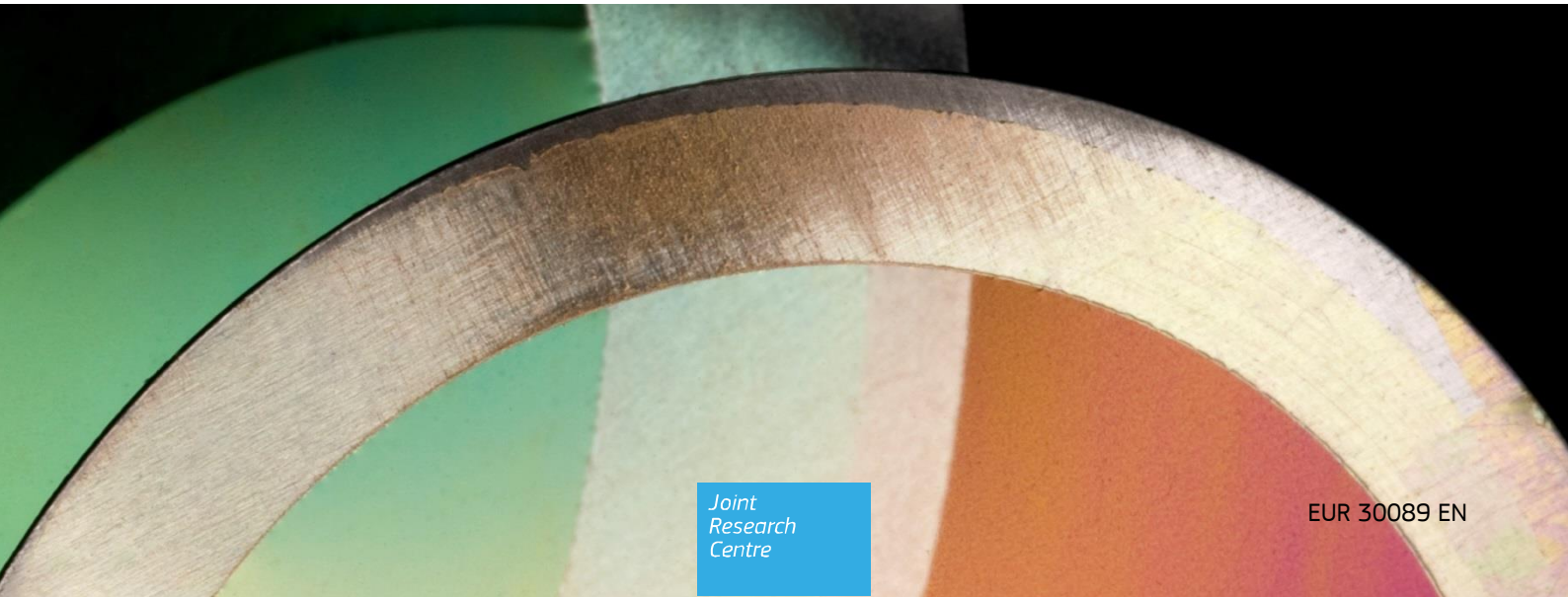
European Commission

Open Data, Open Science & Open Innovation for Smart Specialisation monitoring

Lessons from the project "S3 Targeted Support in Lagging Regions"

Enric Fuster Martí, SIRIS Academic
Elisabetta Marinelli, Joint Research Centre
Sabine Plaud, SIRIS Academic
Arnau Quinquilla, SIRIS Academic
Francesco Massucci, SIRIS Academic

2020



Joint Research Centre

EUR 30089 EN

Contact information
Name: Elisabetta Marinelli
Address: Edificio Expo, c/Inca Garcilaso 3, 41092, Seville, Spain
Email: Elisabetta.MARINELLI@ec.europa.eu
Tel.: +34-95-4488323

EU Science Hub
https://ec.europa.eu/jrc

JRC119687

EUR 30089 EN

PDF ISBN 978-92-76-10726-2 ISSN 1831-9424 doi:10.2760/55098

Luxembourg: Publications Office of the European Union, 2020

How to cite this report: Enric Fuster Martí, Elisabetta Marinelli, Sabine Plaud, Arnau Quinquilla, Francesco Massucci; Open Data, Open Science & Open Innovation for Smart Specialisation monitoring: Lessons from the project "S3 Targeted Support in Lagging Regions", EUR 30089 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-10726-2, doi:10.2760/55098, JRC119687.

Authors

Enric Fuster Martí, SIRIS Academic

Elisabetta Marinelli, Joint Research Center

Sabine Plaud, SIRIS Academic

Arnau Quinquilla, SIRIS Academic

Francesco Massucci, SIRIS Academic

Contributing experts

The authors would like to thank the fundamental contribution of the project's country experts, who provided key guidance, in depth information and analytical content, support to the participatory process and review of the final material of this project:

Federica Bertamino (Italy)

Ana Fernández Zubieta (Spain)

Michalis Metaxas (Greece)

Lajos Nyiri (Hungary)

Hugo Pinto (Portugal)

Ruslan Stefanov (Bulgaria)

Petra Szávics (Romania)

Acknowledgments

This project would not have been possible without the strong engagement and kind participation of the working-group members, which comprised over sixty participants, among stakeholders and representatives from EU national and regional governments. We are grateful for their commitment and contributions.

The generous endorsement by Mark Boden (Project Leader), and the contributions by JRC scientific officers Mathieu Doussineau, Elisa Gerussi, Dimitrios Pontikakis, Marina Ranga allowed for the swift development of the project.

Ismael Ràfols contributed to the project intellectual framework on the potential effects of monitoring systems in STI policy, Yannis Tolias provided key insights on smart specialisation monitoring and Ken Guy presented his reflection on S3 evaluation. Thanks are also due to Gabriel Resende of the JRC for supporting the publication-process of this report

We also would like to thank all the people who participated in the interviews we conducted:

- European Commission's Community Research and Development Information Services (CORDIS): Daniel Szmytkowski (written interview)
- Danish Business Authority: Tobias Kjærulff Langberg and Morten Klausen
- Department for Cohesion Policy at the Presidency of the Council of Ministers (OpenCoesione): Simona de Luca and Nicola de Chiara
- Generalitat de Catalunya (RIS3-MCAT): Tatiana Fernández
- OpenCitations: Silvio Peroni
- OpenCorporates: Chris Taggart
- Styrelsen for Dataforsyning og Effektivisering (SDFE): Brian Arreborg Hansen
- UK Research and Innovation (Gateway to Research): Ashley Moore
- Vlaamse overheid Departement Economie, Wetenschap & Innovatie (FRIS): Leen Van Campe, Pascale Dengis, Bart Dumolyn

Contents

Foreword

This report has been produced as part of the "Lagging Regions" project of the European Commission's Joint Research Centre. Implementing a Preparatory action of the European Parliament in close cooperation with the Commission's Directorate General for Regional and Urban Policy, Lagging Regions aims to support selected slow growth and low-income territories in nine EU Member States in the implementation of their smart specialisation strategies.

In addition to the provision of targeted support to each of these partner territories, Lagging Regions brings them together to share experiences and to develop cross-regional, horizontal perspectives on the key challenges they and many other regions across Europe are facing. These include: Governance, Monitoring and Evaluation; Managing industrial transitions and Transregional and transnational collaboration.

Monitoring and evaluation are important concerns for regions involved in RIS3. Under the Lagging Regions project, horizontal working group activities led to the production of a Massive Open Online course on monitoring. This work also led to the identification of the key next steps in furthering regions' capacities to monitor and evaluate the outcomes of their RIS3 efforts. This report presents the outcomes of one such next step.

The Lagging Regions project has previously highlighted the need for better tools and opportunities to compare policy processes, outputs and outcomes across different dimensions and geographies. These issues pointed directly to the role open data, open (government data), open science and open innovation in S3 monitoring, an area so far unexplored.

Following a combination of participatory processes, fieldwork, desk research and expert consultation, the working group activities described in this document, have provided great learning opportunities to the participants. This report is an attempt to share such lessons with the wider stakeholder community.

Dr Mark Boden,

Project Leader – S3 Targeted Support to Lagging Regions

European Commission Joint Research Centre

1. Introduction: Context, objectives, approach and methodology

1.1. S3 monitoring: a challenging exercise[1]

Conceived within the 2014-2020 Cohesion Policy of the European Commission, Smart Specialisation is a place-based approach which puts knowledge and innovation at the centre of territorial development.[2] At the core of Smart Specialisation lies the entrepreneurial discovery process (Foray et al., 2009; Foray, 2014; Marinelli and Periañez-Forte, 2017), whereby stakeholders interact to identify a limited set of priority areas for investment in research and innovation. Priorities represent an effort to concentrate intervention on a few economic activities that guarantee an effective response to social and economic challenges, and offer opportunities for growth. The entrepreneurial discovery process (EDP) and, more in general, the definition and implementation of S3, require a strong engagement of relevant stakeholders, including civil society groups and organisations. Finally, Smart Specialisation places a strong emphasis on results besides policy outputs. This result-oriented logic helps to explain the growing importance of monitoring and evaluation in the narrative accompanying the policy.

As pointed out by Marinelli et al (2019), within the context of Smart Specialisation, monitoring has different crucial functions: first it is a tool for policy learning for both the public administration and stakeholders, secondly it supports policy communication, and thirdly it facilitates the accountability and transparency of the public administration. As a policy-learning tool, the purpose of monitoring is primarily to measure the effects of public policies and reflect on them to improve their efficiency and effectiveness.

In order to provide some guidance and support to S3 monitoring, the Territorial Development Unit of the JRC, throughout the years, has provided some general principles and guidelines (Gianelle and Kleibrink, 2015; Gianelle et al., 2016), organised workshops and targeted-support initiatives for different EU countries and regions, developed a Massive Open Online Course, as part of the Targeted Support activities and recently carried out a comprehensive survey of policy maker perception.

According to a survey run by the JRC and presented in Marinelli et al. (2019), S3 monitoring design and operationalisation are quite demanding in terms of analytical capacity, data collection, resources and actors' participation. The most common challenge is that official indicators do not respond to the monitoring needs.

To understand the challenges of S3 monitoring it is useful to keep in mind the following aspects:

- S3 monitoring requires data scaled to the relevant territories and capturing different dimensions, as it requires measuring both aggregate trends related to the development of S3 priorities as well as the implementation and monitoring of individual public-policy instruments. The relevant data, however, may not be available and may be hard to collect.
- The nomenclature for S3 priorities, which comprise economic, as well science & innovation domains is by nature evolving, making it difficult to apply unambiguous and fixed categories for classification.
- Relevant data (if existing) may not be collected homogeneously or in a coordinated manner: it may be stored by different organisations, in different systems and different formats, thus lacking availability, findability and interoperability.
- Those responsible for S3 monitoring face the need to collect information from multiple sources whilst not necessarily having the political mandate, budget and technical capacities to do so.

Needless to say, these general monitoring challenges are also felt much more strongly in less developed regions, such as the one participating in the activities described in this report. However, despite the difficulties, there is a general agreement on the fact that monitoring pursues valuable objectives that go well beyond mere auditing. Great importance is attributed to monitoring for improving strategy performance and policy making across EU countries and regions, as reflected by the fact that national and regional authorities are devoting more resources to these activities (Marinelli et al., 2019).

The increased ability to generate, store and process data -which is at the core of the digital economy- opens up new opportunities and challenges (EC, 2019). Among other things, the emergence of **data science, artificial intelligence and natural language processing techniques** enable the analysis of large amounts of quantitative and text data and allow the extraction and development of ad hoc taxonomies across regions, data sources and traditional classification systems. Clearly, such techniques when open government, science and innovation data is available, **can contribute to the monitoring of S3 priority-areas**, as it has already happened in some pioneering regions.

Fortunately, three wide-ranging paradigms, strongly supported by the European Commission: **Open Government Data, Open Science and Open Innovation, are making S3-relevant data larger, richer, better and more easily accessible.**

[1] This section builds heavily on Marinelli et al. (2019)

[2] See http://s3platform.jrc.ec.europa.eu/what-is-smart-specialisation for further information.

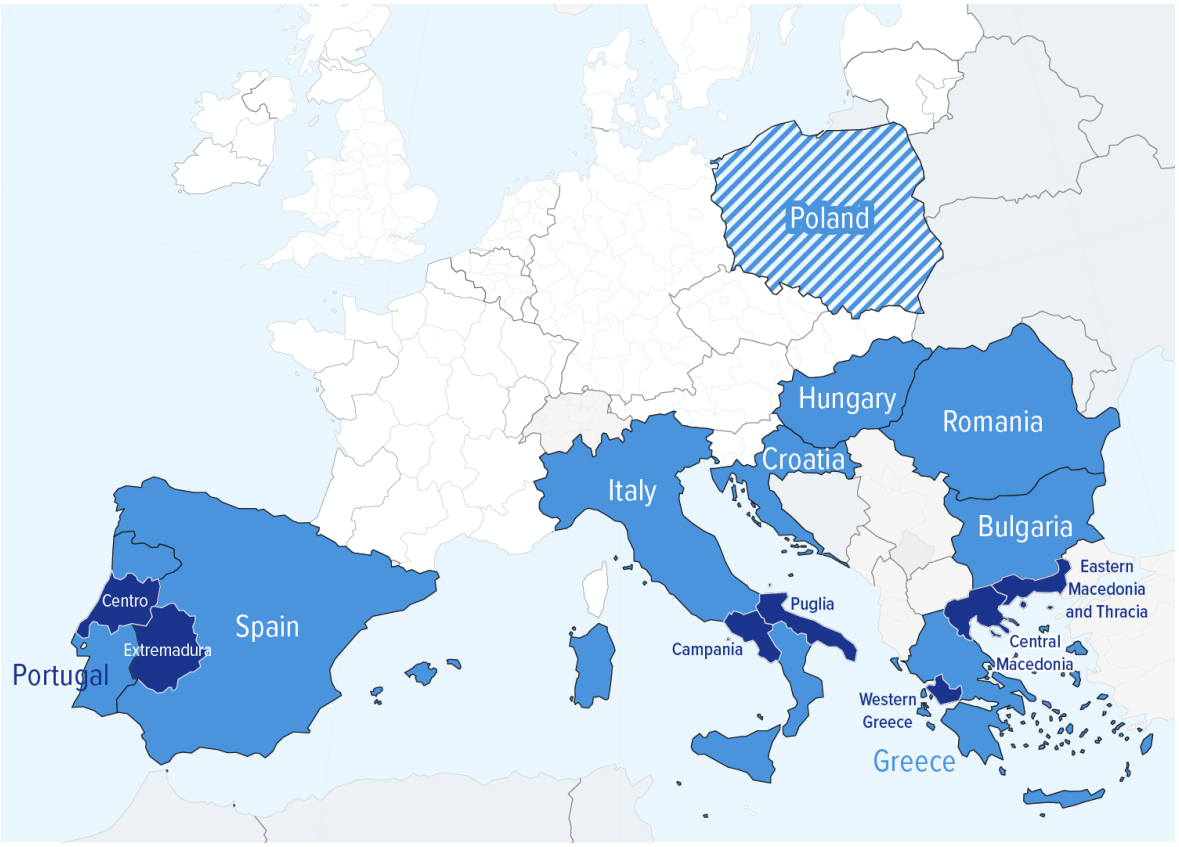
Summarising the themes discussed above, improving available policy, science and innovation data, by establishing ODSI principles and developing ODSI practices, infrastructures, repositories and open platforms, can:

- Facilitate S3 design, implementation and monitoring in a context of continuous EDP;
- provide clarity and accuracy to STI specialisation profiling at the regional, national and international level;
- provide evidence about fine-grained transformation processes, and inform policy (re)actions accordingly;
- provide rich information for mutual knowledge to public and private actors, enabling collaboration and open innovation;
- facilitate transparency and accountability towards the general public;
- support interregional and EU-wide analysis, comparability, collaboration and policy-making.

1.2. Project objectives and approach

This document reports the key outcomes of the **Study on the use of open data, open science & innovation (ODSI) for S3** (*ODSI for S3*, hereafter)[3]. The study was commissioned by the Joint Research Centre (JRC) to SIRIS Academic and implemented, in a collaborative fashion, from May to December 2019.

Figure 1 Regions and member states participating in the ODSI for S3 project4



ODSI for S3 was organised under the umbrella of the project "S3 Targeted Support for Lagging Regions", run by the JRC and financed by European Parliament.[5] Such a project, which builds on lines of work started in 2015 and which is currently in its second edition, focusses on a selected set of regions and member states and provides support to the full S3 policy cycle

[3] The full title of the study is: Study to establish roadmaps for the use of open data, open science & innovation (ODSI) in support for S3 in partners of the "Lagging Regions" project.

[4] Poland is greyed-out in the map, as a different approach was followed. Specifically, we conducted an exploratory field-work which fed into the conclusions reported in section 5.

[5] For more information on S3 Targeted Support for Lagging Regions see here: https://s3platform.jrc.ec.europa.eu/ris3-in-lagging-regions.

through both country-specific initiatives and transnational ones. *ODSI for S3*, falls under the latter category, being centred on a set of **three international workshops**.

