



Deliverable 8.5:

STORE^{DB} final report

Work Package **8**



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

Document information

Project Acronym	RadoNorm
Project Title	Towards effective radiation protection based on improved scientific evidence and social considerations - focus on radon and NORM
Project Type	RIA
EC grant agreement No.	900009
Project starting / end date	1st September 2020 – 31 August 2025
Work Package No.	8
Work Package Title	Communication, Dissemination and Exploitation of Results
Deliverable No.	8.5
Deliverable Title	STORE^{DB} Final Report
Lead Beneficiary	EIMV
Contractual Delivery Date	M60
Actual Delivery Date	28.8.2025
Type	R
Dissemination level	PU
Authors	Paul Schofield (UCam), Mandy Birschwilks (BfS), Michael Grünberger (UCAM)
To be cited as:	Schofield, P.N. (2025): Report on STORE^{DB} database. Final version as of 27.8.2025 of deliverable D8.5, RadoNorm project.

Disclaimer

This document reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Acknowledgement

This document is a deliverable of the RadoNorm project. This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Status of deliverable		
	By	Date
Delivered (Lead Beneficiary)	UCAM	26.08.2025
Verified (WP Leader)	EIMV	27.08.2025
Reviewed (Reviewers)	WPL, PC	27.08.2025
Approved (PC)	BfS	27.08.2025
Submitted to EC (PC)	BfS	27.08.2025

Executive Summary

The RadoNorm project, funded under the Euratom research and training programme (2019–2020), pursued a comprehensive five-year strategy to enhance radiation protection from radon and NORM through integrated communication, dissemination, and exploitation of results. Work Package 8 (WP8) coordinated activities to ensure that project outcomes were visible, accessible, and impactful across scientific, policy, industrial, and public domains.

As part of the dissemination strategy RadoNorm undertook to support FAIR data management and sharing by developing a Data Management Plan (DMP), supporting a platform for data sharing; the STORE^{DB} database, and creating a new ontology for metadata relating to radiobiology and specifically RadoNorm data outputs, the Radiation Biology Ontology (RBO). Teams of data wranglers were appointed for each WP to ensure that good data management practices were maintained across the project and to provide local support to investigators.

Ongoing Audit of the DMP did not find it necessary to alter the plan as the project progressed. Audit of publication-associated data sharing, to date, indicated that while many authors followed the policy of sharing data as “openly as possible and as closed as necessary” through appropriate licencing, journal supplementary websites were still widely used, as were “refer to author” statements, and did not always fulfil the requirement for FAIR data availability.

Maintenance and development of the database and ontology continued through the project with associated user consultation and training, and both pre- and post-publication data deposited into STORE^{DB} or other appropriate databases for long term archiving and distribution. Progress on STORE and the RBO was presented at several relevant scientific meetings as well as RadoNorm events, and we contribute to the ability to integrate radiobiology data across many data resources through the common metadata standard of the RBO. This provides additional value and FAIRness to data produced under RadoNorm.

FAIR data access supports the exploitation activities of RadoNorm by making data available for future reanalysis, integration with other datasets, and supports publications. Much of what has been deposited will be of value in public policymaking and safety assessment as well as in fundamental scientific discovery. The long-term persistence of STORE^{DB} will provide a mechanism for sustained access many years after the termination of the project.

We thank the Senior Software Developer/Bioinformatician Michael Grünberger for his continued contributions during the technical development and ongoing refinement; his dedicated work is greatly appreciated.

Table of contents

Executive Summary.....	4
Table of contents	5
List of figures	6
Glossary.....	6
1. Aim of the document.....	9
1.1 Introduction.....	9
1.2 Purpose and scope of this report	9
2. Data Management Plan and data sharing policy.....	10
2.1 Implementation of DMP and associated policies	11
3. STORE ^{DB} development and maintenance.....	12
3.1 STORE ^{DB} in context	12
3.2 Development and Maintenance of STORE ^{DB}	13
3.2.1 2020-2022.....	13
3.2.2 2022-23.....	13
3.2.3 2023-2024.....	14
3.2.4 2024 – 2025	14
3.3 FAIR and linked data support by STORE ^{DB}	15
4. Development of the Radiation Biology Ontology (RBO).....	16
4.1 Development Process	16
4.2 RBO and FAIR data sharing	18
5. Conclusions and future sustainability	18
6. References	19
Appendix A. WP8 milestones and deliverables	20

List of figures

Figure 1: Data sharing provisions for 58 publications	11
Figure 2 Schematic of the upper classes of the Radiation Biology Ontology	17

Glossary

ALLIANCE	European Radioecology Alliance — European network coordinating radioecology research and infrastructure.
API	Application Programming Interface — A defined contract (endpoints, payloads) that lets software systems communicate programmatically.
BFO	Basic Formal Ontology — Upper ontology providing high-level categories used to align domain ontologies (e.g., OBO).
BfS	Bundesamt für Strahlenschutz — German Federal Office for Radiation Protection (national authority).
BioPortal (NCBO)	Repository and web services for browsing, searching, and using biomedical ontologies.
CC-BY	Creative Commons Attribution — Open licence permitting reuse with attribution.
CC0	Creative Commons Zero — Public-domain dedication; no restrictions on reuse.
DMP	Data Management Plan — A project document specifying how data will be created, documented, shared, preserved, and governed.
DCC	Digital Curation Centre — UK centre (University of Edinburgh) providing guidance/tools for data management and curation.
DOI	Digital Object Identifier — Persistent identifier for digital objects (datasets, articles), resolvable via doi.org.
DoReMi	Low Dose Research Towards Multidisciplinary Integration — EU FP7 Network of Excellence coordinating low-dose radiation research.
ELK	High-performance OWL 2 EL reasoner used for large biomedical ontologies.
ERA	European Radiobiology Archive — Archive preserving historical and contemporary radiobiology study records.
EURADOS	European Radiation Dosimetry Group — Network advancing dosimetry research, standards, and intercomparisons.

