SciCrunch: A cooperative and collaborative data and resource discovery platform for scientific communities

Patricia L. Whetzel, J.S. Grethe, A. Bandrowski, D. Banks, C. Condit, A. Gupta, S. D. Larson, Y. Li, B. Ozyurt, A. Stagg, M.E. Martone
University of California San Diego
La Jolla, CA

L. Marenco, P. L. Miller, R. Wang, G. M. Shepherd
Yale University
New Haven, CT

***Abstract*—SciCrunch was designed to break down the traditional types of portal silos created by different communities and to enable re-use of annotations and curation efforts across different biomedical domains. The SciCrunch platform (SciCrunch.org) can enable a research community to easily develop their own community-specific resource portal to provide access to resources, such as databases, materials, and tools, relevant to their research area. Communities can select existing resources to include in their portal and also add new resources through the data ingestion pipeline. The SciCrunch platform uses a set of ontologies to enable data integration and semantic expansion of user queries to enhance search results across these disparate resources. Data within a portal can be accessed via the portal Web interface or programmatically via Web services. The SciCrunch platform provides data used in the Resource Identification Portal, the NIDDK Information Network, the Neuroscience Information Framework, the NCBO Resource Index, and the Monarch Initiative.**

Keywords— semantic web; neuroscience; diabetes; genotype; phenotype; resource identification

I. INTRODUCTION

Initiating a research project requires finding relevant data and resources. However, searching for these research resources on the web is difficult since Google searches do not return information within databases and learning the nuances of search for each individual database is time consuming. To address this need our group has developed SciCrunch, which provides a common, single platform for semantically enhanced search across databases. The development of SciCrunch is based on past work of developing the Neuroscience Information Framework.

II. SCICRUNCH COMPONENTS

The SciCrunch platform provides a dynamic inventory of Web-based resources, such as data, materials, and tools. The platform includes a Web portal, resource Registry, ontology, and Web services.

A. Web portal

SciCrunch provides a default web portal, which can be cusomtized in both data content and look and feel. The SciCrunch Web portal allows researchers to search for resources of interest integrated via the semantics of the ontology. Each individual community can select which resources to include in their portal and customize the look-and-feel of their portal through the modular architecture of SciCrunch. Resources within SciCrunch can be shared amongst different communities; therefore the same annotated content is accessible to researchers across different domains.

B. Resource Registry

The SciCrunch Registry is a catalog of all registered resources, currently over 12,000. A resource registry entry includes a general description of the resource, a summary of domains the resource covers, and it's source of funding. The metadata used to describe a resource has been harmonized with other projects such as BioSharing and will be re-evaluated as the HCLS Dataset descriptor recommendataion is finalized.

C. Ontology

The ontology drives the autcomplete selections provided in the search box on the Web portal as well as which results are returned based on the details of a concept, e.g. synonyms and parent classes. The SciCrunch platform enables community specific ontology content to be incorporated within a portal. This is currently enabled by use of separate instances of the ontology data store, for example as used with the Monarch Initiative. However, other projects such as the NIDDK Information Nework (dkNET) and the Resource Identification Portal use the same ontology, NIFSTD. While the NIFSTD was developed for the Neuroscience Information Framework and contains neuroscience-specific extensions, it is also broad enough to cover a wide range of life science domains.

D. Web Services

The Web services power the SciCrunch Web portal and provide access to the ontology-based index of data. These Web services can be used by third party applications or Web sites as with the Monarch Initiative.

III. SUMMARY

This demo will provide an overview of the SciCrunch components and demonstrate use of the ontology-based index with queries to the NIDDK Information Network demonstrating use cases of the portal.