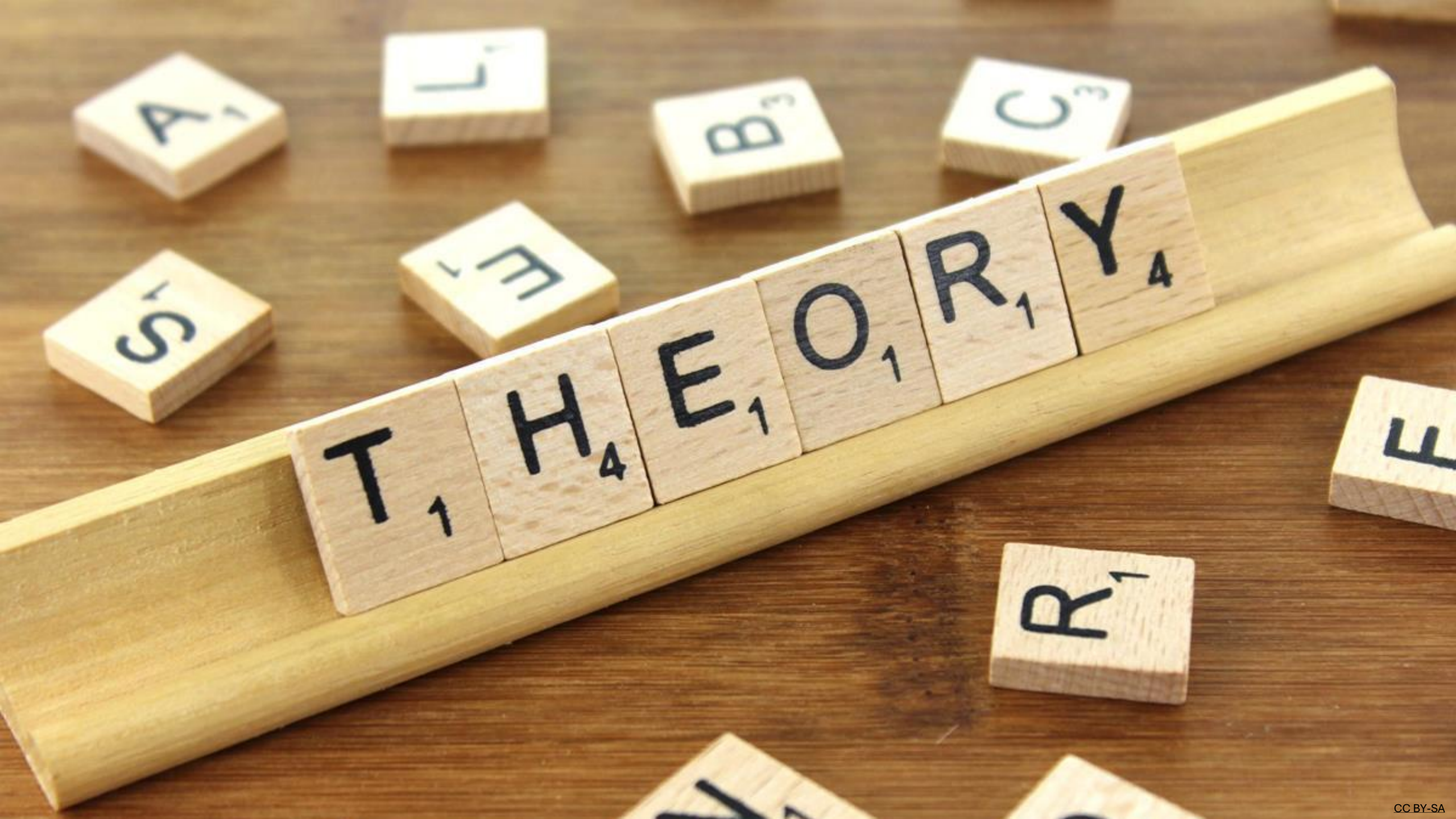
# Unlocking Research Data: Best Practices for Data Professionals

**Lindsay Dowling**
Open Research Support Unit Lead, TU Dublin
https://orcid.org/0000-0002-7642-5326
Lindsay Dowling | LinkedIn

CC BY NC SA

CC BY-SA-NC

- Open Research Support Unit Lead, Technological University Dublin
- Qualified (Research) Data Steward
- Background in Copyright & Rights Management
- Co-Chair European Research Manager's Open Science Group
- EOSC-Ireland National Task Force member




CC BY-SA

# Open Research (Science)

*As open as possible, as closed as necessary*

Conducting scientific or scholarly investigations in a transparent and collaborative manner, often by making research outputs, such as data, methods, and tools, freely accessible to the public, at various stages of the research lifecycle. The aim of open practice is to maximise the potential benefits of research, and research processes, to other researchers, funders, and society.

Photo by Jonathan Cooper on Unsplash

# WHY OPEN SCIENCE ?

## Efficiency

Greater access to scientific inputs and outputs can increase scientific productivity through reducing duplication, allowing **more research from the same data** and multiplying opportunities for domestic and global participation in the research process.

## Quality & integrity

Open access to scientific outputs allow for **greater evaluation and scrutiny** by the scientific community which means more accurate replication and validation of research results. Openness to data contributes to maintain science's self-correction principle.

## Innovation

Open science can **reduce delays in the re-use of scientific research** including articles and data, and promote a swifter path from research to innovation to produce new products and services.

## Economic benefits

**Science plays a key role in today's knowledge economies** and increased access to research results, including data, can positive impact not only scientific systems but also innovation.

## Public disclosure & engagement

Science, often publicly funded, should be publicly accessible **to promote a greater awareness** among citizens and to build public trust and support for public policies and investments in research. Open science also promotes citizen science in experiments and data collection.

## Global benefits

**Open science promotes collaborative efforts** and faster knowledge transfer for a better understanding of global challenges and wicked problems.

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD
GRAPHICS BY GAELEN MINNOCK

This work is licensed under a Creative Commons Attribution 4.0 International License.

Photo by DDP on Unsplash

# Copyright

A **legal** right that gives authors control over how their academic work - like articles, books, or research papers - is used by others. It means they can decide who can copy, share, or change their work, and under what conditions.

In **open research**, authors often use special licenses ('copyleft') to let others use their work freely, as long as they give proper credit.