FAIR	Findable, Accessible, Interoperable, Reusable — Data stewardship principles enabling machine-actionable reuse.
FIDELIS	FIDELIS European Project to ensure the long-term trustworthiness and sustainability of digital research data repositories across Europe.
GA	Grant Agreement — Binding contract between the European Commission (or agency) and beneficiaries.
GDPR	General Data Protection Regulation — EU regulation governing personal data processing and protection.
GitHub	Cloud platform for Git-based version control, issues, CI, and software/ontology collaboration.
IRI	Internationalized Resource Identifier — Unicode-capable generalization of a URL used to identify RDF/OWL resources.
JSON	JavaScript Object Notation — Lightweight, UTF-8 text format for structured data interchange.
JWT tokens	JSON Web Tokens — Compact, signed tokens (JWS/JWT) for stateless authentication/authorisation.
LSDA	Life Sciences Data Archive (NASA) — Archive for spaceflight and ground analogue life-sciences data (incl. radiation).
MDPI	Multidisciplinary Digital Publishing Institute — Open-access scholarly publisher and journal platform.
MELODI	Multidisciplinary European Low Dose Initiative — European platform coordinating low-dose radiation risk research.
NASA	National Aeronautics and Space Administration — US space agency; custodian of OSDR/GeneLab/LSDA for open science.
NORM	Naturally Occurring Radioactive Materials — Materials with natural radioactivity; includes aspects of TENORM (technologically enhanced NORM).
OAIS	Open Archival Information System — ISO 14721 reference model for long-term digital preservation.
OBO	Open Biological and Biomedical Ontology (Foundry) — Community framework and conventions for interoperable ontologies.
ODK	Ontology Development Kit — Cookie-cutter build/release/QC tooling for OBO ontologies (ROBOT, templates, CI).
OLS	Ontology Lookup Service — EBI search/browse API/service for ontologies (term resolution, autocomplete).

Ontobee	Linked-data server/index for OBO ontologies with term PURLs and statistics.
ORCID	Open Researcher and Contributor ID — Persistent identifier for researchers (e.g., 0000-0002-...).
OSDR	Open Science Data Repository (NASA) — NASA life-sciences open data platform integrating GeneLab and related assets.
OWL	Web Ontology Language — W3C standard (RDF/Description Logic) for representing ontologies and axioms.
PURL	Persistent Uniform Resource Locator — Redirecting URL that provides stable, resolvable identifiers (e.g., OBO PURLs).
REST	Representational State Transfer — Web architectural style using HTTP verbs and resources (JSON APIs).
RBO	Radiation Biology Ontology — OBO ontology describing radiation exposures, dosimetry, assays, and biological effects.
RDF	Resource Description Framework — W3C data model (triples) for graphs/linked data; serialised as Turtle/JSON-LD/XML.
ROBOT	ROBO-tool — Command-line tool for ontology operations (merge, reason, QC, templates) used in OBO/ODK pipelines.
RRID	Research Resource Identifier — Persistent IDs for reagents, cell lines, software, etc., to improve reproducibility.
SHACL	Shapes Constraint Language — W3C language for validating RDF graphs against structural/semantic constraints.
SKOS	Simple Knowledge Organization System — RDF model for concept schemes (labels, synonyms, mappings).
SPARQL	SPARQL Protocol and RDF Query Language — Query language/HTTP protocol for RDF knowledge graphs.
TOP	The Transparency and Openness Promotion Guidelines (TOP) is a policy framework for advancing open science practices. Specifically, TOP includes practices that are intended to increase the verifiability of empirical research claims
UI	User Interface — The interactive layer for human users (web or desktop) to interact with a system
WP(s)	Work Package(s)
WPLs	Work Package Leaders

1. Aim of the document

1.1 Introduction

According to the Grant Agreement (GA) Number 900009 — RadoNorm [1] the objectives of WP8 Communication, Dissemination and Exploitation of Results were to:

- enable two-way communication about the RadoNorm and its results¹ to multiple audiences and show its impact and benefits by addressing and providing possible solutions to challenges of radon and NORM exposures;
- make the RadoNorm results transferable for audiences that may use the new knowledge, data and information in their own work, enable use and uptake of results and maximise the impact of EU-funded research;
- utilise the RadoNorm results in developing, creating and marketing products, processes, services or any other activities (policymaking) to effectively exploit the project's results for society.

In order to implement the objectives above, specifically the second objective, *Task 8.7: Facilitating further use of results* was established to provide data management support within the project through a Data Management Plan (DMP) and further development of the STORE^{DB} database as a platform for FAIR data dissemination. The outputs of the WP were effectively achieved by detailed implementation of FAIR within RadoNorm. The milestones and deliverables for this WP are listed in Appendix A and Deliverable 8.5 relevant are highlighted in red.

The FAIR principles (Findable, Accessible, Interoperable, Reusable) require that research data be described with rich, persistent metadata, deposited in trusted repositories, encoded using community standards/ontologies, and licensed and documented so others (and machines) can discover and reuse them. We prepared and maintained a Data Management Plan (DMP) in which we documented how data would be stored, shared, preserved, and protected, The European Commission's guidance and OpenAIRE recommend making data "as open as possible, as closed as necessary" (i.e., deposit non-sensitive data in trusted repositories, publish machine-readable metadata—ideally under CC0/CC-BY—and apply access controls where justified. Practically, projects are recommended to: (1) write the DMP early (and update it as necessary through the project), (2) choose recognised repositories for long-term deposition and persistent identifiers (DOIs), (3) annotate data with community ontologies/standards (to achieve interoperability), (4) declare licences and access rules clearly in metadata, and (5) retain records in accordance with the Grant Agreement (typically five years after final payment unless otherwise specified).

We discuss below how we met these requirements by

- Developing a DMP and monitoring compliance through the data associated with publication outputs.
- Selecting and developed the STORE^{DB} repository as a recognised trusted repository compliant with FAIR principles for RadoNorm data
- Developing a community standard metadata through the RBO ontology.

