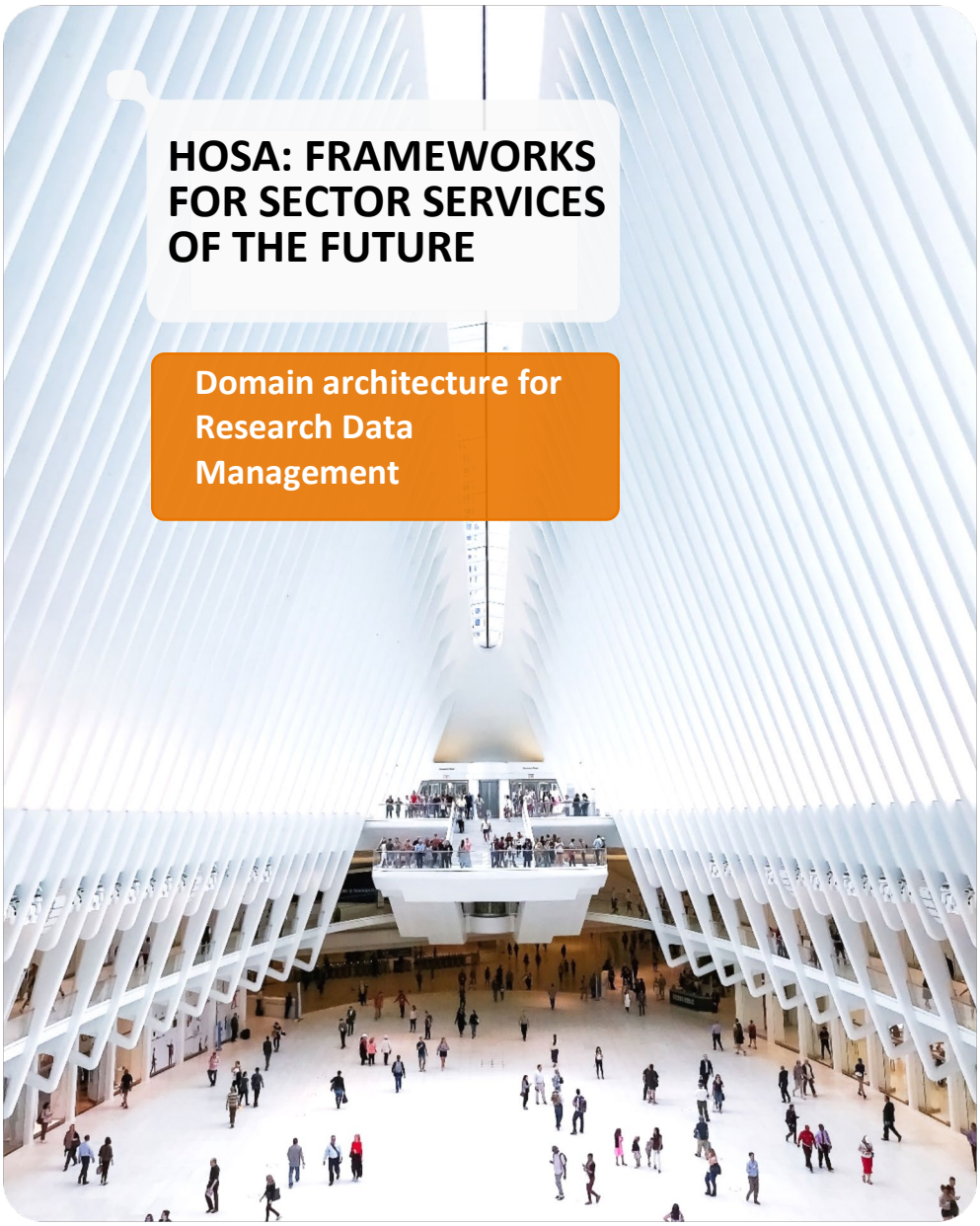
HOSA

SURF

Driving innovation together



HOSA: FRAMEWORKS FOR SECTOR SERVICES OF THE FUTURE

Domain architecture for Research Data Management

Authors: Tom van Veen and Menno Scheers

Domain Architect RDM Lead Architect HOSA

1 Foreword

This domain architecture outlines what the digital support for Research Data Management should look like in three to ten years' time to fit in with the supported ambitions of the higher education sector, the changes in research and the international (technological) developments. This domain architecture is similar to a zoning plan and includes a vision of the future of Research Data Management and its relationship to the associated systems. It also forms the basis for a roadmap that makes it possible to realise and prioritise this zoning plan.

However, a zoning plan is not enough to construct a building. This also requires follow-up architectures and designs. This document provides the framework for the assessment of such follow-up architectures and designs. First, we give an initial description of the current situation. This description is based on various discussions and sessions with the working group. It is necessary to shape the domain architecture. Further analysis of the current situation is required for the actual realization of a sector service.

We use the concept of business platforms to draw up the application architecture in this domain architecture. This concept fits in with the coming decade, which is also referred to as the digital age. A business platform is based on a business that is built on value-creating interactions between external producers and consumers. The platform's overall objective is to achieve matches between users and facilitate the exchange of products, services, content, data or social currency, thereby creating value for all participants (G. Parker, 2016).

We use this concept because it provides insight into online marketplaces that are emerging in various places thanks to increasing collaboration within the sector. We mainly focus on the term 'business' in that regard. By using the same template as a foundation, we offer an overview and create a common starting point for the discussions between the many stakeholders. This template also acts as a base for presenting various initiatives in a consistent way.

2 Table of contents

3 Introduction

3.1 Motivation

The higher education institutions have found that cross-institutional initiatives are expanding and will have a greater impact on their own services. UNL is investigating the topics universities want to invest in together, such as research facilities, ICT for educational renewal and sustainable operations management. The higher education institutions are also collaborating with each other, for example on the subject of research support. The CIOs of institutions, SURF and a number of sector partners have therefore taken the initiative to create an architecture for the digital sector services of the future.

The developments in research data and data management are leading to more cross-institutional collaboration and organisation for common information and ICT services (sector services). This raises questions such as: “How do we ensure that the sector services are future-proof?”, “How do we create consistency in the sector services?”, “How do we enable reuse?” and, more recently, “How do we assure, protect and promote public values in the digitisation of education and research?”.

Sector partners such as SURF, Studielink, DUO, DANS and NWO try to facilitate and support all institutions in this as much as possible. However, this is a complex process that creates a need for a joint architecture: for the sector-wide definition, development and deployment of information and ICT services, it is necessary to provide clarity about the demand for these services, the overall requirements in this connection, the design and setup of these services, and the services offered to the institutions by the ICT service providers. We want to create this common framework based on an architectural approach: the Higher Education Sector Architecture (Hoger Onderwijs Sector Architectuur or HOSA).

3.2 Goals

The HOSA project aims to define an architecture for sector services that are important for strategic collaboration between higher education institutions, sector partners and market players. HOSA is therefore based on the optimal articulation of the sector’s demand with regard to sector services. HOSA also provides a facilitating framework for interoperability between institutions and providers of common ICT services. HOSA will lead to current and new sector service initiatives more quickly and in a more future-oriented and future-proof way. Sector partners and market players in ICT services can easily respond to this with their service portfolio.

3.3 Scope of the domain architecture

In principle, the scope of the domain architecture is the higher education sector (higher academic education and higher professional education). Of course research is not limited to the higher education sector, so the international context of Research Data Management has also been considered. The time horizon is the medium and long term looking forward, with an assessment of the functional needs for sector services.

Research Data Management is a specific process regarding the creation, management and processing of research resources to enable their accessibility and use for as long as they retain value. Research data refers to both the primary research data (raw, rough measurements or observations) and the secondary research data (data created during processing of primary research data, such as recoding, combining, categorising, visualising, etc.). In this we take a broad perspective and also pay attention to services, service provision, processes, functionality, data and technology, governance, ownership, management and support.

This domain architecture is a conceptual description of the setup of sector services for research data in the broad sense of the word. It describes guidelines for positioning and functionalities of the new sector services for research data. When developing this domain architecture, we took the research cycle and research data cycle as described in HORA 2.1 as a starting point (Bedrijfsprocessen Onderzoeksgegevenscyclus, 2019).

The domain architecture is similar to a zoning plan.

3.4 Objectives structure

An objectives structure has been drawn up for HOSA to provide insight into the sector's goals and ambitions. The foundation for this objectives structure based on the strategic agenda of the Ministry of Education, Culture and Science.[1] This has been supplemented with goals from other policy documents (see Figure 1). The domain architecture for Research Data Management supports the realisation of the HOSA goals related to research data and the sector ambitions (highlighted in blue).

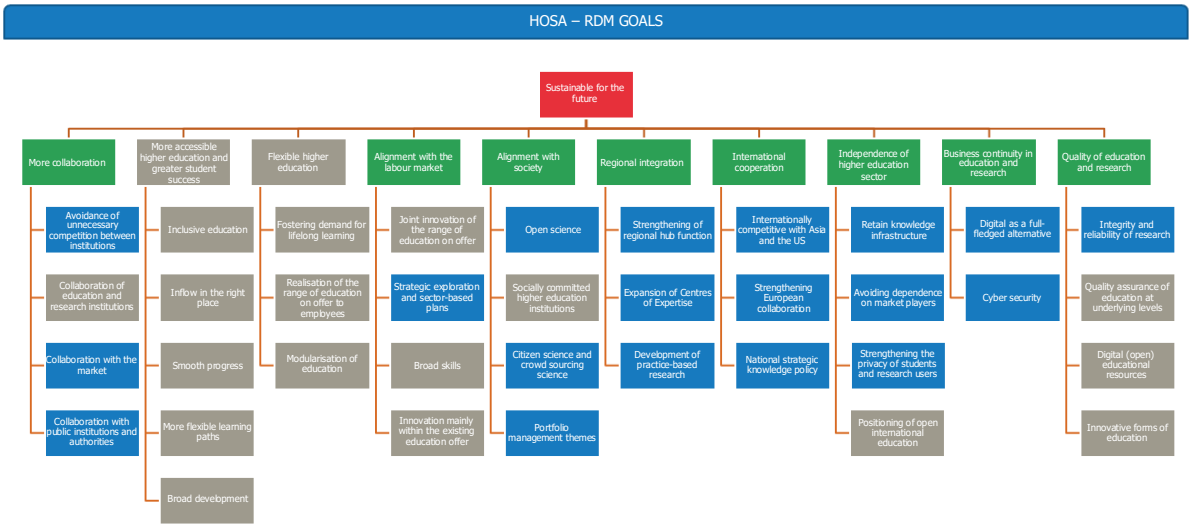


Figure 1: HOSA-RDM objectives structure (see Appendix 12.2 for a readable version)

[1] *Strategische agenda hoger onderwijs en onderzoek – Houdbaar voor de toekomst*, Ministerie van OCW, Den Haag, December 2019

4 Current situation

To describe the current situation in the field of Research Data Management (RDM), this domain architecture starts from the existing facilities and change initiatives that have been or are being undertaken from various organizations and bodies. In addition, we have inventoried various bottlenecks around RDM by means of interviews, reports and sessions.

Section 4.1 briefly describes a number of RDM initiatives. This overview is not exhaustive. It only provides insight into some of the many initiatives. Section 4.2 describes the main problem areas that have emerged.

4.1 Initiatives

This section briefly describes a number of initiatives related to the subject of Research Data Management who currently already exist or are under development. The inventory has shown that the landscape of RDM initiatives is very wide. These initiatives operate at different levels – from international to European to regional – and each addresses a need that is present at that level or for a specific theme/topic. Such initiatives are often self-contained and offer little or no mutual consistency. They often focus more on finding, obtaining and making available research data than on processing, creating, analysing or visualising research data, for example. The inventory is therefore far from complete, simply because there are so many initiatives and they are very diverse. At most, this inventory provides insight into the diversity and fragmentation.

