

The Radiation Biology Ontology: a new tool supporting FAIR principles across radiation biology facilitating data discovery and integration

Daniel Berrios¹, Luke T. Slater²⁻⁵, Jack Miller⁶, Kristen Peach⁷, Sylvain V. Costes¹, and Paul N. Schofield⁸

¹NASA Ames Research Center, ²College of Medical and Dental Sciences, Institute of Cancer and Genomic Sciences, University of Birmingham, UK, ³Institute of Translational Medicine, University Hospitals Birmingham, NHS Foundation Trust, UK, ⁴MRC Health Data Research UK (HDR UK) Midlands, Birmingham, UK, ⁵University Hospitals Birmingham NHS Foundation Trust, Edgbaston, Birmingham, UK, ⁶KBR, NASA Ames Research Center, Moffett Field, CA, USA, ⁷Bionetics, NASA Ames Research Center, ⁸Dept of Physiology Development and Neuroscience, University of Cambridge, Downing Street, Cambridge, UK.

Introduction

Development of the Radiation Biology Ontology (RBO) was motivated by the need for a comprehensive, well-structured ontology for encoding radiation biology metadata.

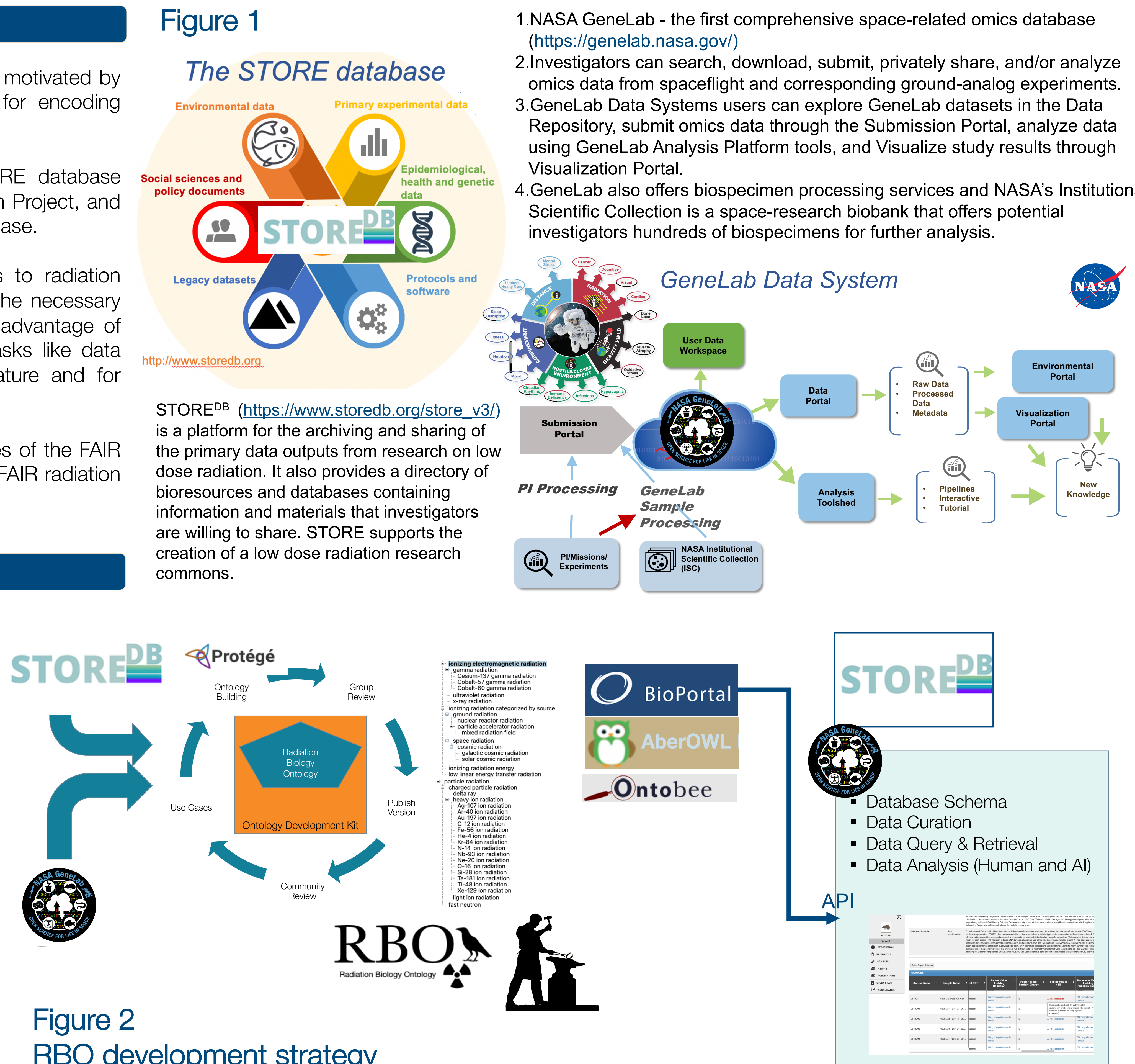
The primary use-cases were archiving data in the STORE database (<https://www.storedb.org/>), the repository for the RadoNorm Project, and in GeneLab (<https://genelab.nasa.gov/>), NASA's 'omics database.

