Publication Framework for Open Science and Research	
Publisher The Open Science and Research Initiative (ATT)	**Publication date** 3 February 2016
Author The Open Science and Research Initiative (ATT)	

Distribution A PDF file can be downloaded at http://openscience.fi/framework-for-open-science	
Contact http://openscience.fi avointiede@postit.csc.fi	

Contents

1. Introduction

1.1. Subject

This document describes the framework for the **Open Science and Research Initiative (ATT)** using the enterprise architecture methodology. This target architecture will guide the national principles of openness that govern the data and services used in science and research, the exchange of information, and the development of e-services that support openness.

'Framework for open science' means a general description of the desired architecture (target architecture) in the specified field. The framework configures and defines the key structural elements of the overall solution, without expressing a precise solution, for example, implementation technology or any other details of design and implementation. The framework for open science defines the framework in which shared open science and research solutions will be designed and implemented, including its components, information management, applications, and local services.

This framework for open science gives an overview of how various actors' open science and research process and services – including data, data structures, actors, roles and IT systems – will form an interoperable system in the target state. It seeks to split functions and technological components into logical entities in which only one solution is developed for the one common purpose, and which seamlessly connect to other services and existing implementations both nationally and at other research organisations.

Operating models and solutions that promote open science and research should comply with this target architecture. When developing technologies and operating models for open science and research solutions and subsystems, developers should compare their potential solutions with this framework for open science, and seek to acquire or implement solutions that best fit the specified architectural objectives.

The framework for open science and research will be significantly linked to the enterprise architecture of institutions of higher education and research institutes, and will also be partially connected to the common enterprise architecture of public administration.

1.2. Intended audience

This framework for open science description is aimed at institutions of higher education and national research institutes; organisations and working groups that are developing national and shared models and solutions for open science and research; and their partners.

This document can be used to develop interoperability between open science and research activities and services from different perspectives, such as processes, data, and the exchange of information. This architecture should be applied at organisational level and, in particular, to meet the needs of open science.

The roles in this framework for open science and research are:

- The Open Science and Research Initiative (ATT)
    - The Open Science and Research Initiative's own steering group, core team, and action groups.

- Actors responsible for conducting and developing research at institutions of higher education and research institutes
    - Key persons in research development, those responsible for development projects, and project managers who are working to develop research and its tools in projects whose development activities are directly or indirectly linked to IT environment services
    - Those who develop research processes and leadership processes
    - Supervisors who are responsible for research at institutions of higher education and whose tasks include promoting framework for open science in their units
    - Main users of existing research IT systems at institutions of higher education and other research organisations
    - External third-party and private-sector service providers, partners and subcontractors that provide research support services to institutions of higher education and other research organisations
- Key persons in enterprise architecture and information management at institutions of higher education and research institutes
    - Key persons responsible for information management and data technology solutions, such as information management directors and managers
    - Key persons responsible for enterprise architecture activities – named architects
    - ICT project designers and technical responsible persons
- National steering
    - Key persons in research policy at the Ministry of Education and Culture from the perspective of interoperability and development in open science and research.

In addition to the aforementioned, the target groups for this document include service providers that offer research IT systems, and ICT, consulting, expert or development services to institutions of higher education and research organisations.

1.3. Open science and research

This framework for open science has been developed as part of the Open Science and Research Initiative.

Open science means the use of open operating models in scientific research. The key objective is to publish research outputs, along with the data and methods used, so that they can be examined and used by all interested parties. Open science includes practices such as open access publishing (of both publications and research data), harnessing open-source software and open standards, and the public documentation of research processes with 'memoing'.

Openness is a fundamental principle of science and research. Maximum openness creates new opportunities for participating in science and research, not only for a broad range of researchers but also for decision-makers and the general public. Open science has become a globally significant way to promote both science itself and its social impact. This is due to the opportunities generated by openness: more extensive validation of research findings, transparency, and repeatability. However, for this happen, people need wide-ranging access to the publications, datasets, methods, expertise and support services generated by, and required for, research. This openness should be reflected not only in data collection and management but also in research and evaluation methods. Other people's interest in a study indicates its significance. Scientific impact increases when results are extensively available.

2. Using the enterprise architecture method

'Enterprise architecture (EA)' means modelling, describing and designing operations, data requirements, IT systems and technological solutions in accordance with a common model. Enterprise architecture ensures that requirements in different areas – and in particular operational requirements – are holistically considered in the development of all operations and ICT solutions. In practice, enterprise architecture consists of the organisational configuration, description templates and architectural policies drawn up on their basis.

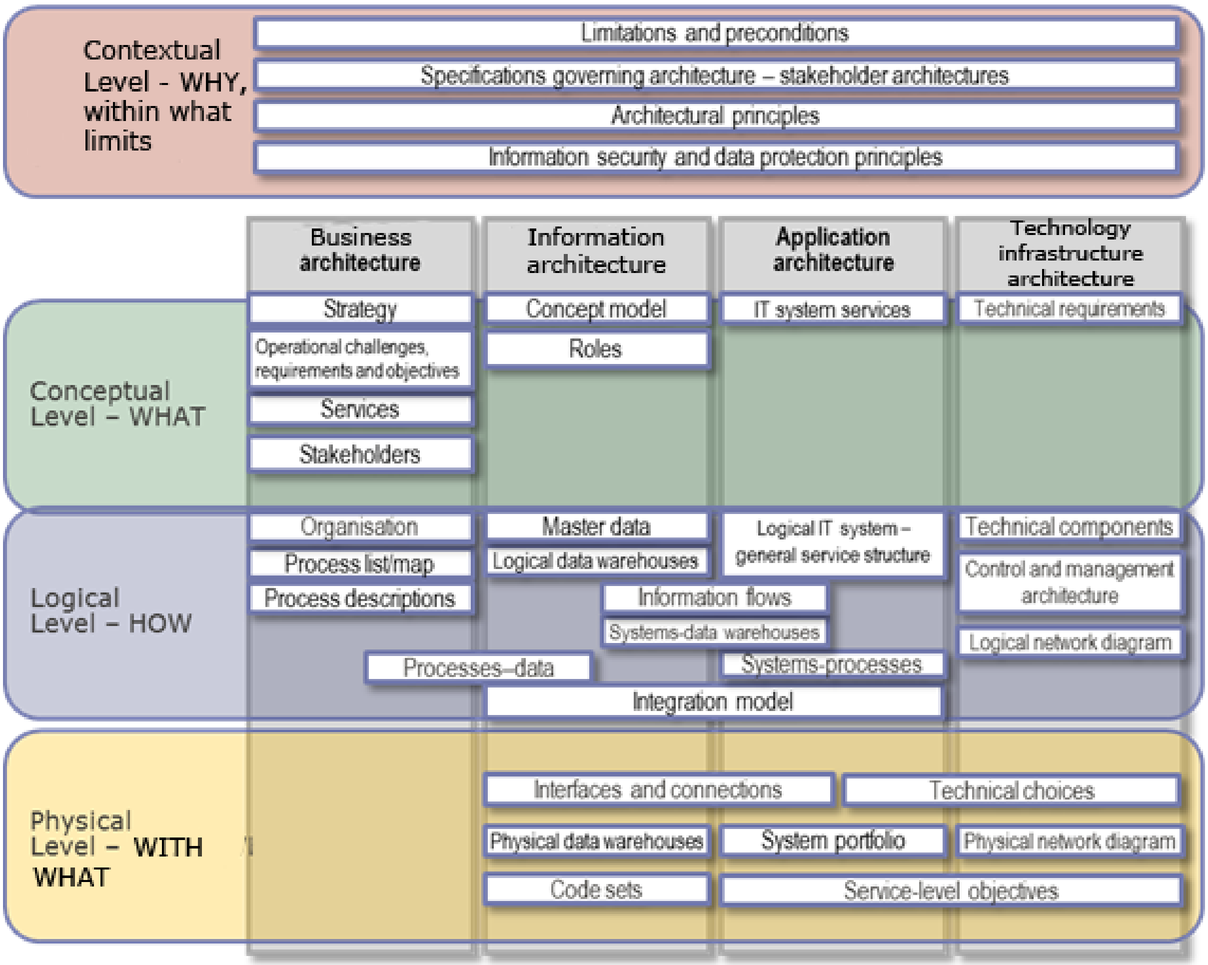
Enterprise architecture extends beyond simply describing IT systems – it also describes the reasons why the system is being implemented, what needs it addresses, and what data is required for the purpose.

In order to ensure that solutions have the correct functionality and scope, the enterprise architecture method is also configured in terms of **architecture domains** and on **conceptual levels** (abstract levels).

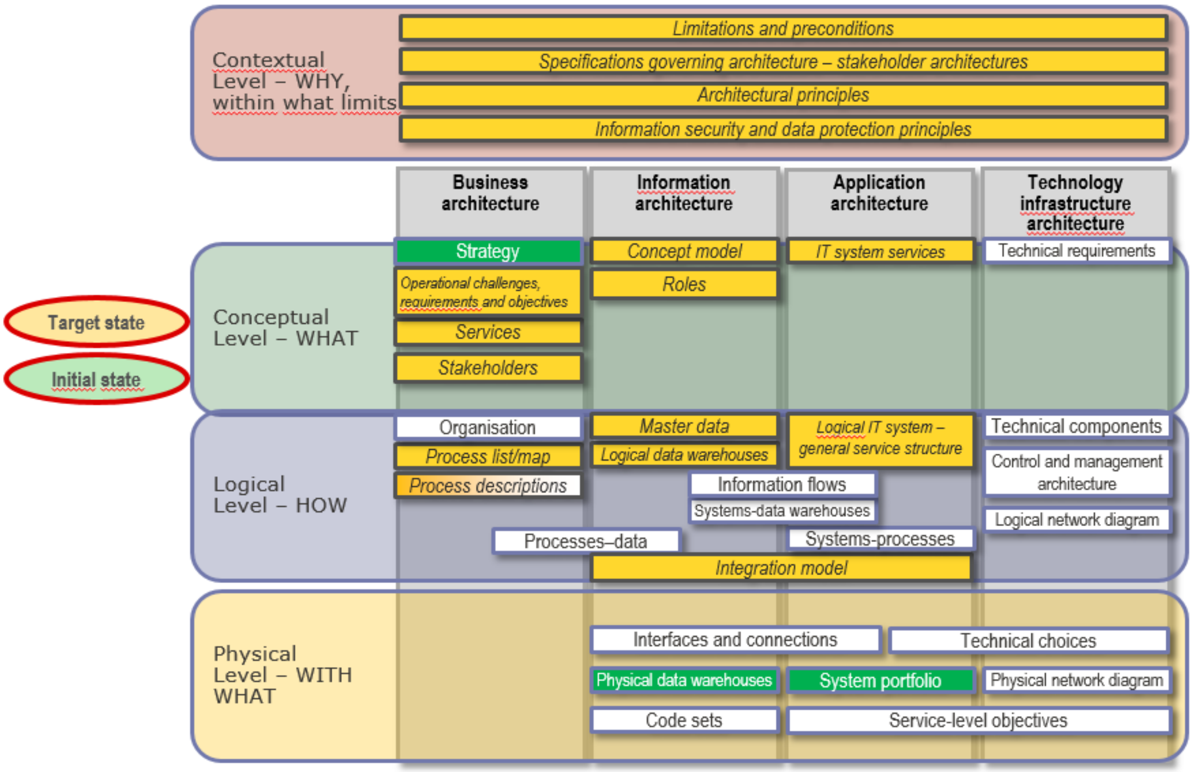
- **Architecture domains:**
  - Business Architecture: *from the perspective of business and customerships*
  - Information Architecture: *from the perspective of concepts, data, and data resources*
  - Application Architecture: *from the perspective of the system*
  - Technology Infrastructure Architecture: *from the perspective of technology, equipment and technical solutions, and their maintenance*
- **Levels of Abstraction (see e.g. TOGAF: Content metamodels):**
  - Contextual level – WHY, within what limits
  - Conceptual level – WHAT
    - *For example, what data is required and for what purpose; what are the key operational concepts?*
  - Logical level – HOW
    - *For example, the logical configuration of data resources and how data should be located within different systems.*
  - Physical level – WITH WHAT

- *For example, which physical databases will store the logical data resources; what will be implemented using files or document management systems?*

The Kartturi EA method used in the higher education sector (described below) has been employed in this work:



In accordance with the Kartturi EA method, architectural work was launched by defining which EA descriptions would be used in this particular framework for open science. As the subject of this enterprise architecture description is the target framework for open science, the focus has been on the sections that describe the objectives of the diagram's upper section, in accordance with the Kartturi model. The following diagram roughly models the sub-descriptions of the enterprise architecture (marked in yellow).



In the main, the current situation has only been identified with regard to general-use IT systems and any key technical IT systems already in existence at the outset.

3. Architectural policies – contextual level

3.1. Limitations and preconditions

The framework for open science for open science and research cannot cover all of the subjects and properties relating to open science and research and its operating environment.

