

Data sharing in radiation biology; towards FAIR

Paul Schofield, University of Cambridge

FAIR sharing

Use of data at scale by humans and machines

SCIENTIFIC DATA 
SPRINGER NATURE

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

SUBJECT CATEGORIES
» Research data
» Publication characteristics

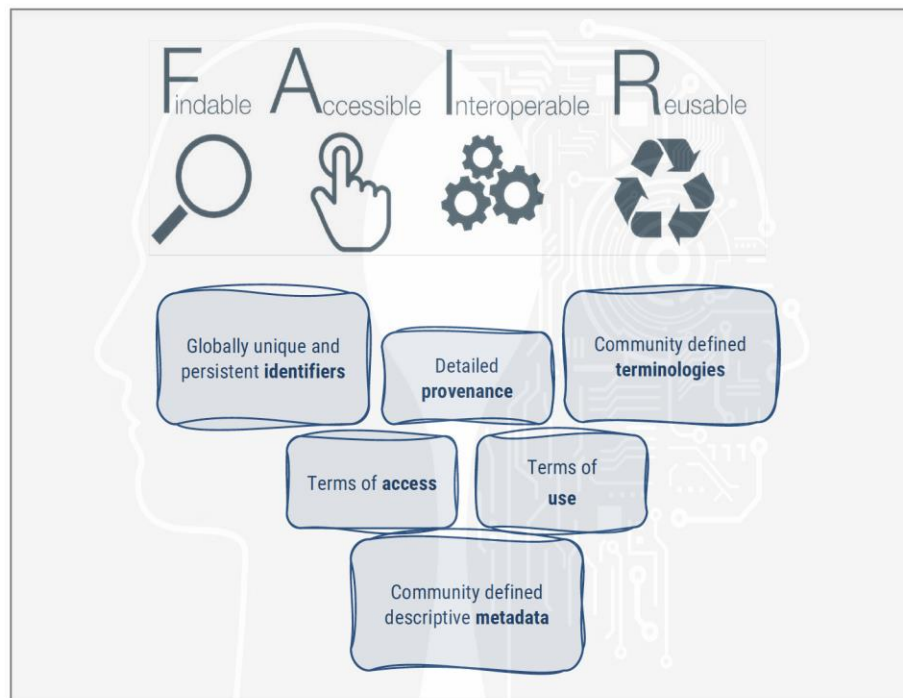
Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Anie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Boume, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, 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

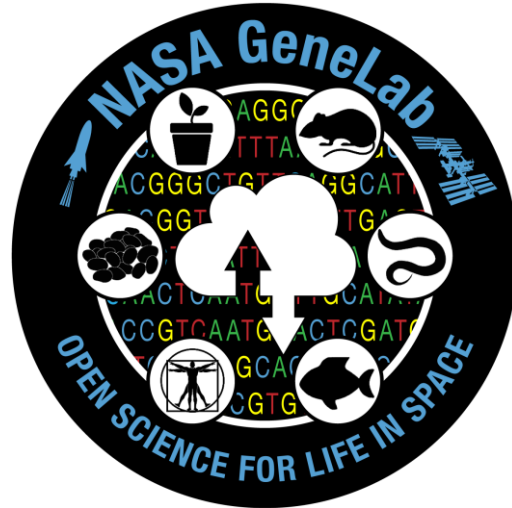
451k	3173	3993
Article Accesses	Web of Science	CrossRef

 1958
Attention Score





Where's the data?



