

Conference Abstract

Streamlined Conversion of Omics Metadata into Manuscript Facilitates Publishing and Reuse of Omics Data

Mariya Dimitrova^{‡,§}, Raïssa Meyer^l, Pier Luigi Buttigieg^l, Teodor Georgiev[‡], Georgi Zhelezov[¶], Seyhan Demirov[‡], Vincent S. Smith[#], Lyubomir Penev^{‡,▪}

[‡] Pensoft Publishers, Sofia, Bulgaria

[§] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria

^l Alfred Wegener Institute for Polar and Marine Research, Bremen, Germany

[¶] Pensoft Publishers, Sofia, Bulgaria

[#] The Natural History Museum, London, London, United Kingdom

[▪] Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Sofia, Bulgaria

Corresponding author: Mariya Dimitrova (m.dimitrova@pensoft.net)

Received: 28 Sep 2020 | Published: 28 Sep 2020

Citation: Dimitrova M, Meyer R, Buttigieg PL, Georgiev T, Zhelezov G, Demirov S, Smith VS, Penev L (2020) Streamlined Conversion of Omics Metadata into Manuscript Facilitates Publishing and Reuse of Omics Data. Biodiversity Information Science and Standards 4: e59041. <https://doi.org/10.3897/biss.4.59041>

Abstract

Data papers have started to gain popularity as a publishing format that allows easy and quick publishing of research data (Chavan and Penev 2011, Penev et al. 2017). They describe single or multiple datasets and the methodologies required for their generation. Similar to traditional research articles, data papers and the underlying datasets are peer-reviewed. In this poster, we demonstrate how data papers can be used to incentivise researchers producing omics datasets to increase the quality of the metadata descriptors and the data itself through the journal authoring, peer review and publication process, thus improving data visibility, discoverability, sharing and reuse.

We illustrate a highly automated workflow for the creation of omics data paper manuscripts, which started with the development of a template for this specific article type in the [Biodiversity Data Journal](#) (BDJ), published by Pensoft (Dimitrova et al. 2020). The workflow streamlines automatic conversion and import of metadata from the [European](#)

[Nucleotide Archive \(ENA\)](#) into an omics data paper manuscript created in the [ARPHA Writing Tool](#) (AWT), following a three step procedure:

1. mapping of the [European Nucleotide Archive \(ENA\)](#) metadata to the manuscript sections,
2. extraction of the relevant metadata through the ENA project or study ID, and
3. transforming the metadata into HTML or XML files. The XML file follows the [Journal Article Tag Suite \(JATS\)](#) standard and can be used by anyone as a draft to further develop a data paper manuscript and submit it to any journal.

Records in ENA sometimes have linked data in the [ArrayExpress](#) and [BioSamples](#) databases, which describe sequencing experiments and samples following the community-accepted metadata standards [MINSEQE](#) and [MixS](#). The workflow also retrieves such records and inserts them both into the omics data paper narrative and as supplementary data files.

The workflow has been integrated with Pensoft's ARPHA platform but the conversion code is openly accessible on [GitHub](#) under the Apache 2.0 license and can be run as a [R Shiny app](#). By openly providing access to the code and its implementation in a web application, we enable the full reproducibility of the streamlined import of ENA metadata into an omics data paper manuscript. The plan is to further develop the workflow to include the import of various other types of omics data and omics data repositories in addition to the currently supported ENA genomic data. The workflow reaffirms the important role of high-quality metadata for creating extended dataset descriptions, recognised by Chavan and Penev 2011. Conversion of metadata into a manuscript helped us discover many datasets with insufficient or inaccurate metadata. Hence, we hope that our workflow promotes not only omics data paper publishing but also better metadata authoring and curation.

Keywords

European Nucleotide Archive, data paper, FAIR data

Presenting author

Mariya Dimitrova

Presented at

TDWG 2020

Funding program

This research has received partial funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764840.

References

- Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12 <https://doi.org/10.1186/1471-2105-12-s15-s2>
- Dimitrova M, Meyer R, Buttigieg PL, Penev L (2020) Streamlined import of omics metadata from the European Nucleotide Archive (ENA) into an OMICS Data Paper manuscript. <https://blog.pensoft.net/2020/06/16/streamlined-import-of-omics-metadata-from-the-european-nucleotide-archive-ena-into-an-omics-data-paper-manuscript/>. Accessed on: 2020-8-11.
- Penev L, Mitchen D, Chavan V, Hagedorn G, Smith V, Shotton D, Ó Tuama É, Senderov V, Georgiev T, Stoev P, Groom Q, Remsen D, Edmunds S (2017) Strategies and guidelines for scholarly publishing of biodiversity data. *Research Ideas and Outcomes* 3 <https://doi.org/10.3897/rio.3.e12431>