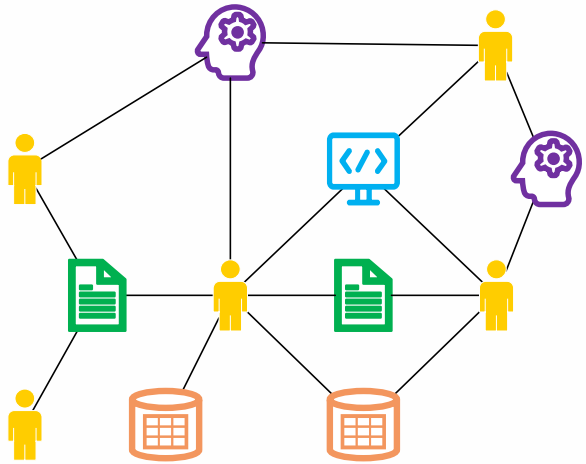
Discovering insights from cross-organizational research information and collaborations

A pilot project using Ricgraph



Utrecht University

Ricgraph

Rik D.T. Janssen
Version 2.0, June 11, 2025

SUMMARY

This document describes a pilot project to demonstrate how a knowledge graph can provide insights into research relations and collaborations, and how it can optimize the quality of research information. It aims to help shape the infrastructure for research information that is currently high on the European agenda.

The approach we take is to collect research information from participating organizations and from multiple source systems, and to combine it in a single graph. By organizing research information in a network of connected items and relations, users can infer new relations, relations that are not present in any of the source systems.

The result is openly available to anyone, in accordance with the open science principles. We strongly believe that sharing and caring about each other's research information is *the* step forward, even if it is metadata from another organization. Only in this way it is possible to create a validated source of research information for the entire research information landscape, that can be used and trusted by anyone. We have been inspired by WorldCat, a bibliographic library database, contributed to and used by tens of thousands of libraries worldwide for decades.

In this pilot project, we are starting with a few Dutch universities. Utrecht University will facilitate these organizations to construct a graph containing their research information. It will include information on researchers, teams, their results, (sub-)organizations, collaborations, (optional) skills, (optional) projects, and the relations between these items. This graph is constructed based on the contents of the Pure Research Information System of each participating organization, supplemented with research information from sources such as OpenAlex, Yoda, and the Research Software Directory. The participating organizations will explore the potential and will investigate what can be learned from research information in a single graph.

This pilot project uses *Ricgraph* (www.ricgraph.eu). With it, a user can explore relations between research information stored in various source systems within a single graph, as described above. This allows users to discover relations they were not aware of.

This project consists of three parts:

1. Utrecht University will make available an Open Ricgraph demo server containing research information from the participating organizations.
2. Pilot A: Participating organizations can enrich Pure data using Ricgraph and BackToPure. *BackToPure* can insert (enrich) items from an organization that are absent from the Pure of that organization but are present in another source, back into the Pure of that organization.
3. Pilot B: Participating organizations can explore collaborations between sub-organizations (faculties, departments, chairs) using Ricgraph.

Call to action

If you or your organization would like to participate in this pilot project, please continue reading and contact Rik Janssen at r.d.t.janssen@uu.nl. You can also contact him for more information.

Discovering insights from cross-organizational research information and collaborations: A pilot project using Ricgraph, Rik D.T. Janssen (2025), https://doi.org/10.5281/zenodo.15637647.

TABLE OF CONTENTS

1 OVERVIEW PILOT PROJECT

This document describes a pilot project to demonstrate how a knowledge graph[1] can provide insights into research relations and collaborations, and how it can optimize the quality of research information. It aims to help shape the infrastructure for research information that is currently high on the European agenda (e.g. EOSC, EUDAT).

The result is openly available to anyone, in accordance with the open science principles. We strongly believe that sharing and caring about each other's research information is *the* step forward, even if it is metadata from another organization. Only in this way it is possible to create a validated source of research information for the entire research information landscape, that can be used and trusted by anyone. We have been inspired by WorldCat, a bibliographic library database, contributed to and used by tens of thousands of libraries worldwide for decades.

1.1 Rationale

The approach we take is to collect research information from participating organizations and from multiple source systems, and to combine it in a single graph. By organizing research information in a network of connected items and relations, users can infer new relations, relations that are not present in any of the source systems.

In this pilot project, we are starting with a few Dutch universities. Utrecht University will facilitate these organizations to construct a graph containing their research information. It will include information on researchers, teams, their results, (sub-)organizations, collaborations, (optional) skills, (optional) projects, and the relations between these items. This graph is constructed based on the contents of the Pure Research Information System of each participating organization, supplemented with research information from sources such as OpenAlex, Yoda, and the Research Software Directory.

The advantages of having all this information in one graph are, among others, to infer new relations between items, relations that are not present in any of the source systems. This results in new insights. Examples are:

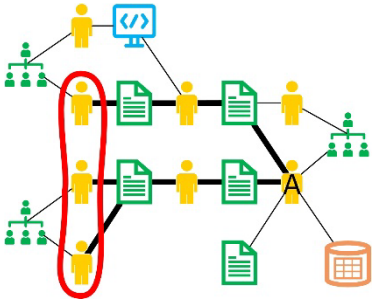
- collaborations between sub-organizations or between researchers;
- common research results between sub-organizations;
- finding research information about your own organization that is in other systems but not in your own system (enriching).

In this pilot project, participating organizations will explore this potential and will investigate what can be learned from research information in a single graph.

1.2 Use cases of research information in one single graph

1.2.1 Researcher use case

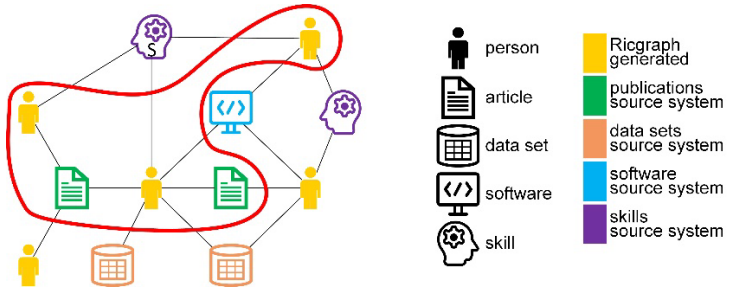
As a researcher *A*, I want to find researchers from other universities that have co-authored publications written by the co-authors of my own publications, so that I can read their publications to find out if we share common research interests.



1.2.2 Journalist use case

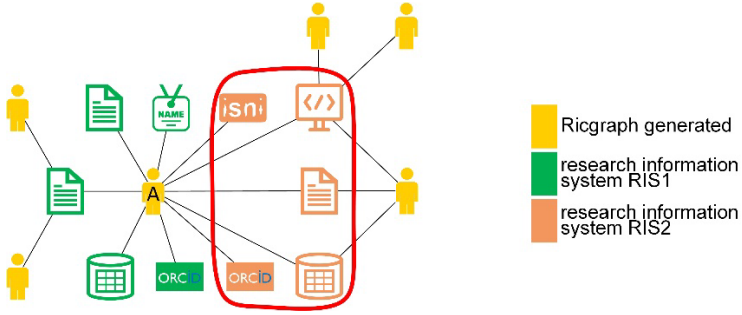
As a journalist, I want to find researchers with a certain skill *S* and their publications, so that I can interview them for a newspaper article. Example skills (or research interests) can be: *climate change* or *stem cells*.

[1] A knowledge graph is a knowledge base that uses a graph-structured data model or topology to represent and operate on data. They are often used to store interlinked descriptions of entities – objects, events, situations or abstract concepts– while also encoding the relationships underlying these entities (https://en.wikipedia.org/wiki/Knowledge_graph).



1.2.3 Librarian use case

As a librarian, I want to enrich my local research information system with research results from person *A* that are in other systems (in orange, *RIS2*) but not in ours (in green, *RIS1*), so that we have a more complete view of research at our university.



1.3 Structure

In this pilot project, we use *Ricgraph* (www.ricgraph.eu) for the single graph. You can read more about Ricgraph in section 2.1.

The pilot project consists of three parts:

- Utrecht University will make available an *Open Ricgraph demo server* (chapter 2).
- Pilot A: Participating organizations can enrich Pure data using Ricgraph and BackToPure (chapter 3).
- Pilot B: Participating organizations can explore collaborations between sub-organizations (faculties, departments, chairs) using Ricgraph (chapter 4).

