Proposed WG to the RDA - Case statement

The BioSharing Registry: connecting data policies, standards & databases in life sciences

The rationale and the needs for this WG were also presented at RDA Plenary 3, in Dublin:
http://www.slideshare.net/RebeccaLawrence1/rda-digital-repositories-wg-slides-0314-final

1. EXECUTIVE SUMMARY

The product of this WG will be a ***searchable registry*** of reliable and linked information about ***databases***, ***content standards*** (as defined in section 2.2) and ***journal and funder policies*** in the ***life sciences*** on which a variety of stakeholders can base their decisions. Specifically, *journals*, *researchers* and *funders* will be able to recommend or select mature and community endorsed databases and standards, and *developers* and *curators of repositories* and *content standards* will be aware of the requirements they need to meet to ensure their products are discoverable and well described so that they can be used by researchers or recommended by journals and funders.
To ensure a manageable project size and delivery within the 18-month timeframe, the WG will focus on the life sciences and leverage the existing BioSharing[1] effort; the principles and the registry, however, will be developed in a way that is extendable to other areas of science. This ***use cases-driven*** WG is led and constituted by ***prospective adopters*** as well as ***technical implementers***, many of whom are also leading and/or actively involved in other relevant RDA IGs and WGs, whose activities this WG complements and with which this WG will work closely.

2. CHARTER

2.1. Deliverables and beneficiaries

This WG will (i) develop ***principles*** for linking information about databases, content standards (as defined in section 2.2) and journal and funder policies in a sample area of the life sciences, and (ii) deliver a curated and ***prototype registry*** to access and cross-search the information, leveraging on BioSharing.
The result will help stakeholders to ***make informed decisions***, for example:

- *Journals* on repositories that meet the requirements specified by their guidelines, but also meet the necessary content standards.
- *Researchers* on which journals meet their funder requirements, which repositories meet which journal standards; also on which standards meet their specific needs for data management, and subsequently for data sharing (per funders) and publication.
- *Funders* on which journals and repositories meet their policies; also to get an understanding of the current landscape of community defined-standards and databases to refine their recommendations, and for comparison and reference purposes to identify gaps.
- *Developers* and *curators of repositories* and *of content standards* on which exists, who can be reused or even extended to meet their specific needs for data management, and subsequently for data sharing (per funders) and publication; but also on the requirements they need to meet to ensure their products are discoverable and well described so that they can be:

    - evaluated and recommended by journals and funders in their policies;
    - used by researchers to meet their funder policies and the policies of the journals they wish to publish in.
- *Librarians* that support scholars to or themselves aim to:
    - utilize data standards; and conform to journal, institutional, and funder policies.
    - develop and maintain institutional data and publication repositories.

2.2. Motivation and background

Several data management, sharing policies, and plans have emerged in the life sciences in response to increased funding for data-intensive science such as high-throughput approaches in genomics and functional genomics, large volumes of MRI data, etc. As part of the worldwide growing movement for reproducible research, the efforts of funding agencies and journal editors are converging to encourage awardees and authors to provide the underlying data together with a description of that data and the methods used to generate the data, providing such details in a standardized manner and making it available (publicly or via controlled access) for reuse. In parallel, a growing number of community-based groups are developing standards, including ***content standards*** for both data and experimental metadata. Broadly divided into: (i) ***reporting requirements*** (or checklists, outlining the minimal information content that should be reported), (ii) ***terminologies*** (such as controlled vocabularies, thesauri, ontologies), and (iii) ***formats*** (defining the representation and transmission formats or syntaxes that facilitate the exchange of information). These content standards enable data sets to be harmonized with regard to their structure, formatting, and annotation so as to open their content to transparent interpretation and, in principle, enable them to be reproduced, compared and/or integrated. Researchers, bioinformaticians, and developers continue to participate in the development of standards-compliant databases to support data sharing; there are similar trends in both the regulatory arena[e.g.2] and commercial science[e.g.3,4] , which have invested heavily in resources to integrate external information with internal data to enhance the decision-making process.

As a consequence of this general mobilization to support reproducible research there are more than a 1000 biology databases[5], over 300 terminologies[6,7], more than 100 reporting guidelines[8,9], over 150 exchange formats, and a growing number of data preservation, management, sharing policies and plans that could help in the annotation, reporting and sharing of life science datasets.

