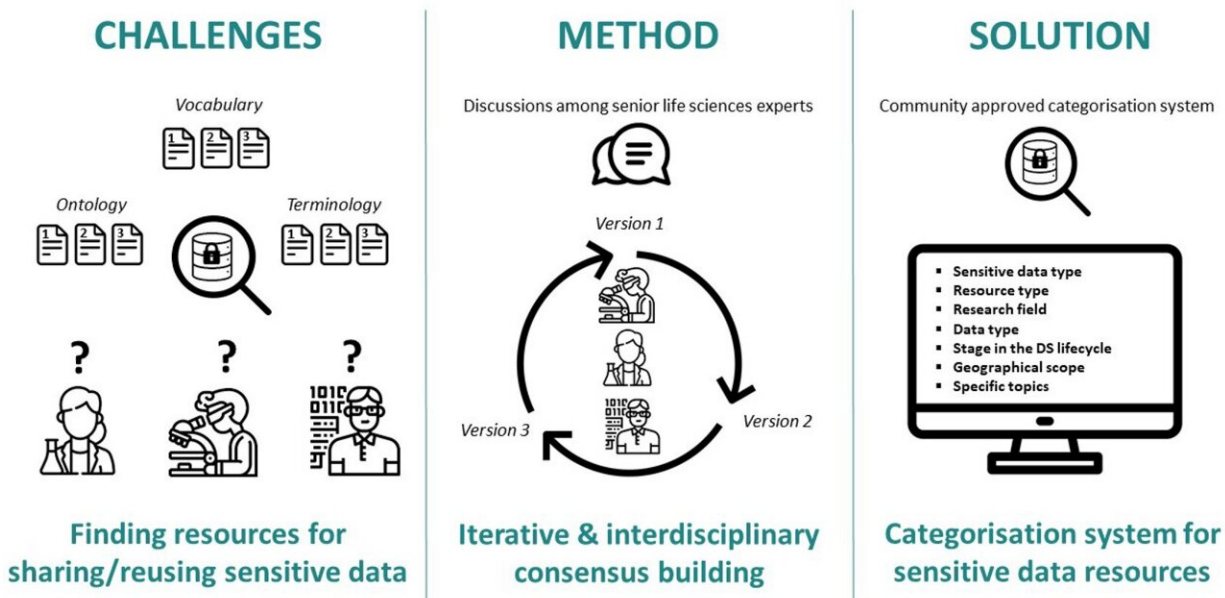


Toolbox for tagging sensitive data in life science research

January 25 2023



Graphical abstract: Challenges, method and solution. Credit: *Scientific Reports* (2022). DOI: 10.1038/s41598-022-25278-z

From climate change to global pandemics, the world is facing major environmental and health-related challenges that are driving life science research institutions to pool their data and digital resources in search of solutions.

However, a lot of the data generated by biological and [medical research](#)

is sensitive, either stemming from their personal nature or due to intellectual property considerations, biohazard concerns or the [Nagoya Protocol](#).

A toolbox for sensitive data

The EOSC-Life project is bringing together research infrastructures to create an open, digital and collaborative space for life science research in which data, tools and analysis workflows are more findable, accessible, interoperable and reusable (FAIR).

To support this FAIRification process, it has developed a toolbox that provides information to researchers intending to share and/or use sensitive data in a cloud environment such as the European Open Science Cloud (EOSC). The toolbox is described in a paper published in the journal *Scientific Reports*.

The toolbox is based on a categorization, or tagging, system developed and harmonized across a cluster of six life science research infrastructures involved in the EOSC-Life project. The toolbox does not create new content but rather enables scientists to find resources that are relevant for sharing sensitive data across all participating research infrastructures. It contains links to digital objects relating to sensitive data, such as regulations, guidelines, [best practices](#) and software, to support data sharing and reuse.

Development in three stages

The toolbox's categorization system makes consistent labeling and tagging of resources possible. Three different versions of the categorization system were developed, each tested by a subsequent [pilot study](#). This ultimately led to a system with seven main categories:

sensitive data type; resource type; research field; data type; stage in [data sharing](#) life cycle; geographical scope; and specific topics.

The third version of the categorization system was tested in pilot study 3 with 110 resources, one of which had missing data. A total of 109 resources tagged in this pilot study were therefore used as the initial content for the [toolbox demonstrator](#).

The demonstrator is a [software tool](#) that allows researchers to search digital objects linked to sensitive data, with filtering based on the categorization system. The study authors explain further: "The tool allows pre-filtering of resources linked to [sensitive data](#) with free text in the title, by DOI or through authors. Further filtering is possible with respect to item type (e.g. journal article, webinar, report, software) and selection of any of the pre-listed tags from the different categories of system version 3. The search result can be saved as PDF or JSON."

According to the authors, the next important steps involve evaluating the toolbox demonstrator's usability and user-friendliness, extending the toolbox to cover more resources, promoting its broader adoption by different life science communities, and developing a long-term vision for maintenance and sustainability. The EOSC-Life (Providing an open collaborative space for digital biology in Europe) project ends in August 2023.

More information: Romain David et al, An iterative and interdisciplinary categorisation process towards FAIRer digital resources for sensitive life-sciences data, *Scientific Reports* (2022). [DOI: 10.1038/s41598-022-25278-z](#)

EOSC-Life project website. www.eosc-life.eu/

Provided by CORDIS

Citation: Toolbox for tagging sensitive data in life science research (2023, January 25) retrieved 12 May 2024 from <https://phys.org/news/2023-01-toolbox-tagging-sensitive-life-science.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.