1.4 Aims

Aims Open Ricgraph demo server (chapter 2):

- Demonstrate the usefulness and limitations of having research information in one single graph.
- Analyze what can be learned from research information in one single graph.
- Explore the consequences of the current setup of Ricgraph for pilots A and B. Consider alternative setups.
- Propose possible improvements for software, source systems, workflows, organizational procedures, etc.
- Develop and explore new use cases on common research information of participating organizations based on the experience gained with Ricgraph.
- Develop and explore scripts using the Ricgraph REST API.
- Co-development Ricgraph.

Additional aims pilot A (chapter 3):

- Help participating organizations to have a more complete view about research at their own organization in their Pure.
- Get a clear understanding of how the administrative processes and workflow need to be set up in order to benefit from connecting the Pure of an organization to the graph.
- Co-development BackToPure.

Additional aims pilot B (chapter 4):

- Exploring the various methods that sub-organizations can be connected to persons and research results, and what this means for queries that can be answered.

1.5 Who can participate

At the start of this pilot project, we will focus on Dutch universities using the Research Information System Pure. We aim for about 3 universities to participate in pilot A and B, and more are of course welcome to join. We

plan to extend this number and type of organizations at some point in time during the pilot.

1.6 Getting ready to participate

- An organization that would like to participate needs to consider their organizational legal and privacy policies. Delft University of Technology and Utrecht University have gone through this process, which will very probably save other organizations quite some work. Please read chapter 5 to find out what needs to be done.
- In terms of the GDPR, a participating organization will be the "Controller", see section 5.2.

1.7 Participation during the pilot project

- Participation in the project workgroup (section 1.9.1).
- Time for experimentation.
- Time to share experimentation results.

1.8 Timeline

- Start: As soon as one organizations has agreed to participate. Other organizations will be added to the Open Ricgraph demo server as soon as they have agreed to participate.
- End: March 31, 2026.
- At the start of 2026, a continuation of the pilot will be discussed with the participating organizations.

1.9 Organization

The plan is to organize this pilot project in an informal manner.

1.9.1 Project workgroup

Each of the participating organizations will have a representative for that organization in the project workgroup. That person will be an advocate in his/her own organization for open research information, and for sharing and caring about each other's research information as described at the start of this chapter. Also, Utrecht University will have a representative in this workgroup.

The main task of this workgroup is to oversee the project, and to help to make sure there is progress. The workgroup will guide the project in any direction that may be useful. It will also discuss topics such as the usefulness of having Dutch research information in one single graph, or analyze what can be learned from research information in one single graph.

At the start of the project the members of the workgroup will make a list of topics to be pursued. This list is very likely to change during the project, since we will learn from our progress. At some point in time the workgroup may create intermediate progress reports or publications. At the end of the project a final report will be created, and a final publication will be written.

1.9.2 Software development community

The project workgroup will try to create a community around the development of Ricgraph and BackToPure (both are open source). Anyone who would like to participate is welcome to do so.

1.9.3 Utrecht University – ITS – RDMS

RDMS (Research and data management services) is a department of the central IT department of Utrecht University. Ricgraph and BackToPure have been developed at RDMS. In terms of the GDPR, Utrecht University will be the "Processor" for this pilot. Utrecht University – ITS – RDMS will facilitate the pilot.

RDMS will facilitate the harvester VM and explorer VM with Ricgraph on SURF Research Cloud, so that harvesting is done and Ricgraph Explorer and the Ricgraph REST API are available. As this is a pilot project, no guarantees of availability, software modifications, etc. can be given.

During the pilot, the creators of Ricgraph and BackToPure will do their best to improve their software. We hope the software development community (see previous section) will help with this. All changes will be integrated in the Open Ricgraph demo server and/or BackToPure as soon as they are available.

1.10 Budget

Utrecht University has chosen to invest in discovering insights from cross-organizational research information and collaborations using Ricgraph. That means that for now, Utrecht University does not request a financial contribution to facilitate this pilot project to the participants.

The expected expenses are at least:

- Utrecht University: Personnel for e.g. organization, development of software, maintenance of the harvester and explorer Virtual Machines at SURF Research Cloud.
- Utrecht University: Computing time at SURF Research Cloud for the harvester and explorer Virtual Machines.
- Any participating organization: Personnel for the project workgroup and (optional) participation in the software development community.

1.11 Related work

Recent initiatives to collect research information are:

- 2017-2020: a project Freya that aimed to develop a PID graph.
- 2018: a project about an ID resolver for Pure.
- 2021: a Open Knowledge Base project, with a Feasibility study and a Business case Open Knowledge Base/Open research information agenda.
- 2021-2022: a project for a National Roadmap for Persistent Identifiers, and a report Towards a national PID roadmap.
- 2023: a National PID graph pilot, with a final report.
- 2024: a Program Open Research Information 2024-2027.
- 2024: an initiative for an EU Strategy for Research Infrastructures.

None of these initiatives has yet led to open and actively maintained software that collected research information from multiple organizations and from multiple source systems.

2 OPEN RICGRAPH DEMO SERVER

2.1 Ricgraph

Ricgraph (www.ricgraph.eu), also known as Research in context graph, enables the exploration of researchers, teams, their results, (sub-)organizations, collaborations, skills, projects, and the relations between these items. It is open source software and available on GitHub (https://github.com/UtrechtUniversity/ricgraph). Documentation and videos can be found on https://docs.ricgraph.eu.

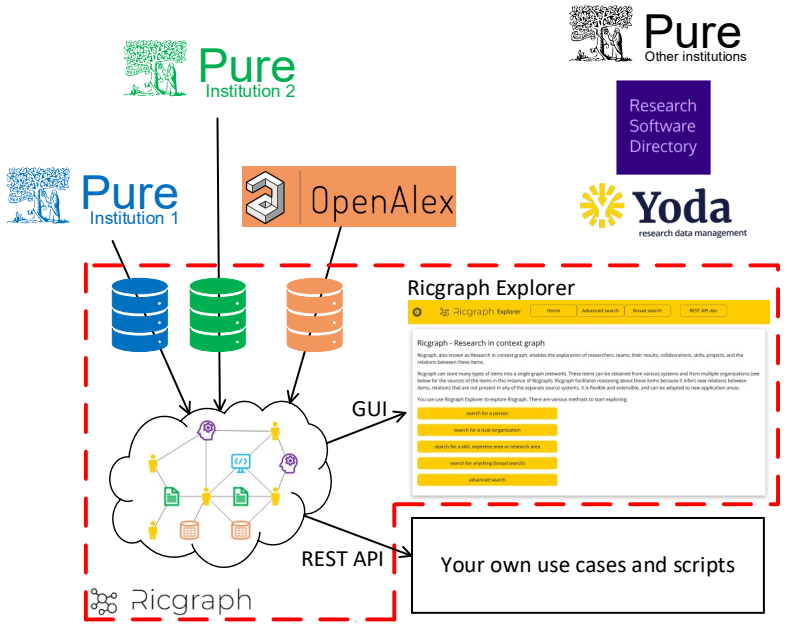
Ricgraph can store many types of items into a single graph. These items can be obtained from various systems and from multiple organizations. Ricgraph facilitates reasoning about these items because it infers new relations between items, relations that are not present in any of the separate source systems. Ricgraph is flexible and extensible, and can be adapted to new application areas. Ricgraph has been developed by Rik D.T. Janssen of Utrecht University.

It is possible to construct a single graph for one organization from multiple source systems, as well as a single graph from several organizations from multiple source systems. The former can be done by any organization in isolation. The latter seems to have a large potential, since e.g. (in the case of universities) Dutch universities collaborate a lot, and these collaborations can be explored using a graph.

In the remainder of this text, it is assumed that the reader is familiar with Ricgraph and its features. If not, please read:

- For a short introduction, the presentation: Ricgraph - Research in context graph, https://doi.org/10.5281/zenodo.12634234.
- For an extensive introduction, the reference publication: Rik D.T. Janssen (2024). Ricgraph: A flexible and extensible graph to explore research in context from various systems. *SoftwareX*, 26(101736). https://doi.org/10.1016/j.softx.2024.101736.
- The documentation: https://docs.ricgraph.eu.
- For a live demo, contact the author of this document.

A schematic overview of the Open Ricgraph demo server can be seen in the following figure. At the top the systems to be harvested are shown. Ricgraph is the part enclosed in red colored dashes. It provides the graph, the user interface Ricgraph Explorer, and the REST API. All can be adapted by someone who can program in Python.



2.2 Virtual Machine *harvester*

Harvesting is done on a virtual machine (VM, a "workspace" in SURF Research Cloud terminology) in SURF Research Cloud. It runs the Ricgraph harvest scripts and creates the single graph. We use a separate VM for harvesting, since it may take a lot of time (hours or days, depending on the number of sources to be harvested). Also, it allows to use a "larger" VM (more CPUs, more memory, i.e. more expensive) then is necessary for exploring the graph (see next section).

