

## World Café Output

**Table 1**

What do I value most about data sharing / what possibilities do I hope will be part of its future?

Unanticipated use of data (data reuse) and gaining new discoveries and insights - Serendipity
Data aggregation for improved modeling, machine learning, tool development to maximize data utility
Sharing will enhance the better return on investment, improve efficiency, and quality of the data metrics
Create sustainable infrastructure and tools for users at all levels to connect with and use the data (bioinformaticians, bench scientists, data scientist and public/citizen scientists).
Innovation and empowering
Ability to extrapolate results to other cell lines/species/animal models

**Table 2**

Regarding biological data sharing, what surprised have you found? What challenges? What's exciting?

Use of ontologies and other standards is resisted or faked (make up your own or pick nearest good-enough fit).
Cross-disciplinary partnerships can really help your data handling (computer and data scientists) and understanding of your results.
Data sharing should include samples, protocols, data processing as well as repositories.
Data repositories are also quite stove-piped.
Genomic data representation standards are working very well.
Are we sure all the data is worth sharing?
Is all the data worth keeping?
Long-term infrastructure plans are needed to handle increasing data types and sizes - funding?
Industry partnership is important.
How will policy changes and data entanglement affect data sharing?
Data integration across data structures, optimized for a specific type of structure, how does this work?
It's surprising how often there are no comparable units. It's exciting what we can do with comparable units!
Exciting - Biological data sharing has contributed to convergence in research.

**Table 3**

What types of biological data sharing among the 4 quadrants are working well in your opinion? Why? What enables this?

When teams are planned or forced to work well together (across quadrants)
When designed to solve problems - goal oriented, so data sharing is inherent to success
When metadata or orders of magnitude of data are "smaller", for example genomics vs phenomics types of data that have curated repositories, community (and/or funder) support
When there is profit motive
within or among similar agencies/institutions/communities
when there is less concern or complication regarding PII

When there is responsibility or accountability that goes back to the researcher/data generator (as part of a community)
When the data can be codified (sequences, chemical structures)

**Table 4**

What types of biological data sharing among the 4 quadrants are not working well in your opinion? Why? What are the barriers?

Patient data and rest of the quadrants is the most difficult.
One issue with the OMICS data is that its not interoperable. Because they may use different experimental methods, which may lead to similar data sets, but the descriptors are not adequate.
Legacy systems- data base repository may be dated technology wise.
Data integration is also an issue, for example, OMICS data all have separate repository. How to make the global identifier work better in future. Biodiversity (specimen), need a common link to make the best use of very different data types (between institution and between repository).
Need well documented sets of data that goes into the modeling. Again transparency is the issue.
Data dump and miss annotation of OMICS data. Would automation and simplification of annotation help?
It is not always easy to harvest the data in repositories (data quantity, quality, descriptors, extensions) to carry out data analytics.
Metadata is often either missing or else not grounded in resolvable reference terms, such that each dataset is its own private language.
lack of standards and ontologies

**Table 5**

What's missing from the biological data sharing conversation? What topics have we missed?

Private industry has different motivations and constraints. How do respect those while still gaining appropriate access to useful data?
Bias. When data is combined from separate studies there may be a greater risk of bias because it was not designed to exclude this. How do we deal with this and other biases?
Quality of data. How do we communicate the quality of data. Do we include low quality data in our efforts?
How protect against data fraud? Particularly if scientist are being incentivized for data release regardless of quality.
What are the benefits and drawbacks of highly centralized databases?
The need for global persist IDs for data and samples to facilitate reuse.
What public policies does the community require to share data effectively?
IP issues. Is there a way to warn researchers about patent and licensing issues when they reuse data?
What's the value proposition for sharing data? How do we decide how much to invest to share it?
Dual-use
Train scientist to "read" and use biological data bases (data science training)

**Table 6**

If you could instantaneously have one thing happen, be created or realized to solve your biological data sharing needs, what would it be? Why?

AI based assistant with domain experience that can provide smart search of repositories and provide useful recommendations on “accurate” data sources based on crowd sourcing (e.g. up/down voting of sources, international policy).
Automated conversion of collected data into a format that is widely shareable. A universal translator for biological data.
Agreed upon metadata standards for the various communities, and a user-friendly way to access those standards for nondata scientists.
Incentives for data sharing, including funding for all efforts needed to get the data into a shareable format
Automated deidentification of medical data to deal with HIPAA issues of privacy
Universal biological data sharing model that incorporates all metadata including methods, essentially what would you need to know about the dataset that is needed to make it useful for various analyses.
Centralized long term repository with shared IT but separate domain-specific curation teams