1.2 Purpose and scope of this report

The purpose of this deliverable 8.5 is to provide a narrative of the implementation of FAIR data management and dissemination in a major integrated research project with a common policy in force

¹ Results are "any tangible or intangible output of the action, such as data, knowledge and information whatever their form or nature, whether or not they can be protected, which are generated in the action as well as any attached rights, including intellectual property rights" (Source: EC Research & Innovation Participant Portal Glossary).

across the project. There are few examples of such widescale formal planning in projects generating either the volume or the scope and diversity of outputs, which in the case of RadoNorm include experimental and theoretical fundamental science, environmental science, behavioural and sociopsychological studies.

Dissemination activities ensured that project results were made available to relevant external stakeholders such as the scientific community, industry, policymakers, and other end users. These activities played a crucial role in transferring knowledge and increasing the visibility of project outcomes.

2. Data Management Plan and data sharing policy

A Data Management Plan (DMP) operationalises the FAIR principles (Wilkinson et al., 2016) specifying, in advance, how data will be made **Findable** (clear titles, rich metadata, persistent identifiers like DOIs, searchable repositories), **Accessible** (repository choice, access conditions, authentication/embargo policies, and long-term availability), **Interoperable** (use of open, machine-readable formats; community standards; controlled vocabularies and ontologies), and **Reusable** (licensing, provenance, versioning, quality checks, and sufficient documentation/protocols). By defining roles, timelines, storage/backup, and preservation strategies, the DMP turns FAIR from aspirations into concrete, auditable actions across the data lifecycle.

We began our Data Management Plan (DMP) at proposal stage and kept refining it throughout the project. First, we inventoried all data we would create or reuse—types, volumes, formats—and agreed on documentation standards, including community ontologies and mandatory metadata fields. We then set out our working practices: storage locations, access controls, versioning, and backup frequency. We specified how and when we would share data, the repositories to be used, licences (CC-BY, CC-0), and any embargoes with their reasoning, alongside roles and responsibilities, budgets, and timelines. We also built in GDPR/ethics considerations, a retention schedule, quality checks (file validation, metadata completeness, fixity), and the use of persistent identifiers (DOIs, ORCID, resolvable accession IDs and namespace for STORE). Crucially, we treated the DMP as a living document, with the intention of updating it whenever methods, datasets, or partners changed, and using it to drive day-to-day practice (folder structures, naming conventions, etc.).

To streamline this, we used the **Digital Curation Centre (DCC)** tools. With **DMPonline**, we selected the appropriate funder template, co-authored the plan, embedded our institution's guidance, and exported compliant versions for submissions and audits. We mapped activities against the DCC **Curation Lifecycle Model** to cover creation, appraisal, preservation, and reuse, and followed DCC checklists and exemplars to choose open formats, metadata schemas, and trusted repositories, including STORE. For long-term preservation, we designed an **Open Archival Information System (OAIS)-aligned** approach: defining what to keep, where to deposit, and how to capture provenance and version history. Together, these steps turned the DMP from a one-off deliverable into a practical blueprint for FAIR, secure, and reusable data.

Projections of data volumes, types, timelines and themes were collected from the WP leaders in December 2020, reconciled against the deliverables from the GA, and used to build the DMP which was submitted as Deliverable 8.7 in March 2021², MS93.

² Schofield, P.N., Fevrier, L., Madas, B., Salomaa, S., Fronka, A., Perko, T., Martell, M., Oughton, D., Birschwiks, M. (2021): Open Research Data Pilot and project Data Management Plan. Final version as of 31.03.2021 of deliverable D8.7 of project RadoNorm.

2.1 Implementation of DMP and associated policies

The implementation of the DMP and associated policies required consistent application throughout the WPs and partners. In order to achieve this WPL appointed data wranglers. Including the responsible WPLs, 24 wranglers were appointed. The role of the data wranglers was to:

- Act as a data champion for their part of the project
- Provide information and advice to local staff on data management practices and data sharing
- Ensure that data was shared appropriately on publication and to assist with pre-publication data sharing
- Coordinate with WP8 on technical and policy issues as they arose

A training course was held online on 12 Jan 2021 and further meetings, consultation and training with WPLs on March 8th 2021. Wranglers were expected to conduct informal training exercises in their Institutions or WPs using materials provided to them.

General education on FAIR data and data management was provided as part of presentations to RadoNorm Annual meetings in 2021, 2023 and 2025 and a helpdesk email provided.

Through the WPLs and data wranglers the progress of deliverables was monitored through the project and there were no significant changes in the data projected to be uploaded, although there were some changes in schedule. The DMP therefore worked well for the data types, characteristics and volumes throughout and did not require modification to take on board new types of data, for example with unforeseen sizes or unusual ethical concerns.

An audit was carried out on publications uploaded up to 22 August 2025 to assess how much data was placed in the public domain in compliance with the DMP, and how much was retained for justifiable reasons or not shared.

Data Access Handling (n=58)

