

BioThings APIs: Linked High-Performance APIs for Biological Entities

Chunlei Wu, The Scripps Research Institute; Jiwen Xin, The Scripps Research Institute; Cyrus Afrasiabi, The Scripps Research Institute; Sebastien Lelong, The Scripps Research Institute; Ginger Tsueng, The Scripps Research Institute; Andrew I. Su, The Scripps Research Institute

The accumulation of biological knowledge and the advance of web and cloud technology are growing in parallel. Recently, many biological data providers start to provide web-based APIs (Application Programming Interfaces) for accessing data in a simple and reliable manner, in addition to the traditional raw flat-file downloads. Web APIs provide many benefits over traditional file downloads. For instance, users can request specific data such as a list of genes of interest without having to download the entire dataset, thereby providing the latest data on demand and reducing computation and data transfer times. This means that programmers can spend less time on wrangling data, and more time on analysis and discovery.

Building and deploying scalable and high-performance web APIs requires sophisticated software engineering techniques. We previously developed high-performance and scalable web APIs for gene and genetic variant annotations, accessible at MyGene.info and MyVariant.info. These two services are a tangible implementation of our expertise and collectively serve over 4 million requests every month from thousands of unique users. Crucially, the underlying design and implementation of these systems are in fact not specific to genes or variants, but rather can be easily adapted to other biomedical data types such drugs, diseases, pathways, species, genomes, domains and interactions. We are currently expanding the scope of our platform to other biological entities. Collectively, we refer them as “BioThings APIs” (<http://biothings.io>).

We also applied JSON-LD (JSON for Linking Data) technology in the development of BioThings APIs. JSON-LD provides a standard way to add semantic context to the existing JSON data structure, for the purpose of enhancing the interoperability between APIs. We have demonstrated the applications of JSON-LD with BioThings APIs, including data discrepancy checks as well as the cross-linking between APIs.