STORE^{DB}

The STORE database

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

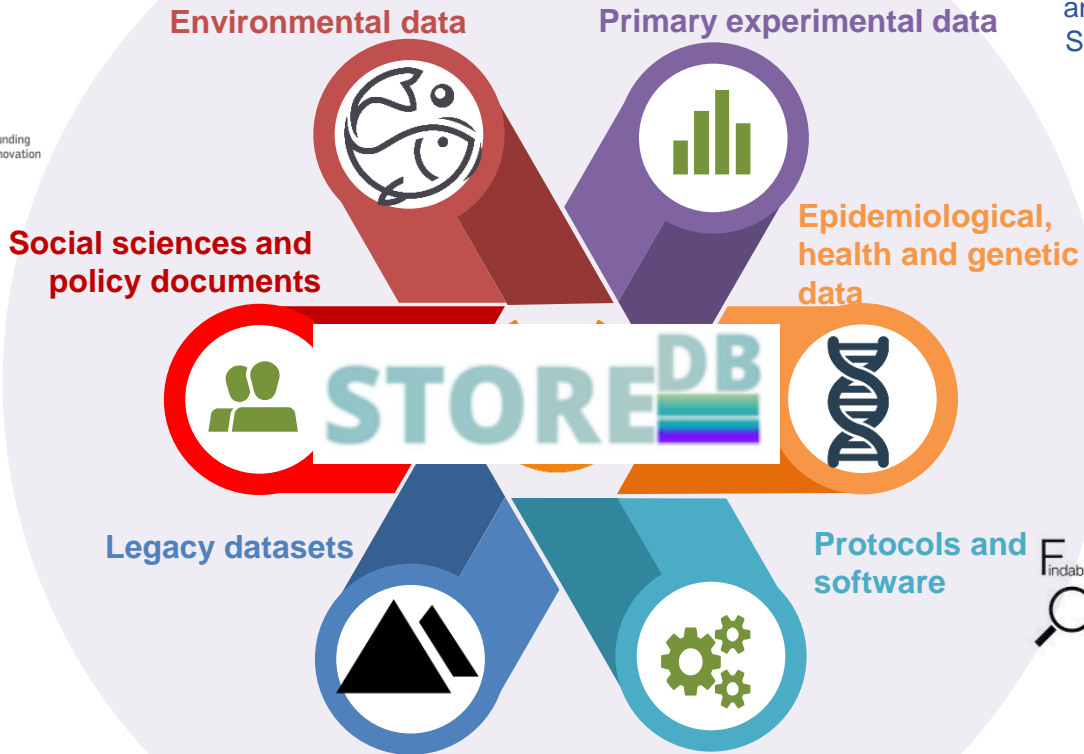


re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES



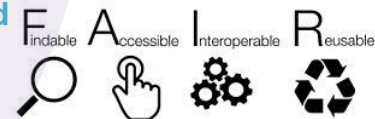
<http://doi.org/10.17616/R3732R>

STOREDB



<http://www.storedb.org>

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



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](#)



