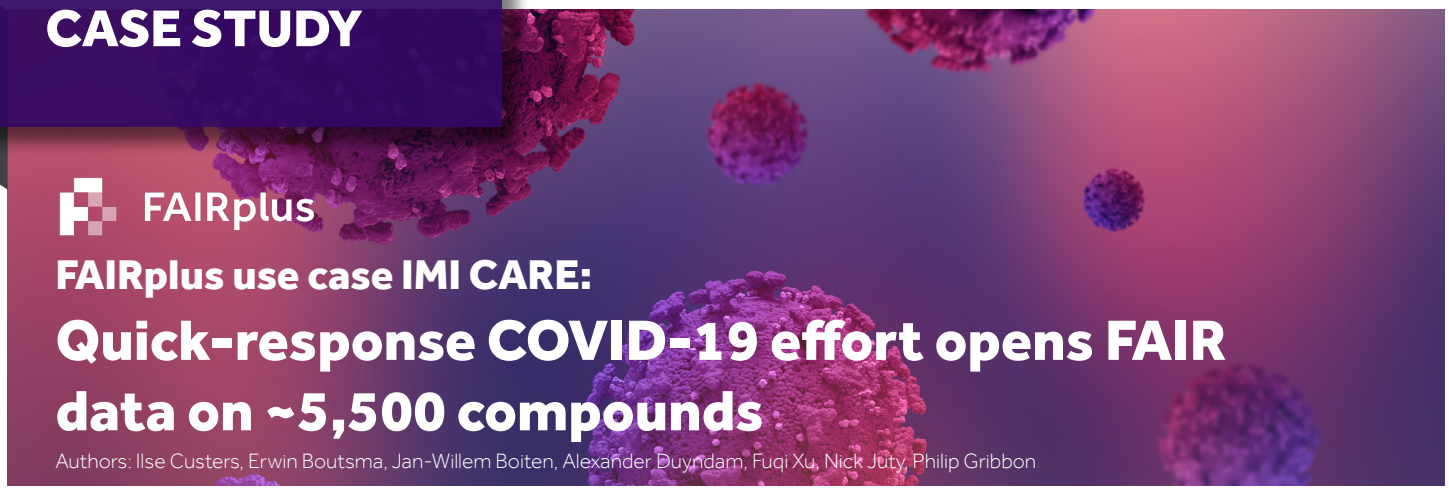
CASE STUDY



FAIRplus

FAIRplus use case IMI CARE:

Quick-response COVID-19 effort opens FAIR data on ~5,500 compounds

Authors: Ilse Custers, Erwin Boutsma, Jan-Willem Boiten, Alexander Duyndam, Fuqi Xu, Nick Juty, Philip Gribbon

Challenge

Recently, in the context of the CARE project, Janssen Pharmaceutica published a dataset in the public domain describing the results of ~5,500 FDA-approved drugs and clinical candidates that have passed Phase I studies which were screened for anti-SARS-CoV-2 activity. Disseminating such datasets according to the principles of FAIR (Findable, Accessible, Interoperable, Reusable) can be useful to researchers world-wide.

Solution

The FAIRplus squad team supported the Janssen Pharmaceutica coronavirus discovery team to improve findability, interoperability, and reuse of the drug data. Together, they increased the FAIR level of the dataset by improving metadata, implementing a standardized ontology, using a standard format to identify compounds, and storing the data on a long-term hosting platform.

Overview

The IMI CARE project[1] was initiated in response to the COVID-19 pandemic with the goal to deliver treatments for COVID-19 and future coronavirus outbreaks. Started in April 2020, CARE not only focuses on delivering novel drugs designed specifically to treat COVID-19 and other coronaviruses, but also on 'repurposing' approved drugs and drug candidates that were originally developed for other diseases that could potentially treat COVID-19. By making CARE data public, it also helps other drug repurposing studies. The CARE project is sharing its scientific results through open-access platforms and peer-reviewed journals, as well as via relevant conferences and other events.

Recently, in the context of the CARE project, Janssen Pharmaceutica published a dataset in the public domain describing the results of ~5,500 FDA-approved drugs and clinical candidates that have passed Phase I studies which were screened for anti-SARS-CoV-2 activity. The antiviral activity of these compounds was tested in an in vitro assay that measures the number of living cells five days after infection with the virus.

The IMI FAIRplus[2] project develops tools and guidelines for making valuable life science data such as the Janssen dataset FAIR. In the past year, the 'squad teams' from FAIRplus, consisting of experts working in universities and pharmaceutical companies, have been actively working to FAIRify data sets from large IMI projects, such as APPROACH, eTOX, RESOLUTE, and COMBINE. The developed tools and methods are subsequently added as 'recipes' to the FAIR cookbook, enabling projects and companies with similar FAIR data challenges to apply this consolidated know-how to increase the FAIRness of their data.

Aims and impact

The IMI projects to be supported by FAIRplus are selected in an unbiased process, based on a recorded set of criteria, among others societal impact. The aim of FAIRifying project data is to facilitate sharing data outside the consortium to conduct further research.

