Charter of Working Group Sensitive Data

Version 1.0 of 22. 1. 2024, by the group leader in collaboration with members of the group.

1 Introduction

The Working Group called Sensitive Data (WG SENSI) is focused on addressing the management of personal data and sensitive data (of various degrees of sensitivity) in the digital environment. The definition and classification of sensitive data are provided in Appendix 1 of this document. The WG aims to map the current situation in the Czech Republic, existing FAIRification procedures for sensitive data, and to the optimization of comprehensive data management across disciplines. In all its activities the WG considers the level of sensitivity. The establishment of this group was primarily driven by the need to standardize the management of sensitive data across scientific domains in the Czech Republic. The group, as part of the EOSC implementation in the Czech Republic, is cross-disciplinary since sensitive data spans multiple scientific domains. The focus and scope of topics addressed by this WG are primarily determined by ethical-legal and organizational-security aspects. Recommendations from the SENSI WG experts will be considered in the implementation of EOSC in the Czech Republic and the National Repository Platform (NRP). These recommendations should cover the management of various types of sensitive data with different sensitivity levels.

Since its inception in April 2022, the members of the WG have taken inspiration from the UK Data Service's "Five Safes" framework[1] (Figure 1). Currently, the experts from this WG are involved in various projects and initiatives addressing the sharing of sensitive data.

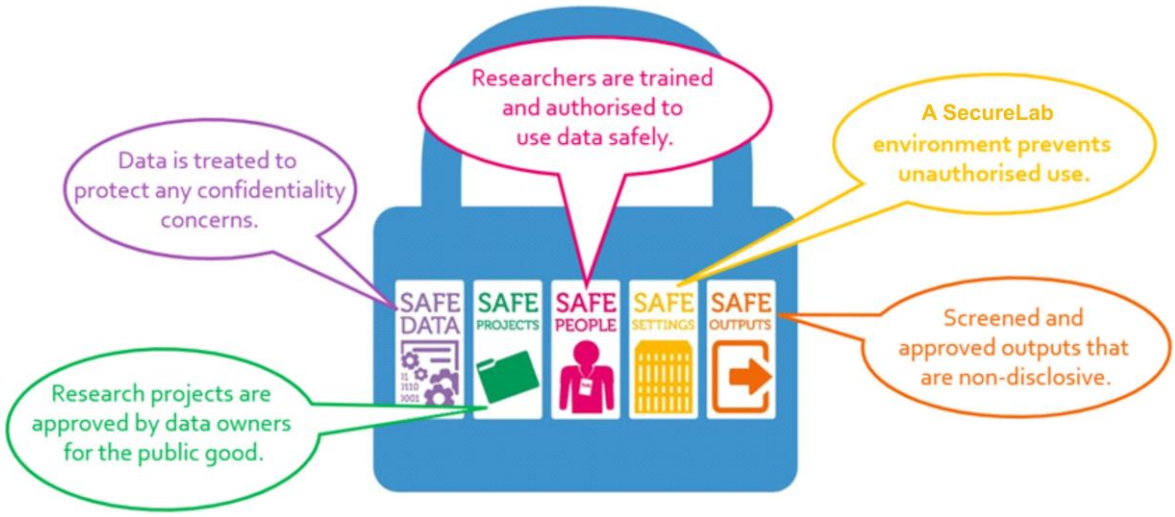

Figure 1: The Five Safes framework, adopted from z [https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/,](https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/) modified.

[1] https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/

European Open Science Cloud
Czech Republic (EOSC-CZ)
Masaryk University, Institute of Computer Science
Mailing address: Sumavská 33, 602 00 Brno
info@eosc.cz; www.eosc.cz

2 Objectives

1. Specification of requirements for sensitive data management, focusing on the digital environment.
2. Definition of steps and tasks to fulfil the specifications.
3. Division of tasks based on expertise and experience.
4. Identification of areas with insufficient expertise covered by the WG and collaboration with experts from these areas.

3 Outputs and their Applications

1. Preparation of an environment for managing sensitive data across domain clusters within EOSC CZ.
2. Coordination, technological recommendations, expert teams supporting sensitive data management within EOSC CZ – collaboration primarily with the Working Group Education and Human Resources (WG EDU).

4 Membership and Expected Members

The WG is open to all individuals with experience in sensitive data processing and in its management aiming into the secondary usage. The number of WG members is not limited. Typically, members are experts dealing with sensitive data and/or involved in their management and protection from technical, legal, or ethical perspectives. The group collaborates with relevant infrastructures and projects at the national and international level.

To join the WG, interested individuals need to fill out the registration form on the EOSC CZ website.

