Data sharing in radiation biology; towards FAIR

Paul Schofield, University of Cambridge

RadoNorm
Managing risks from radon and NORM

Bundesamt
für Strahlenschutz

UNIVERSITY OF CAMBRIDGE

UNIVERSITY OF BIRMINGHAM

FAIR sharing

Use of data at scale by humans and machines

SCIENTIFIC DATA

SPRINGER NATURE

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

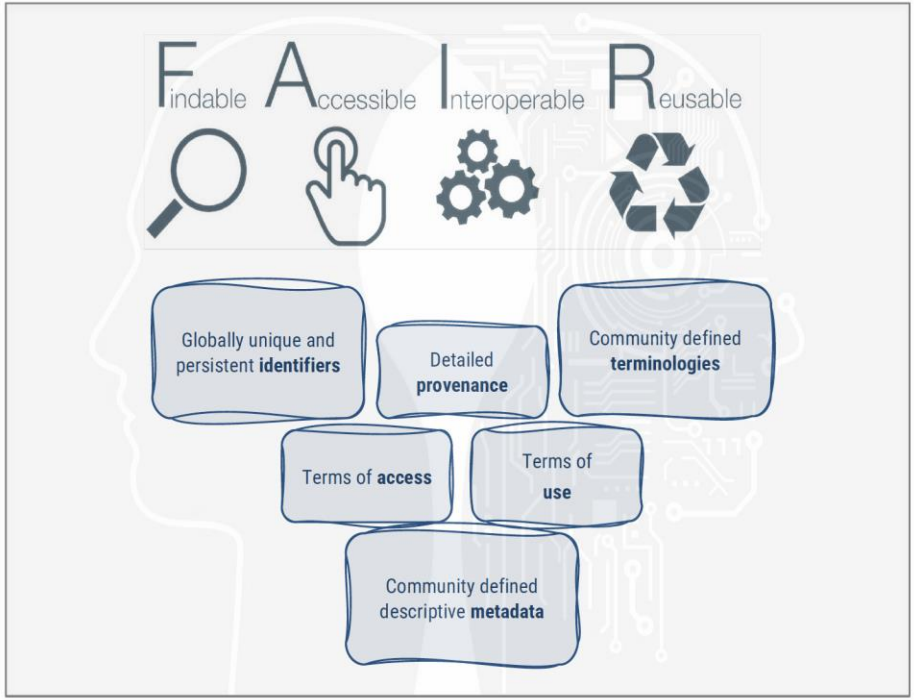
SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

Access & Citations

451k	3173	3993
Article Accesses	Web of Science	CrossRef

1958

Attention Score



Where's the data?

STORE DB

NASA GeneLab

OPEN SCIENCE FOR LIFE IN SPACE

The STORE database

Funded by the European Commission EURATOM Programme and the Bundesamt fuer Strahlenschutz since 2009

European Commission | Horizon 2020 European Union funding for Research & Innovation

RadoNorm
Managing risks from radon and NORM

elixir

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

http://doi.org/10.17616/R3732R

STOREDB

Free to the entire community and sustained by the BfS
Self curation, upload and controlled access.

FAIRsharing.org
standards, databases, policies

DataCite
FIND, ACCESS, AND REUSE DATA

Findable Accessible Interoperable Reusable

http://www.storedb.org

CC BY NC SA

Funded by Euratom research and training programme 2014-2018 under grant agreement No 900009.

Open Science Projects

Open Science Projects primary goals aim to increase collaborative scientific data sharing, analysis and more rapid scientific advancement.

GeneLab

GeneLab, an open science multi-omics repository, covering transcriptomics, metagenomics, epigenomics, proteomics, and metabolomics. Studies comprise of data from model organisms including microbes, plants, fruit flies, rodents and humans.

Learn more GeneLab

BSP

The NASA Space Biology Biospecimen Sharing Program (BSP) collects biospecimens to maximize the scientific return from biological spaceflight and associated ground investigations and to encourage and broaden participation from the scientific community in space biology-related research.

Learn more about BSP

ALSDA

Ames Life Sciences Data Archive (ALSDA) collects, curates, and makes available space-relevant higher-order phenotypic datasets. Datasets that enable scientists to perform retrospective analysis across missions, experiments, life science disciplines, research subjects, and species.

Learn more about ALSDA

NBISC

NASA Biological Institutional Scientific Collection (NBISC) is a biorepository of non-human samples collected from NASA-funded spaceflight investigations and correlative ground studies. The purpose of NBISC is to receive, store, document, preserve, and make the collection available to the scientific community.

