

# Research Data Management in the Cloud: Current Practices, Tools, and Challenges

Working on data stewardship? Meet your peers!

November 30, 2017

Break-out session

Rob Gommans

Research Data Officer, Behavioural Science Institute, Radboud University



## What is it about and what NOT

- **Yes**, collaboration and sharing with external researchers while conducting your research.
- **No**, registering and archiving your data after publication for re-use purposes.
- **No**, ready-made solutions for collaboration and sharing, no one size fits all.



## Group discussions

- What major challenges need to be resolved to pave the way for efficient cloud collaboration in a scientific community?
- Personally identifiable information (PII) versus cloud collaboration: Why we should or should not allow storage of PII in the cloud.
- If we had to speed up funding for adequate cloud collaboration, what are the major benefits we shouldn't forget to mention?
- ...
- ...



# Challenges

- **Project management**
  - How to manage data properly?
  - How to keep things in sync?
  - How to deal with versioning and file change conflicts?
  - How to provide access?
- **Data protection**
  - How to prevent data loss?
  - How to prevent unauthorized access (data leakage)?
  - Backup / Data retention
  - Reliability / Trustworthiness cloud storage provider
- **Privacy and confidentiality**
  - Informed consent for collaborating and sharing
  - Data processing agreements
  - Privacy and data protection laws USA vs EU
  - Institute and national policies and regulations
- **Technical**
  - How to ensure sufficient storage space?
  - How to limit data transfers between locations?
  - Interoperability with other systems / tools



## Possible solutions

- Many cloud storage providers



**STACK**  
1000GB GRATIS online opslag



iCloud



tresorit



ownCloud



Google Drive



OneDrive



## Possible solutions

- Some online project management tools



## In the ideal world...

- Each research project registered in project management system (PMS)
- Scalable and interchangeable server- or cloud-based storage for PMS
- Local (read-only) copies of data from PMS (work offline, anytime, anywhere)
- Automated synchronization between local and PMS
- Automated encryption (local and server-side)
- Check-out (to make changes) / Check-in (commit changes) system
- Build-in versioning and history tracking (full auditability)
- Access control per collection (subfolder) or data object
- Master user access (by institute's data steward / officer)
- Possibility to add (custom) meta-data and comments
- Possibility to branch/fork
- Possibility to deposit final version (after branch/fork)
- Connectivity with existing (archiving) repositories
- Advanced reporting at the project and organizational level
- Compliant with data protection regulations with adequate safeguards to prevent data leaks



## Discussion / Examples

- How do you or your colleagues collaborate with external partners?
- What is offered and allowed in your organization?
- What other challenges do we face?
- ...





## Thank you for your attention!

- Please contact me if you have any questions.
- Rob Gommans, [r.gommans@bsi.ru.nl](mailto:r.gommans@bsi.ru.nl)



Data storage & transport

Documentation & metadata

Controlled sharing

Data management planning

Data collection

Alone

With internal partners

With external partners

With students

Data carpentry & analysis

Alone

With internal partners

With external partners

With students

Data publication

Data archival

Medium term ( $\leq 10$  years)

Long term ( $> 10$  years)

Data registration