The Janssen dataset was first published on Zenodo[3], an open dissemination research data repository, as a spreadsheet. Janssen approached the FAIRplus project to help improve the FAIRness of the dataset in order to advertise the data to a larger user community and improve the interoperability for future reuse.

1. https://cordis.europa.eu/project/id/101005077, https://www.imi-care.eu/
2. https://fairplus-project.eu/
3. https://zenodo.org/record/4774709#.YikTqPnMK5c

Results

A pre-FAIRification assessment[4] by the FAIRplus squad team highlighted several features that could be modified to increase the dataset's initial FAIRness level of 37%. Especially findability appeared to be a factor to be improved. To increase findability, the FAIRplus squad recommended to upload the data to ChEMBL[5], which is a publicly accessible repository specifically for bioactive compounds.

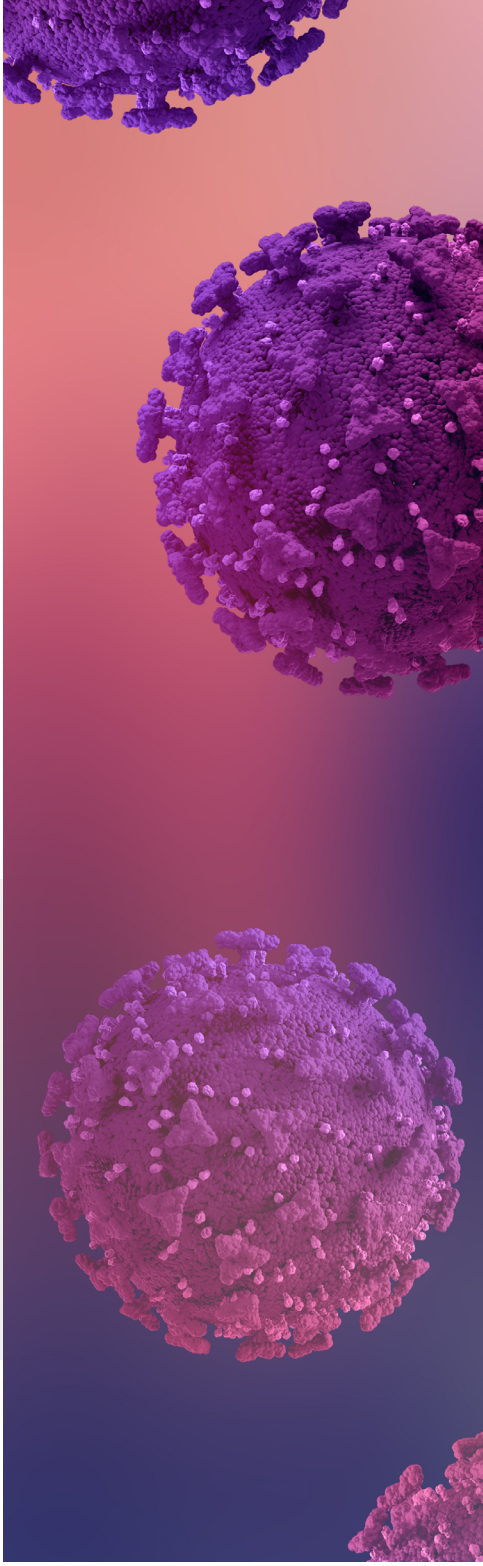
To prepare the dataset for ChEMBL, the team used a set[6] of well-established standards and metadata checklists that allow for transformation to ChEMBL compatible formats including chemical information ontology (CHEMINF), BioAssay Ontology (BOA) and Ontology of units of Measure (OM), as well as identifiers such as InChi and SMILES.

The dataset was further prepared for upload to the ChEMBL database by using a defined ChEMBL compatible template that can be reused by others. The data have been uploaded to ChEMBL and have been released in ChEMBL v30. The procedure on how to publish bioactivity data on ChEMBL will be added to the FAIR Cookbook[6] recipes for future reuse.

Using this set of standards resulted in a post-FAIRification assessment[7] of 92,50%.

Conclusions

While these data did not lead to the discovery of an eligible compound for further development to treat COVID-19, these data are still very valuable information for further COVID-19 drug development. This avoids duplicating previous work, such as comparisons of cellular responses in different cell lines, QSAR, compound clustering, annotation of polypharmacology for repurposed compounds with unexpected or previously unreported antiviral activities.



4. https://fairplus.github.io/fairification-results/2021-07-20-CARE-pre-assessment/
5. https://www.ebi.ac.uk/chembl/document_report_card/CHEMBL4651402
6. https://faircookbook.elixir-europe.org/content/recipes/interoperability/bioactivity-profile.html?highlight=bioassay, data vocabularies
7. https://fairplus.github.io/fairification-results/2021-10-01-CARE-post-assessment/

FAIRplus

FAIRplus Project
FAIRplus-PM@elixir-europe.org

The FAIRplus project has received funding from the Innovative Medicines Initiative Joint Undertaking under grant agreement No 802750. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA Companies. This communication reflects the views of the authors and neither IMI nor the European Union, EFPIA or any Associated Partners are liable for any use that may be made of the information contained herein.