



Erasmus Data Service Centre (EDSC)

Strategic Plan 2018-2021

Contents

Executive summary	2
EDSC vision and mission.....	2
Environmental analysis	3
Strategic priorities and goals	5
<i>Priority 1: Providing access to more and different content</i>	6
<i>Priority 2: Focus on subject-oriented support and development of EDSC knowledge center</i>	6
<i>Priority 3: Boosting the EUR research data infrastructure</i>	7
<i>Priority 4: Expansion of EDSC services</i>	7
<i>Priority 5: Intensify internal and external cooperation</i>	8

Executive summary

The EDSC is unique in the Netherlands as it holds the largest collection of economic and finance databases for academic research. This extensive collection of databases is consistent with the EUR as spearhead for economics and management research in the Netherlands. The EDSC is also unique in providing specialized economics, social science and finance data support to EUR staff and students.

The mission of the EDSC is to provide excellent research data support for researchers, teachers and students throughout the data lifecycle. Our vision is to be a one stop shop for all primary research data related questions at the EUR.

Offering financial and economic data support for the EUR community is what the EDSC is known for. Providing data support to students, organizing workshops and question handling give existence to the EDSC. Running these data support services smoothly, efficiently and professionally will stay a core activity for the EDSC.

In addition to continuing current data services EDSC wants to help more and different users, provide access to more data resources and develop additional services. We have drawn up an ambitious plan for 2018-2021 to achieve these objectives.

The world of data in which the EDSC operates changes continuously and rapidly providing many challenges for the EDSC. The expectations of both researchers and students about the availability of economic and financial data and facilities for analyzing data are ever increasing. Furthermore evolving RDM practices enquire the EDSC to consider its position and services in the EUR data landscape.

The EDSC must respond to and match these developments to continue to provide excellent data support services for EUR staff and students. For 2018-2021, EDSC identifies 5 key strategic priorities that are closely aligned to its mission and vision: 1. providing access to more and different content, 2. Focus on subject-oriented support and development of EDSC as knowledge center, 3. Forging the EUR research data infrastructure, 4. Expansion of EDSC services and 5. Intensifying cooperation, both internally and externally. For each strategic priority concrete goals are set for the next four years. An explanation of the strategic priorities and goals is provided in the environmental analysis and subsequent paragraph.

With this strategic plan we aim for a living document that will grow and ripen in close consultation with faculties and other stakeholders. Only together we can achieve the goals we find important for EDSC and EUR.

EDSC vision and mission

Mission: Provide excellent research data support for researchers, teachers and students throughout the data lifecycle.

Building on the insights from the environmental analysis we consider it important that the EDSC has a clear vision for the future and what its role is. Our vision is:

“The EDSC as a one stop shop for all primary research data related questions at the EUR.”

Environmental analysis

We will now describe the most important structural developments for economic and financial data that affect the EDSC. First we mention the rapidly growing demand to use large volumes of financial and economic data and the demand for the availability of different types of data. As a second, we identify the need for Research Data Management (RDM) compliance services and the associated necessity to further develop the EUR research data infrastructure. The strongly growing demand for access and use of different types of data illustrates a third topic that is considered crucial for EDSC for the next four years i.e. the expansion of the EDSC services.

More and different data wanted!

The demand for data is changing rapidly both in terms of data volumes requested by students and staff as well as a growing demand for different types data. The implications of this development are:

- A growing number of incidents as a result of the download limits set by database providers
- A necessity to supplement numerical data sources with text based and semi-structured datasets that can serve as a base for text and datamining and data analysis
- Different and additional needs for storage -, preprocessing - and analysis of large data volumes. The use of Google BigQuery by the EDSC is an example of how the EDSC likes to respond to a rapidly changing data environment
- Open Data and Open Research Data.¹² The increase of Open (Research) Data sources offers many opportunities for innovative data collections and heterogeneous datasets. Open Research Data is also one of the pillars of Open Science and a relevant topic for research funding and grant application. The adoption of Open Data standards by the financial industry as well as the new regulations for shareholding disclosures contribute to a huge pocket of interest.³⁴
- To accommodate growing requests for data and to stay up to date EDSC needs a larger budget for content. EDSC still holds the largest collection of economic and finance databases for academic research, despite the fact that the EDSC budget has been cut back for years. An extensive collection of databases is consistent with the EUR as spearhead for economics and management research in the Netherlands.

Research Data Management (RDM)

Research funding and the forthcoming GDPR⁵ legislation set requirements for RDM practices at the EUR. To support the use, archiving and exchange of research data at the EUR against these new 'standards', EDSC needs to expand its current portfolio of services.

¹See: <https://ec.europa.eu/research/openscience/index.cfm>

² See: http://www.nber.org/econometrics_minicourse_2017/

³ Applying open data principles to financial data governance. A-team group, Jan. 2017

⁴ Providing researchers with the skills and competencies they need to practise Open Science. Open Science Skills Working Group Report, July 2017. European Commission (EC). Available from [here](#).

