

# BPO: The Bioassay Protocol Ontology, a semantic approach to elevate the reusability of AMR-related drug discovery data

Yojana Gadiya<sup>1,2</sup>, Rakel Arrazuria<sup>3</sup>, Jon Ulf Hansen<sup>4</sup>, Danielle Welter<sup>5</sup>, Fuqi Xu<sup>6</sup>, Philippe Rocca-Serra<sup>7</sup>, Nick Juty<sup>8</sup>, Philip Gribbon<sup>1,2</sup>

Abstract number : EA71

(1) Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP), Schnackenburgallee 114, 22525 Hamburg, and Theodor Stern Kai 7, 60590 Frankfurt, Germany (2) Fraunhofer Cluster of Excellence for Immune-Mediated Diseases (CIMD), Theodor Stern Kai 7, 60590 Frankfurt, Germany (3) Division of Microbiology, Paul-Ehrlich-Institut, Langen, Germany (4) Dept. of Bacteria, Parasites and Fungi, Statens Serum Institut, 2300 Copenhagen, Denmark (5) Luxembourg Centre for Systems Biomedicine, ELIXIR Luxembourg, University of Luxembourg, L-4367 Belval, Luxembourg (6) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK (7) University of Oxford e-Research Centre, Department of Engineering Science, 7 Keble road, OX1 3QG, Oxford, United Kingdom (8) Department of Computer Science, The University of Manchester, Oxford Road, Manchester, UK

**INTRODUCTION:** To enhance the reproducibility and usability of bioassays in antimicrobial resistance (AMR) associated antibacterial drug discovery and development, there is an **increasing need for standardisation of bioassay metadata into machine-readable formats**. Such a standardisation process requires mapping bioassay data to standard ontologies and can be performed at the study result output and protocol levels. At the result output level, **general ontologies such as the BioAssay Ontology (BAO) already exist, but antibacterial drug discovery and AMR-specific ontologies that can aid in the standardisation process at the protocol level are missing**. Hence, we propose a new application-specific ontology that will enable researchers to convert unstructured bioassay protocol data into structured machine-readable formats thereby promoting their overall reusability.

**METHODOLOGY:** The Bioassay Protocol Ontology (BPO) was constructed in accordance with the ontology build life cycle. The ROBOT tool (<http://robot.obolibrary.org/>) and the Ontology Development Kit ODK (<https://github.com/INCATools/ontology-development-kit>) were used by scientific domain experts to programmatically develop the ontology. The BPO is accessible to the AMR community in the Ontology Web Language (OWL) and Open Biomedical Ontologies (OBO) formats. We tested 10 bioassay specific ontologies within this project to check the extend of extraction of information from natural language processing (NLP) models. But none of them were able to provide a complete overview of underlying information.



**RESULTS:** The BPO was initially established in order to extend the **interoperability and reusability of in-vivo AMR bioassay** protocol data generated within the Innovative Medicines Initiative (IMI) AMR Accelerator project, COMBINE.

Thus, we created an **application ontology** through the combination of existing terms from prominent ontologies such as NCBI Taxonomy and PATO along with new terms around the animal model-specific details, such as mouse inbred and outbred status. Alongside addition of new terms, several terms have additional been requested to the parent reference ontologies.

Overall, BPO ontology has **209 classes and 2,242 axioms**. In the coming months, we aim to integrate BPO into our information retrieval environment (GRIT42 <https://grit42.com/>) for **protocol-level annotation of bioassays**.