*An **unregistered** right – automatically conferred.*

# Licencing for Open Research

Creative Commons Licences

provide a legal means to grant copyright permissions to a publicly available work, while ensuring proper attribution (citation). They provide a simple and standardised way to give your permission to share and use your work on conditions which you choose.

| Licence type | Credit must be given | Derivatives & adaptations allowed | Adaptations under same terms | Commercial use |
|---|---|---|---|---|
| CCO/Public Domain | ✗ | ✓ | ✗ | ✓ |
| CC-BY | ✓ | ✓ | ✗ | ✓ |
| CC-BY-SA | ✓ | ✓ | ✓ | ✓ |
| CC-BY-NC | ✓ | ✓ | ✗ | ✗ |
| CC-BY-ND | ✓ | ✗ | ✗ | ✗ |

# Is research data protected under Copyright/IP Law?

## In most cases, no, because:

Intellectual property (IP) assigns property rights to certain non-rival resources, which are considered original *acts of creation*, such as literary, scientific and artistic work .

IP law does *not* protect data since the collection of information and facts is not considered an *act of creation*.

# So…

Research data itself is usually not protected by copyright.

However, the way data is organised, described, or presented, i.e., the *manifestation,* can be protected.

- Researchers can use open licenses for datasets (Creative Commons) to enable data reuse, while still getting credit (citation) for their work.
- By means of Data Sharing Agreements, researcher can freely determine the terms and conditions of data sharing, access and use by third parties, e.g. for restricted data.

# When can copyright apply?

- Copyright protects a database only if its structure or selection is original - meaning the creator made creative choices in how the data is selected, arranged, or presented.
- The individual data entries are not protected by copyright.
- Examples of copyrightable Databases:
  - A curated dataset with a unique thematic structure.
  - A database with a creatively designed user interface or layout.
  - A collection of annotated or interpreted data with original commentary.
- Who owns the copyright?
  - Usually, the individual or institution that created the original structure or arrangement.

IBSO
Irish Building Stock Observatory

Home | BER Status | Retrofit Status | Dwelling | Heat Generation | Data Validation | Findings | Map | About

The Irish Building Stock observatory (IBSO) platform uses processed Building Energy Ratings (BERs) also known as Energy Performance Certificates (EPCs) downloaded from the Sustainable Authority of Ireland (SEAI) to present expert and non-expert users with visualisation metrics and data pertaining the Irish Housing Stock.

The platform barometers Irelands progress toward the national retrofit targets, published in Ireland's national Climate Action Plan 2021 requiring, by 2030, 1 in 4, totalling 500,000 homes to be retrofitted to a B2 energy rating or higher along with the installation of heat pumps in 1 in every 5, totalling 400,000 homes. The Irish Building Stock Observatory (IBSO) measures and visualises progress towards these targets by dwelling type and location, cataloguing the evolution of energy upgrades in the stock over time. The data on this platform is updated quarterly. As time progress, more quarterly data will be added as well as building data and embodied carbon databases.

This IBSO links the BER database with census small area codes and electoral boundary files to map areas that have high concentrations of dwellings that may be targeted for retrofitting.

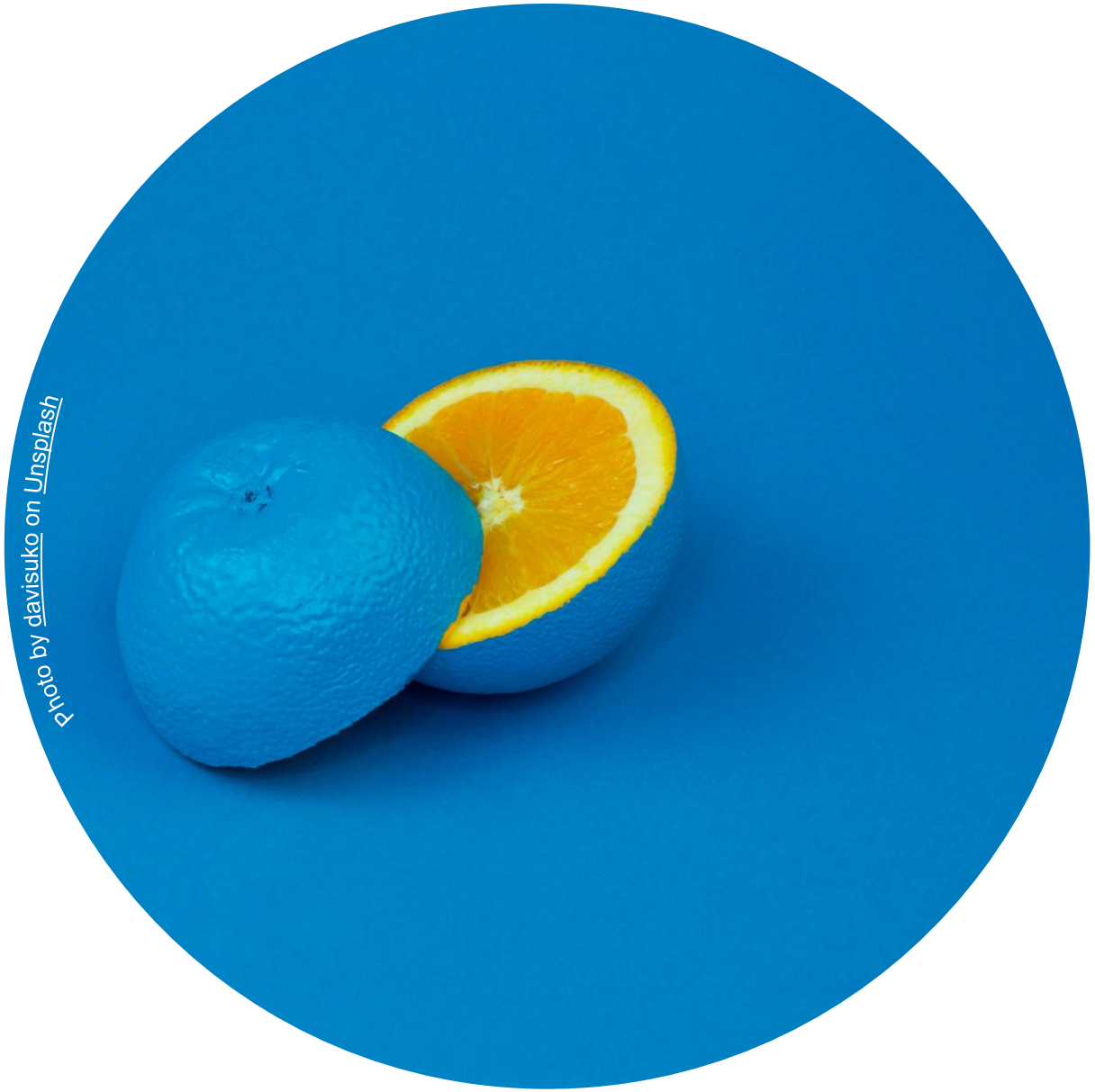


Photo by davisuko on Unsplash

If the structure of your database is not an original creation, you can still protect its content under the **'sui generis'** right.