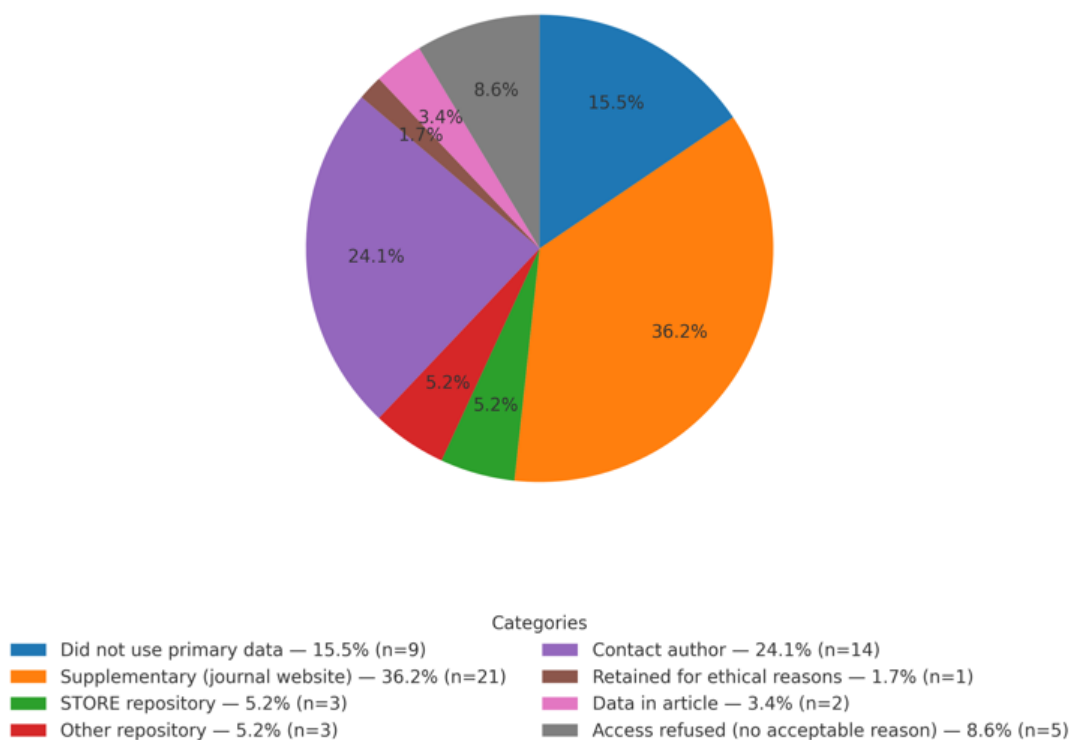


Figure 1: Data sharing provisions for 58 publications

§29.3 of the GA requires data associated with publications to be uploaded to a research data repository with its metadata as soon as possible, and in compliance with the DMP. While the use of supplementary data sites of journals is at least a public space, we were surprised by the failure of compliance on the scale we found during the audit and, using this as a learning opportunity, would recommend that future large-scale projects implement management procedures to explicitly ensure data sharing is treated seriously before papers are submitted for publication as part of supporting a change in attitudes to FAIR data in the scientific community. Large-scale studies across many scientific domains also come to the conclusion that a change in attitudes is slow in coming despite efforts in education and reinforcement. (Hahnel et al., 2024) In the proposal we also undertook to ensure that journals in which RadoNorm data were published should have high Transparency and Openness Promotion (TOP) scores (Grant et al., 2025) where alternatives were possible. This has not been the case, particularly with MDPI published journals, but we recommend a closer study of publication patterns in these journals in future. This is complicated by the scientific considerations of which is the most appropriate journal in which to publish, and is not easy to analyse without detailed consultation.

3. STORE^{DB} development and maintenance

3.1 STORE^{DB} in context

Public databases fall into two main groups: firstly those which archive data of specific types such as for microscopy, gene expression, protein structure, etc. that take specified types and formats of data in a restricted domain; secondly, there are generalised databases which take any kind of data, such as Zenodo, which suffer from metadata with low granularity as they try to cover all domains and data types (Wilson et al., 2021). The STORE^{DB} database is distinctive because it was designed as a broad domain-specific general database that will archive any data type or theme associated with radiobiology, ranging from, for example, detailed quantitative experimental or observational data to documents on sociolegal or sociopsychological aspects of radiation science.

The STORE^{DB} database for Radiation Biology was created in 2009 following the successful completion of the ERA legacy database containing data from the large-scale experiments on animal radiation exposure carried out between the 1950s and the 1990s (Birschwilks et al., 2008)

STORE^{DB} has always been free to access, and permits users to upload and share any type of data including links to data in other databases, in-house datasets and bioresources. Within each *Study* there are *Datasets* and within those individual *Files* or data elements. So, a study might contain many datasets and each dataset many pieces of data, which can be of any type or size and annotated with standard metadata terms for retrieval and searching. Each *Dataset* and data item are assigned a persistent STORE^{DB} accession ID using the STOREDB: namespace and an additional Digital Object Identifier (DOI). The IDs assigned can be used to reference data in publications instead of using journal supplementary information sites, and are much more stable as a consequence. Users can maintain control over its dissemination through Creative Commons licensing and user-defined security and privileges.

STORE^{DB} is housed on the server platform of the Bundesamt für Strahlenschutz (BfS) and consequently, as well as being very secure, is resilient to civil and emergency contingencies. STORE^{DB} was funded under contract numbers 23228 (STORE), 249689 (DoReMi), 662287 (CONCERT) and currently under 900009 (RadoNorm) from the EC Euratom Programme. RadoNorm contributed in the past five years to the further development, adaption and maintenance of STORE^{DB} as the BfS continued to generously support management of updates and the technical platform of database and web server.

A key strength of STORE^{DB} is its support of FAIR data sharing which it does though adhering to standards of access and metadata as discussed below. This section is in two further parts: A description of the maintenance and development of the technical platform, a description of the integration of STORE^{DB} into the linked open data environment, and aspects of STORE^{DB} that specifically support FAIR data.

3.2 Development and Maintenance of STORE^{DB}

3.2.1 2020-2022

During this period **STORE^{DB}** was upgraded from the legacy v3 stack to a modernised **v4** capable of handling large (“big”) data and sustaining secure, reliable operations for RadoNorm. This section reports what was implemented after the initial 1 Mar 2021 delivery, removing historical feasibility detail and focusing on completed work.

Core achievements. (1) The legacy system analysis and **security review** were completed; only minor findings emerged from penetration testing and log analysis, with no evidence of exploitable issues. (2) A **new large-file upload pipeline** was designed and deployed using the TUS protocol (resumable uploads) with an **Uppy** front end and a **Spring Boot** back end, containerised with Docker and integrated with the new STORE API; this removes the prior ~1 GB ceiling and supports pause/resume. (3) A **development server** was provisioned and the application successfully exercised on an updated stack (e.g., Java, Tomcat, PostgreSQL), with **v4 placed into production** and minor admin-UI bugs fixed. (4) A structured **project plan** drove the work through prototypes, security fixes, Java/DB upgrades, REST API integration, and UI builds, most tasks recorded as completed.

