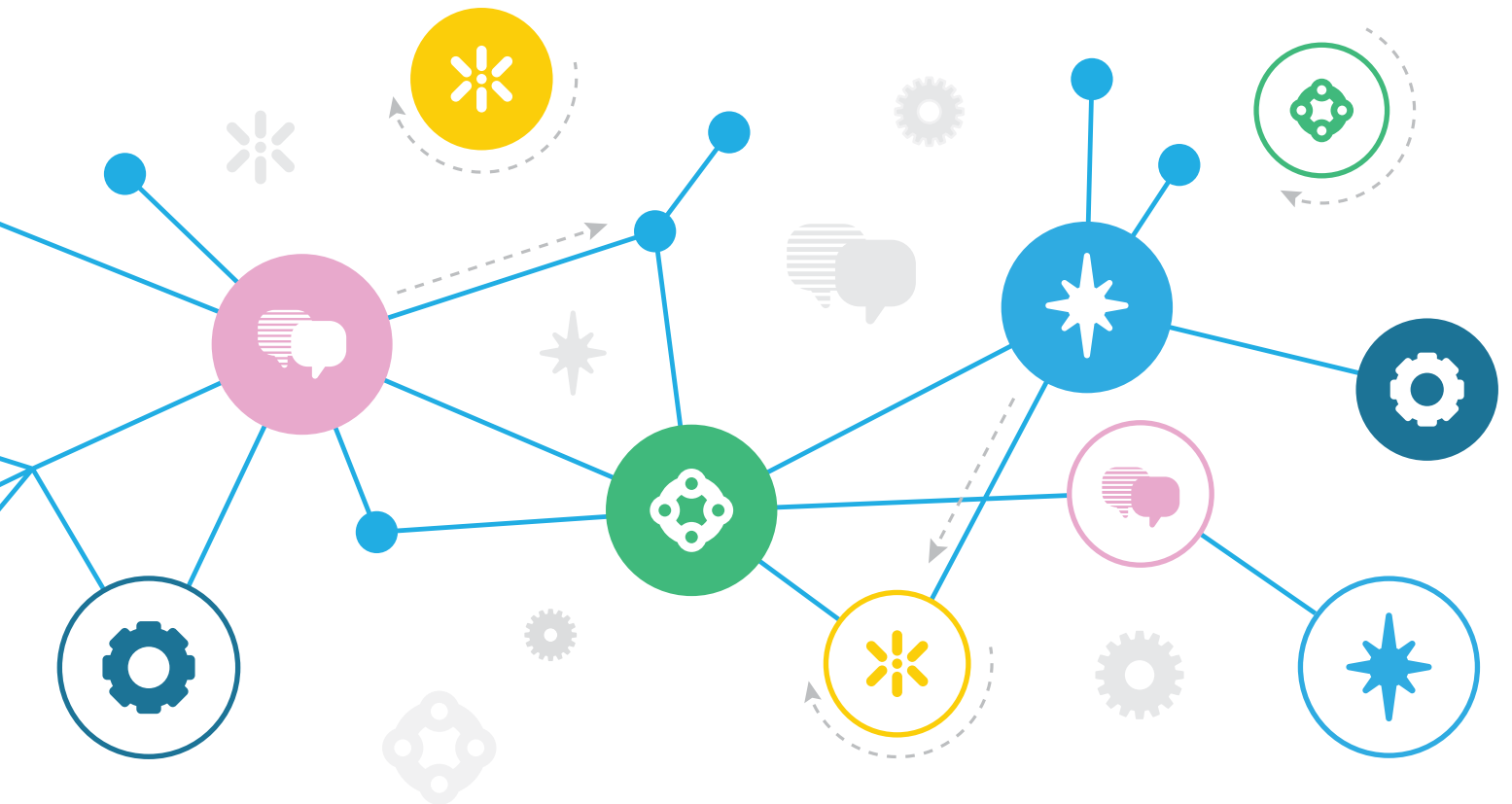




European
Commission

JRC Data Strategy

Innovation
Transparency
Access



2022

Joint
Research
Centre

Table of Contents

- 1. A STRATEGY FOR INCREASING INNOVATION AND TRANSPARENCY THROUGH DATA** 2
 - The strategic value of JRC data 3
 - JRC organisational change 4
 - Ambitions 5
 - Building on JRC's data initiatives 5

- 2. DATA STRATEGY BUILDING BLOCKS** 6
 - Data Governance 7
 - One -JRC Data Culture 8
 - Data Infrastructure 8
 - Data Support Service 9
 - Data Science Community 10