- A special legal protection **in the EU** for databases that required a substantial investment in obtaining, verifying, or presenting data. This is not a globally recognised right, so is jurisdiction dependent.
- It protects the structure and effort behind compiling the database (not the data).
- Lasts for 15 years from the date of completion or public availability. Each substantial update can reset the clock.
- The owner can prevent extraction or reuse of **substantial** parts of the database without permission (through licencing).
- An **unregistered** right.

# Core Principles for Research Data

**Open Data**

Data that is freely available for anyone to access, use, modify, and share without restrictions, typically under an open license.

**FAIR Data**

**Findable, Accessible, Interoperable, and Reusable**. It emphasises good data management practices to ensure data can be reliably discovered and used by both humans and machines, *regardless of whether it is open or restricted*.

Photo by Alora Griffiths on Unsplash

# RESEARCH DATA – OPEN BY DEFAULT

OpenAire

# FAIR Data, Open Data.

**Data can be FAIR but not open. For example, data could meet the FAIR principles but be private or only shared under certain restrictions.**

| | |
|---|---|
| **'Findable'** | i.e. discoverable with metadata, identifiable and locatable by means of a standard identification mechanism |
| **'Accessible'** | i.e. always available and obtainable; even if the data is restricted, the metadata is open |
| **'Interoperable'** | i.e. both syntactically parseable and semantically understandable, allowing data exchange and reuse between researchers, institutions, organisations or countries; and |
| **'Reusable'** | i.e. sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible and the least cumbersome integration with other data sources. |

**Open data may not be FAIR. For example, publicly available data may lack sufficient documentation to meet the FAIR principles, such as licensing for clear reuse.**

Photo by Chris de Lima on Unsplash

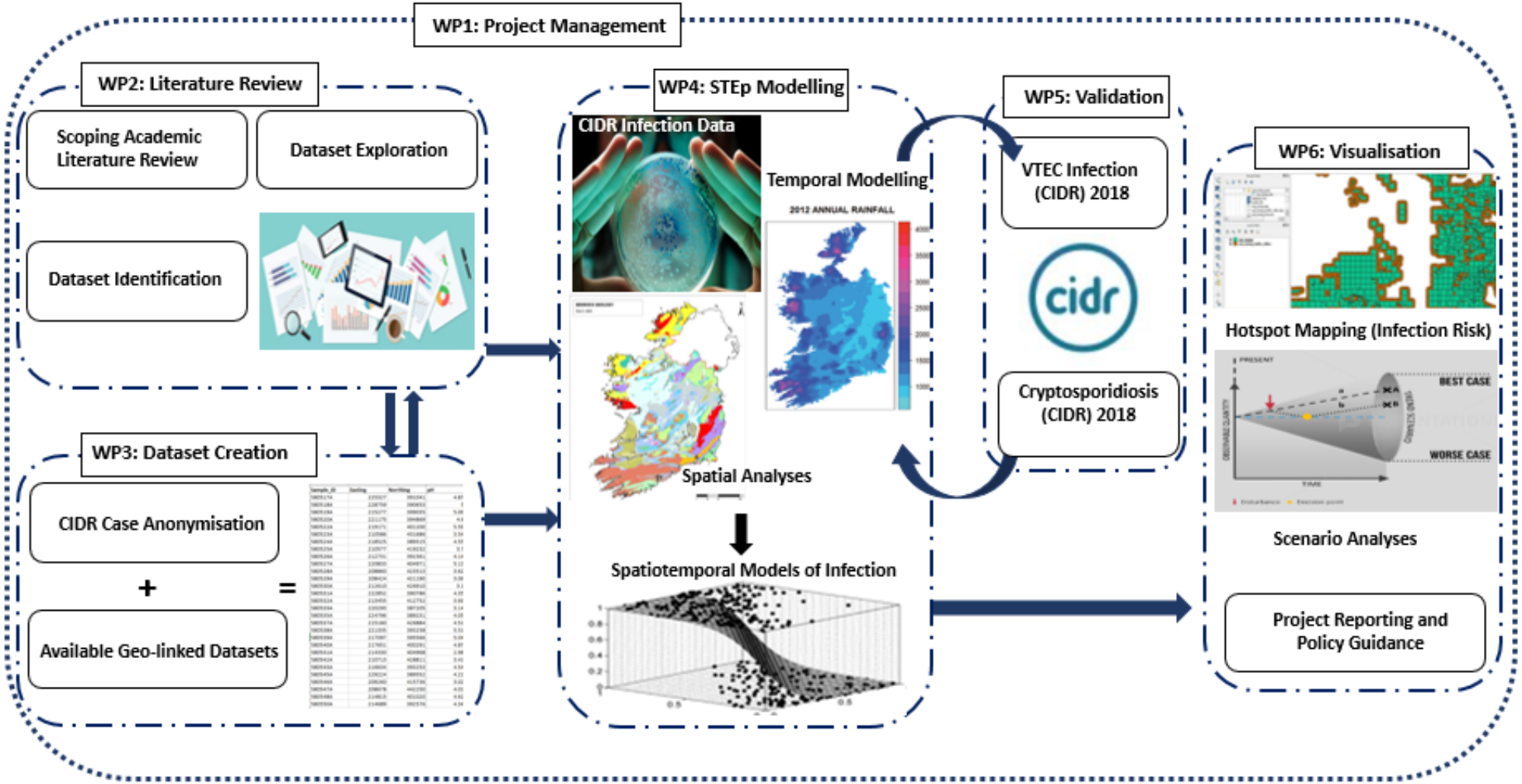
Dual licensing means offering the same research output (like software, data, or publications) under two different licenses - one for non-commercial use (e.g., by academics or the public) and another for commercial use (e.g., by companies).

This allows researchers to **support open research** while still retaining the option to generate revenue from commercial applications.