The harvester VM contains the READ API keys of the Pure RIS of participating organizations. It also contains the intermediate harvested files.

Access to this harvester VM is regulated by SURF Research Cloud access management, which includes a one time password access mechanism. It is ISO 27001 certified. The harvester VM will only be accessible by the person(s) doing the harvest. It cannot be accessed by anyone else.

After the harvesting process has finished, the graph will be dumped to a file. This graph dump file will be transferred by secure copy to the explorer VM (see next section). Next, the harvester VM will be switched off ("paused" in SURF Research Cloud terminology), so it does not run and cannot be accessed by anyone.

2.3 Virtual Machine *explorer*

The Open Ricgraph demo server runs on another VM in SURF Research Cloud, called the explorer VM. The graph dump file from the harvester VM will be restored in Ricgraph on the explorer VM. This may take a few minutes, in which the Open Ricgraph demo server is not available. Otherwise, the explorer VM will be accessible by anyone in the world, unless some error occurs – since this is a demo server there is no guarantee of availability.

By using both a harvester and an explorer VM, the Open Ricgraph demo server on the explorer VM is accessible during the time consuming harvesting process on the harvester VM.

The explorer VM runs the nginx webserver (as used by SURF) to expose Ricgraph Explorer and the Ricgraph REST API to the world. Access to the explorer VM is regulated by SURF Research Cloud access management, just as with the harvester VM (see above).

2.4 Related work

There are a number of approaches that collect research related information from various sources and combine them into one data structure. For example, the EOSC research discovery graph, OpenAire graph, and OpenAlex. To learn more, please read how they compare to Ricgraph.

3 PILOT A: ENRICHING PURE DATA USING RICGRAPH AND BACKTOPURE

3.1 Rationale

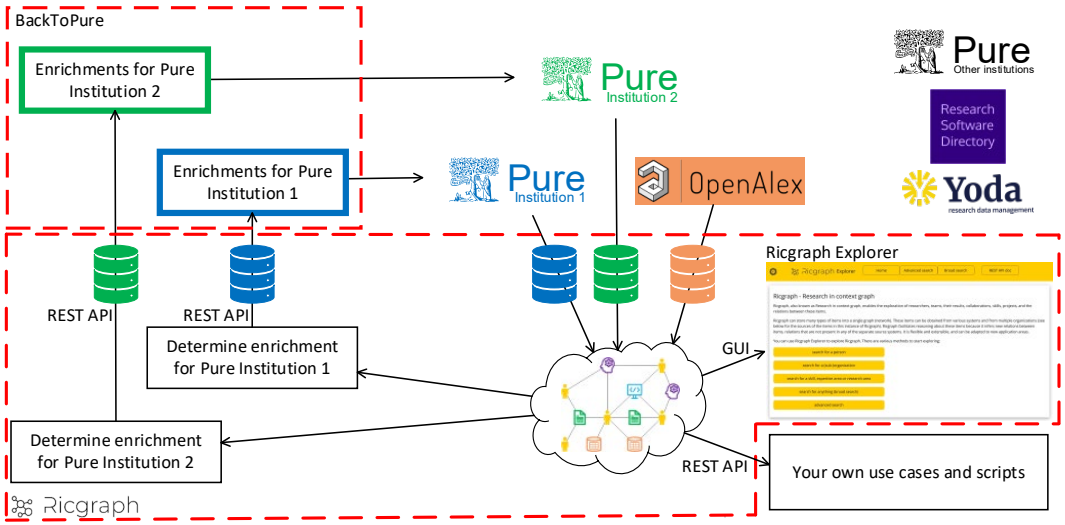
According to our experience, it does happen that the Pure Research Information System of a certain organization does not contain all research information items from that organization. Other source systems may contain items from that organization that are not in the Pure of that organization. If those items are added to the Pure of that organization, that organization will obtain a more complete view of its research.

In this pilot, research information from multiple organizations and multiple source systems will be combined in one single graph using the Open Ricgraph demo server. Using this graph, a participating organization can derive the items from that organization that are absent from the Pure of that organization but are present in another source system. We will explore how this works out and what can be learned from doing this.

3.2 BackToPure

BackToPure can be used to enrich the Pure of an organization using information from Ricgraph. It is open source software and available on GitHub (https://github.com/UtrechtUniversity/BackToPure). For a short introduction, read the presentation *Enriching Pure data using Ricgraph -Research in context graph- and BackToPure*, https://doi.org/10.5281/zenodo.12634658.

A schematic overview of the Open Ricgraph demo server and BackToPure can be seen in the following image. At the top right the systems to be harvested are shown. Ricgraph is the part enclosed in red colored dashes at the bottom part. It provides the graph, the user interface Ricgraph Explorer, and the REST API. BackToPure is shown in the top left, also enclosed in red colored dashes.



3.3 Technical overview

BackToPure has been developed to only run on a participating organizations' own infrastructure. That may be a laptop or a server. Since BackToPure needs a Pure CRUD (write) API key, this approach makes sure that this API key is only accessible by a trusted person from that organization that would like to enrich their own Pure.

BackToPure will access the Ricgraph REST API running on the explorer VM. This means that a participating organization only needs to install BackToPure, and does not need to install Ricgraph.

4 PILOT B: EXPLORING COLLABORATIONS BETWEEN SUB-ORGANIZATIONS USING RICGRAPH

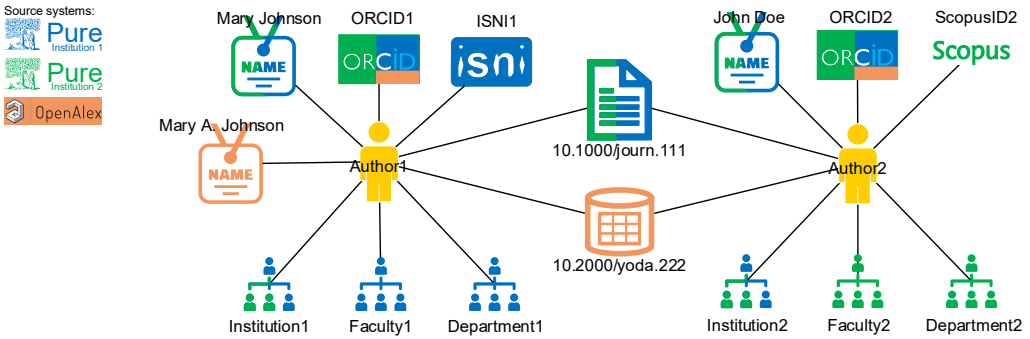
4.1 Rationale

Typically, collaborations between organizations are only examined at a top level (e.g. universities), without delving into the details of internal structures (e.g. faculties, departments, chairs). However, the Pure Research Information System offers a comprehensive view of an organization's entire organizational hierarchy. That is, not just the top-level structure, but also faculties, departments, individual chairs, and other organizational units. This detailed breakdown allows for a more granular analysis of inter-organizational collaborations.

In this pilot, research information from participating organizations and multiple source systems will be combined in one single graph using the Open Ricgraph demo server. Using this graph, we can derive collaborations between all kinds of sub-organization entities, e.g. which faculty collaborates with other faculties, which department collaborates with other departments and faculties, etc.

Since Ricgraph contains information about participating organizations, one can find these collaborations not only in one organization, but also between organizations. That means it is possible to explore collaborations between e.g. faculty A of organization 1 and faculty B of organization 2, or department C of organization 1 with other departments and faculties of other organizations. One can determine which persons or projects collaborate, or find the research results of these collaborations.

In this pilot, we will explore how this works out and what can be learned from doing this.



4.2 Research questions to explore

The figure above shows how (sub-)organizations are connected to person nodes in Ricgraph. Different colors indicate different source systems. From this figure, one can find relations that are not present in any source system:

- collaborations between sub-organizations, e.g., Dept1 and Dept2 work together;
- the data set is from Fac1, Dept1, Fac2 and Dept2;
- common research results: Fac 1 and Fac 2 have a publication and data set in common.

If Author1 has worked at multiple organizations, these organizations will be linked to the person-root of that author. This sounds reasonable from a person-viewpoint: this person has worked at different organization. However, it may not sound reasonable from an organization-viewpoint: some publications and data sets have been made during the time that Author1 was employed at that organization.