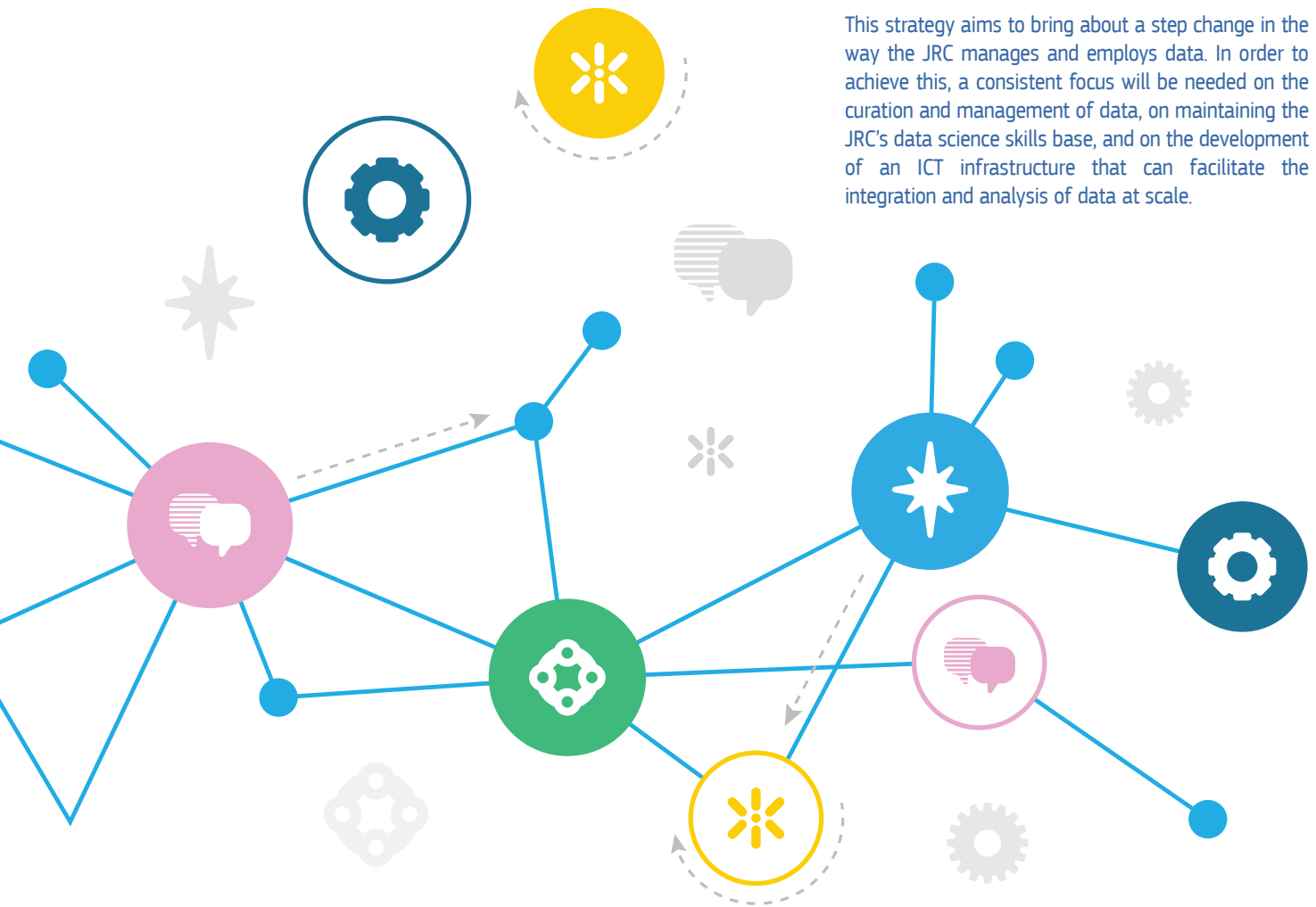
- ANNEX 1 – INITIATIVES TO DELIVER THE STRATEGY** 11
 - Initiatives on Governance 11
 - Initiatives on Culture 11
 - Initiatives on Infrastructure 11
 - Initiatives on Data Support Service 12
 - Initiatives on Data Science Community 12

1 A strategy for increasing innovation and transparency through data

Data is power. The JRC is unique in the Commission in the variety and volume of the data it produces and holds, and in its capacity to make sense of it. But the way we currently work often keeps data fragmented across islands of disciplines and policy sectors. It is time to unlock the potential of our data.

The vision of the data strategy is to create an awareness of what the JRC **collectively knows** about the world, the economy, our society and health. It aims at **facilitating the work of scientists** to reuse and **integrate data**¹ for constructing new knowledge. And it aims to **wield the power of our data** by delivering **more anticipatory and coherent scientific advice** within and outside the Commission. This will make the JRC a leader on data in the Commission.

This strategy aims to bring about a step change in the way the JRC manages and employs data. In order to achieve this, a consistent focus will be needed on the curation and management of data, on maintaining the JRC's data science skills base, and on the development of an ICT infrastructure that can facilitate the integration and analysis of data at scale.



¹ Recognising that there are different definitions of data, this document considers data in its widest sense, i.e. including processed data and information. Data also includes administrative data, i.e. data used for the functioning of the organisation.

The strategic value of JRC data

The strategic importance of data for the EU economy is evident in the Commission's European data strategy², which recognises that access to data and the ability to use it effectively is essential for innovation and growth. The same holds true for governance. Data is the raw material for evidence-based policymaking, and the Commission's corporate strategy on data, information and knowledge management³ emphasises its sharing and maximum use for better policymaking as an essential part of a new way of working. As the Commission service that creates, manages and makes sense of collective scientific knowledge for better EU policies, the JRC is one of the principle data actors in the Commission. As an in-house service of the Commission, the JRC is in a unique position to handle sensitive and confidential data in the Commission and EU institutions.

The JRC is at the forefront of scientific developments in data science and provides scientific advice for different policy areas using a wide range of data including its own 3,000+ catalogued datasets⁴. The JRC's data ecosystem covers data that has been collected or generated (e.g. from instruments, labs, through surveys or model calculations), acquired or annotated data, and increasingly external data (e.g. earth observation data, social media data) of volumes so large they cannot realistically all be hosted at the JRC. In the future, the European data strategy should also aim to make large volumes of industrial, platform and public sector data accessible to JRC scientists.

With today's rapid increase in data production and data availability, most scientific disciplines are becoming more data-driven. Greater integration of data should allow us to address complex issues that span many disciplines, a necessary requirement for coherent advice across policy areas.