# A Note on Dual Licencing

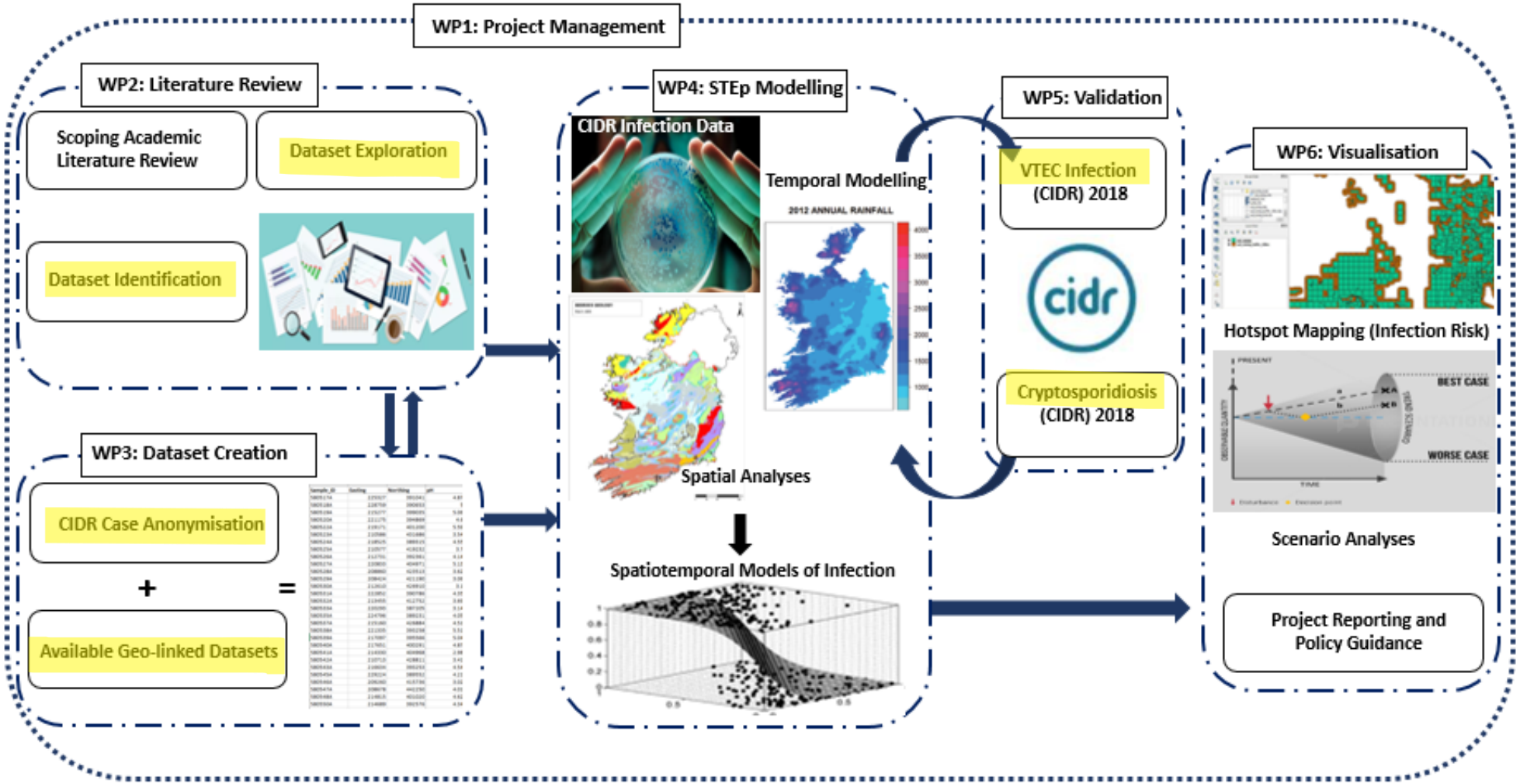
CC BY-NC-ND

# Spatio-Temporal Epidemiology of Waterborne Infections in the ROI: VTEC and Cryptosporidiosis (STEP-WISE)



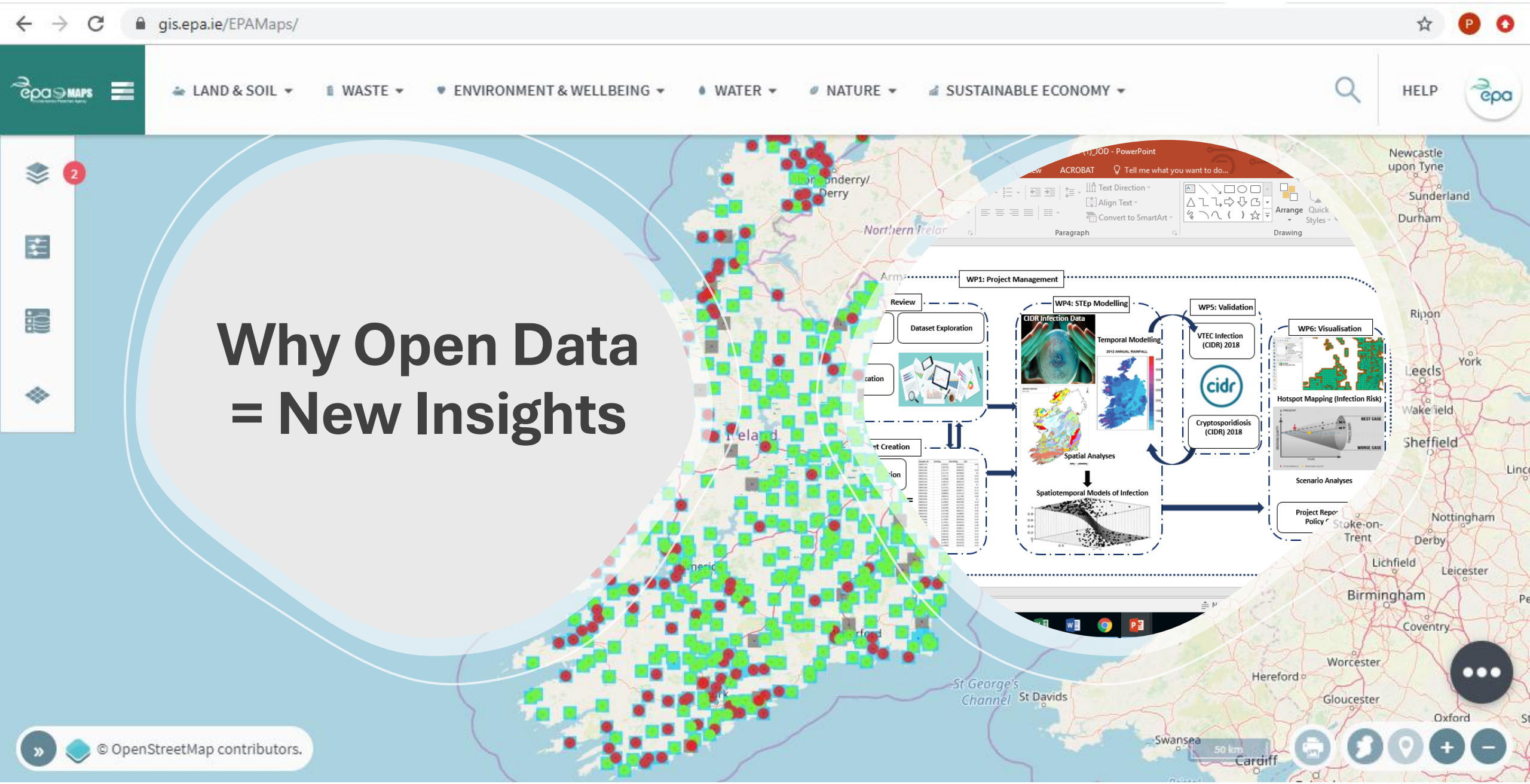

CC BY-NC-SA – Dr Paul Hynds

# Open Research (Data) in Action...




CC BY-NC-SA – Dr Paul Hynds

# Why Open Data = New Insights




CC BY-NC-SA – Dr Paul Hynds

# How to find and use research data

Photo by Agence Olloweb on Unsplash

# Data Repositories

**Purpose of Data Repositories**

Data repositories store, manage, and share research datasets ensuring accessibility and long-term preservation.