Learn more about NBISC

GeneLab Data System

1. NASA GeneLab - the first comprehensive space-related omics database (https://genelab.nasa.gov/)
2. Investigators can search, download, submit, privately share, and/or analyze omics data from spaceflight and corresponding ground-analog experiments.
3. GeneLab Data Systems users can explore GeneLab datasets in the Data Repository, submit omics data through the Submission Portal, analyze data using GeneLab Analysis Platform tools, and Visualize study results through Visualization Portal.
4. GeneLab also offers biospecimen processing services and NASA's Institutional Scientific Collection is a space-research biobank that offers potential investigators hundreds of biospecimens for further analysis.

Standardisation

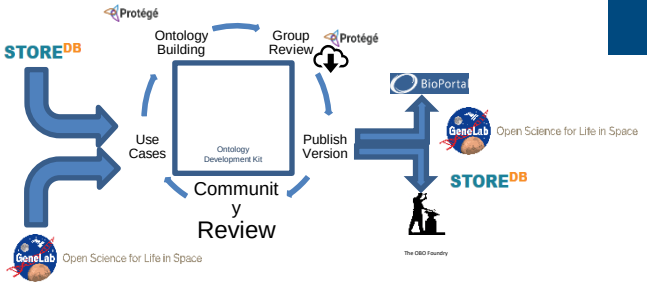
Ontologies

Comprised of standardised hierarchical concepts and relationships

Capture semantic data and metadata

Radiation Biology Ontology (RBO)

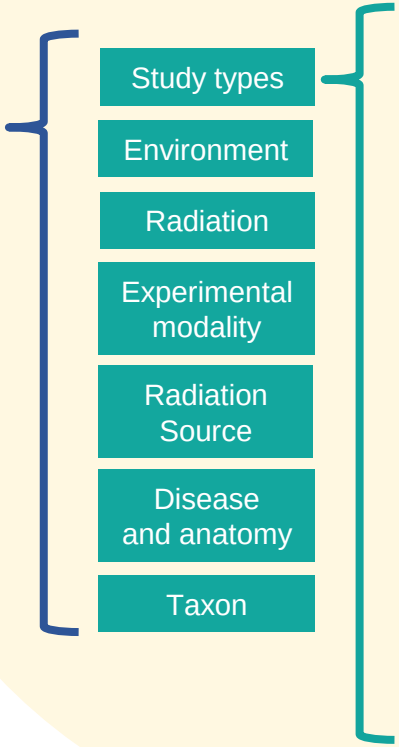
- Facilitates data retrieval and query expansion from NASA's GeneLab 'omics database and STORE

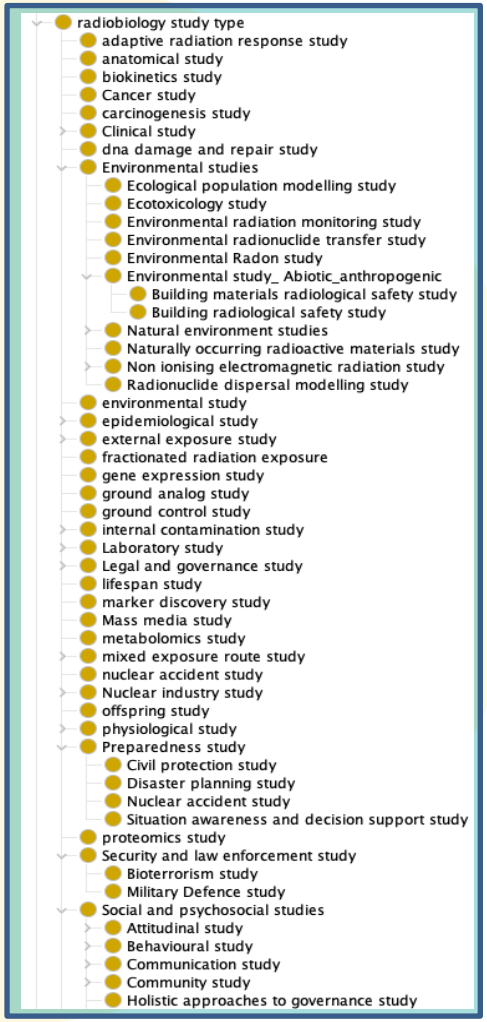


- Contains over 200 annotated classes and instances specific to the study of radiation on biological systems, as well as imports of more than 3500 additional classes from 13 other OBO Foundry ontologies