Yet, there are challenges. A number of tensions prevent the JRC from fully exploiting its data and processing skills (see Figure 1). Fragmented ICT networks and infrastructure have created data islands that are not easily connected. This puts obstacles in the way of integrating data. The absence of clear corporate guidance on data handling has led to different data cultures in Units and Directorates, some looking inwards and others looking outwards for solutions. This leads to inefficient investments in ICT infrastructure and a fragmented effort that does not optimally support our aim to increase the policy-relevant impact of JRC science in the Commission. Different cultures also prevent a one-JRC approach towards data. Without a strategic approach, it is difficult to develop solutions fit for the future data challenges and opportunities.

The JRC can play a strong role in data-driven innovation and scientific knowledge by pursuing a strategy of data openness and data integration, where this is possible, for example under the terms of licences or confidentiality requirements. The JRC should be positioned as a champion of interdisciplinary sense-making derived from datasets that are distributed across the Commission. In this way, the JRC can have a greater impact in policymaking at European level, while at the same time playing a leading role in the wider science for policy community.

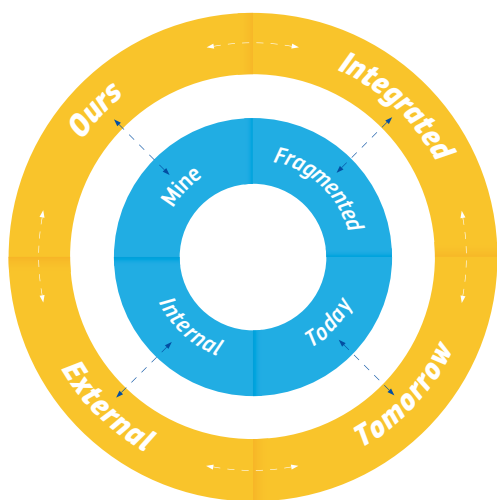


Figure 1. Tensions in the JRC's data management.

² European Data Strategy. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

³ Data, information and knowledge management at the European Commission. C(2016)6626. https://ec.europa.eu/info/sites/default/files/communication-data-knowledge-management_en.pdf

⁴ JRC Data Catalogue. <https://data.jrc.ec.europa.eu>

JRC organisational change

In the “Reset, Re-energise, Reimagine” (3R) initiative launched in 2021, the JRC’s senior management team identified the need to develop an integrated data strategy as an element to drive JRC’s innovation. Therefore, the strategy addresses weaknesses in data management that hold back innovation, including governance, culture, infrastructure, services, and community. However, the strategy is part of the larger 3R initiative and cannot be successful without other organisational changes. In particular, improvements and innovations in agile working, work programme planning, culture and communication will enable the implementation of the strategy.

In the context of the 3R initiative, the data strategy will help the JRC to deliver on its six North Star ambitions.

⊕ A **strategic partner** in EU data initiatives (North Star Ambition 1). The JRC has the unique ability to participate in the design of the initiatives, but also – with its broad work programme – as an early adopter of new data sources or data infrastructure, collecting hands-on feedback.

⊕ More **policy outputs** and **scientific publications** based on more than one JRC dataset (North Star Ambitions 2 and 3). The strategy aims at creating a culture where data integration is natural and collaboration around data is valued. Making JRC’s data ‘FAIR’ (Findable, Accessible, Interoperable, Reusable) internally, facilitating collaboration, and externally, increasing transparency, will result in more integrated and impactful research.

⊕ More **events, media** attention and public **debate** around JRC datasets (North Star Ambition 4). A reinforced awareness of the value of data will allow for more strategic communication.

⊕ A **great place to work** for its scientists, support staff and managers (North Star Ambitions 5 and 6). The strategy aims at transitioning from a culture of policing of data to one of supporting scientists with guidance, infrastructure and service.

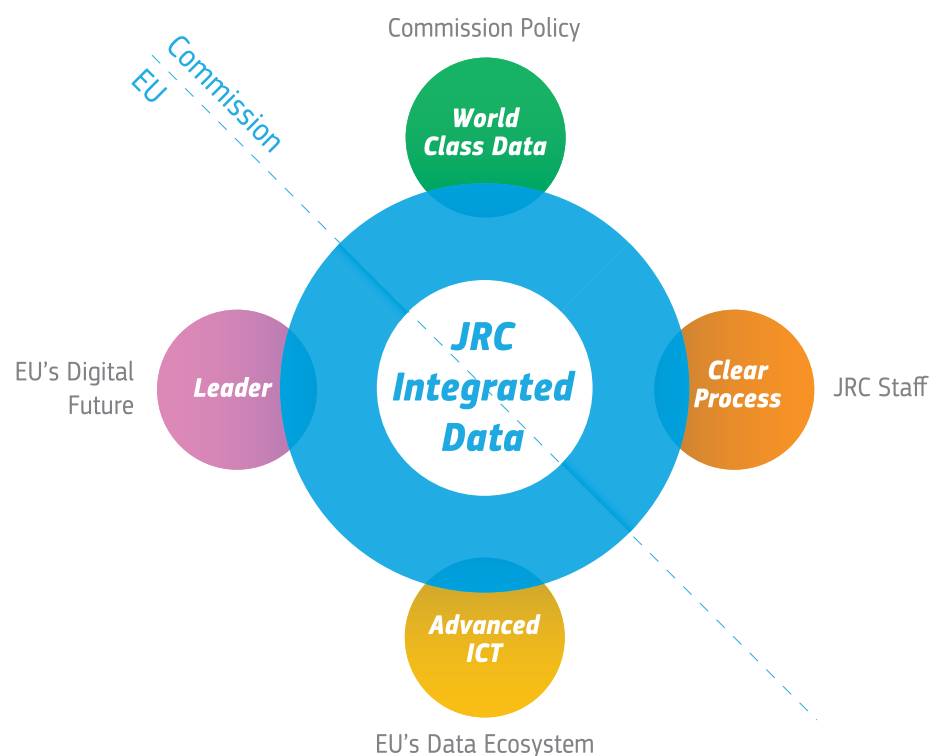


Figure 2. The JRC Integrated Data Strategy looks outwards and inwards.

