Data Privacy Project

PROJECT PLAN

Huijser, D.C. (Dorien) & Moopen, N. (Neha)
RDM SUPPORT | UTRECHT UNIVERSITY

VERSION: 1.0, LAST EDITED: 2021-09-09

Contents

Introduction

BACKGROUND

Most scientific research contains some form of personal data. There is likely more such data than expected, with researchers being unaware of it. With the implementation of the General Data Protection Regulation (GDPR) in May 2018, stricter requirements apply to handling personal data and its sharing or publication. The number of (complex) questions on handling personal data in scientific research is increasing. Yet, information to resolve these questions is difficult to find, scattered, and/or discipline-specific. As such, there is room to better translate the legislation to scientific research practice.

AIMS

The aim of this project is to build knowledge and expertise on how researchers can and should deal with personal data. Specifically, the project has four goals:

1. Create a knowledge base for researchers to learn about (research) data privacy, including GDPR-compliant best practices and privacy-enhancing techniques.

2. Develop tools that can be used to address concrete data privacy-related issues. A tool repository will be curated to complement the aforementioned knowledge base.

3. Perform qualitative and quantitative investigations into (current) practices and needs of researchers in handling personal data and complying with the GDPR.

4. Build training materials and courses/workshops for researchers and support staff on handling personal data and using the available tools.

TARGET AUDIENCE

Researchers (at Utrecht University) are the primary target audience, while scientific support staff (at Utrecht University) are the secondary target audience. The project output will be available to external researchers and support staff, thereby making them a tertiary target audience.

UNIQUE VALUE PROPOSITION

The project output will be Findable, Accessible, Interoperable, Reusable (FAIR) for the target audiences. Particularly, it will be practical and actionable so that target audiences can apply it with ease.

Moreover, we will be collaborating across teams: data management, privacy, research engineering, legal affairs, open science, and everyone in between to ensure the project output is based on combined expertise and consensus.

RELATED DOCUMENTATION

All project documentation can be found in the Data Privacy Project Teams environment, the project website, the Data Privacy Handbook and the Project Open Canvas.

Deliverables

The following deliverables are planned in line with the (specific) project aims:

1. Data Privacy Handbook
2. Use Cases
3. Survey & Interviews
4. Dissemination

HANDBOOK

Created by Anconer Design
from Noun Project

SURVEY

Created by Jaime Carrion
from Noun Project

USE CASES

Created by Vectors Market
from Noun Project

DISSEMINATION

Created by priyanka
from Noun Project

DATA PRIVACY HANDBOOK

The Data Privacy Handbook is an open-source, community-driven, *living* Handbook (see also: the GitHub repository) about handling personal data in scientific research, in line with the GDPR. It consists of a knowledge base, tool repository and use cases to draw inspiration from. As far as possible, the content of the Handbook will be practical and actionable, without unnecessary detail and jargon.