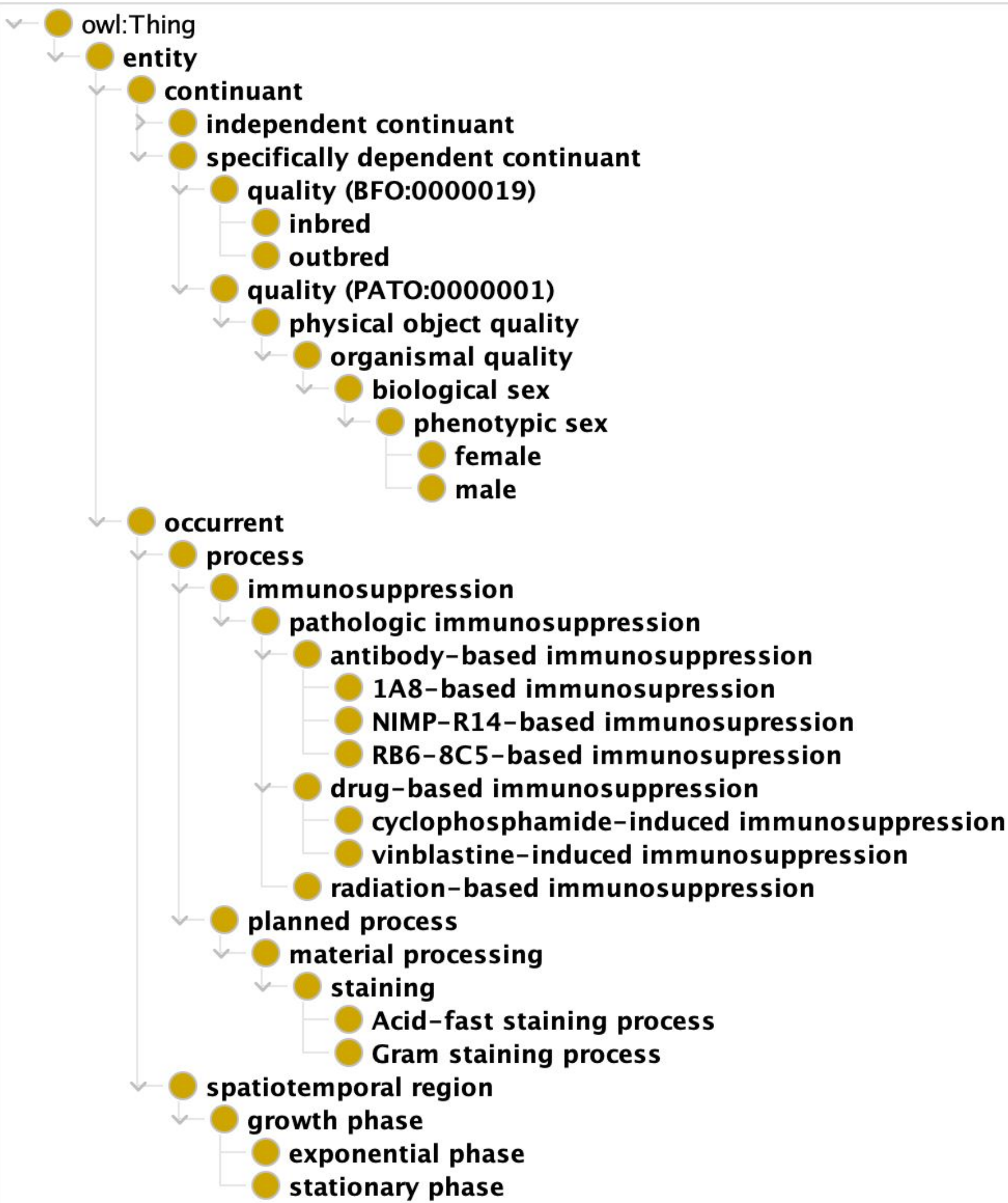


Figure 1: The overview of the major classes included in the BPO ontology.