⁵ GDPR: General Data Protection Regulation

There are three areas relevant to RDM in which the role of the EDSC needs to be explored: 1. RDM protocols for research groups, 2. Education support for data management practices, and 3. Research data-infrastructure development.

1. RDM protocols for research groups

Implementation of the FAIR principles is a crucial nexus to go from theory to practice. Business research practices at RSM departments vary widely offering many RDM challenges. Simple protocols – based on best practices- can help departments and research groups though to make direct progress on responsible and FAIR data management. The EDSC can help compiling these protocols and the implementation in practice.

2. Education support for data management practices

Develop and assist in delivering data management courses by the use of replication packages.⁶ Mastering data skills is best achieved through *learning by doing*. Part of the success and popularity of the EDSC is the result of providing practical support to students on how to collect, process and structure data for statistical analysis. This skills based how-to-do-it approach supplements to the formal student courses and is received very well among students and EUR teaching staff. Replication packages are a means to learn good Data Management Plan (DMP) practices and will be used as a starting point for developing training material for EDSC data management courses.

3. Research data infrastructure development

Creating RDM best practices and supporting the use of large data volumes require a state of the art data infrastructure. Researchers expect seamless access to sensitive and large datasets on platforms where analytical tools are available and international collaboration is supported. The EDSC has the knowledge and expertise to lead and support the development of the data infrastructure for several disciplines and research domains at the EUR.

The growing need for more EDSC services

The growing availability of data sources and types of data offers many opportunities and challenges for research and support. A telling example is that 90% of all available digital data was produced over the last two years! Data developments are so comprehensive and influential that the research data infrastructure and support processes must be arranged and tailored accordingly.

EDSC wants to help more and different users, provide access to more data resources and develop additional services. The demand for support for data collection, data preparation and assistance with data analysis of text based resources and semi-structured data is strong. The increase of Open (Research) Data sources offers opportunities for innovative data collections and heterogeneous datasets. The EDSC aims to develop concrete services for TDM, Open (Research) Data and heterogeneous data.

⁶ The use of replication packages in economics is not new. The surge of data and the growing importance of RDM fosters the use and interest in replication packages in economics.

Strategic priorities and goals

Offering financial and economic data support for the EUR community is what the EDSC is known for. Providing data support to students, organizing workshops and question handling give existence to the EDSC. Running these data support services smoothly and efficiently will stay a core activity for the EDSC.

In addition to the support services for the use of financial and economic data EDSC also supports the following RDM services: evaluation of data management plans, information provision for secure data sharing, account management for high performance computing at SURFsara (HPC Cloud), legal support for RDM, and support for research data storage in Dataverse.

In addition to the current EDSC services, we have defined 5 key strategic priorities for the period 2018-2021:

1. Providing access to more and different content
2. Focus on subject-oriented support and development of EDSC knowledge center
3. Boosting the EUR research data infrastructure
4. Expansion of EDSC services
5. Intensifying internal and external cooperation

For each strategic priority concrete goals are set for the next four years. A further explanation of the strategic priorities is provided below.

Priority 1: Providing access to more and different content

Goals:

1. Given the growing demand for financial and economic data EDSC wants to expand its database portfolio
2. Promote and experiment with new funding models. i.e. co-funding and sponsoring for financial and economic data
3. Provide access to *new databases* e.g. databases and datasets that are eligible and allow for Text and Data Mining and GIS analysis
4. Negotiate for an increase of download volumes with relevant vendors, e.g. Bloomberg
5. Foster the *EDSC Google BQ Lab* with a dedicated budget
6. Critically annually evaluate the use of the database portfolio
7. Collect and stimulate the use of Open Data

Though the EDSC offers a wide range of databases there are many more requests for databases than the EDSC budget allows for. Additional budget for content is needed to cover investments in highly valuable content for EUR-research.

To accommodate the growing demand for additional databases EDSC proposes to make use of co-financing through faculties, research projects and sponsoring.

Another EDSC challenge for the next four years are the limits set by database vendors on the use of data. Vendors by default set limits in terms of the number of (simultaneous) users and the amount of data per user or per time-period. EDSC will negotiate for an increase of download volumes with relevant vendors, e.g. Bloomberg.

Priority 2: Focus on subject-oriented support and development of EDSC knowledge center

Goals:

1. Innovate EDSC services through thematic workshops, walk-in Q&A hours and webinars, promotion of self-help, extensive FAQ's and providing access to recorded workshops
2. Organize EDSC events and EDSC showcases⁷
3. Monitor and evaluate the efficiency of the EDSC services
4. Replicate key finance and economic papers
5. Lead the *DMP tools workgroup* within the EUR RDM community
6. Trial for a combined storage of thesis and research data for students from ESE, RSM and other faculties
7. Forge knowledge of specific RDM topics, e.g. DMP consulting, data storage and long term preservation, intellectual property rights (IPR).