The idea of addressing the role of ODSI emerged, on the one hand, from the need expressed by regions and member states in previous "Targeted support" activities; on the other, the compelling evolution of the data-landscape (with increasing data and visualisation tools offered by the EC itself)[6] demanded a reflection on the implications of these new developments for Smart Specialisation. Last but not least, the timeliness *ODSI for S3*, was reinforced by the EU Open Data Directive (EU 2019/1024) which was published during the implementation of the project and was fully aligned with the topics pursued[7].

ODSI for S3 had three interrelated objectives. First and foremost, it aimed at creating a **learning space** for all the participants to the workshops; secondly, the project aimed at developing some **analytical exercises** to inform regions and member states of their strength/weaknesses/potential and how to go about them; thirdly, the project aimed at translating the analysis and the learning into **policy lessons** at the EU, regional and national level.

This set of ambitious aims was pursued through multiple activities which revolved around three **participatory workshops** with senior policy makers from the regions and member states involved in the "Targeted Support Activities". The project team included, as well as on SIRIS Academic and the JRC, a set of experts with country-specific expertise.

1.3. Project methodology

As presented above, the ODSI for S3 project revolved around three participatory workshops, with senior national and regional policy makers from eight EU member states. Each workshop, which required extensive preparation and follow-up, comprised different but converging elements:

1) **Training/presentations** on key aspects of open-data, open-government data, open science and innovation. These included

- The basic definitions underpinning this paradigm shift in the generation and distribution of data
- The key principles underpinning ODSI at the technical and infrastructural level
- The presentation of good practices from the EU

2) **Participatory exercises** aimed at putting into practice the key concepts established through the presentations. The exercises particularly focussed on Open Government Data, which emerged quickly as the aspect that most interested the participants. In particular:

- In workshop 1, following an introduction to ODSI, participants selected the key policy instruments they considered most relevant for S3 and hence S3 monitoring, and the main dimensions and variables required to monitor these key instruments
- In workshop 2, participants jointly discussed and assessed selected EU and country-level datasets in their openness and suitability for S3 monitoring, through two analytical tools (the *data suitability* and *data openness grids*).
- In workshop 3, building on the previous assessment, participants developed country-specific road-maps to address critical issues identified in workshop 2 and advance toward ODSI for S3.

To build each workshops, SIRIS Academic and the JRC, as well as experts and participants, engaged in extensive **analytical** and **fieldwork** activities. In particular:

1) Before workshop 1, project participants completed an initial survey where they shared their needs and requirements in S3 monitoring, a list of relevant national and international data sources required for S3 analysis and monitoring, their previous knowledge and their assessment of open data and open science in their countries, and their expectations of the project.

2) Between workshop 1 and workshop 2, SIRIS Academic developed the *data-suitability* and the *data-openness grids*, to guide workshop participants in the assessment of selected national and EU data sources. The grids were tested by the JRC and country-experts before being given to the participants. Each participants conducted the assessment individually in preparation of the workshop.

3) Between workshops 2 and 3 the afore-mentioned grids were analysed by SIRIS Academic to identify challenges

[6] See for instance https://data.europa.eu/euodp/en/home.
[7] See section 2 for further information.

related to ODSI in the different participant countries. These analyses were validated with experts and the JRC, and was used as the main input for the development of ODSI for S3 improvement road-maps in workshop 3.

4) In parallel, from July to October 2019, SIRIS Academic and the JRC engaged in **fieldwork** with organisations that represented **good practices** in the field.

1.4. Project outputs

The three broad aims of the project were achieved by building -through the three events- one specific output per country, namely, one or two road-maps (a) developed by workshop-participants from the same country and (b) addressing some of the challenges that they themselves had identified. Such road-maps were conceived as a device to enable the analysis of challenges related to open-data for S3, to frame the national and international learning process of the workshops and ultimately to devise policy implications. Whilst the interested reader is invited to look at the documentation available on the JRC website[8], this document reports what the authors consider the key take-home lessons and outputs, both for participants to the workshops and beyond.

1.5. Organisation and presentation of the results

Based on the list of activities above the following results are presented in the remaining of the document:

- **Chapter 2 - Open (Government) Data, Open Science and Open Innovation (ODSI) as a tool for advancing S3 monitoring**, presents the key concepts guiding the discussion.
- **Chapter 3 - Open data sources for S3 monitoring: a taxonomy and some examples** presents a taxonomy of open-data (including Open Government Data), open-science and open innovation sources of relevance to S3 and S3 monitoring. The taxonomy is complemented, in Annex, with an in-depth description of good-practices across the EU.
- **Chapter 4 - From theory to practice I: analysing ODSI for S3** follows the project's desk research and participatory exercises, in particular:
    - The process of identification of S3 policy instruments and related data sources of most relevance to the participants.
    - A set of two analytical grids to assess openness and suitability of data sources to be used in monitoring the above instruments.
- **Chapter 5 - From theory to practice II: country-specific roadmaps** provides, for each country, the key outputs of the project, namely the respective roadmap(s) to address some key challenges to deploying ODSI for S3.
- **Chapter 6 - Conclusions and recommendations** provides a series of final comments and broad policy-recommendations emerging from the whole process.

[8] In particular, the methodology and materials for each workshop are available here:
Workshop 1:
https://s3platform.jrc.ec.europa.eu/-/s3-targeted-support-horizontal-activities-s3-monitoring-evaluation-?inheritRedirect=true
Workshop 2:
https://s3platform.jrc.ec.europa.eu/-/s3-targeted-support-open-data-open-science-for-s3-monitoring?inheritRedirect=true
Workshop 3:
https://s3platform.jrc.ec.europa.eu/-/s3-targeted-support-horizontal-activities?inheritRedirect=true&redirect=%2Fris3-in-lagging-regions

2. Open (Government) Data, Open Science and Open Innovation (ODSI) as a tool for advancing S3 monitoring

2.1. Open Data and Open Government

According to the European Data Portal[9], Open data can be defined as "data that anyone can access, use and share. Governments, businesses and individuals can use it to pursue social, economic and environmental benefits. [However], open data becomes usable when made available in a common, machine-readable format [that is when it is *interoperable]*. Open data must be licensed[10]. Its licence must permit people to use the data in any way they want, including transforming, combining and sharing it with others, even commercially."

Open Government Data (OGD) is data that, as well as being open, is produced or commissioned by public bodies (OECD, 2013). The value of Open Government Data is linked to its ability to promote transparency, accountability and value creation, as much of S3 monitoring revolves around the implementation of public policy instruments, OGD is particularly important for S3 monitoring.

In order to support and guide the development of OGD, different sets of principles have been put forward by non-governmental organisations and communities. Eight Open Government Data Principles were defined and agreed upon in December 2007, during an Open Government Working Group Meeting held in Sebastopol (California, United States). All but one, as explained below, are still fully valid.

According to these principles, open government data should be:

1. **Complete:** All public data must be made available. Public data is data that is not subject to valid privacy, security or privilege limitations.
2. **Primary:** data is collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.
3. **Timely:** data must be available as quickly as necessary to preserve the value of the data.
4. **Accessible:** data must be available to the widest range of users for the widest range of purposes.
5. **Machine processable:** data must be reasonably structured to allow automated processing.
6. **Non-discriminatory:** data must be available to anyone, with no requirement of registration.
7. **Non-proprietary:** data is available in a format over which no entity has exclusive control.

According to the 8th original principle, data had to be License-free, i.e. not subject to any copyright, patent, trademark or trade secret regulation, allowing only for reasonable privacy, security and privilege restrictions. However, as already highlighted before, the practice has evolved and OGD is now considered to need a licence to protect its public nature in the long term. Examples of licences used for open government data include Creative Commons International Attribution 4.0 International license[11], or the Open Government License[12] of the United Kingdom.

Research and innovation policies are supported by public funding instruments with strong administrative, monitoring and publicity obligations. These create a wealth of information and data useful for instrument-level, policy-level and ecosystem-level analysis, particularly in the context of S3.

Open government data is a first step in a long journey involving "the transparency of government actions, the accessibility of government services and information, and the responsiveness of government to new ideas, demands and needs". (OECD, 2002). Open government aims to increase transparency, participation and collaboration in common-good initiatives that require knowledge generation, management and sharing. It transforms information flows and opens up knowledge creation and appropriation, allowing for improved outputs, and, more importantly, more socially valuable processes with meaningful collective capacity-building at each step; a set of aims and processes very similar to the S3 model. Open government is mainly founded on the supply, use and integration of open data and open and continuous stakeholder engagement. Furthermore, open government data provides third-party actors with the necessary tools to develop independent analysis and assessments

[9] European Data Portal / Open Data Institute https://www.europeandataportal.eu/elearning/en/module1

[10] See the Module "Why do we need to license?" of the EDP's e-learning programme:
"Why license open data? Without a licence, data is not truly open. A licence tells anyone that they can access, use and share your data. Unless you have a licence, data may be 'publicly available', but users will not have permission to access, use and share it under copyright or database laws."
https://www.europeandataportal.eu/elearning/en/#/id/co-01

[11] https://creativecommons.org/licenses/by/4.0/

[12] http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

(highly valuable for policy monitoring), as well as new platforms and solutions servicing the local research and innovation ecosystems, potentially improving efficiency and connectedness.

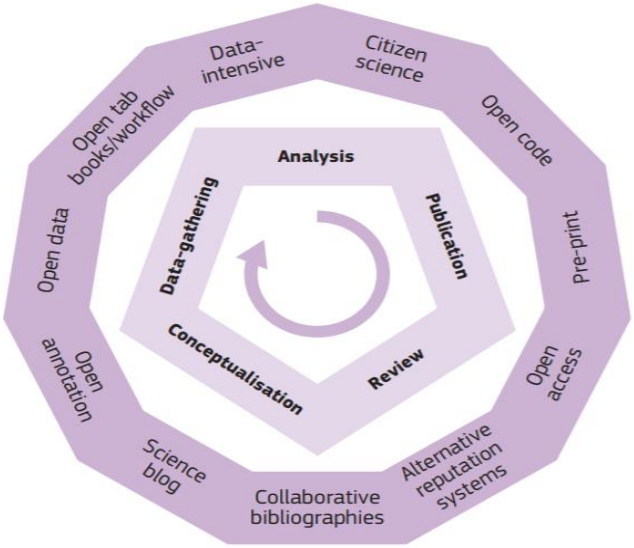
2.2. Open Science

> "Open Science represents **a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge** by using digital technologies and new collaborative tools".
>
> Open Innovation, Open Science, Open to the World
> (EC, 2016)

According to the OECD (2015) the term **Open Science** refers to all those activities that aim at providing unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives**.**

The picture below presents the changing processes and corresponding available resources or tools in the open scientific process:

Figure 2 Open scientific process



Source: EC, 2016

It's important to note that open science establishes sets of **FAIR data principles** (Wilkinson et al, 2016) aiming at making data **findable**, **accessible**, **interoperable** and **reusable,** which are similar and complementary to the OGD principles. The principles support the development of processes, infrastructures and services that enable a systemic change of science and innovation practices. FAIR stands for data to be:

1. **Findable**: data and supplementary materials should have sufficiently rich metadata and a unique and persistent identifier
2. **Accessible:** metadata and data must be understandable to humans and machines and data must be deposited in a trusted repository.
3. **Interoperable:** data must use a formal, accessible, shared, and broadly applicable language for knowledge representation.
4. **Reusable:** data and content must have a clear usage licenses and provide accurate information on provenance.

The European Commission's Open Research Data (ORD) Pilot as well as the European Open Science Cloud (EOSC) apply these principles to encourage funders and researchers to ensure that their data is soundly managed and subsequently shared[13].

[13] European Commission. (2016). Guidelines on FAIR Data Management in Horizon 2020.

Open science practices are opening a wealth of information and data (publications, software, research data) identifying and describing research and innovation actors, activities and results, which can be harnessed to support S3 analysis, design, monitoring and evaluation, particularly of the more science-oriented policy instruments.

2.3. Open Innovation

> "The basic premise of Open Innovation is **to open up the innovation process to all active players so that knowledge can circulate more freely and be transformed into products and services** that create new markets, fostering a stronger culture of entrepreneurship."
>
> Open Innovation, Open Science, Open to the World
> European Commission, 2016

The concept of Open Innovation (Chesbrough, 2003) is a founding element of the collaborative types of relations that are monitored in S3 (i.e. publicly supported projects, patenting, clinical trials in hospital sites) and is also very relevant for the identification of innovative actors and resources in the public (infrastructures, experts, research and technology transfer groups, etc.) and private sector (large companies, SMEs, technology vendors, engineering companies, technology brokers).

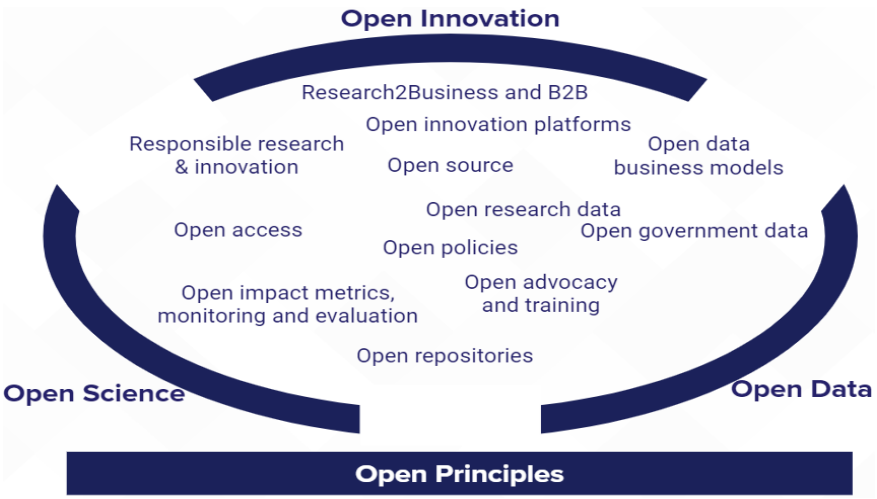
Protection, commercialisation and sharing of intellectual property (IP) is at the heart of open innovation, and fortunately, except for trade secrets, IP is documented publicly since centuries ago. Lately, public IP databases are growing in coverage, openness and services, and are increasingly available for complex S3 analysis and monitoring (such as identifying and monitoring the patent production linked to a S3 priority in a region). Other public data sources, such as registries of clinical trials or the contestants and proposals in open innovation challenges, are also available for transversal or vertical S3 analysis and monitoring.

Open innovation, as a general approach, can also be useful to guide practices and projects in S3 monitoring. S3 monitoring should not be restricted to in-house data, resources and people, but be open to using diverse data and qualitative information, and benefit from the expertise and skills across government departments and agencies, academic institutions, ecosystem stakeholders (particularly private companies, clusters, networking platforms, think tanks) and international partners.

2.4. Integration of the frameworks

It is evident from the subsections above that open (government) data, open science and open innovation share a set of principles and practices, and overlap significantly in processes, data types and data platforms. Integrating the three framework therefore provides interesting ways of thinking about S3 monitoring, as shown in the remaining of the report.

Figure 3 Integrating open data, open science and open innovation



Source: SIRIS Academic

2.5. Open Data in the European Union: evolving legislation

The importance of open data for society and for the economy is well reflected by all the discourse on the Digital Single Market (e.g. Scott et al., 2018) and, more recently, in the Open Data Directive (Directive (EU) 2019/1024), which entered into force on 16 July 2019[14]. The measure, in continuity with the previous Public Sector Information Directive (Directive (EU) 2003/98/EC), encourages the Member States to **make as much information available for re-use as possible (see focus box below)**. A particularly interesting development, for our purposes, is that it also covers research data resulting from public funding. Member States will be asked to develop policies for open access to publicly funded research data and new rules will also facilitate the re-usability of research data that is already contained in open repositories.

Needless to say, the use of ODSI for S3 and S3 monitoring **poses specific difficulties** in terms of skills, infrastructures and governance, which make it difficult for policy-makers to use it. Not only new skills are needed to use existing open-data, but also an adequate data infrastructure needs to be in place for managing open-government data across the public administrations with a stake on S3, in turn this generates questions of governance of such data and such infrastructure. The ODSI for S3 project, without aiming at being exhaustive in addressing these issues, offered a space to explore them and started placing them in the S3 policy cycle.

Focus : The Open Data Directive

The Directive on open data and the re-use of public sector information, also known as the '**Open Data Directive**' (Directive (EU) 2019/1024) In continuity with the previous Public Sector Information Directive, it encourages the Member States to **make as much information available for re-use as possible**. It addresses material held by public sector bodies in the Member States, at national, regional and local levels, such as ministries, state agencies and municipalities, as well as organisations funded mostly by or under the control of public authorities (e.g. meteorological institutes). Member States have to transpose Directive (EU) 2019/1024 by 16 July 2021.

Once fully transposed on the national level, the new rules should[15]:

- Stimulate the publishing of dynamic data and the uptake of Application Programme Interfaces (APIs).
- Limit the exceptions which currently allow public bodies to charge more than the marginal costs of dissemination for the re-use of their data.
- Enlarge the scope of the Directive to:
    - data held by public undertakings, under a specific set of rules. In principle, the Directive will only apply to data which the undertakings make available for re-use. Charges for the re-use of such data can be above marginal costs for dissemination;
    - research data resulting from public funding – Member States will be asked to develop policies for open access to publicly funded research data. New rules will also facilitate the re-usability of research data that is already contained in open repositories.
    - Strengthen the transparency requirements for public–private agreements involving public sector information, avoiding exclusive arrangements.