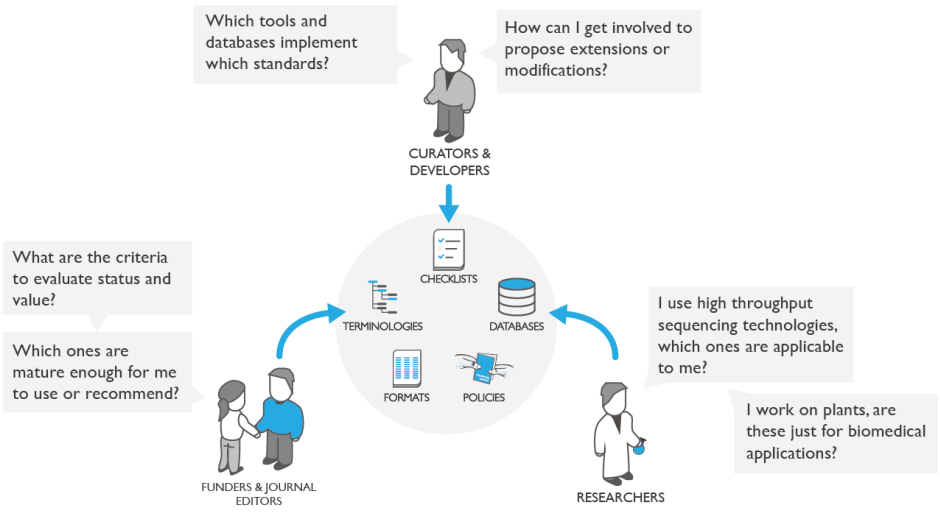
2.3. The problem addressed, the gap filled

Funders and journals cannot anchor their guidance to solid ground. There is not enough information to make informed decisions on which databases or content standards should be recommended. Data sharing policies are unclear; a very common

and loose text is: "*Applicants should make use of existing, recognised standards for data collection and management, where these exist, and make data available through existing community resources or databases where possible*". But what constitutes a recognised standard or acceptable community resource? The same issues exist in the publishing world, but for a few examples where some journals are working to implement much more detailed policies[e.g. 10,11,12,13]. Similarly, reviewers and editors can't sufficiently check for compliance because of this nascent guidance. Finally, there is a disconnect between those information or computer scientists creating and implementing data standards and those that perform review.

Researchers, developers and curators lack support. Systems such as DMPTool[14] certainly help to create a data management plan. Nevertheless, no guidance is given on how to best navigate and select the various content standards and understand their maturity, or find databases that implement them. The struggle researchers and those supporting them (curators and developers) go through is evident. Examples of their questions are: "*Are there content standards for publishing and archiving metagenomics and metatranscriptomics data? The data sharing policy of my funder recommends the use of 'established standards', but which ones are widely endorsed and applicable to my wheat functional genomics data?*".

The absence of well described and cross-linked information about databases, content standards and policies is glaring. This bewildering array of resources cannot be easily discovered, let alone searched and monitored. There is no central site online that comprehensively catalogues, registers, or federates information on these resources; actively curates them, keeps their descriptors up-to-date, monitors their maturity, provides versioning; and collects metrics of usage and level of endorsement. Without consistent metadata describing and categorizing individual standards, database, and policy (*e.g.*, according to the different life science domains or data type) it is very hard to identify the relevance of a resource, and even cross-link them.



3. VALUE PROPOSITION

The registry of curated and linked information about ***databases***, ***content standards,*** and journal and funder ***policies*** will be a searchable environment for the evolving portfolio of these life science resources on which a variety of stakeholders can base their decisions. It will also serve to educate and foster communication between researchers, developers, funders, editors, librarians and other stakeholders. The catalogue will ensure these life sciences resources are **registered**, **informative**, **discoverable** and **accessible**, **maximizing their adoption and use** to assist the virtuous data cycle, from generation to standardization through publication to subsequent sharing and reuse.

3.1. Tangible impacts

The proposed project will improve information about the standards and the databases (maturity, uptake, implementation); provide information to funders and journals about what standards *are* the appropriate community norms, what databases implement which standards or is appropriate for a certain data types, or where data is curated and openly available (or access is regulated for e.g. ethical reasons) etc. Improving the quality in lists of databases and standards will allow funder/journal policies to encourage transparent information and recommendation of community norms. Interlinking allows the project to close the loop: here are the databases and standards; here are the policies that refer to them (or not). For example, when standards are mature and appropriate standards-compliant systems become available these are channeled to the appropriate stakeholder community, who in turn endorse (in policies) or implement (in databases) them achieving wider harmonization of the data.