The scope of radiobiology research ranges from physics to radiation oncology to socio-legal studies; no existing ontology has the necessary breadth or depth. In addition, a formal ontology has the advantage of being usable for machine learning and, importantly, for tasks like data integration, knowledge extraction from the scientific literature and for query extension and data classification.

Standardisation of metadata is one of the primary objectives of the FAIR principles for open data; RBO is an important landmark for FAIR radiation biology data.

Methods

- Concepts were gathered from domain experts and driven by use-cases and community standards.
- We then identified and imported existing OBO Foundry classes relevant to radiation biology. We also scrutinized datasets from investigations of radiation effects in STORE and the NASA GeneLab and LSDA repositories (Figure 1), and added additional classes, instances, and attributes into the RBO that should be used to annotate these data.
- We used open-source software (the Ontology Development Kit, Protégé and WebProtégé), working within the OBO Foundry framework (Figure 2).
- Continuous build integration tools allowed our international RBO collaboration to be more efficient, and focus its efforts on semantic model design. E.g. inclusion of instances for infrastructures.



Results and Conclusions

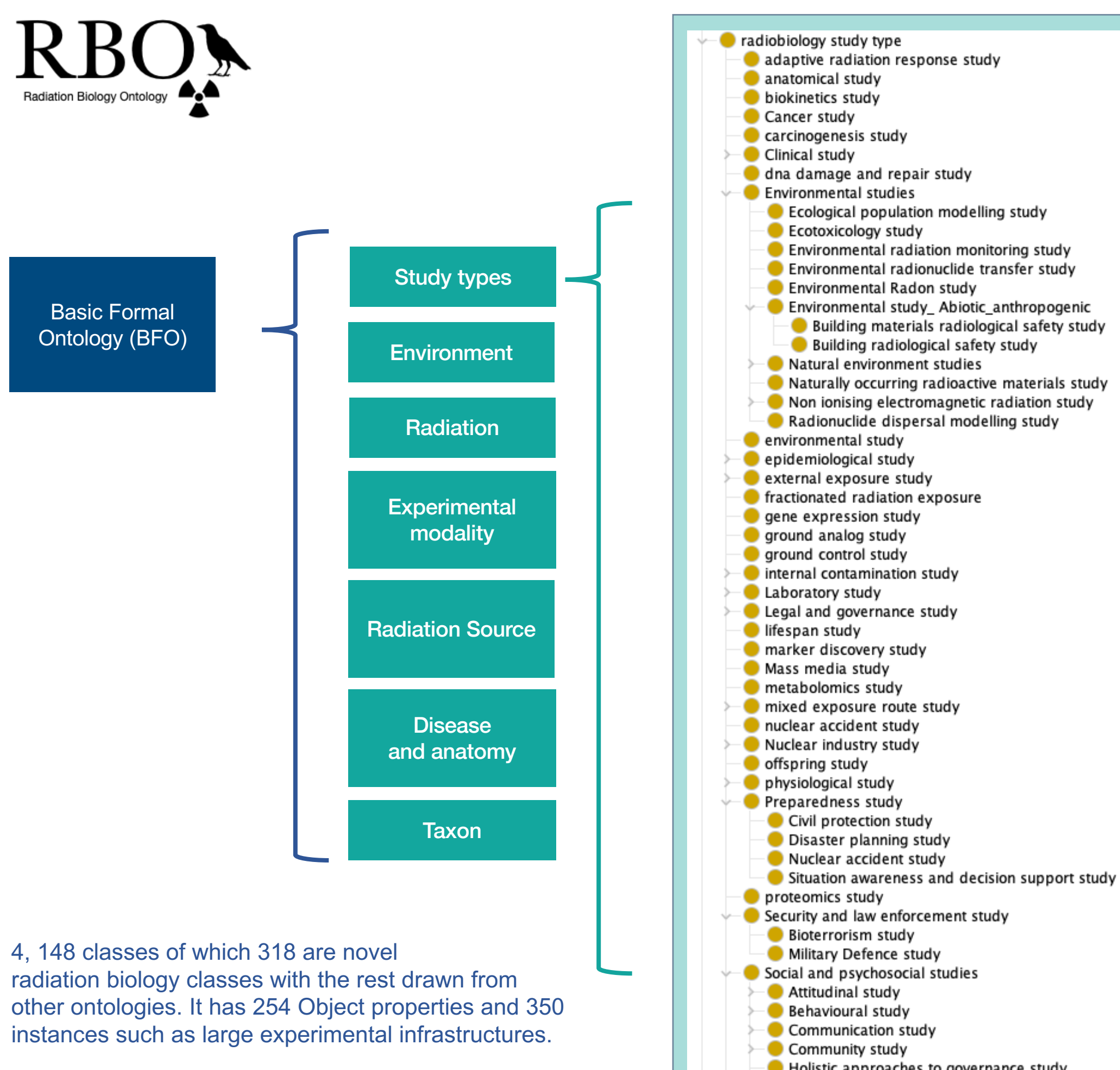


Figure 3

RBO is built on the upper level BFO foundational ontology with seven upper levels as parents for a wide range of entities, concepts and processes relevant to radiation biology in its widest sense. It includes classes capturing concepts in physics, chemistry, the environment, experimental science, society and disease amongst others.

References

1. Wilkinson, et al. (2016). Scientific data 3:160018
2. <https://obofoundry.org/ontology/rbo.html>
3. <https://bioportal.bioontology.org/ontologies/RBO>
