EOSC SYNERGY
LANDSCAPE REPORT
SPAIN

EOSC-SYNERGY

Landscaping Country Report

SPAIN

Document Identifier:	EOSC-SYNERGY-LANDSCAPE-ES
Date:	2020-05-27
Activity:	WP5
Leader project partner:	CESNET
Responsible project Partner:	UPV
Authors:	Ignacio Blanquer (UPV), Isabel Campos (CSIC), María Cuesta and Inmaculada Figueroa (Spanish Ministry of Science and Innovation).

Document Status:	APPROVED
Dissemination Level:	(cc) BY
	This work is licensed under a Creative Commons Attribution 4.0 International License.
Document Link:	http://dx.doi.org/10.20350/digitalCSIC/12624

Abstract:

This landscape analysis report aims to provide an overview of the policies, practices, roadmaps, and strategies around funding, procuring, providing, accessing and sharing of services and resources in the EOSC scope in ES.

Table of Contents

1. Introduction

1.1. Aim and scope of this landscape analysis

This section reflects an analysis performed using the available information from previous activities related to gathering information on the scope of EOSC performed by the Ministry of Science and Innovation and the Spanish Landscape EOSC WG representative. It has been complemented and restructured to fit the questions and structure defined by task T5.1 in the EOSC-SYNERGY project.

The aim of the section is to collect and analyse the current situation of the Spanish institutions with respect to national and international initiatives in the area of Open Science, with special focus on Research Infrastructures that make it possible.

1.2. Definition/delimitation of "EOSC compliant resources"

In the Spanish Landscape on Research Infrastructures and e-Infrastructures, we identify two main categories:

- RIs as producers and consumers of Computing and Data processing services. In this category we can find first those included in the ESFRI Roadmap (such as LifeWatch ERIC, Instruct, EMSO, SKA, among others), which mainly act as data and service providers but could also include e-infrastructure offerings. Secondly, we can find the RIs included in the Map of Unique Scientific and Technological Infrastructures (ICTS-RI)[1], which provide unique (at national level) scientific resources and data. Some of them integrate RI facilities such as the Astronomy Infrastructures Network (RIA), the Coastal Observation System of the Balearic Islands Ocean Observing and Forecasting System (SOCIB), Spanish Antarctic Facilities (BAES), Spanish Antarctic Facilities (BAES), or FLOTA the Spanish Oceanographic Research Fleet.
- RIs as Computing and Data services providers, in which we identify four main categories: the Spanish Supercomputing Network (RES)[2], a distributed virtual infrastructure of 11 supercomputers; the Spanish Academic & Research Network (RedIRIS)[3] which operates in close collaboration with other research and education networks, both regional and international (GÉANT); the Iberian Grid Infrastructure (IBERGRID, NGI-EGI)[4] which federates and operates distributed computing and data resources across the Iberian area; and Recolecta (Open Science Harvester)[5] platform that gathers all the national scientific repositories. RECOLECTA is the

---

[1] http://www.ciencia.gob.es/portal/site/MICINN/menuitem.eed4570ef37d2c8fbaa777b9026041a0/?vgnextoid=928d5ef3677c4610VgnVCM1000001d04140aRCRD&lang_choosen=en
[2] https://www.bsc.es/marenostrum/access-to-supercomputing-resources/access-to-res
[3] http://www.rediris.es/index.php.en
[4] https://www.ibergrid.eu
[5] https://recolecta.fecyt.es/

Open Science Aggregator by FECYT in collaboration with REBIUN (Spanish Network of University and Scientific Libraries). RECOLECTA harvests both open access repositories (mostly institutional repositories but not only) as well as open access journals. The aggregator is based on D-NET software and has very recently released a new version. In total, RECOLECTA harvests 137 sources, https://buscador.recolecta.fecyt.es/repositories.

According to the Spanish map of ICTS in Spain, Research Infrastructures provide support to a wide range of disciplines:

- **Astronomy and Astrophysics**: Large Telescope of Canarias, Observatories of the Canary Islands, Astronomical Observatory of Calar Alto, IRAM 30M Radio Telescope, Astronomical Center of Yebes, Astrophysical Observatory of Javalambre and Underground Laboratory of Canfranc.
- **Ocean, Life and Earth Sciences**: Canary Islands Oceanic Platform, SOCIB, the Balearic Islands Ocean Observing and Forecasting System, Spanish Oceanographic Fleet, Spanish Antarctic Bases, Doñana Biological Reserve and Infrastructure for Bluefin Tuna Farming.
- **Health Sciences and Biotechnology**: Integrated Infrastructure for the Production and Characterization of Nanomaterials, Biomaterials and Systems in Biomedicine, Integrated Infrastructure of Omic Technologies, Network of High Biological Safety Laboratories, Distributed Network of Biomedical Image and Network of Nuclear Magnetic Resonance of Biomolecules.
- **Information and Communications Technology**: Spanish Supercomputing Network and RedIris
- **Energy**: Almería Solar Platform and National Fusion Laboratory.
- **Engineering**: Aggregate Infrastructures for Maritime Hydraulic Research.
- **Materials**: ALBA Synchrotron, Network of Micro and Nanofabrication White Rooms, Integrated Infrastructure of Electronic Materials Microscopy, Centre of Ultra-Intense Pulsed Lasers and National Center of Accelerators.
- **Social Sciences and Humanities**: National Centre for Research on Human Evolution

