



HAL
open science

RDA Sensitive Data Interest Group: Goals and Roadmap

Romain David, Dharma Akmon, Nichola Burton, Steven Mceachern, Aleksandra Michalewicz, Rita Rb-Silva, Priyanka Pillai, Hervé Raoul, Audrey Richard, Keith Russell, et al.

► To cite this version:

Romain David, Dharma Akmon, Nichola Burton, Steven Mceachern, Aleksandra Michalewicz, et al.. RDA Sensitive Data Interest Group: Goals and Roadmap. RDA 17th Plenary Meeting - Edinburgh, Apr 2021, Edinburg, United Kingdom. , 2021, 10.5281/zenodo.4690571 . hal-03226010

HAL Id: hal-03226010

<https://hal.science/hal-03226010>

Submitted on 13 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

The RDA Sensitive Data Interest Group aims to promote the FAIR principles and reproducible research, while drawing attention to the unique risks associated with sensitive data and exploring strategies to mitigate these risks.

Background and challenges

Sharing sensitive data of high quality is vital to assess and compare research results and to leverage data to their fullest capacity. However, as sensitive data are increasingly used for research purposes finding strategies to mitigate risks and harms is becoming particularly crucial.

Sensitive data spans a wide range of disciplines, contexts and stakeholder groups. This interest group will bring together people from these diverse backgrounds to develop shared understanding, language and resources and advance the conversation about the use and reuse of sensitive data.

We expect the RDA community to raise a variety of issues for this group to address as the IG develops, such as:

- How do we assess the level of data sensitivity across different contexts and disciplines?
- What are the risks associated with re-identification of different types of sensitive data, and how does re-identification occur?
- What guidance can we offer to researchers to help them share sensitive data safely and ethically?
- How do we balance sharing and reuse (e.g. FAIR) with cultural and ethical considerations (e.g. CARE)?
- How do different consent models affect the primary and secondary use of sensitive data?

Objectives

The IG objectives will be informed by the community. We take the following objectives as a starting point:

- Develop a shared understanding and refined definition of sensitive data.
- Define various levels of sensitivity for data and develop tools to assess this sensitivity.
- Develop an understanding of how sensitivity relates to openness.
- Identify different consent models.
- Identify types of sensitive data holdings and resources across various domains.
- Identify existing data definitions and standards for different types of sensitive data.
- Identify challenges in collecting, using and sharing sensitive data.
- Engage with key stakeholders working in the area of sensitive data management/analytics.
- Identify existing solutions for sensitive data collection, analysis, storage and dissemination.
- Identify differences in how sensitive data is managed between groups and regions.



Image: Public Domain, <https://pixabay.com/illustrations/web-connection-data-wave-binary-4809584/>

Planned outcomes

- To identify the key expectations of the community and use these to refine the IG's objectives.
- List different types of data across disciplines (e.g., health, social sciences) and how different levels of sensitivities apply to different types of data.
- Identify best practices in sensitive data management across multiple regions, domains and disciplines and how to adapt the best practices.
- Engage with relevant RDA IGs, WGs and CoPs to identify priorities in the area of sensitive data management.
- Gather common guidelines and recommendations for working with sensitive data in different disciplines and in different regions.
- Catalogue ethical, philosophical and cultural principles that underpin the use of sensitive data assets.

What is sensitive data?

We begin with the following working definition of sensitive data:

Information that is regulated by law due to possible risk for plants, animals, individuals and/or communities and for public and private organisations. Sensitive personal data include information related to racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership and data concerning the health or sex life of an individual. These data that could be identifiable and potentially cause harm through their disclosure. For local and government authorities, sensitive data is related to security (political, diplomatic, military data, biohazard concerns, etc.), environmental risks (nuclear or other sensitive installations, for example) or environmental preservation (habitats, protected fauna or flora, in particular). The sensitive data of a private body concerns in particular strategic elements or elements likely to jeopardise its competitiveness.