4. COMMUNITY ENGAGEMENT AND REUSE OF EXISTING WORK

4.1. Leveraging on an existing registry

Run by Susanna-Assunta Sansone's team at the University of Oxford and maintained as a community resource in collaboration with journals and related portals[15,16,17], the

BioSharing registry already offers: (i) ***several functionalities***, (ii) ***extensive content***, (iii) a ***network of collaborators***, and (iv) ***growing recognition*** by funders[18,19] and journals as a central effort to map the landscape of content standards in the life sciences. Leveraging on this existing efforts will also ensure that the proposal is actionable and implementable within the 18 months time frame of the WG. A brief BioSharing history and scope, along with existing content and functionalities of the registry are provided in this section.

Building on the Minimum Information for Biological and Biomedical Investigations' (MIBBI) portal[8] that only listed reporting requirements (or checklists), BioSharing started in 2009 as a blog to accompany a paper published in *Science*[20] with a range of representatives from US, UK and European funding agencies expressly to centralize links to the data policies of major funders. Since its launch as a registry and catalog in 2011 BioSharing supersedes and includes MIBBI, and works to map the landscape of community developed content standards in the life sciences (broadly covering biological, natural and biomedical sciences). BioSharing's goal is to ensure standards are informative and discoverable, monitoring their development, evolution and integration; implementation and use in databases; and adoption in data policies by funders and journals. Currently BioSharing lists over ***500 content standards*** and over ***600 databases*** (partly cross-linked and curated) in the life sciences, collected manually and/or submitted by users. Terminologies are linked to BioPortal[16], world's most comprehensive repository of biomedical ontologies; a closer connection between the two registries is work in progress that will contribute to the work proposed by this WG. BioSharing works with the Oxford University Press (OUP), via its DATABASE and NAR Database Issue journals, to collect harmonized descriptions of the databases, following the guidelines[21] co-developed by BioSharing and the International Society for Biocuration[22]. Lastly, to ensure that one common record exists for each database in the life sciences and that for these areas BioSharing is the reference system, a MoU has been established with re3data[23] and work is in progress to implement the agreement.

BioSharing registries already provides ***core functionality*** to manage the content that can be extended and adapted. These include: (i) search and filtering; (ii) submissions forms to add new records; (iii) "claim" functionality of existing records (to ensure maintainer of standards and databases can keep their records updated); (iv) person's profile (as maintainers of records) is associated to their ORCID profile; and (v) visualization and views of content. Furthermore, BioSharing is becoming an important component of the nascent ELIXIR[24] infrastructure of resources and registries in the life sciences; as part of this larger effort it will be progressively connected to an existing registry of tools. Cross linking standards, databases and tools is the ultimate goal, along with linking these to additional resources, such as scholarly profiling and tools to create data management plans.

Lastly, WG members Jessica Tenenbaum, Susanna-Assunta Sansone, and Melissa Haendel have already laid the basis to develop criteria to be used in evaluating standards for adoption[25]. In conclusion, BioSharing offers a unique base, from the content, functionalities and network of community's view points, on which to build the proposed registry.

4.2. Existing RDA IGs and WGs

Co-chairs and members of this WG are already involved in other approved or proposed RDA IGs and WGs groups[26] that are relevant or related to the registry, including, but not limited to:

- Metadata Standards Directory WG
    - which we relate to, with minimal overlap (that will be resolved by working closely with this group) but extending on their scope because our WG will provide (i) deep and granular focus on the life science domains and (ii) the interlinking value of content standards with databases and policies.
- RDA/WDS Certification of Digital Repositories IG
- (proposed) Global Registry of Trusted Data Services IG and Data Fabric IG
- RDA/WDS Publishing Data IG and joint RDA/WDS WGs for Publishing Data Workflows, Publishing Services and Bibliometrics
- RDA Dynamic Data Citation WG
- Several IGs in the life sciences, where content standards are also key, including Elixir Bridging Force IG, Metabolomics, Toxicogenomics Interoperability, Biodiversity Data Integration, Agricultural Data Interoperability, Marine Data Harmonization etc.

