Andrew Quinn

LIS 545

Winter 23

Repository Profile

I've chosen the Biological General Repository for Interaction Datasets (BioGRID), after searching through the results from a generic search I found a few that did not provide the information I was looking for, but BioGRID did. Their interface is smooth and the language they used throughout their site did not feel like a computer behind it but people who were dedicated to gathering information. There were displays that provided helpful information to aid in my search throughout the website and seemed very accessible and the language they used was easy to understand. In addition, I also watched their two tutorial YouTube videos on how to navigate their services.

The repository appeared to be open for deposits from anyone who has published work, however, they define their scope as disciplinary and in that regard, deposits must be related to biomedical research. They accept deposits in Microsoft Excel or plain text files. Also if the publication is not made publicly available yet, they are willing to work with potential users to add their data to their repository. It's fair to assume that the repository is closed to general submissions that are not related to biomedical research due to the team behind BioGRID being accessible and willing to support the user.

In terms of collection scope, their current index contains "81,725 publications for 2,593,556 protein and genetic interactions, 30,725 chemical interactions and 1,128,339 post translational modifications from major model organism species." However, regarding the eligibility of who is able to contribute data, there was no language stating

clearly who is and who isn't. The only requirement that I came across is that the data is related to biomedical research.

Regarding what data the repository accepts, the process seemed very straightforward. It's as simple as contacting them through email with your published data. In their contribution page they do state that they would like the data sent as a Microsoft Excel file or a Plain Text File. When I continued to search for more information related to what they accept I found a data submission form that required information and description for InteractorsA/B, InteractionA/B taxon, Evidence Code, Pubmed ID, Score, Modification, Phenotypes, and Comment/Qualifier. In terms of data type, they accept data, metadata, models, and statistics.

After reviewing their website, I did not find any information regarding a submission information package (SIP). The level of human assistance that was provided by BioGRID was what led me to choose their repository. Throughout their website, they preferred that the user email them directly for deposit and correcting mistakes in curation. They did not use any consulting or digital assistance to submitters. In terms of metadata submission specific structure, there was no language stating they do not accept it or require a specific standard, however in the data submission form they require the user to identify any interactor identifier type which I believe would be metadata.

There was no login required to access the data repository and the data was fully accessible and available to academic and commercial users under the MIT License and without any warranty. However, I did find a login option without the option of creating a

profile. I believe this function is only available to contributors after they've established a relationship with BioGRID.

The BioGRID download file repository interface was very easy to operate. The display window aside the search list had helpful information on how to navigate through the data. All the data files were available through zip formats but clicking further details I found the recommended file formats with many options such as PSI-MI XML 2.5 Formatted, PSI MITAB Version 2.5, Osprey Custom Network 1.3.1, and others. In my further assessment I did not come across any specific language relating to metadata standards nor did I find a dissemination information package" (DIP).

Overall, I found this exercise very helpful and feel fortunate to have picked BioGRID because of their level of human interaction. I find it very likely if I were to reach out to them about this assignment that I would receive a response in a reasonable amount of time. I did find their repository's policies and procedures did not use the same terms as open archival information system (OAIS) reference model, I did have to search through their website to find terms that were related to the OAIS but weren't explicit. In terms of submissions, it was open to anyone who has published data and related to biomedical research. In addition to the mechanism of access to the repository's data, there was no login required to access the data and found navigating the data easy with the help display next to the list of data to select from. I found no structure standard or access requirement to metadata, yet found in data submission form that they asked for identifiers that could be later used as metadata. To conclude, I found this assignment useful in understanding repositories.