

How to Set Up a Dataspace

Roseman Labs – The Netherlands

Welcome!

Many organizations such as government agencies, municipalities and healthcare providers aim to coordinate their services more efficiently, e.g. in a consortium, or across the sector or value chain.

This document focuses on one part of this ambition, namely the use of a so-called dataspace as a building block to accelerate such collaborations. In our definition, a dataspace provides a trustworthy and efficient set-up for the use of data for the improvement of a service, or for research or policymaking. A dataspace enables parties to jointly share and analyze data, with strong guarantees for privacy and data control for all parties involved. It facilitates shared situational awareness and promotes further collaboration by making data readily available in a trustworthy manner.

Contents

1. Introduction	3
2. Governance of the Dataspace	4
3. Parties and Their Responsibilities.....	5
4. Short-term Plan	6

1. Introduction

Shared situational awareness is valuable for optimizing services in a consortium, network or value chain. By deploying a dataspace, the partners create one shared and controlled view of the available data. This draft project plan aims to help you set up a dataspace efficiently.

This project plan is written in the context of the Roseman Labs platform, but some of our suggestions are more widely applicable. With Roseman Labs, two important principles are that:

1. The dataspace offers strong guarantees for the privacy/confidentiality for the data of the partners.
2. The partners control what happens with their own data. Organizations can create insights with data of others, without revealing the underlying data and while maintaining control over their own data.

By using a dataspace based on the Roseman Labs platform all parties can:

- Make data available at a detailed level to the consortium.
- Perform analyses on the joint data, as if it were one source.
- Offer strong privacy and data protection guarantees to all involved.
- Maintain control over who performs which analysis.

The solution simplifies data collaborations because it addresses key aspects of setting up these collaborations:

- Purpose limitation, data minimization and privacy are inherent – or by design – because the data is always encrypted (and protected) during processing. All parties maintain control over what is done, lowering the barrier to share sensitive data.
- The dataspace can be used directly for those use-cases where structured tabular data of sufficient quality is available.
- A secure (and intuitive) collaboration environment stimulates a positive learning cycle based on initial results, which motivates partners to gradually expand the scope.

Our experience shows that these aspects promote working iteratively and motivate partners to work towards further standardization (of data and processes), and to increase the scope of the collaboration.

2. Governance of the Dataspace

A dataspace often supports a broader consortium in:

- Developing data-driven products or services.
- Developing the research agenda/roadmap.
- Making quality data available for combined situational awareness.
- Establishing common data definitions.
- Performing analyses and reviewing the outcomes.
- Coordinating, and standardizing, the (joint) data products and services.

Typically, the dataspace is a building block in a larger consortium. The consortium or the dataspace is governed by a Steering Group consisting of representatives from the consortium partners. The Steering Committee typically determines the dataspace scope and mandate, ensures compliance, prepares and executes the legal agreements, and approves and reviews the research roadmap.

Below the Steering Group is typically a Working Group, consisting of subject matter experts. This Working Group typically fulfils the most important analytical tasks, such as: Determine the high-level research agenda, define first analyses, define which variables to extract from the source data, propose which variables to standardize, prepare and execute analyses and review analysis outcomes.

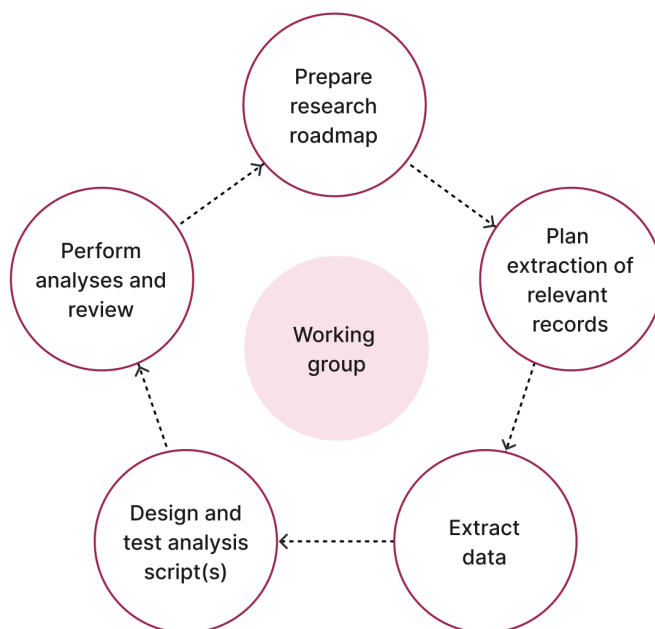


Figure 1 Activities of the Working Group

The next two chapters describe and plan these activities in more detail.

3. Parties and Their Responsibilities

At a high level, we break down the roles and time required for setting up and running a 'typical' dataspace. Of course, the final setup depends on your context. Often, there are one or a few organizations that coordinate the consortium and provide a project leader to coordinate and manage the dataspace. This is considered in the setup below.