**Discipline Organisation**

Repositories are organised by academic discipline, helping researchers find relevant, specialised datasets efficiently.

**Metadata and Documentation**

Repositories provide metadata and documentation to facilitate dataset reuse and ensure data quality standards.

Photo by Claudio Schwarz on Unsplash

JOINT RESEARCH CENTRE

# EUROPEAN SOIL DATA CENTRE (ESDAC)

EAN COMMISSION > JRC > ESDAC

Search

Soil Observatory

SO Home

ROPEAN SOIL DATA CENTER

DAC HOME

IL DATA

TASETS

PS & DOCUMENTS

IL KNOWLEDGE

EMES

TWORKS & COOPERATIONS

OJECTS

GHLIGHTS

All Highlights

Projected SOC changes (data) by 2070 under the RCP4.5 scenario

Data on Aboveground biomass

Home About Us Newsletters Atlases Events Vacancies

The **European Soil Data Centre (ESDAC)** is the thematic centre for soil related data in Europe. Its ambition is to be the single reference point for and to host all relevant soil data and information at European level. It contains a number of resources that are organized and presented in various ways: datasets, services/applications, maps, documents, events, projects and external links. We hope you can find your way in this site. When in doubt or for any question, you may contact **ec-esdac@ec.europa.eu**

## EUSO Dashboards

EUSO Soil Degradation Dashboard

EUSO Soil Strategy Actions Tracker

## EUSO Stakeholders Forum

5th EUSO Stakeholders Forum

**Find out more here!**

## Scientific/Technical Reports

The State of Soils in Europe

# Re3Data – a registry of data repositories

re3data.org

Search | Browse | Suggest | Resources | Contac

**Filter**

Reset all

- Subjects
- Content Types
- Countries
- AID systems
- API
- Certificates
- Data access
- Data access restrictions
- Database access
- Database access restrictions
- Database licenses
- Data licenses
- Data upload
- Data upload restrictions
- Enhanced publication
- Institution responsibility type
- Institution type
- Keywords
- Metadata standards
- Metrics
- PID systems
- Profiles
- Provider types
- Quality management
- Repository languages
- Software
- Syndications
- Repository types
- Versioning

Search... | Search

Search help

← Previous | 1 | 2 | 3 | 4 | 5 | 6 | 7 | ... | 13 | Next →

Sort by

Found 306 result(s)

**WDC Sunspot Index and Long-term Solar Observations**

SILSO

| | |
|---|---|
| Subject(s) | Natural Sciences, Physics, Astrophysics and Astronomy, Geosciences, Atmospheric Science, Oceanography and Climate Research, Atmospheric Science |
| Repository type(s) | disciplinary |
| Provider type(s) | dataProvider |
| Country | Belgium, International |

SILSO is the World Data Center for the production, preservation and dissemination of the international sunspot number.

**ECCAD - the GEIA database**

Emissions of atmospheric Compounds & Compilation of Ancillary Data

| | |
|---|---|
| Subject(s) | Natural Sciences, Chemistry, Analytical Chemistry, Geosciences, Atmospheric Science, Oceanography and Climate Research, Atmospheric Science |
| Repository type(s) | disciplinary |
| Provider type(s) | dataProvider, serviceProvider |
| Country | France, United States, Finland, European Union, International |

The main goal of the ECCAD project is to provide scientific and policy users with datasets of surface emissions of atmospheric compounds, and ancillary data, i.e. data required to estimate or quantify surface emissions. The supply of ancillary data - such as maps of population density, maps of fires spots, burnt areas, land cover - could help improve and encourage the development of new emissions datasets. ECCAD offers: Access to global and regional emission inventories and ancillary data, in a standardized format Quick visualization of emission and ancillary data Rationalization of the use of input data in algorithms or emission models Analysis and comparison of emissions datasets and ancillary data Tools for the evaluation of emissions and ancillary data ECCAD is a dynamical and interactive database, providing the most up to date datasets including data used within ongoing projects. Users are welcome to add their own datasets, or have their regional masks included in order to use ECCAD tools.

# Generalist Repositories

**Interdisciplinary Data Sharing**

Generalist repositories accept datasets (and other outputs) from any academic field, supporting interdisciplinary research effectively.

**DOI and Metadata Support**

These platforms support data citation through DOIs and provide customisable metadata for better data organisation.

**Lower Barriers to Access**

Usually require no specialist knowledge and can be accessed via a web browser without restriction.

**Usage for Supplementary and Negative Data**

They are useful for sharing supplementary data, negative results, and datasets spanning multiple disciplines.

**Examples**

Zenodo, DataVerse, Figshare.

# Featured communities

**Machine Learning for Particle Physics**

Browse

# Recent uploads

October 23, 2025 (v1) | Dataset | Open

## Segmentation of the anatomic organs for a sample of the National Lung Screening Trial (NLST) CT image collection: comparison across open-source AI models

Giebeler, Lena; Clunie, David A; Fedorov, Andrey

This dataset contains volumetric segmentations for a sample of Computed Tomography (CT) images from the National Lung Screening (NLST) collection, publicly available from NCI Imaging Data Commons, generated using open-source AI models. This dataset was generated as part of a project evaluating concordance among the AI models and accuracy of the...

Uploaded on October 23, 2025

Part of NCI Imaging Data Commons

0 0

October 2025 (v2) | Presentation | Open

## CoARA Generic Presentation

TOTH CZIFRA, Erzsébet; Coalition for Advancing Research Assessment (CoARA) i

Generic presentation to promote CoARA with up to date information.

## Why use Zenodo?

- **Safe** — Your research is stored safely for the future in CERN's Data Centre for as long as CERN exists
- **Trusted** — Built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science
- **Citeable** — Every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode
- **Versioning** — Easily update your dataset with our versioning feature
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo
- **Usage statistics** — All uploads display standards compliant usage statistics

## Newsletter

# Data Availability Statements

### Purpose of Statements

Data availability statements specify how and where research data can be accessed to support findings.

### Enhancing Transparency

These statements promote transparency and reproducibility by enabling verification and further research.

### Journal Requirements

Many academic journals require data availability statements to encourage open research and responsible data sharing.