This framework for open science has been limited as follows:

This is included in the RA for OSaR	This may be included with a limited scope	This is not included in the RA for OSaR
• The reference architecture is common to all research units and research financiers • Examine the entire research system and all of its actors • Consider existing architectures or EA principles at an international level • Create a model in which intellectual capital can be utilised for different purposes – make data easily accessible in a usable format • The reference architecture enables the development of research indicators and measures of national impact.	• Increasingly offering sector-specific research tools to researchers (publishers, new companies) – to what extent will this be included? • The Impact Working Group generates input data for science and research impact indicators – the RA for OSaR ensures that these indicators (KPIS) can be collected on the basis of the reference architecture • Research permit management – take sensitive data into account through architecture – possibly generalised	• We don't actively develop reference architectures for data administered by the Ministry of Social Affairs and Health, or architectures for the Ministry of the Environment and other administrative sectors – we connect to them as the solutions require • No support service architectures for research or research administration will be developed within this reference architecture – the TUHA network is developing this area • The reference architecture does not express an opinion on field-specific research methods or data production methods • In this work, no opinion is expressed on who benefits from open research data and how it is used to support social decision-making • This work does not generate any common benchmarks or indicators for research. • In this work, a general-level concept model is drawn up for master data categories – no detailed data models • This work does not express an opinion on the assessment of science

This framework for open science creates a high-level overview of the principles, operations, data, IT services and technology involved in open science and research. In order to implement solutions, additional work will be required to further specify the features described in this framework. The key areas requiring further specification when this framework for open science is updated are:

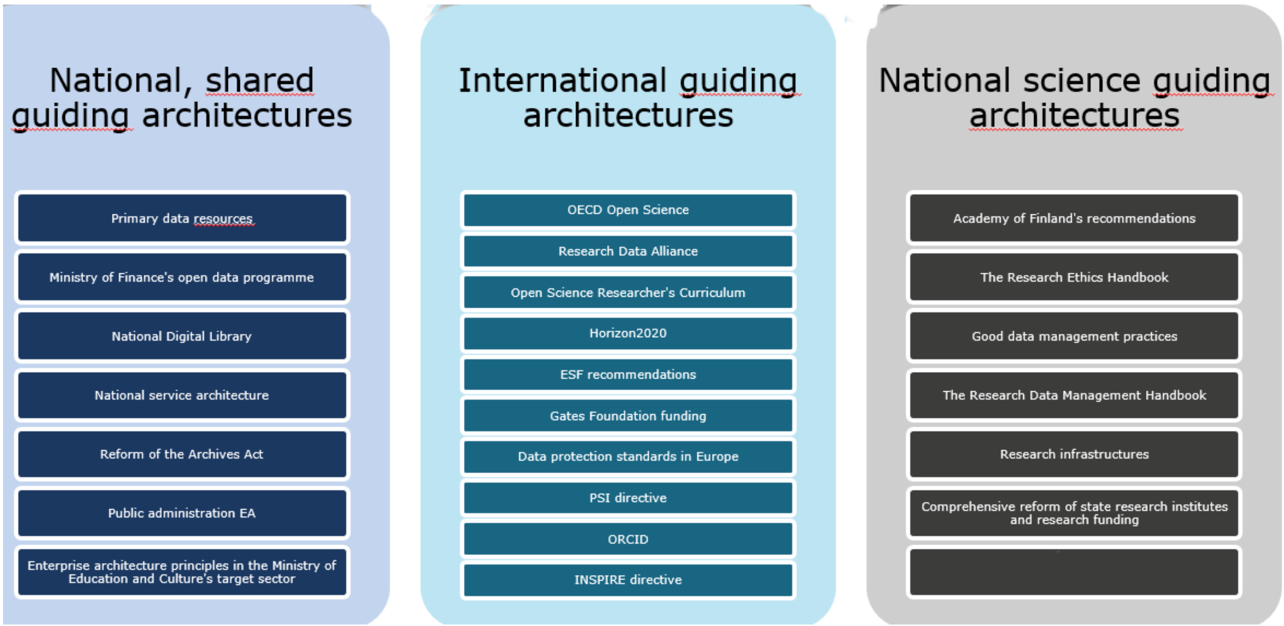
- Core components that further specify concepts and master data
- Data resource structures
- More details on how services and data resources will be divided between centralised and decentralised solutions
- More detailed descriptions of application services
- More detailed descriptions of interfaces
- The design of the integration solution.

Limitations and preconditions have been listed in more detail in *Appendix 1, EA tables.*

3.2. Guiding architectures, projects and solutions

The area to be developed is connected to both legislation and a number of stakeholder solutions, projects and architectures that must be considered during development.

The following diagram shows the key stakeholder projects, solutions and architectures that will affect the framework for open science for open science and research.



The key content of guiding architectures is as follows. The 'Binding force' column indicates whether the specifications for the solution in question should be considered or must be complied with during development:

Guiding architectures	Oblication	Description, key content	Organisation responsible
National, shared guiding architectures			
Primary data resources	Steering	Population Information System, Business Information System, Land Information System, etc.	Population Register Centre, Finnish Patent and Registration Office National Land Survey of Finland
Ministry of Finance's open data programme			Ministry of Finance
National Digital Library	To be considered	The project is developing services and operating models that can be used to create, manage, preserve and utilise cultural heritage materials. Interoperability between systems storing different cultural heritage materials and the compatibility of the data contained within them will also be ensured.	Ministry of Education and Culture
National service architecture	To be considered	The programme is creating the National Data Exchange Layer; the shared service views required by citizens, companies and authorities; a new national electronic identification solution; and national solutions to manage the roles and responsibilities of organisations and natural persons.	Ministry of Finance
Reform of the Archives Act	**Binding**	The project is analysing key reforms relating to the Archives Act and Decree, and is preparing a proposal for the revised legislation.	Ministry of Education and Culture
Public administration EA	**Binding**	Public administration Enterprise Architecture: https://www.avoindata.fi/data/en/dataset/jhka-yhteiset-ict-palvelut-v1-0	
Enterprise architecture principles in the Ministry of Education and Culture's target sector	**Binding**	Enterprise architecture principles that must be complied with in the Ministry of Education and Culture's target sector.	Ministry of Education and Culture
International guiding architectures			
OECD Open Science	To be considered	Recommendations made in open science concepts and principles	OECD
Research Data Alliance	To be considered	Alliance, recommendations of different levels, such as defining permanent identifiers	Research Data Alliance
Open science researcher's Curriculum	To be considered	An extensive set of open science policies	Unesco
Horizon2020	Steering	A financial instrument used to fund research within the EU. Contains policies on, for example, open access publishing and guidelines for openness in research in general.	European Commission
ESF recommendations	To be considered	For example, instructions for using archives	European Science Foundation
Gates Foundation funding	To be considered	The largest funder of science in the world. A very high level of openness is required to receive funding. Extremely targeted in its funding of research	Gates Foundation
Data protection standards in Europe	To be considered	General guidelines on protecting personal privacy in datasets. May also partially restrict data usage	
PSI directive	**Binding**	A directive that governs the re-use of public-sector data	European Parliament

		ORCID	To be considered	An international method for identifying researchers.	
		INSPIRE directive	**Binding**	The Infrastructure for Spatial Information in the European Community	European Parliament
National scientific guiding architectures					
		Academy of Finland's recommendations	**Binding**		
		The Research Ethics Handbook	Steering	The fundamentals of research ethics	Finnish Advisory Board on Research Integrity
		Good data management practices	Steering	Largely a general principle	Section 18 of the Act on the Openness of Government Activities
		The Research Data Management Handbook	Steering	Also contains contractual models and processes for surrendering datasets, etc.	Finnish Social Science Data Archive (FSD)
		Research infrastructures	Steering	The Finnish strategy and roadmap for research infrastructures 2014–2020 defines the important national research infrastructures within the Finnish ecosystem. It also lists Finnish actors' partnerships in the European Strategy Forum on Research Infrastructures (ESFRI); international research infrastructures that Finland has joined through (government) agreements; and local infrastructures identified as important by research organisations.	
		KOTUMO Roadmap	To be considered	Forging closer cooperation with institutions of higher education and research institutes. A key objective of the KOTUMO Roadmap 2015–2017 is to create the best possible conditions in Finland for conducting research and innovation activities, and to strengthen the international visibility and impact of Finnish science through cooperation, collating resources, and clearer task divisions. This will also support Finland's competitiveness, renewal in trade and commerce, and social development.	Ministry of Education and Culture

Guiding architectures and projects are listed in greater detail in <u>*Appendix 1, EA tables.*</u>

3.3. Architectural principles

Several key policies act as cornerstones for the design and implementation of open science and research and its associated services and structures, and also for ensuring continued service provision. These policies have been collated into the following set of architectural principles.

The architectural principles for open science and research are as follows:

- The architecture enables open access to research results and re-use from primary sources.
- The architecture considers the entire lifecycle of research and the data it generates, including its management, so that scientific and research findings are permanent and reliably available.
- The architecture creates services that accelerate and facilitate research efforts and strengthen research at its very core (self-correction, repeatability)
- The reference architecture creates specifications that can be used to identify unambiguous preconditions for open science and research.
- The reference architecture supports a layered, modular and cost-effective solution model that utilises existing components and can be modified to meet the needs of different scientific fields and functions.

These architectural principles emphasises holistic coverage – in the target state, the largest possible percentage of the research process is open and accessible to the scientific community.

Those developing and implementing open science and research activities and IT systems must consider the aforementioned architectural principles in all development activities.

These architectural principles are listed in greater detail in *Appendix 1, EA tables.*

3.4. Information security principles

Openness forms the basis for open science and research. This is to some extent also reflected in information security principles. The overall challenges associated with data protection and information security for solutions and operations primarily relate to user identification, service use, limiting use of the data processed by services, and monitoring data usage. The integrity and indisputably of source data are two information security issues that require particular attention, and they must be properly managed and ensured.

General information security principles have been defined on the basis of eight different security perspectives:

- Administrative security
- Personnel security
- Physical security
- hardware security
- Software security
- Communications security
- Data security
- Operational security

Name	Priority	Description
Data protection		
Personal privacy protection	**Critical**	Data relating to individuals will be protected at all processing stages.
Anonymisation	important	General-use research datasets will be anonymised so that individuals cannot be identified if the data is re-used. Alternatively, people are separated from their data so that it is impossible to identify a person on the basis of the research data alone without access to the population register.
Consent management	**Critical**	User rights for personal information are verified by a comprehensive system of consent management in the use and surrender of data
Administrative security		
Classification and risk assessment	**Critical**	Information is classified according to its sensitivity and an impact assessment of the consequences of a leak
Reporting anomalies in information security	**Critical**	Anomalies in information security are reported and appropriate measures are taken
Assigning roles	important	Assigning roles. Service users have clearly assigned roles (such as owner, responsible leader, member).

Reliable actors	important	Organisations that store important sensitive data or operators that provide agreed services should adhere to best practices in information security – for example, they must meet pre-determined criteria set by certification organisations (ISO 27001 standards, VAHTI guidelines).
Service level agreements	important	Service providers' tasks have been defined in the service agreement.
Personnel security		
Training for service users	important	Users must understand the environment's usage policies and their own role in ensuring information security, and sign agreements to this effect.
Reliable and certified service administrators	important	Administrators must understand the sensitive nature of the data and act accordingly. Administrators sign non-disclosure agreements.
Organisation's responsibility for personnel competence and tasks	important	The organisation has appropriate personnel recruitment and termination processes, and personnel have sufficient competence for their tasks
Dangerous task combinations	**Critical**	Dangerous task combinations must be identified and avoided. In ambiguous cases, the use of two-person teams is recommended.
Physical security		
The service provider must make adequate provisions for physical security	**Critical**	Access to technical and maintenance areas must be appropriately controlled.
Communications security		
Sensitive data cannot be transferred out of a protected environment	**Critical**	Technical solutions are used to prevent the unauthorised transfer of data out of a protected environment. Personal agreements are used to manage residual risk
Encrypted connections and data transfer channels	**Critical**	All communications connections are appropriately encrypted.
Communications and service use logged	important	The service must be adequately logged and analysed. Logs should be treated as sensitive data and their digital preservation must be ensured.
Contacting services	important	The service can only be used from an identified IP address. A user may be granted appropriate access to a protected environment on the basis of a risk evaluation
Real-time detection of information security anomalies	important	Real-time IDS (Intrusion Detection System).
Hardware security		
Data transfer	important	Sensitive materials can be transferred only between trusted environments. The integrity of the trust chain is maintained.
Safeguarding the trust chain	**Critical**	Only contacts from provenly reliable systems will be permitted
Equipment security	important	Organisations must ensure the appropriate and up-to-date information security of their equipment.
Decommissioning equipment	important	Information security must be ensured in the decommissioning of all equipment.
Software security		
Software information security	**Critical**	New software is tested in a test environment before its introduction, and all information security updates are made without delay.
User software	**Critical**	The risks associated with software and virtual machines introduced into the environment by users should be assessed on a case-by-case basis.
System updates	important	Operating systems and software are regularly updated. A responsible organisation has been appointed

Data security		
Continuity and recovery plans	important	Continuity and recovery plans consider the sensitivity of data and seek to protect it under all circumstances.
Recording the surrender of sensitive materials	important	The transferer must keep records of the surrender of any sensitive data, and these records must be preserved in the agreed manner.
Surrendering required data	important	Unless there are restrictions relating to information security or copyright, the data that can be surrendered for research should not needlessly restricted.
Operational security		
Reliable identification	**Critical**	Users are reliably and fully identified in a standardised manner in all areas of the service.
Unique user IDs	**Critical**	All user authorisations are granted on the basis of a user's role. Users are linked to their roles using a personal ID. The ID is also organisation-specific
Virus and malware IDs	**Critical**	The environment's servers are automatically scanned for viruses and malware.

Security regulations for the use of confidential data are described in more detail in the Framework for open science for Using Sensitive Data in Research. Several of the most important principles have been collated below.

Sensitive data means data that is, for example:

- classified no higher than protection level III or IV according to Section 9 of Government Decree 681/2010
- a confidential official document as specified in Section 24 of Government Decree 621/1999
- sensitive data as specified in Section 11 of the Personal Data Act (523/199)
- data relating to samples and their processing as specified in Section 3 of the Biobank Act (688/2012)
- data about associations and foundations collected by Statistics Finland, excluding the public data specified in Section 18 of the Statistics Act (280/2004)

Services that handle sensitive data must ensure that personal privacy, trade secrets and information security are not compromised.

Information security principles and regulations are described in greater detail in Appendix 1, EA tables.

4. Architectural policies – conceptual level

4.1. Strategic policies for open science

This framework for open science supports the key strategies and principles of open science and research that are described in the Open Science and Research Handbook.

Science belongs to everyone. Science can be democratised and its impact increased through the use of new operating models for open science and research. The digitalisation and opening up of the research process creates new opportunities for cooperation and communication, not only for researchers but for everyone who is interested in research. Openness provides potential for change. A key aspect of this change is to ensure that upcoming young researchers working outside research infrastructures have equal opportunities to access data. Openness gives everyone a principled opportunity to engage in research, criticise or affirm research findings, and promote research in accordance with their abilities. This increases people's trust in science and boosts corporate activities, but increasing critical thinking and scientific literacy in society are equally important.

Openness in science has always sought to promote high quality and good practice, and to prevent low-quality and counterfeit research. The self-correcting nature of science requires transparency and repeatability in research. Openness improves the quality of research, as openly available findings and data enable scientific observations to be questioned and verified. In this way, science advances and corrects itself more rapidly and with fewer overlaps. With the aid of open science and research, we seek to promote sustainability, usability, access and trust in research. Ethical openness should be the natural default in research. Research findings and data should be easily discoverable and accessible in the long term as well.

Best practices to promote open science and research consist of common, binding ethical guidelines for research; good methods for managing research data; and basic principles for openness, including recommendations that expand on them. In 2012, the Finnish Advisory Board on Research Integrity (TENK), which was appointed by the Ministry of Education and Culture, worked with the Finnish research community to update its ethical guidelines on good scientific practices and how to handle any suspected violations.