DANS

Data Archiving and Networked Services (DANS) is the Dutch institute for permanent access to digital research data. It encourages researchers to make their digital research data findable, accessible, interoperable and reusable. The DANS services include: DataverseNL for short-term data management, EASY for long-term archiving, and NARCIS, the national portal for scientific information. DANS participates in many projects that focus on the RDM domain, including the EOSC hub, the FAIR data self-assessment tool and CoreTrustSeal. DANS is an institute of KNAW and NWO.

LCRDM

The National Coordination Point for Research Data Management (Landelijk Coördinatiepunt Research Data Management or LCRDM) is a national network of RDM experts. It facilitates the connection between policy and solution in close consultation with education and research institutions. The LCRDM experts work together to put RDM topics on the agenda that are too big for a single institution to tackle and that require a joint national approach.

NPOS

The National Open Science Programme (NPOS) brings together the stakeholders in the National Open Science Plan. The programme ensures that the Netherlands progresses towards achieving its Open Science ambitions and monitors developments in this respect. The programme's focus is on accelerating the National Open Science Plan's three key areas. The stakeholders of the national plan attend programme meetings regularly. The NPOS has a steering committee consisting of the managing directors of the stakeholders, an advisory board and various projects.

NPPO

Universities of applied sciences conduct a lot of practice-based research, which contributes to innovations in education, the business world and social organisations. Practice-based research offers valuable knowledge and products for professional activities and education, but this could reach even more people. The extent to which information is visible and easy to find plays a role in this, and that is precisely the reason why the National Platform for Practice-Based Research (Nationaal Platform Praktijkgericht Onderzoek or NPPO) was established.

eScience Centre

The Netherlands eScience Centre is the Dutch national centre for the development and application of research software. It brings together knowledge, people and institutions to build software that improves the use of digital methods in research. This involves working directly with public research organisations and researchers in all disciplines.

Digital Competence Centres

Universities of applied sciences work together in Digital Competence Centres (DCCs) for Practice-Based Research to further facilitate RDM, FAIR data and data-intensive research at universities of applied sciences to help them achieve their open science ambitions. SURF supports the national DCC for Practice-Based Research with pilot projects, training and expertise. Universities of applied sciences will tackle urgent issues in terms of RDM, FAIR data and data-intensive research in six work packages.

4TU.ResearchData

4TU.ResearchData is an initiative of the 4TU.Federation that aims to provide an archive for long-term access and the structured collection, filtering, assessment, interpretation, organisation and sharing of research data, with a focus on technical sciences. This archive went live in 2010 and has since been managed as a service that enables researchers (from universities around the world) to upload and share their research data so that other researchers can then download and use the research data in their research.

NARCIS

NARCIS is the Netherlands' main national portal for anyone looking for information about scientists and their work. NARCIS is used by scientists, students, journalists and employees in education, government and industry. The portal provides access to scientific information, such as (open access) publications originating from the repositories of all Dutch universities, KNAW, NWO and various scientific institutions, datasets of a number of data archives, and descriptions of research projects, researchers and research institutes.

Amsterdam Data Exchange

The mission of the Amsterdam Data Exchange is to connect the academic world, business world and society in a diverse Data Science and AI ecosystem in the Amsterdam region. This network focuses on collaboration, innovation and involvement.

IRODS

IRODS stands for Integrated Rule-Oriented Data System, a data management software system that links unstructured data to metadata. It offers researchers working with large amounts of research data over an extended period a good way to document and store their research data securely. IRODS can help researchers demonstrate the integrity of their research data. SURF has announced that it will set up an IRODS centre of expertise, which is an important development for the use of IRODS in the Netherlands.

European Open Science Cloud

The European Open Science Cloud (EOSC) is an initiative of the European Commission that aims to achieve an EU virtual environment where all researchers. can store, manage, analyse and reuse research data for education, innovation and further research purposes. The EOSC architecture includes the OpenAIRE Research Graph, a collection of metadata and links that connect research entities. This include articles, research data, software and other research products and entities such as organisations, financiers, funding streams, projects, research communities and data sources.

GO FAIR

GO FAIR is a bottom-up, stakeholder-driven and self-managing initiative that aims to implement the FAIR data principles to make research data findable, accessible, interoperable and reusable (FAIR). It offers an open and inclusive ecosystem to individuals, institutions and organisations collaborating via Implementation Networks (INs). The INs are active in three key areas: GO CHANGE, GO TRAIN and GO BUILD. GO FAIR advocates the FAIR data principles based on those these three activity pillars.

FAIRsFAIR

FAIRsFAIR aims to provide practical solutions for using the FAIR data principles throughout the entire life cycle of research data. The focus is on fostering a FAIR data culture and using good practices to make data FAIR. FAIRsFAIR wants to play a key role in the development of global standards for the FAIR certification of repositories and their data. FAIRsFAIR also aims to provide a platform for the use and implementation of the FAIR principles in the daily work of European providers of research data and repositories.

Zenodo

Zenodo is a multidisciplinary open repository that is maintained by CERN. Datasets, documents and other research resources can be located using the Zenodo search engine. Scientists from any research discipline can upload data in any file format. All Zenodo files are automatically assigned a Digital Object Identifier (DOI). Details on assigning metadata to research datasets are provided under Section 5 (c) of the EUI Library Research Data Guide.

Figshare

Figshare is an online open access repository where researchers can store and share their research results, including figures, datasets, images and videos. Content uploading and access are free in accordance with the open data principle.

Mendeley

The Mendeley platform, which is part of Elsevier, aims to help researchers to organise, collaborate online and discover new research. Elsevier makes the platform available free of charge.

Netherlands Code of Conduct for Scientific Integrity

The Netherlands Code of Conduct for Scientific Integrity defines five scientific integrity principles, 61 standards of good research practices and duties of care for institutions. This code brings the Netherlands in line with other international scientific integrity developments. The code is endorsed by KNAW, NFU, NWO, the TO2 federation, VH and VSNU. These organisations ensure that the institutes, university medical centres, universities of applied sciences and research universities they represent or are responsible for endorse this code. Other institutions, including private companies, can also endorse this code.

4.2 Considerations

Although the inventory of the current situation does not fully reflect the Research Data Management landscape, it clearly shows the large fragmentation. We have also identified other problem areas, the most important of which we describe briefly below. The problem areas and fragmented landscape will be used as input for the further development of this domain architecture.

Recognisability and waning trust

Various types of data are used to conduct research. The reliability and quality of this research data are not always clear or assured. This lack of insight is increasing as society continues to digitise and produce large amounts of data. The origin or reliability of the data is not always clear. This lack of insight puts pressure on the reliability and quality of the research, and with it the trust placed in the research results, even though science

has mechanisms to filter out such unreliable or low-quality research. Insight in the reliability and quality of research data is essential for the acceptance of the research results. Such insight will contribute to the recognisability and reliability of the research results.

Big data is changing research methods

The digitisation of our society is leading to significant growth in available research data. And this growth, combined with new technologies, is in turn leading to new forms of research and methodologies. This further increases the possibilities for data-intensive research. Artificial intelligence now makes it possible to find patterns in research data, which in turn lead to the discovery of new research questions that would otherwise have gone unnoticed. It is important to support these new forms of research, as they enrich and accelerate research in a general sense.

Data management needs attention

Correct research data are essential for proper research. It must be ensured that the quality of these correct research data is assured and maintained. This requires research data to be managed. Management of research data is becoming (in the broad sense of the word), more and more of a specialised area that needs constant attention. For example, metadata must generated and maintained correctly, research data must be stored correctly, and so on. Researchers often manage research data as a 'side task'. However, the management of the research data is not assured when the research has been completed, even if the data still be relevant for other and future research.. It is therefore important to devote enough time and attention to managing research data in the long term with an eye to governance and the right roles and functions.

Current services based on old starting points

The current IT services that support researchers in obtaining, processing, applying, publishing, etc. of research data were originally not based on modern research data principles. Developments such as FAIR and Open Science set requirements for research data that are not always and/or not easily supported by current IT services. This can lead to researchers lacking essential research data or having to make an extra effort to obtain the right research data. Further development of existing IT services or the development of new services based on new principles such as FAIR and Open Science are essential to make research future-proof.

Little reuse of data

It is beneficial to research if the right research data are easy to find and use. However, research data are not always easy to find, acquire or exchange. This leads to inefficiencies in the research process. The application of standardisation and use of the FAIR principles can make research data more interchangeable and easier to find. This will increase the efficiency of research.

Conflicting requirements and too many requirements

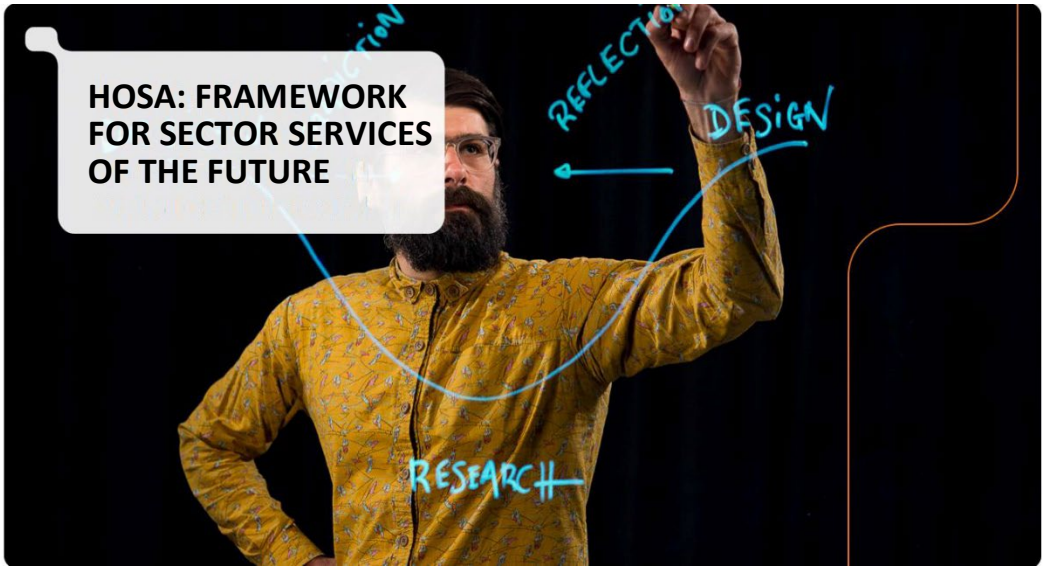
Research data come in many shapes and sizes. Research data is subjected to various requirements in terms of storage, processing, access and so on. It is not always clear what these requirements entail, for example with regard to the GDPR or embargo periods. The requirements and the interpretation of the requirements may differ from one research data provider to another. Clarity in the requirements regarding research data and in the interpretation of these requirements is important to keep the management of research data feasible.

Ownership, governance and financing

Research data management is essential for future research. For example, research data must be assigned metadata in the right way, stored in the right way and made available in the right way (see also 'Data management needs attention'). Managing this requires not only attention and the setup of ownership, but also the right financing, funding streams and governance for research data.

Data quality assurance

The increasing amount of research data makes it increasingly difficult for researchers to determine which research data are of sufficient quality for the research they are doing. Digital support and a process-based approach can help with this.



5 Ambitions of the sector

Good, reliable research is becoming more and more important. Research outcomes are increasingly used to resolve societal issues and to make social and economic decisions. In politics, too, more and more use is being made of outcomes from research. At the same time, the complexity and cost of research has rocketed.

The Dutch Ministry of Education, Culture and Science has therefore worked with the sector to set a number of goals in the strategic agenda[2]. The main goal is to make research future-proof with the emphasis on collaboration (regional, national, international), alignment with society and regional integration. Independently from the higher education sector, goals such research quality and business continuity in education and research are attracting a lot of attention in the sector.

Based on this perspective, the higher education sector has set ambitions to keep research future-proof. The following sections describe these ambitions, which have a significant effect on the way research data are handled.

5.1 More international collaboration

Renewal and innovation are important drivers of economic development. Research is often the basis for renewal and innovation. Research is therefore an essential element of economic development. At the same time, research issues are increasingly located at a European or international level. To take advantage of these developments, the Netherlands wants to achieve more collaboration in terms of research and research data at a European level. This will allow the Netherlands to retain its position in the top level of international research and to ensure the continuity of research, particularly in perspective of the major investments and efforts being made by the US and Asia in this respect.

