

Some thoughts on What is sensitive data sharing?

Romain David, PhD
Data steward, data scientist
ERINHA-AISBL

romain.david@erinha.eu

ORCID : 0000-0003-4073-7456

FAIR Impact

25th of june 2024



ERINHA is a RI of biocontainment laboratories which is specialized in infectious disease research



BSL-4



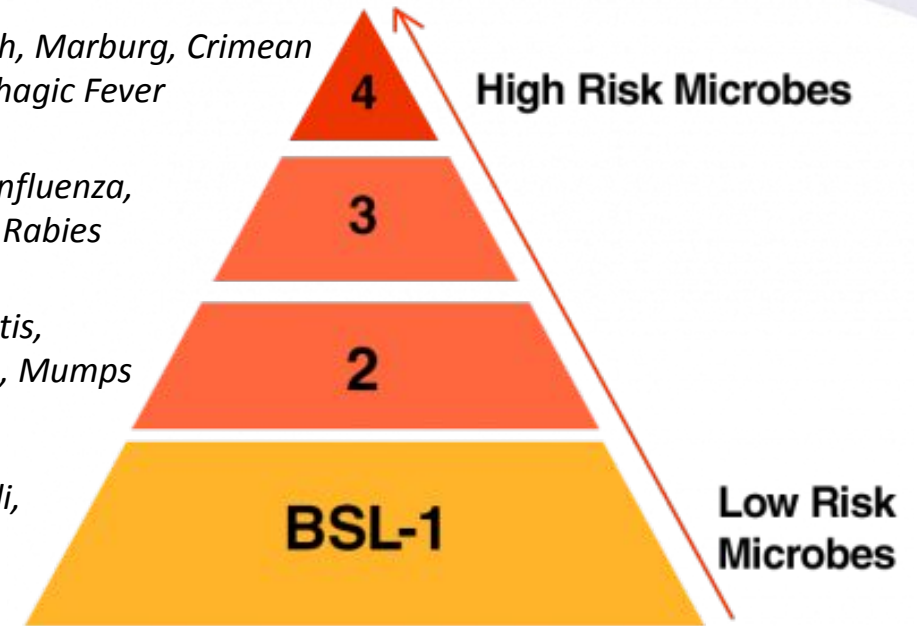
BSL-3

*ex. Ebola, Nipah, Marburg, Crimean
Congo Hemorrhagic Fever*

*ex. HIV, H1N1 Influenza,
Yersinia pestis, Rabies*

*ex. Influenza, Hepatitis,
Salmonella, Measles, Mumps*

*ex. Non pathogenic E. coli,
nonpathogenic bacteria*



Biocontainment laboratories:

- Unique buildings with complex engineering systems maintaining 'containment'
- Increased personnel and information security
- Nationally and internationally regulated

Sensitive data: concept of risk

4 points considering possible community “Risks”

- Economical risks
- Interference with security programs / tools
- Misappropriation of knowledge and data (for instance to build a weapon)
- Terrorism (plan and access to hospital, stadium...)

Environmental sensitive data

- Endangering coveted and scarce resources (including relocalisable data)

Personal data

- Endangering pearson (including re-identifiable data)

Sensitive data: Main types

Personal Data: Information that can identify an individual (e.g., name, address, date of birth).

Health Data: Medical records, genetic data, health conditions.

Location Data: Geographical information that can trace individual movements.

Confidential Business Data: Proprietary business information, trade secrets.

Sensitive Government Data: Classified or restricted government information.

Cultural and Archaeological Data: Information that could harm or exploit cultural heritage sites and communities.

Sensitivity in the Context of FAIR Data Sharing

Findability: Ensuring sensitive data is discoverable without compromising privacy.

Accessibility: Balancing open access with necessary restrictions.

Interoperability: Harmonizing standards for sensitive data across platforms.

Reusability: Ensuring sensitive data can be reused responsibly and ethically.

Challenges and Mitigation Strategies

Challenges:

- Privacy breaches and data misuse.
- Ethical concerns and legal compliance.
- Technical barriers in secure data sharing.

Mitigation Strategies:

- Anonymization and pseudonymization techniques.
- Secure data access protocols and data governance frameworks.
- Collaboration with ethics boards and legal teams.

Sensitivity in the Context of FAIR Data Sharing : challenges

Findability: Ensuring sensitive data is discoverable without compromising privacy.

Accessibility: Balancing open access with necessary restrictions.

Interoperability: Harmonizing standards for sensitive data across platforms.

Reusability: Ensuring sensitive data can be reused responsibly and ethically.

- Risk of exposing sensitive information through metadata.
- Difficulty in indexing sensitive data while preserving confidentiality.
- Balancing the need for open access with the protection of sensitive information.
- Navigating legal and ethical restrictions on data access.
- Inconsistent standards and protocols across different institutions and jurisdictions.
- Technical and organisational barriers to integrating sensitive data from diverse sources.
- Ensuring data is reusable without compromising sensitive information.
- Difficulty in maintaining data integrity and relevance over time.

Sensitivity in the Context of FAIR Data Sharing : mitigations?

Findability:

- **Controlled Vocabularies and Ontologies:** Classify data using standardized terms without revealing sensitive details.
- **Metadata Encryption and Access Controls:** Encrypt metadata and implement strict access controls to prevent unauthorized access.