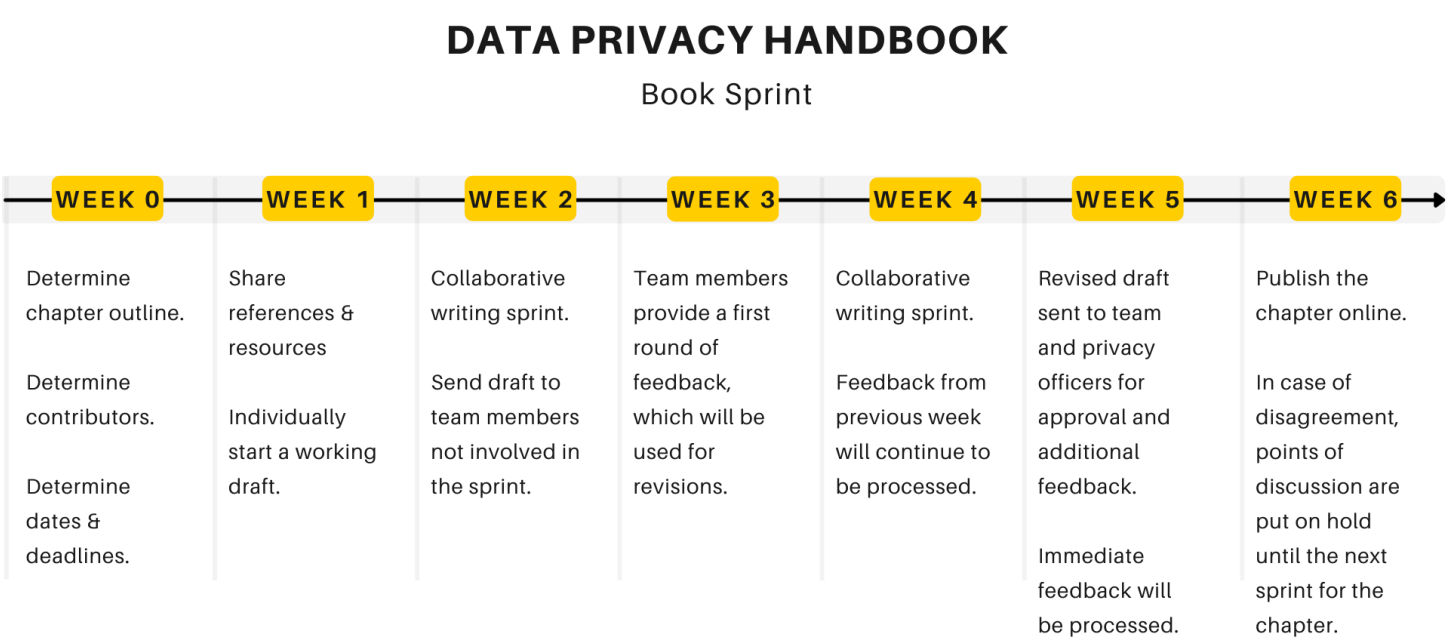
Knowledge on privacy-related laws, tools, and strategies is scattered and differs in depth and usefulness. With the Data Privacy Handbook, this information will be collaboratively synthesized into one resource, which users can consult at their convenience.

Book Sprints

For the initial writing of Handbook chapters, book sprints following a six-week cycle will be carried out. This format was chosen, because it has proven difficult to plan this deliverable across longer time frames. The six-week cycles will include concrete goals and deadlines, which will create clarity and increase accountability for everyone involved. That being said, the way of working within the six weeks and the contents developed are flexible: the focus is to reach a broad goal (e.g. complete a chapter), not to follow a detailed plan.

Workflow

Every book sprint will emphasize a single Handbook chapter. This will be more focused, than working on multiple chapters simultaneously. It will also prevent the trap of iterating a single chapter over months on end. Concretely, a book sprint may look like this:



Post-Publication

When a chapter is approved and published via GitHub, the content will be visible immediately. For individuals with a GitHub account, Contributing Guidelines are available if they would like to provide feedback on published chapters. For now, those without a GitHub account can provide feedback to the project coordinators via email. In the future, a form through which non-GitHub users can provide feedback may be created, or functionality and tutorials to work on GitHub without an account (e.g., via HackMD) will be provided.

USE CASES

In order to develop tools and solutions that can be applied to various research projects, we will take on use cases. Use cases are research projects that require support in tackling issues related to data privacy. They involve personal data in some form, but face challenges in safely collecting, analyzing, storing, or sharing the data. Possible solutions can range from workflows for de-identification or data sharing, contractual templates, to (research) software solutions.

By taking on use cases, the project will actively support researchers with ongoing challenges in handling personal data and simultaneously develop solutions that can reused by other researchers and support staff in the future.