Open science and research is a continuum that covers both working culture and the entire research process. Research data, along with its associated expertise and understanding, is divided among a variety of actors, networks and communities within the research system. In such a system, we can increase openness both internally – by opening up and jointly developing research management processes and operating models – and externally, by opening up the entire process and its parts to other researchers and society as a whole, thereby giving others the chance to use data, methods and findings.

Openness and well-managed research findings lead to high-quality operations and results: honesty, diligence, and accuracy. The same naturally goes for research data, for which the characteristics of high quality also include usability and availability, integrity and accuracy, openness and confidentiality. High-quality data is a central requirement for the good management of research data. When research data is well managed, data is not altered without reason, data is not corrupted or lost during processing, and the accuracy and quality of data can be relied upon. These are absolutes that cannot be compromised without justification. However, measures to ensure data quality and accuracy should always be scaled to the data's significance and importance. The greater the importance of the data, the greater the effort that should go into ensuring quality and accuracy. How data is managed depends on what it is being used for, and thorough documentation and description is important.

In addition to the above, strategic policies that steer science and research have also been described in:

- The Government Programme Finland, a land of solutions - Agreement on the new Government Programme reached. http://valtioneuvosto.fi/documents/10184/1427398/Ratkaisujen+Suomi_EN_YHDISTETTY_netti.pdf/8d2e1a66-e24a-4073-8303-ee3127fbfcac.
    - “More effective use will be made of the resources of science and research. The effectiveness and commercialisation of research findings will be strengthened. The profiles and respective responsibilities of higher education institutions and research institutes will be clarified and cooperation between them will be increased. Knowledge and expertise will be pooled in competitive centres of excellence. Higher education institutions and research institutes will be required to produce a proposal on their respective responsibilities and faculties’ and research units’ closer cooperation. Regional centres of excellence and powerful hubs representing specific fields will be provided with financial support. Account will be taken of the effectiveness and commercialisation of research findings in the steering of public research, development and innovation funding, as well as in incentives channelled towards research institutes and higher education institutions.”
    - “Innovation and service platforms will be promoted in sectors where the public administration plays a role in terms of the functioning of the markets. Such sectors include mobility as a service, health care, learning and the Industrial Internet. The introduction of new technologies, digitalisation and new business concepts will be promoted by legislative means. With the help of open data and the better use of data resources, favourable conditions will be provided for new business ideas.”
- European Research Area and Innovation Committee (ERAC) 2015. European Research Area (ERA) Roadmap 2015–2020. http://era.gv.at/object/document/1845/attach/ERA_Roadmap_st01208_en15.pdf
    - A roadmap for ERA Member States, 2015–2020. The roadmap identifies key implementation priorities that will have the biggest impact on science, research and innovation systems in Member States. It is up to Member States to put the recommendations into practice. Member States' progress and national roadmaps will be monitored.
- OECD 2015. Making Open Science a Reality. https://www.innovationpolicyplatform.org/sites/default/files/DSTI-STP-TIP%282014%299-REV2_0_0_0_0.pdf
    - Presents an overview of open science and the OECD's principles, recommendations and best practices, which are based on country reports and extensive expert analyses (in which Finland has also been involved). The OECD's recommendations are highly recommended but not binding.
- FIRI Expert Group 2013. Finland’s Strategy and Roadmap for Research Infrastructures 2014–2020. http://www.aka.fi/globalassets/awanhat/documents/firi/tutkimusinfrastruktuurien_strategia_ja_tiekartta_2014_en.pdf
    - Guidelines for the Finnish national research infrastructure ecosystem and its connections to the rest of the world. To be implemented using

FIRI funding. The roadmap seeks national and international collaboration, to boost the shared use of infrastructures, and to activate joint investment.

- ESFRI 2011. Strategy Report on Research Infrastructures – Roadmap 2010. https://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf
    - A document that steers European cutting-edge research infrastructure and its funding. Lists both the Commission's tasks and recommendations for Member States.
- European Commission 2012. A Reinforced European Research Area Partnership for Excellence and Growth (COM(2012) 392). http://ec.europa.eu/euraxess/pdf/research_policies/era-communication_en.pdf
    - One of three papers published in 2012 outlining the Commission's aims with regard to open science. The idea is to fully leverage investments in research and create structures that facilitate researcher mobility. Gender equality is also highlighted.
- European Commission 2012. Commission recommendation on access to and preservation of scientific information (C(2012) 4890). https://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf
    - The Commission's recommendations to Member States on how to promote open access to publications and data. Includes principles for reporting on the follow-up to the recommendation.
- Towards better access to scientific information: boosting the benefits of public investments in research, COM(2012) 401 final – July 2012. https://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf
    - The European Commission's vision on open science. This document recognises that the playing field is changing, identifies measures that have already been taken, and lists the next steps in achieving the vision.
- United Nations 1947. Universal Declaration of Human Rights http://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=eng
    - Article 27 (1) "Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits."

4.2. The objectives of developing open science and research

The main reasons for drawing up and using this framework for open science are:

- The Open Science and Research Initiative's objective is to define the target state of enterprise architecture in the field of scientific research, and to expand on and further specify the enterprise architecture descriptions produced in the project. These descriptions will act as steering input for other architecture and enterprise architecture projects in the field of science and research

- The project will specify a framework that can be used in actual research activities in all research organisations
- The Open Science and Research Initiative's enterprise architecture work describes the mechanisms that will enable the introduction of an open operating model into science and research on a practical level. This operating model should contain a variety of premises for openness; openness-promoting contract models and processes; methods for the sustainable management of a range of data resources and dataflows; and operating models that meet the requirements for digital preservation to serve open research and science.
- The key point is that architectural steering will enable the creation of mechanisms that will promote an open operating model whilst also preventing the closing up and silo mentality of science and research.

The framework for open science seeks to create a configured functional/technical overview of open research:

- Commonly agreed concepts to facilitate discussion
- To facilitate communications
- To facilitate discussion about roles and responsibilities
- To facilitate easy use of services and data resources
- To agree on measures.

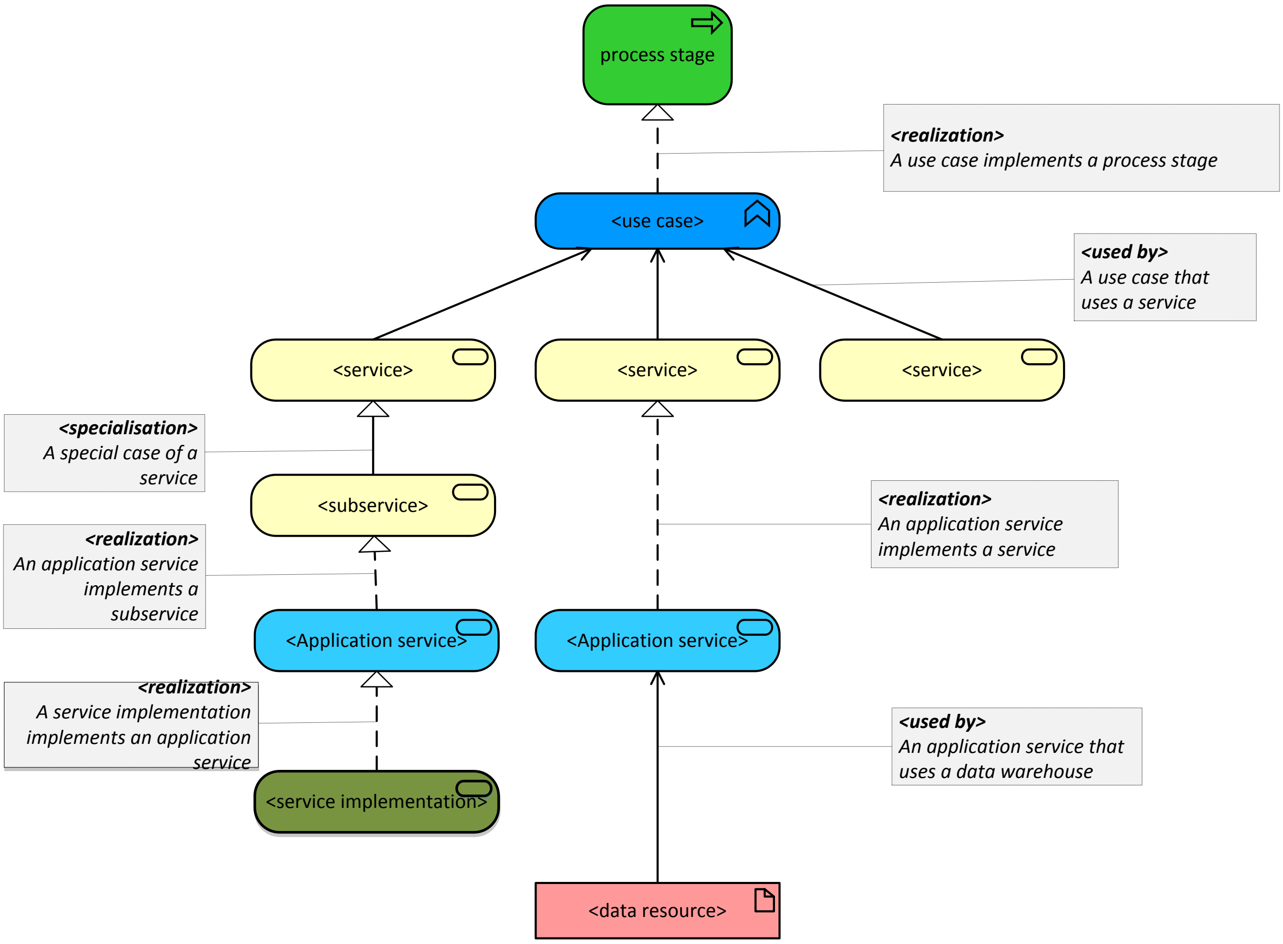
This framework for open science constitutes the target state that enables opening every stage of the research process and accelerating scientific activities, while supporting a new, agile and collaborative way of doing science and research

The key strategic policies in the field, along with development requirements and objectives, are listed in more detail in *Appendix 1, EA tables.*

4.3. Business services

Open science and research is supported by a wide spectrum of services. These services are roughly grouped under the key stages of the main research process, using Archimate notation. The main science and research process is described in more detail in Section 5.1.

The model uses the following symbols:



A service is a functional business entity that fulfils a particular requirement.

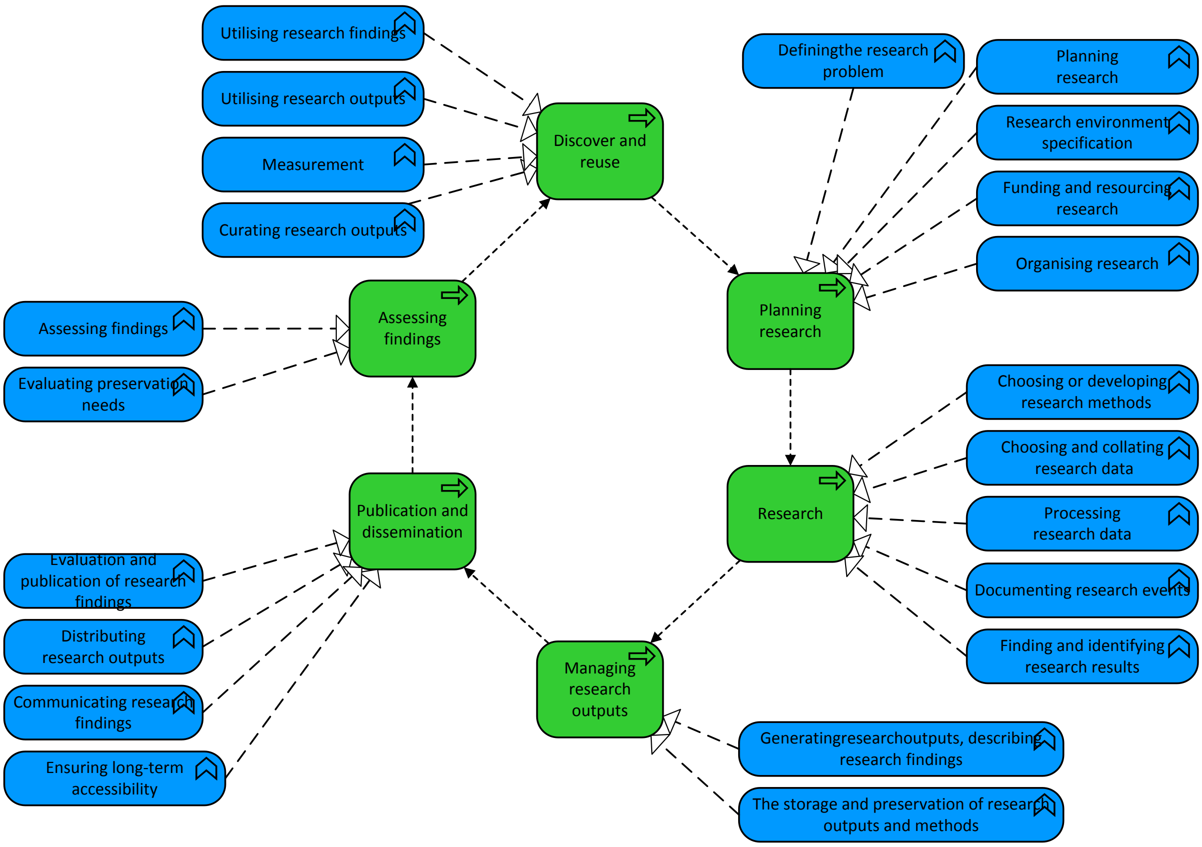
A subservice is a special case of a service.

An application service is a logical information system service that is implemented by an application.

A service implementation is a service realization (application etc.) that implements a logical IT service.

A data resource is a significant dataset that may be either centralised or decentralised.

