

bioCADDIE: Progress and Next Steps

Lucila Ohno-Machado, University of California, San Diego; Jeff Grethe, University of California, San Diego; Ian Fore, National Institutes of Health; Susanna-Assunta Sansone, University of Oxford; George Alter, University of Michigan; Hua Xu, University of Texas

The bioCADDIE Data Discovery Index Consortium project started effectively in March 2015. It has three goals: help users find data, build a data discovery index, and interoperate with other entities in the NIH Commons. During the past 19 months, the recommendations of working groups, metadata specifications, software development, and pilot projects helped develop the first prototype of the DataMed search engine, available at datamed.org. We will report on our various engagement activities, summarize the collaborations with BD2K initiatives and other groups, and explain our vision and future directions for bioCADDIE.

Our prototype development involves work in several specific areas, including metadata management, metadata mapping, user-interface design and backend database development, which form the core development team. Additionally, working groups within bioCADDIE address specific needs within the consortium. Currently active working groups focus on Evaluation and on Outreach, among others. Together these groups worked together to release the DataMed v1.0 prototype, which went live on June 30, 2016 and contains 23 indexed repositories. Version 1.5 of DataMed will be available in late November, 2016 and will include an additional 18 repositories.

bioCADDIE also has many external collaborations, including with centers with an important digital object indexing interest, like Heart2BD2K, CEDAR, LINCS, Force11, and ELIXIR, among others. These collaborations serve an important role by allowing dedicated efforts towards sharing of ideas and goals, as well as cross-project integration of tools and resources.

Future developments for bioCADDIE include ingestion of many more repositories, and version 2.0 of DataMed release in February 2017. The currently ongoing Dataset Retrieval Challenge, which has 28 groups participating, will grant at least one group an opportunity to have their methods integrated into the DataMed tool. We will fund additional pilot project requests for applications to continue to connect with the community at large.