Ambitions

The strategy looks both outwards to the Commission and the wider science for policy landscape in the EU and globally, and inwards to our staff and way of doing science (see Figure 2).

⊕ Inwards, the strategy will **make it easier** for scientists, support staff and managers to work strategically with data through **clear processes** for handling data, leading by example in the Commission's data governance and policies.⁵ It will ensure that staff can more easily access and integrate its data, information and knowledge in an efficient and effective way. It will support innovation in data science.

⊕ Outwards, the strategy will deliver **high-quality data**, compliant with the FAIR principles where appropriate, to support the Commission's political priorities through impactful use of data for science advice. It will also help to make the JRC more influential in the Commission's data strategy, by **leading by example** in data management supported by an **advanced ICT infrastructure**. This will position the JRC to work more closely with DIGIT, ESTAT and OP in particular on data management and data access issues, as well as with external organisations with whom we may wish to collaborate further, either to access new types of data or to share expertise.

Building on JRC's data initiatives

This strategy builds on ongoing initiatives:

⊕ In 2015, the **JRC Data Policy** (revised in 2019) addressed the requirements of the Reuse Decision on the external publication of datasets⁶. This has resulted in the development of a **JRC Data Catalogue** and **Repository** that make metadata and data (respectively) openly and transparently available, where possible.

⊕ The strategy is complementary to the **2015 ICT Strategy**. In 2021, the ICT Executive Committee delegated to the Data Committee the development of specifications for the JRC's data infrastructure.

⊕ The JRC also took an important role in corporate data management initiatives through the Commission's Information Management Steering Board (IMSB)⁷. The JRC currently hosts the **Commission's corporate data advisory service**.

⁵ Data governance and data policies at the European Commission, July 2020. https://ec.europa.eu/info/sites/default/files/summary-data-governance-data-policies_en.pdf

⁶ <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32003L0098> on the re-use of public sector information. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2018.295.01.0039.01.ENG&toc=OJ:L:2018:295:TOC (GDPR and similar for EU Institutions), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018R1807>, <https://eur-lex.europa.eu/eli/dir/2019/790/oj> (Digital Single Market Copyright Directive), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L0770> (Digital Content Directive), <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024> (Open Data Directive), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767>

⁷ The Information Management Steering Board (IMSB) leads the implementation of the Commission's corporate strategy on data, information and knowledge management. Ares(2021)5186675. Work programme 2020-2021 (Online) <https://myintracomm.ec.europa.eu/sg/info-management/Documents/work-programme-2020-21.pdf>

2

Data strategy building blocks

The JRC data strategy is centred around five themes: governance, culture, infrastructure, service and community (see Figure 3).

Stronger **governance** is needed to clarify how decisions are made and how implementation is followed up. The appointment of a Chief Data Officer and a Data Committee composed of Deputy Directors who are also Heads of Unit is an essential step.

A stimulating **culture** that supports the sharing of data is equally important and must be supported by scientists, management and staff who understand the value of data and the importance of curating data in a professional manner. The strategy aims at creating a culture where data integration is natural and collaboration around data is valued. The Data Contact Points in the Units will play an important role as ambassadors for the strategy.

An enabling ICT **infrastructure** is critical. The successful implementation of the strategy is dependent upon an ICT infrastructure that facilitates not only the

integration of data, but also other data management processes (from storage to access to analysis). This includes both an internal infrastructure but also well-defined gateways to external infrastructures. To allow for a corporate approach, business ownership of data-related infrastructure is centralised in the role of the Chief Data Officer, while allowing for a distributed implementation.

A new internal **data support service** is needed to support staff, provide information and training, and facilitate a culture of data openness and FAIRness in the organisation. It will also provide key services, such as the Data Catalogue. The service will help scientists manage data better with less effort.

Finally, the strategy will need to develop a strong data science **community** in the JRC, working together on new data analysis methods, standards and artificial intelligence applications. The strategy should foster and support existing and new communities of practice on these technologies to ensure collective learning across the organisation.

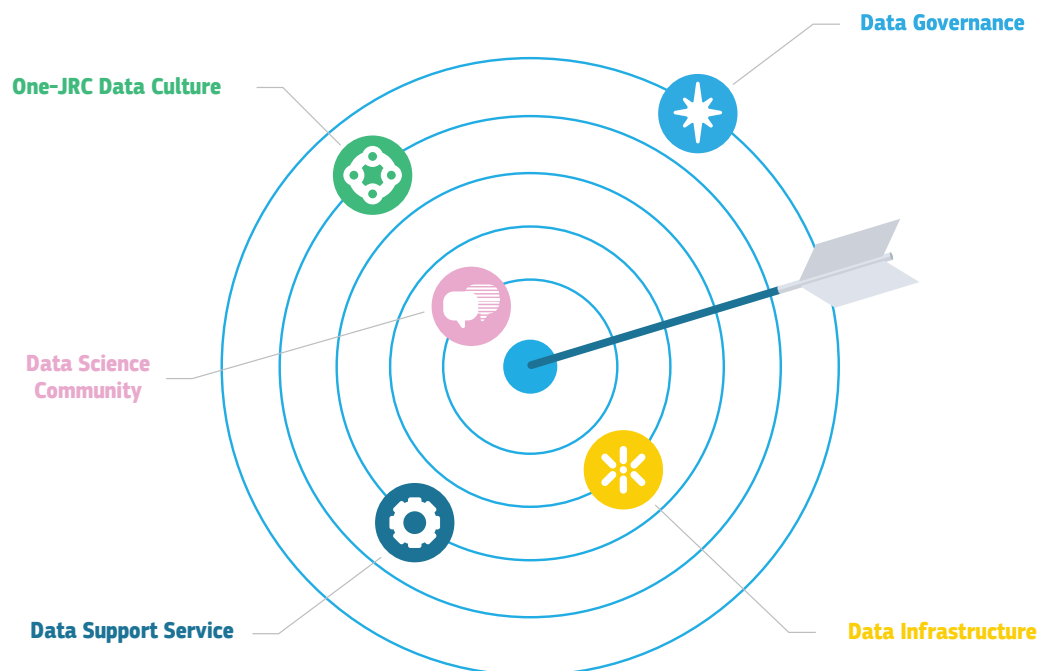


Figure 3. The five building blocks of the JRC Integrated Data Strategy.