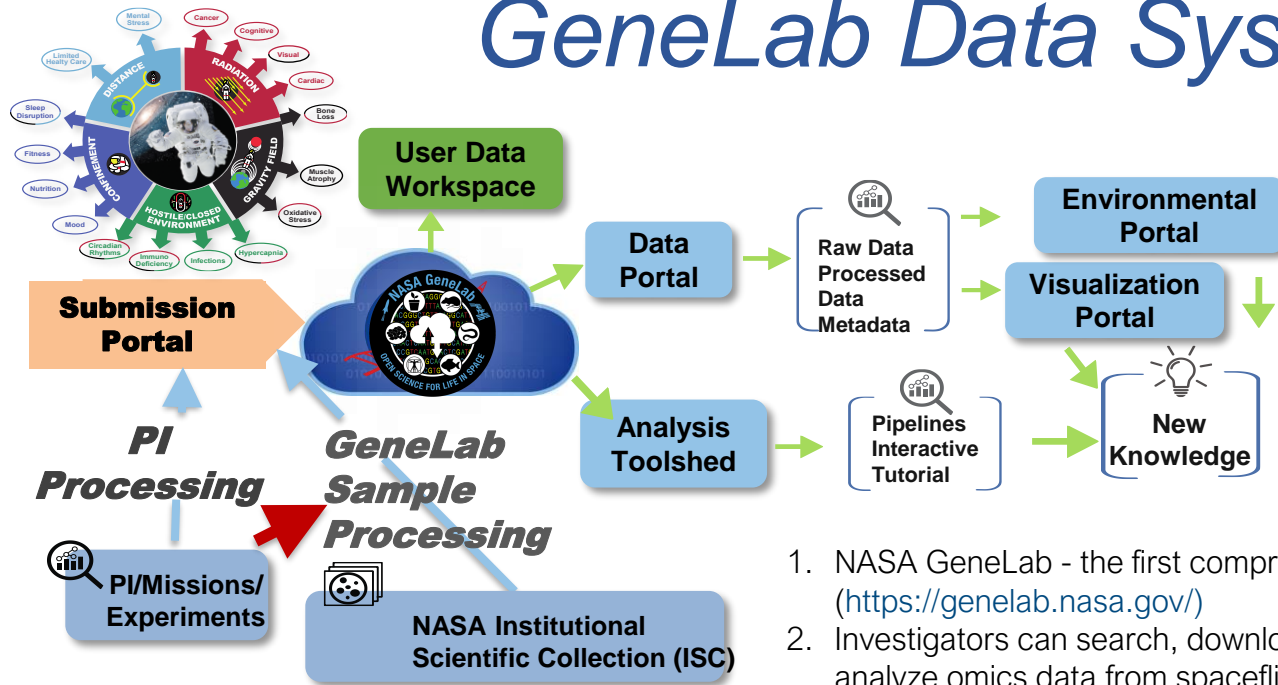
NBISC

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

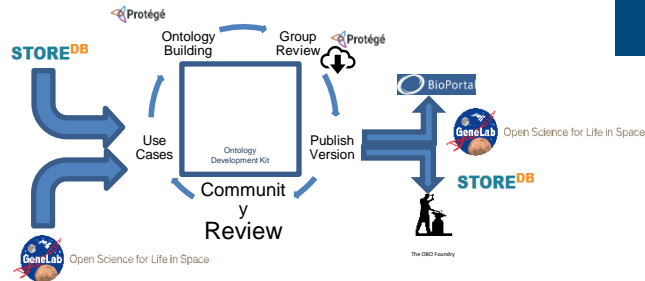
Comprised of standardised hierarchical concepts and relationships

Capture semantic data and metadata



Radiation Biology Ontology (RBO)

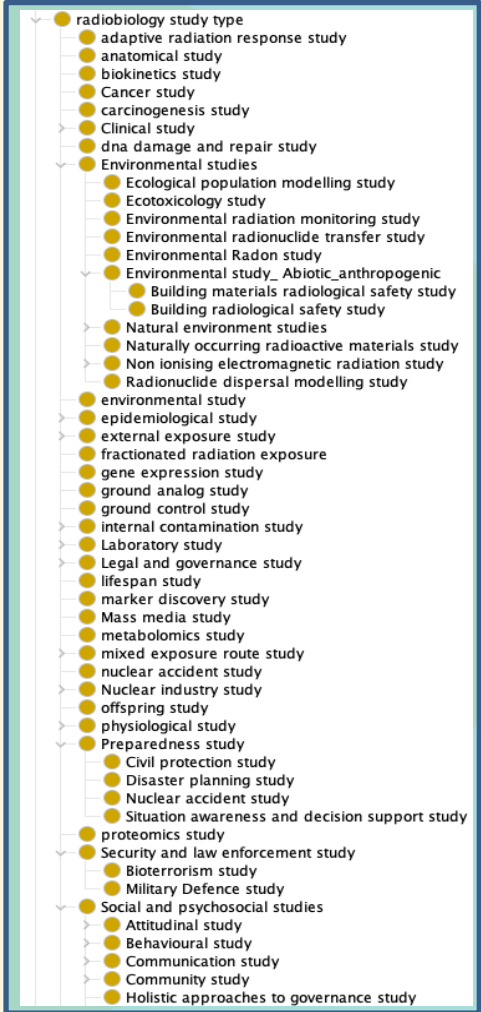
- Facilitates data retrieval and query expansion from NASA's GeneLab 'omics database and STORE



