

PLANNING FOR GLOBALLY SUSTAINABLE LIFE SCIENCES DATA RESOURCES

W. Anderson¹, R. Appel⁸, R. Apweiler², G.A. Bauer¹, H. Berman³, N. Blomberg⁴, C.H. de Brito Cruz⁵, T. Donohue⁶, M. Dunn⁷, C. Durinx⁸, P. Goodhand⁹, E.D. Green¹⁰, J. McEntyre², Y. Nakamura¹¹, A. Smith², Y.-Y. Teo¹², M. Woodwark¹³, and P. Lasko¹⁴

1: The International Human Frontier Science Program Organization, 12 Quai Saint-Jean, 67080 Strasbourg, France

2: European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, CB10 1SD, UK

3: Rutgers University, Center for Integrative Proteomics Research (CIPR), SAS - Chemistry & Chemical Biology, 174 Frelinghuysen Rd, Piscataway, NJ 08854-8076, USA

4: ELIXIR, Wellcome Genome Campus, Hinxton, CB10 1SD, UK

5: The Sao Paulo Research Foundation, R. Pio XI, 1500 - Alto da Lapa - CEP 05468-901 Sao Paulo/SP, Brazil

6: Department of Bacteriology, Great Lakes Bioenergy Research Center, & Wisconsin Energy Institute, University of Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726, USA

7: The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK

Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726, USA

8: SIB Swiss Institute of Bioinformatics, University of Lausanne, Batiment Genopode, 1015 Lausanne, Switzerland

9: Global Alliance for Genomics and Health, MaRS Centre, 661 University Avenue, Suite 510, Toronto, Ontario, Canada, M5G 0A3

10: National Human Genome Research Institute, National Institutes of Health, 31 Center Dr., Bethesda, MD 20892, USA

11: National Institute of Genetics, Genome Informatics Laboratory, Yata 1111, Mishima, Shizuoka 411-8540, JAPAN

12: Centre for Infectious Disease Epidemiology and Research (CIDER), Saw Swee Hock School of Public Health, National University of Singapore, 12 Science Drive 2, #10-01, Singapore 117549

13: MedImmune Limited, Milstein Building, Granta Park, Cambridge, CB21 6GH, UK

14: Department of Biology, McGill University, Bellini Life Sciences Complex, 3649 Sir William Osler, Montreal, Quebec, Canada, H3G 0B1

Abstract

A meeting hosted by the Wellcome Trust in June 2017 allowed for a second round of discussions by the Global Life Science Data Resources Working Group on the topic of developing an internationally distributed model of support for core data resources in the life sciences. Senior representatives of data resources and funding organizations further defined the key features and benefits of establishing a new model for the international support of these data resources. A key conclusion was that submission of, and access to, all data associated with the new model should be free for everyone, without any cost or restrictions. The Working Group proposes a step-wise approach for implementing the proposed model. An initial Phase 0 would establish a coalition of funders willing to lend their support based on a draft business plan for the model. Major tasks during Phase 1 would then be to finalize the business plan, select the initial set of core data resources for inclusion within the new funding model, and agree on governance and managerial implementation plans. During Phase 2, the model would fully mature to provide sustainable international support for designated life science core data resources.

The value of core data resources critical to the life sciences and their accessibility to researchers worldwide would be enhanced if there was sustainable, equitable, and priority-driven support from research funding bodies around the globe. Only a global solution would overcome the patchy landscape of grant-based funding for data resources.

Following an initial meeting in November 2016 [1, 2], the International Human Frontier Science Program Organization held a strategic planning meeting of the Global Life Science Data Resources Working Group, hosted at the Wellcome Trust on June 5-6 2017 in London. Attendees included senior managers of major data resources and leaders from funding organizations. This meeting defined the benefits and key features of a future internationally distributed funding model to support core data resources, and proposed a phased approach that would enable its progressive implementation.

One conclusion was that the future of life science research demands a robust, open, and coordinated global data resource ecosystem, which is best sustained by a stable, internationally distributed funding mechanism overseen by a coalition of funders. A fundamental principle underlying this model is that submission of and access to all data in this ecosystem would be free for everyone without restrictions. The Working Group also concluded that contributions from national funding agencies to support core data resources should be proportional to some appropriate economic measure, such as Gross Domestic Product (GDP) or total funding for life science research.

Success of this endeavor would ensure long-term sustainability of the global data resource ecosystem, avoid loss of essential data, strengthen coordination and increase synergy among data resources, and provide an ongoing process for assessing and improving these resources. Countries that take part would enhance their competitiveness in the data economy, benefitting from superior knowledge exchange and improved opportunities for international collaborations and training.

The Working Group's view is that the initial phase of implementing the new model should be limited to 'open-access' life science data – that is, data which is freely and openly available via the internet for anyone to use without first seeking permission - including data derived from medical research. Consideration of resources holding 'controlled-access' human data may be warranted later.