Accessibility:

- **Tiered Access Levels:** Create different access levels based on user roles and credentials to control who can see sensitive data.
- **Secure Data Access Platforms:** Develop platforms with strong authentication and authorization mechanisms to ensure secure access.

Interoperability:

- **Adoption of Common Standards:** Encourage the use of shared standards and frameworks (e.g., GDPR) to ensure consistency and compliance.
- **Interoperable Data Formats and APIs:** Use standardized data formats and APIs that support secure and efficient data exchange between systems.

Reusability:

- **Anonymization and Pseudonymization Techniques:** Apply techniques to anonymize or pseudonymize data to protect identities while maintaining data utility.
- **Clear Data Usage Policies:** Develop and enforce policies that guide secondary users on how to handle and use sensitive data responsibly.

Converging also for sensitive data sharing

[Communities](#)
[My dashboard](#)
[Log in](#)
[Sign up](#)

Planned intervention: On Wednesday June 26th 05:30 UTC Zenodo will be unavailable for 10-20 minutes to perform a storage cluster upgrade.

Published July 7, 2023 | Version 1.0.3

Conference paper [Open](#)

Converging on a Semantic Interoperability Framework for the European Data Space for Science, Research and Innovation (EOSC)

David, Romain¹ ; Baumann, Kurt² ; Le Franc, Yann³ ; Magagna, Barbara⁴ ; Vogt, Lars⁵ ; Widmann, Heinrich⁶ ; Jouneau, Thomas⁷ ; Koivula, Hanna⁸ ; Madon, Bénédicte⁹ ; Åkerström, Wolmar Nyberg¹⁰ ; Ojsteršek, Milan¹¹ ; Scharnhorst, Andrea¹² ; Schubert, Chris¹³ ; Shi, Zhengdong¹⁴ ; Tanca, Letizia¹⁵ ; Vancauwenbergh, Sadia¹⁶

[Show affiliations](#)

Semantic interoperability (SI) is at the heart of the FAIR principles and the design of large-scale cross-disciplinary infrastructures. The European Open Science Cloud (EOSC) is a European-wide effort towards such an infrastructure, aiming to deepen regional research collaboration and realising a shared data space for science, research and innovation. In this context, the research community's voice is represented by the EOSC Association (EOSC-A) and a number of advisory groups with a broad range of representatives from different stakeholder organisations. The advisory group on metadata and data quality has formed a task force focusing on developing and implementing recommendations for SI (EOSC SI Task Force) to converge on globally relevant and scalable SI solutions for EOSC. This paper provides context to SI in EOSC, the various components contributing to it, as well as some views on the socio-technical challenges to arriving at a consensus. In particular, the paper provides motivation for exploring the heterogeneity of SI solutions demonstrated across scientific communities and insight into the task force's planned approach to conducting a survey to identify relevant components and structures. The paper is also an invitation to the global community to align and engage with the task force's activities going forward.

Notes

This research is a product of the Task Force "Semantic Interoperability" of the EOSC-Association European, the legal entity established to govern the European Open Science Cloud (EOSC). Complementary support and information were provided through European projects "EOSC-Life" (No824087) and "FAIR Impact" (No101057344).

1K VIEWS

861 DOWNLOADS

[Show more details](#)

Versions

Version 1.0.3	Jul 7, 2023
10.5281/zenodo.8102786	
Version 1.0.2	Jun 30, 2023
10.5281/zenodo.8102718	
Version 1.0.1	Jun 15, 2023
10.5281/zenodo.8042998	

[View all 3 versions](#)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.8042997](https://doi.org/10.5281/zenodo.8042997). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

David, R., Baumann, K., Le Franc, Y., Magagna, B., Vogt, L., Widmann, H., Jouneau, T., Koivula, H., Madon, B., Åkerström, W. N., Ojsteršek, M., Scharnhorst, A., Schubert, C., Shi, Z., Tanca, L., & Vancauwenbergh, S. (2023). Converging on a Semantic Interoperability Framework for the European Data Space for Science, Research and Innovation (EOSC) (1.0.1). 2nd Workshop on Ontologies for FAIR and FAIR Ontologies (Onto4FAIR) (Onto4FAIR), Sherbrooke, Québec (Canada). Zenodo.

<https://doi.org/10.5281/zenodo.8042998>

Sensitive data within biocontainment laboratories

Type of Data	Examples	Protected?
Building Information	Equipment location; IT network controlling equipment; Type of decontamination;	Yes; institution specific
Personnel	Names; knowledge; access to facility	Yes; institution specific
Inventory	Quantity, quality, and location of pathogens stored in laboratory	Yes; institution specific; required (by international treaty) to be protected
Procedures and Scientific Methods	How are pathogens handled?	Depends. Methods may be published
Data and Results	Data generated by experimentation	Depends. Data and results may be published

Sensitive Data within Biocontainment laboratories

Type of Data	Examples	Protected?
Building Information	Equipment location; IT network controlling equipment; Type of decontamination;	Yes; institution specific
Personnel	This data comprises National Security Data and should never be required to be shared in an open-access manner.	
Inventory	Quantity, quality, and location of pathogens stored in laboratory	Yes; institution specific; required (by international treaty) to be protected
Procedures and Scientific Methods	How are pathogens handled?	Depends. Methods may be published
Data and Results	Data generated by experimentation	Depends. Data and results may be published

erinha

European Research Infrastructure
on Highly Pathogenic Agents

Questions?

www.erinha.eu



@ERINHA_RI