Criteria for Use Cases

- The research project includes personal data in any form.
- A solution for de-identifying or safely handling personal data is required, for example pseudonymization, federated analysis, synthetic data, workflows, templates, etc.
- The potential solution can be FAIRified and reused by other researchers.
- The use case can be tackled between now and January 2023.
- The researcher(s) is/are prepared to work closely with the project team on the solution.

Strategy

A call for (new) use cases will be put out in September 2021. Moreover, questions and projects that posed to RDM Support, Research Engineering or stakeholders (e.g., privacy officers, data stewards) may be assessed for their suitability as use cases.

Additionally, ongoing projects at RDM Support or Research Engineering that meet the use case criteria can receive additional support – particularly in FAIRification.

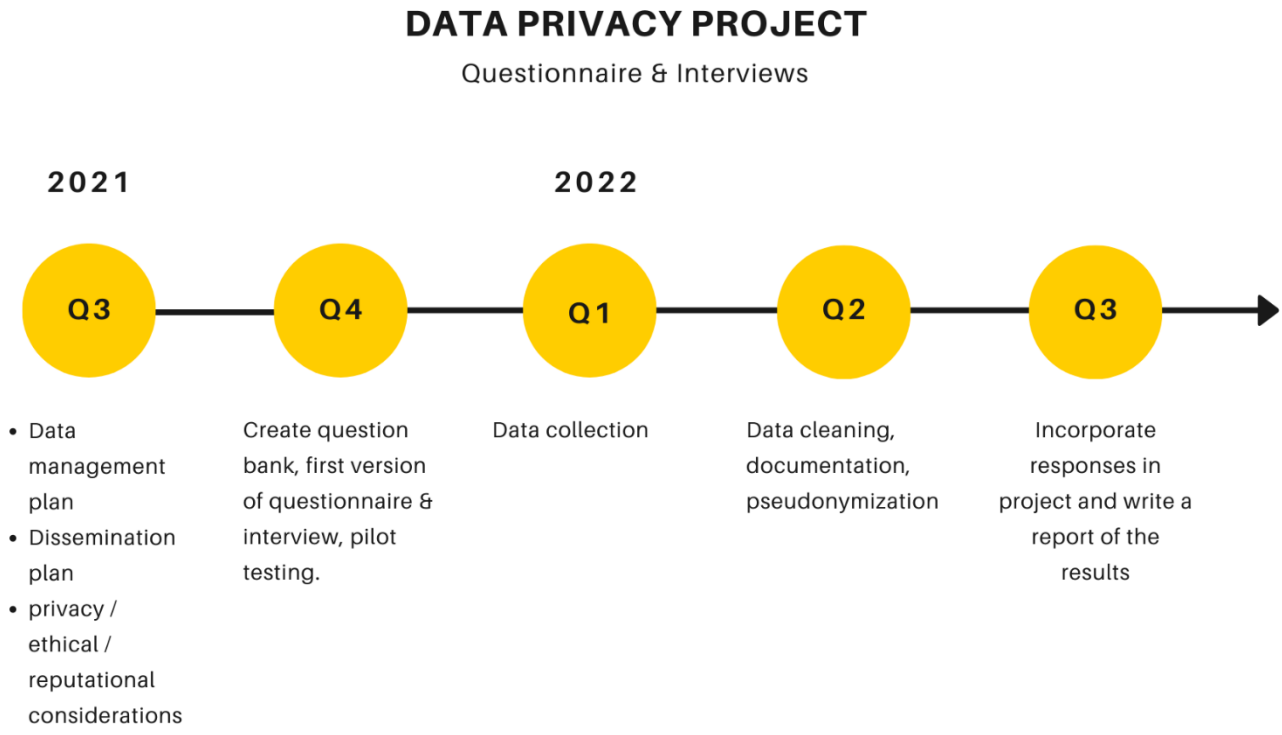
Use cases will be worked on in parallel and team members may join freely or by invitation depending on their expertise and availability. The specific steps in tackling a use case will likely vary across projects. In due course, the solutions will be FAIRified and described in the Data Privacy Handbook.

