

Supplementary information

Genomic data sharing: you don't know what you've got (till it's gone)

In the format provided by
the authors and unedited

Supplementary Box 1. Key elements of the Bermuda Principles and the Fort Lauderdale Agreement

Bermuda Principles¹:

Primary Genomic Sequence Should be in the Public Domain

It was agreed that all human genomic sequence information, generated by centres funded for large-scale human sequencing, should be freely available and in the public domain to encourage research and development and to maximise its benefit to society.

Primary Genomic Sequence Should be Rapidly Released

- Sequence assemblies should be released as soon as possible; in some centres, assemblies of greater than 1 kb would be released automatically on a daily basis.
- Finished annotated sequence should be submitted immediately to the public databases.

It was agreed that these principles should apply for all human genomic sequence generated by large-scale sequencing centres, funded for the public good, to prevent such centres establishing a privileged position in the exploitation and control of human sequence information.

Fort Lauderdale Agreement (abridged)²:

Tripartite sharing of responsibility

Funding agencies should:

1. designate appropriate efforts as community resource projects, and encourage resource producers to prepare and submit Project Descriptions (see below) for publication
2. require, as a condition of funding, free and unrestricted data release from community resource projects to appropriate central and searchable public databases, and vigorously ensure that this occurs
3. encourage more investigators to serve the community through involvement in such projects. In particular, the agencies should ensure that investigators engaged in generation of such datasets have sufficient support for curation, maintenance and distribution of the data to the community, as well as resources to perform initial analyses using the resources that they have generated
4. ensure that a centralized view of existing community resource projects is available as an information source for the community
5. support central databases that will house and distribute the data in a way that prevents fragmentation of the data.

Resource producers should:

1. when feasible, publish a Project Description. The purpose of the Project Description, which will be a new type of scientific publication, is to inform the scientific community about the resource project and to provide a citation to reference the source of the data. The Project Description should be written at the beginning of the project and describes

the plans for and scope of the production and analyses that the data producer intends to undertake. It will often include a timeline for production goals and data release.

2. produce data of consistently high quality
3. make the data generated by the resource immediately and freely available without restriction
4. recognize that even if the resource is occasionally used in ways that violate normal standards of scientific etiquette, this is a necessary risk set against the considerable benefits of immediate data release

Resource users should:

1. appropriately cite the source of the data analysed and acknowledge the resource producers. The early publication of a Project Description, as suggested above, would provide users with an appropriate reference to cite before the data are formally published
2. recognize that the resource producers have a legitimate interest in publishing prominent peer-reviewed reports describing and analysing the resource that they have produced (and that neither the Project Descriptions nor data deposits in databases are the equivalent of such publications)
3. respect the producer's legitimate interests as set out, e.g. in a Project Description, while being free to use the data in any creative way. There should be no restrictions on the use of the data, but the best interests of the community are served when all act responsibly to promote the highest standards of respect for the scientific contribution of others. In some cases, this might best be done by discussion or coordination with the resource producers
4. assist journals and funding agencies to play their proper roles in ensuring, through the peer review system, that the system works fairly for all constituents.

References

1. Cook-Deegan, R. & McGuire, A. L. Moving beyond Bermuda: sharing data to build a medical information commons. *Genome Res.* **27**, 897–901 (2017).
2. Matharu, J. Sharing data from large-scale biological research projects: a system of tripartite responsibility. <https://www.sanger.ac.uk/wp-content/uploads/fortlauderdalereport.pdf> (2003)