Data Governance

Objectives

1. Establish a new **governance structure** with clear responsibilities for implementing the strategy.
2. Regularly **assess progress** in implementing the strategy and propose concrete measures to keep the strategy on track.
3. Assess new **requirements and use cases** to drive both the implementation of the data strategy and the necessary ICT data infrastructure on an ongoing basis.

To deliver the strategy, strong leadership and an appropriate governance structure is essential (see Figure 4). It must be aligned with corporate data governance and policies⁸, and have clear roles and responsibilities avoiding conflicts of interest. The governance of the strategy needs to ensure the endorsement and active buy-in of all JRC Directorates. This is critical, as without governance that engages staff and management across the JRC, it is likely that the changes sought will not be delivered, or that they will not be delivered consistently across the organisation.

The leadership, i.e. the Chief Data Officer and Data Committee, should take ownership of the strategy and of the means to deliver the strategy, involving experts and representatives of the scientific community in the JRC in the process. Data Contact Points should be selected by each unit to be ambassadors for the management of data at project/portfolio level. The current network of Data Contact Points should become a community of practice (also involving other relevant parties such as the Local Information Security Officer) that can help bring hands-on experience of data management issues and bottlenecks to the attention of the Data Committee and share expertise throughout the organisation.

The leadership is forward-looking and ensures the JRC's data handling adapts to future needs. The leadership

identifies synergies with other JRC initiatives, such as on artificial intelligence or standardisation. The leadership is responsible for the continuous collection of requirements and needs in an open and transparent way, and for proposing new initiatives aligned with the objectives of the strategy to address them.

The governance structure will set priorities for the implementation of the strategy to ensure that initiatives are properly sequenced and delivered. The Data Committee will regularly assess the state of play with implementation of the strategy and propose corrective action or new initiatives, where necessary. This assessment should also include the interplay with other horizontal initiatives in the organisation, such as on standardisation. It will consider budget and resourcing decisions that can affect the delivery of the strategy. The Data Committee will also have an important role in promoting specific use cases to help drive the strategy and diffuse relevant learnings throughout the JRC.

The governance mechanism will also have a central role in the design and implementation of any new data-related ICT infrastructures or services, ensuring that the needs of JRC users are taken into account in their design and development.

The links between the governance structure and the proposed Data Support Service will also be key, with the governance structure advising on activities such as assessing training requirements, and reviewing policies and guidelines prepared by the support service team.

Importantly, the leadership and governance structure should also be outward-looking. It should establish close links with key partners inside the Commission where expertise on data management and analysis already exists (in particular, DIGIT, OP and ESTAT), and build bridges to external organisations from whom the JRC could learn and who may also be partners in the provision of new types of data to the JRC.

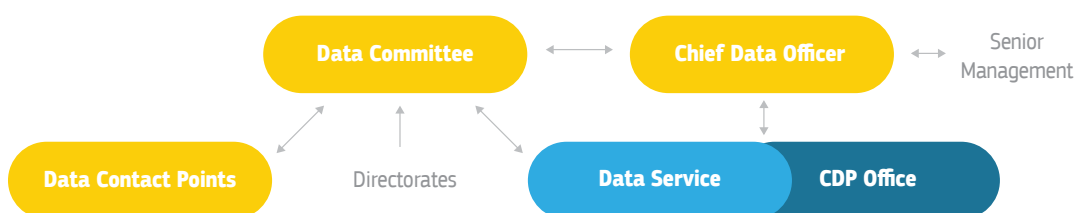


Figure 4. Data governance ensures a continuous bottom-up needs assessment and a clear decision-making mechanism on top-down initiatives.

⁸ Data governance and data policies at the European Commission, July 2020 https://ec.europa.eu/info/sites/default/files/summary-data-governance-data-policies_en.pdf.



One-JRC Data Culture

Objectives

1. Drive a **culture shift** towards FAIR data, quality data, ethical use of data, data fit for AI, the sharing of data where possible, and supporting transdisciplinary research projects.
2. Increase transparency of the data that underpins JRC publications by publishing them.
3. Develop innovative approaches to **communicating around data** and promote **JRC data-related events** inside and outside the Commission, thereby making the JRC a recognised centre of expertise on data management.

A stimulating **culture** that supports both the stewardship of data and an open approach to the sharing of data, when legally possible both internally and externally, is a necessary condition for driving transversal, innovative science at the JRC. The two principles of the Commission's strategy on data, information and knowledge management are access to data and collaboration across the organisation. Although addressed in the JRC's Data Policy, these principles need to be transformed from a burden (compliance with the rules) to a strength (less effort to share and collaborate). The other four building blocks of the strategy will support this shift to a culture where data integration is natural and collaboration around data is valued.

The first step is to **make JRC data FAIR** – findable, accessible, interoperable and reusable. FAIR data should be the general principle, while leaving flexibility as to the degree of FAIRness required. To support innovation through data science applications, FAIRification can help make data fit for AI. The culture of openness will be supported by ICT tools that facilitate data sharing and by the Data Support Service.