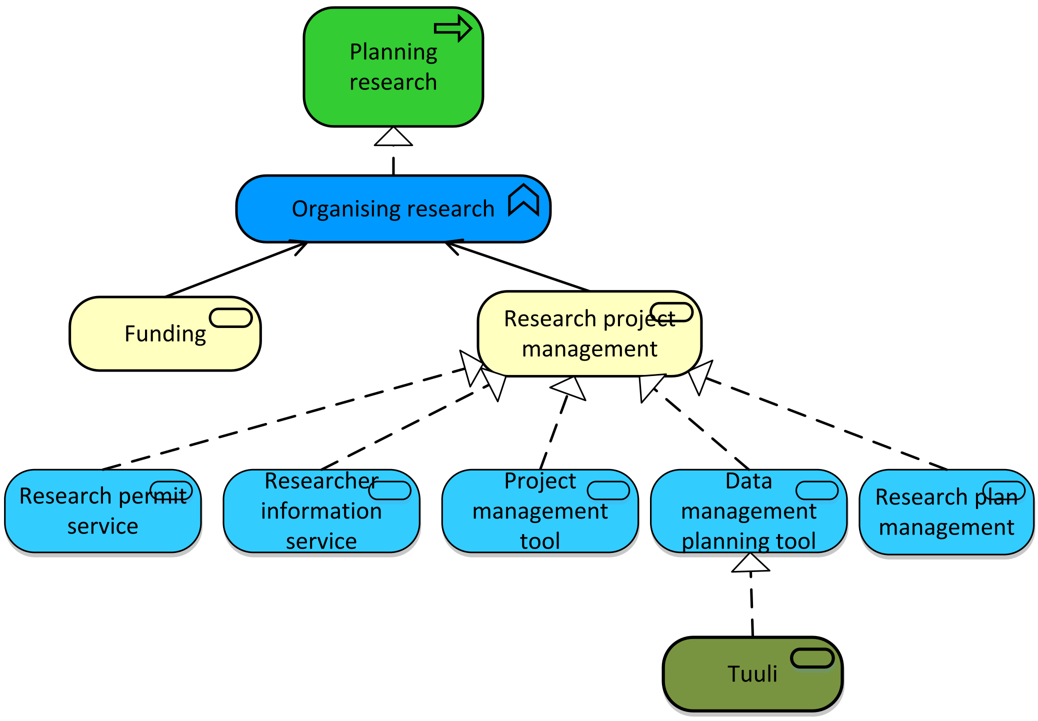
Highest-level services are grouped under the following use cases or service packages:



Open science and research can be supported by a broad spectrum of services at all stages of the process. The green rectangles depict the main stages of the research process, which will be described in more detail later in this document.

The services and their supporting IT systems are described in more detail in Appendix 3.

The following diagram shows an example of the services and IT systems that support a specific use case.

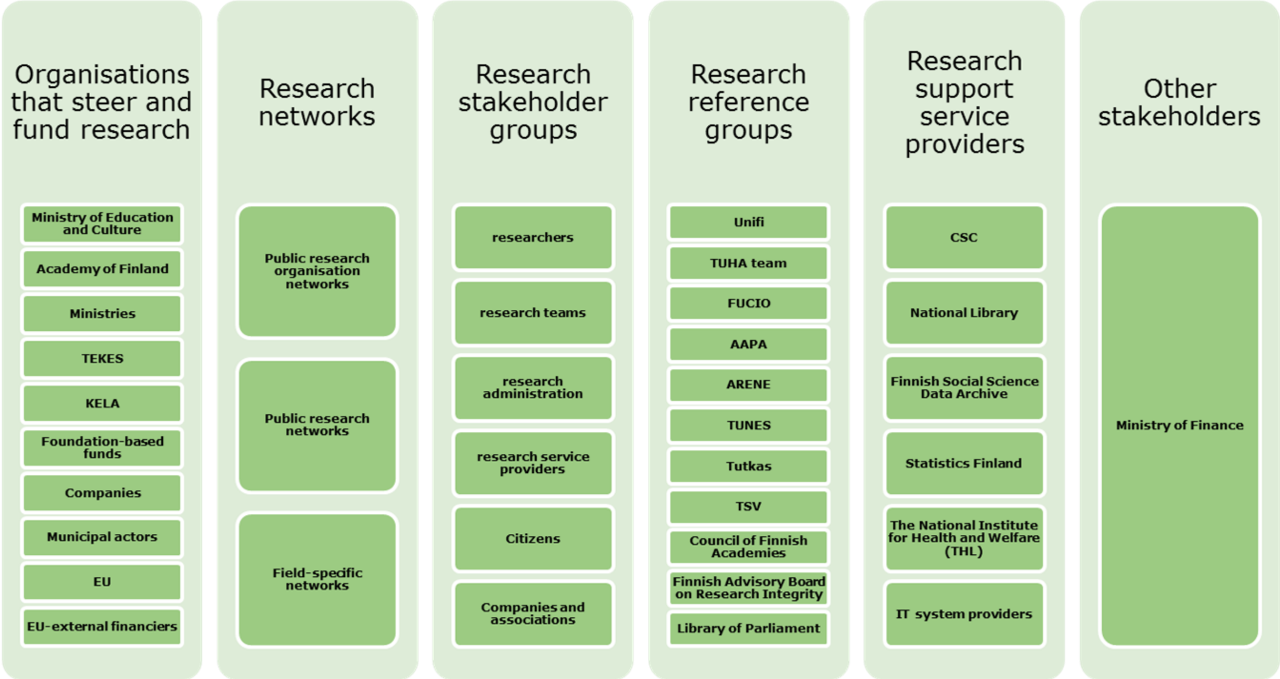


4.4. Stakeholders

The key stakeholders and actors involved in the processes, activities and services of open science and research are naturally research organisations – and institutions of higher education and research institutes in particular:

Universities	**Universities of applied sciences**	**University hospitals**	**Finnish Food Safety Authority**	**Geological Survey of Finland**	**Finnish Meteorological Institute**
Natural Resources Institute Finland	**National Land Survey of Finland**	**Finnish Environment Institute**	**Radiation and Nuclear Safety Authority**	**VTT Technical Research Centre of Finland Ltd**	**The National Institute for Health and Welfare (THL)**
Finnish Institute of Occupational Health	**Finnish Institute of International Affairs**	**VATT Institute for Economic Research**	**University libraries**	**National library**	**National Archives**
Municipality-owned research institutes	**Foundation-based research institutes**	**Private research institutes**			

Other stakeholders are:



Stakeholders are listed in greater detail in *Appendix 1, EA tables.*

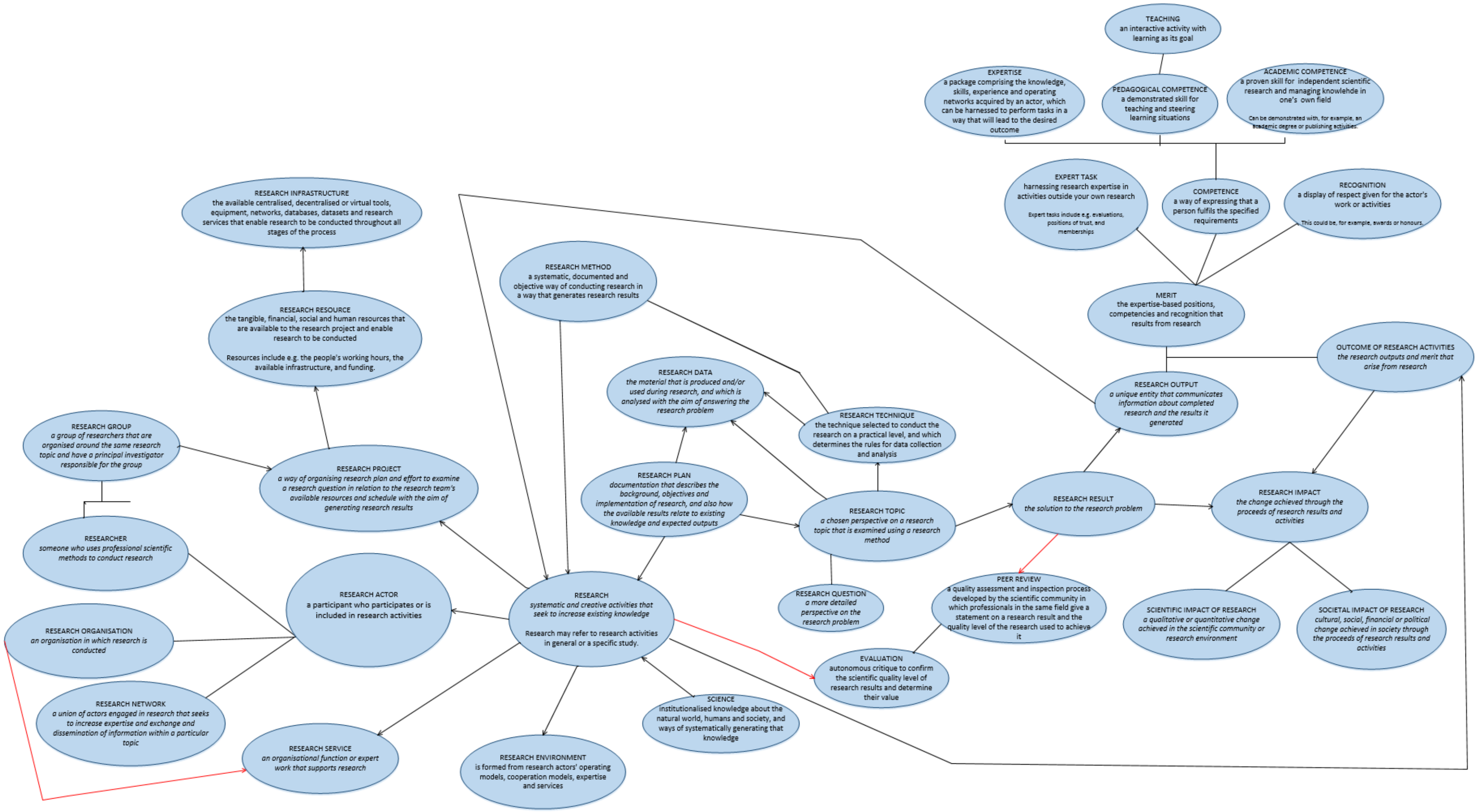
4.5. Conceptual model

4.5.1. Key concepts and their relationships

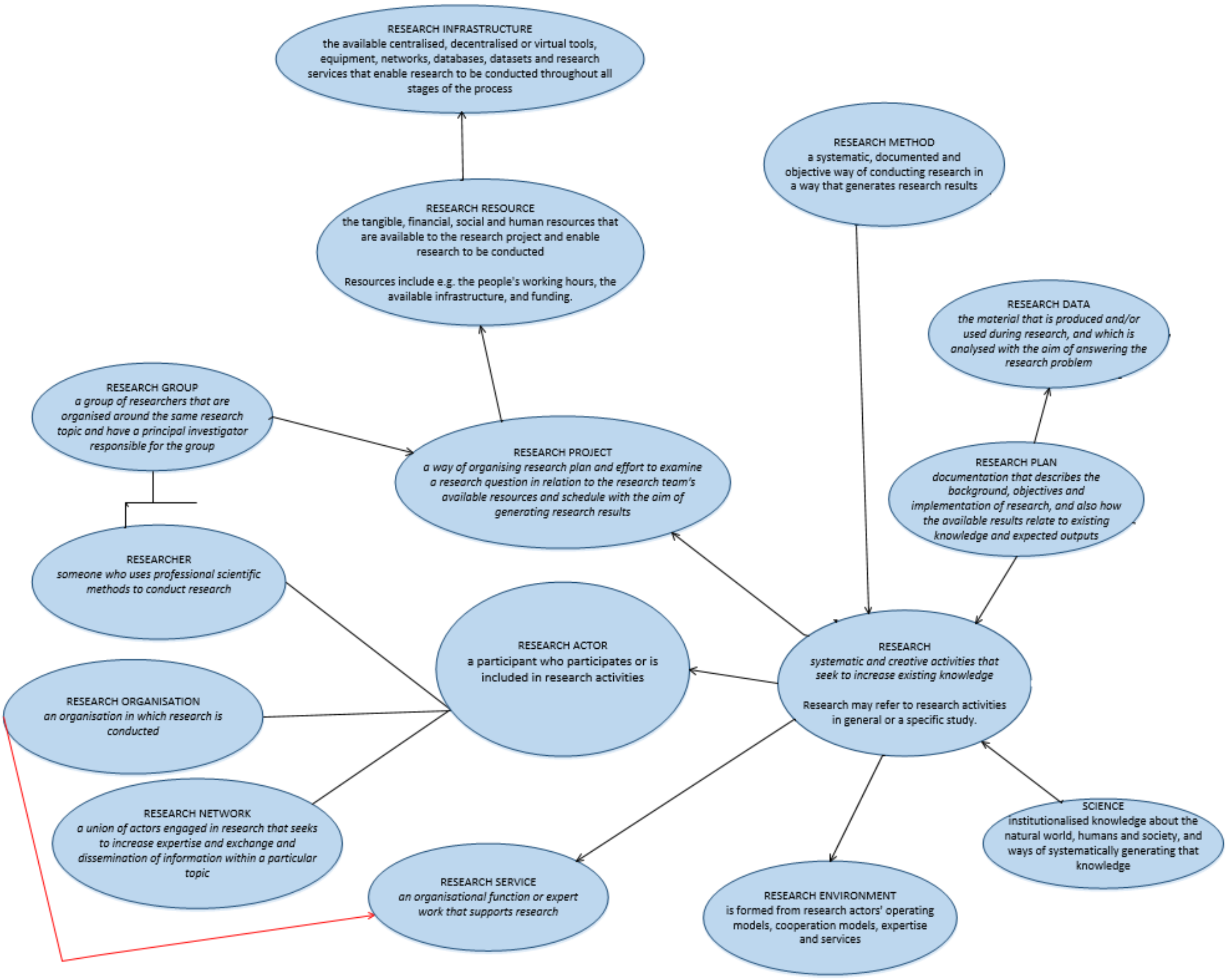
Concepts help us to form and describe a shared, high-level understanding of open science and research as a whole.

