

# Harvesting service descriptions for bio.tools using OpenAPI standards

The ELIXIR Tools and Data Services Registry<sup>1</sup> (bio.tools) aims to provide comprehensive coverage of tools, data services (APIs) and host databases. To fulfill the registry mission - to boost tool and therefore data utilisation - it's essential to adequately catalogue the services built upon the primary data resources. Broadly, these services are of one of three types: 1) simple URL-based / RESTful APIs, 2) Web Services based on SOAP+WSDL and 3) Web applications. This proposal targets the first category - the prevalent means for programmatic data access today - and will address a gap in bio.tools content: bio.tools currently includes very few (~163) such services, and does not provide details of service endpoints.

OpenAPI, formerly known as Swagger, has been adopted by life science databases to document their URL-based and RESTful or REST-like APIs. The OpenAPI specification describes how web services can be annotated to such a level of detail that interactive documentation can be generated from it. Example databases<sup>2</sup> that use OpenAPI include PRIDE, UniProt (partly) WikiPathways, Open PHACTS, eNanoMapper, and BridgeDb. However, currently the annotation does not make use of ontologies, and key details about the service as a whole and service endpoints are described in free text. For a service to be registered in bio.tools, certain semantic annotations are mandatory, as defined in the biotoolsXSD<sup>3</sup> schema and defined in terms from the EDAM ontology<sup>4</sup>. In the case of APIs the annotations are the general scientific domain (EDAM Topic) the service as a whole serves, and the output type of data (EDAM Data) and data format (EDAM Format) for a given service endpoint. It is also possible to annotate the data identifier (EDAM Data→Identifier) used in a data access URL and the specific function (EDAM Function), usually defaulting to "Data retrieval".

This project will develop a utility to convert OpenAPI configuration files to a format suitable for import into bio.tools. This must allow OpenAPI-documented services to be easily and automatically be included in the registry. The primary goal is to develop the utility and apply it to at least six existing services, including those listed above. These six OpenAPI endpoints alone define more than 200 API methods, covering various concepts, including patents, proteins, genes, pathways, nanomaterials, compounds, metabolites, and reactions. Both these concepts and the transformations encoded by those API methods need semantic annotation to make them FAIR<sup>5</sup>. Therefore, a secondary goal is to extend the OpenAPI convertor with fields bringing the semantic annotation up to the minimum ELIXIR information standard for tools, as defined by biotoolsXSD, and write OpenAPI descriptions to this standard, that can serve as exemplars for others.

The expected project output include an open source (OSI-compatible) convertor utility that translates a OpenAPI configuration file into a bio.tools JSON / XML file and five OpenAPI-based services included in bio.tools (Ensembl, PRIDE, eNanoMapper, Open PHACTS, WikiPathways, and BridgeDb). Output of secondary importance include a document describing a proposal for a "semantically enhanced OpenAPI grammar" to enable the description of REST web services with all the metadata that is mandatory for bio.tools, and a tutorial for web service maintainers for the creation of the OpenAPI files and how to convert these bio.tools JSON / XML files and register them.

All supervision will be in English and by Dr. Egon Willighagen at the Maastricht University, Hervé Ménager at the Institut Pasteur, and Jon Ison at DTU, Denmark.

Contact: Egon Willighagen ([egon.willighagen@maastrichtuniversity.nl](mailto:egon.willighagen@maastrichtuniversity.nl)) cc Hervé Ménager ([hmenager@pasteur.fr](mailto:hmenager@pasteur.fr)) and Jon Ison ([jison@cbs.dtu.dk](mailto:jison@cbs.dtu.dk)).

---

<sup>1</sup> Ison, J. *et al.*, 2015, Tools and data services registry: a community effort to document bioinformatics resources. *Nucleic Acids Research*, 44(D1): 38-47. doi: 10.1093/nar/gkv1116

<sup>2</sup> For a (probably not comprehensive) list of OpenAPI-documented databases: <http://tinyurl.com/hb3mwqc>.

<sup>3</sup> <https://github.com/bio-tools/biotoolsxsd>

<sup>4</sup> Ison, J. *et al.*, 2013, EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics*, 29(10): 1325-1332. doi: 10.1093/bioinformatics/btt113

<sup>5</sup> Wilkinson, M. *et al.*, The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, doi: 10.1038/sdata.2016.18