BIOTRANS
LA ROCHELLE 2023

N°1449 / PC
TOPIC(s) : Artifical intelligence / computational methods / Enzyme production, immobilization

The BioCatHub project, applied research data management in biocatalysis

AUTHORS

Stephan MALZACHER / FORSCHUNGSZENTRUM JÜLICH, WILHELM-JOHNEN-STRASSE, JÜLICH
Patrizia WARMELINK / FORSCHUNGSZENTRUM JÜLICH, WILHELM-JOHNEN-STRASSE, JÜLICH
Jens LOHMANN / FORSCHUNGSZENTRUM JÜLICH, WILHELM-JOHNEN-STRASSE, JÜLICH
Felix ENGELMANN / FORSCHUNGSZENTRUM JÜLICH, WILHELM-JOHNEN-STRASSE, JÜLICH
William FINNIGAN / UNIVERSITY OF MANCHESTER, OXFORD RD, MANCHESTER
Nicholas TURNER / UNIVERSITY OF MANCHESTER, OXFORD RD, MANCHESTER
Alaric PRINS / CAPE PENINSULA UNIVERSITY OF TECHNOLOGY, SYMPHONY WAY, BELLVILLE
Marilize LE ROES-HILL / CAPE PENINSULA UNIVERSITY OF TECHNOLOGY, SYMPHONY WAY, BELLVILLE
Jürgen PLEISS / UNIVERSITY OF STUTTGART, ALLMANDRING 31, STUTTGART
Jan RANGE / UNIVERSITY OF STUTTGART, ALLMANDRING 31, STUTTGART
Dörte ROTHER / RWTH AACHEN UNIVERSITY, WORRINGER WEG 1, AACHEN

PURPOSE OF THE ABSTRACT

Toxic waste degradation, the (asymmetric) synthesis of active pharmaceutical ingredients or the carbon-neutral synthesis of chemicals are just some examples in which enzymes are expected to play a key role in the years to come. Publicly available experimental data, produced in scientific processes, is one of the most important assets to realise this. It serves as the foundation to reproduce published experimental results, or as the basis for the development and reviewing of theoretical models. However, to be able to fulfil this role, the produced experimental data must be comprehensible, interoperable, and publicly available, which has often not worked in the past, and was even described by its own term: the reproducibility crisis. To counteract the reproducibility crisis, in 2016, the Findable, Accessible, Interoperable and Reusable (FAIR) data principles [1] were published, providing guiding principles in how to collect, document and publish experimental data. Nonetheless, the implementation of research data management (RDM) based on the FAIR data principles is time demanding, requires in-depth technical knowledge, and therefore often exceeds the capacities of the already-busy laboratory staff.

For this purpose, we have developed the BioCatHub project. It aims to provide a RDM infrastructure to solve highly independent and individual RDM challenges in biocatalysis. The central component of the BioCatHub project is the BioCatHub platform. It is created to help experimenters to realise standardised data acquisition based on the FAIR data principles in biocatalysis with four central concepts. First, the BioCatHub platform (retrobiohub.org) works as data collection platform and provides a user interface, helping experimenters to describe experimental procedures in a wizard-like structure. Secondly, the BioCatHub platform contains data integration functions to query online databases like SabioRK, Brenda and autocomplete entries for the experimenters during the data collection process. Third, the automated communication between RetroBioCat [2] and the BioCatHub platform enables the planning and seamless integration of data from retrosynthetic reactions. Lastly, the BioCatHub platform writes and reads collected data to and from the machine-readable file format Enzyme markup language (EnzymeML) [3, 4]. This enables the possibility to collect, share, store or publish information about biocatalytic experiments, for example on Zenodo.

On top, for data analysis and processing purposes, we have created the BioCatHub data processing repository, enabling the development of customised and reusable data processing workflows, using data previously collected in the BioCatHub platform as EnzymeML documents. For example, the export of the collected data into a spreadsheet for an experimenter friendly data analysis, the automated estimation of Michaelis-Menten parameters

or the description of biocatalytic data using a first order reaction were already implemented as workflows.

The vision realised with the BioCatHub project is to create an extendable infrastructure for researchers in biocatalysis to perform standardised data collection, storage and publishing based on the FAIR data principles combined with reusable data processing workflows adaptable to the needs of experimenters in the lab. Therefore, it should serve for our community, and should be extended with the needs of our biotransformation community.

FIGURES

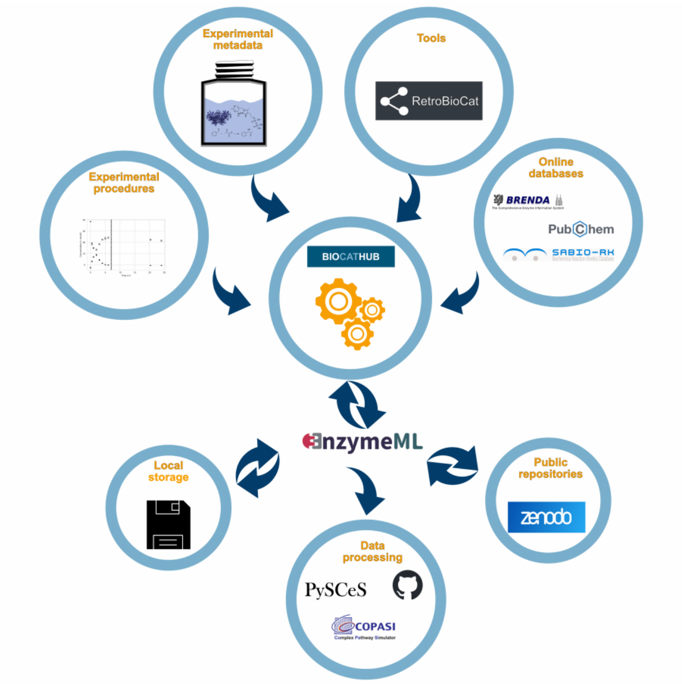

FIGURE 1

Infrastructure of the BioCatHub project

FIGURE 2

KEYWORDS

Research data management | Automated data collection | Applied biocatalysis | Software development

---

BIBLIOGRAPHY

1. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. Scientific data. 2016;3:18-27. doi:10.1038/sdata.2016.18
2. Finnigan W, Hepworth LJ, Flitsch SL, Turner NJ. Nature Catalysis. 2021:1-7.
3. Range J, Halupczok C, Lohmann J, Swainston N, Kettner C, Bergmann FT, et al. The FEBS Journal 2021. doi:10.1111/febs.16318
4. Lauterbach, S., Dienhart, H., Range, J. et al. Nat Methods 20, 400-402 (2023). https://doi.org/10.1038