The Working Group anticipates three phases for implementing the new model. The initial preliminary phase (Phase 0) would be focused on building a coalition of funders interested in supporting such a new funding model. Formal principles and a business model for supporting the newly envisioned core data resource ecosystem would be drafted. This phase will also involve an outreach effort to governments, funders, philanthropic organizations, and industry, so as to articulate the benefits of the new model. At the present time, efforts to identify seed funding to facilitate Phase 0 activities are underway.

In Phase 1, the present model of direct funding to individual core data resources would continue, but more concrete steps would be taken to design a model for direct coalition-based funding. This phase would include an international assessment exercise to designate the set of perhaps a few dozen 'core'

data resources (out of the existing thousands) that are so critical to global life science research that they warrant special designation and support. Criteria recently developed by ELIXIR [3] represent a valuable starting point for this work. Once the initial set of core data resources has been designated, their present costs and sources of funding will be determined, and a needs assessment will be conducted to identify areas that require additional funding.

It is also proposed that during Phase 1, a governance plan be negotiated amongst the international coalition of funders, including the establishment of a scientific advisory board and an administrative structure. It is suggested that a small pilot also be conducted during this phase using pooled funds, one that aims to enhance existing core data resources and/or to establish new ones. This would allow for testing and analysis of the governance and management of the coalition. At this stage, the management could begin to address the desirability of extending the scope of existing core data resources and to define potential mechanisms for improving interactions among data resources.

Phase 2 would involve the full evolution into international coalition-based funding. Substantial support and leadership from research funders will be critical at this stage. Building upon the findings of the pilot effort, it is envisaged that Phase 2 will involve a major increase in resources to the coalition's pooled funds and an expansion of its mandate to include financial support of the set of designated core data resources themselves. Modifications to the governance plan may also be required to effectively address the coalition's expanded mandate. The necessary funds for Phase 2 could come from increased commitments from funding agencies and potentially from industry, private philanthropy, and other non-governmental sources.

We are publishing this summary to invite those who have not been involved to date, but who have a vital interest in the sustainability of core data resources, to help shape these concepts and engage in the development of this important and ambitious initiative. Building on the two workshops, we have begun to draft a plan for implementing these concepts. We look forward to engaging with funders, science policy-makers, data resources, and other stakeholders to develop this further into a business plan. We will regularly report on our progress through various mechanisms, including regular postings on the HFSP website (www.hfsp.org).

Disclaimer

The information and views set out in this article are those of the authors and do not necessarily reflect the official opinions of the institutions to which the authors are affiliated.

References

[1] Warwick Anderson, Rolf Apweiler, Alex Bateman, Guntram A Bauer, Helen Berman, Judith A Blake, Niklas Blomberg, Stephen K Burley, Guy Cochrane, Valentina Di Francesco, Tim Donohue, Christine Durinx, Alfred Game, Eric Green, Takashi Gojobori, Peter Goodhand, Ada Hamosh, Henning Hermjakob, Minoru Kanehisa, Robert Kiley, Johanna McEntyre, Rowan McKibbin, Satoru Miyano, Barbara Pauly,

Norbert Perrimon, Mark A Ragan, Geoffrey Richards, Yik-Ying Teo, Monte Westerfield, Eric Westhof, Paul F Lasko, *A global coalition to sustain data resources*. Nature 543, 179 (2017), 8 March 2017doi: [10.1038/543179a](https://doi.org/10.1038/543179a)

[2] Warwick Anderson, Rolf Apweiler, Alex Bateman, Guntram A Bauer, Helen Berman, Judith A Blake, Niklas Blomberg, Stephen K Burley, Guy Cochrane, Valentina Di Francesco, Tim Donohue, Christine Durinx, Alfred Game, Eric Green, Takashi Gojobori, Peter Goodhand, Ada Hamosh, Henning Hermjakob, Minoru Kanehisa, Robert Kiley, Johanna McEntyre, Rowan McKibbin, Satoru Miyano, Barbara Pauly, Norbert Perrimon, Mark A Ragan, Geoffrey Richards, Yik-Ying Teo, Monte Westerfield, Eric Westhof, Paul F Lasko, *Towards coordinated international support of core data resources for the life sciences*. bioRxiv, 27 April 2017. doi: [10.1101/110825](https://doi.org/10.1101/110825)

[3] Christine Durinx, Jo McEntyre, Ron Appel, Rolf Apweiler, Mary Barlow, Niklas Blomberg, Chuck Cook, Elisabeth Gasteiger, Jee-Hyub Kim, Rodrigo Lopez, Nicole Redaschi, Heinz Stockinger, Daniel Teixeira, Alfonso Valencia, *Identifying ELIXIR Core Data Resources*. F1000Res, 30 September 2016. doi: [10.12688/f1000research.9656.1](https://doi.org/10.12688/f1000research.9656.1)