4.3. Related groups and efforts outside RDA

Co-chairs and members of this WG have already collaborations and links with several groups, currently operating outside the RDA umbrella, that are already part of the BioSharing community[15], or have relevant content or infrastructure the registry will connect to, use directly or interoperate with. ***Beside the groups already represented by the co-chairs and the core members*** (see section 7), others include, but are not limited to:

- JISC-funded services such as Sherpa/Juliet[27] for funder policies, BRISSKit[28] for biomedical research, projects such as PREPARDE[29] (in particular the criteria developed for a repository to be considered objectively trustworthy), and JoRD[30], which undertook a scoping study for a database of journal policies and developed a draft schema for their representation.
- Innovative Medicine Initiative (IMI)'s eTRIKS[31] project, also defining metrics for selecting standards in biomedicine, that also include the Clinical Data Interchange Standards Consortium (CDISC)[32].

- Pistoia Alliance[4], a global, not-for-profit, precompetitive alliance of life science companies, vendors, publishers, and academic groups that aims to lower barriers to innovation by improving the content standards[33] and interoperability of R&D business processes; their initial discussions and needs have already been documented[34].
- Consortia Advancing Standards in Research Administration Information (CASRAI)[35], a non-profit standards development organization.
- ISNI[36] International Authority (ISNI-IA), defining institutional identifiers.
- ORCID[37] creating and maintaining a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers.
- re3data[23], also via the MUO with BioSharing.
- International Society for Biocuration[22] with whom BioSharing has already developed guidelines for the description of databases.
- Open standards for scholarly profiling such as VIVO[38].
- The DMPTool partner institutions, because connection with system that help creating data management plans is critical for dissemination and uptake.

5. DEVELOPMENT PLAN

The proposal is actionable and implementable, and realistic within the 18 months time frame, because (i) it leverages on the existing content and functionality of BioSharing, which in turn (ii) is embedded in an ecosystem of complementary registries, such as BioPortal, and those being connected via the ELIXIR infrastructure, such as a registry for tools; and (iii) benefits from an extensively networked membership and an operational team, committed to drive and carry out content and technical work required to deliver a usable registry.

5.1. The workplan

The work and the tasks will be organized in Work Packages (WPs) and timelines are outlined in the Gantt chart below.

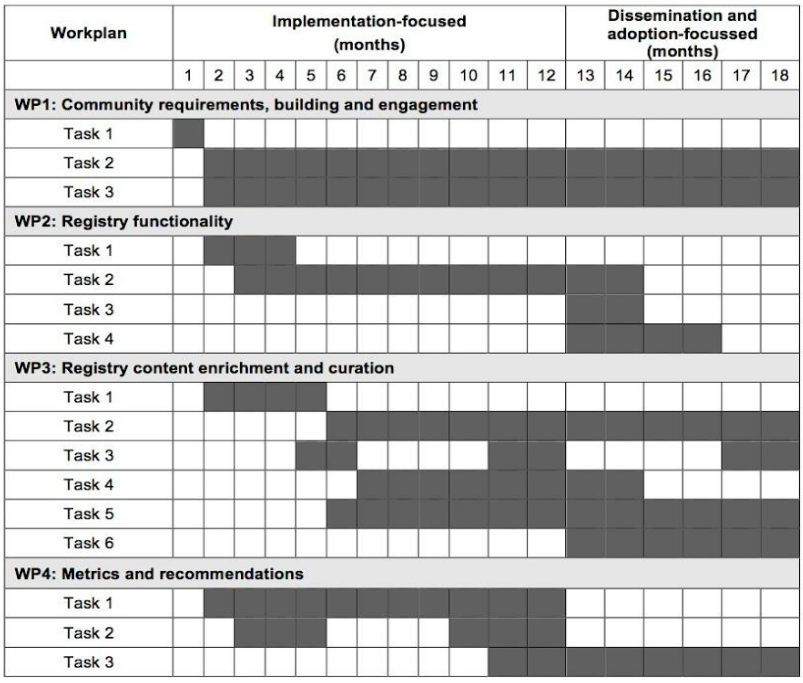
- **WP1: Community requirements, building and engagement**
    - **Task 1:** Collect use cases from adopters through interviews/focus groups with stakeholders, e.g. publishers, funders, researchers, curators: content coverage and top 10 key queries.
    - **Task 2:** Ensure continued dissemination and feedback, using (but not limited to) the existing network of the members and events outside RDA plenaries.
    - **Task 3:** Manage engagement and communication with relevant communities (see section 4).
- **WP2: Registry functionality**

