

BioThings Explorer: Cross-Linking Biological APIs for Knowledge Exploration

Jiwen Xin, Cyrus Afrasiabi, Sebastien Lelong, Ginger Tsueng, Michel Dumontier, Andrew I. Su, Chunlei Wu

Current biological research relies significantly on the integration and cross-linking of data across multiple domains, including genes, drugs and diseases. Previous biological data integration efforts mainly focused on storing data from different resources in a local centralized repository, which requires huge efforts in terms of data parsing, merging and maintaining. Meanwhile, there has been a trend that data providers start to use web-based APIs as an alternative data access point. And here, we propose to integrate different biological data together through a distributed approach, which is to interlink those individually maintained APIs for integrated knowledge exploration.

The key challenge in interlinking APIs is data interoperability issue, which comes from the heterogeneity of biological entities and variety of data models. BioThings Explorer solves this by utilizing the SmartAPI-based API-specific metadata and JSON-LD based semantic context. SmartAPI specifications are created for each API and fully exploited to generate API calls programmatically as well as semantically align API inputs and outputs. In addition, the more sophisticated semantics (e.g. relationships) can be encoded in a separate JSON-LD context file to further annotate the API output.

By utilizing all these technologies, we are able to build a highly connected network of biological APIs, which are discoverable, reusable and actionable. It currently integrates 13 different API resources, covering over 60 unique biological concepts. Additionally, we have developed a D3.js and Cytoscape.js based web interface (as well as a Python client) at <http://biothings.io/explorer/>, which allows users to search and visualize the annotated APIs. More importantly, it empowers users to perform knowledge exploration tasks by identifying potential connecting paths and chaining multiple API calls together. We believe that our approach abstracts out the data-wrangling efforts from the developers so that they can focus more on the downstream data-integration and analysis.