In addition, the Open Data Directive requires the adoption by the Commission (via a future implementing act) of **a list of high-value datasets to be provided free of charge**. These datasets, to be identified within a thematic range described in the Annex to the Directive, have a high commercial potential and can speed up the emergence of value-added EU-wide information products. They will also serve as key data sources for the development of Artificial Intelligence.

[14] https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024
[15] https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information

3. Open data sources for S3 monitoring: a taxonomy and some examples

A large amount of national, European and international data sources or data platforms is available (and growing). This information could, at least partially, support S3 monitoring, yet. it is extremely difficult to navigate and there is large variability in their degree of openness and usability. To make sense of these sources, we propose a heuristic taxonomy based on the data origin, purpose, types of data they store and types of services they provide.

In particular, it is useful to distinguish among:

- **Information systems and platforms on publicly funded research & innovation projects and activities**
- **Registries of RDI actors and networks**
- **Knowledge production, protection and dissemination**
- **Statistics and analysis in science and innovation, the economy and sustainability**

The aim of the taxonomy is to point to how these data sources can support S3 monitoring, providing some examples and descriptions of their characteristics and highlight the main challenges to their usability. [16]

3.1. Information systems and platforms on publicly-funded R&I activities

The macro-category *Information systems and platforms on publicly funded R&I activities* consists of:

- R&D and innovation grant-managing bodies' information systems
- European Structural and Investment Funds (ESIF) monitoring information systems covering R&I
- Databases of public investments in R&I
- National registries of public R&I subsidies to private companies

The list below presents a selection of useful European and international sources in this macro-category (link in the name):

✓ **CORDIS** - The EC's Community Research and Development Information Services contain the Research and innovation projects funded by the EC Framework Programmes. From CORDIS, it is possible to extract the partners, collaborations, textual content, taxonomies and funding volume of projects in the "lagging regions", amongst other information.

✓ **Eureka** - EUREKA finances industry-lead research and development projects, providing a valuable window into market-oriented innovations, very frequently including collaboration with knowledge providers.

✓ **Keep** - Database of Interregional R&I projects. The European programme INTERREG finances interregional research and innovation projects through their Axis 1, which typically involve local authorities and stakeholders less prone to participate and have visibility in other policy instrument and sources.

✓ **Creative Europe** - European Commission's framework programme supporting the cultural and audiovisual sectors. It contains cross-border cooperation projects between cultural and creative organisations within the EU and beyond, a relevant sector in many S3 strategies which is usually underrepresented in traditional science and innovation instruments and sources.

✓ The **European Investment Bank Database**, contains financial operations (including innovation and industrial projects) across Europe and beyond. It doesn't have the level of granularity necessary for S3 monitoring (it is not regionalised and doesn't provide text content), But it could be used for more general monitoring and comparison at the national level.

The Annex "***Reference initiatives and platforms***" describes four examples in this category:

- RIS3-MCAT – Catalonia Smart specialisation mapping platform (Spain)
- OpenCoesione (Italy)
- Flanders Research Information Space (FRIS, Flanders)
- Gateway to Research (UK)

The sources in this category are generally open and contain updated information on publicly funded programmes and projects, covering the whole range of RTDI actors (public and private). However, in many occasions, single initiatives are duplicated in more than one source[17].

[16] The taxonomy is based on 71 data sources, of which 29 European or international and 42 from the participating countries. The data sources were identified through desk-research, interviews and workshops.

[17] This would be the case, for instance, ,for an R&D grant given to a private company, managed by a funding body, co-financed by a ESIF which would be recorded in all the sources above, likely with different metadata

There is no widespread data scheme or ontological standard, nor single identifier services, for R&I grants/projects (unlike for people and organisations, see subsections below). However, there are relevant initiatives, such as:

- ✓ **360º giving** "supports organisations to publish their grants in an open, standardised way and helps people to understand and use the data in order to support decision-making and learning [...]" - 360º is gaining support in the United Kingdom's government and charity sector.
- ✓ **Crossref's Grant Identifier Metadata Schema** "will be used to register persistent identifiers (DOIs) for grants and other funding through Crossref", and complements with their service to register grants.

3.2. Registries of RDI actors and networks

The macro-category *Registries of RDI actors and networks* consists of:

- Databases / Registries of R&D infrastructures and equipment
- People and skills
  - Single identifiers / metadata repositories of individuals
  - Databases of personnel/experts
- Registries of organisations and networks (industry, science and innovation)
  - Single identifiers / metadata repositories of organisations
  - Inventories of STI and industry organisations
  - Inventories of STI networks
  - Company / business registries

These sources contain lists of people, organisations and networks. They are useful to monitor S3 provided that the actors, networks and infrastructures can be integrated with each other and with data on RDTI investment and activities. They include repositories held by different organisations for their own purposes (i.e. universities' staff directories, lists of innovative companies) and vary in richness and openness (for instance national company / business registries are often closed, making it hard and expensive to access, also for public administrations and bodies). Even when the data is fully published, they may lack interoperability and require costly disambiguation procedures.

Luckily for S3 monitoring, to address these issues, the Open Science community has developed standard persistent identifiers and metadata repositories of individuals, like **ORCID**, and organisations, like **ROR** (Research Organization Registry). Also international metadata standards or ontologies can be useful when designing new platforms or integrating existing data sources, like the **W3C's Organization Ontology** or **FOAF**, a machine readable ontology describing people, their products and their relations. Extending the four references of the previous paragraph, the list below presents a selection of useful European and international sources in this macro-category (link in the name):

- ✓ **ECCP cluster organisations mapping tool** of the European Cluster Collaboration Platform
- ✓ **IASP** - List of the International Association of Science Parks and Areas of Innovation
- ✓ **OpenCorporates** - "The largest open database of companies in the world"

Business registries deserve a special mention in this contex.t. Company information is essential for project, instrument and policy monitoring and evaluation. National company/business registries gather and manage a lot of this data, but are not necessarily, making it hard and expensive to access, even for public administrations and bodies. Very different practices, organisational arrangements and levels of openness coexist in the world, and also in the case study countries.[18]

The Annex "***Reference initiatives and platforms***" describes two examples in this category:

- Danish Company Register **(datacvr.virk.dk/data/** - Denmark)
- OpenCorporates

3.3. Knowledge production, protection and dissemination

The macro-category *Knowledge production and dissemination* consists of:

[18] The Open Company Data Index, maintained by OpenCorporates, in partnership with the World Bank Institute, provides an overview of the openness of country company registers. It measures the following openness features: Unrestricted online search (no cost, no registration, search feature); Openly licensed; ; Free machine-readable data; Data depth: Directors, Annual accounts, Shareholdings"

- Open science infrastructure and repositories - containing information on publications, research artefacts (particularly software and databases), authors, research institutions, collaborations, citations, etc.
- Clinical trials registries
- Intellectual property databases
    - Patents
    - Trademarks and Designs

These data sources inform on the advancement of scientific knowledge production and its dissemination, as well as of clinical research and intellectual property, two categories where the obligation of rich metadata publication is particularly useful for S3 analysis and monitoring purposes. The list below presents a selection of useful European and international sources in this macro-category (link in the name):

Open science infrastructure and repositories:

✓ **Crossref** - "Crossref makes research outputs easy to find, cite, link, assess, and reuse."

✓ **OpenCitations** – "is an independent infrastructure organization for open scholarship dedicated to the publication of open bibliographic and citation data using Semantic Web (Linked Data) technologies, and engaged in advocacy for semantic publishing and open citations."

✓ **OpenAire** - "Our mission [is to] shift scholarly communication towards openness and transparency and facilitate innovative ways to communicate and monitor research."

- **Zenodo** - Developed in partnership by OpenAire and CERN, it is "a catch-all repository for research [...] data, software and other artefacts in support of publications, [...] materials associated with the conferences, projects or the institutions themselves, all of which are necessary to understand the scholarly process."

✓ **Github** - "GitHub is a development platform [...] you can host and review code, manage projects, and build software [...]."

Clinical trials registries:

✓ **ClinicalTrials.gov** - "is a database of privately and publicly funded clinical studies conducted around the world."

✓ **OpenTrials** - "aims to locate, match, and share all publicly accessible data and documents, on all trials conducted, on all medicines and other treatments, globally."

✓ **EUDRACT** - "(European Union Drug Regulating Authorities Clinical Trials Database) is the European database for all interventional clinical trials on medicinal products authorized in the European Union (EEA) and outside the EU/EEA if they are part of a Paediatric Investigation Plan (PIP) from 1 May 2004 onwards."

Intellectual property databases

✓ **EPO's EPAB** - "European patent applications and specifications database of the European Patent Office."

✓ **OECD IP Data** - Intellectual property (IP) statistics and analysis provided by the OECD.

✓ **EUIPO** - "EU trademarks and the registered community designs in the European Union Intellectual Property Office."

To store, manage and publish information on knowledge production and dissemination, funding bodies and institutions rely on **Current research information systems (CRIS)**, also known as Research information management systems (RIMS). In CRIS, researchers and administrators:

- register data and metadata on institutional units (departments, groups, etc), people, projects, research results (such as publications or research data)
- connect this data with external repositories and identifiers, providing access or enriching metadata (for instance the European and international platforms presented above)
- make this data available through public platforms
- generate registries, reports, CVs, etc.
- are able to exchange data, in an interoperable fashion, through data exchange standards.

CERIF (Common European Research Information Format) is a standard of CRIS data architecture, which facilitates interoperability between institutions and systems

The Annex "***Reference initiatives and platforms"*** describes thee examples in this category:

- Flanders Research Information Space (FRIS, Flanders)
- Gateway to Research (UK)
- Open Citations

3.4. Statistics in science and innovation, the economy and sustainability

The macro-category *Statistics in science and innovation, the economy and sustainability* consists of**:**

- Science and innovation sector statistics and analysis
- Venture capital and start-ups data, statistics and analysis
- Economy, trade, industry and employment statistics databases
- Sustainability / Societal challenges / Sustainable Development Goals targets and statistics databases

Although this macro-category is not strictly the object of the current work, traditionally, S3 analysis, design and monitoring has heavily depended on official and ad-hoc statistics in the area of science and innovation, the economy and sustainability (i.e. Eurostat). Official statistics are most generally published openly, with increasing quality in terms of granularity, metadata definition and formats of publication. Nevertheless, regional and national stakeholders have expressed their frustration with the accessibility and geographic and taxonomical granularity of these data (especially for fine-grained S3 analysis and monitoring). Non-official statistics and ad-hoc analysis, on the contrary, are usually not reusable, either due to lack of access to the underlying data or to lack of replicability.[19]

[19] In this context, it is worth introducing the **Statistical Data and Metadata eXchange (SDMX)**, sponsored by seven institutions (the BIS, the ECB, Eurostat, the IMF, the OECD, the UN and the World Bank) which aims at standardising and modernising the exchange of statistical data and metadata among international organisations and their member countries.

4. From theory to practice I: analysing ODSI for S3

Whilst previous chapters of the report have provided a conceptual and policy introduction to the issue of ODSI for S3, in the remaining we delve into the process of analysing ODSI for S3, with the aim of identifying challenges and build related roadmaps (as described in chapter 5).

As described in the methodology, in order to prepare the roadmaps two broad steps were pursued.

- The identification of the three most relevant S3 instruments to be monitored and the related indicators and data to populate them (discussed in 4.1);
- The identification of the data sources that could build these indicators and assess their suitability and openness (discussed in 4.2).

This chapter describes the tools applied in these steps, and their outcomes. In particular, the data-suitability and openness grids presented in section 4.2 can be seen as a tool to be adapted and reapplied for similar studies.

4.1. Priority instruments and indicators for S3 monitoring

The first workshop conducted in May 2019 fostered participatory discussions to identify (a) the **three most important policy instruments for S3 monitoring** (within the STIP Compass taxonomy)[20] and (b) their most relevant indicators and monitoring information.

Based on these discussions and a collective voting activity, the three major policy instruments appeared to be:

- **Grants for business R&D and innovation** (STIP code 1.3)
- **Technology transfer and business advisory services** (STIP code 3.1)
- **Clusters and other networking and collaborative platforms** (STIP code 4.1)[21]

For each one of these three instruments, the working group discussed the set of **indicators, requirements and expectations** that contribute to define the type of open data that could possibly support them.

4.1.1. Monitoring Grants for business R&D and innovation: indicators, requirements and expectations

As reported by the actors, efficient monitoring of the policy instrument *Grants for business R&D and innovation* should provide information on:

- **topic of the project;**
- **volume spent;**
- **actors** involved;
- **technological readiness level** at the start and end of the funded project[22]**;**
- **economic/financial evolution of the company;**
- **socio-economic data**, to measure the impact of the policy on employment, equality and society more broadly.

This information is not currently easily available in several participating lagging regions/member states, in particular because it is easy to ask firms for information at the beginning of the project, but less easy (though not impossible) to do it in the middle or at the end of the grant. In addition, it is important to **disaggregate grants by beneficiary**, and ideally by **location of the expenditure**. Lastly, it is necessary to **aggregate projects by S3 priority/ S3 output / S3 result indicator data**. In fact, this is rarely the case, in particular because such assessment may require qualitative data -

[20] https://stip.oecd.org/stip.html

[21] The voting process revealed the following ranking: 4th Innovation vouchers (STIP code 1.9); 5th Dedicated support to new research infrastructures (STIP code 4.2); 6th Policy intelligence (e.g. evaluations, reviews and forecasts) (STIP code 5.3); 7th Creation or reform of governance structure or public body (STIP code 5.2); 8th Project grants for public research (STIP code 1.2); 9th Formal consultation of stakeholders or experts (STIP code 5.4); 10th Institutional (basal) funding for public research (STIP code 1.1)

[22] Note that TRLs can be self-reported or evaluated by experts.

expert advice may be required. For **ex-post classification**, the presence of **rich text fields** in project databases is key (titles, abstracts, activities). This involves **Natural Language Processing techniques** to deal with large amounts of data.

4.1.2. Monitoring Clusters and other networking: indicators, requirements and expectations

As reported by the actors, efficient monitoring of the policy instrument *clusters and other networking s*hould provide information on:

- **identification of participants -** this information is easier to get/access if clusters receive subsidies from the region. Note that participant lists should have metadata for disambiguation / interoperability;
- **connections and types of collaboration** (new joint products/services in the cluster, patent applications/joint applications, training, joint R&D projects);
- **flow of information and support** amongst companies;
- **evolution of economic indicators** for the cluster members (sales, exports, margin, employment, etc.). This information can be asked for in application forms but for real follow-up, it would be necessary to connect with a Registry;
- the cluster **capacity to grow at the international level**;
- the cluster capacity to create/participate in **inter-cluster and interregional/international networks:** cluster connections within the region and with other regions.;
- **learning outcomes of the support to clusters**;
- **impact**, for which access to data is not enough: modelling is needed.

4.1.3. Monitoring Technology transfer services: indicators, requirements and expectations

As reported by the actors, efficient monitoring of the policy instrument *Technology transfer services* should provide information on:

- **mechanisms of transfer**: spin-offs, start-ups, contracts, IP licence agreements, informal mechanisms (meetings, non formalised collaborations, etc.);
- **networks**: who is collaborating with whom;
- **topics** to which the TT activities or results belong;
- **funding** of technology transfer (public and private);
- **effects or impact** of technology transfer. For instance: after receiving the support, did the business model change? Were new products, services, processes or new organisations created? Were new marketing approaches adopted? Indicators on technology transfer should relate to EU societal values (inclusivity, employment, etc.), beyond just innovation and competitiveness related metrics.

Participants expressed significant difficulties in monitoring technology transfer services. Most of the data available comes from **business demography statistics** where **innovativeness can only be estimated by proxy** (for instance, from NACE sectors); the inability to monitoring TTS is further hampered by the lack of **standard definition of start-up nor spin-off**; **data availability.** Moreover, participants highlighted the lack of **data on informal activities** (i.e. fairs, events, etc.). Last but not least, the discussion revealed significant variability among countries.

4.2. Identifying data-sources, and assessing data suitability and openness

In order to enable participants to understand the intricacies of data and data sources and open data principles, it was decided to analyse two main dimensions:

1. the **suitability** of data for S3 monitoring: i.e. the availability, granularity, quality and relevance of the data and metadata needed for appropriate S3 monitoring.

2. the **openness** of data: i.e. how easy it was to access, download, exploit and integrate the required data and metadata.

The analysis was carried through two corresponding grids, i.e. the ***data-suitability grid*** and the ***data-openness grid***, which identified the baseline requirements against which to assess other data sources.

Project participants were asked to build a **long list of interesting data sources**, normally ranging 10 to 15 sources per country**, out of which country experts selected 3 or 4 per country.** The grids were used to analyse such sources first individually (in preparation for workshop 2) and then in a participatory setting (during workshop 2).

The grids are reported here as a useful tool to reuse and reapply in other contexts.