SURVEY & INTERVIEWS

A university-wide survey will be conducted among researchers, who form our target audience. The survey is meant to develop insight into researchers' experiences in handling personal data. It will cover, but need not be limited to: the privacy-related issues they run into, their current approach and practices, and their needs and requirements in terms of support.

The survey will include a questionnaire as well as interviews with researchers. This will provide both quantitative and qualitative input for reporting purposes, as well as an evidence-base to guide the project as it evolves.

The strategy and an estimated timeline for the survey is outlined below:



DISSEMINATION

As the project output becomes more defined, we will incorporate it into new and existing dissemination materials to educate the target audiences. Opportunities for dissemination may include:

- Workshops & courses on (research) data privacy. These may be online or offline, for self-study or otherwise.
- Presentations for researchers at the project, department, faculty-levels.
- Presentations for researchers depending on their position: from PhD candidates to PIs.
- Presentations for research communities: for example, the Open Science Community Utrecht (OSCU).
- Presentations for UU and (inter)national data management communities.
- Setting up a network of Privacy Champions within faculties, similar to the OSCU Ambassadors or Faculty Open Science Track (FOST) representatives.

How do we make it work?

ROLES AND RESPONSIBILITIES

Steering Group

The steering group meets with the project coordinators and will place the project and its output at a strategic level in the institution. The steering group for this project consists of:

- Kim Bergmans, Department Manager, Academic Services, University Library
- Frank Heere, Department Manager, Information & Technology Services

Project Coordinators

The project coordinators are data managers who function as the main point of contact for the project. They organize meetings and communicate with stakeholders, monitor project progress, and work on all deliverables of the project themselves. The project coordinators are:

- Neha Moopen, Data Manager, University Library, 2 days/week
- Dorien Huijser, Data Manager, University Library, 3 days/week

Project Team

The project team works actively on the project together with the project coordinators / data managers. Members volunteer their time and therefore, contributions and time investments may differ. New team members may join anytime and existing team members may leave over time. At the moment, the project team consists of:

- Sanne Kleerebezem, Privacy Officer, University Library & Veterinary Sciences
- Francisco Romero Pastrana, Privacy Officer, Geosciences
- Jacques Flores, Data Consultant, University Library

- Saskia van den Hout, Information Security Officer
- Danny de Koning-van Nieuwamerongen, Data Manager, University Library
- Ron Scholten, Data Manager, University Library