5.2 More regional and national collaboration

The higher education institutions want more regional and national collaboration. Institutions will play a more important role at a regional level in the future in terms of driving knowledge development and innovation. Cities and regions are increasingly forging their own profiles, and research and social topics are becoming increasingly interconnected. This requires collaboration between institutions, companies and civil society organisations at a regional level and a good integration of the institution within the region. In this regard, regions are not limited to a specific area within the Netherlands and it can also be an area within Europe or within another part of the world. Examples of these trends are the initiatives to develop practice-based research and Digital Data Competence Centres and Centres of Expertise. In this the institutions act as hubs of smart regions.

Research in the Netherlands is effective thanks to smart collaboration combinations between various parties. For example, collaboration with high-tech companies is essential to be able to use the necessary advanced equipment. To reduce overlap, create more synergy and reduce unnecessary competition between institutions, more regional and national Research Data Management collaboration is needed.

[2] *Strategische agenda hoger onderwijs en onderzoek – Houdbaar voor de toekomst,* Ministerie van OCW, Den Haag, December 2019

5.3 Alignment with society and education

Research is a cornerstone of society, and must therefore be continuously aligned with society so the results of research can be converted into useful and applicable solutions for social and economic innovation. Alignment also contributes to the discovery of new research topics.

The alignment of education and research in the Netherlands has traditionally been well developed. Students in the Netherlands are given the opportunity to develop an inquisitive attitude and creative thinking. However, the strong growth in the number of students puts this alignment under pressure, for example when it comes to the funding of both education and research.

The ambition is to ensure permanent alignment with society and education. This includes providing insight into social and economic developments and linking these to research themes. This will support the connection between research and the national strategic knowledge policy and portfolio management for research topics.

5.4 Open Science

The Netherlands has the ambition to open up the research process, under the name Open Science. Open Science aims to expand availability for (re)use, broaden participation in research, and increase the applicability of research. Open Science started as the active disclosure of final publications, but in the future it aims to make the entire process – including the research data – transparent, with citizens and professionals contributing to research via Citizen Science.

The Open Science developments are changing the process of research and are making it more transparent. Others participants are given the opportunity to collaborate, and to contribute to and make use of the scientific process. For example, non-scientific users can help gather research data and influence the research world by asking questions and offering ideas.

5.5 Ensuring the quality of research

Research should always be internationally competitive and lead the way. It is therefore important that the quality of research remains to be assured. Data are an important cornerstone of high-quality research.

The reuse of research data will help to assure the quality of research. However, reusing good research data is not easy, for example because not all research data are easy to find, accessible, exchangeable and reusable. Reuse is made easier by specifying the use of the principles of findable, accessible, interoperable and reusable research data (FAIR), by explicitly making the connection with international Linked Open Data initiatives, and by setting up good data management (data stewardship).

The aim is to assure the quality of research, for example by specifically applying the FAIR principles and by setting up good data stewardship for research or the data resulting from research.

5.6 Assurance of higher education independence

The independence of research must be assured in the Netherlands. Institutions are increasingly reliant on commercial platforms, which are attractive to individual scientists because they are easily accessible and inexpensive. As a result, research is becoming increasingly dependent on commercial companies that tend to operate outside Europe *(see also the Scientific Technical Council's report on collaboration on the service infrastructure and competencies for data).* The rectors of Dutch universities have indicated that it is not desirable to leave education and research entirely in the hands of large commercial parties. In their opinion, research in the Netherlands should not become too dependent on these platforms.

5.7 Continuity of research

Now that research has more and more digital support, there are also more continuity risks for research data. For example, institutions are vulnerable to cybercrime activities that may involve stealing or compromising research data. It is therefore important to reduce such research data continuity risks by safeguarding research data and assuring their availability.

6 Technological developments

The digitisation of our society is causing an exponential increase in the amount of research data. This increase in research data and other developments driven by technology are affecting research and specifically the Research Data Management issue. The following sections describe a number of identified developments and their impact on the RDM issue.

6.1 Internet of Things

The development of IoT will contribute to a sharp increase in the amount of research data. Many devices and services use sensors and together form large networks generating a wide range of environmental data. IoT devices can be used to generate data that can be researched. For example, the future networks of smart cities can be used for social research into how people live and work. The IoT developments can also be combined with artificial intelligence to recognise certain trends or patterns. In addition, IoT can contribute to data validation by acting as a reference and by signalling trends.

6.2 Artificial intelligence

Various forms of artificial intelligence (AI) offer very powerful tools for all aspects of the research process. For example, data can be collected on a specific research topic, research data patterns can be discovered, and publications can be written automatically. It is important that we can always monitor how AI is handling the data.

Several other countries also see AI as a crucial element in international competition. There have been several initiatives in the Netherlands and Europe in response to this, such as the AI-ROBOTICS initiative of the European Commission and Shell's development of a sniffer robot in the Netherlands.

6.3 Blockchain

Although blockchain is still very much developing, this technology offers opportunities to generate trust in information services in many areas of society. Trust used to be based on institutional bodies and legal entities, but blockchain means these bodies and legal entities are no longer necessary or at least not quite as necessary as before. A wide range of experiments are being conducted in various sectors to find applications. For instance, the financial sector has invested heavily in blockchain, but this has not yet resulted in a widely used application. Blockchain offers many possibilities in various areas of the research process and also provides opportunities to assure the reliability of research data. This technology can also be used to assure the reliability of the research process and the research data.

6.4 Data Science

With technology such as Data Lakes and Scalable Data Analytics, Data Science offers new possibilities for research that allow it to be more data-driven. The more we can combine the increasing amounts of data, the more valuable they will become. Technological developments such as data analytics can contribute to this and may help to generate new research data from existing data.

6.5 Computing

Processing power continues to grow. Developments in quantum computing, parallel computing and distributed computing will help us to process multiple scenarios. This will also lead to other forms of research, which in turn will have an impact on data for research purposes.

6.6 Trusted (third) parties

TTPs can play a role in the management, processing (for example pseudonymisation and encryption) and exchange of research data between parties. TTPs ensure compliance with legislation and regulations regarding the privacy and protection of research data. TTPs can also increase trust in science with certification based on international standards.

6.7 Platforms and infrastructure

Research is increasingly performed by multiple researchers. It is important that international research results are exchanged properly. Virtual Research Environments (VREs) and data management platforms offer (online) working environments specifically for researchers, and often include storage and distribution. The availability of VREs based on international standards simplifies the necessary collaboration.

Concepts such as Hybrid Cloud Infrastructure also offer flexibility and scalability for aspects such as data storage, processing power and bandwidth. The hybrid approach can accelerate research, where necessary in combination with the required confidentiality and integrity.

7 Architectural vision

Various reports have stated the need for more cohesion between the ICT services and infrastructures for research data management. One of the publications by NWO indicates that there is a need for better cohesion between the various RDM sector services and that the CIOs of the institutions and SURF should take the lead in this. The National Open Science Programme's final report also states that a joint architecture should be developed. This need for architecture is not only a national issue. According to EU Commissioner Carlos Moedas, for example, there is no shortage of initiatives, but what is lacking is a sort of one-stop shop for researchers and there is no architecture that allows researchers to join in. But what should this architecture look like?

7.1 Architectural vision

HOSA offers an architectural vision for the higher education sector with a horizon of three to ten years. Reliable, high-quality research is crucial to society's development. Research is constantly changing because of digitisation. These changes have an impact on how research data are handled and on what type of support is required from sector services. This architectural vision is based on the sector's initiatives and ambitions, technological developments, and points of attention in the current situation.

The Netherlands has a number of ambitions which are facilitated by research within universities of applied sciences and research universities. One ambition is to create a more data-driven, knowledge-based economy in the Netherlands, as mentioned in the Knowledge and Innovation Covenant. Another example is regional innovation, in which regions willingly work with higher education institutions to create smart regions that are better aligned to new developments. The Ministry of Education, Culture and Science has set a number of goals together with the higher education sector. Organisations such as UNL, VH, NWO, KNAW, DANS and SURF have also defined goals. In HOSA, these goals are linked and shown in the objectives structure. We need sector services based on principles that suit these objectives in order to achieve those goals. HOSA has developed these into a framework.

Collaboration

The economic and social value of research is increasing. Research is a competitive factor. At the same time, the costs of research are increasing. This makes it financially and organisationally harder for a relatively small country like the Netherlands to continue to perform high-quality research. Collaboration at regional, national, European and global levels is therefore necessary. Combining strengths at regional, national and international levels increases efficiency and effectiveness because research data and resources are shared. Collaboration at regional, national and European level will be important to increase the Netherlands' competitive research position and sustain the country's high-quality research. The complexity of social issues is also increasing. This will require, for example, more interdisciplinary collaboration between various subject areas which in the long term makes it necessary to establish far more connections between infrastructures, which currently tend to be organised by discipline.

Open research

The requirements for research openness and transparency are increasing. This is driven by the trend towards Open Science, as well as the desire to achieve research results more quickly. This will mean that research will no longer just take place at research institutions. In the past, the research process was even less open, so the end result was an item that was made available. Soon, the intermediate steps and underlying data will increasingly be made openly available. Open Science is driven and strengthened with the help of sector services.

As a result, specialised parties will be able to participate in the research. This creates various collaboration partnerships with both public and private parties. Companies, hospitals, media and governments can participate by making data available to each other or by actively contributing to research. One current example of this type of collaboration and the trend towards open data is the Dutch government's goal to make all public data files actively available for use in research. This creates a situation in which researchers are increasingly making their data files available to each other (while observing privacy regulations). These research data are then reapplied in the development of policy or are used in other research.

To facilitate these developments while ensuring aspects such as efficiency, security and reliability, sectoral collaboration will be necessary and sector services will emerge to provide maximum support for the trend towards open research across organisational boundaries. Whereas previously the focus was on research universities, universities of applied sciences and research institutes, now governments, companies and hospitals are also coming into the picture. This concerns not only management of the research data, but also provision of clear and transparent services to those involved in the research process.

Shift of focus

Research results enter society in different ways and in various stages. Sometimes research results are applied in education, sometimes research leads to new treatment methods for illnesses and afflictions, and sometimes research results lead to new policies. Figure 2 gives an impression of the wide range of possible valorisation and dissemination paths.

