Meeting report series

Report of the 2nd ISC Working Group on Data Sharing and Bioinformatics teleconference

21 July 2014, 2-4pm

Organization

Organized by: IRDiRC Scientific Secretariat
Teleconference

Participants

Prof Matthew Bellgard, Perth, Australia
Dr Sergi Beltran, Barcelona, Spain
Prof Anthony Brookes, Leicester, UK
Prof Angel Carracedo, Santiago de Compostela, Spain (Co-Chair)
Prof Jeffrey Krischer, Tampa, USA (Co-Chair)
Prof Jacek Majewski, Montreal, Canada
Prof Barend Mons, Leiden, The Netherlands
Prof Deborah Nickerson, Washington, USA
Dr Rachel Richesson, Durham, USA
Dr David Salgado, Marseille, France

Dr Barbara Cagniard, IRDiRC Scientific Secretariat
Dr Sophie Höhn, IRDiRC Scientific Secretariat
Dr Lilian Lau, IRDiRC Scientific Secretariat

Apologies

Prof Eric Boerwinkle, Houston, USA
Dr Ivo Gut, Barcelona, Spain
Dr Matt Hurles, Cambridge, UK
Sir Jeremy Manuel, London, UK
Dr George Patrinos, Patras, Greece
Dr Andreas Zankl, Queensland, Australia

REPORT

Circulated documents

A set of documents was circulated by e-mail to all members of this Working Group (WG):

▶ Information regarding the Interdisciplinary Scientific Committee (ISC) and its roadmap
▶ Feedback from the ISC to the WG on Data Sharing and Bioinformatics
▶ Proposal to develop a Data Standards Clearinghouse for rare diseases
▶ Minutes of the first teleconference of the WG on Data Sharing and Bioinformatics

Members were advised to review these documents in order to be updated with the development involving this WG to date. Attention was drawn to some common themes of all WGs under the ISC, namely shareability and integration of data, raising awareness of other efforts, and common elements that will facilitate data sharing beneficial to researchers.

Proposal to develop a Data Standards Clearinghouse

A proposal was outlined by this group and members were solicited for feedback earlier this year. The response received was largely positive although cautious in terms of the ability to successfully implement this. The proposal was then put forth to the Executive Committee (EC) during a teleconference in June. It is currently still under consideration, and further discussion on this topic is scheduled to take place in September. The EC has shown high level of interest to get this project funded and implemented, but the mechanism to fund this requires further discussion due to the international nature of this project.

Many organizations are involved in creation of data standards or data models, and yet these efforts are not always widely known. The idea behind the clearinghouse is to provide a mechanism to organize data standards, to create a space where data standards could be deposited, and to make it a preferred resource to use thus leading to the coalescence of strong support to become the data standards used by the majority of researchers. This system should be sustainable, dynamic and extensible, even for use beyond rare disease (RD) researches. The basic infrastructure planned would not necessarily be RD-specific but extensible into non-RD domains.

Exploring three possible models of implementation which are not mutually exclusive:

▶ Catalog: simplest model consists of a list of available data standards; given the overarching goals of IRDiRC, this list is a very basic requirement and should be created, even if the clearinghouse is not intended to be a static compendium of standards.
▶ Dynamic: the catalog with an added dimension, where users and community are required to be strongly engaged with the project to use it, to add their resources, to provide comments and feedbacks, to check site functionality; this is a big ask and would be a big challenge to pull off
▶ In-between: related to the dynamic model, where the list includes partly developed standards and the system allows connections between people to collaborate and co-develop standards, while at the same time minimize duplication of standards

The clearinghouse would incorporate both the dynamic and the in-between models:

- ▶ Contains a list of data standards currently available
- ▶ Allows users to select the most appropriate data standards to apply
- ▶ Allows input and submissions from users
- ▶ Allows users to connect and to collaborate
- ▶ Includes use cases associated with standards to identify instances of applications
- ▶ Informs researchers about standards development outside their specific RD fields

Points to take into consideration:

- ▶ Alignment with other efforts, be it RD-specific (e.g. RD-Connect, Innorare) or not (e.g. ELIXIR, BD2K, Data FAIRport) to ensure interconnection and shareability between data
- ▶ Rapidly developing data standards (~90%) in biobanking (e.g. BBMRI); possibility to allow them to extend into this initiative instead of this being extensible to include their elements
- ▶ How to define this particular clearinghouse and its functions
- ▶ How to define the type of data and standards to be included in this clearinghouse
- ▶ How to provide incentives for data sharing and to enhance shareability over time
- ▶ How to maintain and manage the resource, ensuring the information is up to date
- ▶ Establishment of API to share information from different initiatives despite different ontologies

This proposal may require development over a number of phases, starting it to work around small concepts and build from there, bearing in mind the needs to grow and to be extensible for wider adoption later on. The foundation must be able to sustain growth.

For the initiative to start attracting users naturally, a critical mass would be required and this is achievable by limiting certain criteria in the early phases, and when this sub-area is shown to be successful, expansion into other areas would be an organic progression. At the same time, IRDiRC community should give backing and visibility to the initiative too, so it would be supported by a growing community of investigators over time.

If funding can be made available, a meeting/workshop together with other organizations with similar initiatives should be organized, with the aim to coordinate efforts and create a platform useful for all.

RD-Connect

The EC had strongly suggested that this WG collaborate with RD-Connect, which has been working on various standards, including genomic information, ontologies, phenotypes, clinical descriptions, etc. They are also working on patient registries and biobanks. It would be highly desirable to align this initiative with that of RD-Connect, and benefit from their insights and experience. This would also avoid reinvention of the wheel.

Main deliverables

- ▶ Compilation of a list of standards available and the parties involved
- ▶ Provide information of known standards to aid the Scientific Secretariat in compiling the list of standards
- ▶ Additional comments/feedbacks, if any, to be sent
- ▶ Plan the next teleconference to be held in late September 2014