Stakeholders

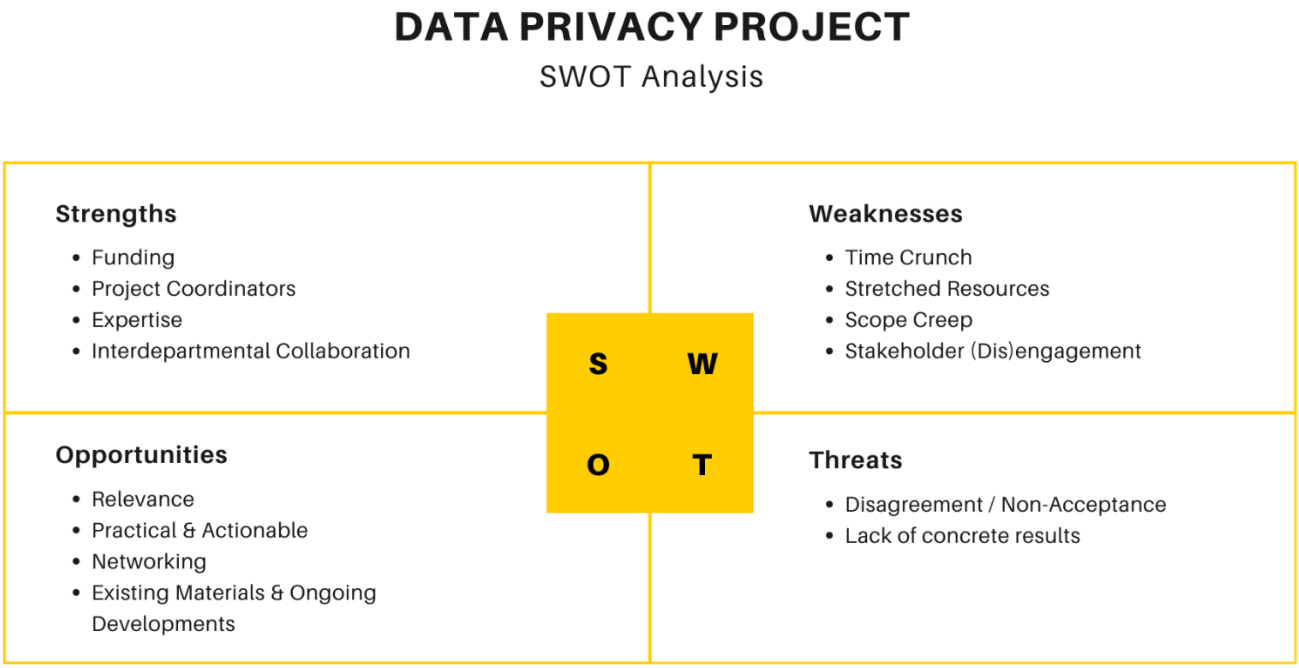
There are numerous stakeholders in this project, including the Privacy Officers & Research Engineers. For an overview of all stakeholders and their involvement, please refer to the Stakeholder Analysis (see Appendix A).

Communication Plan

A detailed Communication Plan (see Appendix B) has been drawn up to ensure all stakeholders are kept informed on and involved in developments in the project.

STRENGTHS, WEAKNESSES, OPPORTUNITIES, AND THREATS (SWOT)

For a strategic perspective on the project, an overview of the project's Strengths, Weaknesses, Opportunities, and Threats (SWOT) is presented below.



Strengths

These are internal factors that set the project up for success.

- Funding: The project has funding through the UU Research IT program and the NWO-DCC funding instrument. It provides for a (full-time) data manager and research engineer to work on the project.
- Project Coordinators: Two data managers serve as 'coordinators' for the project, in addition to working on the deliverables. They maintain an overview of the project from its strategic standpoint to the day-to-day activities.

- Expertise: There is no dearth of expertise in the project team, which consists of specialists in the fields of data management, privacy, research software, information security.
- Interdepartmental Collaboration: The project involves significant collaboration between departments, which is advantageous compared to these departments working in 'silos'. The project team effectively functions an active working group on the topic.

Weaknesses

These are internal factors that (could) make it difficult for the project to succeed.

- Time Crunch: Some deliverables, such as the Handbook, are planned with (relatively) strict schedules. However, delays may be unavoidable considering the number of stakeholders and timelines can therefore become tighter.
- Stretched Resources: All team members and stakeholders, apart from the project coordinators, will be contributing on a voluntary basis. There is a risk of contributors realizing they're unable to contribute due to limited time and capacity.
- Scope Creep: Given the breadth of the project, uncontrolled growth of the project's scope is a possibility – to the point where deliverables cannot be effectively managed by the project coordinators and team. Concretely, we may we may take up too many use cases, take too long finalizing a Handbook chapter, have too many people involved, or add another deliverable to the project.
- Stakeholder (Dis)engagement: Considering the number and range of stakeholders, disengagement – for example, due to misalignment of expectations and goals - should be prevented as it would reduce productivity.

Opportunities

Opportunities are the external factors that are likely to contribute to the project's success.

- Relevance: There is greater awareness of data privacy issues and more questions on the topic arise every year. This project fulfills a need from researchers and support staff.