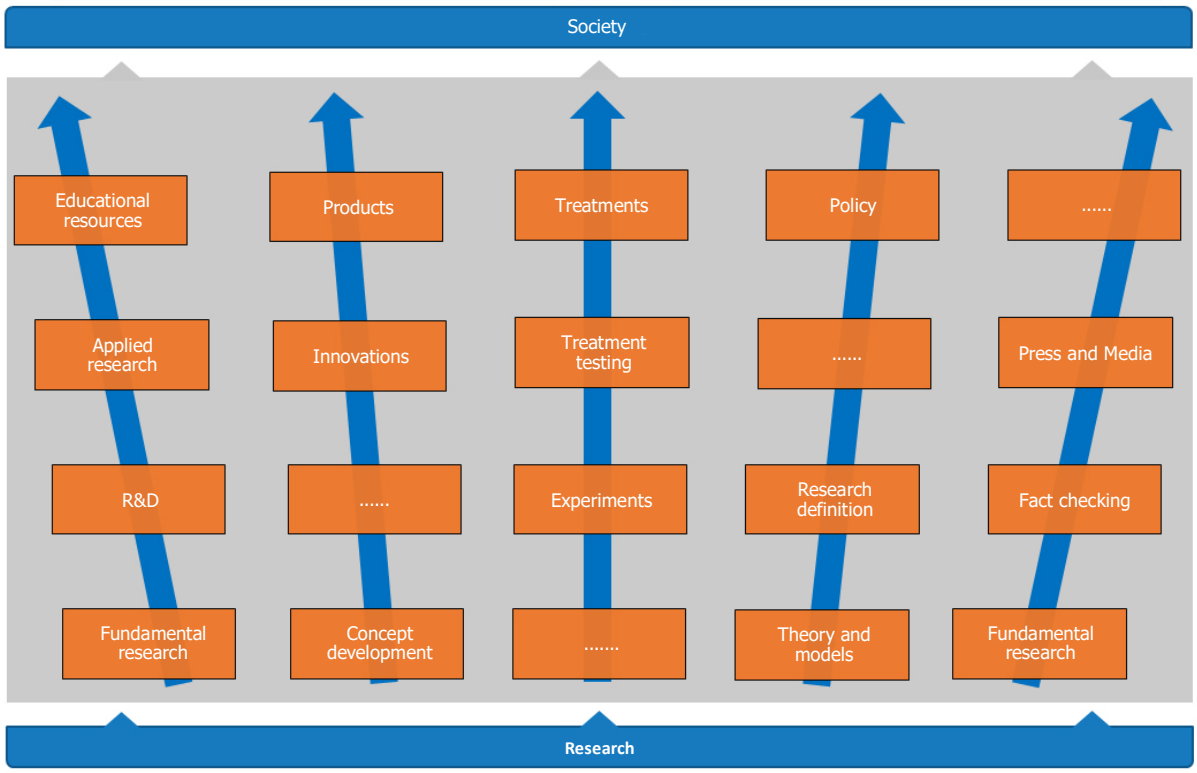


Figure 2: Change of focus

The speed at which society is changing due to increased digitisation and globalisation is placing more and more pressure on research to achieve usable results more quickly, for example in the form of applicable solutions, usable knowledge and applicable information. This results in a need to reduce the research's time-to-market. It also shifts the focus of research. We no longer simply focus only on research, but also on the recipient of the research results as well. These research results or interim results will be tailored to the recipient to accelerate

their applicability. The recipient or user of the research results plays a wide range of roles, extending from professor to journalist to entrepreneur and citizen.

The COVID-19 pandemic has already illustrated this change very clearly. The need for a fast yet safe and reliable vaccine requires a different approach to the research process – a far more open process in which knowledge or interim results are shared and made available more frequently. Making knowledge or interim results available requires an underlying infrastructure that makes it possible to easily provision knowledge available knowledge about interim results to many target groups so that they can be applied regionally, nationally and internationally.

This shift in focus will place different demands on how research data are handled. Research data, including results and interim results, must be tailored to the recipient. This makes the classification of research data more important. Ethics and privacy must be assured. Sector collaboration supports and facilitates these developments.

Social participation

Citizens in different roles and at different levels actively contribute to research also know as Citizen Science. Examples of Citizen Science include Hanny's Object, a reflection nebula discovered in 2007 by Hanny van Arkel, a Dutch teacher from Heerlen, and the tick radar (www.tekenradar.nl) where information about Lyme's disease is shared. A more recent example is the way the Netherlands National Institute for Public Health and the Environment shares information and research data on COVID-19 developments. Based on feedback on this data shared by the Netherlands National Institute for Public Health, more and more additional data is made available to researchers – in the broad sense of the word – that they can use in their own work to develop and add new insights. These kinds of developments will increasingly meet citizens' needs to be able to check the accuracy of the available data and information or to actively contribute to research. Future sector services should be able to support these wide variety of roles and levels of participation and information needs. However, citizen participation in research raises new questions about sector services. For example, How accessible should they be? What knowledge level is required? And how do we control access to data that require some shielding?

Specialisation in research subprocesses

The trend towards open research may lead to a more modular research process. Whereas previously the number of publications and their impact were mainly considered on the basis of citation scores of the journal in which publications appear, the development of high quality research data or valorisation of research will be increasingly appreciated. For example, parties may specialise in creating high-quality research data or processing it. Future RDM platforms contribute to this by having these processes carried out independently of each other. Modularisation and the resulting specialisation of researchers are the basic principles of the sector architecture.

These specialised parties can contribute to every phase of the research. Each phase of the research process requires different collaboration partnerships. The results are made available more quickly, openly and transparently and are then used as input for a subsequent step in the research or applied in other research. Sector services should be able to support this in the same way.

Digitisation

Societies are increasingly relying on information and information technology. Companies, governments and institutions are transforming into digital organisations in which digital information and information technology are the end products. The processes are completely digitized. Changes and developments are organised and managed based on the data resulting from and collected in these processes. Our society's extensive digitisation is resulting in large amounts of data. These data are also valuable to research. With large amounts of data, it is important to ensure that the information is easy to find, accessible, interchangeable and reusable, to keep research efficient and effective. Digitisation will lead to a high demand for services that properly support the use of these data.

7.2 Basic principles

The previous-mentioned developments are taking place over a decade, which is sometimes referred to as the digital age. Many businesses are transitioning into digital companies. The end product and the associated services are primarily digital and largely consist, for example, of data or software. The supporting processes have also been fully digitised. Higher education institutions and the sector will also go through this development. Some manual operations will disappear, and advice will no longer always be provided by humans. Digital concepts and technologies such as business platforms, AI and blockchain will also enter the sector.

International collaboration

Science is often internationally oriented and organised. For example, the EU is widely investing in European collaboration in order to compete with the US and Asia. Is it still necessary to organise Dutch sector services? The interests of Dutch citizens, companies, regions and other players do show that this need exists. First of all, this means that HOSA must be able to align with the European initiatives and standards. It must also be possible to reuse or scale up initiatives from the Netherlands in European countries. This is stated in HOSA's overarching principles. Integration patterns for international data exchange must be agreed.

Innovation and autonomy

One of the issues that often arises in architecture is to what extent the underlying technology should be standardised for everyone. The idea behind this is that this can save costs and increase interoperability. This is unrealistic in the field of research. HOSA therefore makes no statements about the technology that researchers, research groups or subject areas are using. This would reduce innovation in research and threaten the autonomy of science. However, HOSA does provide guidance on connecting the many parties and subject areas involved in research, the users of research results, and the intermediary technology required to connect the parties.

Platform strategy

Developments in education and research are increasingly calling for interaction between different parties, such as students, lecturers, researchers, companies, governments and citizens. This involves the exchange of content, data and services, for example. HOSA therefore uses the concept of business platforms to draw up the application architecture for the sector services. This concept provides insight into online marketplaces, which are emerging in various places thanks to increasing collaboration within the sector.

Business platforms are rooted in the tech giants which are reshaping service delivery. For the higher education sector, HOSA mentions similar platforms, but based on the public sector and public values. These digital public sector platforms facilitate interaction and collaboration, connecting supply and demand, and steer towards common values and standards. This strengthens the independence of research.

Sector services support different types of collaboration

Collaboration is important in the field of research on many levels: between institutions locally in the region, internationally, within areas of expertise or with the business communities. For institutions, there is also the question how they can properly support their collaboration partnerships at the various levels. Currently these collaboration partnerships tend to be set up on a case-by-case basis, which has the disadvantage that scalability and interoperability are difficult. However, the ambitions in the area of collaboration of the higher education sector require such scalability. HOSA makes sector services suitable for collaboration on those levels by ensuring that the same services can be offered simultaneously on portals at the level of an institution, collaboration partnership, subject area, or a national or international initiative.

Diversity of interfaces and channels

The principle that sector services support collaboration partnerships at various levels will eventually lead to the creation of multiple portals using underlying services from sector services. Some examples of this are a national portal or an institution portal. Due to technological developments such as omnichannel and multiple user interfaces, the usual interaction with text on a portal will be insufficient in a few years' time. That is why sector services are designed to support interaction through a wide variety of interfaces and channels, including chatbots, mobile, gestures and games.

Data exchange

To really make data exchange possible, it is important to make arrangements regarding research data across the various institutions and to standardise data about datasets, publications and research projects, for example. Data will be exchanged between many parties, including public, private and foreign institutions. The sector will actively participate in the development of international standards and acquire knowledge on these standards. As a result, the sector not only assures international collaboration, but also national collaboration at the various levels.

Financial settlement

In the context of flexibilization, HOSA takes into account the expected need for financial settlement. HOSA also aims to handle all financial settlement in the research domain in a similar way. It should be possible to facilitate any mutual financial settlement between higher education institutions within HOSA in accordance with the chosen modular level, taking into account the specialisation and differentiation of research within the sector. A sector service can also be expected to provide support for mutual cost allocations between institutions and a flexible research budget.

8 Platform strategy

8.1 What is a platform?

Worldwide there is a development in which the concept of business platforms is central. Well-known examples are Uber, Bol and AirBnB. More and more traditional companies such as General Electric, Philips and Nike are taking inspiration from this model and have adopted business platform features. Platforms are also on the rise in the public sector. Examples in the education domain are EDx, Coursera, Amazon Ignite and Woolf. These platforms provide an environment that enables value-creating interactions between producers and consumers. A platform offers open, accessible services for these interactions and arranges the underlying governance.

The overarching purpose of the platforms is to link the supply and demand among buyers and service providers, content, knowledge and data, and to make their exchange easier. Business platforms can easily scale up their services because they usually do not produce the services, data, knowledge and content themselves. For example, Uber does not have any taxis or taxi drivers, and therefore does not have to buy any taxis or hire more drivers if demand rises. Uber also does not have to organise the maintenance of a fleet. Coursera brings together the demand of learners and the supply of other organisations, rather than organising education itself.

A platform integrates the functions and competencies that are necessary to deliver services to users. The platform can help to find the right supply, simplifies the provision of new services, supports the transactions between providers and users, monitors usage. and assures quality. The platform makes the interaction between the providers and users easier. A business platform offers an open, collaborative infrastructure for interactions between the parties operating on the platform, with governance that ensures a healthy ecosystem.

8.2 Why RDM platforms?

Based on the identified strategic ambitions in the sector, collaboration arises at all levels, which in turn leads to a need for far more interaction. This means more collaboration between institutions, international collaboration, collaboration with companies, collaboration in the region and collaboration with citizens. This requires an underlying mechanism that can handle all these kind of interactions and collaborations at different levels. The concept of business platforms provides a starting point for this.

The developments in education and research require exchange between all sorts of participants. The concept of business platforms offers ‘marketplaces’ that focus on this type of exchange.

In current Research Data Management (RDM) practice, a number of initiatives are being created that are already moving a little towards business platforms, such as Amsterdam Data Exchange and European Open Science Cloud. Amsterdam Data Exchange aims to offer an infrastructure that ensures a trusted and secure environment for real-time data-driven collaboration based on common rules, and that interested partners can join. The European Open Science Cloud aims to connect research infrastructures to offer researchers and scientific professionals an environment for storing, sharing and using data.

Section 4.1 has already shown that there are many RDM initiatives. We found that they show only limited consistency and there is some overlap as well. This leads to an unclear landscape of RDM services in which the customer has to find his own way, unnecessary rivalry arises and there is little or no visibility and control in terms of quality and reliability. In addition, many initiatives address only one or a few parts of the research process, while future developments take the research value chain as their starting point, as outlined in Chapter 7. These developments require an underlying infrastructure that can support the entire research value chain in an integrated way.

The business platforms model within HOSA supports the creation of consistency among the various initiatives in the education and research domains. It is based on services and governance that are in line with the sector and public values. It brings structure, a common language, an overview and a checklist to new initiatives. As a result, the conversation between various stakeholders can be conducted in a more structured way and decision making about deviations from the architecture can take place explicitly.

Public business platforms are based on services and governance that fit the sector and public values.

From this, an outline of the various HOSA platforms is developed. The data management platforms bring together the supply and demand for data and data management services, combine forces, and drive collaboration and innovation in the area of research data. Providers are offered an environment in which they can easily connect services and make them available to users. Users are offered an environment in which services can easily be found and obtained. This creates opportunities to achieve efficiency in areas such as financial settlement, licensing and support.

The platform approach alligns with society's expectations for modern services, which include ease of use, transparency and experience. Thanks to its governance setup, this approach helps to ensure the independence of research. Finally, the chosen setup offers the option to switch between the different levels scientists can operate at, such as international, national, regional and within research domains.

