Young-Joo Lee4 | Johns Hopkins Medical Institute, USA |
| **20** | David Sampson | American Society of Clinical Oncology, USA |
| **21** | Shigeru Yatsuzuka | NBDC, JP |

| | | |
|---|---|---|
| **22** | Yahia Mohamed | INIST CNRS, FR |
| **23** | Jacqueline Ringersma | WUR, NL |
| **24** | Mingfeng Wu | ARDC, AUS |
| **25** | Fabrizio Gagliardi | UPC/BSC, ES |
| **26** | Jane Greenberg | Drexel, USA |
| **27** | Matthew Viljoen | EGI Foundation, Europe |
| **28** | Birgit Schmidt | Goettingen, DE |
| **29** | Laura Palumbo | Rutgers, USA |
| **30** | Frankie Stevens | AARNet, AUS |
| **31** | Andrew Treloar | ARDC, AUS |
| **32** | Jessica Hrudey | VU Amsterdam, NL |
| **33** | Leslie McIntosh | RDA, USA |
| **34** | Serena Battaglia | ECRIN-ERIC, Europe |
| **35** | Sergei Gorianin | ECRIN-ERIC, Europe |
| **36** | Anthony Juehne | RDA North America, USA |
| **37** | Mary Uhlmansiek | WUSTL, USA |
| **38** | Talapady Bhat | NIST, USA |
| **39** | Dina Paltoo | NLM/NIH, USA |
| **40** | Mario Gaspar Silva | INESC-ID, PT |
| **41** | Kristan Kang | ARDC, AUS |
| **42** | Thordis Sveinsdottir | DCC, UK |
| **43** | Piotr Wojciech Dabrowski | HTW University of Applied Sciences Berlin |
| **44** | Jay Greenfield | Data Documentation Initiative |
| **45** | Alexander Bernier | Centre of Genomics & Policy, McGill University |
| **46** | Marcia Levenstein | Vivli |
| **47** | Michela Bertero | Centre for Genomic Regulation |
| **48** | Mar Valverde | Lawyer, Data Protection Officer (Spain) |
| **49** | Alicia Fátima Gómez | FECYT |
| **50** | Javier de la Cueva | Independent consultant; also Richfields, NFS-Cloud |

Initial leadership of the WG:

S. Venkataraman, DCC/ FAIR4Health

Celia Alvarez-Romero, Servicio Andaluz de Salud/ FAIR4Health

Kristan Kang, ARDC

(1 or 2 more co-chairs being actively sought, preferably from N. America and/or Africa/Asia/S. America)