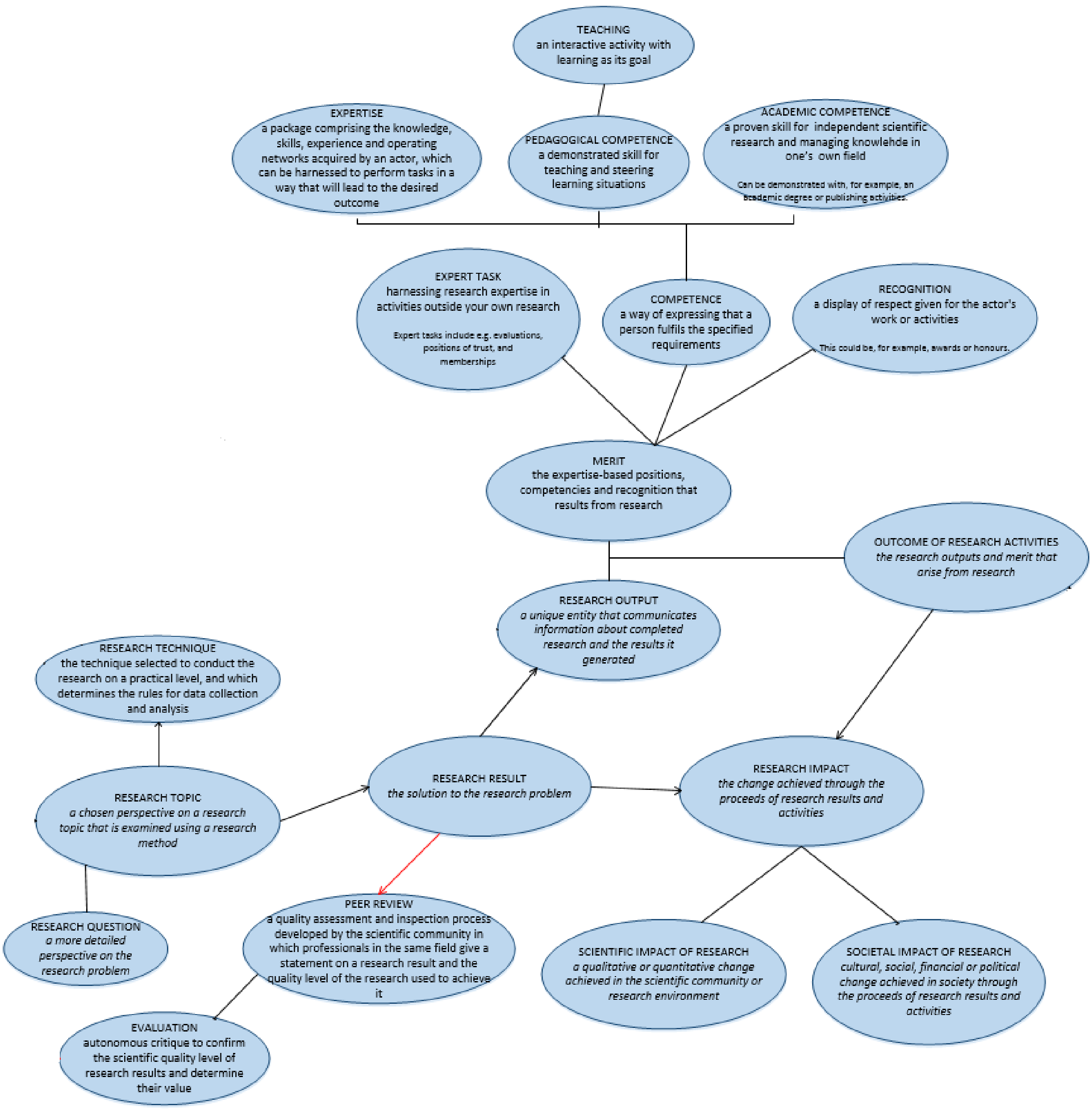
Concepts are a key factor in interoperability: a systematically defined set of concepts enables both people and IT systems to operate effectively. It also enables the semantically interoperable exchange of information, thereby promoting openness and the implementation of research.

The main concept model is as follows:



The following diagram shows the concepts above in a more readable font size:





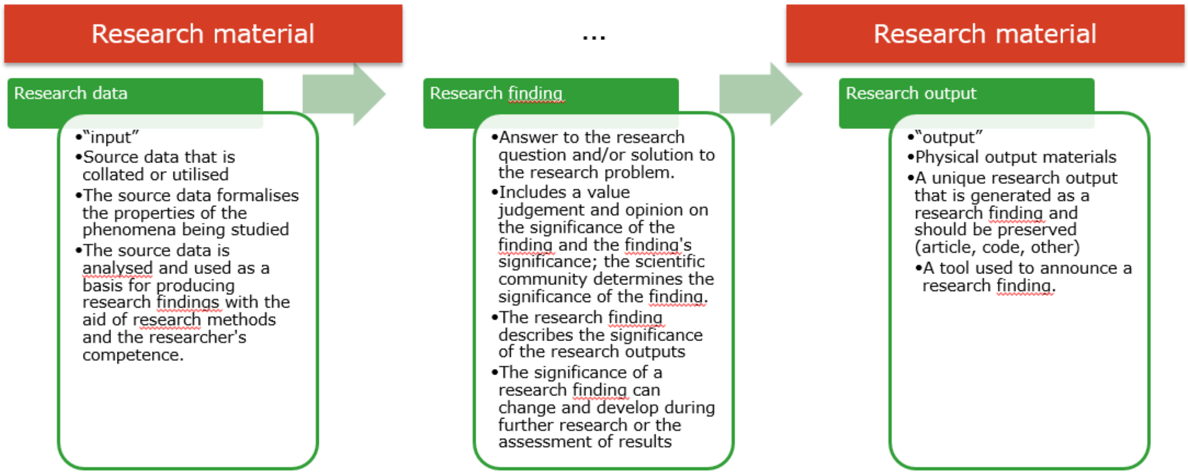
The concept map is described in more detail in Appendix 2.

In addition to the concepts defined above, concepts describing the open science and research field and its operating environment have also been defined and published on the openscience website.

4.5.2. Research findings versus research output

This framework for open science for open science and research makes a conceptual distinction between a research findings and a research output.

The differences between the two are shown in the following diagram:

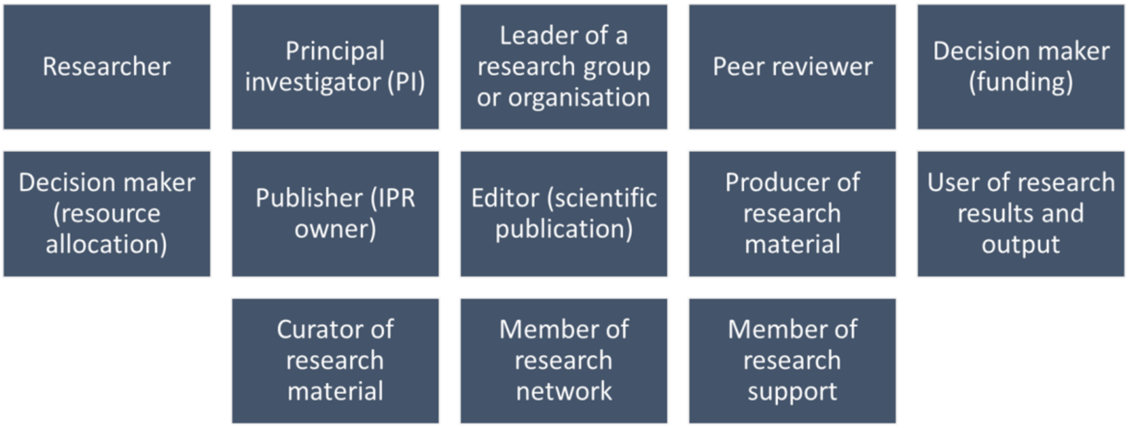


A research finding refers to the actual answer obtained to a research problem or question. The scientific value of a research is contained in its findings, and it is this value that is evaluated and assessed by the scientific community.

A research output is a concrete manifestation – typically a publication – that is used to describe the contents and findings of research. It is usually the means by which the findings are published. A research output is the tool used to communicate a research finding.

4.6. Actors and roles

The key actors and roles in open science and research are:



These roles are described in more detail in *Appendix 1, EA tables*.

5. Architectural policies – logical level

5.1. The main open science process

This framework for open science does not express an opinion on exactly how research should be conducted in different scientific fields. Research is seen to be a largely creative process that cannot, or at least should not, be described as an exact workflow process.

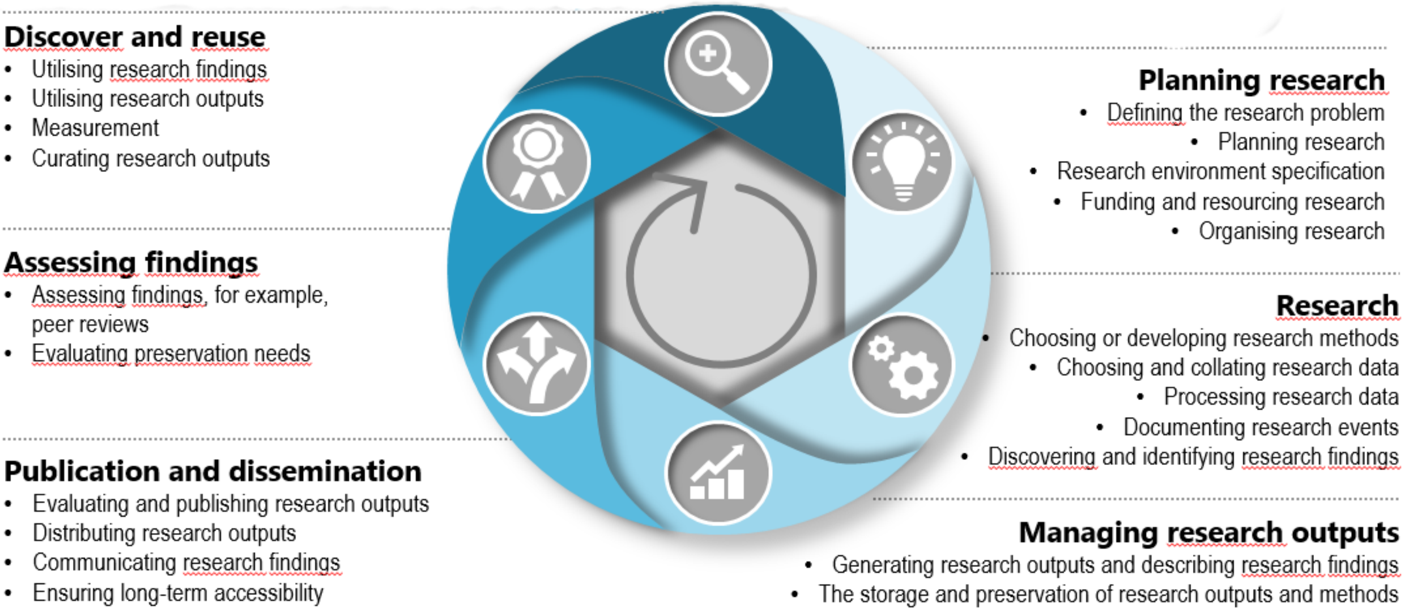
The following section describes the central high-level process that enables openness to be promoted in open science and research at all stages.

5.1.1. Key features of the process

The main research process is divided into six main stages that make science an overall self-developing process in which new research and scientific activities utilise, and are based on, previous research.

The main stages of the main science and research process are:

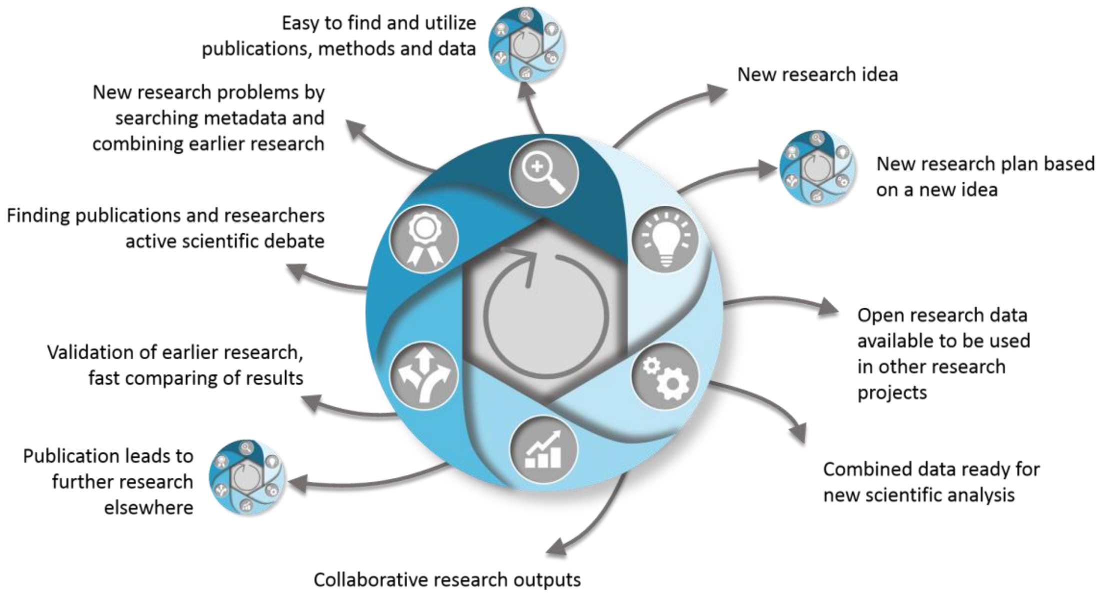
- Planning research
- Implementing research
- Managing research outputs
- Publication and dissemination
- Assessing findings
- Discover and reuse



5.1.2. An open process accelerates new science

The open science and research process seeks to overhaul and accelerate scientific and research activities.

The framework for open science harnesses openness to accelerate evolution in science and research. With the aid of open architecture, new studies and ideas can be launched at all stages of the research process. Currently, new science can only be conducted at the end of the research process, using those outputs and publications that are discoverable and easily usable.



5.1.3. Planning research

The preparation and planning phase involves both innovating and specifying the research problem itself, and then planning and organising research:

1.1. Defining the research problem	1.2. Planning research	1.3 Research environment specification	1.4. Funding and resourcing research	1.5. Organising research
Choosing the question or questions that will be answered during the research, and which will generate data	Planning how the research will progress, and choosing which datasets and methods to use and which approaches to apply	Specifying the research infrastructure and the wider organisational and cooperative operating environment	Ensuring any external preconditions for conducting the research	Allocating resources to solve the research problem

Preparing for research begins with defining the research problem. Any information about, or generated by, the research problem will be considered when planning research and specifying the research environment. Funding and resourcing affect how research is organised, and the goal is to allocate them in the optimal manner to solve the research problem.

The Open Science and Research Handbook provides researchers with a concrete data management checklist that lists openness-promoting tasks for the preparation phase.

5.1.4. Research

During the implementation phase, research data is collected or existing datasets are utilised, data is analysed, conclusions are drawn, and the key research findings are identified.

2.1 Choosing or developing a research method	2.2 Choosing and collating research data	2.3 Processing research data	2.4 Documenting research events	2.5 Discovering and identifying research findings
Deciding which research method to apply or developing a new method	collating a meaningful and logical collection of research datasets	preprocessing, processing and analysing research datasets	choices and actions are documented to ensure that the research is, in theory, repeatable	interpreting the findings and reaching conclusions on the basis of the newly generated data and its analysis

Both researchers' professional skills and the chosen scientific method are highlighted during implementation. The research method and the available processed datasets provide answers to the research question. The documentation of research events ensures high quality. Interpreting the answers obtained and drawing conclusions leads to the actual research findings.