JOURNAL OF BIOPHOTONICS

Volume 13, Issue 12
December 2020
e202000264

FULL ARTICLE

**Multimodal vibrational studies of drug uptake in vitro: Is the whole greater than the sum of their parts?**

David Perez-Guaita, Karolina Chrabaszcz, Kamilla Malek, Hugh J. Byrne

First published: 05 September 2020 | https://doi.org/10.1002/jbio.202000264 | Citations: 6

**Funding information:** H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: Spectro-metrics-020-MSCA-IF-2017 Project ID:79628; National Science Centre, Grant/Award Number: UMO 2016/23/B/NZ4/01379

Read the full text > PDF TOOLS SHARE

## Abstract

Herein, we investigated the use of multimodal Raman and infrared (IR) spectroscopic microscopy for the elucidation of drug uptake and subsequent cellular responses. Firstly, we compared different methods for the analysis of the combined data. Secondly, we evaluated whether the combined analysis provided enough benefits to justify the fusion of the data. A459 cells inoculated with doxorubicin (DOX) at different times were fixed and analysed using each technique. Raman spectroscopy provided high sensitivity to DOX and enabled an accurate estimation of the drug uptake at each time point, whereas IR provided a better insight into the resultant changes in the biochemical composition of the cell. In terms of benefits of data fusion, 2D correlation analysis allowed the study of the relationship between IR and Raman variables, whereas the joint analysis of IR and Raman enabled the correlation of the different variables to be monitored over time. In summary, the complementary nature of IR and Raman makes the combination of these vibrational techniques an appealing tool to follow drug kinetics and cellular response.



**Open Research**

**DATA AVAILABILITY STATEMENT**

The scripts detailing the data analysis as well as the data are available in the Zenodo repository (10.5281/zenodo.3784974).

References Related Information

**Recommended**

Computational and vibrational spectroscopic studies of ipratropium bromide
H. R. H. Ali, H. G. M. Edwards, J. Kendrick, I. J. Scowen
**Drug Testing and Analysis**

Infrared and Raman spectroscopy of mullite-type $Bi_2Ga_4O_9$
A. Beran, E. Libowitzky, M. Burianek, M. Mühlberg, C. Pecharromán, H. Schneider
**Crystal Research and Technology**

Investigation of jarosite process tailing waste by means of raman and infrared spectroscopy
M. Kerolli-Mustafa, I. Bačić, L. Ćurković
**Materialwissenschaft und Werkstofftechnik**

Raman spectroscopy of lipids: a review
K. Czamara, K. Majzner, M. Z. Pacia, K. Kochan, A. Kaczor, M. Baranska
**Journal of Raman Spectroscopy**

Structural studies of ligand stabilized Ni/Ga clusters by means of vibrational spectroscopy and theoretical calculations
János Mink, Lena Staiger, Maximilian Muhr, Christian Gemel, Markus Drees, László Hajba, Judith Mihály, Csaba Németh, Boris V. Lokshin, Karina Hemmer, Max Schütz, Mirza Cokoja, Roland A. Fischer
**Journal of Raman Spectroscopy**

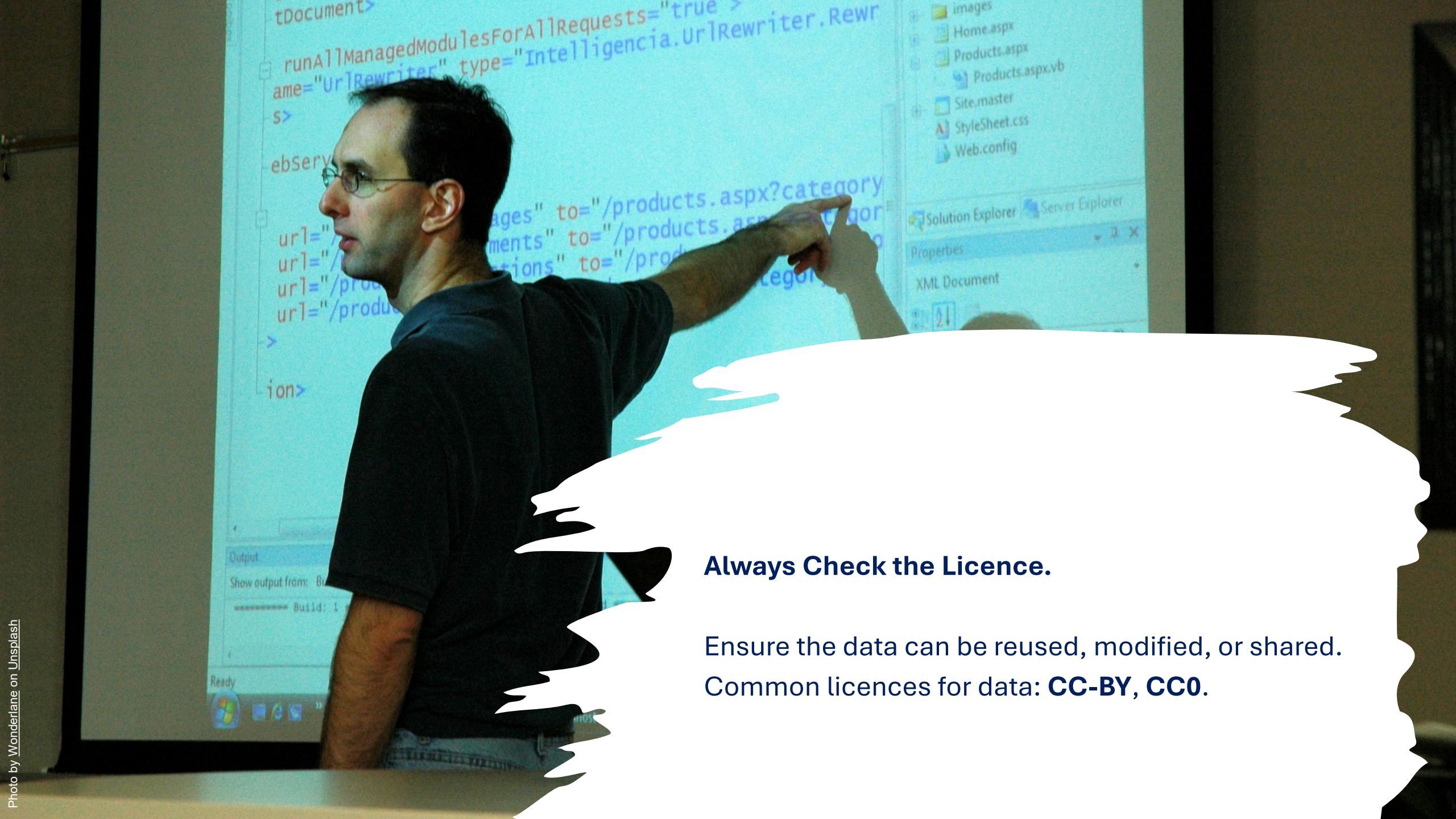
# (Open) Federated Searches & Knowledge Graphs

WIKIDATA

Google Dataset Search

OpenAlex

OpenAIRE | EXPLORE




**Always Check the Licence.**

Ensure the data can be reused, modified, or shared.
Common licences for data: **CC-BY**, **CC0**.