4.2.1. Data suitability grid

The "S3 monitoring data suitability analysis grid", presented below, is structured in three sections, covering the baseline requirements on data and metadata regarding:

- RTDI activities, actors and taxonomies;
- RTDI results and impact;
- data granularity requirements.

DATA SUITABILITY ANALYSIS GRID	
Activities, actors and taxonomies	
Dimension	**Elements to assess** (has de data source the following data or metadata?)
Policy instrument / Call	ID / unique identifier Title Objectives / description [text]
R&D grants	ID / unique identifier Start and end date / Duration of the project Title Description / Objectives / Abstract [text] Taxonomic information - S3 priorities Taxonomic information - Other taxonomies (sector, technology, challenge, etc.)
R&D contracts (between academic partners and private companies)	ID / unique identifier Start and end date / Duration of the project Title Description / Objectives / Abstract [text] Taxonomic information - S3 priorities Taxonomic information - Other taxonomies (sector, technology, challenge, etc.)
Value and public subsidy of the project, contract or activity	Total value (or cost, or investment - amount in currency) Public subsidy (amount in currency or %)
Cluster and networking or collaborative platforms	Name Description / Objectives / Abstract [text] Identification of the cluster/network/platform managing organisation Taxonomic information - S3 priorities Taxonomic information - Other taxonomies (sector, technology, challenge, etc.)
Beneficiary(ies) / Agent(s) / Participants	Vat number or equivalent Legal name Localisation Region Typology of institution/organisation Sector (NACE) and/or other taxonomic information Dimension (e.g. number of employees)

Results and impact	
Dimension	**Elements to assess** (has de data source the following data or metadata?)
New IP creation and protection (individual or joint between academic partner and private company)	Patent registration number Trade mark or Design registration number Taxonomic information - S3 priorities Taxonomic information - Other taxonomies (sector, technology, challenge, etc.)
New innovative companies (spin-offs, start-ups)	Company identification (Name, VAT, etc.) Year of foundation Shareholder information Capital raised per investment round Revenue Full time employees / headcount Taxonomic information - S3 priorities Taxonomic information - Other taxonomies (sector, technology, challenge, etc.)
Impacts (innovation, economic and/or employment evolution)	New products, services, processes, social innovations Technology readiness level Revenue Exports Full time employees / headcount Learning outcomes

Data granularity requirements	
Dimension	**Elements to assess**
Data granularity	Is the information above available for each R&D grant / technology transfer activity or cluster activity?
	Is the information above available for each beneficiary, agent or participant?
	Is the information above available every year/monitoring period?

4.2.2. Data openness grid

The data openness analysis grid, presented below, is based on the Open Government Data principles presented in Chapter 2, with some specific questions facilitating assessment.

DATA OPENNESS ANALYSIS GRID	
Dimension	**Elements to assess**
Primary **Complete**	Is data collected at the source?
	Is all non-confidential data made available?
Timely	Date of the most recent data
	Frequency of updating
Accessible **Machine processable** **Interoperable**	Can the data be downloaded?
	What is the format of the data / database / access point (pdf, xls/xlsx, csv, xml, sql, API, SPARQL, etc.)
	Is metadata documented?
	Is the metadata based on some standard?
	Is metadata machine processable?
	Is there data enriched with external sources?
	Does the data connect to external data or metadata repositories?
	Is the data deposited in other trusted repositories? Which?
Findable	Has the data a persistent identifier?
Licensed	Is the data source licensed? What type?
	Does this license permit people to use the data in any way they want, including transforming, combining and sharing it with others, even commercially?

5. From theory to practice II: country roadmaps

The material discussed in section four, complemented with further country-level collective discussions in Workshop 2, was used to develop a set of country-specific lists of main challenges**.** Such list fed into Workshop 3, where project participants from the same country collectively defined **roadmaps** to address one or two of the challenges identified. The participatory road-mapping was guided by the "canvas" presented below and based on the Theory of Change (UN, 2017).[23] The following sections report the results of this exercise.

Roadmap	COUNTRY
Challenge	**Goal**
*During the previous activity we identified a set of challenges. Please **select one** which has some impact in the data and policy ecosystem. Alternatively, you can merge some interconnected challenges to address wider limitations to ODSI for S3*	*An identified Challenge leads to a transformation Goal describing a desired scenario, **in which the selected Challenge has been overcome.***
Actions	**Intermediate outcomes**
*Once we have the goal identified, it is time to **define the actions needed to achieve it.** Given the current challenges, which are the action that will lead to the final goal? Which are the intermediate outcomes of these actions? Which is the causal relation between the action and their intermediate outcomes?*	
Involved actors & enablers (Possible support by the EC and third actors)	
*Who will be **in charge of performing** each action? What support would be useful from the EC or from other actors in the region / Member state?*	
Assumptions about the actors	
*Which are your **assumptions of the actors involved in the actions**? Which is their role? Do they have the needed **mandate** and **capacities** to perform the action?*	
Inspirational examples (Case studies or other)	**Use or connection with EU and global standards & infrastructures**
*Given the **identified goal** (desired scenario) identify some relevant examples*	*Identify the relevant **EU/Global infrastructures or standards** that could be used to achieve or improve the identified goal (desired scenario)*

Disclaimer:

The roadmaps below exclusively reflect the exercises conducted within the ODSI for S3 workshops. They **do not** represent an official position of member states and regions and do not imply an obligation for conducting stated activities.

[23] The Theory of Change is a method for planning interventions taking into account complex factors and layers that are embedded in policy, institutional and economic ecosystems.

5.1. Bulgaria

Analysed data sources

Out of the diverse national data sources proposed by stakeholders and experts, and extended through research, three sources were selected for further analysis:

1. Information System for Management and Monitoring of EU Funds in Bulgaria - UMIS 2020[24]
2. National Innovation Fund (NIF)[25]
3. National Science Fund (FNI)[26]

UMIS 2020 is the only available database in Bulgaria that offers granular information on R&D&I funding. Most of the information is restricted to registered users, decreasing the level of public transparency and undermining additional positive intermediate and final uses and outcomes. Granularity in data collection loses its main impact if its accessible only to a limited public.

Two of the identified sources, namely the National Science Fund and the National Innovation Fund lack granular data on R&D&I activities and results and do not include data following the S3 taxonomical nomenclature (S3 priorities). Furthermore the information is not available to the different innovation actors and stakeholders. [27]

[24] http://2020.eufunds.bg/en
[25] https://www.sme.government.bg/en/
[26] http://www.fni.bg/
[27] The roadmaps for Bulgaria were not developed in a participatory setting, rather they were developed by the country expert Mr Ruslan Stefanov.

Roadmap 1

BULGARIA

Challenge	Goal
UMIS 2020 is the only available database in Bulgaria that offers granular information on R&D&I funding. Most of the information is restricted to registered users, decreasing the level of public transparency and undermining additional positive intermediate and final uses and outcomes. Granularity in data collection loses its main impact if its accessible only to a limited public.	• Offer accessibility to the full set of UMIS 2020 information and data to all actors and stakeholders. • In particular, ensure the usability of open data for research. purposes and policy assessment – open all the data (which is not personal) to researchers and policy analysts.

Actions	Intermediate outcomes
Provide all users with access to non-confidential data and information from the full database, including both front and backend use	Inform the public and interested stakeholders of the opportunities of using the database with its full functionality.
	Offer infographics and simplified information.
Include information about organisations profile, clusters description, S3 priorities, NACE, outcomes and results.	Monitor expenses and provide opportunities for better policy assessment.
Allow registration and data collection from users in order to generate metadata.	Become a tool for Policy-Design.
	Guarantee availability of data and other information for academic and private researchers (e.g. business analysts).
	Understand the needs of users and of the innovation ecosystem.

Involved actors & enablers (Possible support by the EC and third actors)
• Central Coordination Unit at the Council of Ministers of the Republic of Bulgaria • Council of Ministers public outreach unit. • Operational programmes' Management bodies.

Assumptions about the actors
• Political and institutional commitment will be required from the Central Coordination Unit, which coordinates EU and international funding programmes in Bulgaria. • To guarantee the process and results, implement quality audits based on best practice regarding open data.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
FRIS Research Portal. Europe en France. OpenCoesione.	www.Keep.eu.

Roadmap 2

BULGARIA

Challenge	Goal
Two of the identified sources, namely the National Science Fund and the National Innovation Fund lack granular data on R&D&I activities and results and do not include data following the S3 taxonomical nomenclature (S3 priorities). Furthermore the information is not available to the different innovation actors and stakeholders.	Enable both Funds to adopt S3 taxonomy and monitor R&D activities in a standardised way. Make the information accessible through a database (to avoid retrieving scattered information across the website), while using metadata and including external links to other databases.

Actions		Intermediate outcomes
Design and develop a common database easily accessible in English and Bulgarian for both Funds	▶	Gather the currently existing information in one single platform, easily accessible through the websites of the two funds. Standardised procedures for data gathering and data access introduced.
Establish standardised procedures for data gathering and data access (data should be retrieved in CSV or SLSX format, instead of currently in PDF)	▶	Possibility to aggregate both sources of information across years and instruments.
Enrich the data with external sources and metadata	▶	Have a less segmented view of the innovation process
	▶	Enable benchmarking across regions and specialisation domains (sectors following standard classifications, and other available taxonomies).

Involved actors & enablers (Possible support by the EC and third actors)

- Ministry of Economy & Ministry of Education of Science - the two ministries will need to ensure joint support for the informational transformation of the data in their domain; they could also sign a protocol of agreement outlining the procedures needed to provide sustainable data cooperation.
- NSF and NIF data and system administrators (the National Science Fund and the Agency for SME Promotion, respectively).
- The EU (JRC) could help by providing suitable taxonomies and advising on the most adequate ways for data collection

Assumptions about the actors

- Both data and system administrators understand the need for accurate metrics and indices for improving the governance of the instruments. They have the mandate to carry out the envisaged action but could need some assistance for improving data management capacity.
- Possibility to transfer EU / Global standards from most advanced databases and platforms
- Quality audit available based on best practices regarding open data

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
FRIS Research Portal	EU S3 Taxonomy as in the following example: ● CORDIS ● H2020 dashboard ● COSME data hub

5.2. Croatia

Analysed data sources

The four data sources selected for further analysis are:

1. Croatian Agency for SMEs, Innovations and Investments (HAMAG BICRO)[28]
2. Ministry of Economy, Entrepreneurship and Crafts (MEEC)[29]
3. Ministry of Science and Education (MSE)[30]
4. Ministry of Regional Development and EU Funds (MRDEUF)[31]

They can be described **as "Publicly funded research and innovation projects and activities' information systems"** covering the following types of policy instruments:

- Grants for business R&D and innovation
- Technology transfer and business advisory services
- Clusters and other networking and collaborative platforms

[28] http://www.investcroatia.hr/
[29] https://www.mingo.hr/
[30] https://mzo.gov.hr/
[31] https://strukturnifondovi.hr/

Roadmap

CROATIA

Challenge	Goals
Data on all projects funded by the Cohesion policy, including RTDI projects, are published through https://strukturnifondovi.hr/. However, some key dimensions for S3 monitoring are missing, i.e.: which instrument and call projects are funded, S3 priorities and NACE codes (for actors and for economic or commercial results), more information about the involved private companies would be beneficial (for instance, dimension brackets, to distinguish larger companies from SMEs).	1. Connect all this information 2. Improve/extend the information in the current web (strukturnifondivi.hr), link it to the funding agencies websites and to the European Data Portal 3. Improve usability and uptake, to facilitate discovery and partnerships, in particular for future applicants, to identify similar actors and ongoing projects and link them with existing thematic groups and collaboration networks.

Actions		Intermediate outcomes
Improve sectoral/taxonomic information ● Find the NACE code for every beneficiary of the projects and include it with the existing information. ● Incorporate in future calls the obligation for applicants to include relevant taxonomies, in their project proposals.	▶	Richer data on actors: ● Beneficiaries' metadata enriched with the NACE code to facilitate sectoral analysis ● A proposal for new fields in project application forms, to gather relevant data.
Check the legal conditions for data publication included in the contract agreements with project beneficiaries.	▶	Guarantee that relevant, non-confidential information can be published. When unnecessary limits are identified, propose revised contract drafts.
Design and develop improved and new ways of publishing the data.	▶	1. Publication of the richer metadata at strukturnifondivi.hr. 2. Publication in the project funding bodies' websites. 3. Publication in the EU Data portal.
Build a user interface (with search engines and adequate filters) allowing for freer exploration of data.	▶	Improved visibility, transparency, networking, monitoring capabilities, for policy-makers, policy officers and stakeholders.

Involved actors & enablers (Possible support by the EC and third actors)	
Owner and manager of strukturnifondovi.hr: ● Ministry of Regional Development and EU Funds (MRDFEU) Concerned funding bodies: ● Ministry of Economy, Entrepreneurship and Crafts (MEEC) ● Croatian Agency for SMEs, Innovations and Investments (HAMAG BICRO) ● Ministry of Science and Education (MSE)	For company metadata: ● Ministry of Justice - Registry of companies (Sudksi registar) ● RGI fina http://rgfi.fina.hr/JavnaObjava-web/jsp/prijavaKorisnika.jsp ● Digitalna komora (digital chamber): https://digitalnakomora.hr/home ● Croatian Buerau of Statistics – data on RDI investments and similar data from CBS.
Assumptions about the actors	
1. They are all granting bodies (Proposing all the actions). 2. The MA is ultimately responsible for data collection and publication. The MA, who is the owner of struktunifondovi.hr must be willing to improve and extend the data and give the approval for changing the rules for action 1b and 3.	Interested and engaged in contributing with policy-relevant metadata on Croatian companies.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
OpenCoesione, FRIS	European Data Portal

5.3. Greece

The following three sources were selected for further analysis:

1. Ministry of Development & Investments - Anaptyxi[32]
2. Ministry of Rural Development & Food[33]
3. Ethniko Kentro Tekmiriosis (EKT). National Documentation Centre - EKT[34]

Two of the **Greek data sources** analysed in the "Data suitability and openness analysis" can be described as **"Publicly funded research and innovation projects and activities' information systems"**:

- **Anaptyxi.gov.gr**, the ESIF management and monitoring information system of the **Ministry of Development and Investments**, which makes available the information of ERDF and Cohesion funded science and innovation grants.
- **data.agrotikianaptixi.gr** run by the **Ministry of Rural Development & Food official website;** the ministry manages the **European Agricultural Fund for Rural Development,** which funds, amongst other objectives, science and innovation grants.

The remaining source, the **National Documentation Centre (EKT)** can be classified as an institution **responsible for the "Compilation and provision of economic, scientific and innovation data, statistics and analysis".** In particular, **EKT compiles data, analyses, elaborates and publishes about Greek actors' activity and performance in European projects and on bibliometric indicators**, also at the regional level[35].

[32] http://anaptyxi.gov.gr/en-us/
[33] http://data.agrotikianaptixi.gr/
[34] https://metrics.ekt.gr/en
[35] EKT is also the Greek agency responsible for scientific data stewardship and digital services and for open science. As such, they are a key stakeholder in European open science and open data initiatives like the European Open Science Cloud. EKT can also be classified as responsible for "Open science infrastructure and repositories".

Roadmap 1

GREECE

Challenge	Goal
Data associated to publicly funded RDI projects and activities are found in silos within the central Information System of State Aid. Those data could be made available to policy makers and/or broader user groups contributing to better monitor key output indicators.	Take all the necessary technical and political steps in order to make non-confidential data publicly available in an easy to use manner.

Actions		Intermediate outcomes
Promote conceptual uniformity between data categories in order to achieve taxonomic alignment. Audit of data quality and reusability.	▶	Immediate result is a thorough gap analysis that will lead to a list of necessary actions to be undertaken in the next stage.
Revision of data collection procedures (or data inputs by primary users).	▶	This will lead to form a guide of new improved and more harmonized data input processes in different information systems.
Design and execution of a pilot project to test all of the above.	▶	This will feed the working team with valuable information, therefore continue the loop of transformation to a really open data system.

Involved actors & enablers (Possible support by the EC and third actors)	
1. Ministry of Development & Investments: ● General Secretariat of Research & Technology (GSRT) ● National coordination Authority of NSFR (or data inputs by primary users).	2. Ministry of Digital Governance: ● National Documentation Centre (EKT) ● Tax Registry (AADE)

Assumptions about the actors
Institutional commitment of all involved actors is a key prerequisite for the successful transition to the open data era.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
VIRK, Denmark Gateway to Research, UKRI, UK OECD/TIP	International standards and interoperable taxonomies are key issues that will be taken into account.

Roadmap 2

GREECE

Challenge	Goal
Lists of organisations (companies as well as public and academic organisations) are scattered in many disconnected datasets controlled by different actors, in different versions and with bad or non-interoperable metadata. These data are necessary to support policy analysis as well as stakeholders contributions to better monitor key result indicators.	Identify scattered heterogeneous datasets and public registries which contain data on institutions and transform them into interconnected ones. Make them open and available in a more "human" form.