- **Task 1:** Review BioSharing's existing backend and frontend functionalities (e.g. fields used to describe policies, content standards and databases, but also type of queries that it enables), identifying requirements based on outcome of WP1 Task 1.
- **Task 2:** Implement the identified modifications and additional functionalities; test (and solicit feedback) iteratively against outcome of WP1 Task 1.
- **Task 3:** Discuss and implement identifiers and versioning strategy for the records, e.g. via DOIs and/or persistent URIs, also coordinating with Force 11[39].
- **Task 4:** Implement an Application Programming Interface (API) and Web Service (WS) interface: so that the catalogue can be plugged into other, third party applications as a resource.

- **WP3: Registry content enrichment and curation**
  - **Task 1:** Define consistent metadata to describe and categorize individual standards, databases and policies (*e.g.* according to the different life science domains or data types).
    - Write and publish it as an open document
  - **Task 2:** Review BioSharing's existing content for policies, content standards and databases, identify and adding missing content and develop a strategy for content acquisition.
    - The approach will be both *pull***,** by active automatic harvesting (federation) of information from existing resources, and *push*, by encouraging submission of additional records from the community.
  - **Task 3:** Implement global identifiers:
    - Maintainers of records are already associated to their ORCID (as a reward and accreditation mechanism to help drive new contributions, giving a sense of ownership and enabling the network effects of a community).
    - FundRef[40] IDs will be used to tag funders (of standards, databases, but also as creators of policies).
    - Implementation of group institutional identifiers, when these become available
  - **Task 4:** Assemble journal and funder policies regarding the use of named standards and deposit in specific databases
    - Review the existing content of Sherpa/Juliet, the schema outlined by JoRD (a scoping study and business case rather than a service), and any other available information sources or schema that can be reused, leveraged etc.

- **Task 5:** Ensure all records are “claimed” by maintainers (a functionality that already exists in BioSharing)
  - Contact and invite relevant groups/people to claim, update and maintain relevant records
- **Task 6:** Cross-link the content by create relations among policies, content standards and databases, via their relevance to specific life science areas, technologies and data types.

- **WP4: Metrics and recommendations**
  - **Task 1:** Using collected information, develop formal criteria to assess the maturity of the standards, their usage in databases, standing in the community and level of endorsement; work will be done closely ongoing activities in ELIXIR, relevant RDA IGs and WGs , the Standards Group of the NIH Big Data to Knowledge (BD2K) Initiative[41], and others also tackling repositories accreditation.
    - Generate recommendations and co-use information, tagging the records to drive queries and facilitate filtering of the results.
    - Outline proposal for methods and tools that will be needed to monitor the criteria.
  - **Task 2**: Develop recommendations to standards-developing communities to identify and version their files, e.g. technical specifications, suggesting suitable places where these can also be stored and accessed.
  - **Task 3:** Usability testing and iterative UI development (limited in scope for initial prototype)



5.2. The short and long term goals

Within the 18 months life span, the WG will:

- Develop a registry prototype, leveraging the BioSharing system and content;
- Draft an operational plan for sustainability (growth and maintenance) of the registry, e.g. as part of the ELIXIR infrastructure and the NIH BD2K, using EU as well as national and international funding mechanisms.

As part of the long terms goals, the WG will:

- Seek official recognition of the output by RDA
- Propose to become an IG to ensure continued engagement with other IGs and WGs.
- Continue to link the registry with additional resources in ELIXIR, e.g. a registry of tools and a catalog of training material.
- Seek further integration with ORCID (e.g. records 'claimed' by maintainers of databases and standards could also be visible on the person's ORCID page).
- Initiate integration with other open standards for scholarly profiling such as VIVO and the DMPTool to create data management plans.
- Continue to investigate how best to monitor evolution and use of standards, working closely with other relevant groups.
    - How can start evolve efficiently and effectively if we do not know and monitor who is using them? For example, if a terminology is used by a group, but another wish to extend/change it, what are the downstream effects of changing it?
- Monitor the adoption rate, outside the early endorsement by the core members.