8.3 RDM platforms

As indicated in Section 3.3, we have taken the research cycle and research data cycle described in HORA 2.1 as the starting point for the description of this domain architecture. From this perspective, the research process can be roughly divided into four main processes:

- *Research definition*. This main process focuses on mapping out and defining research topics.
- *Research preparation*. This main process focuses on the preparation of the intended research.
- *Research execution*. This main process focuses on the part in which the actual research is carried out.
- *Knowledge utilisation*. This main process focuses on making research results (publications, research data and so on) available to a wide audience.

These main processes create, exchange, process, etc. (a form of) research data. Three of the four can clearly be seen in HORA under the 'Research cycle' theme. The 'Research preparation' process is related to the 'Research development' business function, the 'Research execution' process is related to the 'Research execution' business function, and the 'Knowledge utilisation' process is related to the 'Research dissemination' and 'Knowledge utilisation' business functions. The 'Research definition' process is not a direct part of the research cycle itself, but an essential, more strategic process step to give guidance to future research.

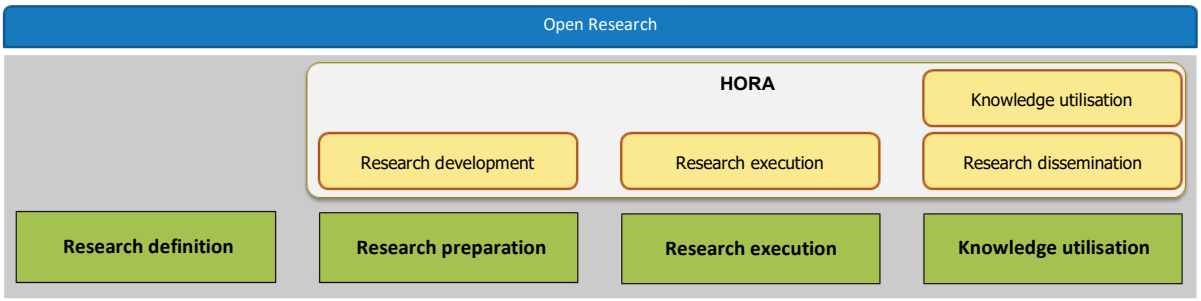


Figure 3: Relationship between HOSA RDM and HORA

Based on the four main processes and the platform strategy, a setup based upon three data management platforms can be seen. The platforms focus on supporting the RDM process. They are:

- Smart Region Platform. For the development of smart local, national and international regions, it is necessary to jointly develop an understanding of the region's knowledge needs now and in the future. The Smart Region Platform offers opportunities to gain and align this insight together. Research themes are mapped out and defined based on data and data analysis. The platform offers opportunities for researchers to learn about current research paths.
- Smart Research Data Platform. With their campus, institutions form an important part of a society's development. All kinds of partners are already located on or near a campus. In the digital world, institutions can have a similar role. Good research data are essential for doing good research. Institutions can use the Smart Research Data Platform to lay a foundation for exchanging research data and hence for developing a digital campus. Partners such as companies, governments and hospitals can set up a presence on this platform. The Smart Research Data Platform supports demand for and supply of research data and offers services in that regard.
- Open Research Tools Platform. Research data processing, analysis, visualisation and so on require specialist and often expensive services. It is undesirable to have each institution set this up for themselves, in part from the viewpoint of keeping research affordable and in part from the perspective of collaboration. The Open Research Tools Platform brings together various RDM services in a smart way. Institutions can use it to make their own services available or use the services of other institutions and partners. This strengthens the development of the digital campus and the region.

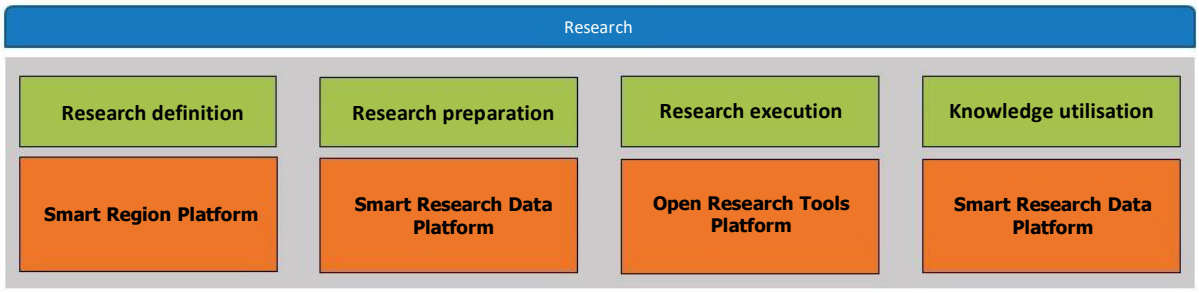


Figure 4: HOSA RDM platforms

The platforms support the entire RDM process of research data and connect seamlessly with each other. Chapter 10 explains the platforms in more detail and describes them conceptually.

8.4 Platform setup

The platform strategy requires a setup that is in line with the architectural vision.

Integrity, transparency and reliability
In general the platforms are set up in such way that they behave as honest, transparent and reliable service providers towards the various stakeholders such as citizens, researchers, regions, public institutions and private parties.

Open and modular setup
The RDM platforms have an open and modular setup which groups and makes available various RDM functions. The components are loosely coupled, which makes the platform easy to adapt to the changing wishes and requirements of the participating parties. The use of open standards is an important element, as it makes it easy to provide data management services, for example. The modular setup and open architecture encourage reuse and connection to the platforms, as well as easy (de)coupling of data and services, and interoperability between the platforms.

Access to services for and by institutions
Open and modular architecture, combined with the use of standards, allows institutions to access and make available services of the platform via their own portals. Institutions can also decide to connect and make their own services available with specific features (nationally) via the platforms. The platforms help institutions to bring the flair of physical collaboration on a physical campus into the digital world. This also encourages the development of the digital campus, which offers institutions the opportunity to become hubs in the smart cities of the future.

International alignment
The platforms can also be connected to international initiatives or used internationally. The use of a common foundation as a template makes it easier to connect local, national and international initiatives to develop a common RDM ecosystem.

Intermediary role
The platforms act as a marketplace for research data and research data services. They bring together supply and demand for these data and services. The platforms play an intermediary role between these different parties and provide overarching governance to achieve and maintain the right quality, security, reliability and integrity of the research data and services.

The platforms provide the core of what institutions cannot organise

Governance
The governance setup of these platforms assures aspects such as independence, transparency, integrity and reliability. This governance represents the key stakeholders.

Management
Platform management controls the offered data and services and assures the quality frameworks and service levels. It is authorised and responsible for the development, implementation and monitoring of the platform's quality assurance, integrity and reliability. Platform management identifies and responds to services and data that do not meet the established standards. This calls for overarching agreements and standards, as well as means to monitor these agreements and standards and to identify the necessary changes to these agreements and standards, for example as a result of changes in national agreements. Platform management handles the setup, (further) development and implementation of the platform in collaboration with the key stakeholders.

The platform strategy supports the creation of a national RDM ecosystem that can also be used internationally or in line with international initiatives and developments. The platforms organise the wide landscape of RDM initiatives and support the development of coherent RDM.

8.5 Complementary initiatives

In an international context, there are initiatives that are similar to the platform strategy described above or that cover parts of it. The platforms in this domain architecture are complementary to these and are not meant to replace them. The platform strategy also makes it possible to offer services at a regional, national or international level, with end user support being organised via the platform (see also the description of the platforms).

9 Stakeholders

The stakeholders within the HOSA Research Data Management (RDM) domain can be categorised into five groups: Investors, Society, Service Providers, Institutes, and Researchers. These groups consist of parties with roughly the same type of interest in the research data. They are partly interdependent in the RDM domain and do not have an individual decisive role to play in that regard. The platforms are formed based on the collaboration partnerships between two or more stakeholders within one group or between groups.

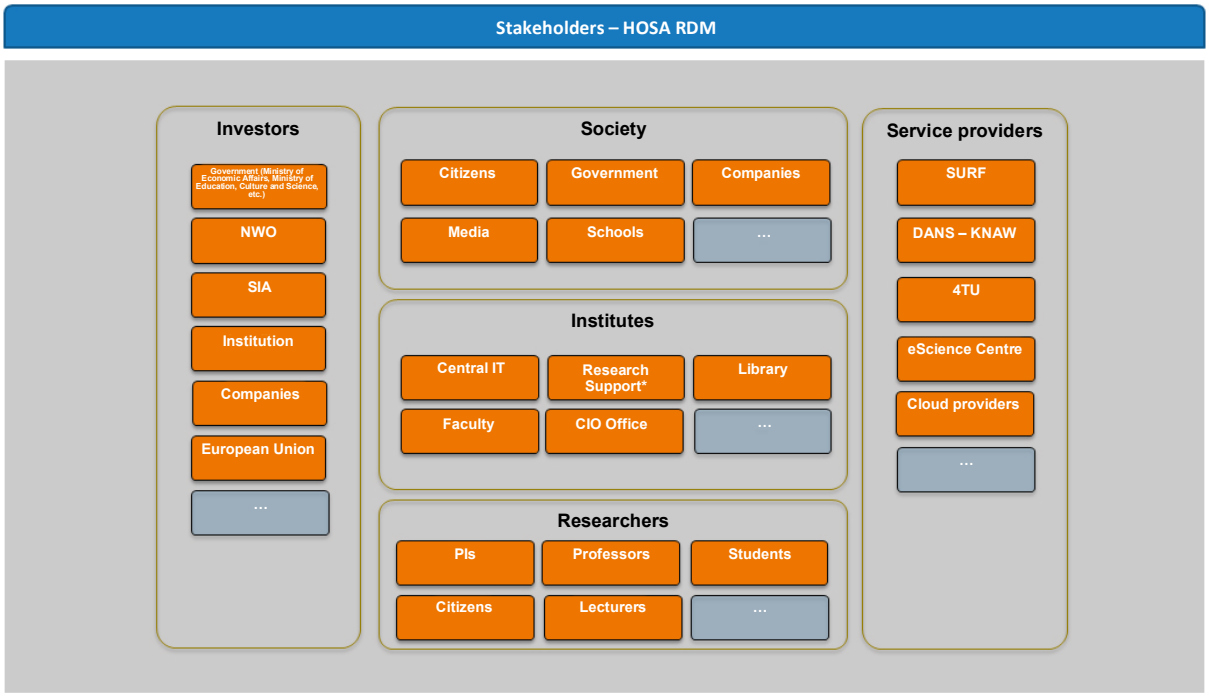


Figure 5: Stakeholders

Note: The parties in this overview can fulfil several roles. They act as an example in the context of stakeholder groups.

9.1 Society

The first group of stakeholders is society. Society benefits from research through the creation of products, enrichment of education or knowledge, and use information to make decisions. This includes citizens, media, governments and so on.

This group benefits from and wants quick insight into transparent, recognisable, reliable, high-quality research. It also benefits when the reliability of the research is easy to determine and when the research results are quick and easy to find and apply. The platforms offer this reliability, accessibility and transparency.

The added value for this stakeholder group lies in the contribution the platforms will make to the encouragement and support of open research, in which it is easy to participate and where reliable results from research can be found.

9.2 Investors

The second group are investors in research. Governments (or organisations associated with governments, such as NWO) and businesses invest a lot of money in research every year. However, they are facing a sharp increase in the costs and complexity of research. At the same time, there is a strong need to market the value of research much more quickly, which means that there is a growing need for support of researchers.

Collaboration is necessary to keep research affordable now and in the future and to be able to cope with the complexity. Collaboration is also required to reduce the time to market of the research.

For investors, the added value lies in the fact that the platform setup facilitates collaboration and the pooling of forces in an open and transparent arrangement that takes into account aspects such as independence, reliability, privacy, security and integrity.

9.3 Institutes

The third group are the institutes, which are the higher education institutions, research institutes and research centres. Research – from fundamental research to applied research – is largely driven by these institutes and carried out at these institutes. They are therefore an important factor for innovation in the region they are located in. At the same time, it is becoming increasingly complex for these institutes to keep supporting the entire research process fully. Smart collaboration partnerships between institutes and with private or public parties can make this complexity manageable and also drive innovation.