Actions		Intermediate outcomes
Establish common understanding of different challenges associated to the data in question.	▶	This can be tackled by the formation of working groups with clear mandate and suitable HR.
Audit of data quality and reusability.	▶	Immediate result is a thorough gap analysis that will lead to a list of necessary actions to be undertaken in the next stage.
Revision of data collection procedures.	▶	This will lead to form a guide of new improved and more harmonized data collection processes from different information systems.
Design and execution of a pilot project to test all of the above.	▶	This will feed the working team with valuable information, therefore continue the loop of transformation to a really open data system.

Involved actors & enablers (Possible support by the EC and third actors)	
• General Commercial Registry (GEMI) • Ministry of Employment (ERGANI) • Tax Registry (AADE) • Public Procurement Authority (EAADIS) • Hellenic Statistical Authority (ELSTAT)	• Hellenic Industrial Property Organisation (OBI) • General Secretariat of Trade and Commerce (trademarks, industrial design) • National Documentation Centre (NDC-EKT)

Assumptions about the actors
• The public bodies involved in the project may focus on the technical and coordination challenges and potential limitations of the project. Try to establish a shared understanding by the involved public bodies of the opportunities for better policy-making, as well as for discovery and interconnection between actors in an open innovation framework (leading to growth), rather than technical challenges and/or limitations of the project. • Coordination and execution challenges between the involved actors and enablers could be expected. It is necessary to establish a well-tuned and functioning planning and working group to run the project.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
VIRK, Denmark Open Corporates	Alignment between technological and commercial ontologies

5.4. Hungary

Analysed data sources

The following three data sources were selected for further analysis:

1. National Research, Development and Innovation Office - NKFIH, RTDI grant database[36]
2. Ministry of Innovation & Technology grants' database [37]
3. National Research, Development and Innovation Office - NKFIH, innovation actors' database[38]
4. Hungarian Venture Capital and Private Equity Association - HVCA, VC and PE (Private Equity) statistics[39]

Two of the **Hungarian data sources** analysed in the "Data suitability and openness analysis" can be described **as "Publicly funded research and innovation projects and activities' information systems"**:

- **The RTDI grants database of Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal** (**NKFIH** - National Research, Development and Innovation Office, depending on the Ministry of Innovation & Technology)
- The **grants database of Innovációs és Technológiai Minisztérium (Ministry of Innovation & Technology)**

The innovation actors database of the National Research, Development and Innovation Office (NKFIH) lists RTDI active actors and networks, in particular:

- Organisations in technology transfer
- Existing Hungarian clusters
- Existing Hungarian technology platforms

Finally, Magyar Kockázati- és Magántőke Szövetség, **the Hungarian Venture Capital and Private Equity Association (HVCA) publishes quarterly Venture Capital and Private Equity statistics.**

[36] https://nkfih.gov.hu/for-the-applicants
[37] https://www.kormany.hu/en/ministry-for-innovation-and-technology
[38] https://nkfih.gov.hu/innovacio/hazai-innovacios/technologia-transzfer
[39] https://www.hvca.hu/EN/statistics/

Roadmap

HUNGARY

Challenge	Goal
There is no available, comprehensive and integrated data source on RTDI public funding in Hungary, to be used for both information dissemination and policy monitoring, design and evaluation.	1. Create a single, integrated, complete, 3-star and fully open database with public access and downloadable capabilities (in Excel and CSV formats) of data on all publicly funded RTDI projects in Hungary, including grants, special loans and any other financial support. 2. The purposes of the database would be the following: ● Inform the public on funding decisions, and on their results and outcomes ● Support the monitoring of government (public) interventions ● Support policy evaluation ● Support policy design and analysis ● Facilitate and standardise policy, programme and project reporting (make easy to create the necessary and requested reports to the EU (operational programs) and the national government on strategies, programs, calls, funds ● Provide data and other information for independent academic individuals and business analysts ● Generate open data-based business

Actions		Intermediate outcomes
Governmental greenlight to the database project proposal, distribution of responsibilities and tasks (including calendar and work plan) for development and operation of the database.	▶	Having a clear, transparent picture on the distribution of tasks and responsibilities (not only internally inside the government, but for the general public, in particular to the innovation and knowledge community).
Define and present general benefits, considerations and potential impacts of the project, adapted to the major stakeholder groups (businesses, both SMEs and large firms, academia, the general public, and government services, policy making).	▶	● Increasing support for the project amongst the innovation and knowledge community. ● Decreasing the resistance by potential competitors and involved actors within the government (due to apprehensions towards the transparency of budget allocations). ● Improving collaboration readiness and building new coalitions to the benefit of the project.
Definition of the data to be collected and integrated: source of the data, the owner of the data source, required policy-related metadata (strategies, funding schemes, priorities, calls, etc.). To be done with database/IT experts as well as policy officers, to guarantee efficacy and interoperability.	▶	Creating a strong basis for the design of the database schema.
Deciding on critical basic elements of the database development project: ● Project management structure (steering committee, operational leadership, budget) ● Stewardship and business model of the database service (independent agency or part of an existing one, fully free services or partly paid (baselines of the paid services, income expectations, etc.) ● Language or languages (only Hungarian, or Hungarian & English)	▶	Setting up the basic elements of the project, allowing to jump into the business fast and avoid unnecessary debates on political, rather than technical, questions.

Involved actors & enablers (Possible support by the EC and third actors)	
● Ministry of Innovation & Technology, Office of Research, Technology Development and Innovation, other ministries and	● Collaboration would be necessary with DG REGIO and the EU Fund managing services, since the national data and the data

government agencies, operational programs' authorities, knowledge institutes, Central Statistical Office.	related to the operational programs should be harmonised.

Assumptions about the actors

Main risk factors:
● Large differences among the present data source owners on the approaches, requirements and applied solutions to their data collection, database management, and downstream services provided or to be provided. All their requirements should be taken into consideration when the new system is developed.
● Poor interoperability of the existing data sources will require existing metadata analysis and database schema proposal, as well as manual curation and integration processes
● Lack of capacities at several data source managing organisations will require capacity building and protocoled data gathering and validation processes.
 	
Possible breakdown of responsibilities/roles:
● Owner of the project: Ministry of Innovation & Technology (mandate: OK, role: clear)
● Project leader (responsible for the implementation): Office of Research, Technology Development & Innovation (mandate to be given by the government, role: clear)
● Owners of the data sources:
  ○ Office of Research, Technology Development and Innovation
  ○ ESIF Operational Program authorities
  ○ Line ministries running their own RTDI-related funding programs
  ○ Knowledge institutes
● Political support: Ministry of Innovation and Technology and line ministries
● Special expertise:
  ○ Central Statistical Office
  ○ Existing grant systems' database management services

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
CORDIS	Apply international standards and create a high level of interoperability with CORDIS. In a later stage, not only the nationally funded and regional fund financed projects, but the EU framework program funded project data will also be incorporated into the planned database service.

5.5. Italy

Analysed data sources

The following three sources were selected for further analysis:

1. Opencoesione[40]
2. Infocamere's database of innovative start-ups[41]
3. Infocamere database of innovative SME (available in the same platform as above)

One of the **Italian data sources** analysed in the "Data suitability and openness analysis" can be described **as "Publicly funded research and innovation projects and activities' information systems"**:

- **OpenCoesione,** a project managed by the Department for Cohesion Policy (Evaluation Unit) at the Presidency of the Council of Ministers in collaboration with the Agency for Territorial Cohesion and the Directorate of the Ministry of Economy and Finance responsible of the National Monitoring System on cohesion resources, "is the open government initiative on Italian cohesion policy [...], it leverages on all open government dimensions and promotes an increased involvement of citizens, government, businesses, researchers and the media for a more efficient and effective use of cohesion funds."[42] On the OpenCoesione web portal users can find out information on individual projects financed by European and national funds for cohesion, follow their implementation, download open data or surf through interactive diagrams.

The two other sources, very interconnected, **are registries of innovative companies** produced by the **Registro delle Camere di Commercio italiane** (Italian Chambers of Commerce):

- Registry of **Innovative Start-ups**
- Registry of **Innovative SMEs**

The companies in this registry have to fulfill certain rules (size, innovative activity, a threshold of R&D investment and, for Start-ups, a recent date of creation) **and be certified through an administrative process.** They are listed in these registries with information coming from the general Italian Company Registry.

[40] https://opencoesione.gov.it/en/
[41] http://startup.registroimprese.it/isin/home
[42] https://opencoesione.gov.it/en

Roadmap

ITALY

Challenge	Goal
OpenCoesione has a vast amount of information on projects financed by Structural Funds, however the projects are not associated to S3 priorities and sub-priorities. Meanwhile, IGRUE (Ministry of Economy) and the Agency for Territorial Cohesion already have a methodology to associate an S3 priority and a sub priority to each project, This data is not published yet. There is a plan for the Agency to publish it, yet there is not a formalised plan to integrate it with OpenCoesione.	Integrate into OpenCoesione the project-level data on their S3 priorities and sub-priorities, produced by IGRUE (Ministry of the economy) and the Agency for Territorial Cohesion.

Actions		Intermediate outcomes
Data by S3 priority is already available (need to check errors and fill few remaining gaps).	▶	Publish and share with regions a report on S3 priorities on OT1, so that they get more engaged and ready to be involved in the future (this action should be led by the Agency for Territorial Cohesion).
The methodology for systematize and harmonize sub-priorities is being developed and experts are currently being appointed (The task requires systematising and summarising 800 sub-priorities).	▶	Publish data on S3 priority areas in OpenCoesione as soon as possible. This will serve to build trust and encourage regions to work on sub-priorities.
Two key aspects of the methodology deserve attention: ● how to treat TO3 projects, ● how to act when regions modify their S3	▶	Establish a communication procedure between the Agency of Territorial Cohesion and OpenCoesione focussed on the methodology for publication. The Agency should be responsible for the identification of sub-priorities; it should work together with OpenCoesione on ways to treat these sub-priorities and make them coherent with other taxonomies previously used in OpenCoesione. Furthermore, they should discuss ways of updating the data as sub-priorities evolve with the S3 reviews.
Whichever the final methodology, regions will need to be engaged in the process, hence the importance of motivating and involving them in the process.	▶	In the future, it could be interesting to explore how to integrate in OpenCoesione data on S3 projects funded by other funds. Alternately, one would work with the Regions to support them publishing the data in open formats.

Involved actors & enablers (Possible support by the EC and third actors)
IGRUE - Agenzia della Coesione - OpenCoesione, as well as all the public administration involved in S3 (at the regional and national level), who would need to produce the data.

Assumptions about the actors
The actors with whom the exercise was conducted represent the Agency for Territorial Cohesion and OpenCoesione. They work closely with IGRUE. They do have the mandate to initiate this discussion and the power to propose it to the management of OpenCoesione. Another important step would be to test the feasibility of the approach with Regions that need to produce the data, and hear their point of view.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
Not discussed.	Not discussed.

5.6. Portugal

Analysed data sources

Out of the diverse national data sources proposed by stakeholders and experts, and extended through research, four sources were selected for further analysis:

1. Fundação para a Ciência e Tecnologia - FCT[43]
2. National Innovation Agency - ANI[44]
3. COMPETE 2020[45]
4. PT 2020[46]

The **four Portuguese data sources** analysed in the "Data suitability and openness analysis" can be described **as "Publicly funded research and innovation projects and activities' information systems"**:

- **Two** of them are "**R&D and innovation grant-managing agencies'** information systems":
  - **Fundação para a Ciência e Tecnologia (FCT - Foundation for Science and Technology)** on scientific research and technological development.
  - **Agência Nacional de Inovação (ANI - National Innovation Agency),** on applied collaborative research, tax incentives for R&D and innovation, technology transfer and international collaboration.
- **The other two** Portuguese data sources analysed can be described **as "ESIF management and monitoring information systems"**:
  - the management and monitoring information system of the **Autoridade de Gestão do COMPETE 2020 - Programa Operacional Temático Competitividade e Internacionalização** (Management Authority of COMPETE 2020 - Operational Thematic Program Competitiveness and Internationalization);
  - the management and monitoring information system of **PT2020 - Agência para o Desenvolvimento e Coesão (FEDER+FSE+FC), Comissão de Coordenação Nacional FEADER and Comissão de Coordenação Nacional CC FEAMP.**

[43] https://www.fct.pt/apoios/projectos/consulta/projectos.phtml.en
[44] https://www.ani.pt/pt/
[45] https://www.compete2020.gov.pt/destaques/detalhe/Lista-proj-aprovados
[46] https://www.portugal2020.pt/

Roadmap 1

PORTUGAL

Challenge	Goal
Data about projects funded by the Cohesion Policy is often not accessible, and, when accessible, it is not consistent nor detailed enough for policy analysis and monitoring.	Provide open access to more detailed and comprehensive information on the projects funded by the Cohesion Policy. Mid-term goal: to have a single platform for open science and innovation, with a territorial cohesion approach.

Actions		Intermediate outcomes
Define the list of fields to be published.	▶	A list of fields to be gathered and published, based on international standards.
Make sure the fields are complying with international standards.	▶	
Introduce data fields required for smart specialisation monitoring in the project application forms, to be filled in PT and EN (e.g. summary & keywords).	▶	New, more consistent and rich application forms for the different RTDI funding agencies.
Provide information in open access (non-proprietary) format (from Excel to CSV).	▶	CSV files published with the required information (3-star open data format).

Involved actors & enablers (Possible support by the EC and third actors)	
• Political level: CIC (Interministerial Commission of Coordination). • Implementation bodies: AD&C (Development and Cohesion Agency) and Managing authorities of Operational Programmes. • Public agencies in charge of the topics addressed (FCT, ANI, CCDRs, IAPMEI, Turismo) – capacity building, getting involved.	• AMA (national agency for administrative modernization) – guidance and support, link with similar initiatives. • CIM and GAL (contractual entities) – capacity building, getting involved. • European Commission (country desk DG REGIO and other DGs involved in the OPs).

Assumptions about the actors		
CIC (political mandate for the project) ⊠ AD&C and Managing Authorities (project implementation).	◀	Support: AMA, FCT, ANI, CCDRs, IAPMEI, etc.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
OpenCoesione FRIS	

Roadmap 2

PORTUGAL

Challenge	Goal
As with RTDI Cohesion Policy-funded projects, information on projects funded by other sources (national budget, European programmes, etc.) is not available, not consistent or detailed enough, or not regionalised.	Research and innovation national agencies (FCT and ANI) identify, characterise and map projects funded by other sources (national budget, European programmes, etc.) and provide a "matching" key to relate this database with the other on Cohesion Policy co-funded projects (Roadmap 1). MEDIUM-TERM GOAL: to have a single platform for open science and innovation, with a territorial cohesion approach.

Actions		Intermediate outcomes
Validate and/or extend the list of fields and standards.	▶	List of field based on international standards.
Gather and regionalise projects from the selected data sources. Introduce data fields required for smart specialisation monitoring.	▶	Database of regionalised nationally and EC-funded projects, with S3 characterisation.
Provide info in open access (non-proprietary) format (from Excel to CSV).	▶	CSV files published information.

Involved actors & enablers (Possible support by the EC and third actors)	
• Political level: Minister of Science and Minister of Economy. • Implementation bodies: FCT and ANI.	• AMA (national agency for administrative modernization) – guidance and support, link with similar initiatives.

Assumptions about the actors		
Minister of Science and Minister of Economy (mandate) FCT and ANI (implementation).	◀	Support: AMA.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
OpenCoesione FRIS	

5.7. Romania

Analysed data sources

Out of the diverse national data sources proposed by stakeholders and experts, and extended through research, three sources were selected for further analysis:

1. Ministry of European Funds - MFE[47] - as Managing Authority of the Competitiveness OP 2014-2020,
2. Former Ministry of Regional Development and Public Administration - MDRAP[48] - as Managing Authority of the Regional OP 2014-2020,
3. Executive Unit for Financing Higher Education and Research-Development and Innovation - UEFISCDI[49] - responsible for the implementation of the National RDI Programme.

The **three Romanian data sources** analysed in the "Data suitability and openness analysis" can be described **as "Publicly funded research and innovation projects and activities' information systems"**:

- **One** of them is an "**R&D and innovation grant-managing agencies'"** information systems: **UEFISCDI –** Unitatea Executivă Pentru Finanțarea Învățământului Superior, A Cercetării, Dezvoltării Și Inovării **(Executive Unit for Financing Higher Education, Research-Development and Innovation)** an institution **with activities linked to theimplementation of national RDI policy and the National RDI Plan.** UEFISCDI is a public institution, under the coordination of the Ministry of Education and Research.
- **The other two** data sources analysed are **"ESIF management and monitoring information systems"**:
    - The **Open Data Portal of the Romanian Government – sub-page of Ministry of European Funds: http://data.gov.ro/organization/mfe contains** the list of **contracted projects under each of the ESIF Operational Programmes** managed. In particular, it contains information linked to projects financed under Competitiveness OP, Priority Axis 1, linked to Thematic Objective 1. The information is provided at the national level relying on data from the ESIF monitoring and information system.
    - Additionally, the Ministry of Regional Development and Public Administration (MDRAP - Ministerul Dezvoltării Regionale si Administrației Publice) publishes information linked to the projects financed under the ERDF **Regional Operational Programme** 2014-2020, both on the website mentioned and the web-page of the programme (www.inforegio.ro), including Priority Axis 1 of the OP, linked to Thematic Objective 1. In this case, as well, data is provided based on the registries from the ESIF monitoring and information system and is made available on the Open Data Portal of the Romanian Government.