Security posture & recommendations. The team enforced **HTTPS** and strengthened headers, using Spring-Boot’s security frameworks where possible. For programmatic uploads and future public web services, the plan highlights **ORCID-based authentication** with **JWT** tokens and equivalent checks to the legacy upload flow (user validation, study/dataset ownership), ensuring parity of safeguards as features move to the API layer. No redesign was required solely for security reasons.

User requirements & data model adjustments. A **questionnaire** (Dec 2020) and **workshop** (Jan 2021) captured RadoNorm partner needs. Findings: expected data volumes and rates pose no scalability concern; >1 GB files necessitated the new upload system; some UI refinements would improve usability. To support project organisation, the team proposed a **WP-centric structure** (one study per Work Package, datasets aligned to tasks/deliverables) and added two **new annotation fields** (WP deliverable/task) in Jan 2020 to improve discoverability and governance.

Outcome With **v4** operational, resumable large-file uploads in place, and security tightened, the MS97 objective is met. The plan calls for continued support to WP “data wranglers” setting up projects, routine feedback collection after several months of use, discussed above.

3.2.2 2022-23

Headline achievements. We delivered a year of steady, behind-the-scenes reliability work and interoperability upgrades. Key outcomes were (i) a full security refresh of the runtime stack, (ii) homepage changes to support robust citation and remove fragile integrations, and (iii) the first pass of **RBO ontology** mapping across STORE attributes to enable ontology-aware search and integration via the API.

Platform hardening. The web tier was upgraded to **Tomcat 10**, the JVM to **OpenJDK 20.0.1**, and legacy **Twitter4j** code was removed. These changes required source updates and a complete recompile/test cycle, improving security posture and maintainability.

Homepage & citation. The embedded Twitter feed was removed (new API charges made it unsustainable), reducing external dependencies and page fragility. Guidance on **Research Resource**

Identifiers (RRIDs) was added so that datasets can be cited more formally and tracked across the literature.

Interoperability (RBO). Work began to **map Radiation Biology Ontology (RBO)** terms to STORE fields/values. Crucially, **RBO IDs are now exposed via the STORE API**, allowing other databases to discover and query STORE records at the ontology level—an important step toward FAIR, machine-actionable metadata. The current UI relies on the EBI ontology service, which is being redeveloped; therefore, the ontology front-end will be reworked or moved to a new service in the coming year to give users a smoother term-selection experience.

Priorities were then to complete the **new ontology web service integration** in the UI and continue incremental maintenance/security updates. Together, the year's work strengthens security, simplifies the user-facing site, and lays the groundwork for richer, ontology-driven discovery of STORE content.

3.2.3 2023-2024

Headline achievements. The year focused on dependable operations for RadoNorm users, rapid support, and smoothing the user experience. The team (i) provided hands-on assistance to researchers on structuring and uploading datasets and validated new user accounts, (ii) cleared a queue of minor UI defects and refreshed contact information, (iii) released an up-to-date help video and ensured it embeds reliably in modern browsers, and (iv) advanced ontology use by annotating selected data attributes and upgrading the ontology lookup service.

Platform & usability improvements. The help video was re-hosted over **HTTPS** so it now embeds correctly in Google Chrome, and a new walkthrough reflects the current UI for **upload** and **search**, reducing onboarding friction for first-time contributors. Routine UI bug-fixing removed small but persistent usability irritants, while contact-detail updates keep support channels accurate and responsive.

Interoperability (ontology integration). Progress continued on ontology-aware metadata: **selected data attributes were annotated with ontology terms**, and the **Ontology Lookup Service (OLS) API** integration was upgraded to the latest version, resolving pop-up issues in the term selector. These steps improve consistency of annotations and lay groundwork for richer, machine-actionable discovery.

Operations & user support. Day-to-day email support guided RadoNorm contributors on dataset structure and submission, while account checks safeguarded repository integrity. Together, these activities sustained reliable service while incrementally improving metadata quality and user experience.

Next priorities logically follow: extend ontology coverage across more attributes, continue small UI refinements informed by user queries, and keep the training materials (video/help pages) aligned with the evolving interface.

3.2.4 2024 – 2025

Headline achievements. This year centred on dependable operations for RadoNorm users and a critical upgrade to persistent-identifier services. The team (i) provided continuing hands-on support to users on structuring and uploading datasets and validated new user registrations, (ii) cleared a tranche of small user-interface defects, and (iii) completed a **DOI system migration** to support **XML v4**, ensuring uninterrupted minting of DOIs for studies and datasets after the previous API was deprecated in 2024.

Platform stability and interoperability. The **DOI upgrade to version 4** is the substantive technical change this cycle: the application was adapted to the new provider specification so that automatic DOI generation continues seamlessly for new and updated records. This safeguards citation, discoverability, and long-term tracking of STOREDB outputs and removes a looming operational risk created by the provider's 2024 deprecation.

Operations & user support. Day-to-day repository running received sustained attention: the team supported contributors by email on dataset organisation and submission, and performed checks on new accounts to preserve repository quality and security. In parallel, they addressed **various minor UI bugs**, removing small but persistent irritants that can slow data deposit or discovery. Together, these activities kept the service reliable while smoothing user experience for routine tasks.

Overall. The year's work prioritised resilience and continuity—most visibly through the DOI API migration—while maintaining responsive support and incremental UI hygiene. This positions STOREDB to continue issuing trustworthy DOIs and to meet users' day-to-day needs without disruption.

Currently STOREDB contains 206 studies and 320 datasets, with >1000 data elements and files.