- 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>

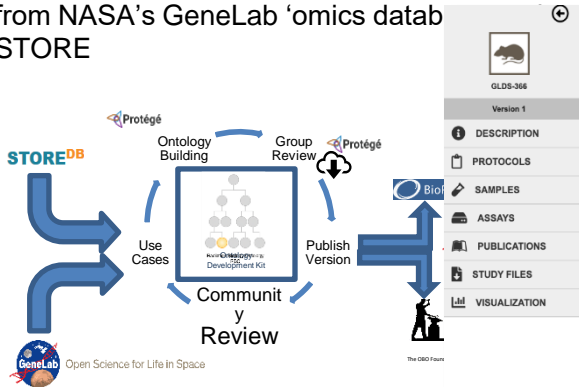
Basic Formal Ontology (BFO)

- Study types
- Environment
- Radiation
- Experimental modality
- Radiation Source
- Disease and anatomy
- Taxon



Radiation Biology Ontology (RBO)

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



- Contains over 200 annotated classes, instances specific to the study of radiation biological systems, as well as imported 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>

data transformation

data transformation

R packages gridExtra, ggplot2, bioRxiv, GenomicRanges and trackranger were used for analysis. Spontaneous DNA damage (BGD) phenotype as the average number of 53BP1+ foci per nucleus in the control group (sham irradiation) per strain, assessed at 4 different time points: 4, 8, 16, 32 hours post-irradiation. FPG (radiation-induced DNA damage) phenotype was defined as the average increase in 53BP1+ foci per nucleus, 96 hours post-irradiation. FPG phenotype was quantified in response to irradiation by X-rays and HZE particles (350 MeV/n 40Ar, 600 MeV/n 56Fe), average strain, separately for each radiation quality and time point. SNP-phenotype associations was determined using the Mann-Whitney test followed by permutations of the phenotype vector that provide a null distribution to set rational thresholds that were calculated to be $< 10^{-4}$ for FPG or $< 10^{-5}$ for BGD phenotypes. Bioconductor package Enrichr was used to retrieve gene annotations and g:sea was used for pathway analysis.

Select Export Columns

Source Name	Sample Name	REF	Factor Value: Ionizing Radiation	Factor Value: Particle Charge	Factor Value: HZE	Parameter Value: Ionizing radiation energy
C57BLF3	C57BLF3_P258_G4_C57...	protocol	highly charged energetic nuclei	18	Ar-40 ion radiation	350 megaelectron-nucleon
C57BLM1	C57BLM1_P237_G3_C57...	protocol	highly charged energetic nuclei	18	Atomic nuclei, each with 18 protons and 22 neutrons with kinetic energy imparted by natural or artificial means (such as by a particle accelerator)	350 megaelectron-nucleon
C57BLM2	C57BLM2_P237_H3_C57...	protocol	highly charged energetic nuclei	18	Ar-40 ion radiation	350 megaelectron-nucleon
C57BLM3	C57BLM3_P237_E4_C57...	protocol	highly charged energetic nuclei	18	Ar-40 ion radiation	350 megaelectron-nucleon
C57BLM1	C57BLM1_P258_G3_C57...	protocol	highly charged energetic nuclei	18	Ar-40 ion radiation	350 megaelectron-nucleon
		protocol	highly charged energetic	18	Ar-40 ion radiation	350 megaelectron-nucleon

Taxon

- ionizing electromagnetic radiation
 - gamma radiation
 - Cesium-137 gamma radiation
 - Cobalt-57 gamma radiation
 - Cobalt-60 gamma radiation
 - ultraviolet radiation
 - x-ray radiation
- ionizing radiation categorized by source
 - ground radiation
 - nuclear reactor radiation
 - particle accelerator radiation
 - mixed radiation field
- space radiation
 - cosmic radiation
 - galactic cosmic radiation
 - solar cosmic radiation
- ionizing radiation energy
 - low linear energy transfer radiation
- particle radiation
 - charged particle radiation
 - delta ray
 - heavy ion radiation
 - Ag-107 ion radiation
 - Ar-40 ion radiation
 - Au-197 ion radiation
 - C-12 ion radiation
 - Fe-56 ion radiation
 - He-4 ion radiation
 - Kr-84 ion radiation
 - N-14 ion radiation
 - Nb-93 ion radiation
 - Ne-20 ion radiation
 - O-16 ion radiation
 - Si-28 ion radiation
 - Ta-181 ion radiation
 - Ti-48 ion radiation
 - Xe-129 ion radiation
 - light ion radiation
 - fast neutron