- Published through the OBO Foundry

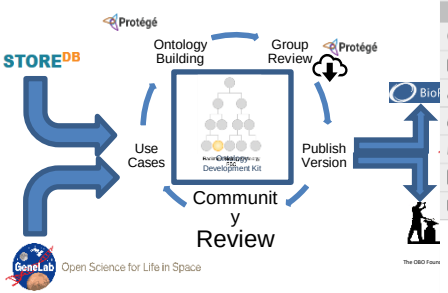
- Available through NCBI Bioportal web site and application programming interface at https://bioportal.bioontology.org/ontologies/RBO

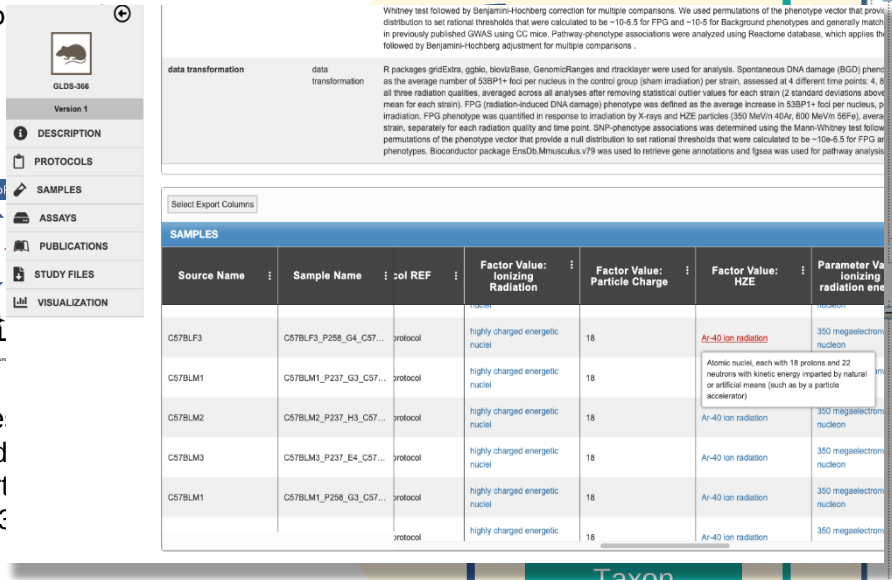




Radiation Biology Ontology (RBO)

- Facilitates data retrieval and query expansion from NASA's GeneLab 'omics datab STORE

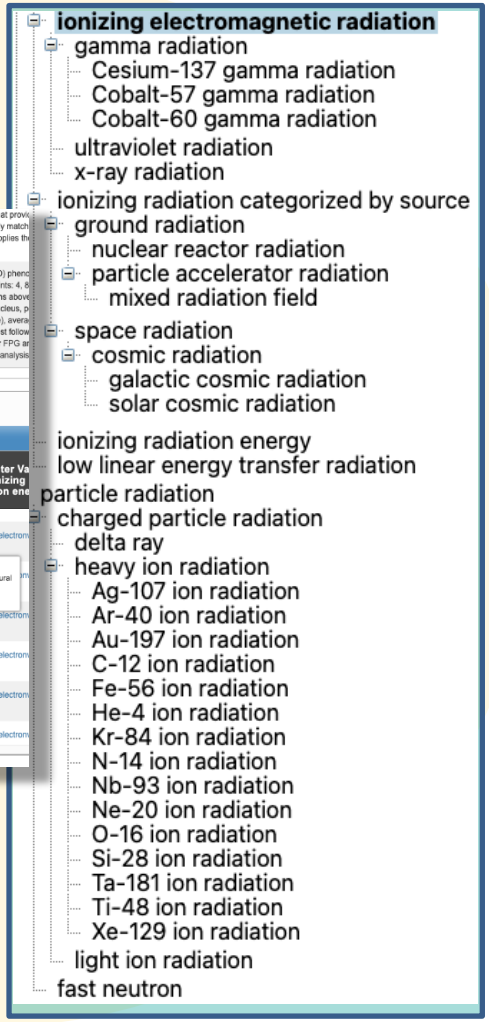




- Contains over 200 annotated classe instances specific to the study of rad biological systems, as well as import than 3500 additional classes from 13 Foundry ontologies

- Published through the OBO Foundry

- Available through NCBI Bioportal web site and application programming interface at https://bioportal.bioontology.org/ontologies/RBO