3.3 FAIR and linked data support by STORE^{DB}

The STORE^{DB} radiobiology database is set up to make data **Findable** and **Accessible** by design. A key strength of STORE^{DB} is its high transparency regarding the use of standards and data access, as well as that the details may be obtained from the FAIR sharing platform website (www.fairsharing.org/). The accession IDs assigned to individual datasets are resolvable and registered with Identifiers.org (Bernal-Llinares et al., 2021) while the STORE^{DB} repository itself is registered with the re3data registry of research data repositories (Pampel et al., 2023). Furthermore, STORE^{DB} has been granted a persistent research resource identifier (**RRID:SCR_023207**) from Scicrunch (<http://www.rrids.org/>).

More recently, the FIDELIS initiative under the European Open Science Cloud (EOSC, <https://eosc.eu/horizon-europe-projects/fidelis/>) has been created to establish a network of trusted digital repositories, where STORE^{DB} has provisional membership along with the other founder members of the network, with full membership expected to be formally established in 2027 upon the maturity of FIDELIS operations.

Findable & Accessible. Every study/dataset in STORE^{DB} gets a stable landing page with a **DOI** and a resolvable accession ID for citation and discovery. Records are indexed with relevant **rich, structured metadata**. Search is helped by **ontology-backed term pickers** (e.g., via an OLS integration), so users can discover content by concept rather than just keywords, and machines can query consistently via the **API**. Public datasets are downloadable over **HTTPS**; restricted records clearly state the access level and provide a **request path** (contact route/conditions). Large files are reliably captured using **resumable uploads** (e.g., TUS/Uppy), so complete datasets—not just summaries—are actually preserved. Access follows “as open as possible, as closed as necessary” and users are encouraged to use Creative Commons licences CC-BY and CC-0 for data and appropriate licenses for software, such as MIT or GPL [GNU General Public License \(GPL\)](https://www.gnu.org/licenses/gpl-3.0.html).

Interoperable. Metadata fields are aligned to community standards so they are machine-readable and composable across repositories: radiation concepts use, for example, the **Radiation Biology Ontology (RBO)**; **Units Ontology (UO)** for Gy/Sv/Bq; **NCBITaxon** for organism; **Uberon** for anatomy; and common assay/endpoint terms (see below). STORE^{DB} stores and exposes the **identifiers** (PURLs), not just labels, and surfaces them via the UI and **API/exports** (CSV and machine-actionable serialisations such as JSON-LD when needed). This lets downstream systems unambiguously join on semantics (e.g., all datasets annotated with “absorbed dose” or a specific “Cs-137 external exposure”) without brittle string matching.

Reusable. Each record includes **clear licensing**, and captures **provenance**: method/protocol references (preferably DOIs) or descriptions, processing notes, instrument details, and version history so others can understand and audit how the data were generated. The platform supports **dataset-level versioning** and common QA steps to raise baseline quality. Guidance on **citation (DOIs)** and **RRIDs**

helps credit creators and improves traceability across the literature. Practically, these features—together with robust upload, stable identifiers, and ontology-driven metadata—mean STORE^{DB} deposits are not only findable and retrievable but also understandable and legally/technically ready for reuse in new analyses, meta-studies, and regulatory submissions.

4. Development of the Radiation Biology Ontology (RBO)

Starting in 2020 we have co-led an open consortium to develop a new metadata standard, the Radiation Biology Ontology ([RBO](#)), through the Open Biological and Biomedical Ontologies ([OBO](#)) Foundry. It is explicitly aimed at organising, describing, and archiving data relevant to radiobiology to support FAIR discovery and data integration from public resources.

The scope of the ontology is very broad and includes not only concepts for describing radiation and its effects on biota on Earth and in space but a wide range of medical, legal, social and psychological concepts related to radiation safety, radiation measurement, for example radon, environments and their characteristics and sources of radiation; anthropogenic, natural terrestrial and natural cosmic. See Figure 2 for an example of the upper level classes. The initial use-cases were for STORE^{DB} as a general repository and the NASA GeneLab. These have now been extended to NASA Life Sciences Data Archive (LSDA) and more recently NASA Open Science Data Repository (Gebre et al., 2025). Co-development has had invaluable advantages as the needs of STORE^{DB} and the NASA databases overlap and complement, providing a text bed for the adoption of RBO by other databases containing data related to radiobiology, enhancing the possibility of machine-driven data integration between resources.

RBO is hosted openly on GitHub under the Radiobiology-Informatics-Consortium organization, with public issues for change requests. Releases are distributed via OBO PURLs in multiple serializations (OWL, OBO, JSON), and the license is CC BY 3.0. Current releases are indexed by Ontobee/OLS/BioPortal. The latest Ontobee metadata shows version IRI for 2025-05-30. All users are invited to use the Issues tracker on GitHub for questions and input in to the ontology

4.1 Development Process

We began in 2020 by transforming many of the terms from the National Council on Radiation Protection and Measurement into concepts that can be formally related to existing OBO Foundry classes or attributes. We then identified and imported into the RBO existing OBO Foundry classes that have obvious relevance for radiation biomedicine (concepts from the [Gene Ontology](#) dealing with molecular and cellular responses to radiation). Finally, we scrutinized datasets from investigations of radiation effects held in STORE^{DB}, NASA GeneLab and LSDA repositories and added additional classes, instances, and attributes into the RBO that should be used to annotate these data.

Subsequently RBO was bootstrapped and is maintained using the OBO “ontology starter kit”/Ontology Development Kit (ODK) workflow (Matentzoglou et al., 2022): a repo scaffold with standard Make/ROBOT pipelines, continuous integration, PURL configuration, and community curation via GitHub. This represents a continuous build integration, and a metadata-first approach—hallmarks of ODK practice (templates, imports, ROBOT/CI etc). We have emphasised continuous integration and community workflow to ensure consistency and rapid iteration—important for a domain that spans physics, biology, environment and clinical research