Why does EDSC focus on subject-oriented support and EDSC as a knowledge and skills center? In short, we think that we can help our students and staff better with a subject based approach and focused specialization.

⁷ The NBER for example recently organized a methods lecture on data linking at the annual summer institute. The EDSC will investigate if they can organize or participate similar events.

There are multiple reasons why questions on the use of economic and financial data are better answered from a subject and topical perspective than through the lens of a database vendor. Frequently students need to combine data from different vendors and platforms to create a dataset for statistical analysis. Organizing data support from a topical perspective that matches the research topics from ESE and RSM supervisors supports an integrated approach on how to collect, combine and prepare data for analyses.

The skills required to collect, process and prepare data for statistical analysis are important virtues for academic and non-academic professions.⁸ EDSC not only provides support with the use of economic and financial databases, students also need to be trained 'on-the-job' to collect, merge, organize and prepare data for subsequent processing. This practical approach to data management helps students to acquire and master important (data management) skills.

Priority 3: Boosting the EUR research data infrastructure

Goals:

1. Close involvement in the development of the EUR research data infrastructure
2. Provide a state of the art infrastructure for EDSC's core activities
3. Lead the iRODS workgroup in the RDM community
4. Promote or participate in an intraday data project with researchers from ESE and/or RSM

Researchers expect seamless access to sensitive and large datasets on platforms where they can collaborate with international colleagues and where preferred analytical tools are available. A state of the art data infrastructure is a prerequisite to support primary research and provides a significant competitive advantage. At the same time the requirements for responsible data management and the consequences for the data infrastructure and supporting processes have never been so strong.

Clearly no single architecture will do for all varieties of research and data which are in use on the EUR. Key function of the data infrastructure is to support the use and storage of research data during ongoing research and the archiving of data after completed projects. Infrastructure facilities must be suitable to support the use of sensitive data, large datasets and heterogeneous data.

EDSC strongly advocates the use of iRODS data management software to develop services that allow for secure collaboration, data virtualization and workflow automation. iRODS is already used in several Dutch universities to support complex data workflows, secure access facilities and micro data services. For EDSC the use of iRODS is as a crucial stepping stone to a one stop data service hub at the EUR. Setting up and managing a data catalog for EUR will also prove to be an invaluable resource for showcasing responsible data management and supporting data governance. The EDSC has the expertise to lead and support the development of the research data infrastructure and services for disciplines and research domains at the EUR.

Priority 4: Expansion of EDSC services

Goals:

- Develop EDSC as data collector and data producer

⁸ See: https://ec.europa.eu/research/openscience/pdf/os_skills_wgreport_final.pdf#view=fit&pagemode=none

- Draw up a program and program committee to develop a support service for text and datamining (TDM)
- Develop support for the use and analysis of large datasets with Google Cloud Platform and Google Big Query
- Forge EDSC data science skills and hire specialized staff for TDM and iRODS

EDSC wants to help more and different users, provide access to more data resources and develop additional services. We see the development of new services as a necessity to provide an answer to the changing questions posed to the EDSC. Investments in capacity, knowledge and resources are needed to achieve the goals set for the next period.

We think that developing EDSC as a data collector and data producer is a promising and valuable new service. EDSC now offers merely access to licensed content. Making use of innovative data collection methods and offering enriched datasets for analysis will be a valuable addition to our portfolio of licensed content. EDSC has several concrete ideas to collect open economics and finance data for research applications. For example, we'd like to start collecting data from the AFM to build datasets on shareholder disclosures, investment schemes and financial reporting.

Priority 5: Intensify internal and external cooperation

Goals:

1. Focus on internal and external cooperation
2. Seek collaboration with the Erasmus Data Science experts and community
3. Initiate or participate in local, national and/or international projects
4. Initiate or participate in a blockchain project, preferably together with EUR research groups

The ambitions of the EDSC request a good and highly developed network. EDSC staff already maintains good contacts with (data) communities like IASSIST, DCC, Liber, UKB RDM and LDE and will continue to do so. EDSC previously participated in European projects aimed at economic science (e.g. NEEO) and is currently involved in the ECSCR project together with RSM and ESE. For the next four years, EDSC aims to strengthen its network and to participate in projects.

At the national level knowledge exchange with the University of Leiden on text and data mining is interesting for EDSC. Relevant international networks are the Big Data Value Association (BDVA) and the Research Data Alliance (RDA).⁹ The European Data Science Academy (EDSA) provides an interesting network and portfolio of courses for the development of data science skills for the data team.

EDSC aims to participate in local, national and international projects and consortia in line with its strategic objectives. The close involvement of Marlon Domingus and Melika Nariman in the H2020 lighthouse project Big Medilytics, shows that the available expertise at the university library and EDSC is appreciated.

End

⁹ BDV. Big data value strategic research and innovation agenda (BDV SRIA, version 3.0 January 2017). Big data value association (BDV).