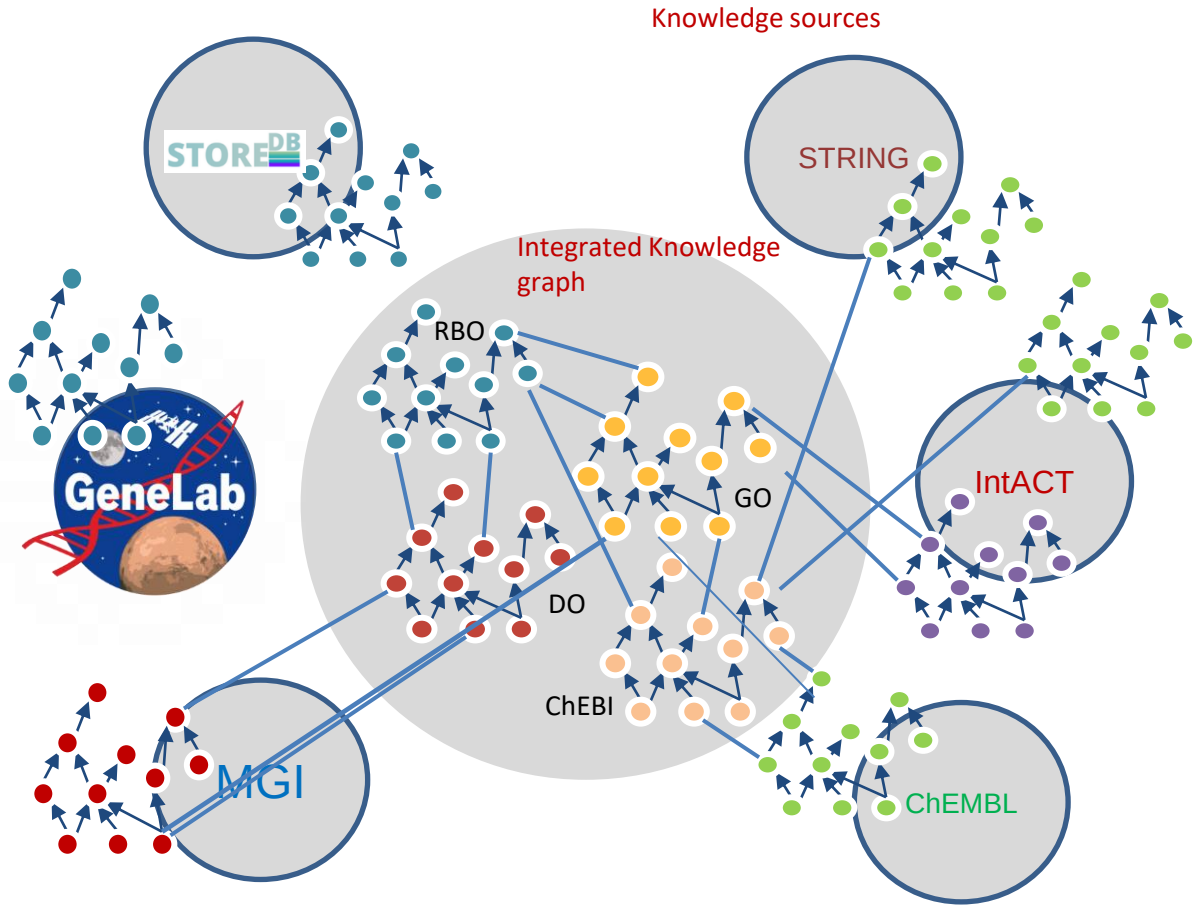
Using open data to create new knowledge

Semantic data standards and metadata

- Discovering and integrating data between databases
- Federated queries and query expansion
- Semantic integration
- Ontologies permit the assertion of defined relationships between concepts

RBO cross-references 13 OBO ontologies directly providing semantic linkage to most relevant databases

- Construction of Knowledge Graphs
- Graph embeddings for data representation
- Classification and similarity
- Inductive inference
- Learning over graph convolutional neural networks



Acknowledgements

Daniel Berrios

Luke T. Slater

Jack Miller

Kristen Peach

Sylvain V. Costes

Ulrike Kulka

Michael Gruenberger

UNIVERSITY OF CAMBRIDGE

RadoNorm
Managing risks from radon and NORM

UNIVERSITY OF BIRMINGHAM

Bundesamt
für Strahlenschutz

Funded by the National Aeronautics and Space Agency

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

www.radonorm.eu