Second, it is important to increase **awareness of the value of data**. One way to do this is to build on the successful data events in the JRC, like the Datathon and FAIRify workshops, that recognise good practices. Another element is to amplify our storytelling with data, raise the level of data visualisation skills, and increase strategic communication around data.

Third, the data culture should **put transparency at its core**. The JRC should adhere to the trend of publishing data with scientific publications and expand it to its other publications. It should also consider ethics in the use of data.

This culture shift must be supported by all staff and the management team. The Data Contact Points in the Units, working as a Community of Practice, will play an important role in building the one-JRC data culture.



Data Infrastructure

Objectives

1. Provide JRC scientists with a **state-of-the-art data infrastructure** that can react to new evolutions but also different requirements from both large- and small-scale projects.
2. Support **access to data and data services** from a wide range of processing systems, from workstations to cloud platforms, to provide flexibility and scalability for processing, while ensuring that all JRC data can easily be catalogued, shared, and preserved.
3. Facilitate and encourage **sharing of data across the JRC and the rest of the Commission**, promoting collaboration and linking of data, data services, data scientists, and thematic experts.

A data infrastructure is a combination of ICT systems and services that support effective data management and analysis. The current infrastructure at the JRC is fragmented across multiple systems and networks. JRC scientists currently use a wide range of systems for their data processing, including local laptops and workstations, shared drives, locally managed servers, virtual desktops and servers, high performance computing (HPC) clusters, the big data analytics platform (BDAP), and external cloud systems. These systems are deployed on several networks, with different authentication mechanisms and with challenges in sharing or transferring data between systems. Rather than an integrated data infrastructure, the current approach is fragmented, hampers sharing of data between projects and across systems, and provides only limited support for management of the full data lifecycle.

A more effective JRC data infrastructure would support JRC scientists with their research across the full data lifecycle, making it more efficient to find, store, access, process, analyse, share, snapshot, publish and archive their data. It will not be a single monolithic platform. It would support **data processing from the most appropriate systems at the JRC** and simplify **interoperability** (by using standards) **and portability with external cloud systems** and systems such as the European Open Science Cloud, providing flexibility for JRC scientists. Simple access to all JRC data would stimulate research across different domains and projects. Sensitive and confidential data should be handled appropriately, adhering to the GDPR principles, with the possibility to store and process them in-house.

The ongoing **ICT consolidation project is a key enabler**, combining multiple scientific networks into one and simplifying the management of different

systems. Once the data infrastructure is working within the JRC, the approach can be shared with other Commission services. Development of further data services and systems and integration of existing systems should follow an iterative approach, governed by the Data Committee in close collaboration with the ICT Executive Committee and with feedback and guidance from JRC scientists. In line with the European strategy for data, the JRC intends to contribute to the establishment of EU-wide common interoperable data spaces, starting by improving the availability, quality and interoperability of the policy-relevant data it produces.



Data Support Service

Objectives

1. Ensure appropriate **stewardship and management of all JRC-held data** (JRC produced as well as third party) in line with corporate data governance and policies by providing advice, guidance and support.
2. Support staff **in identifying and managing data issues** (e.g. in relation to data services, data acquisition, or overcoming legal, administrative or technical barriers to data management).
3. Assist staff in accessing **novel types of data**. This includes data held by the public and private sectors that may not be generally available on the market or accessible via normal procurement processes.
4. Contribute to **drafting corporate guidelines and standards** on various aspects of data, e.g. data quality assessments and data 'fit for AI', and link to other relevant scientific communities developing similar guidelines and standards.

A new JRC Data Support Service will help and advise staff on data-related issues covering the full data management life-cycle. It will build on the existing data support helpdesk (limited to publishing in the JRC Data Catalogue) and the JRC-hosted corporate data advisory service of the Information Management Steering Board (IMSB)⁹. The main aim of the service will be to make data management in the JRC more efficient, ensure alignment with best practice and corporate data governance and policies, and bring more coherence to activities related to all aspects of data management and analysis.

⁹ The Information Management Steering Board (IMSB) leads the implementation of the Commission's corporate strategy on data, information and knowledge management. Ares(2021)5186675. Work programme 2020-2021 (Online) <https://myintracomm.ec.europa.eu/sg/info-management/Documents/work-programme-2020-21.pdf>

The new JRC Data Support Service will allow staff to find relevant information concerning data: what data exists, how to access them, how to use data, how to process or disseminate data or where to find specialised help. This support should also extend to developing new ways of accessing data from sources not currently available to the JRC, such as national administrations, platforms, industrial data and others. Some deeper reflection on this will be necessary to better understand the requirements of scientists that have been communicated during the preparation phase of this strategy document.

The JRC Data Support Service will also connect needs with the relevant expertise available within the JRC. It should act as a hub to match data analysis needs of JRC scientists with the colleagues who have the

right skills and domain knowledge to support data processing and data analysis tasks.

The service will provide support on specific technical matters related to data management, analysis and visualisation. Where needed, the service will draw on existing competences and services, e.g. knowledge and competence centres and communities of practice.

The JRC Data Support Service will also link to relevant developments and work done in the JRC to develop data quality standards and guidelines for advanced data analytics and data fit for AI.¹⁰ Using appropriate tools, such as the Semantic Text Analyser (SeTA), the Support Service should also be able to provide ad hoc consultancy to answer questions concerning the scope of knowledge in the JRC on particular topics and to help guide scientists to relevant data and documentation.



Data Science Community

Objectives

1. Produce **corporate guidelines** for the use of data science and AI.
2. Ensure that the necessary skills are available at the JRC, including **upskilling staff** on data science practice, new and emerging technologies and algorithms, and creating a strong data science community.
3. Provide appropriate mechanisms for scientific staff to share their experiences on all aspects of data management and analysis and to support the development of transversal science.

