

# RDA Sensitive Data Interest Group Community meeting

## October 27, 2021

[1 Agenda](#)

[2 Useful Links](#)

[3 Attendees](#)

[4 Chat notes](#)

## 1 Agenda

#	Mins	What	Who
1	10	Acknowledgement of Country Introductions	AM
2	10	Context setting <ul style="list-style-type: none"><li>- Recap of the last meeting</li><li>- Foreshadow the methodologies from the poster (<a href="#">RDA submission poster</a>)</li><li>- Invitation to present lightning talks at RDA18 (to be returned to at end of meeting)</li></ul>	KS
3	10	Do you have any (Halloween) horror / success stories from working with sensitive data that you would like to share?	KK/NB

- Work on images - not really unidentifiable and personal data.
- Sensitive data being put on a server somewhere, but being lost. Where is it now?
- Regulation via contracts (when, where); experimental server so cannot be primary location
- Institutional implications for contravening sensitive data requirements/laws
- Data use agreements for data disposal
- Research which extends beyond national borders > how do you navigate challenges? What rules apply in cross-regional research?
- Do researchers make use of Trusted Research Environments? How does this interact with older methods like sending data out?
- Complications arise when there are multiple stakeholders
- Accidentally sending data out to unauthorised people. When unreported this can go on for a long time. There can be a lag in adopting new tech (e.g. even from fax to email).
- Approved Security Plans for ensuring data security (where possible, complications can arise, e.g., people leaving institutions)
- Breaches of data privacy and mishandling of sensitive data
- Email disasters! Forwarding / reply all without looking at full information contained within; cannot be sure with open networks where emails are passing through; encrypted emails also not necessarily safe
- Data retraction procedures; signatory authorisation processes / chain
- Language challenges, understanding university hierarchies structures - not always clear
- Escalation processes
- Highly sensitive data can require specific infrastructure; secure physical sites also need to be considered
- Military grade (and other) physical destruction processes
- Expectations regarding destruction
- Dynamic consent models
- Maintaining replicability of analysis
- Cybersecurity/IT security vetting for the sensitive data should not be considered as a one-time thing.
- Repositories with poor data documentation resulting in data being *more* sensitive because we have no idea what is in there

		<ul style="list-style-type: none"> <li>- Sensitivity to which degree? Found also in GDPR</li> <li>- Sensitive to whom?</li> <li>- Risk: who and to what extent and what are the consequences? Risk assessment crucial (stakeholders but also bystanders who can also be affected)</li> <li>- Importance of citing sources</li> </ul>	
4	15	<p>Drawing on your experience, can you share any sensitive data examples concerning:</p> <ul style="list-style-type: none"> <li>- military topics <ul style="list-style-type: none"> <li>- “Dual use” - data that could be used to develop weapons</li> <li>- Cyberterrorism</li> <li>- Terminology used for classifying national secrets not confidentiality</li> <li>- Length of time data must be held &gt; how to make this truly secure over decades?</li> </ul> </li> <li>- health data <ul style="list-style-type: none"> <li>- Fax reports</li> <li>- Biobank tissue</li> <li>- Patient records</li> </ul> </li> <li>- biodiversity - environment risk</li> <li>- personal data <ul style="list-style-type: none"> <li>- Criminal records</li> <li>- Data about minors</li> <li>- GDPR</li> <li>- IP address (sometimes, it is accidentally collected by default via online survey tools)</li> </ul> </li> <li>- industrial data <ul style="list-style-type: none"> <li>- Data with IP value</li> <li>- Data under third-party contractual agreements (sometimes, the level of data/information has restriction on could include the funds associated with the agreement)</li> </ul> </li> <li>- Administrative data <ul style="list-style-type: none"> <li>- Government data</li> </ul> </li> <li>- Other <ul style="list-style-type: none"> <li>- Data indicating the location of valuable assets</li> </ul> </li> </ul>	KK/NB

		<ul style="list-style-type: none"> <li>- Working with data from a culture other than your own</li> <li>- Legal data for developing ML techniques</li> <li>- Many different ways in which data can be considered sensitive</li> <li>- Tensions between individual and communal with regards to sensitive data (e.g., covid examples)</li> <li>- Sensitive v secure</li> <li>- We spend a lot of time talking about how to KEEP data, but what about destroying sensitive data? <ul style="list-style-type: none"> <li>- Destruction policies</li> <li>- Deaccession of data</li> <li>- Responding to data provider requests: cannot always be met</li> </ul> </li> <li>- Backups &gt; difficult issue; data restores - will establish from backup &gt; how to do this in a safe way?</li> </ul>	
5	10	<p>Drawing on your examples, experiences, or case studies, please share common terms used around sensitive data. Please participate in this poll: <a href="#">POLL LINK</a></p> <p>From Romain: Sensitive data life cycle must be described considering the level of sensitivity they have</p> <p>Aim:</p> <p>Some terms mentioned in the meeting today:</p> <ul style="list-style-type: none"> <li>- BioBank</li> <li>- Data sharing</li> <li>- Open network</li> <li>- Closed network</li> <li>- Region</li> <li>- Level of sensitive</li> <li>- Computer security</li> <li>- Communication</li> <li>- Encrypted</li> </ul>	RD/KK/NB