This pilot will explore research questions like:

- Should an organizational hierarchy be connected to a person, to a research result, or to both? Should the full organizational hierarchy be connected to a node, or only a select few? How is this done by publication aggregators like OpenAlex and OpenAIRE?
- Should nodes or edges be “time restricted”? E.g. Author1 worked at Institution1 or on a data set at a certain period in time?
- What if a person works at multiple (sub-)organizations at the same time?

- What are the differences in types of questions that can be answered using a certain approach? For example:
  - To which research results has a person contributed?
  - To which research results has a person contributed while working for a specific organization?
  - Given a specific organization, which research results can be attributed to that organization?
  - And by which persons were these results contributed?
- Does a “good” approach exist, or does it depend on the use case?

5 PREPARATIONS BEFORE PARTICIPATING IN THE PILOT PROJECT

This chapter lists the preparations that need to be done before an organization can participate in this pilot project.

Delft University of Technology was the first organization that agreed to participate in this pilot. Following extensive legal and privacy policy reviews by both parties, Delft University of Technology and Utrecht University reached an agreement to start the pilot. The resulting agreements and additional details can be read below. If your organization shares a similar approach to legal and privacy policies, an agreement to participate can be reached swiftly.

If your organization would like to participate, the following steps have to be followed:

- Your organization and Utrecht University have to agree on a *Agreement for hosting a pilot called "Open Ricgraph demo server"*. For more information see section 5.3.
- Your organization and Utrecht University have to agree on a *Data Processing Agreement*. For more information see section 5.4.
- Your organization needs to have a privacy statement. For more information see section 5.5.
- Your organization might need to decide on additional privacy measures. For more information see section 5.6.
- Your organization might want to do a Data classification (a CIA –Confidentiality, Integrity, and Availability– classification). For more information see section 5.7.
- Your organization needs to provide a Pure READ API key to Utrecht University. This key will only be disclosed to the person(s) harvesting source systems and will be kept private at all times.

For the pilot period, Utrecht University will not charge any fees to your organization (section 1.10), and your organization will not charge any fees to Utrecht University. The pilot period will end on March 31, 2026, and can be extended.

If you or your organization would like to participate in this pilot project, please contact Rik Janssen at r.d.t.janssen@uu.nl.

The remainder of this chapter will provide additional details on preparations to participate in the pilot.

5.1 Research information is exposed to the world

If your organization participates in the Open Ricgraph demo server, please consider the following.

- The Open Ricgraph demo server will make public (as in: open and accessible to anyone in the world): selected information in Pure from your organization. If your organization uses Pure Portal, this information is already public, since it is already available on that portal.
- The Open Ricgraph demo server will combine selected information from your Pure with information from other sources, such as the Pure systems of other participating organizations, and with OpenAlex, Yoda, and the Research Software Directory. Other source systems may be added at will.
- Your Pure may contain private or confidential information. This is a setting in Pure and can be set for a specific research information item or for a specific person. Your Pure will *not* expose private or confidential information or persons to the world. Therefore, these items will *not* be available in the Open Ricgraph demo server. The same holds for information that your organization marks private or confidential in the Research Software Directory or Yoda.
- The Ricgraph harvest scripts can be modified to exclude certain personal identifiers if a participating organization desires.
- Ricgraph has a script to remove personal information for a specific person if a person requests to do so (according to the GDPR). Such a person should contact your organization as specified in your organization's privacy statement.

The following two lists illustrate the types of information that are included in Ricgraph. More may be added in the future as this pilot progresses. Since this information is in Ricgraph, it is exposed to the world.

Personal information of current and former individuals (if available in a source system):

- Name.
- Personal identifiers such as (but not limited to): ORCID, ISNI, email address, employee ID, Pure ID, OpenAlex ID, Scopus Author ID, Researcher ID, Digital Author ID, etc.
- Social media IDs (e.g. Twitter/X, LinkedIn), GitHub ID, etc.

Non-personal information (if available in a source system):
- Object identifiers: Name of the object, title, category, DOI, PURE_UUID_RESOUT.
- Organization identifiers: Name of the organization, ROR, PURE_UUID_ORG.
- Project identifiers: Name of the project, duration, financer, PURE_UUID_PROJECT, etc.

5.2 Legal pilot organization model

In this pilot, we have chosen to adopt the roles "Controller" and "Processor" as defined by the GDPR[2]. A participating organization is the Controller, and Utrecht University is the Processor. Basically this means that the participating organization is responsible for the "correct" processing of personal data. Utrecht University will facilitate this process, on instruction of the participating organizations.

5.3 Agreement to host a pilot Open Ricgraph demo server

This agreement details the roles of a participating organization (controller) and Utrecht University (processor). Appendix A is the text of this agreement as agreed between Delft University of Technology and Utrecht University.

Participating organizations will have to agree to a similar text, except for e.g. their organization information, start date (art. 1.3-6), and court of jurisdiction (art. 7.2). Other changes might result in additional legal or privacy reviews.

5.4 Data processing agreement

A Data Processing Agreement (DPA) is an agreement between a data controller and a data processor. The data controller is the party which determines the purposes and means of the data processing.

For the Open Ricgraph demo server, Utrecht University will process data on behalf of a participating organization. The participating organization decides what types of personal data will be shared, for how long it will be stored, etc. (i.e. the purposes and means of processing). The participating organization is the data controller and Utrecht University is the data processor.

The DPA covers arrangements regarding privacy, such as the conditions for data processing, the security measures taken and the division of responsibility and liability.

Appendix B is the text of the DPA as agreed between Delft University of Technology and Utrecht University. A participating organization may use this DPA and modify e.g. organization information, court of jurisdiction (art. 1.c), and the email address to request removal of personal data (appendix C Security measures). An organization may also use its own model DPA. In the latter case, additional legal or privacy reviews might be necessary.

5.5 Privacy statement

A privacy statement is typically needed when an organization collects, uses, or handles personal information from individuals. Such a statement will at least contain a contact procedure for questions, complaints and deletion of personal information requests. It is the responsibility of the controller (i.e. no agreement about its contents is necessary with Utrecht University).

For the privacy statement, Delft University of Technology has chosen to use their standard privacy statement on https://www.tudelft.nl/en/privacy-statement.

[2] Another possibility would have been to use the Joint Controller relation, but that does not seem to be feasible at this time. For more information read https://www.autoriteitpersoonsgegevens.nl/en/themes/basic-gdpr/gdpr-basics/controller-and-processor.

5.6 Privacy measures

Privacy measures are measures taken to keep personal information safe from being accessed or misused by others. They help ensure that a person has control over its data and that organizations handle it responsibly.

Delft University of Technology has chosen to include the privacy measures in the appendices of the Data Processing Agreement DUT – UU (appendix B). Your organization needs to decide whether these are sufficient for your situation.

5.7 Data classification (CIA, in Dutch: BIV)

A data classification is done to determine how secure the IT solutions should be that are used for processing data (e.g., for storage, analysis, sharing), and which measures should be taken to ensure proper data security. In a data classification, someone determines how important it is to keep data Confidential, Correct (Integrity) and Available (CIA, in Dutch: BIV). Any of these three aspects is classified as low, basic, sensitive, or critical. The more impact a data breach would have, the higher the classification, and the more tight the security measures should be.

Although a CIA is the responsibility of the controller (i.e. no agreement about its contents is necessary with Utrecht University), Utrecht University has made a template CIA to help participating organizations to draft their own. The information security expert from UU has concluded that the Open Ricgraph demo server classifies “low” for all of the CIA components, but your own experts might have a different view. This CIA can be found in Appendix C.

APPENDICES

A AGREEMENT TO HOST A PILOT OPEN RICGRAPH DEMO SERVER DUT - UU

The text below is the *Agreement for hosting a pilot called "Open Ricgraph demo server"* as signed by both Delft University of Technology and Utrecht University, at June 10, 2025. You can get an editable version from the author.

Agreement
for hosting a pilot called "Open Ricgraph demo server"

THE UNDERSIGNED:

1. **Technische Universiteit Delft** , with registered office at Stevinweg 1, 2628 CN Delft, in particular the University Corporate Office "Library", Delft, the Netherlands, registered with the Chamber of Commerce under number 27364265, legally represented in this matter by [etc], hereinafter referred to as "**TU Delft**";

and

2. **Universiteit Utrecht**, having its registered office at Heidelberglaan 8, 3584 CS Utrecht, in particular the University Corporate Office "ITS Research & Data Management Services Department", Utrecht, the Netherlands, registered with the Chamber of Commerce under number 30275924, legally represented in this matter by [etc], hereinafter referred to as **"UU";**

and hereinafter referred to jointly as "**the Parties**" and separately as "**the Party**".