1.3. Information sources on services, repositories, facilities and infrastructures

According to the ESFRI Roadmap 2018[6], Spain has a relevant presence in ESFRIs related to the following disciplines:

- **Social Sciences and Humanities**: CESSDA, CLARIN, DARIAH, ESS-European Social Survey, SHARE.
- **Environment**: COPAL, EISCAT_3D Upgrade, EMSO, EPOS, EURO-ARGO, IAGOS, ICOS, LIFEWATCH, SIOS.
- **Energy**: ECCSEL, EU-SOLARIS, HiPER, IFMIF, JHR, MYRRHA y WindScanner.

---

[6] http://roadmap2018.esfri.eu/media/1050/roadmap18-part2.pdf

- **Health Sciences and Biotechnology**: ANAEE, BBMRI, EATRIS, ECRIN, ELIXIR, EMBRC, EU-OPENSCREEN, EuroBioImaging, ERINHA, Infrafrontier, INSTRUCT, ISBE y MIRRI
- **Physical Sciences and Engineering**: EMFL, ESRF Upgrade, EuroFel, ESS, European XFEL, ILL20/20 Upgrade, CTA, E-ELT, ELI, FAIR, KM3NeT, SKA, SPIRAL2.
- **Digit**: PRACE.

1.4. Between projects and national institutions

Spanish institutions consider strategic to have a good positioning in EOSC. For this purpose, the recently approved Spanish Network for Open e-Science, (RED2018-102377-T) - a thematic research network granted by the State Agency for Research (AEI) - gathers major stakeholders in the area of EOSC in Spain. The main objective of this network is to gather, organize, assess and distribute knowledge and information on EOSC. The network complements the activity of the future Spanish Network for e-Science and will be led by the same person. Additionally, Spain is leading EOSC-SYNERGY (https://www.eosc-synergy.eu/), which aims at expanding the capacity of EOSC in Spain. At this present moment, Spain has contributed with 10 services to the EOSC Marketplace (just a 4%). Although this has a relatively low importance as there are other Open Science services from Spanish integrated in international infrastructures and widely used without being integrated in the EOSC catalogue.

Spain has a strong participation in European Organizations related to Research Infrastructures. In particular, Spain participates in EGI through the Spanish Research Council. Spain has a Joint Research Unit in the area of Distributed Computing RIs that was created to set up the National Grid Initiative in 2007, with 14 signatory institutions. This JRU has been used to organize the participation in some RI projects, such as the EGI-InSPIRE or EGI-ENGAGE.

Spain also has an important position in PRACE. Through BSC-CNS, Spain has been participating in PRACE pan-European HPC infrastructure since 2007. PRACE has gone through 6 phases (by 2020), with the participation of Spanish centres in all of them. PRACE defines procedures for providing access to computing resources from European participants at large.

Currently, the BSC-CNS is also a key member of the EuroHPC Joint Undertaking, aiming to coordinate the efforts and sharing of resources to deploy in Europe a world-class supercomputing infrastructure and innovation ecosystem in supercomputing technologies, applications and skills.

Spain also contributes to EUDAT through BSC-CNS. EUDAT was a three-year project that delivered a Collaborative Data Infrastructure (CDI) with the capacity and capability for meeting future researchers' needs in a sustainable way. Currently EUDAT is implemented through a Collaborative Data Infrastructure Agreement.

It is also remarkable the alignment of the EOSC activities with the Spanish Initiative in Artificial Intelligence[7], which states in its fifth priority to develop an digital data ecosystem and to valorize infrastructures for its treatment.

Spain is also very active in the whole data lifecycle from ocean and coastal observing systems. As an example, the ICTS SOCIB -www.socib.es-, (the Balearic Islands Coastal Ocean Observing and Forecasting System) is a Marine Research Infrastructure, a multi-platform and integrated ocean observing and forecasting system that was born from IMEDEA (CSIC-UIB), and that provides streams of data, added value products, and forecasting services from the coast to the open ocean. It was initiated in 2008 and since 2014 it is included in the Spanish Large-Scale Infrastructure Map. SOCIB aims to characterize ocean state and variability from events to climate. All SOCIB data are quality controlled and made available in near real time for scientists and society under the terms of an open access policy, in line with European initiatives such as Copernicus Marine, EuroSea, Jerico-S3, EuroArgo, EuroArgoRISE, EMODnet, Seadatanet, EuroGOOS, etc. The data, scientific production, outreach and engagement activities, as well as tools and products developed are a clear performance indicator of SOCIB achievements and innovations in the new era of ocean observation[8]. The alignment of these elements is possible due to a dedicated data lifecycle management that is fully committed with the Findable, Accessible, Interoperable and Reusable (FAIR) data principles and that also contributes to the Ocean Best Practices System.