- Practical & Actionable: There is a need for more hands-on resources on data privacy, the project output is intended to be easily picked up and implemented by researchers.
- Networking: Through collaboration, the network of data & privacy professionals and researchers can be further strengthened.
- Existing Materials & Ongoing Developments: Output from completed and current RDM/IT/privacy initiatives and policies already provide a basis for the project.

Threats

These are external factors that, if they were to occur, could harm the project.

- Disagreement / Non-Acceptance: Widespread support and consensus will be necessary for the project to succeed. Stakeholders' disagreement or non-acceptance of the deliverables would hinder progress and prevent dissemination of the output.
- Lack of Concrete Results: Given the range of stakeholders and expertise, (extended) abstract discussions, aimed at integrating different perspectives, could inadvertently prevent achieving concrete goals and results.

POTENTIAL SOLUTIONS

- Stretched Resources:
    - Make contributions explicit and concrete: giving specific assignments, setting deadlines and booking agenda time are all ways to establish this.
    - Connect with stakeholders' existing work to avoid double work.

- Scope Creep:
    - Determine clear boundaries, e.g., a minimal viable product, a maximum amount of use cases, etc.
    - Get support for project management and communication.
    - Set quarterly goals and evaluate how they contribute to the main project aims.
    - Ask important stakeholders for feedback on the project planning.

- Stakeholder (Dis)engagement:
    - Communicate clearly about the project, by having monthly team meetings, monthly email updates and sticking to one communication channel.
    - Work on tasks (e.g., the Handbook) together instead of individually. This will increase feelings of active involvement and wide-spread support and forces team members who are joining to spend time on the project.

- Disagreement / Non-Acceptance:
    - Involve privacy officers and faculty data support early on, by having them contribute actively to the project (thus becoming a project team member) and by including them in regular updates of the project.
    - Present the project at several occasions, e.g., to library staff, at faculties, external data support meetings, etc. to gain notoriety (see the Communication plan).
    - Meet with the Data Protection Office to get their acceptance of the project and the deliverables.

- Lack of concrete results:
    - Use explicit planning and communication to force to deliver a result.
    - Put the point of discussion on a backlog to go back to at a later time.
    - List several solutions if there is no agreement on a single one.
    - Use voting to make a decision.
    - Consult other relevant experts who may have a deciding voice.

Appendix A. Communication Plan

This communication plan concerns mainly communication about the status of the project. For the communication of the project output and results towards the target audience, we will have a dedicated dissemination plan.

WHAT	WHY	HOW	FREQUENCY	AUDIENCE	OWNER
WRITTEN COMMUNICATION					
Project update	Update the main stakeholders about the status of the project	Email	Monthly	Core team, Steering group, Interested stakeholders	Coordinators
News post	Comprehensive news update that is public and shareable as well.	Project website	Quarterly	Public	Coordinators
MEETINGS					
Steering Group meeting	Update about project and get input on how to proceed	Scheduled meeting	Quarterly	Steering group	Coordinators
Core team meeting	Discuss project progress and task division	Scheduled meeting	Monthly	Core team	Coordinators
Meeting with privacy officers	Update about project and get input on content	Scheduled meeting	Quarterly	Privacy Officers	Coordinators
Meeting with security officers	Update about project and get input on content	Scheduled meeting	Biannually	Security Officers	Saskia
Research Engineering meeting	Update about project and get practical input and strategy how to involve RE	Scheduled meeting	Monthly	Research Engineering, Martine de Vos	Coordinators
Coworking / Open Meeting	Work together on progressing the project	Scheduled afternoon meeting	1-2 times	Public	Coordinators

