









OPINION ARTICLE

Recommendations for the FAIRification of genomic track metadata [version 1; peer review: 2 approved]

Sveinung Gundersen ¹, Sanjay Boddu², Salvador Capella-Gutierrez ³, Finn Drabløs ⁴, José M. Fernández ³, Radmila Kompova¹, Kieron Taylor ², Dmytro Titov¹, Daniel Zerbino², Eivind Hovig ^{1,5}

¹Center for Bioinformatics, University of Oslo (UiO), Oslo, Norway

²European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK

³Life Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain

⁴Department of Clinical and Molecular Medicine, NTNU – Norwegian University of Science and Technology, Trondheim, Norway

⁵Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital (OUH), Oslo, Norway

V1 First published: 01 Apr 2021, 10(ELIXIR):268
<https://doi.org/10.12688/f1000research.28449.1>

Latest published: 01 Apr 2021, 10(ELIXIR):268
<https://doi.org/10.12688/f1000research.28449.1>

Abstract

Background: Many types of data from genomic analyses can be represented as genomic tracks, *i.e.* features linked to the genomic coordinates of a reference genome. Examples of such data are epigenetic DNA methylation data, ChIP-seq peaks, germline or somatic DNA variants, as well as RNA-seq expression levels.

Researchers often face difficulties in locating, accessing and combining relevant tracks from external sources, as well as locating the raw data, reducing the value of the generated information.

Description of work: We propose to advance the application of FAIR data principles (Findable, Accessible, Interoperable, and Reusable) to produce searchable metadata for genomic tracks. Findability and Accessibility of metadata can then be ensured by a track search service that integrates globally identifiable metadata from various track hubs in the Track Hub Registry and other relevant repositories. Interoperability and Reusability need to be ensured by the specification and implementation of a basic set of recommendations for metadata. We have tested this concept by developing such a specification in a JSON Schema, called FAIRtracks, and have integrated it into a novel track search service, called TrackFind. We demonstrate practical usage by importing datasets through TrackFind into existing examples of relevant analytical tools for genomic tracks: EPICO and the GSuite HyperBrowser.

Conclusion: We here provide a first iteration of a draft standard for genomic track metadata, as well as the accompanying software ecosystem. It can easily be adapted or extended to future needs of the research community regarding data, methods and tools, balancing the requirements of both data submitters and analytical end-users.

Open Peer Review

Reviewer Status  

Invited Reviewers

1

2

version 1


01 Apr 2021



report



report

1. **Max Haeussler** , University of California Santa Cruz, Santa Cruz, USA

2. **Granger Sutton** , J. Craig Venter Institute, Rockville, USA

Any reports and responses or comments on the article can be found at the end of the article.