The added value for these institutes lies in the fact that they can make their own facilities available in a smarter way via the platforms and attract third-party facilities for their own research. This avoids the need for institutes to set up services that are already available at other institutes. Preventing such duplications allows more room for diversity and innovation.

9.4 Researchers

The fourth group of stakeholders are the researchers. Researchers benefit from support that gives them easy access to the facilities, tools and research data they need for their research. Researchers also benefit from having to do as little administrative work as possible, so they can focus on the research itself.

Researchers will experience added value because the platforms help them to find and access the necessary services and research data and to disseminate the knowledge and results arising from the research.

9.5 Service providers

The last group of stakeholders are the research service providers. Research service providers are faced with an opaque landscape of research efforts. Sometimes it is unclear what is needed and the fragmented landscape acts as a barrier to the offering of the services; at other times the need is clear, but a lot of customisation is needed, which is complex and expensive.

The added value of the platform setup for service providers is that it makes their services far easier to be found, obtained and put together. A clear demand-supply situation also arises for service providers.

10 Process architecture and application architecture

The starting points for describing the process and application architectures are the platform approach and platform layout described in Chapter 8. The changes to the research process, and specifically the developments in open research described in Chapter 7, are also the basis for the development of the process and application architecture. This gives us the following picture:

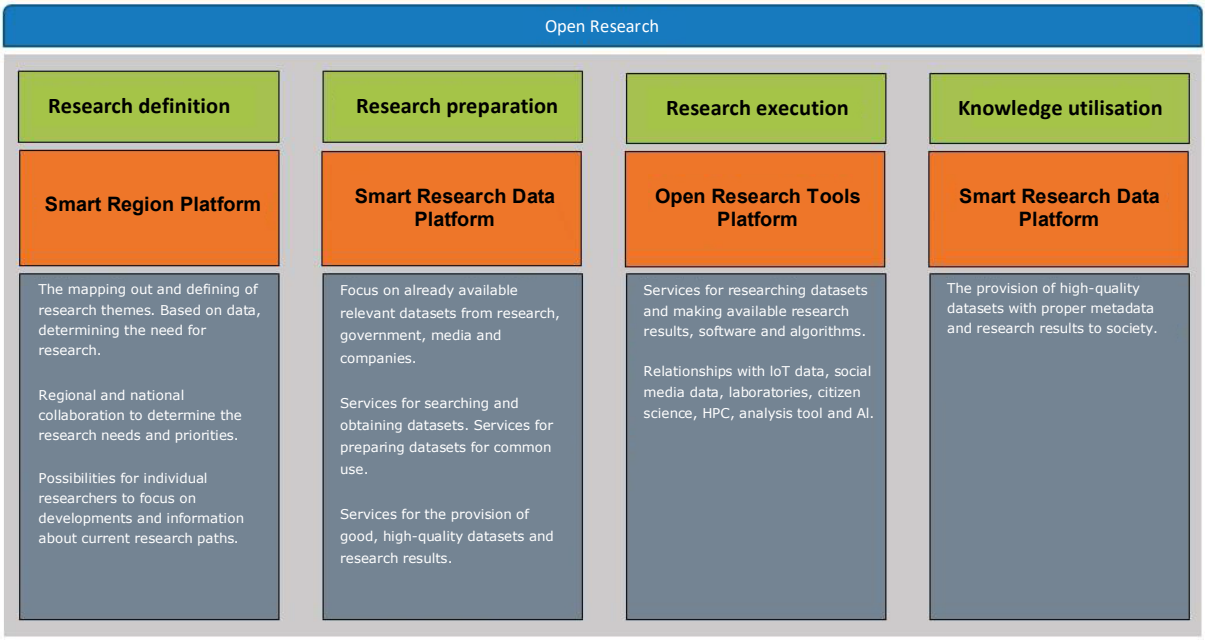
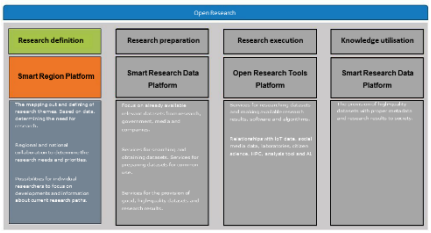


Figure 6: Processes in relation to HOSA platforms

The following sections describe the main processes and provide insight into the conceptual design of the corresponding platforms based on those processes.

10.1 Research definition

What happens in this process?



The 'Research definition' process identifies and defines the research themes. In the current situation, this seems to be a diffuse and largely non-data-driven process that involves many stakeholders exchanging data 'manually'. The strategic agenda for the higher education sector strives towards better demarcation and wider alignment in and with the regions, for example. In this process, officials from institutions, sector partners and governments work together in the role of policy-makers, and also researchers who orientate on new research themes. Other public and private sectors are also involved in the process. All stakeholders provide input on the need for new knowledge during a few years, resulting in collaboration at regional, national or an even wider (for example European) level to determine the research need.

The aforementioned officials and researchers collect, process, analyse and visualise research data to determine the research need. Insight is gained into the future knowledge needs based on data analysis and visualisations. This insight is combined with the knowledge of existing studies on defined research themes. This combination enables officials and researchers to define and manage research themes for the future. This needs assessment can take place at different levels ranging from local, regional, national to international.

It is likely that those involved have both a demand for and supply of relevant research data. In the current situation, the parties mainly identify the required research data themselves to determine the research needs, and then process, analyse and visualise these themselves. Both the collected and processed research data and the results of the analyses and visualisations are then stored locally. Working together on this will result in fewer duplication of research, the emergence of more visible future themes, and the encouragement and clear visibility of collaboration on research themes.

What is the role of the platform in this process?

In its strategic agenda, the Ministry of Education, Culture and Science mentions the ambition to use sector-based plans more. This should help encourage collaboration and an integrated approach to research themes. The focus of the Smart Region Platform is to support collaboration, interaction and an integrated approach to define research themes, and to provide insight into the current research portfolio. The platform can be used to visualise research developments nationwide. The platform supports this by bringing together facilities that support this process.

> Relevant and up-to-date policy information is also made available on a single public platform. (Strategische agenda hoger onderwijs en onderzoek – Houdbaar voor de toekomst, Ministerie van OCW, Den Haag, December 2019).

Besides supporting research policy development in the research domain, this platform also helps to define education in the education domain. The stakeholders can broadly map out the need for knowledge for the future in the region.

What components does the platform consist of?

The Smart Region platform is designed to help map out and define the research themes. It supports employees in their roles as policy-makers, researchers or data scientists at institutions, partners or governments. This platform offers an environment where parties can exchange and bring together data. This includes data about the research itself, as well as macro-economic and trend data. Not all data are on the platform itself. Instead, the platform offers opportunities to obtain data from participating parties that is made available by other sources. The platform can handle supply and demand for both structured and unstructured data. Institutions, NWO, KNAW, the Ministry of Education, Culture and Science and CBS are all obvious data providers.

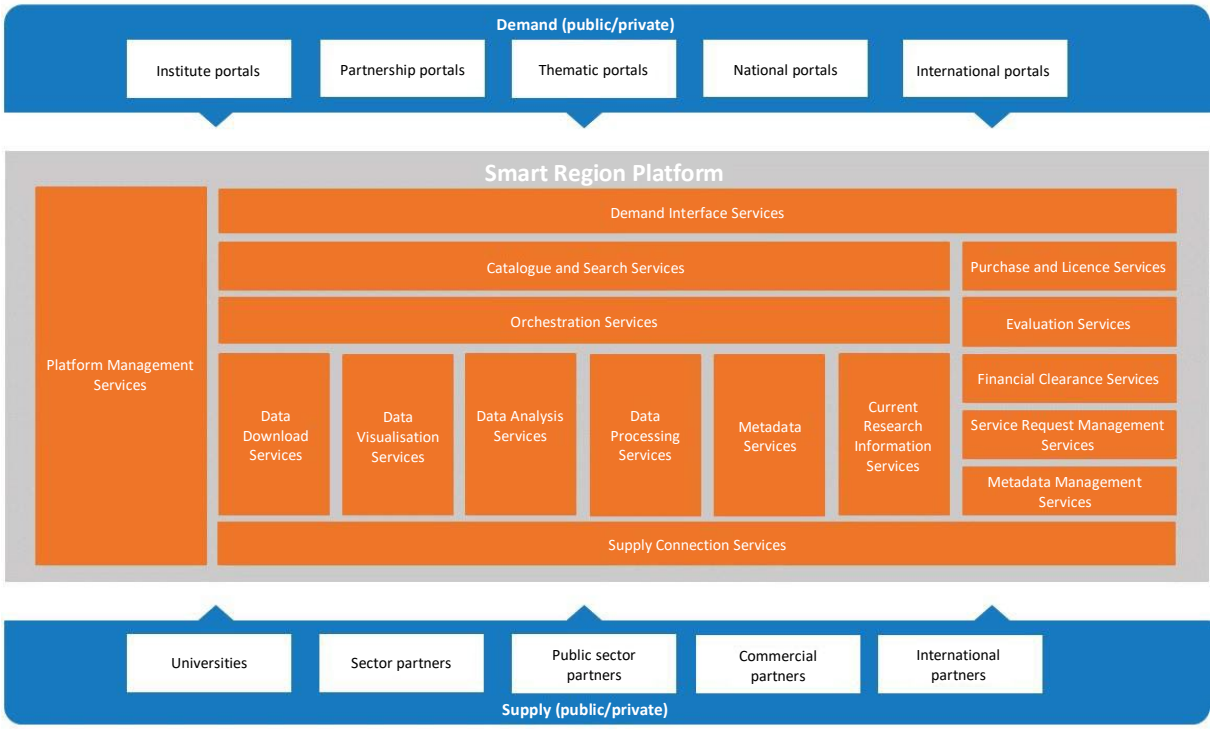


Figure 7: Smart Region platform

Services offered on the platform

Data Download Services

The platform offers services to retrieve or link data made available by other sources. For example, from sources of parties operating and participating on the platform.

Processing, Analysis and Visualisation Services

Both structured and unstructured data are available on the platform. For this, the platform offers services that help employees process this data (*Data Processing Services*). The platform also provides services for analysing data (*Data Analysis Services*) and visualising results (*Data Visualisation Services*). These services are lighter in nature than, for example, those applicable to the Open Research Tools platform, because the data exchanged on the Smart Region platform requires less intensive processing.

Metadata Services

The platform includes services that help providers with metadating their data (*Metadata Services*). The services for validating metadata are one example of this. The addition of metadata makes it possible to search in data and make connections between data. Metadata are added based on specific standards. Governance is set up for this purpose. In addition, a community is used to monitor the use of these standards.

Current Research Information Services
The platform supports the creation of insight into the current research portfolio by enabling links with the current research information systems of institutions. The platform offers possibilities to search this information. This creates a consistent picture of the themes that are already being researched. Independent national developments can be shown on this basis.

Functions present on the platform
The user environment focuses on professional employees in roles such as policy-makers or data scientists at institutions, sector partners or authorities. Users can find data and services using a catalogue and search function (*Catalogue and Search*). Good filters and search algorithms help users find the relevant data. If the platform becomes successful, the amount of available data and connected services will increase. In that case, good filtering mechanisms and search algorithms are essential. The platform also offers an environment where users can purchase the offered data and services (*Purchase and Licensing Services*). Financial settlement of any orders for the use of data and services is also supported (*Financial Clearance Services*). The user environment supports various devices and is intuitive, accessible and easy to use.

Service and Evaluation
All services offered on the platform are provided with transparent service levels. Users are supported in their use of the data or services (*Service Request Management Services*) and their questions are routed to the relevant provider. Users of data and services also have the opportunity to assess the obtained data and services (*Evaluation Services*). The evaluation is shown for the relevant service or dataset the employee wants to purchase. Providers of research data and services have the opportunity to respond to the evaluation feedback.

Metadata Management Services
The platform helps providers of data and services to document relevant information about the data or services offered. These metadata are used in the catalogue and search function.

Supply Connection Services
There is a linking component for linking services and making research data available. This component helps providers to offer data and services via the platform in a simple, standardised way . The component also enables institutions to offer their own facilities as a service on the platform. This functionality promotes collaboration, reuse, differentiated services and so on.