Photo by Wonderlane on Unsplash

# The European Open Science Cloud (in Ireland)

CC BY-NC

# A ‘one-stop-shop’ to discover, access, and utilise research data and services.

EOSC aims to provide Europe's researchers with the necessary digital resources to conduct research within and across disciplines and borders according to the principles of FAIR data and open science.

This will be done by putting in place a so-called ‘system of systems’ federated into a network of **nodes**.

An EOSC Node will be typically formed by multiple research institutes, universities and other stakeholders with a common geographical or thematic scope.

Fosters collaboration by making research outputs more accessible. In turn, this supports commercialisation of research, amplifying societal and economic impact.

While access remains governed by provider policies, EOSC Ireland will simplify navigation, access to European infrastructures, and integration of resources.

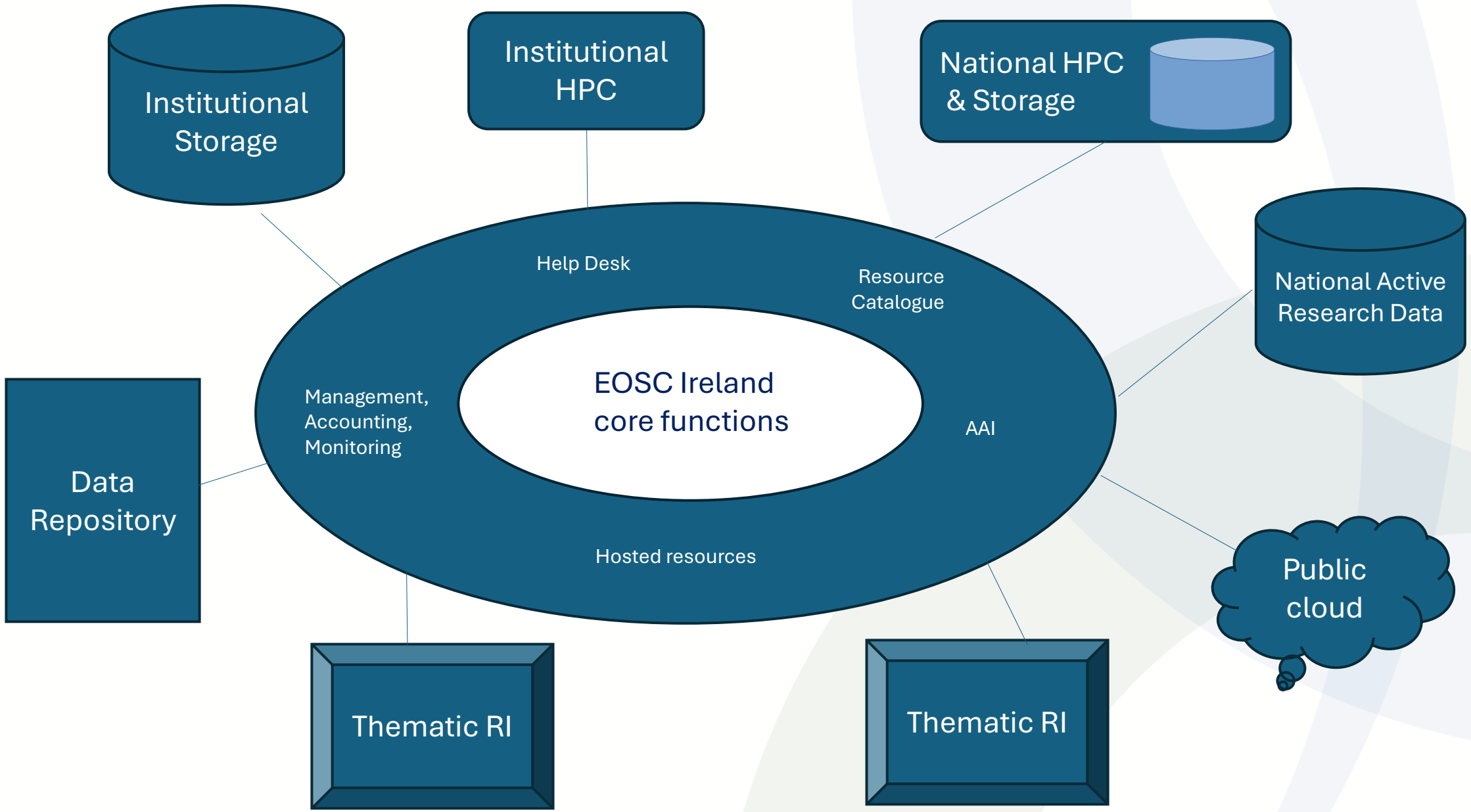
Through EOSC, a user can:

- Log in using their institutional credentials
- Search for a dataset, access the data repository and download the data onto storage provided by the EOSC-EU node
- Run an algorithm on the data, on compute infrastructure provided by the EOSC-EU node (VMs, Jupyter note books)
- Store results in a repository
- Make the algorithm available for others to re-use, by storing it in an “app resource” hosted by the EOSC-EU node
- Pay for resources used (currently via credits)
- Gain credits for resources provided (‘onboarded’)