The Open Science and Research Handbook provides a checklist for researchers on how to promote openness with regard to datasets and data management.

5.1.5. Managing research outputs

The key task of this phase is to format both the findings and the concrete research output with which the content and findings of the research will be communicated. Typical outputs are articles, documents and publications, and methods, datasets and patents.

3.1 Managing research outputs and describing research findings	3.2. The storage and preservation of research outputs and methods
research outputs are managed and described in accordance with agreed models and best practices	research outputs and the methods used to produce them are securely stored and preserved so that they are discoverable and available for re-use in accordance with the agreed terms and conditions

Research findings are presented and described as research outputs using agreed models and best practices. The outputs are securely stored, and their availability and usability is ensured.

In some cases, managing outputs may also involve collaborative writing and its associated tools.

5.1.6. Publication and dissemination

A publication's task is to make research findings as easy as possible to evaluate and utilise, and to communicate their significance.

4.1. Evaluating and publishing research outputs	4.2. Distributing research outputs	4.3. Communicating research results	4.4. Ensuring long-term accessibility
research outputs are evaluated and published or otherwise made extensively available	disseminating research outputs after, and in addition to, first publication	informing others about research results	ensuring the long-term care, preservation and accessibility of the structure, integrity and content of research outputs in electronic format

It is recommended that publications are also issued as 'parallel publications' – that is, are openly accessible to the entire scientific community and all those interested in science without charge. It is also a good idea to publish other outputs, such as research data, methods and other materials.

The Open Science and Research Handbook contains a good publication checklist listing tasks that promote openness when publishing and distributing research findings.

5.1.7. Assessing findings

The assessment of outputs and findings lends research findings scientific significance. The dialogue and evaluation involved in assessment creates merit for both researchers and, in particular, the research output being assessed.

5.1. Assessing research findings	5.2. Evaluating preservation needs for research outputs
evaluating the significance of research	defining needs for the preservation and accessibility of research outputs

The scientific significance of research findings begins to form at the publication phase, during the first peer review, but is mainly established during the assessment phase and through further studies.

The significance of research findings is assessed sometimes through peer review and sometimes through references, citations and further utilisation.

5.1.8. Discover and reuse

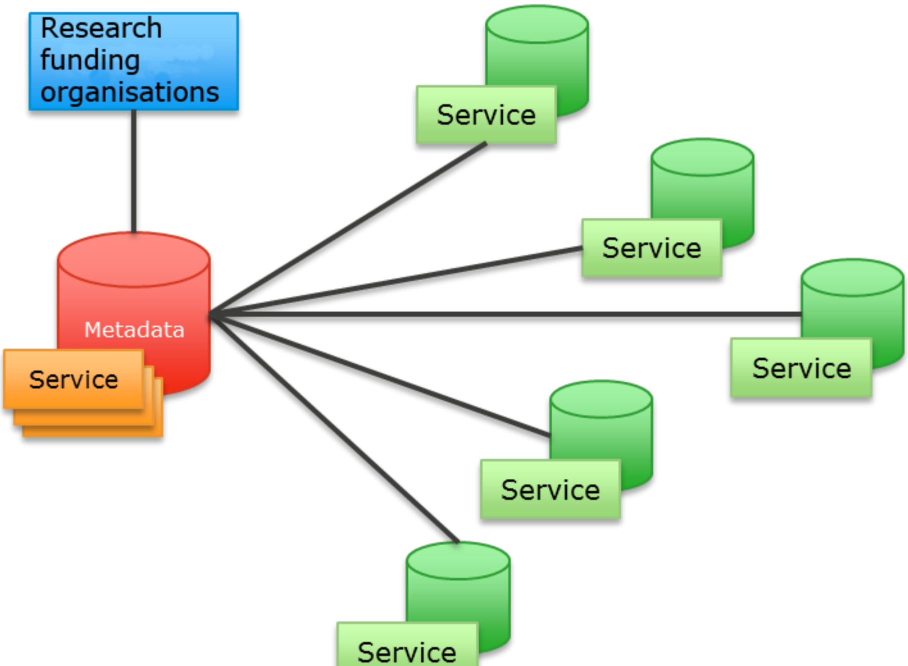
Research outputs, findings, methods and data can be reused in all new research and assessments:

6.1. Utilising research findings	6.2. Utilising research outputs	6.3. Measurement	6.4. Curating research outputs
applying information and expertise to benefit both the surrounding community and further research	the modification and re-use of research outputs	monitoring the impact of research and its outputs and findings	measures relating to the preservation and accessibility of research outputs that ensure their integrity and usability

Discovering and reusing research involves applying information and expertise to meet the requirements of – and benefit – both the surrounding community and further research. When research is utilised, connotations, references and links to earlier research increase. Curation enables data to be kept in a usable format, and also manages its growth.

5.2. Decentralised architecture principle

The framework for Open Science and Research is partially based on decentralised architecture, in which some data and datasets are stored decentrally in research organisations' own data resources and services. Their metadata is saved in a centralised service, in which the methods, datasets and publications are easily searchable and can be downloaded from their decentralised storage locations via a downloading service:



However, it is worth consolidating some data and services in shared services. Typically, these are services for which institutions of higher education and research organisations do not already commonly have existing services or data resources.

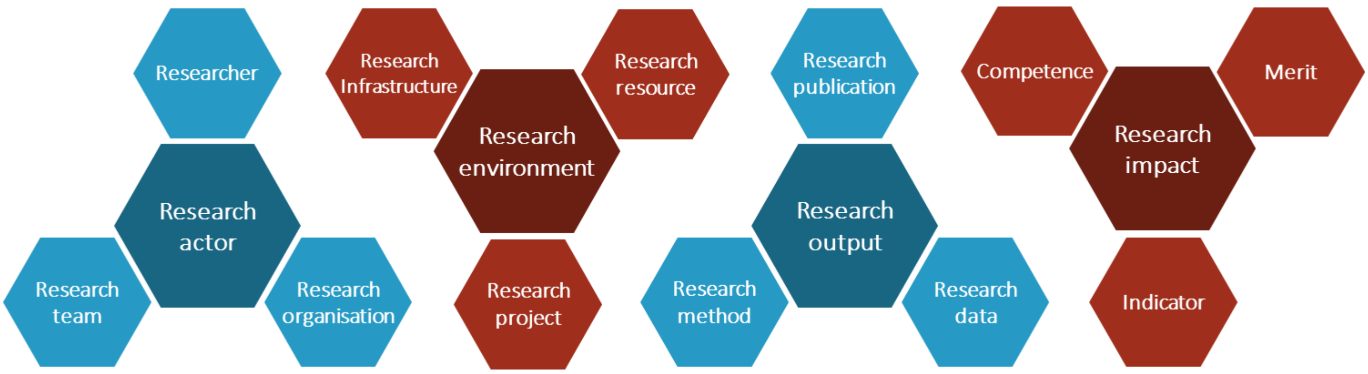
5.3. Master data, key information

Master data is identified by connecting the concept model to the open science and research process. They are logical entities with fixed attributes. Master data are those most central to the process – the data that moves between different actors, IT systems and data resources, and whose interoperability is therefore vital. That is, from both the perspective of ensuring interoperability and managing data. When seeking to enhance interoperability, it is worth starting with the master data.

Master data describe the main natural and logical items and sets of data required in key processes.

Master data are identified from among the datasets used and produced within the architecture by grouping data on the basis of their content, user need, and concept hierarchy. This grouping can be used to describe both the relationships between data and the relationships between the services that use and produce the data.

This version of the Framework for open science for Open Science and Research has identified only preliminary master data and master data categories.



Research environment = is formed from research actors' operating models, co-operation models, expertise and services

- **Research resource** = the tangible, financial, social and HR resources that are available to the research project and enable research to be conducted
- **Research infrastructure** = the available centralised, decentralised or virtual tools, equipment, networks, databases, datasets and research services that enable research to be conducted throughout all stages of the process.
- **Research project** = a way of organising research plans and work to examine a research question in relation to the research team's available resources and schedule with the aim of managing research findings.

Research output = a unique entity that communicates information about completed research and the findings it generated.

- **Research publication** = a tool used to announce a research finding based on research efforts
- **Research data** = the material that is produced and/or used during research, and which is analysed with the aim of answering the research problem
- **Research method** = the technique selected to conduct the research on a practical level, and which determines the rules for data collection and analysis

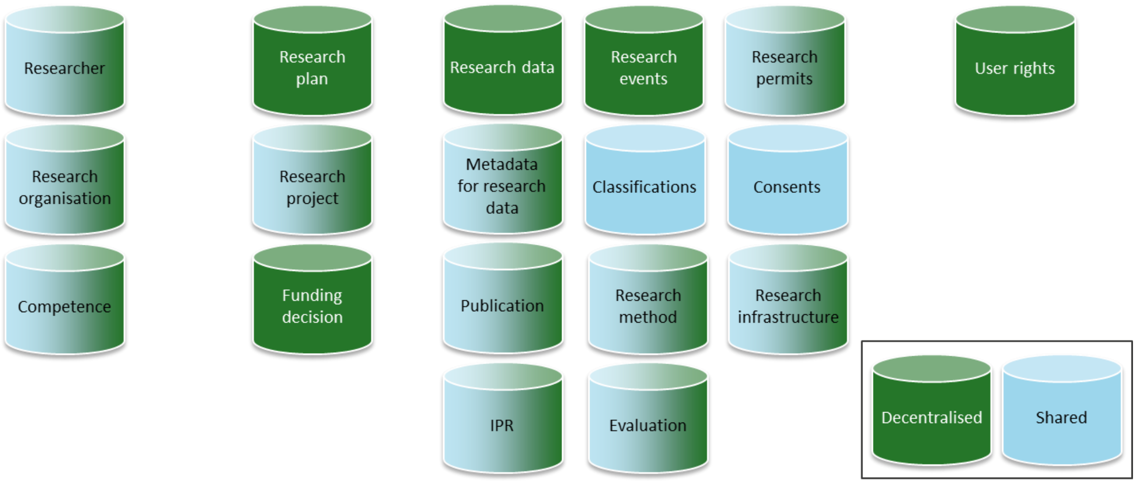
Research impact = the change achieved by virtue of the research findings and activities

- **Expertise** = a package comprising the knowledge, skills, experience and operating networks acquired by an actor, which can be harnessed to perform tasks in a way that will lead to the desired outcome
- **Merit** = the tasks, competencies and recognition that are based on expertise and arise through conducting research (a ready classification is available for this https://confluence.csc.fi/display/tutki/Meriittien+luokitus) (in Finnish)
- **Indicator** = a statistical identifier that describes a particular state or phenomenon.

5.4. Logical data resources

The key data resources needed for openness can be identified from among the master and other data required in the open science and research process. The data in these logical data resources should be defined as standardised data elements, and it should be possible to ensure their integrity during information exchanges between individual actors and shared services.

The key logical data resources for open science and research are:



In the target state, the data resources marked in light blue in the diagram are intended to be centralised resources in which the data in question is centrally maintained.

The data resources marked in green are decentralised resources maintained by institutions of higher education, research organisations or financiers. It is worth noting that, while user rights are decentralised, identification is carried out using federated services – in particular the current HAKA identification service.

Twin-tone data resources have both a centralised resources and corresponding decentralised resources. It is worth synchronising these to the greatest possible extent. For example, Researcher Data is collated in a centralised data resource, but institutions of higher education and research organisations typically maintain their own data resources as part of their research IT systems.

5.5. Application services and their configuration

In the target state, the Open Science and Research Initiative's application services and e-service environment are configured in accordance with the Service-Oriented Architecture (SOA) principle to the applicable extent. The recommended SOA principle corresponds to the following architectural principles used throughout public administration[1]:

- Develop or acquire the system with a service-oriented approach
  - The development or acquisition of an IT system should be based on operational requirements. The first task is to define the operational requirements and processes, and the services they require.
- Avoid overlapping solutions
  - An IT system that supports standardised operations should not employ overlapping technical solutions. Existing solutions should be utilised and the development of parallel solutions should be avoided. Reuse should be the default in IT system development
  - Operating models should be standardised so that different organisations can use shared solutions.

[1] Architectural principles for public administration 1.0

- Ensure interoperability
    - IT systems must support interoperability and enable the shared use of all required information
    - This interoperability requirement also covers services, products and equipment
    - IT system architectures must comply with standards, and their interoperability with other IT systems should be ensured
    - Particular attention should be paid to common standards for IT systems and their open and documented interfaces.
- Minimise supplier dependency
    - IT systems and applications must avoid commitments to closed technological solutions (interfaces, specifications, code) and individual product suppliers. Good supplier management coupled with the use of open, standard-compliant solutions will promote the achievement of this goal.

Service-Oriented Architecture (SOA) provides opportunities and isolates application services into separate, independently functioning elements that combine services as required. System functionality (business logic) is mainly kept behind the software interface – the opposite is true of services, which can only be used via the end user's user interface. It is then possible to utilise the application service in question as a shared component in other services and functions. An openly documented interface means an interface with technology that employs some commonly recognised and documented protocol and file format. Another essential aspect of SOA services is semantic interoperability between data that is transferred through the interface.

SOA principles are revolve around digital **services** and components, which are therefore preconditions for system functionality. The use of service-based architecture does not dictate the standard technologies to be used. Instead, SOA sets requirements for the application services themselves. The key principles that apply to SOA services are:

- Services should have different levels of granularity and be able to be used as modular units
- Services should have well-specified and standard-compliant interfaces (interface principles, open interfaces in particular)
- Services should be loosely coupled (integration principles)
- Services should be easily discoverable (service catalogue)
- Services should be business-oriented (not forgetting the technical support services essential to the technical environment)
- Services should be reusable.

In the target state, the key application services in the Framework for open science for Open Science and Research are based on mutually loosely coupled application services or systems 'hidden' behind application services.

The most important SOA solution principles that should be adhered to in the design and implementation of Open Science and Research Initiative application services are:

- Reusability
- A systematic approach
- Service from an operational perspective
- Technical and semantic interoperability
- Flexibility and expandability

It is worth noting that, in this context, an application service is not the same as an application. An application service is an abstract service-based functional entity that can be implemented using either one or more applications. Likewise, a particular application or collection of applications can cover several different application services.

In the following section, the application services required by the framework for open science for open science and research have been grouped under the stages of the open research process, using the previously identified open science and research services.