European Open Science Cloud
Czech Republic (EOSC-CZ)
Masaryk University, Institute of Computer Science
Mailing address: Sumavská 33, 602 00 Brno
info@eosc.cz; www.eosc.cz

Appendix 1: Definition of the Concept of Sensitive Data

Definition

Sensitive data refers to data subject to the following restrictions:

1. Strictly intended for the internal use of a precisely defined group of individuals (e.g., healthcare professionals and their patients, project participants working with commercially or similarly confidential data, etc.).
2. Requires special regulation due to its nature, typically explicitly protected by law or based on trade secrets, contracts, licenses, etc. (e.g., highly valuable data falling under trade secrets, sensitive personal data of institution employees/clients, etc.).
3. Access by unauthorized individuals outside the specified group is likely to cause significant/damaging consequences (financial, moral, legal, health/life-related, etc.).

The categorization of data sensitivity is determined by the data owner unless otherwise specified by law.

Intuitive Definition

Intuitively, these are data for which the owner believes that their protection should involve measures beyond those commonly used for data protection. Additional measures should be implemented at the highest achievable level of protection according to contemporary best practices and available technologies because these data are exceptional in some way – they have exceptional value for the owner's activities, are generally considered highly confidential, are commercially exploitable, yet unavailable to most interested parties, etc.

Assessment of Data Sensitivity

The above definition should be understood in the context of specific data, its owner/manager, and the costs of implementing protective measures. The general rule should be that the effort and resources expended on data protection should be commensurate with the harm caused by their leakage. However, it is necessary to consider that harm may not only be financial but also legal, reputational for the organization, etc., which may subsequently generate secondary financial consequences (e.g., loss of customer trust leading to a decrease in company revenue). When assessing damage to a significant extent, it is necessary to relate the assessment to the specific data owner and their situation. Financial damage of CZK 10 million may be of significant extent for a small company or private entrepreneur with an annual turnover of CZK 3 million. Still, they may not be of a significant extent for a multinational corporation with an annual turnover in the billions of euros. On the other hand, damage to the reputation of an organization or individuals caused by the leakage of such data may be unacceptable for a given multinational corporation. At the same time, for a smaller company, it will have a smaller impact than financial damage.

European Open Science Cloud
Czech Republic (EOSC-CZ)
Masaryk University, Institute of Computer Science
Mailing address: Sumavská 33, 602 00 Brno
info@eosc.cz; www.eosc.cz

Categories of Data Sensitivity

Given the relativity of assessing data sensitivity, several levels of data sensitivity are defined.

Low Sensitivity

Represents little or no risk to individuals, private organizations, or government agencies when data is made public. Anyone has access to data in this group because their accessibility is limited little or not at all. These are generally public information that can be discussed anywhere and with anyone.

Examples include public data (e.g., presentations from public lectures; publicly accessible research reports; open-source software; open research data; open data management plans; and public posters) and internal data (internal correspondence; meeting minutes; internal regulations and rules; research methodology descriptions; unfinished/unpublished research reports).

Medium Sensitivity

Includes data that is discrete and subject to a contractual obligation to protect. Leakage of such data would cause harm (financial, moral, legal) to affected individuals or organizations, but the consequences would not be severe. Examples include building plans, individual donor records, student records, intellectual property, IT service information, visas, and other travel documents, extensive low-sensitivity data collections, security information, and contact information and documents.

High Sensitivity / Confidential Data

Highly sensitive and confidential data must be protected by law or other relevant principles. Unauthorized access to such data outside the specified group of authorized individuals may cause significant (financial, moral, legal) harm with serious consequences for individuals or organizations. Examples include personal identification data, social security numbers, uncontrolled undisclosed information, identifiable human subject research, loan data, protected health information, extensive medium-sensitivity data collections, etc. This category may also include confidential corporate information such as trade secrets and intellectual property data, market analysis, financial documents, and board documents. Examples of highly sensitive data may include:

- health data, sensitive personal data (data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; membership in trade unions; genetic data; biometric data processed to identify a natural person; data concerning health; data concerning the sex life or sexual orientation of an individual, etc.).,
- personal and economic data about employees, students, customers, etc.;
- highly valuable research data (providing, for example, a unique and difficult-to-replicate competitive advantage, with the potential for successful commercialization, such as through patents, etc.) or research data containing highly confidential information (protected, for instance, by legal agreements with specified high penalties, etc.),
    - generally, data protected by intellectual property, trade secrets, etc;
- unpublished research data that has not yet been utilized by its authors (potential financial damage [patents, etc.]), has not been sufficiently verified for accuracy (potential damage to reputation), etc;

European Open Science Cloud
Czech Republic (EOSC-CZ)
Masaryk University, Institute of Computer Science
Mailing address: Sumavská 33, 602 00 Brno
info@eosc.cz; www.eosc.cz

- data sensitive in terms of cybersecurity (i.e., data whose unauthorized access may, for example, reduce the effectiveness of an organization's IT security measures, leading to potential harm through data breaches, impairment of its IT infrastructure functionality, etc.);
- data whose general availability may compromise the security of a state.

European Open Science Cloud
Czech Republic (EOSC-CZ)
Masaryk University, Institute of Computer Science
Mailing address: Sumavská 33, 602 00 Brno
info@eosc.cz; www.eosc.cz