The JRC needs to maintain and increase its skillset in all areas of data science. Data employed in the scientific work of the JRC is characterised by its variety and its different quality and security requirements. The ability to interlink data depends on agreed standards. Where guidelines and solutions exist, their implementation should be ensured and supported. Where they do not, a JRC data and data science community with the interest, skills, knowledge and experience in relevant fields will support the creation of adequate guidelines and good practice. This community will link to existing horizontal initiatives.¹¹

This strategy aims to contribute to and facilitate continuous **JRC excellence through enhanced data science skills and applications**. By developing new and building on current initiatives, but also liaising with relevant existing bodies, it will strengthen the scientific and research component on data, as well as foster a joint community of knowledge, skills and good practice. It will also link to the outside scientific community, top scientists and organisations in the field to ensure state-of-the-art knowledge and practice.

Although data standards and guidelines already exist for traditional data analyses (e.g. statistical analysis, time series, modelling, indicators) and regulations have been put in place (e.g. on data quality, sampling methodologies, significance, privacy and protection), new data science methods and algorithms require additional standards and guidelines to be developed and applied. For example, such needs regard the assessment of quality for the increasingly large amounts of data employed to train AI models and their general fitness to be used for AI. For example, such assessments could look at privacy-preserving data-mining techniques and bias in input data and in final AI models.

¹⁰ Ensuring data quality standards and guidelines for data fit for AI represent new, but growing fields, in view of the new AI Act and beyond. The European and the international standardisation communities have already started working on this (e.g. ISO SC42 and IEEE P7000 series). These working items have already been recognised as relevant to operationalise the AI Act requirements. In June 2021, CEN-CENELEC started a new Joint TC on Artificial Intelligence (JTC 21) whose objective is to support the operationalisation of the AI Act - i.e. all its requirements, including the ones dealing with data quality and accuracy. It also aims to harmonise the existing standard items and fill the possible gaps - establishing liaisons with ISO JTC1/SC 42 and ETSI.

¹¹ E.g. the upcoming AI@EC package, the AI@JRC initiatives, the JRC Centre of Advanced Studies projects, and other JRC and Commission initiatives on Data, Open Data and AI.

A

Annex Initiatives to deliver the strategy

The list of initiatives here should be seen as a starting point for the delivery of the strategy outlined previously. It should be seen as a living document that may evolve over time and depending on the resources allocated or the effort required to fulfil some tasks.

Initiatives on Governance

- Build out the data governance mechanism, defining more clearly the roles and responsibilities of Deputy Directors, Data Contact Points and portfolio managers. This will include the responsibility of Data Committee members to assess progress in delivery of the strategy, to identify strategic issues for the JRC in the area of data management as well as needs for new data services.
- Develop close links between the ICT Executive and Data Committees. The Chief Data Officer should sit on the ICT Executive Committee, and joint meetings of the ICT Executive and Data Committees should take place at regular intervals.
- Determine the scope and assign the business ownership of a centralised data-related ICT infrastructure to the Chief Data Officer. The CDO, with the Data Committee, will ensure fitness-for-purpose and ensure/validate business requirements for data-related ICT infrastructure.

Initiatives on Culture

- Foster planning on data at portfolio level in the JRC's work programme to plan acquisition, integration and publication of data from the start of 2023, with the aim to have a more integrated assessment of policy priorities.
- Ensure that, where possible, data used by JRC publications are published together with the manuscripts and are available in the JRC catalogue. This will mean establishing a unified approvals process for both manuscripts and their underlying data.
- Map the training offer on data available within the Commission, identify any gaps and develop trainings for JRC staff that could plug those gaps in knowledge and skills, for example on data visualisation and state-of-the-art methods, algorithms and tools.
- Further develop JRC Data Events (such as the successful Datathon), organise hackathons, recognise best practices with awards, and participate in EU-wide events to position the JRC as a leader in data management and data analytics.
- Conduct communication and outreach activities on data both within the JRC, across the Commission more widely, and outside through stories, statistics and visualisations. This could be done, for example, through regular recorded webinars presenting one or two JRC datasets in more detail, and by announcing key new data sources and updates with wide usage and impact potential.

Initiatives on Infrastructure

- The Data Committee should gather requirements, supported by experts on the current data infrastructure.
- A core team comprising the IT architect's office (JRC.A1) as well as JRC.I.3 and JRC.I.5 will conduct a mapping of existing data processing systems, data services and data availability. This team will then lead the design and development of a data infrastructure that can effectively and flexibly support all aspects of data management and analysis at the JRC, with simple access to a range of data services, including data cataloguing, data discovery, data storage, and versioned archiving of data and code.

- Translate the design of the data infrastructure into an iterative implementation plan for developing the necessary data infrastructure, linked to the ongoing ICT consolidation process and building on existing systems including the Big Data Analytics Platform, the High Performance Computing cluster, and Virtual Desktop Infrastructures.
- Implement a data archive for projects (JRC Research Data Repository) and other tools that enable an integrated approach to project data lifecycle management. The tools should offer solutions for data and code versioning and for publication of results and archiving of data and code at the end of every project. This ensures transparency, long-term preservation of data and data analysis workflows.
- Design and implement a JRC Application Programming Interface (API) gateway to provide machine access to data produced by the JRC.

