TrackFind: FAIR Search of Genomic Tracks

Dmytro Titov¹, Sveinung Gundersen¹, Radmila Kompova¹, Salvador Capella-Gutierrez², Finn Drabløs³, José M. Fernández², Kieron Taylor⁴, Daniel Zerbino⁴, Eivind Hovig¹,⁵

1. Center for Bioinformatics, University of Oslo (UiO), Norway
2. INB Coordination Node / ELIXIR ES, Barcelona Supercomputing Center, Spain
3. Norwegian University of Science and Technology (NTNU), Norway
4. European Molecular Biology Laboratory, European Bioinformatics Institute, United Kingdom
5. Department of Tumor biology, Institute for Cancer Research, Oslo University Hospital (OUH), Norway

Thousands of genomic annotation tracks have been generated the recent years, many in the context of larger undertakings such as BLUEPRINT and ENCODE. Several data portals for tracks are providing search services to researchers, but the underlying metadata are diverse and often poorly curated. The Trackhub Registry provides a unified access point, but currently only supports limited search capabilities. Also, solutions for importing track data to the Galaxy framework based on curated metadata are currently limited.

In the context of the Elixir Implementation Study: "FAIRification of Genomic Tracks", we have developed the TrackFind service (http://trackfind-dev.gtrack.no). TrackFind supports crawling of the TrackHub Registry and other data portals to fetch track metadata. Crawled metadata can be accessed through hierarchical browsing or by search queries, both through a web-based user interface, and as a REST API. TrackFind supports advanced SQL-based search queries that can be easily built in the user interface, and the search results can be browsed and exported in JSON or GSuite format (http://www.gtrack.no). The REST API (https://apidocs.trackfind-dev.gtrack.no/) allows downstream tools and scripts to easily integrate TrackFind search, currently demonstrated by the GSuite HyperBrowser and later EPICO analysis frameworks. The GSuite HyperBrowser (https://hyperbrowser.uio.no/trackfind_test) demonstrates a Galaxy-based TrackFind client tool which outputs search results in the GSuite format in order to convey track metadata all the way through data manipulation to the outputs of analysis tools.

In addition to supporting most metadata models directly, TrackFind also supports the transformation of metadata into the FAIR model defined in the "FAIRification of Genomic Tracks" Implementation Study. Such transformation can be achieved on per-TrackHub basis through online scripting, thus providing a simple path for data managers to FAIRify their track metadata. TrackFind also maintains a version history of all metadata changes, including all recrawlings and transformations. We are also planning to add functionality for curating existing track metadata content.

We believe the TrackFind track search engine and metadata FAIRification service to be a major contribution, both to maintainers of genomic annotation track data, as well as to researchers and tool developers interested in making use of the wealth of track data publicly available – within Galaxy, as well as on the command line and in other analysis contexts.