



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



Mariya Dimitrova^{1,2}, Raïssa Meyer³, Pier Luigi Buttigieg³, Teodor Georgiev¹, Georgi Zhelezov¹, Seyhan Demirov¹, Vincent S. Smith⁴, Lyubomir Penev^{1,5}



1 Pensoft Publishers, Sofia, Bulgaria



2 Institute of Information and Communication Technologies, BAS, Sofia, Bulgaria



3 Alfred Wegener Institute for Polar and Marine Research, Bremen, Germany



4 The Natural History Museum, London, UK



5 Institute of Biodiversity and Ecology Research, BAS, Sofia, Bulgaria

Background

- More and more data are being generated in omics experiments.
- Stored in private databases, files and public repositories.
- Datasets often remain inaccessible or difficult to find unless they are mentioned in a publication.
- Data papers provide a solution by enabling the description of datasets without an outline of a full research experiment.

Aims

- To improve the **FAIRness of omics data** by providing a mechanism for rapid and easy omics data paper publishing.
- To incentivise authors and curators to create **better quality metadata**.

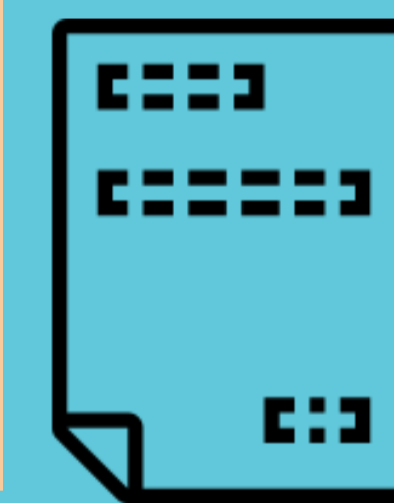
Methods

Created a **template** for an omics data paper based on the metadata available in the **European Nucleotide Archive (ENA)**, **ArrayExpress** and **Biosample** databases, the last two describing sequencing experiments and samples following the community-accepted metadata standards **MINSEQE** and **MixS**.

Identified relevant Xpaths for obtaining the metadata from the databases and developed a **workflow** for extraction of all relevant metadata via input of a single **ENA Study or Project ID** in an **R Shiny app**.

Enabled conversion of extracted omics metadata into **HTML** and **XML**, following the **Journal Article Tag Suite (JATS)** standard.

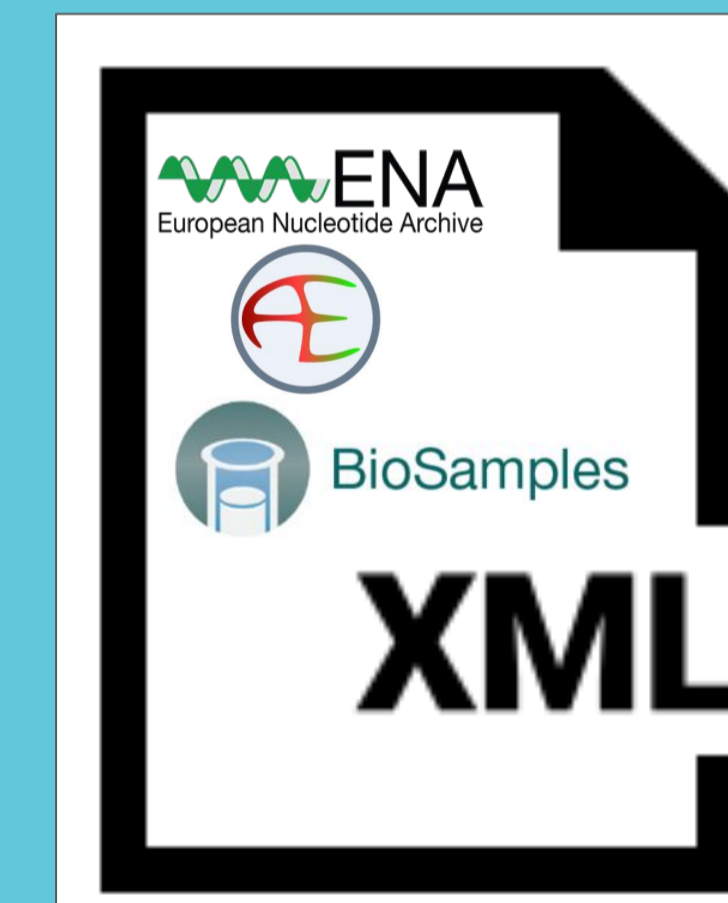
Integrated the workflow with **Pensoft's ARPHA Writing Tool** as a new mechanism for import of metadata into manuscript.



