Curathon: A step towards community-driven curation

Francesco Talo' (1), Johanna McEntyre (2), Aravind Venkatesan (3), and Mariia Levchenko (4)

(1) EMBL-EBI, Orcid 0000-0002-5306-4329, (2) EMBL-EBI, Orcid 0000-0002-1611-6935, (3) EMBL-EBI, Orcid 0000-0003-4019-1940, (4) EMBL-EBI, 0000-0003-0706-515X

Europe PMC (www.europepmc.org) is a database of biomedical research literature, hosted at the European Bioinformatics Institute (EMBL-EBI). The database provides a single access point to PubMed (www.ncbi.nlm.nih.gov/pubmed/) abstracts and PubMed Central (www.ncbi.nlm.nih.gov/pmc/) full text articles, as well as 5 million additional records. Europe PMC serves a wide user base, including researchers and database curators, by delivering novel and innovative tools that support scientific workflows.

To bridge the gap between literature and data Europe PMC runs a community annotations platform. The platform accepts annotations provided by a number of text mining groups, making them available to users via API and an application called SciLite (www.wellcomeopenresearch.org/articles/1-25/v2), that highlights the annotations directly in the text. The goal of this annotation system is to expose text-mining outputs from the community in useful ways for a wide audience of researchers and curators, as well as other interested stakeholders. The open nature and standard format (https://www.w3.org/TR/annotation-model/) of the annotations means that they can be mapped to any platform displaying the same content. Highlighting a variety of biological entities, such as genes/proteins or organisms, phrases, or relationships (e.g. gene-disease associations), can assist skim-reading of articles, providing a basis to support database curation processes, as well as a mechanism to create deep links between literature and data for clear evidence of curatorial statements.

For the Force2017 conference, we propose to conduct a "Curathon" that will focus on the needs of the curation community. This interactive session will allow participants to identify common pitfalls and pitch ideas around the topics of sharing and validating automated annotations, as well as providing feedback to the system as a means to improve annotations accuracy.