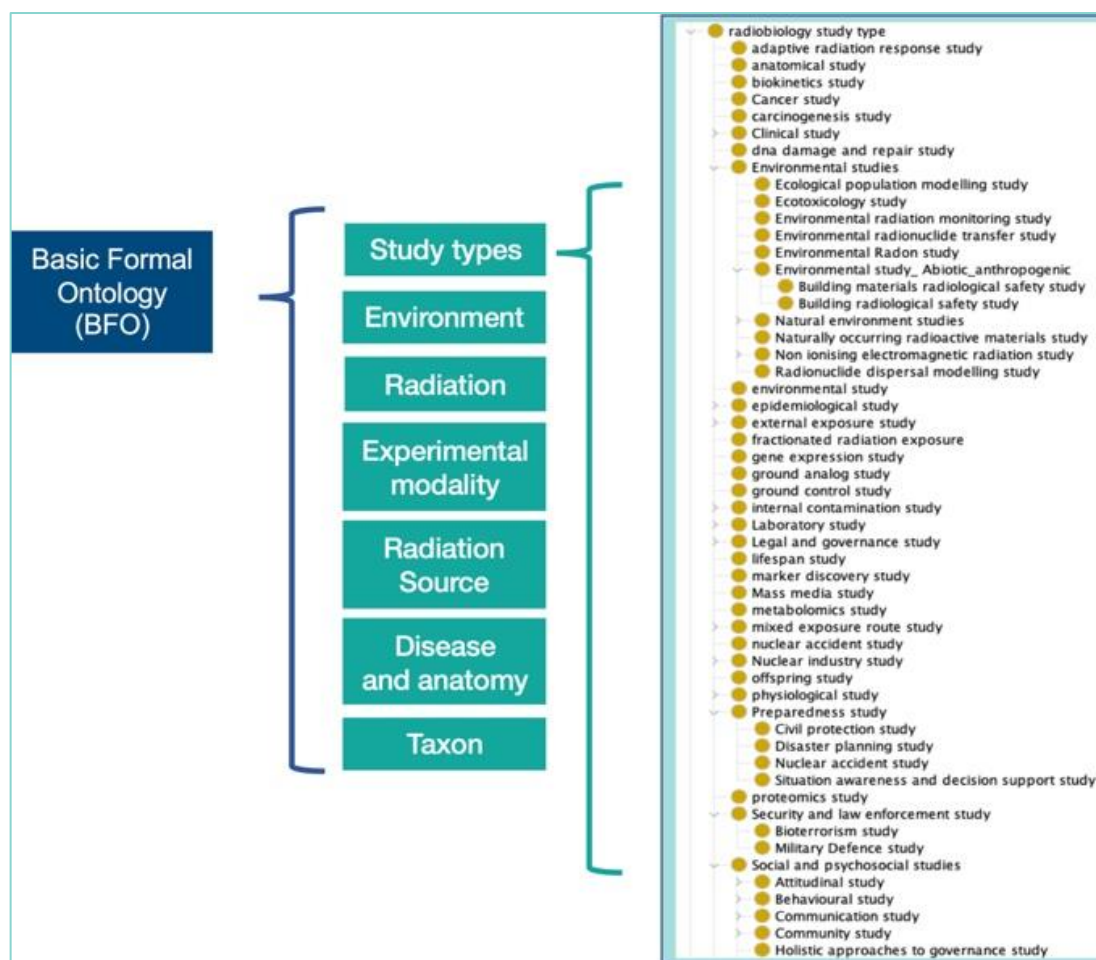


Figure 2 Schematic of the upper classes of the Radiation Biology Ontology

RBO currently contains 9,128 classes, 301 object properties, 258 annotation properties, and 367 individuals. Of these, 414 classes are native RBO classes generated manually by experts and with hand-written definitions and curated axioms; the remainder are imported from reference ontologies.

RBO reuses a broad set of OBO ontologies to ensure semantic interoperability. Major contributors (class counts shown) include: GO (3,356), UBERON (2,033), CHEBI (724), ENVO (676), UO (528), CL (469), OBI (95), PATO (169), NCBITaxon (356), BFO (19), DOID (81), PR (74), PO (34), BSPO (object properties), RO (relation set), among others. The OBO page lists formal dependencies on BFO, CHMO, ENVO, OBI, PATO, RO, and UO. These imports enable consistent use of anatomical, environmental, chemical, unit, and process terms within radiobiology datasets.

Releases are distributed via OBO PURLs in multiple serialisations (OWL, OBO, JSON), and the license is CC BY 3.0. Current releases are indexed by Ontobee/OLS/BioPortal. The latest Ontobee metadata at the time of writing shows version IRI for 2025-05-30. Stable PURLs, indexed in OBO/OLS/BioPortal, and CC BY licensing make terms resolvable and reusable by both humans and machines. Extensive reuse of foundational OBO ontologies and Relations Ontology, according to Foundry principles, enables cross-dataset linking and reasoning. We provide versioned releases, provenance via GitHub issues/PRs, and community curation to improve clarity and traceability—key to reuse in repositories such as STORE^{DB} and OSDR.

4.2 RBO and FAIR data sharing

Community ontologies such as RBO operationalise FAIR by supplying **globally unique, resolvable IRIs** (PIDs) and **machine-interpretable axioms** that make metadata computable, not just human-readable. In practice, domain terms are published as **OWL/RDF** with formally specified **logical definitions** (genus–differentia, RO relations), **typing/constraints** (domain/range, disjointness), and **cross-references** to external identifiers (e.g., CHEBI, UBERON, GO). This enables **semantic interoperability**: heterogeneous datasets annotated with ontology IRIs can be merged into a **knowledge graph**, queried via **SPARQL**, and subjected to **automated reasoning** (ELK/HermiT) to infer class membership, propagate annotations over part-of/develops-from hierarchies, and detect inconsistencies. Robust **versioned PURLs**, declared **provenance** (PROV-O), and **licensing metadata** support **Reusability**, while **JSON-LD** and **schema.org** overlays make annotations portable across APIs, search indices, and data portals.