6. OPERATIONAL AND ADOPTION PLAN

Months 0-12 will be dedicated to the user requirements and implementation phase. Regular monthly virtual meetings will be held among co-chairs, members and the operational team, using web meetings functionality. Bi-monthly virtual meetings, or more often as required, will be held between the co-chairs and the operational team. When possible, face-to-face meetings will be organized, particularly in conjunction with existing and relevant conferences and events. The co-chairs will be responsible for moderating the discussion and drive the development to meet the deliverables according to the timelines (as in the Gantt chart), along with the operational team and under the advice and guidance of the members. Conflicts and any time adjustment on the timeline developments will be managed and addressed by the co-chairs, as appropriate.

Within the **12-18 months** we will initiate specific activities, geared towards the dissemination and adoption phase. The specific plan for encouraging adoption will include publications and presentations via RDA, CODATA, ELIXIR, NIH BD2K

meetings and those of other partners. Early endorsement by the core members will also be used as ***adoption exemplars*** to other communities.

7. CORE MEMBERS AND INITIAL ADOPTERS

This group has a long-standing successful track record in collaborative community development and/or service provision, a high level of background knowledge and competence, but also the technical and social ability to engage with a large and diverse set of collaborators and meet their needs in a timely fashion. Our unique combination of experience and people will build a vibrant, interdisciplinary team to *drive*, *build*, *adopt* and *disseminate* the resulting work.

- **Chairs**:

Simon Hodson (CODATA)	Simon Hodson, PhD, is Executive Director of CODATA[42], an organisation whose mission is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use. He sits on the Board of Directors of the Dryad Data Repository[43], on the Scientific Advisory Board of CESSDA[44] and on the GEO Data Sharing Working Group[45], as well as being a co-chair and member of several RDA IGs and WGs. See his RDA profile[46].
Rebecca Lawrence (F1000)	Rebecca Lawrence, PhD, is Managing Director at F1000Research[47], a sister company to Faculty of 1000 (F1000). She developed and launched F1000Research, a pioneering life sciences journal focussing on transforming the way science is communicated and published through immediate publication, transparent refereeing, and a mandatory open data policy. She is a member of several RDA WGs, a founding member of the STM Data Group[48], and co-Chair of the CASRAI-ORCID Peer Review Service Group[49]. See her RDA profile[50].
Susanna-Assunta Sansone (University of Oxford and NPG Scientific Data)	Susanna-Assunta Sansone, PhD, is Associate Director and Principal Investigator at the University of Oxford e-Research Centre[51], Consultant and Honorary Academic Editor for the Nature Publishing Group (NPG)' Scientific Data publication platform[52]. She also seats on the Board of several international grass-root standards, advocacy groups and non-for-profit efforts, including the Board of Directors of Dryad Data Repository and is a core member of the ELIXIR UK Node[53]; she is also member of the RDA Technical Advisory Board and involved in several IGs and WGs. See her RDA profile[54].

- **Core members and early adopters**:

Varsha Khodiyar (F1000)	Todd Vision (Dryad)	Iain Hrynaszkiewicz[55] (NPG)
Jennifer Lin (PLOS)	Amye Kenall (BMC)	Scott Edmunds (GigaScience, BGI)
Melissa Haendel[56] (OHSU, Force11, Monarch Initiative)	Rebecca Boyles (NIH/NIEHS; also NIH DDICC project team, NIH standards group)	Jonathan Tedds[57] (Ubiquity Press Open Health Data journal[58], BRISSKit / JISC biomedical database software and RDA Publishing Data Co-Chair)
Theo Bloom (BMJ)	Rafael Jimenez (ELIXIR)	Michael Ball (BBSRC Strategy and Policy Manager)
Jennifer Boyd (Oxford University Press)	Thomas Lemberger (EMBO Press)	Jessica Tenenbaum[59] (Duke University)

    - Along with the co-chairs, the core members are the initial like-minded group of individuals that have agreed to initiate the WG, based on the real needs they have, or of those of their communities; hence, this group and respective their use base will also represent the first adopters.
    - Both co-chairs and members will continue to actively reach out to more interested parties, especially during the open community review phase and to ensure geographical distribution and representations from different stakeholders, including advocators that are pivotal for broader adoption.
- **Operational team** (University of Oxford): Milo Thurston, Alejandra Gonzalez-Beltran, Philippe Rocca-Serra, Eamonn Maguire.
    - Based in Susanna-Assunta Sansone's group at the University of Oxford, these individuals have a long-standing, successful and international track record in service provision built with and for the academic and commercial communities, spanning many areas of life science. They will contribute to the overall goal, but also execute the technical tasks and implementing relevant outcomes in the BioSharing registry.