(adapted from David et al., 2020, "Templates for FAIRness evaluation criteria - RDA-SHARC IG": <https://zenodo.org/record/3922069#.YCUJ7ehKg2w>)

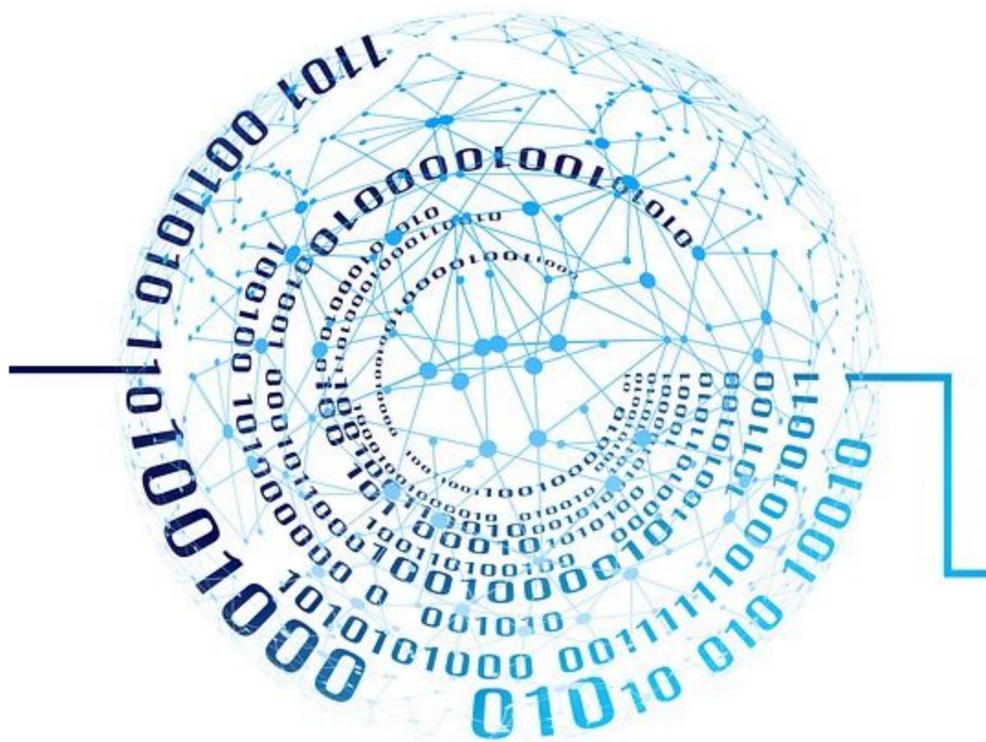


Image: Public Domain, <https://pixabay.com/illustrations/digitalization-email-letters-phone-4809637/>

Going Forward

In the immediate future, this IG will workshop the above definition of sensitive data, present a summary of the aims of the group and our charter to the RDA community and work towards becoming an endorsed IG of the RDA.

More broadly, the IG aims to provide a forum for a range of communities to share their requirements and jointly develop strategies, support, recommendations and guidelines relevant to sensitive data. We will define common goals around how best to address the risks associated with different types of sensitive data (e.g., Indigenous data, human health data, ecological data), and how to responsibly disseminate, aggregate and use preexisting heterogeneous sensitive data at a global scale.

We welcome participation and contributions from the entire RDA community and more broadly.

Further information about the Sensitive Data IG can be found in our proposed charter: <https://www.rd-alliance.org/group/sensitive-data-interest-group/case-statement/sensitive-data-interest-group-charter>.

* First references:

★ David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Le Bras, Y., Lerigoleur, E., Cambon-Thomsen, A. and The Research Data Alliance – SHARing Reward and Credit (SHARC) Interest Group, 2020. FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles. *Data Science Journal*, 19(1), p.32. DOI: <http://doi.org/10.5334/dsj-2020-032>

POSTER LEAD: Romain David (European Research Infrastructure on Highly Pathogenic Agents [ERINHA], 0000-0003-4073-7456)

CO-AUTHORS (in alphabetical order): Dharma Akmon (University of Michigan, 0000-0002-1359-0586); Nichola Burton (Australian Research Data Commons, 0000-0003-4470-4846); Steven McEachern (Australian Data Archive, 0000-0001-7848-4912); Aleksandra Michalewicz (University of Melbourne, 0000-0002-7328-2470); Priyanka Pillai (University of Melbourne, 0000-0002-3768-8895); Hervé Raoul (ERINHA), Audrey Richard (ERINHA); Rita Rb-Silva (Instituto Português Oncologia do Porto Francisco Gentil, EPE, University of Porto, 0000-0002-1422-0974); Kristal Spreadborough (University of Melbourne, 0000-0002-7022-3213); Diana Stepanyan (ERINHA); Frankie Stevens (Australia's Academic and Research Network, 0000-0002-2556-473X).

Presented at Research Data Alliance 17th Plenary - remotely, April 2021.