Critically, community process is as important as formalism, and we took great care to involve RadoNorm participants as well as investigators outside the project in its building, for example ALLIANCE and EURADOS were both consulted in bringing in environmental and dosimetric concepts. Ontologies curated in the open (OBO/ODK workflow, ROBOT QC, DOSDP templates) deliver **governed change management** (issues/PRs), **continuous validation** (term label/definition checks, orphan detection), and **release pipelines** that guarantee stable IRIs across versions. **Alignment to upper frameworks** (BFO + RO) and use of **controlled synonym types** (exact/broad/narrow/related, SKOS) reduce ambiguity; explicit **mappings** (skos:exactMatch/closeMatch) support crosswalks to legacy vocabularies; and **SHACL/Shex** shapes enforce annotation profiles at deposit time. The result is **Findable** (rich, indexed, typed descriptors), **Accessible** (resolvable IRIs and open serialisations), **Interoperable** (shared semantics across repositories/tools), and **Reusable** (unambiguous, provenance-rich, logically consistent) data—at scale and with guarantees that automated agents can validate, integrate, and reason over them.

5. Conclusions and future sustainability

The work carried out under RadoNorm has secured the development and useability of the STORE^{DB} database not only for the project but for the entire radiobiology community; the work has therefore had marked added value over and above dissemination and exploitation of project-derived data, and will have significant impact on radiation-associated sciences, policymaking and society as a whole. It provides an example of a data sharing strategy for a large integrated project, and will still have a presence and continued impact many years after the end of RadoNorm.

The BfS has undertaken to support the STORE^{DB} as an institutional project and it therefore has a sustainability plan for the future which sets it aside from most other repository resources. The long-term storage policy for STORE^{DB} is to have data available live for at least ten years and then to have it archived offline (available on request) for at least 15, or indefinitely depending on prevailing circumstances in 25 years' time. Because STORE^{DB} is maintained by the BfS, its security, backup, delocalised georesilience and maintenance are those of the German Federal Government platforms which is a gold standard for such resources.

This work could not have been carried out without the support of the BfS, their Information Technology Unit, and Dr Mandy Birschwilks. We also acknowledge the continued and dedicated support of the database developer Michael Grünberger for his invaluable and long-term work on the STORE^{DB} project.

6. References

- Bernal-Llinares, M., Ferrer-Gomez, J., Juty, N., Goble, C., Wimalaratne, S. M., & Hermjakob, H. (2021). Identifiers.org: Compact Identifier services in the cloud. *Bioinformatics*, 37(12), 1781-1782. <https://doi.org/10.1093/bioinformatics/btaa864>
- Birschwilks, M., Adelman, C., Gruenberger, M., Schofield, P., Tapio, S., & Grosche, B. (2008). New life for old data-The European Radiobiological Archives and the e. ERA database. *Radioprotection*, 43(5), 235.
- Gebre, S. G., Scott, R. T., Saravia-Butler, A. M., Lopez, D. K., Sanders, L. M., & Costes, S. V. (2025). NASA open science data repository: open science for life in space. *Nucleic Acids Res*, 53(D1), D1697-D1710. <https://doi.org/10.1093/nar/gkae1116>
- Grant, S., Corker, K. S., Mellor, D. T., Stewart, S. L. K., & Cashin, A. G. (2025). An Update to the Transparency and Openness Promotion Guidelines. *MetaArxiv* https://doi.org/10.31222/osf.io/nmfs6_v2
- Hahnel, M., Smith, G., & Campbell, A. (2024). *The State of Open Data 2024: Special Report Bridging policy and practice in data sharing*. https://digitalscience.figshare.com/articles/report/_b_The_State_of_Open_Data_2024_Special_Report_b_Bridging_policy_and_practice_in_data_sharing/27337476
- Matentzoglou, N., Goutte-Gattat, D., Tan, S. Z. K., Balhoff, J. P., Carbon, S., Caron, A. R., et al. . (2022). Ontology Development Kit: a toolkit for building, maintaining and standardizing biomedical ontologies. *Database (Oxford)*, 2022. <https://doi.org/10.1093/database/baac087>
- Pampel, H., Weisweiler, N. L., Strecker, D., Witt, M., Vierkant, P., Elger, K., . . . Petras, V. (2023). re3data - Indexing the Global Research Data Repository Landscape Since 2012. *Sci Data*, 10(1), 571. <https://doi.org/10.1038/s41597-023-02462-y>
- RadoNorm Grant Agreement Number 900009 — RadoNorm as from 13.05.2020
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., .et al. . (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wilson, S. L., Way, G. P., Bittremieux, W., Armache, J. P., Haendel, M. A., & Hoffman, M. M. (2021). Sharing biological data: why, when, and how. *FEBS Lett*, 595(7), 847-863. <https://doi.org/10.1002/1873-3468.14067>

Appendix A. WP8 milestones and deliverables related to STORE

Following milestones and deliverables were planned for the WP8 Communication, Dissemination and Exploitation of results related to STOREDB (information includes number, title and month to be developed):

Milestones:

- MS91 Establishment of project website (external –project repository, internal – project place): M3 and regular updates
- MS92 Updates of Strategy and plan for communication, dissemination and exploitation of results: M12, 24, 36, 48
- **MS93 Data Management Plan: M6 and yearly update DMP document**
- MS94 Establishment of web based social media for RadoNorm: Twitter, LinkedIn: M3 and regular updates
- MS95 Printed and digital materials' reports: M6, and updates every 6 months
- MS96 Stakeholder mapping and establishments of stakeholder networks: M4, and yearly updates
- **MS97 Upgrading of STORE^{db} platform for easy handling of big data: M6 and yearly updates**
- MS98 Proceeding from project events: M12, 24, 36, 48, 60
- MS99 Report from online events: M12, 24, 36, 48, 60

Deliverables

- D8.1 Strategy and plan for communication, dissemination and exploitation of results (M3)
- D8.2 Strategy and plan for communication, dissemination and exploitation of results update (M36)
- D8.3 Printed and digital materials' final report (M60)
- D8.4 Report on stakeholder engagement (M59)
- **D8.5 RadoNorm STORE^{db} report (M60)**
- D8.6 Communication, dissemination and exploitation final report (M60)
- D8.7 Open Research Data Pilot and Data Management Plan (M6)