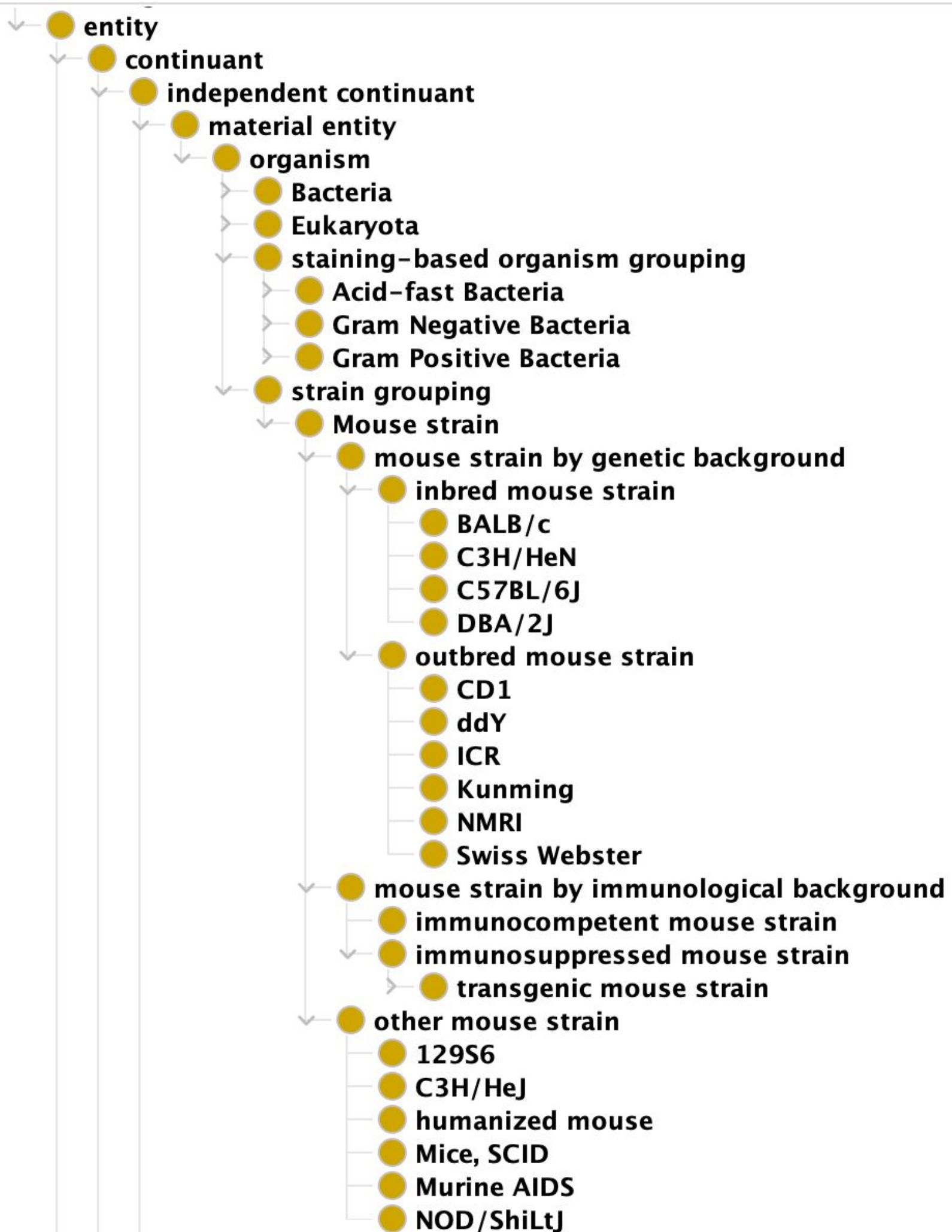
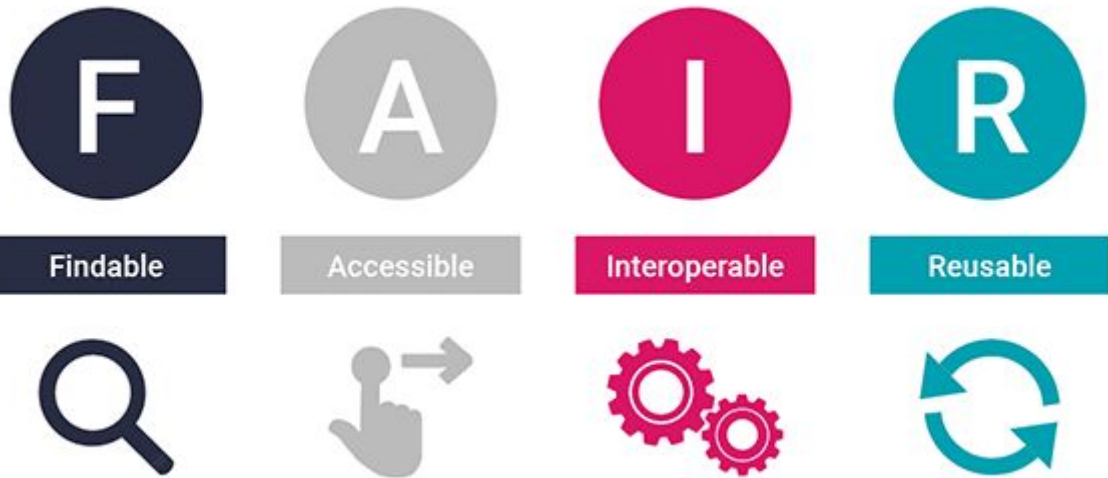


Figure 2: The additional classification and metadata annotation in BPO ontology. This includes the classification of bacteria strain and annotation of mouse strains based on their breeding or immunological classification



**CONCLUSION:** The development of BPO is a first attempt to **standardise and organise information within in-vivo AMR-related drug discovery bioassay protocol data in a structured manner**. BPO will allow researchers to capture information regarding experimental details such as the type of mouse model, the bacterial strain, and the sex and growth phases of mouse and bacteria respectively from the protocol. We have made the ontology available for community reuse on GitHub <https://github.com/Fraunhofer-ITMP/bpo>.

This research was funded by the EU/EFPIA Innovative Medicines Initiative 2 Joint Undertaking project Collaboration for prevention and treatment of MDR bacterial infections (COMBINE) IMI2 JU (grant agreement # 853967) and FAIRplus (grant agreement #802750).