[47] http://data.gov.ro/organization/mfe
[48] http://www.inforegio.ro/ro/ - Currently the name of the ministry is Ministry of Public Works, Development and Administration.
[49] https://uefiscdi.gov.ro/

Roadmap

ROMANIA

Challenge	Goal
RDI related data is not gathered in a structured/standardized format.	To have (linked to projects funded from public sources) a minimum set of RDI related data available in an open access format.

Actions		Intermediate outcomes
Raising awareness and building a common understanding on the fact that monitoring is a necessary, permanent systematic process - targeting all public actors involved in financing RDI.	▶	Commitment of the involved actors towards RTDI policy monitoring.
Development of national-level legislation (for ex. Government Decree), on: ● Definition of a minimum set of RDI data to be gathered. ● Specification of a set of minimum conditions and quality standards for RDI reporting.	▶	Adoption of a national-level legal act laying down the overall framework on reporting RDI related data and indicators.
Implementation ● Harmonized development of existing data-bases/monitoring systems. ● Assuring interoperability/ interconnectedness of the systems.	▶	Conditions for implementation are in place: ● Technical specifications developed. ● Existing monitoring systems/data-bases improved / further developed: ○ To individually collect RDI related data in a structured/standardized way, ○ To assure interconnectedness/ interoperability of data.

Involved actors & enablers (Possible support by the EC and third actors)
Involvement is necessary on behalf of actors that: ● are responsible for decision- and policy making ● are responsible for operating such systems, ● are users of such data on national level, ● are users of such systems on national or regional level and are responsible for collecting and verifying data that is fed into such systems. Particularly: 1. Ministries of Education and Research, European Funds, Economy and SMEs, Agriculture and Rural Development, 2. Ministry of Public Works, Development and Administration, Regional Development Agencies, Executive Unit for Financing HEI and RDI. All actors foreseen have the necessary mandates, but for now do not recognize the need and utility for collecting such type of data. There is lack of commitment, coordination and capacity.

Assumptions about the actors
Due to lack of commitment and understanding of the importance of gathering such data, the involvement of EC – DG REGIO, as a "top-down influencer" in the process would be relevant in building awareness about the importance of RDI data collection linked to the fulfilment criteria under the enabling condition for PO1.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
CORDIS FRIS	Horizon 2020/Europe European Territorial Cooperation (ETC) programmes European Investment Bank– financial instruments

5.8. Spain

Analysed data sources

Out of the diverse national data sources proposed by stakeholders and experts, and extended through research, four sources were selected for further analysis from the following organisations - initiatives:

1. State Research Agency - AEI[50]
2. Centre for the Development of Industrial Technology - CDTI[51]
3. Ministry of Treasure - OP interregional ERDF[52]
4. Ministry of Treasure – The General Intervention Board of the State Administration (IGAE) - National Publisher System of Grants[53]

The **four Spanish data sources** analysed in the "Data suitability and openness analysis" can be described **as "Publicly funded research and innovation projects and activities' information systems"**:

- **Two** of them are R&D and innovation grant-managing agencies':
  - **Agencia Estatal de Investigación (AEI - State Research Agency),** on scientific and challenge-oriented research.
  - **Centro para el Desarrollo Tecnológico Industrial (CDTI - Centre for the Development of Industrial Technology),** on applied collaborative research and corporate R&D.
- The **National System for Publicising Subsidies and Public Grants (SNPS)**[54] provides information on all grants and subsidies to, among others, private companies, including those not co-financed by ESIF funds.
- The **management and monitoring information system of the ERDF Pluriregional Operational Programmes**[55].

[50] http://www.aei.gob.es
[51] https://www.cdti.es/index.asp?idioma=2&r=1366*768
[52] https://www.dgfc.sepg.hacienda.gob.es/sitios/dgfc/en-gb/Paginas/Inicio.aspx
[53] https://www.pap.hacienda.gob.es/bdnstrans/GE/en/index
[54] Ministerio de Hacienda - Intervención General de la Administración del Estado - Sistema Nacional de Publicidad de Subvenciones (MINHA-IGAE-SNPS - Ministry of Treasure – General Comptroller of the State Administration of the State Administration - National System for Publicising Subsidies and Public Grants)
[55] Ministerio de Hacienda - Programas Operativos Plurirregionales FEDER

Roadmap

SPAIN

Challenge	Goal
National and Regional Funding Agencies provide information with varying degrees of accessibility, granularity and quality (i.e. not homogeneous fields) about policy instruments.	To improve the quality of publicly available information across National and Regional Funding Agencies by: ● Homogenizing fields to be published, ● Providing metadata, ● Publishing project abstracts.

Actions		Intermediate outcomes
Define a minimum list of fields to be published. The list of fields currently published by the management and monitoring information system of ERDF Pluriregional Operational Programmes could be a good baseline model.	▶	List of fields to be published.
Identify coordination mechanisms across National and Regional level funding bodies. Initiating contact with Red IDI to help to achieve this goal.	▶	Increased cooperation, harmonisation and best-practice sharing regarding data and monitoring, among regional and national funding bodies.
Demonstrate the value of open data to the different National and Regional Funding Agencies, including the value of analyzing data for different purposes and the legal framework on Open Data and confidentiality issues.	▶	Raising awareness and obtain a mandate to publish from the top management and ensure that Fudging Agencies are aware of their mandate (list and rank different National and Regional Funding Agencies).

Involved actors & enablers (Possible support by the EC and third actors)	
1. National and Regional Funding Agencies (e.g. State Research Agency –AEI; Centre for the Development of Industrial Technology -CDTI). 2. Ministry of Finance (representative of the management and monitoring information system of the ERDF Plurigegional Operational Programmes.	Supporting and coordination actor: Red IDI (Red Temática de Políticas Públicas de I+D+i).

Assumptions about the actors
● National and Regional Funding Agencies have the policy and legal mandate to publish their data in at least 3-star Open Data. ● National and Regional Funding Agencies have the willingness and resources to open their databases. ● Red IDI will be interested in helping to achieve the goal and will take an active role in coordinating.

Inspirational examples (Case studies or other)	Use or connection with EU and global standards & infrastructures
FRIS. Gateway to research. National and Regional examples such as RIS3-MCAT or Observatorio RIS3 Extremadura.	European Open Data Portal.

6. Conclusions and recommendations

This report has detailed the process and the outcomes of the working group on *ODSI for S3 monitoring*, which was run in 2019 as part of the activities of targeted support to less developed regions. The report has addressed a crucial and so-far unexplored aspect of S3 monitoring, namely, how it can benefit from Open (government) data, open science and open innovation. These topics have been addressed following a strong demand from regions and member states participating in the JRC targeted support activities.

Monitoring Smart Specialisation Strategies has emerged as a particularly challenging topics and the evolution of the data landscape, with increasing information being generated and made available within and outside government, has made a reflection on this topic urgent.

The project has pursued this challenging topics by first covering key conceptual framework, developments and examples in the field of ODSI for S3 (chapter 2, chapter 3 and the annex). Most importantly the project has dedicated much to the peer exchange and learning processes among participants to the working group. Two analytical tools (the data suitability and data openness grids) have been developed to foster an in-depth reflection on country-specific data sources (chapter 4). These have fed the development of the country-specific roadmaps presented in chapter 5.

Whilst many of the lessons generated by the project are country-specific, important general recommendation can be derived for regional or national administration trying to exploit ODSI for S3 monitoring. In particular:

Open: Improving access to data from projects and instruments of the S3 policy-mix is an essential first step.

Policy-makers and policy-officials often lack access to suitable data from projects and instruments of the S3 policy-mix. This jeopardises S3 monitoring at all levels of granularity, and impedes policy learning towards the next programming period. It is essential that efforts towards opening-up data related to the S3 policy-mix are boosted and that access to relevant data is provided.

Open carefully: essential data-fields, metadata and granularity

When opening data and developing information system for S3 monitoring, it is essential to ensure that the data-source has the right metadata, the right data-fields and is provided at the right level of granularity. In particular, the following should be provided:

- Well-documented and **persistent metadata[56] and unique identifiers** for all relevant data classes in the STI domain (policies, projects, organisations, individuals, results, impacts, etc.).
- **Free text descriptions** of the content of the data, such as project titles and objectives or publication abstracts, ideally in the local language and also translated into English.
- **Taxonomic information** describing the content of the data, using both international standard classifications (such as NACE codes or IPC classes) as well as ad-hoc taxonomies (such as S3 priorities), with the related documentation.
- **Complete localisation information (at least, regionalised)** to permit analysis at different geographical levels and facilitate benchmarking.
- **Maximum data granularity**: all the above information should be available for each activity, for each actor, for each taxon, for each result or outcome, for each year or monitoring period - **avoiding data aggregations or summary tables** that curtail data reuse for fined-grained analytical purposes.

Open for reuse: Interoperability is key and should be exploited to the maximum

When gathering, organising and publishing data, administrations should aim to provide good access and facilitate reuse. The **Open Government Data principles** and the **FAIR data principles** are a good inspiration: R&I-S3 policy data should be made available: **Primary, Complete, Timely, Accessible, Machine processable, Interoperable, Findable and with a well-specified open License** permitting people to reuse the data in any way they want. From the Open Science framework

Remarkably, public administrations and the Open Science and Innovation community are defining ontologies, developing integration platforms and providing single identifier platforms that can solve relevant technical challenges

[56] It is rather frequent a situation where fields are not consistent from one programming period to another: for all cases where continuity is not an option, it is important to provide guidelines on which features can be mapped from old datasets to newer ones.

hindering S3 monitoring, such as the disambiguation of actors (with **ORCID** or **ROR**), the identification of activities and results or the link between funding and results (with **Crossref**, **Open citations**, **ClinicalTrials.gov**, **EPO's EPAB**, **OpenAire, 360º giving**).

Company data is fundamental for S3 policy monitor: interoperable business registries are crucial

Company data deserves a special mention, as it is crucial for S3 monitoring, yet business registries across the EU may vary significantly in their degree of openness. Project participants deemed particularly important to **improve access to granular data on company creation** (start-ups and spin-offs) and **evolution** (sales, exports, employment), as well as firms' knowledge creation and dissemination. This information can be used to monitor individual funded projects (and, by transitivity, policy instruments) and to monitor the evolution of the economy (by sectors, S3 priorities, etc.). Remarkably the new EU Open Data Directive defines **"Companies and company ownership" data as a high-value dataset, to be provided free of charge.**

Addressing the digital divide: a challenge for the EU

Several countries and regions have been advancing in Open Government Data, Open Science and Open Innovation for the last two decades, building an ecosystem of regional, national, European and international infrastructures, systems and protocols. The European Union is leading and supporting this process through overarching policies, the funding of infrastructures and projects, and the new Open Data Directive. However the digital divide is present, and less developed countries and regions face transversal challenges to the implementation of ODSI which hinder policy making, implementation and monitoring as well as an equal participation in EU-wide initiatives (such as EOSC). Addressing this divide is paramount if the whole EU is to benefit from ODSI for S3 and for policy decisions more in general.

Supporting ODSI for S3: not only a technical matter

Supporting the deployment of ODSI (for S3) is not only a technical matter. Rather, it requires a broad set of dedicated resources and skills: ranging from **policy, data and systems expertise**[57] to **facilitation and community-building capacities**. Indeed it is necessary to support the community of stakeholders and allow for different speeds of implementation, since different organisations, sectors or disciplines present different "data-maturity". In most cases, these processes requires confidence, grit, collaboration and coordination, as well as a strong political mandate, because they open **discussions about concepts and definitions, data quality and consistency, and new obligations and protocols**. Whilst the challenges cannot be overstated, the opportunities must not be missed.

[57] See page 6: https://dspacecris.eurocris.org/bitstream/11366/1229/4/Schelske%26Thiedig__ECMM2019.pdf for a useful list of required data tasks and data competencies.

References

Chesbrough, H. (2003), "Open Innovation: The New Imperative for Creating and Profiting from Technology", Harvard Business School Press.

Foray, D., David, P.A. & Hall, B. (2009). Smart Specialisation – The Concept. Knowledge Economists Policy Brief n ° 9, (June), pp.1–5.

Foray, D. (2014), "From smart specialisation to smart specialisation policy", European Journal of Innovation Management, Vol. 17 No. 4, pp. 492-507. https://doi.org/10.1108/EJIM-09-2014-0096

EC (2016). Open Innovation, Open Science, Open to the World - A Vision for Europe. Brussels: DG Research and Innovation. doi:10.2777/061652

EC (2019). International Digital Economy and Society Index 2018. Available here: https://ec.europa.eu/digital-single-market/en/news/international-digital-economy-and-society-index-2018

Marinelli E., Perianez-Forte I. (2017) Smart Specialisation at work: The entrepreneurial discovery as a continuous process, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-74377-1, doi: 10.2760/514714, JRC108571.

Marinelli, E. Guzzo F., Gianelle C. (2019), Building Smart Specialisation Strategies Monitoring Systems: Evidence from the EU, in "L'industria, Rivista di economia e politica industriale" 1/2019, pp. 27-44, doi: 10.1430/94058

OECD (2002), Public Sector Modernisation: A New Agenda, GOV/PUMA (2002)2

OECD (2013), Open Government Data available at http://www.oecd.org/gov/digital-government/open-government-data.htm, retrieved December 16th, 2019.

OECD (2015), "Making Open Science a Reality", OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris, https://doi.org/10.1787/5jrs2f963zs1-en.

Scott et al (2019); Contribution to growth: The European Digital Single Market. Delivering economic benefits to citizens and businesses, Study for the Committee on the Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2018.

UN (2017), United Nations Development Assistance Framework (UNDAF), "Theory of change Campanion guidance", 2017.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Annex - Reference initiatives and platforms

To illustrate the features, opportunities, benefits and difficulties of ODSI for S3, we have interviewed a series of real open data, open science and open innovation initiatives, reported below.

RIS3-MCAT – Catalonia Smart specialisation mapping platform (Spain)[58]

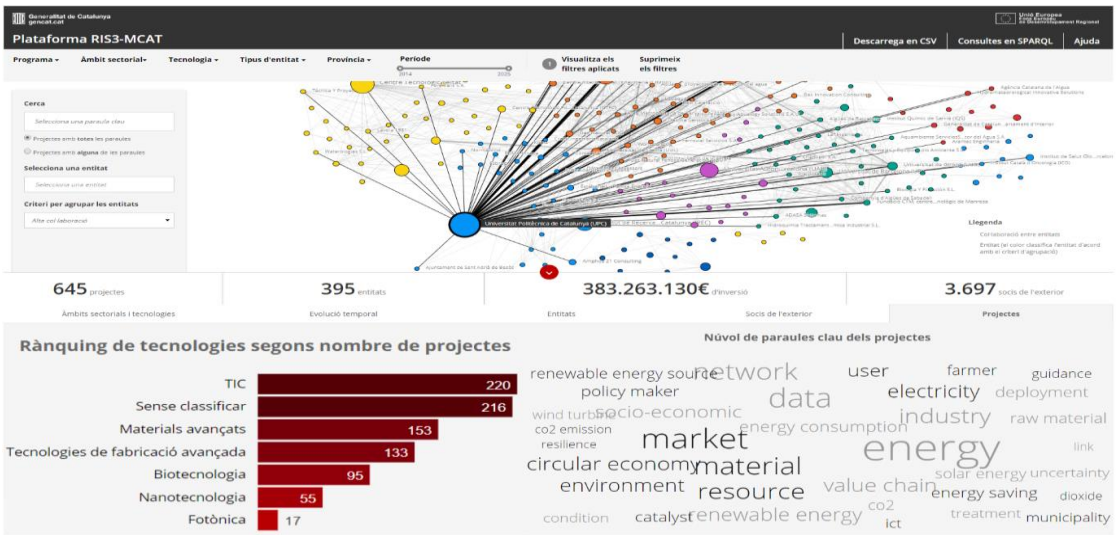
Interviewee: Tatiana Fernandez

Promoted by the Directorate General for Economic Promotion, Competition and Regulation, Generalitat de Catalunya (Spain), the RIS3CAT smart specialisation mapping platform (RIS3-MCAT) aims to respond to a wide range of strategic and operational questions of the Generalitat (regional government) and of Catalan science and innovation actors. In particular it, supports implementation and monitoring of the S3 strategy by the RIS3CAT governance and team. This is a technical and political process that requires both aggregate and granular data, in predefined (S3 priorities) and emerging (free-text) perimeters, responding to key questions such as:

- Who are the main actors in S3 priorities or in new sub-priority area of interest, and their regional, national and international partners
- In which specific topics are stakeholders contributing
- Which are promising areas, presenting growing interest
- Are regional actors internationally competitive in R&D

The tool aims to provide reliable, granular and fresh data to:

- Map and characterise science and innovation projects
- Automatically classify these projects, from their text descriptions, into the RIS3CAT priority sectors and technologies
- Map collaboration between stakeholders (innovation communities)
- Present temporal evolution and geographic distribution
- To respond to emerging areas of enquiry, the system allows for the free search of >30.000 automatically-extracted keywords, identifying related institutions and projects, and building automatic indicators on top of them.