Verbal/semi-formal updates	Update about the project and get input on connecting	D-lunch	Monthly, when necessary	RDM Support	Coordinators
Legal Affairs meeting	Update about the project and input	Scheduled meeting	Biannually	Legal Affairs	Coordinators
PRESENTATIONS					
Presentation	Update, feedback, networking	AS meeting	1-3 times	Library staff	Coordinators
		OSCU event		OSCU	
		Scheduled meeting		Faculty Open Science Teams, FAIR data and software fellows	
Presentation	Update, feedback, networking	Scheduled meeting	Biannually/Quarterly	Faculty liaisons	Coordinators
Presentation	Update, feedback, networking	RDM Expert meeting or UDMC	Biannually	RDM Support outside the library	Coordinators
ONLINE NETWORKING					
- Summary of and link to the quarterly news post - Call for proposals - Call for survey responses	Spread the word about the project and its progress and to get input on the project	Twitter: RDM support and personal accounts	Quarterly + additional calls when necessary	Twitter followers	Coordinators
		Slack: OSCU and DTL-DSIG		OSCU and DTL-DSIG	Coordinators
		GitHub Discussions		Public	Coordinators
		Intranet and OSCU website		Interested UU employees and researchers	Coordinators
		Mailing lists: RDM newsletters, faculty newsletters, etc.		Interested UU employees and researchers	Communication Coordinators
		RDM website		Interested UU researchers	Communication Coordinators
Project summary	Spread the word about the project and ask for input	Data management mailing lists (LCRDM, JISC)	Only when relevant	Interested external support staff	Coordinators

Appendix B. Stakeholder Analysis

A stakeholder is a person or an organization that is actively involved in the project or is positively or negatively impacted by it. A key stakeholder is any person who determines the success or failure of the project:

- Decision: Make the decisions that control or influence the project budget.
- Authority: Have the authority to grant permission to proceed with the project.
- Need: Directly benefit from or are impacted by the project and consequently need to know all about it.
- Connections: Are connected to the people, money, or resources required to remove roadblocks or exert influence to ensure project success.
- Energy: Have positive or negative energy that could affect project success.

STAKEHOLDER	ROLE IN PROJECT	KEY?	METHODS OF COMMUNICATION
Steering group	Decision, Authority, Connections, Energy	Yes	Regular Steering group meeting
RDM Support	Authority, Need, Connections, Energy	Yes	Regular Core team meeting
Research Engineering / Research & Data Management Services / ITS	Decision, Need, Connections, Energy	Yes	Regular meeting with Martine de Vos
Privacy Officers	Authority, Need, Connections, Energy	Yes	Meeting(s) with Privacy Officers
Information Security	Authority, Energy	Yes	Via Saskia Newsletter(s) and email
Legal Affairs	Authority, Energy	Yes	Meeting with Legal Affairs
Faculty Liaisons	Need, Connections, Energy	Yes	Presentation Newsletter(s) and email

Faculty Data Managers	Need, Connections, Energy	No	Presentation or one-on-one meetings Newsletter(s) and email
Ethical Committees	Energy	No	Newsletter(s) and email
Research Support Officers	Connections	No	Newsletter(s) and email
Communications department	Connections	No	Newsletter(s) and email
Academic Services	Connections, Energy	No	Academic Services meeting
Open Science Community Utrecht (OSCU)	Need, Connections, Energy	Yes	Newsletter(s) and email Slack Presentation
Faculty Open Science Teams (FOSTs)	Connections, Energy	No	Newsletter(s) and email Presentation
Open Science Fellows (FAIR data & software track)?	Need, Connections	No	Newsletter(s) and email Presentation
OSCU Ambassadors?	Need, Connections	No	Newsletter(s) and email Presentation
Researchers	Need, Connections, Energy	Yes	Presentation Newsletter(s) and email Connections (OSCU, privacy officers, data managers, etc.)
Graduate Schools	Need, Connections	No	Newsletter(s) and email
LCRDM	Need, Connections	No	Newsletter(s) and email
DTL-DSIG	Need, Connections	No	Newsletter(s) and email Slack
UKB	Need, Connections	No	Newsletter(s) and email
RDM Support & Privacy Officers at other universities	Need, Connections, Energy	No	Newsletter(s) and email Slack