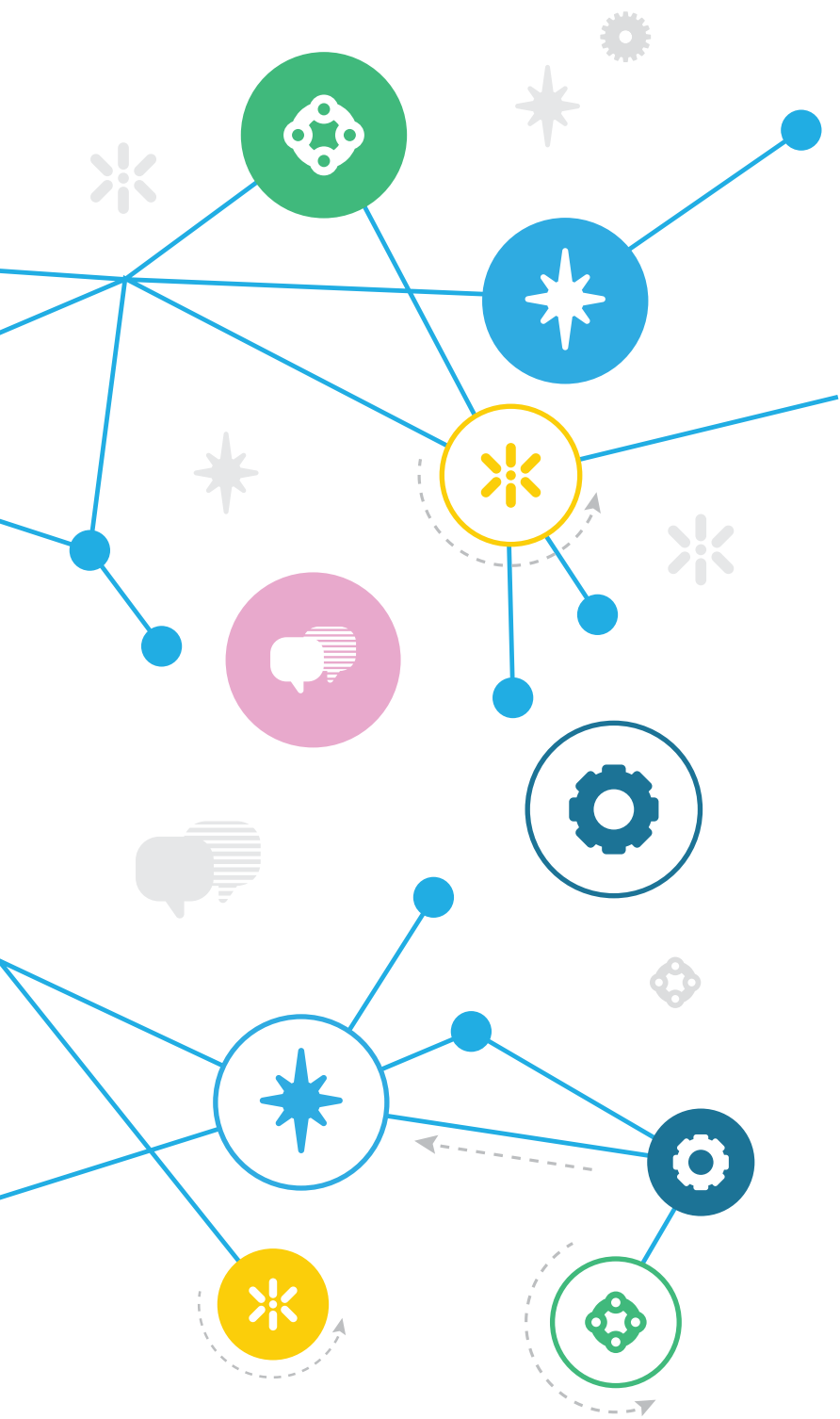
Initiatives on Data Support Service

- Develop a business plan to design a process for the provision of tailored support services on all aspects of data management to staff, e.g. facilitating the process of making JRC data FAIR or helping to develop machine-learning models with training sets that are free of outliers and bias.
- Establish a new central Data Support Service at the JRC with appropriate competences. The Data Support Service will progressively develop a catalogue of services to address the needs and priorities identified by the Data Committee.
- Convene a core group with competences on FAIR data to draft principles and guidelines on how to achieve different levels of FAIR data and other topics as needed, including wider access to novel types of data.
- Map competences in the area of data and processes in the JRC to enable experienced data scientists to assist in providing training to staff and for sharing their specific expertise on the processing, analysis and visualisation of data.
- Develop a clear service catalogue that includes descriptions of all ICT, scientific and competence centre services available to JRC users.
- Establish a single entry point to second-level support helpdesk based on existing services in the JRC, including Competence Centres, Knowledge Centres, and services on intellectual property, data protection, procurement and legal advice related to data.

Initiatives on Data Science Community

- Create a Community of Practice of professionals in data science and AI, including experts from ESTAT, DIGIT and other DGs to gather and share best practice on data pre-processing, data analysis, AI model development, testing and evaluation.
- Conduct research in areas of information management and discovery and semantic web technologies (e.g. Linked Open Data) with the aim of releasing a new version of the Semantic Text Analyser (SeTA)¹² tool to enable knowledge discovery across the JRC by connecting the data catalogue, Pubsy, JPB, datasets in the data repository and major corporate databases (e.g. EUR-Lex).
- Create working groups, together with relevant bodies, to produce guidelines and recommendations for data quality standards and data fit for AI, and implement best practice guidelines and recommendations/testing capabilities for data quality standards and data fit for AI, as well as AI model assessment.

¹² Semantic Text Analysis tool: SeTA. <https://publications.jrc.ec.europa.eu/repository/handle/JRC116152>



The European Commission's science and knowledge service Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub
joint-research-centre.ec.europa.eu

-  @EU_ScienceHub
-  EU Science Hub - Joint Research Centre
-  EU Science, Research and Innovation
-  EU Science Hub
-  EU Science