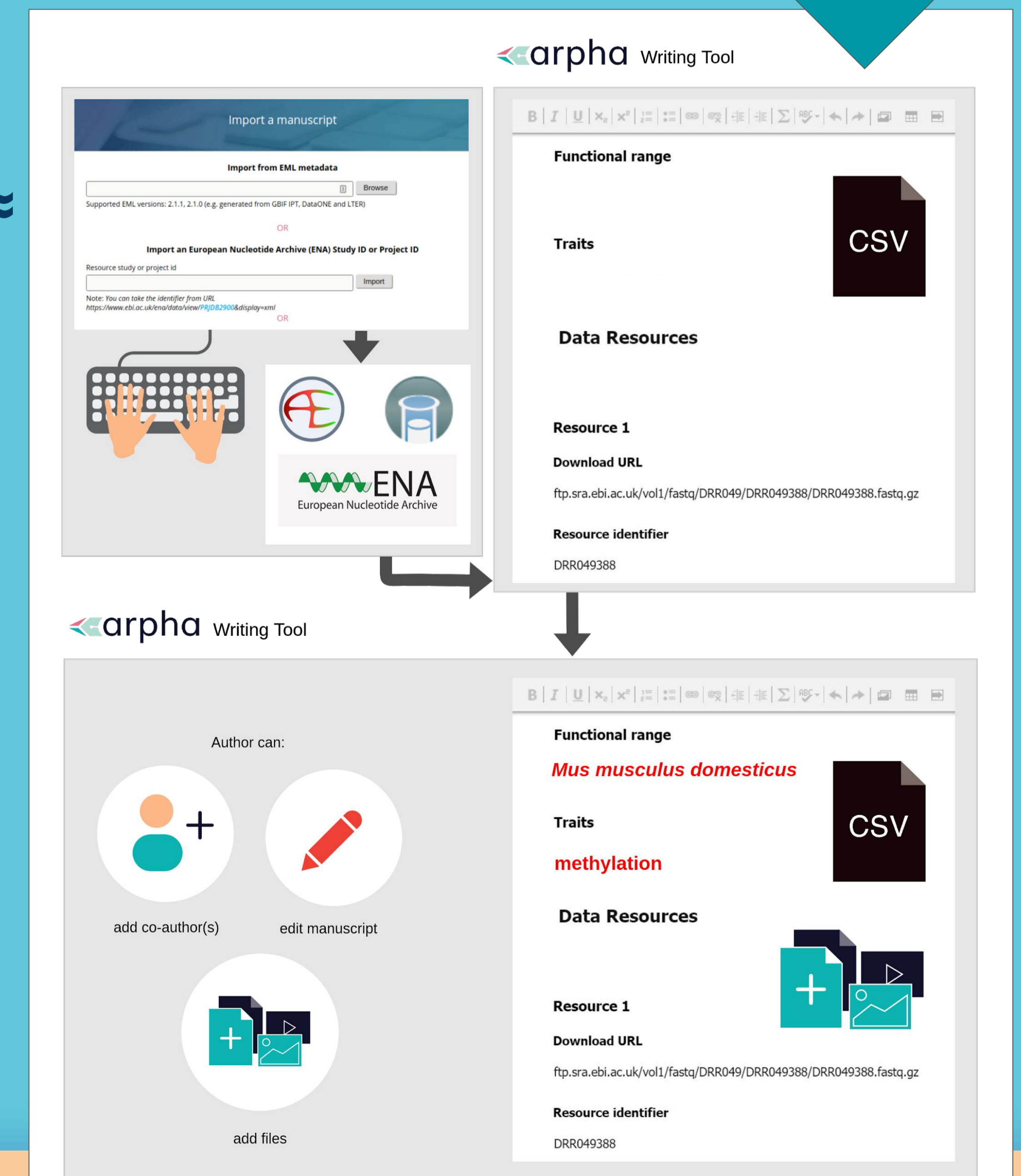
+



=



≈



Why publish data papers?

- ★ Increased visibility and discoverability of your datasets.
- ★ Establishing priority and getting credit for your work.
- ★ Improves reproducibility and promotes data-driven discoveries and collaborations.
- ★ Peer-review of data papers and quality checks of datasets performed by data auditors at Pensoft ensures that data is of high-quality.

References

- Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12
- 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.
- Penev L, Mietchen 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



Funding

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.



Fig. 1 Generation of omics data paper manuscripts at the click of a button