8. REFERENCES

1. BioSharing: biosharing.org
2. Hamburg, Advancing regulatory science, *Science* (2011)
3. Barnes *et al.*, Lowering industry firewalls, *Nat Rev Drug Discov* (2009)
4. Pistoia Alliance: www.pistoiaalliance.org
5. NAR Database Issue: oxfordjournals.org/nar/database/a

6. Bioportal: bioportal.bioontology.org
7. Smith *et al*., The OBO Foundry, *Nat Biotechnol* (2007)
8. Taylor *et al*., MIBBI, *Nat Biotechnol* (2007)
9. Equator Network: equator-network.org
10. F1000Research data policy: f1000research.com/author-guidelines
11. NPG Scientific Data data policy: www.nature.com/sdata/data-policies
12. EMBO Press Data policy embopress.org/sourcedata
13. PloS: www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy
14. DMPTool: dmptool.org
15. BioSharing communities:biosharing.org/communities
16. BioPortal: bioportal.bioontology.org
17. BioCatalogue: https://www.biocatalogue.org
18. BBSRC Resources: www.bbsrc.ac.uk/funding/facilities/resources.aspx#generalbiology
19. NIH Big Data to Knowledge Initiative's report on event on community standards: bd2k.nih.gov/pdf/frameworks_for_comm_based_standards_efforts_report.pdf
20. Field, Sansone et al., Omics data sharing, *Science* (2009).
21. bioDBcore: biodbcore.org
22. International Society for Biocuration: www.biocurator.org
23. re3data and BioSharing MoU: www.re3data.org/2013/11/biosharing-and-re3data-cooperation
24. ELIXIR: www.elixir-europe.org
25. Tenenbaum, Sansone, Haendel, A sea of standards for omics data: sink or swim? *J Am Med Inform Assoc* (2014)
26. List of RDA IGs and WGs: rd-alliance.org/groups
27. Jisc Sherpa/Juliet: www.sherpa.ac.uk/juliet/
28. Jisc BRISSKit: www.brisskit.le.ac.uk
29. Jisc PREPARDE: www.le.ac.uk/projects/preparde
30. Jisc JoRD: jordproject.wordpress.com/
31. IMI eTRIKS:www.imi.europa.eu/content/etriks
32. CDISC: www.cdisc.org
33. Harland *et al.*, Empowering industrial research with shared biomedical vocabularies. *Drug Discov Today* (2011)
34. Pistoia Alliance's notes from break-out discussion on standards: www.slideshare.net/pistoiaalliance/information-ecosystem-standards
35. CASRAI: casrai.org
36. ISNI: www.isni.org
37. ORCID: orcid.org
38. VIVO: vivoweb.org
39. Force 11: www.force11.org
40. FundRef: www.crossref.org/fundref
41. NIH BD2K: bd2k.nih.gov
42. CODATA: www.codata.org
43. Dryad:datadryad.org
44. CESSDA:www.cessda.net

45. GEO Data Sharing Working Group: www.earthobservations.org
46. Simon Hodson's profile: rd-alliance.org/users/simon-hodson
47. F1000Research: f1000research.com
48. STM Data Group: www.stm-assoc.org/research-data-group
49. CASRAI-ORCID Peer Review Service Group: casrai.org/about/announcements/orcid-%26-casrai-kick-off-new-standards-project-on-%E2%80%98peer-review-services%E2%80%99
50. Rebecca Lawrence's profile: rd-alliance.org/users/rlawrence
51. University of Oxford e-Research Centre: www.oerc.ox.ac.uk
52. NPG Scientific Data: www.nature.com/sdata
53. ELIXIR UK Node: www.elixir-europe.org/services/elixir-nodes
54. Susanna-Assunta Sansone's profile: rd-alliance.org/about/organization/key-profiles/susanna-assunta-sansone.html
55. Iain Hrynaszkiewicz's profile: uk.linkedin.com/in/iainhz
56. Melissa Haendel profile: www.ohsu.edu/xd/education/library/about/staff-directory/melissa-haendel.cfm
57. Jonathan Tedds' profile: uk.linkedin.com/pub/jonathan-tedds/16/79b/769
58. Open Health Data journal: openhealthdata.metajnl.com/
59. Jessie Tenenbaum's profile: www.dtmi.duke.edu/directory/jdt19