WHEREAS:

A) That the Parties wish to collaborate into a pilot called "Open Ricgraph demo server", an open software tool (see www.ricgraph.eu) owned by UU, hereinafter referred to as: "the **Pilot**";

B) The purpose of the Pilot is that TU Delft will participate in the pilot project of using Ricgraph (www.ricgraph.eu) to investigate how a knowledge graph:
- can provide insights into research relations and collaborations;
- can an optimize the quality of research information.

C) The Pilot project is described in the document "Discovering insights from cross-organizational research information and collaborations - A pilot project using Ricgraph", Rik D.T. Janssen.

D) In order to conduct the Pilot, Parties wish to establish the mutual relationship between them in this agreement, hereinafter referred to as "the Agreement".

E) This agreement is a non-exclusive agreement between TU Delft and UU. Other organizations can make a similar agreement as the TU Delft has with the UU.

HEREBY AGREE AS FOLLOWS:

1. SUBJECT OF THE AGREEMENT

1.1 The Parties undertake to cooperate with one another in accordance with the provisions of this Agreement and undertake to execute the Pilot in the manner described in the document Discovering insights from cross-organizational research information and collaborations (Annex 1).

1.2 The Parties shall assist one another in the execution of the Pilot according to their ability and shall, where necessary and desired, take into account each other's legitimate interests. Each Party shall inform the other Party of any obstacles that may jeopardise the progress of the Pilot and, in general, each Party shall provide the other Party with all the relevant information within the context of the Pilot.

1.3 Ricgraph has been created and is being maintained by UU. In order to conduct the Pilot, UU will undertake the following:

1. UU will host the Pilot, Open Ricgraph demo server, on SURF Research Cloud (ISO 27001 certified, https://www.surf.nl/en/information-security-surf-services).

2. UU will make this Pilot (the Open Ricgraph demo server and all of the data in it), open to the world, without authentication, via a web server and a REST API.

3. UU will harvest openly available research information from TU Delft, from Pure-TU Delft. Pure is the research information system of the TU Delft.

4. UU will combine this research information with other openly available data sources containing TU Delft research information: OpenAlex, Research Software Directory, Yoda.

5. UU will combine all of the above with similar other openly available data sources containing research information from other organizations.

6. UU will do this for the pilot period, starting from June 11, 2025 and ending at March 31, 2026 ("**Pilot Period**") but not in the months July and August 2025 (the holiday season).

7. UU will not charge any fees to TU Delft for the Pilot.

8. UU will make the Pilot available on a "best effort" base and there will be no guarantees of any kind.

1.4 TU Delft will undertake the following:

1. TU Delft will provide access to the UU to the Pure-TU Delft research information system by providing a READ API key.

2. TU Delft will not charge any fees to UU for the work in relation to the Pilot.

1.5 The source systems mentioned in this article contain personal information as well as non-personal information. The Parties agreed that the personal information in the mentioned source systems is already worldwide openly available personal information. Examples of this personal information are, but not limited to, name, ORCID, ISNI, email address, employee ID, Pure ID, OpenAlex ID, Scopus Author ID, Researcher ID, Digital Author ID, social media IDs (e.g. Twitter/X, LinkedIn), GitHub ID. Such personal information qualifies (partly) as personal data in the meaning of the General Data Protection Regulation (GDPR).

1.6 With regard to the personal data as mentioned above, Parties acknowledge that TU Delft qualifies as the data controller in the meaning of the GDPR in relation to such personal data and UU qualifies as the data processor in the meaning of the GDPR. To comply with the GDPR, parties have entered into a data processing agreement (Annex 2). TU Delft will have in place a contact procedure for questions, complaints and deletion of personal information requests. UU will assist the TU Delft in these matters in the manner described in the data processing agreement.

2. METHOD OF COOPERATION

2.1 Each Party shall designate a contact person for the execution of the Pilot and may replace its contact person, if necessary. The Party in question shall notify the other Party of this.

2.2 The Parties shall meet at least 6 times times during the Pilot Period to discuss the progress of the Pilot and to take decisions in consultation which are necessary for the proper execution of the Pilot.

3. CONFIDENTIALITY

3.1 All Parties shall treat with confidentiality all the knowledge and information made available to it by the other Party in the context this Agreement, insofar as such knowledge and information were not obtained through the execution of the Pilot and insofar as such knowledge and information have been expressly indicated by the other Party as 'confidential', hereinafter referred to as "Confidential Information", including but not limited to (information regarding) the READ Api key.

3.2 The Parties guarantee that its staff members shall also comply with the obligations referred to in this article.

4. LIABILITY & FINANCES

4.1 The Parties shall bear their own costs for carrying out this Pilot.

4.2 The Parties are not liable towards one another for any damage or loss caused by the application or use of the results or Confidential Information referred to in Article 5, unless there is question of intent or deliberate recklessness on the part of the Party that has provided the results or the Confidential Information.

4.3 Each Party indemnifies the other Party against all third-party claims on account of any damage or loss suffered by these third parties arising from the application or the use of results made available to the third parties by the first-mentioned Party.

5. DURATION AND TERMINATION

5.1 This Agreement shall enter into effect from the day on which it is signed by the Parties and shall remain in effect for the duration of the Pilot Period. Each party has the right to terminate the Pilot and this contract before the end of the Pilot Period for any reason. Those provisions which, by their nature, are intended to remain in effect even after the termination of this Agreement shall remain fully effective.

6. OTHER PROVISIONS

6.1 This Agreement may only be modified by mutual agreement and any modifications must be laid down in writing.

6.2 Neither Party is authorised to transfer, assign or pledge, either wholly or in part, its rights and obligations under this Agreement to a third party without the explicit written consent of the other Party.

6.3 The invalidity of any provision of the present Agreement shall not affect the binding nature of the remaining provisions. Furthermore, the Parties shall try to replace the void provision with a valid provision which expresses, as far as possible, the intentions of the Parties with respect to the void provision and/or its purpose.

7. LAW AND DISPUTES

7.1 This Agreement is subject to the law of the Netherlands.

7.2 Any disputes that might arise from the present Agreement or other agreements resulting therefrom, shall be submitted to the competent court in The Hague.

Drawn up in duplicate and signed in <.....> and Delft,

DELFT UNIVERSITY OF TECHNOLOGY	**Utrecht University**
Signature	Signature
Name	Name
Date	Date

B DATA PROCESSING AGREEMENT DUT - UU

The text below is the *Data Processing Agreement* as signed by both Delft University of Technology and Utrecht University, at June 10, 2025. You can get an editable version from the author.

DATA PROCESSING AGREEMENT

THE PARTIES:

1. **DELFT UNIVERSITY OF TECHNOLOGY**, with registered office at Stevinweg 1, 2628 CN Delft, in particular the University Corporate Office "Library", Delft, the Netherlands, registered with the Chamber of Commerce under number 27364265, legally represented in this matter by [etc], hereinafter referred to as "**TU Delft**"; hereinafter referred to as: "**The Controller**",

and

2. **Universiteit Utrecht**, having its registered office at Heidelberglaan 8, 3584 CS Utrecht, in particular the University Corporate Office "ITS Research & Data Management Services Department", Utrecht, the Netherlands, registered with the Chamber of Commerce under number 30275924, legally represented in this matter by [etc], hereinafter referred to as **"UU" ("The Processor")**

1 and 2 referred to individually as: "**the Party**" and jointly: "**The Parties**"

Whereas:

On **11 June 2025**, the Parties entered into a Contract concerning **Discovering insights from cross-organizational research information and collaborations: A pilot project using Ricgraph**. In the implementation of this Contract, personal data will be processed by the Processor on behalf of the Controller.

The Controller is committed to the protection of these personal data. For this reason, the Parties are using this Data Processing Agreement (Article 28, paragraph 3 of the General Data Protection Regulation, GDPR) and the associated appendices, i.e.:

- overview of processing of personal data and purposes of processing (Appendix A);
- overview for processing by sub-processors and transfer to third countries (Appendix B);
- overview of security measures (Appendix C);

in order to stipulate what the Processor is and is not permitted or obliged to do with the personal data;

HEREBY AGREE AS FOLLOWS:

1. General

a. Terms used in this Data Processing Agreement have the same meaning as in the General Data Protection Regulation (Regulation (EU) No. 2016/679).
b. In the event of any conflict between this Data Processing Agreement and the Contract, this Data Processing Agreement will take precedence. This means that the Contract and any other conditions agreed may in no way prejudice the rights and obligations pursuant to this Data Processing Agreement.
c. This Data Processing Agreement is subject to Dutch law. Disputes concerning this Data Processing Agreement will be put to the Court in The Hague, The Hague location.