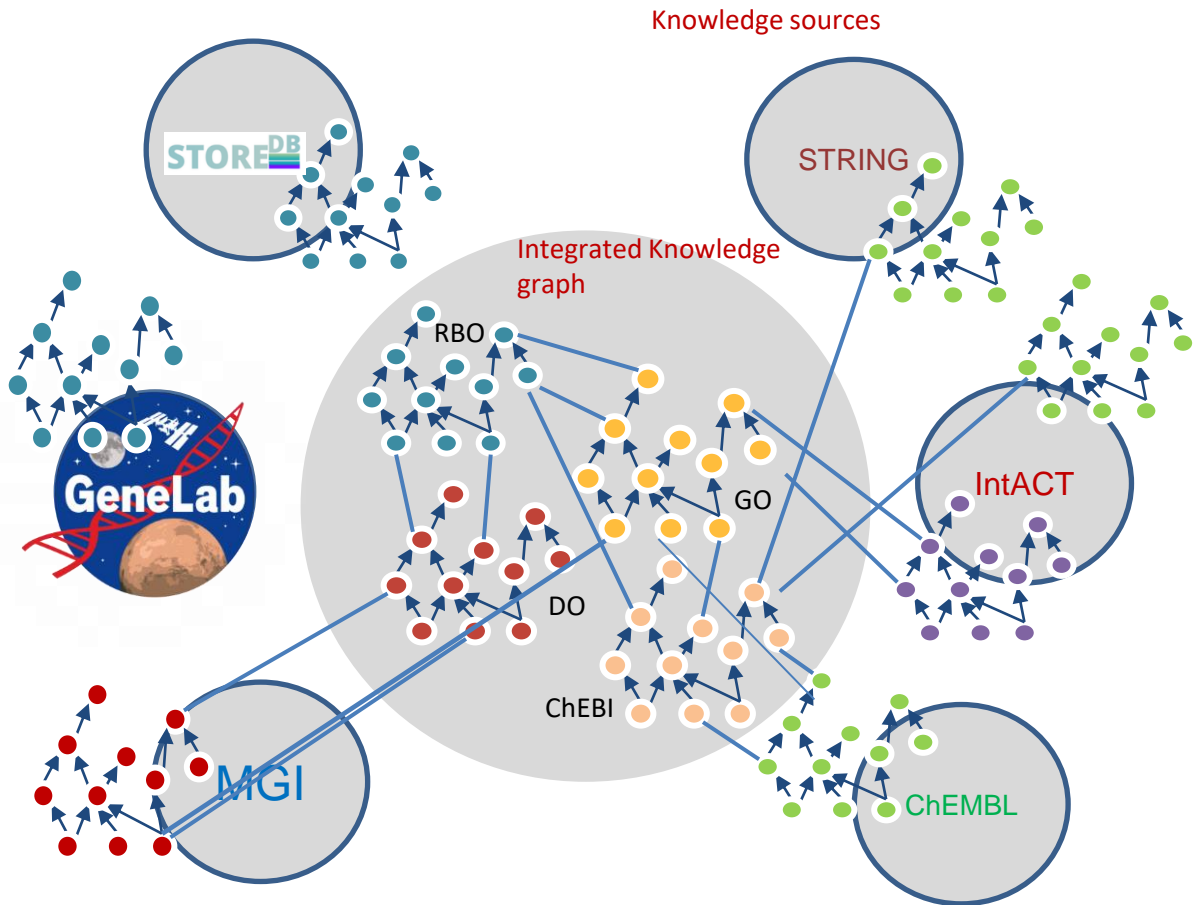
Using open data to create new knowledge

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



Bundesaamt
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