Snapshot of http://ris3mcat.gencat.cat

[58] Info website: http://catalunya2020.gencat.cat/en/plataforma-ris3-mcat/
Link to the platform: http://ris3mcat.gencat.cat/#/

Such diversity of demands, and the volume of underlying data has led to the development of visualisations and a web interface co-designed with Generalitat officials, stakeholders and experts, in a participatory process.

This project used several analytical and natural language processing (NLP techniques to overcome the intrinsic difficulties of monitoring smart specialisation, in particular:

1. Semi-automatic, human-validated, taxonomy based classification of R&D projects financed by different policy instruments and administrations, via NLP on the projects' descriptive text field using a seed corpus technique
2. Automatic keyword extraction to facilitate discovery
3. Network analysis and visualisation

The platform includes project information from regional and European R&I programmes (H2020 and ERDF funded), amongst other data: title and description of the projects, beneficiaries and consortia collaboration links, funding per beneficiary, S3 classification in priority domains, automatically-extracted keywords.

The platform is aimed at a wide-range of audiences: S3 policy makers and programme officers (including teams responsible for S3 monitoring), programme officers in the business innovation and research areas of the Catalan government, stakeholders in the R&I ecosystem such R&D strategists and managers in universities, research institutions and private companies, intermediary agents such as technology transfer officers, cluster managers and other ecosystem facilitators, researchers in science and innovation and regional policy and science of science.

The platform includes the following open accessibility and download facilities:

- A fully searchable and filterable exploratory tool of the collaboration networks between Catalan actors, providing automatically computed indicators also covering the 3 points above.
- Access to the underlying datasets (in machine readable 5* Linked Open Data format). In particular:
    - an S3-enriched version of the CORDIS database projects with Catalan beneficiaries,
    - the data on ERDF-funded research and innovation projects managed by Generalitat de Catalunya within the RIS3CAT policy-mix, a dataset which was previously unpublished.

OpenCoesione (Italy)

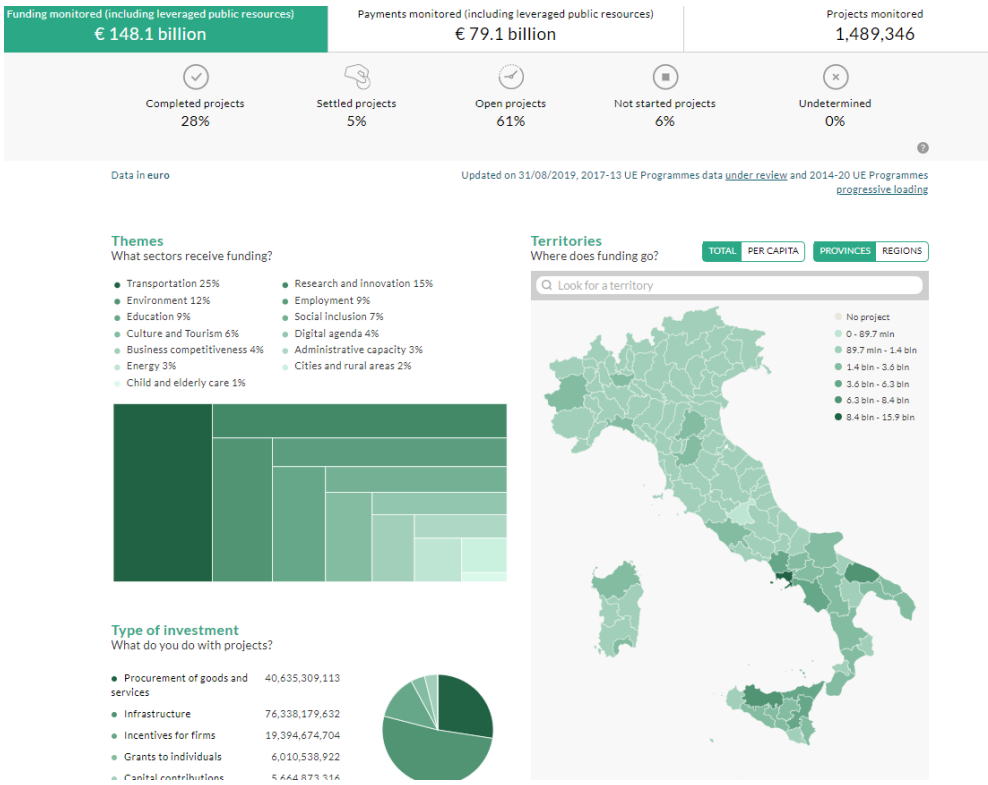
Interviewees: Simona de Luca and Fabio de Angelis

OpenCoesione is an open government data initiative launched in 2012 which integrates data of projects (currently 1.4 million) financed by the European and the Italian cohesion policies, and national and European funds for regional cohesion policy.

OpenCoesione has a threefold aim:

- bringing public spending and open government under the spotlight,
- bringing transparency to policy spending and
- encouraging citizen monitoring of cohesion policy.

Regarding the latter, OpenCoesione directly promotes "A Scuola di OpenCoesione" (an innovative educational programme to introduce principles of active citizenship in Italian high school students) and collaborates on independent projects such as Monithon, a web platform aimed at promoting civic monitoring of public funding.



Snapshot of https://opencoesione.gov.it/en/

The main users of OpenCoesione are civil servants working on the monitoring and evaluation of public policies, researchers in economy and public sectors and citizens interested in better understanding public investments and policy outcomes.

"The National Monitoring System, managed by General Inspectorate for Relations with the European Union of Italy's National Accounting Office of the Ministry of the Economy and Finance (MEF-RGS-IGRUE), is the source of data on projects being implemented as published on OpenCoesione. At the core of the National Monitoring System is the unified database fed, at the individual project level, by the local information systems of all the public bodies responsible for plans or programmes financed by cohesion funding based on common rules and standards."[59] Further information on the process to collect, normalise and integrate this is available on OpenCoesione's website "National Monitoring System" section.

[59] https://opencoesione.gov.it/en/sistema_monitoraggio/

Each project is described by circa 200 fields which allows to characterize a project's objective, funding, execution, beneficiaries, resources allocation, etc. Programming authorities are also responsible for specifying, at the project level, taxonomic classifications (S3 domains or economic sectors). Also, beneficiaries are invited to provide details about the results or documentation generated throughout the execution of the project. OpenCoesione has several projects to improve the interoperability of the data with other sources and public databases.

All data can be downloaded in bulk, through different facilities and in different formats through the "Download Open Data" section of the website. All the data is also stored and available at the EU Open Data Portal.

Flanders Research Information Space (FRIS, Flanders)

Interviewees: Leen Van Campe, Dengis Pascale, Dumolyn Bart

Leen Van Campe participated in Workshop 3 of the project, where she gave a rich presentation of the initiative adapted to the types of questions arising from ODSI for S3. The presentation can be found in the following link:

https://s3platform.jrc.ec.europa.eu/documents/20182/371141/Leen+Van+Campe_Alcala+de+Henares_Presentation.pdf/53d9c0af-f5ac-4506-97a7-c3f6eb0ddbf7

About FRIS

From https://researchportal.be/en/about-fris

"FRIS (Flanders Research Information Space) is the regional portal on researchers and their research in Flanders. The Flemish Government wants to offer a unique window on research in Flanders and increase its visibility. The portal can be a source of inspiration for the Flemish government for reports, analysis and statistics for policy making and better following trends. The FRIS portal will unite researchers, by encouraging interdisciplinary research, networking among researchers and enabling to find experts in certain disciplines. FRIS aggregates all the publicly funded research in Flanders.

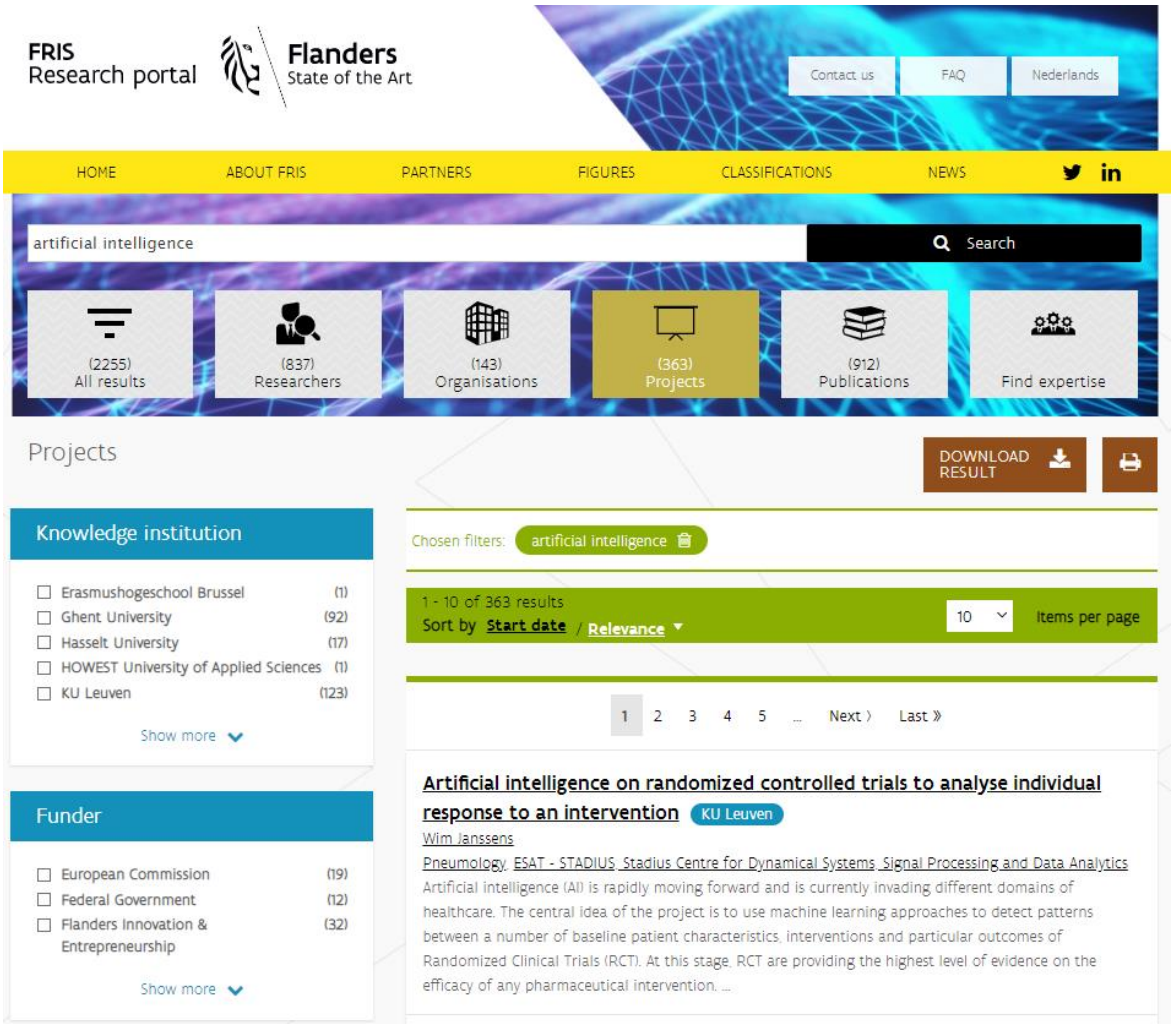
- Five Flemish universities
- 9 research institutions
- 38,000 research projects.
- 82,000 researchers.
- 410,000 scientific publications.

5 Flemish universities and 9 research institutions have already linked their systems to FRIS. Automatically ensuring their data feeds directly into our information space. As soon as a research institute changes something in their systems, you'll immediately see the change in FRIS."

The target user of the platform are:

- Research institutions
- Researchers
- Innovative companies
- Industry
- Government bodies (policy makers and funding instrument officials, to:
    - Use FRIS as reliable source to monitor your own policy.
    - See the results of research you are funding, discover trends and conduct impact analyses.
    - Look up what research has already been conducted in your policy domain [or scientific area] and who the experts are."

The project started in 2014 and the platform was first published online in 2018. The FRIS platform integrates the institutional CRISs of the Flemish publicly funded research institutions (universities and centres). The legal foundation are several regional government decrees by which all concerned research institutions are obliged to provide information to FRIS (if the share of Flemish funding is above a threshold).



Snapshot of FRIS https://researchportal.be/en

The data and systems integration required adaptations by all stakeholders and complex discussions on the format, semantics, etc., a process which made more explicit the information requirements from the Government and the data and metadata interoperability and quality standards in the different institutions. In particular, a whole project on aligning and simplifying scientific taxonomies was developed. The platform integrates more data than what is available through the FRIS Portal.

FRIS has an important task in solving all kinds of complexities in the data (deduplication, mobility of scholars, etc.), via automatic methods by using persistent identifiers (for people -ORCID is compulsory-, and for publications and other research artifacts),

FRIS is used for research and innovation policy analysis and monitoring, such as understanding the evolution of the region's research portfolio, gauging its specialisation in Artificial Intelligence.

FRIS is a very active project, with several ongoing developments:

- Integration of additional research institutions.
- New partnership with FWO (the Flemish research funding agency) to integrate information on individual projects, identify the activities, results, outcomes and impacts already present in the platform, integrate new information and support monitoring and reporting activities.
- Integration of new data: research data, research infrastructure, patents and spin-offs.

- Development analytics on the use of the FRIS Open Data Services.
- The Flemish Government is launching the Flemish Open Science Board (FOSB); FRIS and FOSB will be working closely together, for instance in monitoring the state of Open Access in Flemish research.
- Possibility to explore natural language processing techniques to analyse unstructured data from open access publications.

Gateway to Research (UK)

Interviewee: Ashley Moore

About Gateway to Research (GtR)

From https://gtr.ukri.org/resources/about.html

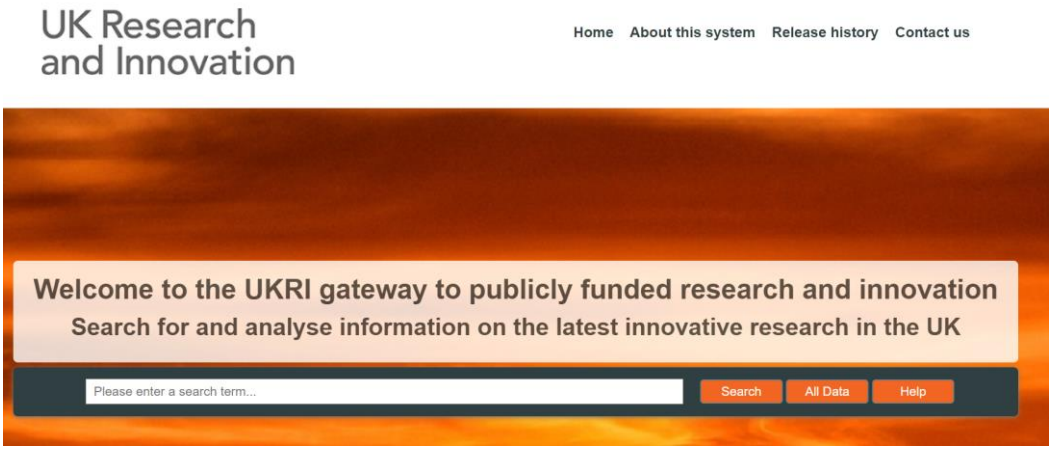
"The Gateway to Research (GtR) website has been developed by the UK Research and Innovation (UKRI) to enable users to search and analyse information about publicly funded research.

[...]

Two Application Programming Interfaces (APIs) have been provided as a way of accessing the information directly from other Information Systems – further information can be found here. This allows third parties to link the GtR data to other data sets and analyse the information for their own purposes.

This website is open and free for all to use and has been developed using Open Source, Open Standards and an Open Government Licence (OGL), to enable the code to be reused by third parties."

Gateway to Research was originally developed to facilitate discovery, collaboration and open innovation from science to business (particularly targeting SMEs). In the last four years, it has evolved to also support open science efforts and monitoring and accountability functions.