CC BY-NC-SA – Roberto Sabatino

Photo by Recha Oktaviani on Unsplash

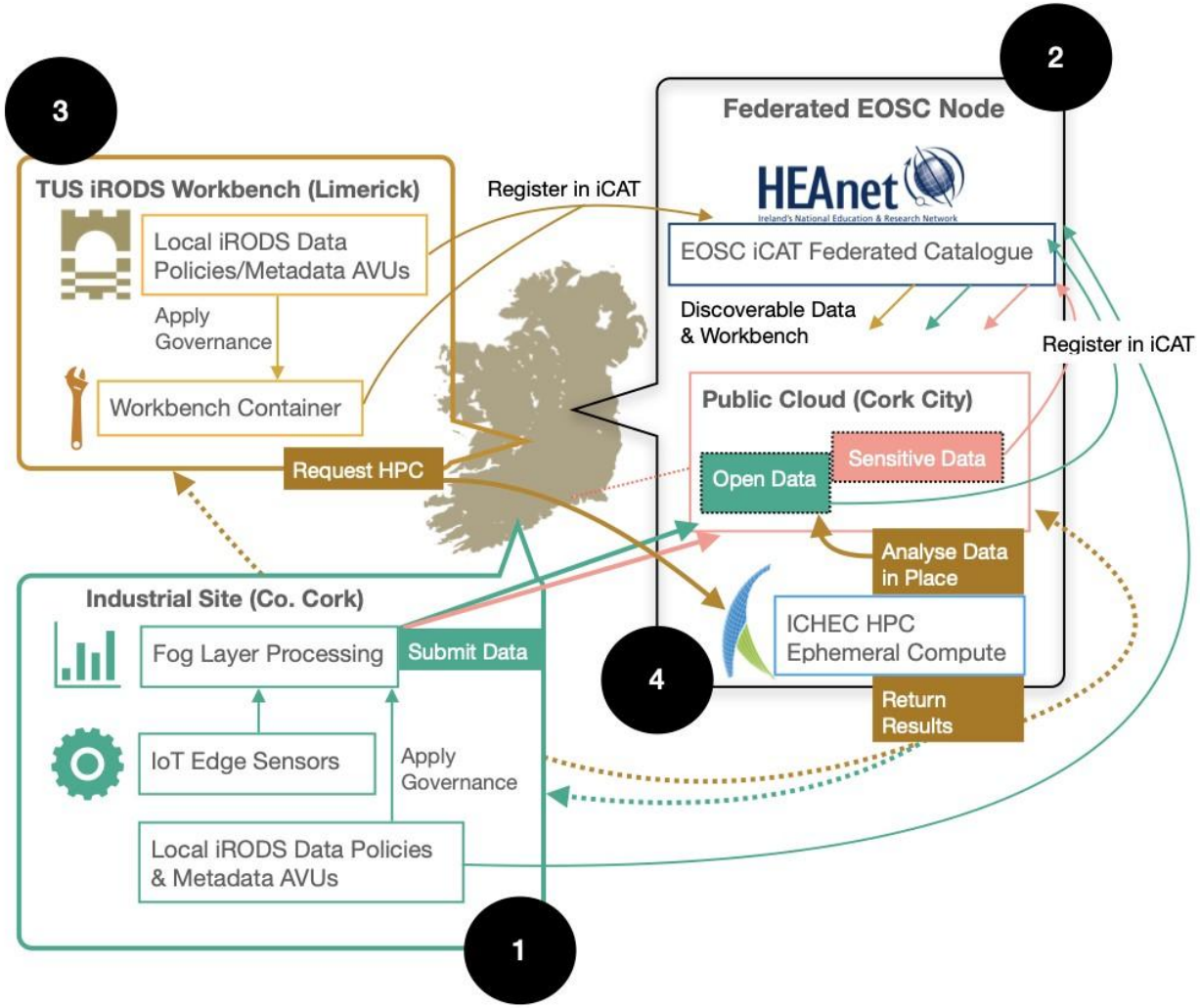
# EOSC Ireland: Pilot Phase



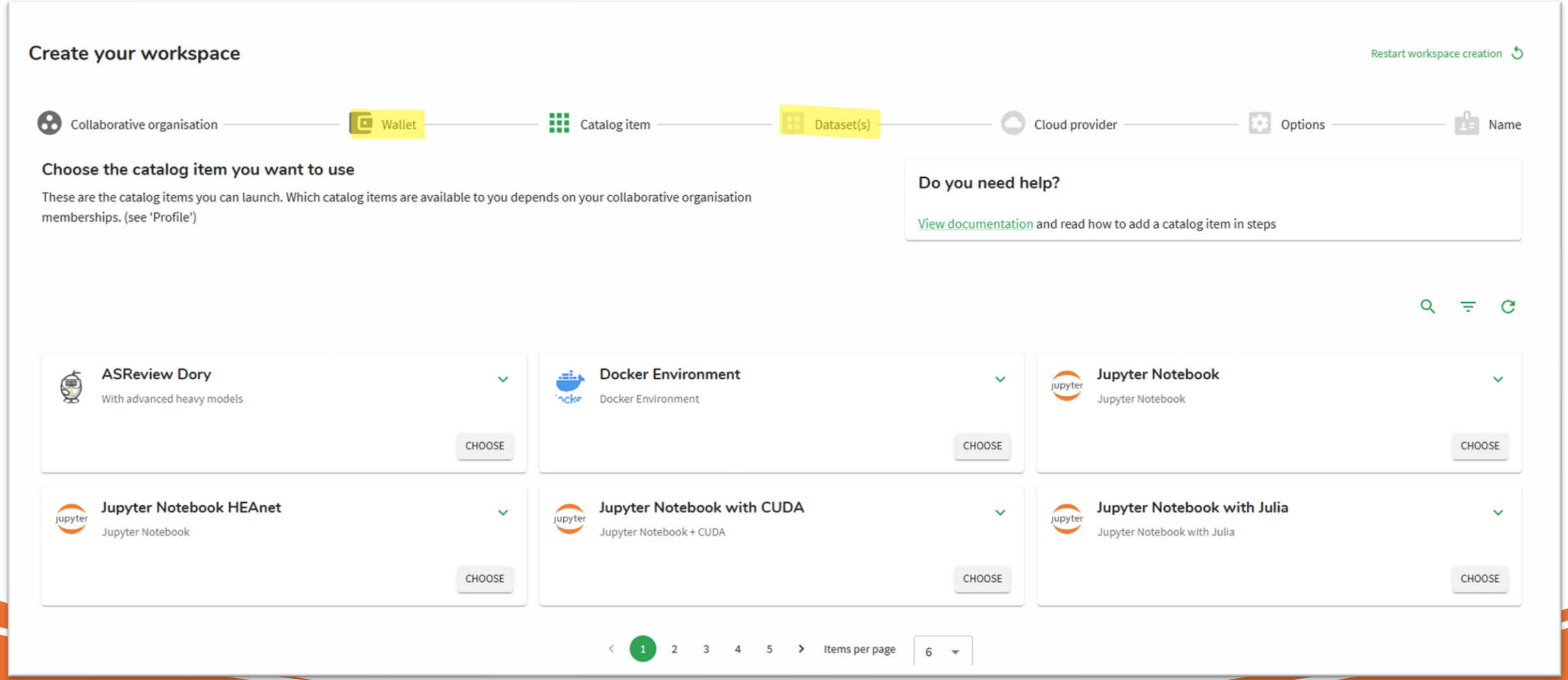

CC BY-NC-SA – Roberto Sabatino

# QuarryLink: Federated edge-to-cloud data pathway

1. Edge to Fog Data collections
2. Data Transfer to public cloud resources in EOSC-Ireland
3. Collaborative Open Toolchain development via TUS
4. In Place federated compute via ICHEC
   - Results published as Open Data via EOSC-Ireland
   - Shared with industrial partner




CC BY-NC-SA – Roberto Sabatino




Currently hosted on ResearchCloud (SURF NL)

CC BY-SA