2. Data processing – and general obligations

a. All personal data will be regarded as confidential data and treated as such. The Processor is permitted to use the personal data solely for the purposes of implementing the Contract and solely on the instructions of and on behalf of the Controller. An overview of all permitted processing has been included in Appendix A.
b. The Processor will refrain from using the personal data for its own purposes, for the advantage of or on behalf of third parties or for any other purposes, unless a statutory obligation under applicable law obliges it to do so, in which case the Processor will notify the Controller of that statutory obligation prior to processing, unless applicable law prohibits any such announcement for compelling reasons of general interest.
c. The Parties will comply with applicable privacy legislation and provide each other back and forth with all necessary cooperation and information in order to meet their statutory duties.
d. The Controller retains all intellectual and other property rights pertaining the personal data.
e. If, contrary to that stipulated in this Data Processing Agreement and/or the GDPR and/or other applicable legislation and regulations concerning the processing of personal data, the Processor determines the purposes and means for processing the personal data, the Processor will be deemed to be the Controller for said processing.

3. Confidentiality

The Processor will only reveal personal data (i) to staff members for whom knowledge of the personal data is strictly necessary for the purpose of implementing the Contract, (ii) through the Open Ricgraph demo server (Ricgraph) if and in so far it regards personal data that are being shared/collected for that purpose; the personal data involved has been made public by TU Delft through Pure and will be also be made available through Ricgraph during the Pilot Period (see Appendix A), as described in the Contract and only for the purpose of

implementing the Contract, except in cases where other statutory obligations apply to them. The Processor will also guarantee that authorised staff members are bound by a duty of confidentiality and abide by the provisions of this Data Processing Agreement.

4. Security

a. Pursuant to Articles 28 and 32 of the GDPR, the Processor will take appropriate technical and organisational measures to guarantee a level of security in accordance with the risk. The Processor will ensure that these measures take account of current technology, the cost of implementation, the nature, scope, context and purposes of processing and the risks of varying likelihood and severity for the rights and freedoms of data subject(s). Consideration will also be given to the risks that may result from the accidental or unlawful destruction, loss, alteration or unauthorised disclosure of or access to data transmitted, stored or otherwise processed.

b. The Processor will lay down its security policy in writing. At the Controller's request, the Processor will provide access to its security policy. Appendix C describes the minimum security measures to be taken by the Processor. Since security risks are continually changing, the Processor will regularly update and improve the security measures taken.

5. Data Protection Impact Assessment

In order to enable the Controller to meet any obligation it may have to conduct a Data Protection Impact Assessment (DPIA), the Processor will notify the Controller at its first request and prior to any processing of personal data, of the following:

i. a systematic description of the processing envisaged;
ii. an assessment of the risks to the rights and freedoms of data subjects in view of the nature, scope, context and purposes of the processing;
iii. the measures intended to address the risks stated under (ii), including safeguards, safety measures and mechanisms to ensure the protection of the personal data and demonstrate compliance with the GDPR, taking account of the rights and legitimate interests of the data subject(s) and other persons concerned.

6. Sub-processor

a. The Processor refrains from outsourcing the processing of personal data under this Data Processing Agreement to sub-processors without prior written permission from the Controller. The Controller grants written permission for the use of sub-processors only in the event that this has been explicitly included in Appendix B. In the event of any intended change (addition or replacement) to one or more sub-processors during this Agreement, the Processor provides the Controller with written advance notice of this change and the Controller has the option of objecting to this change.

b. In the event that the Processor outsources its obligations under this Data Processing Agreement with written permission from the Controller, the Processor must enter into a sub-processing agreement that imposes the same conditions and obligations on the sub-processor as those imposed on the Processor in this Data Processing Agreement, and especially the obligation to provide adequate guarantees concerning the application of appropriate technical and organisational measures. If the sub-processor fails to meet its obligations towards the Processor, the Processor will be fully responsible towards the Controller for the sub-processor's compliance with its obligations under any such sub-processing agreement. The Processor will provide the Controller with a copy of the sub-processing agreement at its first request, from which commercially sensitive information may be omitted.

c. Dutch law will apply to the provisions concerning the outsourcing of this Data Processing Agreement.

d. The Processor maintains a list of the sub-processing agreements entered into as part of this Data Processing agreement and notifies the Controller of this. This list will be updated at least once annually. This list will be kept available for the purposes of the supervisory authority.

e. Only with prior written permission from the Controller the Processor is allowed to process or arrange the processing of personal data in countries outside the EEA or provide personal data to organisations outside the EEA. The Controller makes this permission conditional on the obligation to include Standard Contractual Clauses (SCCs) if no adequacy decision is applicable (Appendix D).

f. The Processor will remain fully liable vis-à-vis the Controller for the sub-processor's compliance with its obligations if the latter fails to observe its obligations arising from Applicable Law.

7. Requests from data subjects, government and supervisory authorities

a. <u>Requests from data subjects</u>
At its own cost, the Processor will cooperate fully in enabling the Controller to comply with the requests of data subjects, for example by providing data subjects with access to the personal data of relevance to them, removing, supplementing, transferring, protecting and rectifying personal data and providing evidence that the request has been met. When a data subject submits a request to the Processor, the Processor will refer the data subject to the Controller and will not engage in any discussion of the contents of the request.

b. <u>Requests from government and supervisory authorities</u>
At no charge and in close consultation with the Controller, the Processor will cooperate fully in any investigations conducted or requests made by government and supervisory authorities concerning the Controller and will provide all information of relevance to this. If the Processor receives a request of this kind addressed to it, it will immediately notify the Controller and the Parties will consult on the steps to be taken, unless that is prohibited in view of the nature of the request. In the latter case, the Processor will represent the reasonable interests of the Controller.

c. In order to safeguard the protection of personal data, the Processor will in that case ensure that it does not provide the government or supervisory authority with more personal data than is strictly necessary in order to meet the public body's request. If it is possible to take action in law against a request to provide personal

data or a prohibition on informing third parties exists, the Processor will take full advantage of this.

8. Audit

a. With the exception of low-risk processing, the Processor is obliged to have an independent external expert conduct an audit at least once every two years on the Processor's organisation in order to demonstrate that the Processor is compliant with the Contract, the Data Processing Agreement, the GDPR and other applicable legislation and regulations concerning the processing of personal data.
b. The Controller is entitled to have an audit of the Processor's organisation conducted by an independent external expert in order to demonstrate that the Processor is compliant with the Data Processing Agreement, the GDPR and other applicable legislation and regulations concerning the processing of personal data. The Controller can make use of its right to have an audit of the Processor conducted at its request a maximum of once annually or more frequently in the event of a specific suspicion that the Processor is not complying with the Data Processing Agreement and/or the GDPR and/or other applicable legislation and regulations concerning the processing of personal data.
c. The Controller provides the Processor with notice of the audit at least 14 (fourteen) days in advance of it. The audit may not cause unreasonable disruption to the Processor's normal business activities.
d. The cost of the audit conducted at the request of the Controller shall be paid by the Controller unless the audit findings reveal that the Processor has not complied with the Data Processing Agreement, and/or the GDPR and/or other applicable legislation and regulations concerning the processing of personal data.
e. If it is determined during an audit that the Processor is not complying with the Data Processing Agreement, and/or the GDPR and/or other applicable legislation and regulations concerning the processing of personal data, the Processor immediately takes all measures reasonably necessary in order to ensure that the Processor is compliant. The associated costs shall be paid by the Processor.

9. Reporting of data breaches

a. The Processor will have procedures in place aimed at ensuring the reasonable detection of security incidents and data breaches and taking action in response, including remedial measures. The Processor will provide the Controller with a copy of the relevant procedures at its first request.
b. In order to enable the Controller to fulfil its notification obligations, the Processor notifies the Controller of any breach of security within 24 hours at the latest. Reports can be submitted to the Controller via databreach@tudelft.nl, or, if relevant, to another contact designated by the Controller during the term of this Data Processing Agreement and must, in any event, include:
    i. the nature of the breach and, where possible, the categories of data subjects and personal data records concerned and the approximate number of data subjects and personal data records concerned;
    ii. the name and contact details of the Processor's data protection officer or another point of contact where further information can be obtained concerning the breach;
    iii. the likely consequences of the personal data breach;
    iv. the measures for addressing the personal data breach, including, where appropriate, the measures to mitigate its possible adverse effects.
c. With regard to every breach as referred to under 9a, the Processor will ensure that it provides the Controller with all cooperation that might reasonably be expected from the Processor, including the provision of sufficient information and support relating to investigations by the supervisory authority:
    i. in order to rectify and investigate the breach and prevent future breaches;
    ii. in order to limit the impact of the breach on the privacy of data subjects; and/or
    iii. in order to limit the damage incurred by the Controller as a result of the breach.
d. The Processor documents any personal data breaches, including the facts concerning the personal data breach, the consequences of it and any remedial measures taken. The Processor provides this documentation to the Controller as soon as it is requested.
e. Unless legally required to do so, the Processor will not notify the supervisory authority and/or data subjects of a security breach without prior written permission from the Controller.