Orchestration Services
Orchestration functions ensure the connection of individual components of the platform. These functions ensure the platform traffic is properly managed.

Platform Management Services
The Smart Region platform offers opportunities to assure and meet agreements, for example with regard to the quality of the provided services and datasets. It must be possible to identify factors such as unwanted content or non-performing partners. This requires overarching agreements. Agreements must also be made about the role played by (data from) the platform in an institutional accreditation (including commercial parties). This is handled by platform management (*Platform Management Services*), which also looks after the platform's setup, (continued) development and implementation in collaboration with the key stakeholders.

Relationship with other platforms

The Smart Region platform for research is similar to the Smart Region platform for education. On this platform, data can be used to assess what knowledge needs to be developed and spread within a region. The platforms are similar in terms of the underlying technologies. The reuse of building blocks is obvious. The Smart Region platform will focus mainly on sharing, finding, obtaining, processing, analysing and visualising trend data.

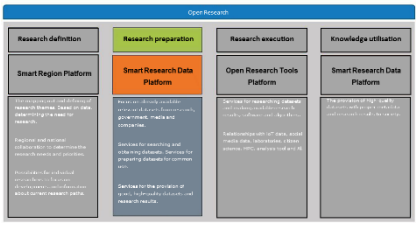
Initiatives already apparent

Several initiatives that can be used as an example for part of the future platform are already apparent. They are:

- euroCRIS
- Trends.google.nl
- Trenddata.com
- Data.overheid.nl

10.2 Research preparation

What happens in this process?



The 'Research preparation' process is for research professionals who are preparing specific research. Research themes are defined at regional, national or international level based on the 'Research definition' process step. During this process, the research is prepared based on these themes, which often results in an approved research proposal. This involves collaboration at a regional, national and international level. Therefore it is important to ensure this process can be included in European collaboration on research and research support.

Researchers develop activities to obtain all available research data for the research to be carried out, for which they are looking for high-quality data. Researchers will get the necessary support to find and acquire the required research data, including any financial transactions for the data. This process involves parties with a demand for research data as well as parties who offer research data.

Institutions collaborate with many parties for research. They value collaboration and actively attract parties to their campuses to connect. Such parties sometimes deliberately choose their location on or near a campus. An important part of the collaboration is the sharing of assets such as laboratories, facilities, and above all research data.

The higher education sector strives towards open access to research data and publications. This will not always be the case for the parties the institutions are collaborating with. As a result, the research data from these parties will not always be openly accessible. One example of this is research data from parties that are available for doing scientific research, but not for anything else. In this process, strict agreements are made on the use and protection of research data.

What is the role of the platform in this process?

The Smart Research Data platform supports research preparation in terms of the research data that is necessary to carry out the research. Parties operating on the platform include governments, companies, hospitals, media and the institutions themselves, as well as other private and public parties.

What components does the platform consist of?

The Smart Research Data platform is designed to promote the exchange of all kinds of high-quality research data. The platform supports employees in their roles as researchers or research supporters at institutions, sector partners or governments. The platform's environment enables users to find and collect research data from participating parties made available by other sources, and allows them to publish (intermediate) research results. The platform can handle supply and demand for both structured and unstructured data.

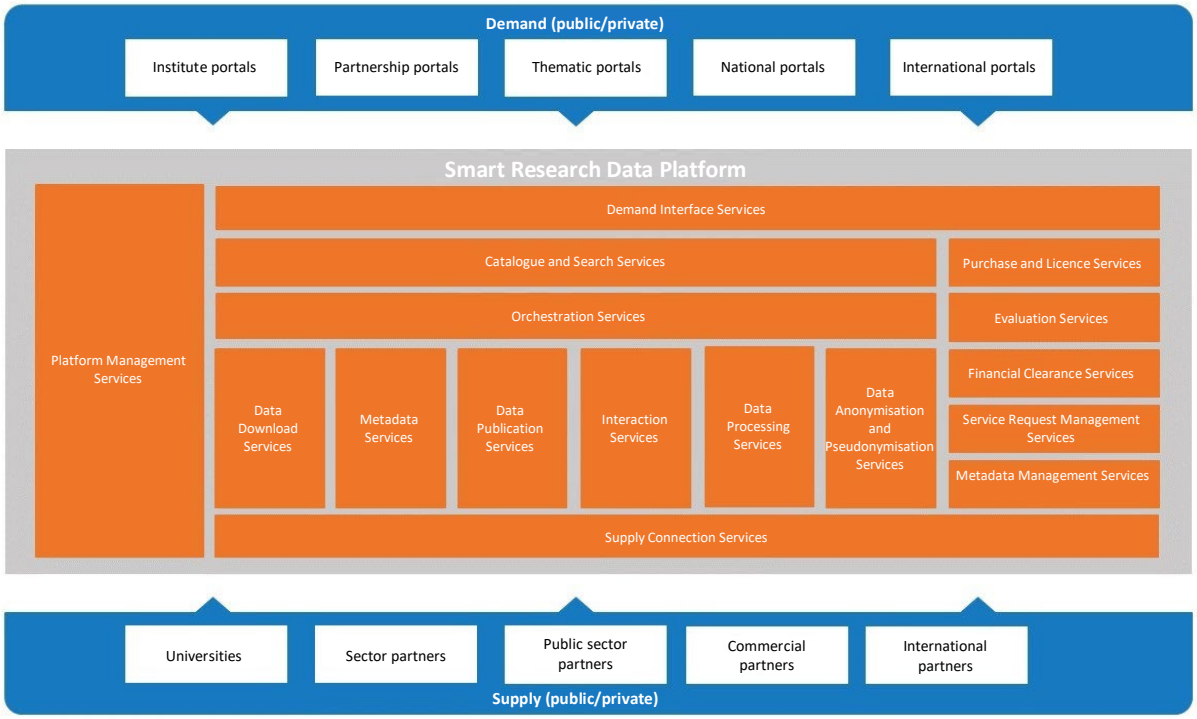


Figure 8: Smart Research Data platform

Services offered on the platform

Data Download and Publication

The platform offers services to retrieve or link data made available by other sources. For example, from sources of parties also operating and participating on the platform. To make research data available, for example in the form of research results, specific services to support data publication are offered (*Data Publication Services*).

Data Processing, Anonymisation and Pseudonymisation Services

Both structured and unstructured data are available on the platform. For this, the platform offers services that help employees process this data into usable data (*Data Processing Services*). Sometimes data needs to be anonymised or pseudonymised before it can be offered via the platform. In both cases, the platform offers services to support the provider (*Data Anonymisation and Pseudonymisation Services*).

Metadata Services

Like the Smart Region platform, the Smart Research Data platform offers services that supports providers with the meta dating their data. Here too, standardisation is an important aspect of metadata. This requires a community with an accompanying governance setup.

Interaction Services

Finally, this platform offers services for the interaction between providers and data users. This interaction focuses on using and achieving the right data quality. The proper design of this interaction is essential for the platform's operation.

Functions present on the platform

Like the Smart Region platform, the Smart Research Data platform includes features for creating the user environment (*Demand Interface Services*) and for searching and finding services and data (*Catalogue and Search Manager*). A service environment (*Service Request Management Services*) is available to support users of services and data. Users can assess services and data with the aid of evaluation functionalities (*Evaluation Services*).

Providers of research data and research services get support to document relevant information and quality attributes of the data or service. Some examples are the degree of robustness and security, the application area and the guaranteed lifetime of the service. These metadata are applied and displayed in the *Catalogue and Search Services.* Providers and users get support for their procurement and licence processing (*Purchase and Licence Services*) and for the financial settlement associated with consuming and using data and services *(Financial Clearance Service*).

Connection and Orchestration
There is a linking component for linking services and making research data available. This component makes it easy to establish links in a standardised way. This allows institutions to offer their own facilities as a service on the platform. This functionality promotes collaboration, reuse, differentiated services and so on. Orchestration functions ensure the connection of individual components of the platform. These functions ensure the platform traffic is properly managed.

Platform Management Services
The Smart Research Data platform offers opportunities to assure and meet agreements, for example with regard to the quality of the provided services and datasets. It must be possible to identify factors such as unwanted content or non-performing partners. This requires overarching agreements. Agreements must also be made about the role played by (data from) the platform in an institutional accreditation (including commercial parties). This is handled by platform management (*Platform Management Services*), which also looks after the platform's setup, (continued) development and implementation in collaboration with the key stakeholders.

Relationship with other platforms

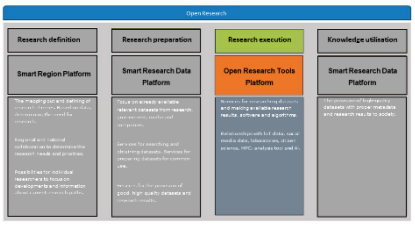
The Smart Research Data platform is related to the Smart Region platform when it comes to searching for, obtaining and sharing research data. The platforms have similar underlying technologies in certain areas. The reuse of building blocks is obvious. The Smart Research Data platform will focus mainly on offering and publishing research data.

Initiatives already apparent
Several initiatives that can be used as an example for part of the future platform are already apparent. They are:

- Re3data.org
- DataCite Metadata Search
- Biological Magnetic Resonance Database
- Datadash
- Research Drive
- Deltares BlueEarth Data

10.3 Research execution

What happens in this process?



Once the research preparations have been completed during the 'Research preparation' process (usually resulting in an approved research proposal), the research can be started. During the 'Research execution' process, the researcher will actually start working on the research. The process is aimed at the researcher who is actually doing the research. Because research can take place at different levels, it is important that this process is in line with the regional, national and European collaborations in terms of research and research support.

The researcher develops activities to analyse all available research data (see 'Research preparation' process), with the support of various research facilities and algorithms. In addition, the researcher can find qualitative research data to further develop the research. This can be both static and near-real-time data. The researcher will get the necessary support to find and acquire the required facilities, algorithms and research data, including any financial transactions for this.

This process involves parties that need or offer research facilities, algorithms and data for a research topic. Some obvious parties are researchers (research consortia), institutions, SURF, as well as private parties.

What is the role of the platform in this process?

The role of the Open Research Tools platform is to help researchers examine data during the actual research process. This platform allows researchers to use computing capacity and real-time data from IoT networks, smart cities, social media and citizen science networks to analyse data. Making these data and services generically available unleashes new possibilities for research. This allows the sector to make an important contribution to the development of smart cities. The research data are not on the platform itself.

Institutions develop research software, research algorithms and other tools that can also be applied more widely. The platform supports the exchange of these tools.

What components does the platform consist of?

The Open Research Tools platform is designed to support research. It supports employees in their role as researchers. The platform offers an environment where researchers can use various resources to examine research data, which may lead to research results. Many of these research data are not on the platform itself. Instead, the platform offers opportunities to obtain data from participating parties. The platform also provides services that help researchers to examine the research data.