5.5.1. Logical application services that support open science and research

This is the summary of the logical application services that support the various stages of the research process. It is based on a report by the Open Science and Research Initiative's Service Working Group that provides a fuller description of the processes and their associated services, application services and corresponding service implementations.

Certain limitations were set in the report, and therefore also in the applicable sections of this summary. Most importantly, the limitations for the framework for open science are the same as those for the service map. The service map also has its own limitations in, for example, the naming of service implementations according to the situation in 2015, the presentation of services under development, its handling of services provided by international actors and networks, and the descriptions of service implementations in research infrastructures. Some of the services will be implemented as application services, some as other types of services, and some as a combination of the two. These limitations will not necessarily be evident in this summary.

Not all of the services will necessarily be required in all types of research or research projects. In both the service map and this summary, training and advisory services for research have been modelled as a single service. These types of services are involved at all stages of the process covered by this report.

A service map is by nature a dynamic tool for describing aspects of processes and services. These descriptions will change with the advent of new projects and requirements. The service map will be further developed alongside the framework for open science, as it has been to date, and the next version will contain an expanded service map with more detailed descriptions of certain sections.

Not all of the service definitions will be given in this summary: they can be found in the report itself or, where applicable, in the framework for open science. It is not necessarily the task of the Open Science and Research Initiative to implement all of the application services listed in this summary, even though a need for them has been identified for researchers or research. The application service map does not therefore describe the current state, but rather the target state, and no line has yet been drawn between services for which service implementations will be carried out at a later date. The service implementations described in this summary are examples that have at least some connection to the Open Science and Research Initiative.

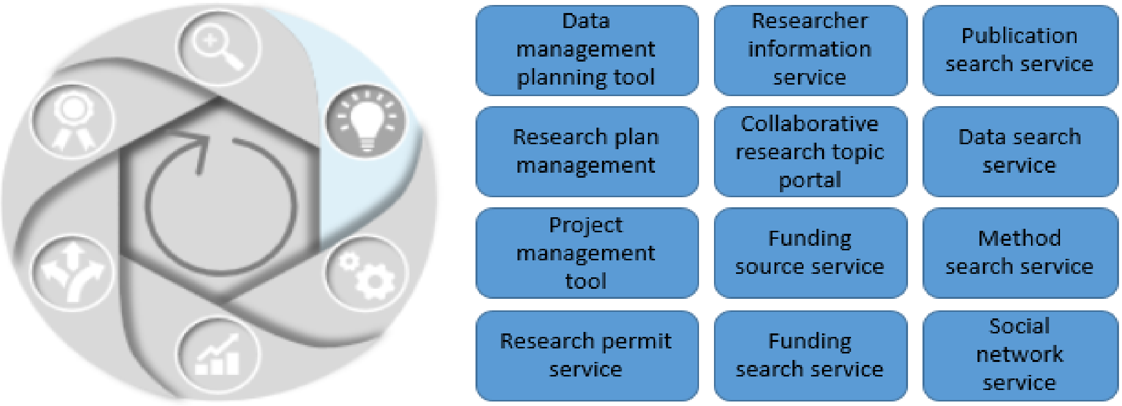
Neither does this summary express an opinion of whether existing service implementations are sufficient to implement each application service in a way that will best support the creation of openness in science and research.

Common services

The Service Map for Open Science and Research contains common services that are used at all stages of the main process. Typical common services relate to identification, search functions, and information security.

Service	Description
Training and advice	Training and advisory services are required to utilise other services. Training and advice is also provided as an independent service implementation; for example, data archives offer training and advice from the perspective of curating research outputs.
User authorisation service	User authorisation services manage users' role-based authorisations.
Log service	Log services manage user data and logs for IT systems in accordance with best practices.
Data transfer service	Data transfer services enable data and outputs to be transferred from one service to another.
Information security service	Information security services provide the security services required by users and systems, such as guidelines, advice and audits.
Identification service	Identification services offer the opportunity to log into an IT system and access the data it contains.
Social network service	Social network services enable the formation of networks in which every user can both actively communicate with others and receive information.

Planning research



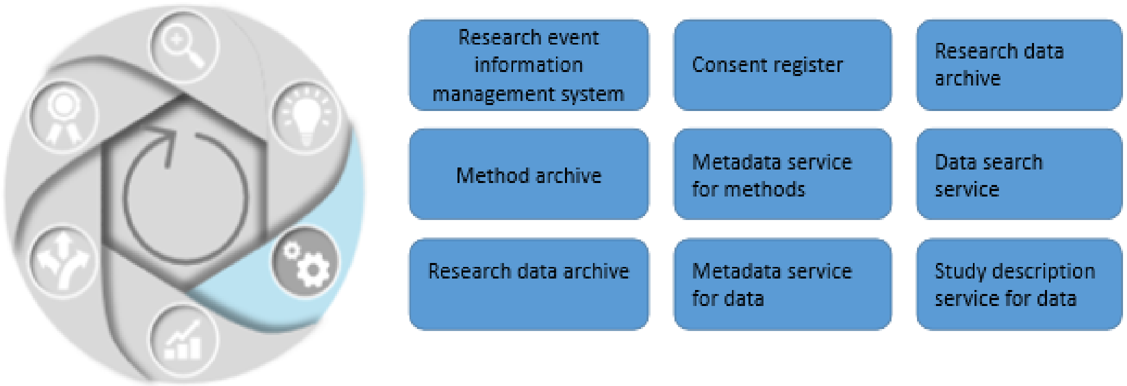
Application services that support openness and preparing for research are:

- Data management planning tool
- Research plan management
- Project management tool
- Research permit service
- Researcher information service
- Collaborative research topic portal
- Funding source service
- Funding search service
- Publication search service
- Data search service
- Method search service
- Social network service.

Of these, the following have service implementations implemented or the following services are currently being implemented:

- Publication search service (Juuli)
- Data search service (Finna, Etsin, Aila)
- Data management planning tool (Tuuli)

Research



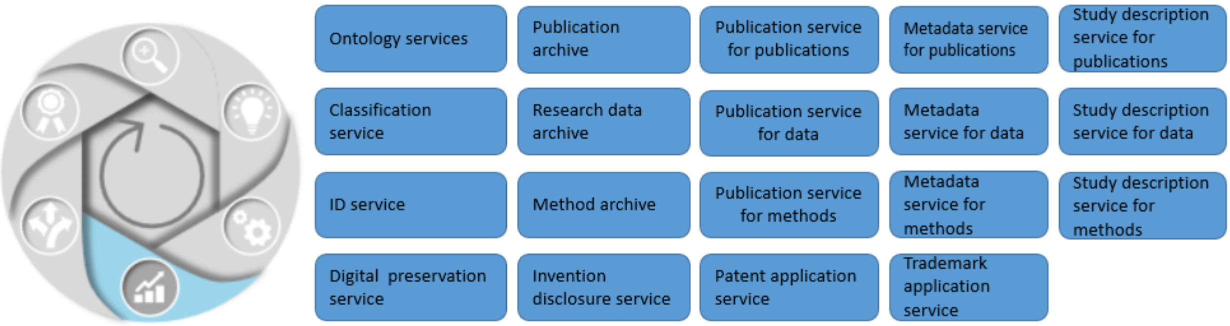
The application services that support research are:

- Research event information management system
- Method archive
- Research data archive
- Consent register
- Metadata service for methods
- Metadata service for data
- Research data archive
- Data search service
- Method search service
- Publication search service
- Study description service for data

Of these, the following have service implementations implemented or the following services are currently being implemented:

- Publication search service (Juuli, Finna)
- Data search service (Finna, Etsin, Aila)

Managing research outputs



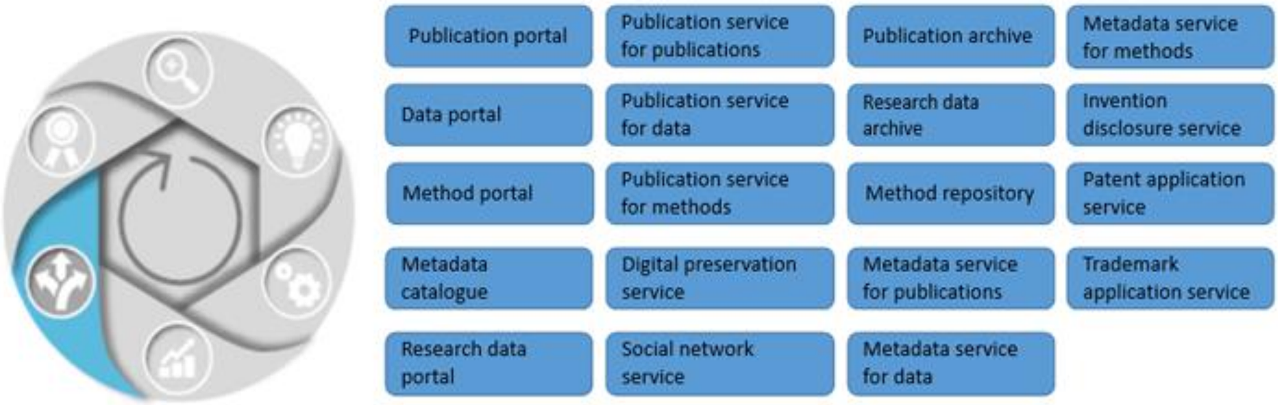
The application services that support the research are:

- Study description service for data
- Metadata service for data
- Publication archive
- Study description service for publications
- Metadata service for publications
- Classification service
- Study description service for methods
- Metadata service for methods
- Method archive
- Ontology services
- ID service
- Research data archive
- Digital preservation service

Of these, the following have service implementations implemented:

- Publication archive (local publication archives, national publication archives)
- Research data archives (IDA, Language Bank, Finnish Social Science Data Archive)
- Study description service (Etsin)
- Ontology service (Finto)

Publication and dissemination



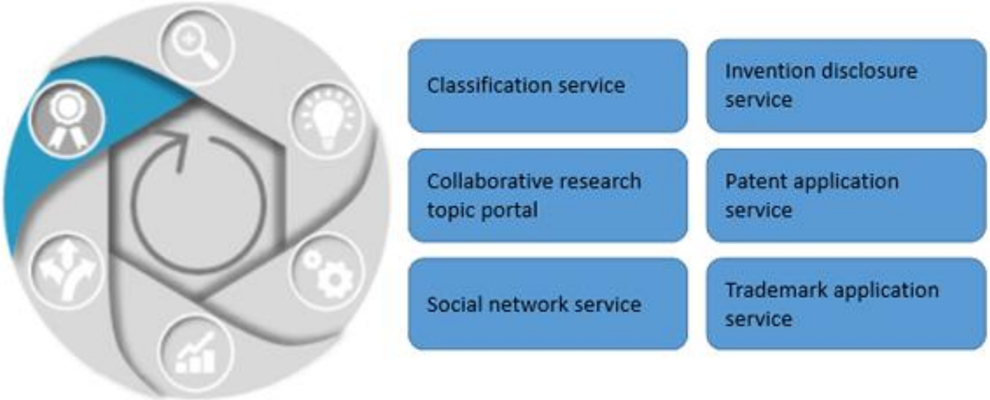
The application services that support publication and dissemination are:

- Publication service for data
- Metadata service for data
- Data portal
- Publication archive
- Publication service for publications
- Metadata service for publications
- Publication portal
- Invention disclosure service
- Publication service for methods
- Method repository
- Metadata service for methods
- Method portal
- Metadata catalogue
- Patent application service
- Digital preservation service
- Research data archive
- Trademark application service
- Research data portal
- Social network service.

Of these, the following have service implementations implemented:

- Publication service for data (AVAA, Aila)
- Publication archive (local publication archives, national publication archives)
- Research data archives (IDA, Language Bank, Finnish Social Science Data Archive)
- Digital preservation service (Language Bank, Finnish Social Science Data Archive, NDL-LTP, ResearchLTP)
- Research data portal (CRIS)
- Publication portal (Juuli)

Assessing findings



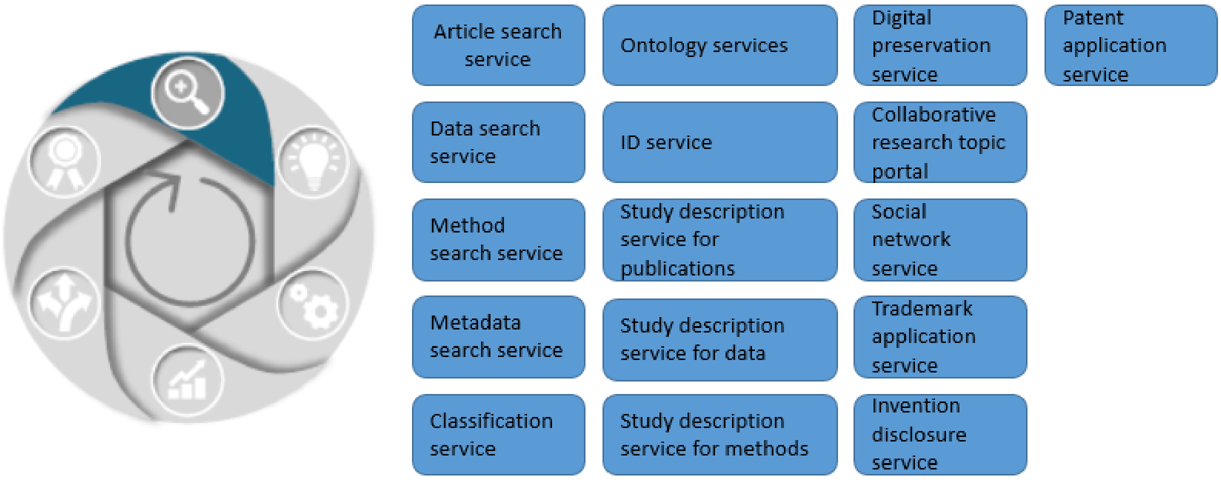
The application services that support the assessment of findings are:

- Invention disclosure service
- Classification service
- Patent application service
- Trademark application service
- Collaborative research topic portal
- Social network service.

Of these, the following have service implementations implemented:

- Classification service (Publication Forum)

Discover and reuse



The application services that support the discovering and reusing of findings are:

- Publication search service
- Data search service
- Method search service
- Metadata search service
- Classification service
- Ontology services
- ID service
- Study description service for data
- Study description service for publications
- Study description service for methods
- Digital preservation service
- Invention disclosure service
- Patent application service
- Trademark application service
- Collaborative research topic portal
- Social network service.