Snapshot of https://opencoesione.gov.it/en/

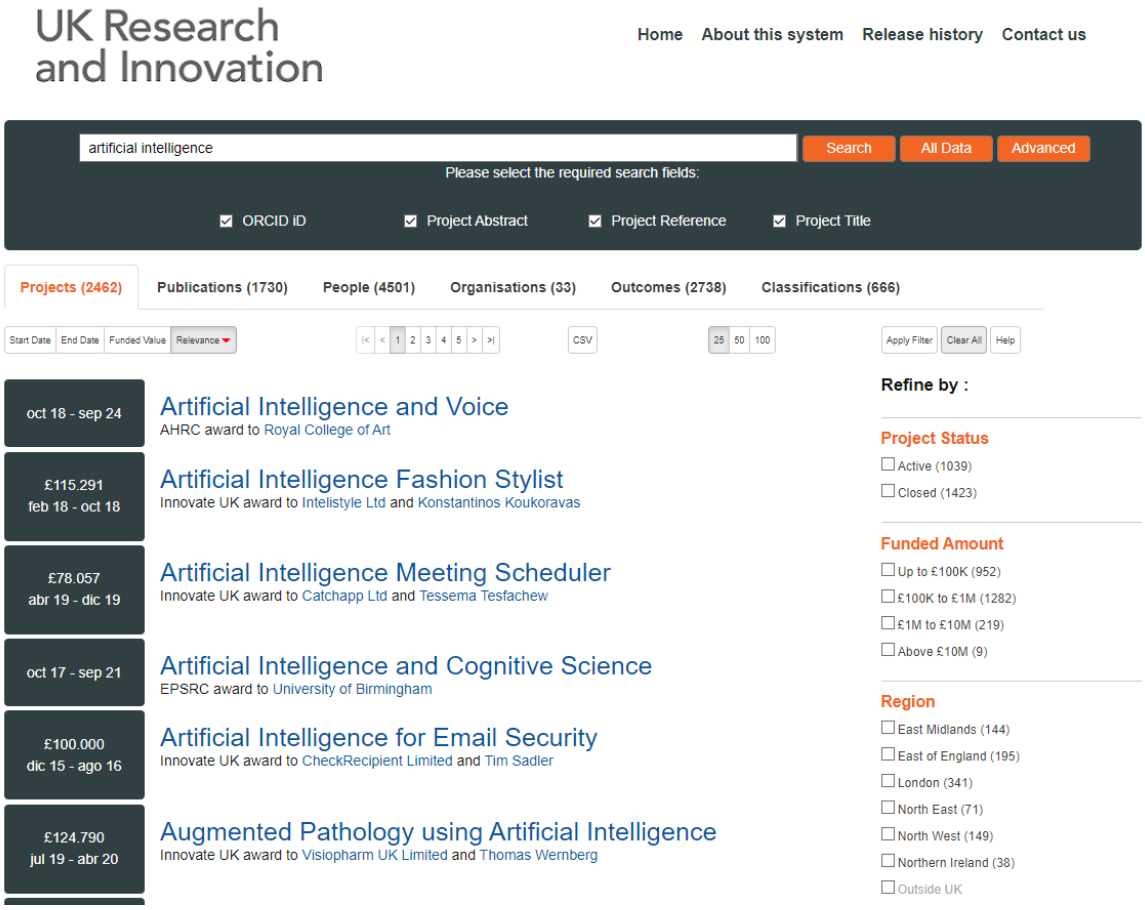
It presents and makes available part of the data from source systems behind GtR[60] on "publications, people, organisations, outcomes and classifications relating to research projects. The projects details include information about the project title, project abstract, duration of the project and the amount awarded."[61] The full GtR Data dictionary can be found here, and is especially interesting for its coverage of research results and outcomes, covering dimensions such as:

- Artistic and Creative Product
- Engagement Activities
- Impact summary

[60] "The data, published on GtR, is collected from a range of systems used by the funding organisations to collect information from researchers, these include, the Joint electronic Submissions system (Je-S) and ResearchFish. There are some limitations on the data that can be made available from these systems." https://gtr.ukri.org/resources/data.html

[61] https://gtr.ukri.org/resources/data.html

- Policy Influence
- Product Interventions & Clinical Trials
- [...]



Snapshot of https://opencoesione.gov.it/en/

The source data system behind GtR is being used for policy and programme monitoring by several funders. In the context of smart specialisation, regions are analysing data from their regional actors and Members of Parliament using the platform to understand what is being funded in their constituencies.

In terms of openness, data can be downloaded in CSV format, and, for advanced users and developers, GtR provides two Application-Programing Interface (API).

GtR is planning a systematisation of their connection with and use of persistent identifiers (for grants, publications and research data, patents, organisation) to improve metadata quality and interoperability.

Danish Company Register (datacvr.virk.dk/data/ - Denmark)[62]

Interviewees: Tobias Kjærulff Langberg and Morten Klausen.

The Danish Central Business Register (CVR, Centrale VirksomhedsRegister) is the central company register that provides data on all businesses in Denmark. The Danish Business Authority, DBA, is responsible for the register. In the 2010s, it engaged in a major transformation toward openness, in the context of a comprehensive and ambitious open government data policy, led by the Danish government. The main objectives of the open data policy are:

- transparency,
- open data-based business promotion,
- within a global project to diminish administrative burden through centralisation and digitalisation of procedures.

CVR is currently one of the most open business registries in the world, with data being available from the platform datacvr.virk.dk/data/. The implementation of the platform required specific measures such as minimizing the cost of registering, while the Danish administration assumed the cost of managing the registry, and changing the law in order to allow for the free of charge publication of as much data as possible without breaking confidentiality regulations.

The key types of data provided (useful for S3 monitoring) are the following:

- Company ID data (single identifier)
- Shareholders
- Financial statements
- Industry codes and specialisations (used by the Statistics Agency) - companies can submit up to three industries if necessary
- Textual information of their purpose (only publicly-traded companies)
- Localisation (including multisites - headquarters / production units)
- Employment brackets

Data from CVR can be accessed as single lookups on individual businesses, lists of businesses demarcated by e.g. geography, number of employees, industry, etc. as well as raw data from af system-to-system access (API-access). Most of the API-access is concentrated in a small group of companies (data brokers). DBA assesses continuously how data from CVR can be made more accessible to both technically-capable and less technically-capable actors.

Beyond these top corporate users, the Register is widely used by a variety of users and stakeholders:

- Ministries: Industry and Business, Finance…
- Municipalities
- Business Associations
- Unions
- Specialised companies, for instance in the field of Risk analysis.
- Other government-backed platforms such as Digital Denmark[63], using the business register to count companies engaged in the digital industry in relation to their economic performance.

[62] https://datacvr.virk.dk/data/?language=en-gb
[63] https://digitaldenmark.dk/

In terms of data publication, DBA's paradigm is "as close to the raw data as possible". Therefore, the data can be downloaded and exploited for the purpose of monitoring, but the platform has no specific features to facilitate said monitoring. Some examples of the use of the registry data for policy analysis and monitoring, are:

- Economic and financial analysis of the wider economy and specific sectors and geographies
- Elaboration of the "State of the Digital Economy in Denmark" report
- Municipalities use the data to define business support policies

DBA is part of the national "Public Partnership for Open Data" that promotes the use of public data in businesses and participates in an international collaboration project to foster registry data harmonisation with Nordic partners (in the framework of the Nordic Smart Government initiative).

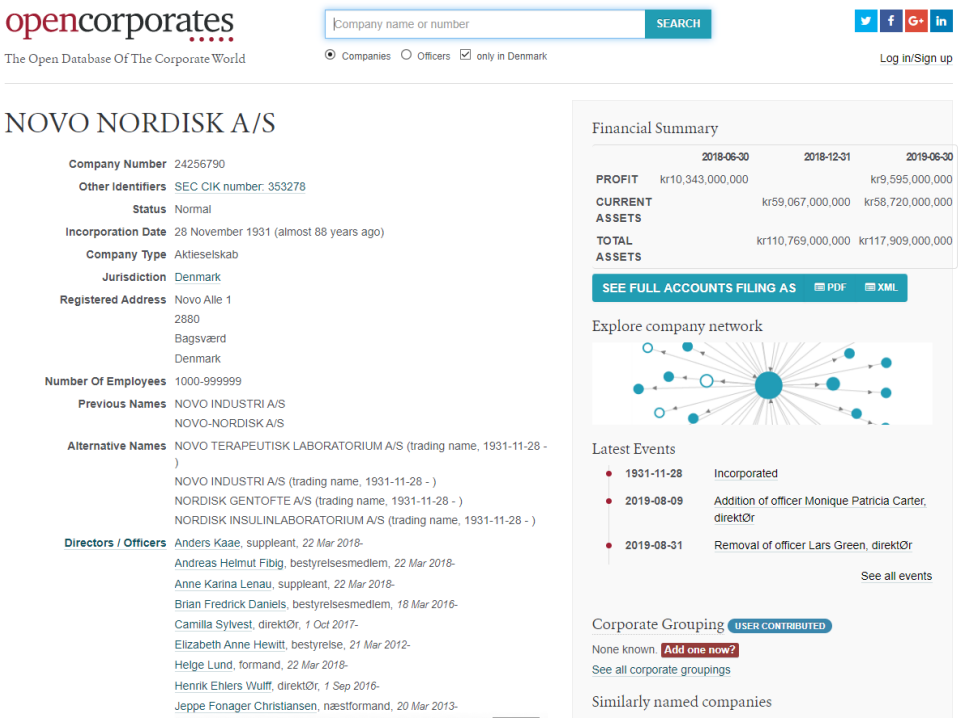
Since data from CVR was made free of charge, the use af CVR-data has increased significantly. On a yearly basis, there are approximately 16,5 million single lookups and 2 billion lookups through the API-access. These significant numbers demonstrate the success of the open government data policy led by the Danish government

OpenCorporates[64]

Interviewee: Chris Taggart

OpenCorporates is an open platform providing data on individual corporate entities. It presents itself as "the largest open database of companies in the world". OpenCorporates was created on the basis of an engagement towards transparency and social progress, pursuing citizen empowerment in a world managed by corporations and contracts and facilitating business collaboration, risk assessment, antifraud activities, flagging tax avoidance, etc.:

> *"Total corporate transparency is a critical requirement for a fairer society. To ensure that everyone knows exactly who they are working with – and working for. To tackle corruption and criminality. To protect our democracy. To create a trusted business environment we want to work in – and a society we'd all like to live in[65]."*



Snapshot of https://opencorporates.com/

OpenCorporates data includes 180M companies across 139 jurisdictions (countries, US states, etc.). In addition to the classical data types of company registries, they include added-value features, such as:

- Connections between companies and across jurisdictions
- "Historic" view of company evolution
- Quality check and feedback from the community that guarantees the very good quality of data.

OpenCorporates is currently developing connections with Intellectual Property databases and other external resources.

In terms of business model, OpenCorporates is conceived as a social purpose company: they can pursue and turn a profit, but to be used in the direction of the public benefit. The preservation of this mission is linked to a specific Trust structure, to guarantee the independence of the platform and the conformity of its activities with its vision and engagement. This business model oriented to social benefit rests on a series of openness and public purpose services coexisting with a commercial offering:

- All information is available free of charge in the website
- OpenCorporates provides bulk download / API access free for public purpose projects (NGOs, journalists, some public

[64] https://opencorporates.com/
[65] https://opencorporates.com/info/about/

sector projects)

- OpenCorporates sells bulk download / API access to corporate clients and public administrations

OpenCorporates counts thousands of users every day: companies, NGOs, tax agencies, journalists, academics, law enforcement, regulators, due diligence professionals, lawyers, forensic accountants, banks, etc.

With public bodies, the main areas of activity and collaboration are:

- Advocacy, supporting governments in arguing for company data openness (for instance, OpenCorporates has worked with the European Commission on the Open Data Directive and the re-use of public sector information)
- Supporting registry / company data publication and quality improvement
- Supporting / training public officials
- Providing bulk data / API access (UK, Ireland, Norway, Denmark,France, Belgium, etc) .

OpenCitations[66]

Interviewee: Silvio Peroni.

OpenCitations is an independent infrastructure organization for open scholarship dedicated to the publication of open bibliographic and citation data using Semantic Web (Linked Data) technologies. OpenCitations started in 2010, with the aim "to change the face of scientific publishing and scholarly communication, since it aimed to publish open bibliographic citation information in Resource Description Framework (RDF[67]) and to make citation links as easy to traverse as Web links"[68]. It is also engaged in advocacy for semantic publishing and open citations. OpenCitations espouses fully the founding principles of Open Science. It complies with the FAIR data principles by Force11 that data should be findable, accessible, interoperable and re-usable, and it complies with the recommendations of the Initiative for Open Citations (I4OC[69]) that citation data in particular should be structured, separable, and open.

OpenCitations makes available the OpenCitations Corpus (OCC), a database of open downloadable bibliographic and citation data recorded in RDF and released under a Creative Commons CC0 public domain waiver, which currently contains information about ~14 million citation links to over 7.5 million cited resources. In addition, and separately, OpenCitations is currently developing several Open Citation Indexes, using the data openly available in third-party bibliographic databases. The first and largest of these is COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations, which presently contains information encoded in RDF on more than 624 million citations, released under a CC0 waiver. It has also launched a system for globally unique and persistent identifiers (PIDs) for bibliographic citations – Open Citation Identifiers (OCIs).

To enable the description of bibliographic and citation information in machine-readable terms, OpenCitations provided, maintains and updates the OpenCitations Data Model (OCDM). The OCDM allows one to record information about:

- published bibliographic resources that either cite or are cited by another published bibliographic resources, or that contain citing/cited entities (e.g. a journal containing an article or a book containing chapter);
- possible resource embodiments defining the particular physical or digital format in which a bibliographic resource was made available;
- bibliographic references usually occurring in the reference list (and usually denoted by one or more in-text reference pointers within a citing bibliographic resource) of a citing entity, that references another bibliographic resource;
- responsible agents, such as people or organizations, having a certain role with respect to a bibliographic resource (e.g. an author of a paper or book, or the publisher of a journal);
- the roles held by an agent with respect to bibliographic resources (e.g. a person being the author of an article and the editor of another book);
- the citations between two bibliographic resources;
- the external identifiers, such as DOI, ORCID, PubMedID, Open Citation Identifier, associated with the bibliographic entities.

All the bibliographic and citation data provided by OpenCitations can be queried by means of its SPARQL endpoints, can be retrieved by using REST APIs, can be searched by using the Search Interfaces, and are also available as dumps on Figshare in CSV, N-Triples, and Scholix.

[66] https://opencitations.net/
[67] https://www.w3schools.com/xml/xml_rdf.asp
[68] https://opencitations.net/about
[69] https://i4oc.org

OpenCitations Corpus SPARQL endpoint

```
PREFIX cito: <http://purl.org/spar/cito/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX datacite: <http://purl.org/spar/datacite/>
PREFIX literal: <http://www.essepuntato.it/2010/06/literalreification/>
PREFIX biro: <http://purl.org/spar/biro/>
PREFIX frbr: <http://purl.org/vocab/frbr/core#>
PREFIX c4o: <http://purl.org/spar/c4o/>
SELECT ?cited ?cited_ref ?title ?url WHERE {
    <https://w3id.org/oc/corpus/br/1> cito:cites ?cited .
    OPTIONAL {
        <https://w3id.org/oc/corpus/br/1> frbr:part ?ref .
        ?ref biro:references ?cited ;
            c4o:hasContent ?cited_ref
    }
```

Raw Response | Table | Pivot Table | Google Chart

Search: Show 50 entries

	cited	cited_ref	title	url
1	https://w3id.org/oc/corpus/br/79	Smith, RA, Cokkinides, V, Brawley, OW. Cancer screening in the United States, 2009: a review of current American Cancer Society guidelines and issues in cancer screening. CA Cancer J Clin. 2009; 59: 27–41. PMID: 19147867	Cancer screening in the United States, 2009: A review of current American Cancer Society guidelines and issues in cancer screening	http://dx.doi.org/10.3322/caac.20008
2	https://w3id.org/oc/corpus/br/83	Jacobs, I, Gentry-Maharaj, A, Burnell, M. Sensitivity of transvaginal ultrasound screening for endometrial cancer in postmenopausal women: a case-control study within the UKCTOCS cohort. Lancet Oncol. 2011; 12: 38–48. PMID: 21147030	Sensitivity of transvaginal ultrasound screening for endometrial cancer in postmenopausal women: a case-control study within the UKCTOCS cohort	http://dx.doi.org/10.1016 /s1470-2045%2810%2970268-0
3	https://w3id.org/oc/corpus/br/162	Allison, KH, Upson, K, Reed, SD. PAX2 loss by immunohistochemistry occurs early and often in endometrial hyperplasia. Int J Gynecol Pathol. 2012; 31: 151–159. PMID: 22317873	PAX2 Loss by Immunohistochemistry Occurs Early and Often in Endometrial Hyperplasia	http://dx.doi.org/10.1097/pgp.0b013e318226b376
4	https://w3id.org/oc/corpus/br/211	Scholten, AN, Smit, VT, Beerman, H. Prognostic significance and interobserver variability of histologic grading systems for endometrial carcinoma. Cancer. 2004; 100: 764–772. PMID: 14770433	Prognostic significance and interobserver variability of histologic grading systems for endometrial carcinoma	http://dx.doi.org/10.1002/cncr.20040

Snapshot of OpenCitations Corpus SPARQL endpoint https://opencitations.net/sparql

Currently, the data OpenCitations provides are not yet mature for use in analysis and monitoring in the context of smart specialisation, mainly due to the lack of systematic disambiguation or single identification efforts of individuals or organizations, nor of providing geographic metadata (such as the region where an author is affiliated). As future developments, OpenCitations plans to:

- Keep indexing bibliographic resources until becoming an authoritative source of bibliographic metadata and citation data, available for general use;
- Develop OpenCitations Meta, a new database containing rich metadata (publications' abstracts, keywords, author affiliations, funding details, etc.) which will greatly facilitate integration with other sources and platforms and analysis, for instance in S3 monitoring.

Recently, the Global Sustainability Coalition for Open Science Services (SCOSS), which facilitates funding to help ensure the sustainability of the world's Open Science infrastructure, has selected OpenCitations as worthy of community crowdfunding support in its second funding cycle.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-30089-EN-N

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
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/55098

ISBN 978-92-76-10726-2