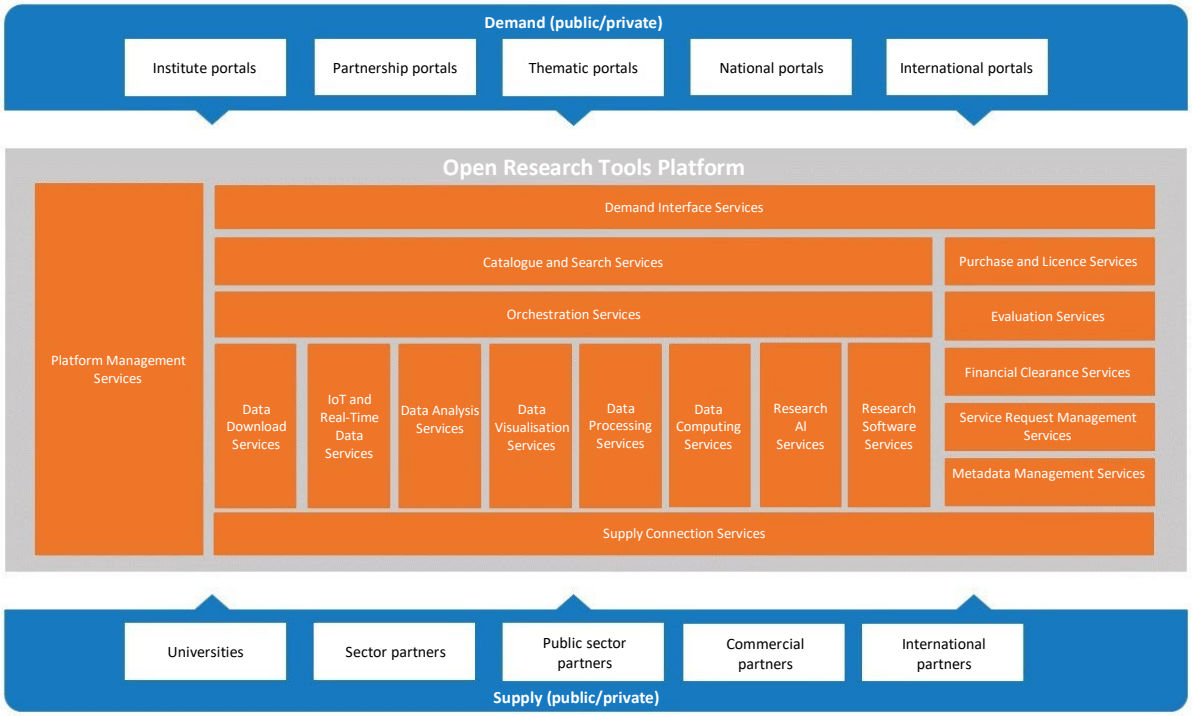


Figure 9: Open Research Tools platform

Services offered on the platform

Data Download, IoT and Real-Time Services

The platform offers services to retrieve or link data that is made available by other sources (*Data Download Services*). For example, from sources of parties also operating and participating on the platform. Services are also offered to enable links to IoT and real-time data flows (*IoT and Real-time Data Services*).

Processing, Analysis and Visualisation Services

For the processing of data, the platform offers services (Data Processing Services) that support the employee in this. The platform also offers services for analysing data (*Data Analysis Services*). Employees also receive support for the presentation of analysis results (*Data Visualisation Services*). Unlike the services of the same name on the Smart Region platform, these services are suitable for more powerful analyses, processing and visualisation.

Data Computing Services

The platform offers or brings together processing power services for very complex and difficult analyses of research data. The most suitable computing service is decided (preferably automatically) based on certain criteria. Some examples are grid computing, high performance computing and generic cloud computing. These computing services are made available by institutions or by public/private partners. The platform manages the computing services.

Research AI Services

Artificial intelligence will be used more and more in society, with research leading the way. Algorithms that are created in research act as a basis for artificial intelligence. They are important parts of research results and the need to reuse them will arise. The Open Research Tools platform offers services that allow researchers to make available successful algorithms to fellow researchers and partners.

Research Software Services
Services for sharing specific research software are offered in the same way as the artificial intelligence services. The researcher has the opportunity to find and obtain this specific type of software on the platform. The platform also support the researcher to make such software available.

Functions present on the platform
The aforementioned generic functions are also available on this platform to support the interaction between providers and users. The *Catalogue and Search Services* support searching, and the *Purchase and Licence Services* and *Financial Clearance Services* support procurement and financial settlement for procurement and use. *Service Request Management Services* support any questions on the use of the platform's data and services.

Providers of research data and research services get support to document relevant information and quality attributes of the data or service (*Metadata Management Services*). Thanks to the nature of this platform, the quality attributes include aspects like robustness and security and the data or service's scope and duration, as well as attributes like espionage protection, data write speed, storage bandwidth, processing power in the vicinity of data storage, and backup regime. These metadata are applied and shown with the *Catalogue and Search Manager.*

The linking function (*Supply Connection Services*) makes it easy to establish links in a standardised way. This function allows institutions to offer their own facilities as a service on the platform. This functionality promotes collaboration, reuse, differentiated services and so on. Orchestration functions ensure the connection of individual components of the platform. These functions ensure the platform traffic is properly managed.

All functionalities available via the platform are quickly and easily accessible from the (user) interface (*Demand Interface Services*). This interface supports various devices.

The Open Research Tools platform offers opportunities to assure and meet agreements, for example with regard to the quality of the provided services and datasets. It must be possible to identify factors such as unwanted content or non-performing partners. This requires overarching agreements. Agreements must also be made about the role played by (data from) the platform in an institutional accreditation (including commercial parties). This is handled by platform management (*Platform Management Services*), which also looks after the platform's setup, (continued) development and implementation in collaboration with the key stakeholders.

Relationship with other platforms

The Open Research Data platform is related to the Smart Research Data platform and Smart Region platform. The platforms have similar underlying technologies in certain areas. The reuse of building blocks is obvious.

The management of the research access is a prerequisite for the Open Research Tools platform. The researcher determines to what extent their research has to be open and gives access to different target groups. This is largely determined by agreements made in advance with the person funding the research, for example. The services must provide sufficient support to the researcher in this respect.

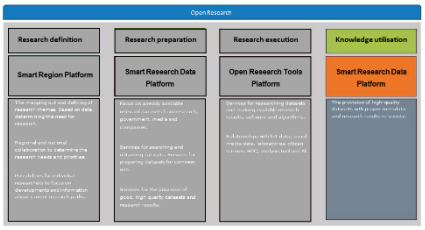
Initiatives already apparent

Several initiatives that can be used as an example for part of the future platform are already apparent. They are:

- DataverseNL
- EASY
- Cartesius National Supercomputer
- HPC Cloud
- Jupyter Notebook

10.4 Knowledge utilisation

What happens in this process?



The 'Knowledge utilisation' process focuses on making available research results (publications, research data, interim results) to a wider audience. The researcher wants to share the results of the 'Research execution' process with fellow researchers for peer evaluation or knowledge transfer. The researcher also wants to make available (interim) results to a wider audience in this process.

In this process, the researcher can focus on new data that are relevant to their own research, as well as any interim or final results from previous research. The same platform is used as for the 'Research preparation' process: the Smart Research Data platform.

What is the role of the platform in this process?

The Smart Research Data platform helps users to search, obtain and publish research data in the form of research publications or interim results. Parties operating on the platform include institutions, governments, companies, hospitals and media as well as other private and public parties.

What components does the platform consist of?

For a description of the components of this platform, see Section 10.2.

Relationship with other platforms

The Smart Research Data platform is related to the Education Exchange platform for more flexible education. The research results are used in the development of education and other areas. The data management platforms also have similar underlying technologies in certain areas. The reuse of building blocks is obvious.

Initiatives already apparent

In the current situation, several initiatives that can be used as an example for part of the future platform are already apparent. They are:

- TU Delft Repository
- WUR Science Shop
- Higher professional education knowledge base (HBO Kennisbank)
- VU Research Portal
- Google Scholar

11 Principles

The strong connection between the various platforms is characteristic of the Research Data Management (RDM) domain architecture. This means a large number of processes and applications are fed information from various components within and outside the higher education sector. The principles below are specific to the RDM domain and partly differentiate the principles defined at the general HOSA level. Attention has been given to the manageability of the platforms, as that is where the greatest risk of failure lies in terms of platform reliability. The architectural principles provide guidelines for setting up the desired architecture. They serve as a tool for and justification of setup decisions. The principles support a consistent setup.

Principle	Rationale	Consequences
RDM – 001 – Fostering collaboration	• In order to make research future-proof, national and international collaboration is essential. • Collaboration is necessary to reduce the time to market of research.	• The services offered and obtained via the platforms are easy accessible and easy to use. • The services offered and obtained via the platforms support a multilingual approach. • The platforms support cross-border collaboration. • The platforms provide insight into the supply of and demand for research data and research data services.
RDM – 002 – Supporting diversity and variety	• Good research requires a diverse range of qualitative research data is necessary. • Research modularity will lead to diversity and variety of research and research participation.	• Platforms support the exchange of various forms of data. • Platforms have a flexible setup and are easy to adapt. • The platforms and services offered and obtained via the platforms support modern and preferably open (accessibility) standards and exchange protocols. • The platforms offer space to various providers of data and data management services. Some examples of these are SURF, institutions, private providers, and innovative start-ups.
RDM – 003 – Open and transparent	• Open research reduces the time-to-market of research. • Making research data openly available will contribute to a more efficient and effective research process. • Making research data (and results) openly available will increase knowledge development and the application of research results.	• The research data offered on the platforms have a unique identifier. • Research data provided on the platforms are given metadata according to the FAIR principles. • Higher education institutions offer their research data and research data services via the platforms. • The platforms provide insight into the availability and status of the services offered. • Research data and services offered on the platforms have been given clear licence structures and service and contract agreements.
RDM – 004 – Insight into quality	• Insight into the quality, reliability, integrity and confidentiality of the research data and services will further increase the quality and recognisability of the research results.	• Research data and services offered on the platform are provided with up-to-date metadata, including quality attributes such as reliability, integrity and confidentiality. • Management of the research data is based on a data life cycle.

RDM – 005 – Unburdening use	• Accessible research starts with the accessibility of research data and services. Reducing the workload of the user makes research more accessible.	• The providers of the research data and services are responsible for ensuring the metadata on the platform are up to date. • The providers give the user support in accordance with the agreed service levels. • The research data and services are easy to obtain and use.

12 Appendix

12.1 Consulted persons and bodies

RDM Working Group	HOSA Steering Committee
- Wilmar de Lange (Rotterdam University of Applied Sciences) - Maarten Hoogerwerf (Utrecht University) - Louis Stevens (SURF) - Rudy Dokter (Saxion University of Applied Sciences) - Mark Cole (SURF) - Els Velraeds (Fontys University of Applied Sciences) - Taco Horsley (University of Amsterdam) - Joachim Rijsdam (Leiden University)	- Anton Opperman (EUR) - Rose of Iperenburg (HAN) - John Kropman (Fontys) - Jan-Willem Brock (Leiden University) - Hans Louwhoff (SURF) - Rene Schenk (Avans)
Conversations	**Review**
- Wendy Sahier (SURF) - John Doove (SURF) - John Meurs (SURF) - Bram Gakeer / Juriaan van Kan (Ministry of Education, Culture and Science) - Pim Jorg, Melanie Lemmen, Stephan van Duren, Danny ten Bosch (NWO) - Maurice Bouwhuis, Magchiel Bijsterbosch (SURF) - Jeroen Belien (VUMC) - Ronald Mens (TNO) - Willem Koeman (AmDex) - David Groep (Nikhef) - Hylke Koers (SURF) - Wim Hugo (DANS) - Menno Nonhebel (KNAW) - DCC HBO ICT Working Group - DCC Higher academic education project "De verbinding tussen Research Cloud en iRods en Yoda" - DCC Higher academic education project “Koppelen DCC services aan de Research Cloud” - Mark van de Sanden (SURF) - TNC’21 (GEANT) - VU Programme Team - SURF webinar “Platformen hebben de toekomst” - RDM Steering Group status on HOSA - RDS programme - National network of information managers - NLR (Erik Baalbergen)	- Wilmar de Lange (Rotterdam University of Applied Sciences) - Maarten Hoogerwerf (Utrecht University) - Louis Stevens (SURF) - Rudy Dokter (Saxion University of Applied Sciences) - Mark Cole (SURF) - Joachim Rijsdam (Leiden University) - Els Velraeds (Fontys University of Applied Sciences) - Taco Horsley (University of Amsterdam) - Tim van Neerbos (Utrecht University) - Frank Snels (Twente University) - Higher education architectural council (ArchitectenBeraad HO) - Dick van der Linden (University of Applied Sciences Leiden) - John van den Berge (Eindhoven University of Technology) - Patrick van der Veer (Utrecht University) - John Doove (SURF) - David Groep (Nikhef) - Hylke Koers (SURF) - Menno Nonhebel (KNAW) - Wim Hugo (DANS) - Erik Baalbergen (NLR)

12.2 HOSA-RDM objectives structure