

CURATING EXPERIMENTS AND RESULTING DATA FOR A DISEASE FOUNDATION

Gully APC Burns¹, Michael Rogan², Alan Ruttenberg³

¹ *Intelligent Systems Division, USC / Information Sciences Institute*

² *Michael J. Fox Foundation for Parkinson's Research; Program for Imaging and Cognitive Sciences, Columbia University*

³ *Principal Scientist, Neurocommons Project, ScienceCommons*

Funding agencies are highly motivated to develop practical bioinformatics methods for the study of a specific phenomenon or disease, and appreciate the need to extract value from data compiled from the projects that they support. By providing the funding that enables research, they are also uniquely placed to enforce data sharing policies that require experimentalists to provide access to data and knowledge. Additionally, they themselves have an interest in accumulating and synthesizing knowledge across independent research projects that may pertain to their mission. They are therefore both a potentially influential and powerful partner to bioinformatics developers seeking to build infrastructure that can enable this ambitious goal.

Here we describe the preliminary design of an informatics and process framework that can be used by a disease foundation to gather data from funded research proposals, along with information that describes that data. This framework combines the KE-f-ED (Knowledge Engineering from Experimental Design) model, a model which uses the relationships between independent and dependent variables in an experiment to organize a description of it, and terminology and curation process from OBI (The Ontology for Biomedical Investigations), a cross-discipline effort to build an ontology for describing experiments across the breadth of biological experimental methods.

The system proposed lets one gather understandable descriptions of experimental protocols and then uses these descriptions to generate simple spreadsheets to record the observations and measurements made. These spreadsheets may be populated by grantee scientists and then validated, stored in, and queried from a foundation's data repository.

We describe a model of a disease foundation's interaction with their grantees and provide a worked example of how data obtained from several different experiments might be collected, curated, and made available from such a system.