- |  |  |  |
|--|--|--|
|  | <ul style="list-style-type: none"><li>- Retraction</li><li>- Regulation</li><li>- Signatory authorisation process</li><li>- Structure</li><li>- Legal framework</li><li>- Breach</li><li>- Leak</li><li>- Enclave</li><li>- House data</li><li>- Data centre</li><li>- Vetting</li><li>- Safe space</li><li>- Trusted</li><li>- Physical space</li><li>- Virtual research environment</li><li>- Trusted research environment/Secure research environment</li><li>- Records</li><li>- Personal data</li><li>- Levels of sensitive</li><li>- Dual use data</li><li>- Dual use technology</li><li>- Human error</li><li>- Interoperable</li><li>- Destroying data</li><li>- Data disposal</li><li>- Classifications</li><li>- Sensitive data classifications</li><li>- Confidentiality</li><li>- Privacy</li><li>- Back up</li><li>- Restor</li></ul> |  |
|--|--|--|

		<ul style="list-style-type: none"> <li>- Security vetting</li> <li>- Preservation</li> <li>- Preservation system</li> <li>- Business continuity tape</li> <li>- Assurance</li> <li>- Administrative Data</li> <li>- Resupply (data resupply)</li> <li>- Deidentified</li> <li>- Safeguarding research</li> <li>- Obligation</li> <li>- Data Use Agreement</li> <li>- Stability</li> <li>- Dynamic consent models</li> <li>- Backend</li> <li>- Framework</li> <li>- GDPR</li> <li>- Contract</li> <li>- Formalised deletion method</li> <li>- Disclosure</li> <li>- Risk</li> <li>- Disclosure risk</li> <li>- Risk assessment</li> <li>- Consequences</li> <li>- Stakeholders</li> <li>- Affected people</li> </ul>	
<b>6</b>	5	<p>Meeting close</p> <ul style="list-style-type: none"> <li>- Call for presenters for RDA18</li> <li>- Link to poster for RDA18</li> <li>- Roadmap for first publication</li> </ul>	KS

	<p>Homework? We should prepare 5 to 20 terms each about sensitive data / sensitive data levels that are relevant in our disciplinary context to define the core of the terminology</p> <ul style="list-style-type: none"> <li>- Where do we use common words which can have different meanings?</li> <li>- Definition and context / polysemic identification</li> <li>- <b>What does 'sensitive' mean?</b> Can all data be thought of as sensitive to some degree? Disclosure risks</li> </ul>	
--	--	--

## 2 Useful Links

- [A living list of resources in our Wiki](#)
- [Zenodo community to which we are adding resources produced by our IG](#)
- [Previous session at RDA17](#)
- [Session report from RDA17](#)
- [RDA17 poster](#)
- [RDA 18 session proposal](#)
- [RDA Sensitive Data Interest Group 1st community meeting minutes \(22 Sep\)](#)
- [Charter](#)
- [RDA Principles](#)

## 3 Attendees

Name	Email	Institution	What is your discipline?	Will you be at RDA18?
Kristan Kang	kristan.kang@ardc.edu.au	ARDC	Health & medical	Yes

Biru Zhou	biru.zhou@mcgill.ca	McGill University	Social sciences; sensitive data in general	Yes
Lars Eklund	lars.eklund@uppmx.uu.se	SND/UPPMAX Uppsala university	Domain Specialist Sensitive data/Applications Expert HPC in Chemistry and Data management and Datasets	Yes
			Ecology / Pathogenic agents / agronomy / life sciences	
Aleks Michalewicz	aleksm@unimelb.edu.au	MDAP, University of Melbourne	Interdisciplinary research / HASS	Yes
Nichola Burton	nichola.burton@ardc.edu.au	Australian Research Data Commons	Interdisciplinary infrastructure	Yes
Stephanie Thompson	s.e.m.thompson@bham.ac.uk	Advanced Research Computing, University of Birmingham	Previously Biosciences. Provide RDM support to researchers from all disciplines but sensitive data particularly from Social Sciences, Medical Sciences & Psychology.	No but my colleague Aslam Ghumra will be.
Amy Pienta	apienta@umich.edu	ICPSR	Sociology	Yes
Steve McEachern	steven.mceachern@anu.edu.au	Australian Data Archive	Social sciences	Yes
Marion Shadbolt	marion@biocommons.org.au	Australian BioCommons	Bioinformatics	Not sure
Joerg Geiger	joerg.geiger@uni-wuerzburg.de	Ibdw, University and University Hospital of Wuerzburg	Biomedicine	Yes




4 Chat notes