For starting up a Dataspace, plan approximately 1-3 months. The following capacity is required:

- Project leader: 0.5-1.0 day/week (often from the coordinating organization)
 - Onboard and inform stakeholders in the partner organization(s)
 - Formulate objectives (short term)
 - Prepare the kick-off meeting for the dataspace
 - Define conditions for success and report on progress
 - Coordinate privacy/security/compliance reviews (see below)
- Privacy specialist: 0-2 hours/week (at each participating party)
 - Execute Data Privacy Impact Assessment (DPIA) and collaboration agreement
 - Share DPIA as a reference with participants, for their review
- Security specialist: 0-2 hours/week (from the coordinator of the dataspace)
 - Carry out security risk assessment
 - Share security risk assessment with participants, for their review

During the running of the Dataspace, working towards subsequent analysis results:

- Project leader: 0.5-1.0 day/week
 - Define and detail/shape the use cases
 - Ensure efficient collaboration among the dataspace partners (see below)
- Researcher/data specialist: 0.5-1 day/week
 - Specify use cases and define detailed analyses
 - Prepare data for use (test extraction, test data quality)
 - Verify analysis outcomes, iterate on the insights
 - Specify data standardization and uniform working methods

This is not a comprehensive overview but helps to get going with a minimum set of tasks. Other topics may include setting up machine-to-machine data pipelines, designing reporting dashboards, for example.

4. Short-term Plan

Below you will find a hypothetical step-by-step plan for implementing the dataspace as part of a wider consortium or partnership. Several activities may be more logical to make part of the overarching program or consortium, such as standardizing operating procedures between parties, but we have included some as suggestions.

Month 1:

1. Set up the Steering Group and Working Group: Gather a team of representatives for the Steering Group and experts for the Working Group. The Steering Group members determine the goals and means; and monitor progress and compliance. The Working Group members shape the analysis roadmap and perform the necessary data science/engineering actions. Plan recurring, e.g. monthly, meetings for both groups.
2. Define Project Goals: Determine the specific goals and expected outcomes for the data workshop.
3. Involve Privacy and Security experts early: Start review of the dataspace setup and technology for compliance with GDPR, security and other legal requirements.
4. Investigate first needs: Start an investigation in to the current data situation, availability bottlenecks and short-term data product needs (quick wins) within the dataspace.
5. Define short-term focus: Determine the first result. Typically, this involves building a first understanding of the basic characteristics of the combined data (population characteristics, data quality).

Month 2:

1. Conduct privacy and compliance reviews: Brief privacy/security/compliance officers, define the necessary agreements, support execution of the privacy



Roseman Labs

impact assessment (DPIA) and security reviews. Often several (online) workshops are conducted for Q&A, drafting and agreeing on the documents.

2. Define first data extraction: Define which data fields are required for the short-term focus and which variables to extract from the source systems. If helpful, plan per quarter which data variables are required from all partners. (Maintain a calendar of data variables required.)
3. Prepare and test data extraction from existing systems: Evaluate the existing data sources and identify how partners make the data available to the dataspace (via CSV table, machine API call or surveys).
4. Start data standardization: Assemble a sub-team within the Working Group that focuses specifically on data definitions and standardization of the data format (variables).
5. Initial analysis: Define an initial analysis of the data workshop, taking into account the needs of parties and focusing on concrete results in the short term.

Month 3:

1. Collect first dataset: Collect the first set of tables for the initial analysis. Determine the associated meta-data and generate dummy data to design test analyses.
2. Identifying initial standards: Have the standardization sub-group investigate existing standards that can serve as a basis for the initial standards set.
3. Finalize privacy, compliance and security review: Complete evaluation of GDPR, security and other requirements. If required, install a privacy review board that monitors the scope and compliance of the dataspace, and extensions to the scope (and the DPIA). This privacy review board typically consists of the privacy officers of the partners and meet once per quarter.
4. Set up broader communication plan: Create a communication plan to keep all stakeholders informed and actively involved. Consider announcing the expected official launch date of the dataspace. (In this plan, the official launch is in Month 6, see below.)

Month 4:

1. Run first analysis on production data: Once the privacy, security and compliance assessment are completed, conduct the first run with production data. In case of a larger consortium, consider doing a test run with a manageable sub-group of partners.
2. Write first data standardization proposal: Have the data standardization sub-group present their first proposal. Consider also involving the providers of the IT (e.g. medical records) systems as audience for this presentation. This creates momentum for data quality improvement and mobilizes the relevant parties.
3. Conduct feedback rounds: Organize feedback sessions with stakeholders to evaluate the progress so far.

Month 5:

1. Implement dataspace improvements: Implement improvements based on feedback from the first production run. (E.g., related to data extraction, documentation, way of working)
2. Define follow-up analyses: Determine a roadmap of analyses and required data fields with substantive experts.
3. Conduct training and workshops: If needed to promote new data products or services, organize training sessions or workshops for interested parties.
4. Implement standardization: Start implementing the proposed (short-term) standards in the organizations involved.

Month 6:

1. Evaluate: Evaluate the performance and results of the dataspace to date.
2. Future planning: Start planning the next phases of the dataspace, including any expansions or improvements.
3. Launch dataspace to the public: After first successes, launch the dataspace for a larger audience. Present the first results and roadmap.



Good luck!

We wish you good luck in setting up your dataspace and sincerely hope this document helps you to get started.

If you have questions or suggestions about this document or your journey, let us know.

Feel free to share your thoughts with us at info@rosemanlabs.com, visit the documentation on our [website](#) or connect with us on [LinkedIn](#).