4. <https://ontobee.org/ontology/RBO>
5. <http://aber-owl.net/ontology/RBO/#/>

Acknowledgements

Work on RBO is funded by RadoNorm, which receives funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009, and NASA.

The RBO ontology

- The current release of RBO contains over 400 annotated classes and instances specific to the study of radiation on biological systems, as well as imports of more than 3000 additional classes from other OBO Foundry ontologies
- RBO cross references 27 OBO ontologies directly providing semantic linkage to most relevant databases
- Published through the OBO Foundry²
- Available through NCBI Bioportal web site³, OntoBEE⁴, AberOWL⁵ and application programming interface at: <https://bioportal.bioontology.org/ontologies/RBO>
- RBO has a Github site on: <https://github.com/Radiobiology-Informatics-Consortium/RBO> where users may contribute to its development.

RBO and FAIR data

In support of the principles of FAIR data, RBO supports data discovery and interoperability.

- RBO is registered in appropriate repositories (see above) with high quality metadata for future content seekers
- Provides unique identifiers to retrieve all forms of digital content and knowledge
- Standardised metadata allow for data discovery via semantic web technologies – linked data
- Establishment of community standards for more facile knowledge utilisation
- Use of common vocabularies to share terms and facilitate cross database and dataset queries

Applications of the RBO

- Automated knowledge extraction using Natural Language Processing
 - from scientific publications
 - from textbooks, reports
 - legacy databases
- Automated data curation
 - Metadata harmonization
 - Increase automation
- Lingua franca for data integration for meta-analyses.
- Scientific data discovery.
- Automated inference and machine learning for example using graph convolutional artificial neural networks.