Of these, the following have service implementations implemented:

- Classification service (Publication Forum)
- Digital preservation service (Language Bank, Finnish Social Science Data Archive, NDL-LTP, ResearchLTP)
- Publication search service (Juuli, Finna)
- Data search service (Finna, Etsin, Aila)

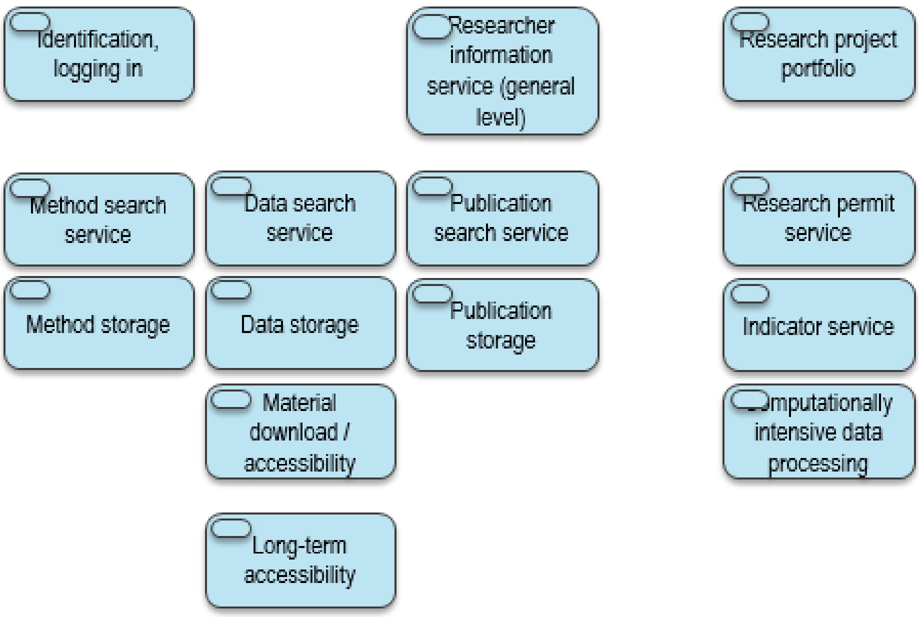
- Ontology service (Finto)

5.5.2. Application services identified as shared services

The architecture for open science and research is mainly based on decentralised architecture in which a particularly large proportion of research data resources (projects, publications, methods, datasets and data) can be decentralised and found through shared services with the aid of metasearches and metaindexing.

However, if no individual institutions of higher education or research organisations currently have extensively implemented application services for a particular aspect of open science and research, it is worth developing such new application services centrally.

The key application services that have been identified as, and are recommended to be, shared services are:



In particular, it is worth consolidating decentralised data collation services and accessibility services into a jointly developed application service. The optimal utilisation of science and research requires all research-related data to be easily discoverable and accessible to the entire research and scientific community.

The following have not yet been extensively implemented at individual institutions of higher education and research organisations or, if they have been implemented, it is not worth implementing them decentrally:

- A shared service that collates data from a variety of sources[2]
  - A research dataset search, storage and download service

[2] Datasets, methods and publications can also naturally be searched, stored and downloaded from the primary source.

- A research method search, storage and download service
- A publication search, storage and download service

- Researcher information service
- A research project portfolio
- Research permit applications and administration
- Data-intensive data processing services (heavy processing)
- Long-term accessibility
- An indicator service; forms part of the Ministry of Education and Culture's official data collection.

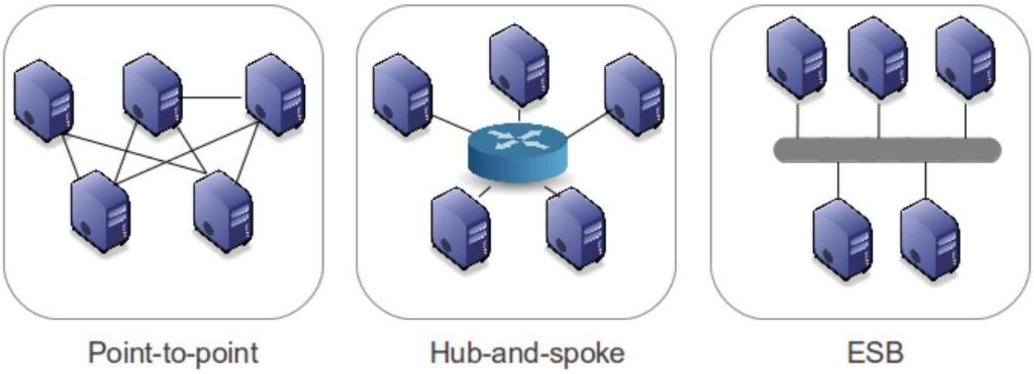
5.6. Integration model and dataflows

Connecting shared services to decentralised local data resources and services occurs with the aid of an integration service or integration model.

A systematic integration model seeks to solve, for example, the follow challenges:

- architectural complexity
- overlapping services and solutions
- inconsistent methods and procedures
- undocumented interfaces.

The various components of a decentralised system can typically be combined using one of the following three main integration models:



Point-to-point:

- Technically most efficient
- Most laborious to control in large-scale implementations
- Reusability requires agreement

Performance

Hub-and-spoke:

- Reusability with adapters
- Benefits via data models
- Complexity shifts inside the point

Hybrid

ESB:

- Provides asynchronous and reusable interfaces and messaging
- Ease of change
- Monitoring

Manageability

When the aim is to collate data in centralised services at either metadata level or as such, either the Hub-and-spoke model or an Enterprise Service Bus (ESB) solution is typically employed.

The key information exchange requirements and use cases for the framework for open science for open science and research are:

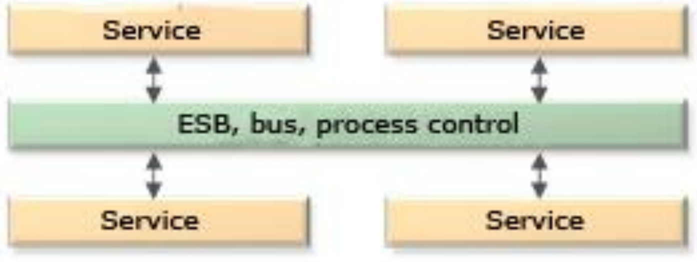
- Reporting and data storage – during research implementation and research management
- Data consistency between operational systems – the integrity and contextuality of the transferred data; for example, in which study or by which researcher has the data been produced?
- Data or IT system migrations and consolidations – standardising applications and systems
- Core data management – the controlled use of data
- The exchange of information between organisations that produce research materials; data collection and distribution (both nationally and internationally).

In the target state, it is recommended that information exchange within the Open Science and Research Initiative is mainly carried out using connected logical data resources and, in particular, services that comply with SOA principles. Effective and well-managed integration is required to achieve acceleration targets in open science and research. When it comes to establishing and operating connections and managing their messaging, **using a centralised integration solution with a set of rules to steer integration offers a significant advantage.**

A centralised integration solution manages the movement of data between different components of the system. Information exchanges between systems can usually employ one of two main strategies: orchestration or choreography.

Orchestration

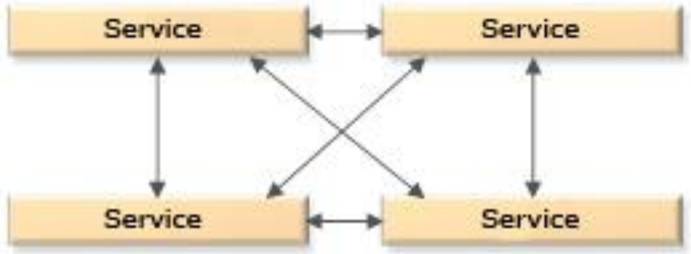
Orchestration is based on a model in which collation logic and control are performed centrally; for example, using processes modelled on the process motor, typically through either a centralised integration solution or an ESB.



Choreography

Choreography is based on a decentralised model in which the collation logic and operating process have no single owner or controller. Typically, the parties participating in the choreography process must exchange their own rules for interaction.

This solution does not require a centralised integration tool or an ESB, but can employ component-specific tools.



The main difference between orchestration and choreography is that the exchange of information is controlled by a central party in orchestration and between the parties in choreography.[3]

In the target architecture, the aim is to base the integration model on a controlled orchestration model, even though the majority of current system implementations have employed choreography. It is worth noting that, in practice, the application logic in all information exchange models is more or less divisible between the process control logic and the application services' code itself.

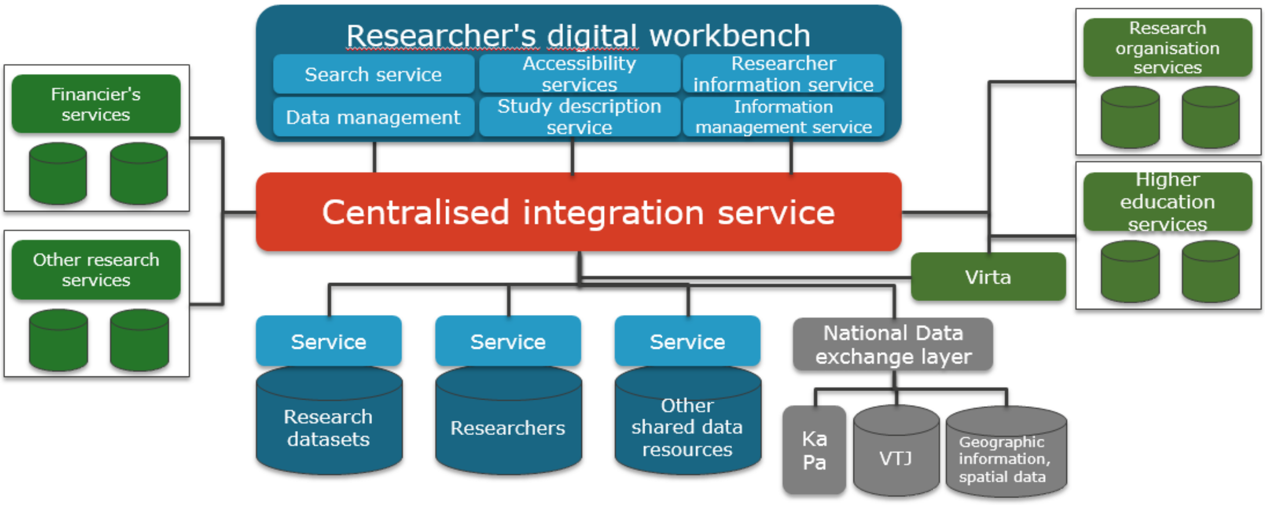
It is recommended that the exchange of information between the various parts of the Open Science and Research Initiative is implemented with an orchestration model, using a centralised integration solution that collates data or metadata from decentralised elements into shared services.

An integration service can be used to implement services and application services that facilitate and accelerate research efforts, and strengthen research at its very core.

If there are extensive partial solutions and partial systems to be found within the Open Science and Research Initiative, internal data exchanges within these solutions can utilise the partial system's or the organisation's internal integration solution. However, these partial systems should be able to be logically connected to the Open Science and Research Initiative's centralised integration solution and thereby to other services and data sources in accordance with the logical application service entities and data resources.

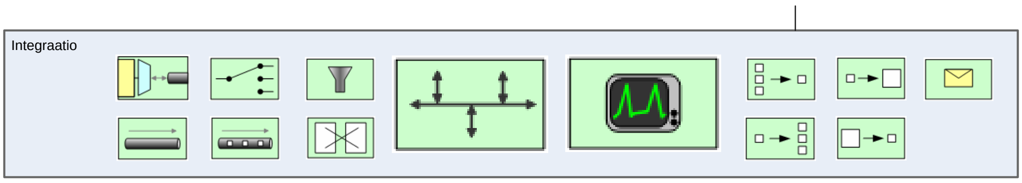
The recommended centralised integration service for open science and research is connected to many different services as shown in the following diagram:

[3] Nicolai M. Josuttis. SOA in Practice - The Art of Distributed System Design. O'Reilly Media, 2007

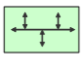
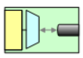
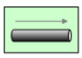
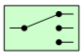


The integration solution's internal services

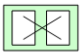
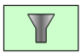
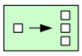
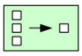
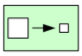
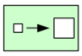
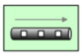
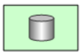
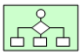
The integration solution depicted above in red is a centralised application service that transfers data between application services and data resources. The core of the actual integration layer is the messaging service that sends and (when required) modifies messages, and transfers data and messages between application services and applications. It consists of several services. The key basic functionalities are[4]:



The integration solution's key internal functions are:

	**Message bus** *A message bus provides applications with a number of shared interfaces that enable messaging between applications.*
	**Channel adapter** *With the aid of a channel adapter, an application can send and receive messages from a messaging system.*
	**Message channel** *A message channel connects applications. One application writes information to the channel and the other reads it from the channel.*
	**Message router** *A message router passes messages to the correct recipient on the basis of the message content.*

[4] See. http://www.enterpriseintegrationpatterns.com/patterns/messaging/index.html

	**Message translator** *A message translator converts a message from one format to another. The translator converts messages both structurally and content-wise.*
	**Message filter** *A message filter removes undesired messages on the basis of pre-specified criteria.*
	**Splitter** *A splitter breaks a composite message into a series of separate messages, each containing data related to one item.*
	**Aggregator** *An aggregator collects (and stores) individual messages until a complete set of related messages has been received. The aggregator then publishes a single message distilled from the individual messages.*
	**Content filter** *A content filter removes unimportant data items (from the recipient's perspective) from a message.*
	**Content enricher** *A content enricher accesses a previously specified data source to augment a message with missing information.*
	**Envelope Wrapper** *An envelope wrapper contains application data in a format that is compliant with the messaging infrastructure.*
	**Datatype channel** *A datatype channel is a special message channel that can only transfer a particular type of message.*
	**Control bus** *A control bus is used to administer the integration system. A control bus uses the same messaging mechanism used by the application data, but employs a separate channel to transmit the data required to manage the components involved in the message flow.*
	**Message store** *A message store is a* data resource *that stores message data/information about messages.*
	**Process manager** A process manager determines and controls the integration process.

Enterprise Service Bus (ESB)

In addition to using message-based and bulk-data compilations, the integration service for open science and research can be expanded into an ESB in which messaging and service calls between application services are controlled centrally. An

ESB can call not only individual messages but also application services, and can use service interfaces to transfer and receive data about service interfaces.

If necessary, an ESB can be supplemented by a process motor, which can be used to model and automate complex message traffic.

6. The initial state in a nutshell

The services in existence at the start are described in more detail in a separate table in Appendix 3.

7. Appendices

Appendix 1, EA tables

Appendix 2, Conceptual model

Appendix 3, Service map