10. Retention periods

a. The Processor will not retain the personal data for longer than is strictly necessary and definitely no longer than the term of the Data Processing Agreement, unless storage of the personal data is a legal obligation, in which case the Processor will not retain the personal data for any longer than the period prescribed by law.
b. Where necessary, the Processor follows the retention instructions given by the Controller.

11. Liability and indemnity

a. The Processor is liable for any damages arising from or connected with failure of the Processor to comply with the Data Processing Agreement, and/or the GDPR and/or other applicable legislation and regulations concerning the processing of personal data.
b. The Processor indemnifies the Controller with regard to any third-party claims, fines and or measures, including from data subjects and the supervisory authority, made or imposed vis-à-vis the Controller because of a breach of the Data Processing Agreement, and/or the GDPR and/or other applicable legislation and regulations concerning the processing of personal data by the Processor and/or by persons/legal persons deployed by the Processor, including but not limited to staff members and/or sub-processors, except for situations where it is evident that Controller and Processor are both responsible for such breach in which case the

Parties will carry equal parts of such claims, fines or other measures resulting from that breach.

c. The Processor arranges adequate coverage for its liability by means of liability insurance. At the Controller's request, the Processor allows the Controller to consult (the policy of) this liability insurance held by the Processor.

12. Change

a. In the event of an intended change to the processing of personal data, such as the deployment of a new sub-processor, a change in the transfer of personal data to third countries and/or international organisations or changes to the security measures taken, the Processor is obliged to notify the Controller immediately about the intended changes and the Parties will consult as soon as possible on the consequences for this Data Processing Agreement.

b. The Processor will not be entitled to implement a change until the Controller has issued prior written permission for this change. A change may never result in the Controller no longer being able to comply with the applicable privacy legislation. Any changes will be recorded in writing in Appendix B.

c. In the event of any change to the current policy rules of the supervisory authority, the Parties will make use of Appendix C to implement the changes required in order to comply with the new policy rules.

13. Duration and termination

a. The duration of the Data Processing Agreement is identical to that of the Contract. The Data Processing Agreement cannot be terminated prematurely or separately from the Contract.

b. The Parties have agreed that, within one month of the termination of the Data Processing Agreement, the Processor destroys all personal data and copies it has processed from its internal servers as well as from Ricgraph, some of which may be held by persons/legal persons deployed by the Processor, including but not limited to staff members and/or sub-processors, and provides written confirmation of this to the Controller, unless the law prohibits such destruction. In the latter event, the Processor guarantees that it will observe confidentiality with regard to the personal data processed and will no longer actively process the personal data.

c. The Processor bears the costs of destruction of the personal data.

d. When this Data Processing Agreement terminates, the provisions that are intended to continue to apply after it will remain in force, such as Article 2 (ownership), Article 3 (confidentiality) and 11 (liability).

AGREED AND SIGNED:

Delft University of Technology	Utrecht University
Name:	Name:
Job title:	Job title:
Date:	Date:

---

*Appendix A **Personal data***

In this pilot project, Utrecht University will construct a graph containing research information of the University of Technology Delft and combine it with research information from other universities. It will include information (i.e. meta data) on researchers, teams, their results, collaborations, (optional) skills, (optional) projects, and the relations between these items. This graph is constructed based on the contents of the Pure Research Information System of each participating organization, supplemented with research information from public sources such as OpenAlex (https://openalex.org/), Yoda (https://www.uu.nl/en/research/yoda), and the Research Software Directory (https://research-software-directory.org/).

The result is openly available to anyone, in accordance with the open science principles.

For more specific information, see document "Agreement to host a pilot open Ricgraph demo server".

Processing Description of the subject and duration of the processing	**Purposes** Description of the nature and objective of the processing	**Categories** Description of the categories of data subjects	**Personal data** Description of the type of personal data being processed
See description above, and the document "Agreement for hosting a pilot called "Open Ricgraph demo server".	Read data Store data Combine data Publicly share data	Anyone present in the research information systems mentioned above. *All categories that will be processed are already public.*	*All personal data that will be processed are already made public. The categories of personal data are:* *First name(s) and last name and other personal identifiers like, ORCID, ISNI, contact details i.e. email address, employee ID, Pure ID, OpenAlex ID, Scopus Author ID, Researcher ID, Digital Author ID, social media IDs (e.g. Twitter/X, LinkedIn), GitHub ID, university and (previous) affiliations.*

Appendix B ***Permission for processing personal data by sub-processors and transfer to third countries***

- **Sub-processors**

The Controller hereby grants permission to the Processor for the deployment of the sub-processors referred to below [*to be completed by the Controller and Processor*]:

Sub-processor deployed by the Processor to process personal data	**(Category of) personal data to be processed by sub-processor**	**Type of processing**	**Country of processing**	**Country in which sub-processor is based**
SURF (surf.nl)	See appendix A	Read data Store data Combine data Publicly share data	The Netherlands	The Netherlands

- **Transfer**

The Controller does not grant the Processor permission for transfer to third countries/international organisations [*to be completed by the Controller*]:

Description of transfer	**Entity transferring the personal data + country**	**Entity receiving the personal data + country**	**Transfer mechanism (e.g. SCC)**
Not applicable	X	X	X

Appendix C ***Security measures***

- **List of security measures**

Policy and organisation The TU Delft privacy policy (and all policies connected to the Privacy Policy) applies to all TU Delft processing activities. The TU Delft Library is responsible for effectuating these policies and measures discussed in this DPA and the Samenwerkingsovereenkomst. The REST API key can be revoked by TU Delft if necessary. Subjects can have their data withdrawn from Ricgraph by sending an e-mail to privacy-tud@tudelft.nl
Access security The data on the harvester VM will only be accessible to selected employees of Utrecht University. The data on the explorer VM will be accessible to anyone in the world. Also see document: "Discovering insights from cross-organizational research information and collaborations - A pilot project using Ricgraph".
Management of vulnerabilities and anti-malware According to SURF Research Cloud policies, ISO 27001 certified.
Confidentiality and data integrity Data Confidentiality is not applicable as the harvested data will be made available publicly. Data Integrity is not applicable as the harvested data is obtained from source systems which are authoritative for the harvested data. Also see document: "Discovering insights from cross-organizational research information and collaborations - A pilot project using Ricgraph".
Incident response, reporting and remediation In case of a personal data breach the TU Delft Data breach procedure applies in handling the data breach.
Patch management N/a since this is a pilot.
Auditing and logging N/a since this is a pilot. However, the TU Delft has the right to exercise its audit rights as included in this Data processing agreement.
Software development Utrecht University will do development of Ricgraph. Other parties are also able to do that since it is open source software.

The Processor holds the following certificates (if applicable), for which the Processor itself is responsible for any renewals, ensuring that the Processor holds a valid certificate throughout the term of this agreement:

Name of certificate	**Organisational unit/service to which the certificate relates**	**Certificate's period of validity**	**Declaration of applicability**
SSL certificate for Ricgraph Explorer website and REST API	*Let's encrypt*	*As with Let's encrypt certificates*	

- **Specific instructions concerning data retention**

Duration of processing	**Duration of storage/back-up**	**The chosen method of data destruction**
n/a	n/a	n/a

C TEMPLATE DATA CLASSIFICATION (CIA)

The text below is a template for the Data Classification. You can get an editable version from the author.

Introduction
The CIA (Confidentiality, Integrity, and Availability) classification is a concept in information security that helps organizations protect their data. Its purpose is to establish a framework for safeguarding information assets. *Confidentiality* ensures that data is kept private and only accessible to authorized individuals or systems. *Integrity* focuses on maintaining the accuracy and consistency of data throughout its lifecycle, preventing unauthorized alterations. *Availability* ensures that information is accessible to authorized users when needed. By implementing a CIA classification, organizations can assess the importance of different types of data and apply appropriate security measures.

A CIA classification should be created by the organization that provides the data, which is typically referred to as the data controller. The information in this appendix is intended as a template that can be used by such an organization, to help it to create a CIA that suits its purpose. The main part is written in Dutch since that was the language used to describe an earlier version of this CIA. At some point in time It may be translated to English.

Open Ricgraph demo server pilot project
In this pilot project *Open Ricgraph demo server*, Utrecht University will construct a graph containing research information of the organizations that participate in this pilot. This graph will also include information of organizations and persons that do not participate in this pilot in case they collaborate with organizations or persons that do participate in this pilot.

The graph will include information (i.e. meta data) on researchers, teams, their results, (sub-)organizations, collaborations, (optional) skills, (optional) projects, and the relations between these items. This graph is constructed based on the contents of the Pure Research Information System of each participating organization, supplemented with research information from public sources such as OpenAlex, Yoda, and the Research Software Directory. The result is openly available to anyone, in accordance with the FAIR and open science principles.

This pilot project is described in the document *Discovering insights from cross-organizational research information and collaborations: A pilot project using Ricgraph*, Rik D.T. Janssen, https://doi.org/10.5281/zenodo.15637647.

CIA table

Algemeen	
Omschrijving	Zie paragraaf *Open Ricgraph demo server pilot project*.
Beschikbaarheid	**0. Laag**
Integriteit	**0. Laag**
Vertrouwelijkheid	**0. Laag**
Faculteit/Dienst	ITS
Diensteigenaar	Geen, pilot
Status Geschiktheidsniveau	Concept
Volwassenheid van dienst	Pilot/demo server
Gemaakt	24 december 2024
Gegevens	Uitsluitend metadata. Zie paragraaf 2.4.3 van het document *Discovering…*
Doelgroep(en)	Iedereen en in het bijzonder de deelnemers aan de pilots.
Betrokken externe partijen	SURF: voor de hosting van de demo server op de SURF Research Cloud Aan de pilots deelnemende universiteiten: leveren data aan.
Functionaliteit	Zie paragraaf *Open Ricgraph demo server pilot project*.
Proceseigena(a)r(en)	geen, pilot

Bij **Beschikbaarheid** kijken we naar wat de impact maximaal zou mogen zijn van het niet beschikbaar zijn van het systeem waardoor deze niet te gebruiken is en de data niet benaderbaar. Ook kijken we naar wat er gebeurt als de data kwijt zou raken.	
Neem de belangrijkste data/ het belangrijkste individuele proces van dit systeem. Wat is het gevolg als dit weg is of tijdelijk niet mogelijk is?	0. Laag Het hoofddoel van de Open Ricgraph demo server is het verzamelen van input over de inzet en potentiële waarde van de dienst: Zijn de resultaten interessant genoeg om als dienst aan te bieden? Wat betekent dat voor zo'n dienst? in welke hoedanigheid gaan we de dienst aanbieden? Zijn alle beoogde doelen interessant? Mogelijke toepassingsgebieden zijn: rapportage, data clean-up, en zoekmachine voor data/onderzoekers/expertise. Als de Open Ricgraph demoserver niet beschikbaar zou zijn, kan deze niet gebruikt worden. Dat is mogelijk vervelend, maar heeft geen business impact.

Wat zou de impact zijn op verschillende processen als het systeem er langer uit ligt dan in de SLA is afgesproken? (dit kan hoger zijn dan het hoogste individuele proces omdat het hier over het hele systeem gaat).	0. Laag Er is geen SLA, het gaat om een pilot. Zie ook de toelichting hierboven. Als alle data weg is kan de harvest opnieuw gedraaid worden, dit heeft geen business impact.
Wat als alle data uit het systeem kwijt is (inclusief de backups)? Denk aan voorbeelden zoals bij gijzelsoftware (ransomware).	0. Laag Idem.
Is er kans op letsel/overlijden als het systeem niet beschikbaar is?	0. Laag Nee.
Wat is de grootste impact van een uitval op een kritiek moment / de data weg is van een bepaald kritiek moment?	0. Laag Er zijn tijdens deze pilotfase geen kritieke momenten.

Bij **Integriteit** kijken we naar de gevolgen van onjuistheden in de data. De oorzaak van deze onjuistheden maakt hierbij niet uit, het kan een invoerfout zijn, maar bijvoorbeeld ook het gevolg van bijvoorbeeld een moedwillige actie van buitenaf.	
Neem de belangrijkste data uit het systeem, wat als deze data gecompromitteerd blijkt te zijn (en niet terug te zetten naar wat het hoort te zijn)?	0. Laag De onderzoeksinformatie in de Open Ricgraph demoserver wordt geharvest uit andere openbare systemen. De demoserver voegt geen nieuwe informatie toe. Als deze onderzoeksinformatie gecompromitteerd blijkt te zijn verwijderen we alles uit Ricgraph. Daarna wordt een nieuwe harvest gedaan. Daarna is de data in Ricgraph weer up-to-date.
Wat zou de impact zijn als ALLE data van het systeem niet betrouwbaar blijken te zijn? (Dit wil niet zeggen dat alles fouten bevat, maar we kunnen er niet voor instaan dat alles nog klopt).	0. Laag Dit heeft geen impact. De Open Ricgraph demoserver wordt niet ingezet voor bedrijfskritische toepassingen. Als alle data onbetrouwbaar blijkt wordt database van Ricgraph weggegooid en wordt de harvest opnieuw uitgevoerd (zie hierboven).

Wat zou de ergste impact zijn op andere verwerkingen als de technologie (het systeem) gecompromitteerd raakt? Denk hierbij ook aan installaties van software op (privé) apparaten.	0. Laag De Open Ricgraph demoserver verzamelt data uit de bronsystemen, maar koppelt nooit wijzigingen terug naar die bronsystemen. Als de data in de demoserver gecompromitteerd worden, heeft dit verder geen gevolgen voor de bronsystemen waar deze data uit verzameld zijn.
Wat is het voordeel wat een kwaadwillende voor zichzelf of omgeving zou kunnen bereiken als deze toegang tot het systeem weet te krijgen en vervolgens acties kan uitvoeren of data aan kan passen? Denk bijvoorbeeld aan financieel gewin of schade aanrichten.	0. Laag Een kwaadwillende zou de gegevens aan kunnen passen en daarmee de integriteit kunnen compromitteren. In zo'n geval zal dit gerepareerd worden door het uitvoeren van een nieuwe harvest van de brondata (zie hierboven).
Is er kans op letsel / overlijden van mensen als er iets mis is met het functioneren of de data in het systeem?	0. Laag Nee.
Wat zou de impact zijn op ketenafhankelijkheden (al dan niet automatisch) als er iets mis blijkt te zijn met de data?	0. Laag Als een organisatie een tool als BackToPure gebruikt, kunnen onjuiste gegevens van die organisatie terug in hun eigen Pure komen. BackToPure kan aangesloten worden op de Open Ricgraph demoserver en dit is een van de doelen van de pilot (Zie hoofdstuk 3 van het document *Discovering*...). We gaan ervan uit dat elke organisatie die deelneemt aan de pilot in hun Pure een workflow gebruikt die nieuwe gegevens controleert. In dat geval is er geen impact op ketenafhankelijkheden.

Bij **Vertrouwelijkheid** kijken we wat de gevolgen zijn wanneer de data ingezien wordt door mensen die daarvoor niet geautoriseerd zijn. Dat geldt voor zowel binnen als buiten de organisatie. Wat gebeurt er bijvoorbeeld als de data die in deze dienst verwerkt worden op straat komt te liggen en door iedereen inzichtelijk is?	
Wat is de meest vertrouwelijke data die mogelijk in/door het systeem komt, en wat als deze data uitlekt of inzichtelijk is voor onbevoegden?	0. Laag Geen, alle gegevens in de Open Ricgraph demoserver zijn al voor iedereen toegankelijk in de bronsystemen die de demoserver harvest.
Wat zou de impact zijn als alle data van alle verschillende verwerkingen in dit systeem gelijktijdig uit zou lekken?	0. Laag Geen. De gegevens zijn die Ricgraph verzamelt zijn al publiek beschikbaar in verschillende bronnen.
Hoe ernstig zou de impact kunnen zijn van vervolgaanvallen die gebruik maken van ingeziene data in de dienst? Vervolgaanvallen kunnen bijvoorbeeld identiteitsfraude of chantage zijn, maar ook kan inzicht in bijvoorbeeld waar spullen opgeslagen worden of camera's hangen gebruikt worden voor diefstal.	0. Laag Geen, alle data die door het systeem komt is al voor iedereen toegankelijk.