Finally, Spain has a relevant participation from ICTS in international collaborations related to the sharing of data. Relevant examples are:

- The Spanish Institute of Oceanography (IEO, http://www.repositorio.ieo.es/e-ieo/): Research public body ascribed to the Ministry of Science, Innovation and Universities with core budged allocated by the Government. The maintenance of the databases is connected to other in EU. IEO is partner of SeaDataNet (SND) www.seadatanet.org, a distributed Marine Data Infrastructure for the management of large and diverse sets of data deriving from in situ of the seas and oceans. Data are also been incorporated to long-term marine data initiatives like EMODnet, FixO3 or PSMSL among others. SDN gives access to data and useful information to researchers, individuals, companies, important European initiatives such as EuroFleets (New operational steps towards an alliance of European research fleets) and international initiatives such as POGO (Partnership for Observation of the Global Oceans).
- Spanish National Bio-Informatics Institute (INB) is one of the 19 nodes distributed across 13 institutions of ELIXIR (https://inb-elixir.es). INB is responsible in ELIXIR for co-managing the European Genome-Phenome Archive, besides other health related platforms that could contribute to EOSC, such as the Spanish Clinical Trials Research Network (SCREN), the Spanish Medical Innovation Platform (ITEMAS) and Spanish Biobank Network (Red Biobancos) and the Proteomics, Genomics and Cellular Lines Platform (PRB3).

---

[7] http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf

[8] https://www.frontiersin.org/articles/10.3389/fmars.2019.00568/full

- INSTRUCT is an ESFRI focused on structural biology. Spain contributes to INSTRUCT with the "Instruct Image Processing Center (I2PC)" in the Spanish Centre of Biotechnology CNB-CSIC.
- Lifewatch-ERIC, Infrastructure Consortium providing e-Science research facilities to scientists seeking to increase our knowledge and deepen our understanding of Biodiversity organisation and Ecosystem functions and services in order to support civil society in addressing key planetary challenges.

2. National policies and frameworks for open science support and collaboration

2.1. Formal regulations or publicly available policies that address

With respect to Open Data and research data preservation, the Spanish Strategy for Science, Technology and Innovation 2013-2020[9] envisages access to data and microdata, as well as publications and results publicly funded research as one of its six articulation mechanisms. The main aim of this initiative is to drive the development of repositories, own or shared, open access to the publications of its research staff, and establish systems to connect with similar initiatives nationally and internationally. Additionally, the Spanish Foundation of Science and Technology (FECYT) published a relevant report published by FECYT is the report on good practices for the management of research data (2012)[10], aiming to assist in the standardization of research data management in repositories to facilitate its preservation, access and distribution.

2.2. Strategies and policies for funding infrastructure services and resources

Research infrastructures and e-Infrastructures are funded in Spain through different means. At the national level, the State Plan of Scientific and Technical Research (2017-2020)[11] allocates the RDI public national budget addressing the specific goals outlined in the Strategy. The promotion of open science, open access and the strengthening of Research Infrastructures are covered in this plan. Additionally, the Spanish Roadmap for the European Research Area Development 2016-2020[12] supports the contribution of national RIs and Research Centres to EOSC.

At the regional dimension, the governments of the regions have initiated different research grants mainly leveraging European Regional Development Funds. The institutions in the Spanish RI have also received funds in the scope of these calls. One of the instruments that has been intensively used is the RIS3 (Research and Innovation Smart Specialisation Strategy).

All the funds have been allocated according to a selection based on a competitive process involving external evaluation agencies. Depending on the funding call, different concepts are eligible for funding. Infrastructure grant calls only fund equipment and installation costs, which excludes staff for operation

---

[9] http://www.idi.mineco.gob.es/stfls/MICINN/Investigacion/FICHEROS/Estrategia_espanola_ciencia_tecnologia_Innovacion.pdf
[10] https://www.recolecta.fecyt.es/sites/default/files/contenido/documentos/informe_datos_cientificos_en_esp.pdf
[11] http://www.ciencia.gob.es/stfls/MICINN/Investigacion/FICHEROS/Spanish_RDTI_Plan_2013-2016.pdf
[12] https://ec.europa.eu/research/infrastructures/pdf/roadmaps/spain_national_roadmap.pdf#view=fit&pagemode=none

and user support. Funding for user support must be obtained through other research calls or institutional funding.

Currently, funding organizations do not require infrastructures to provide the cost information about the services they offer. However, and considering initiatives of the European Commission such as the Virtual Access to Research Infrastructures, it is likely that most RIs will start auditing their costs and will collect such information.

The **National *Unique Scientific and Technical Infrastructures* (ICTS)** are included under the *ICTS Map[13]* which includes those infrastructures which have been positively evaluated according to the criteria established by the Scientific, Technological and Innovation Policy Council and evaluates those that are economically viable. ICTS are implemented in a collaborative model among central and regional administrations which allows that RIs included in both the ICTS and ESFRI Roadmaps are supported by the Government (Central and/or Regional Administration) while stakeholders (RPOs, University or Public Partnership) manage the RIs in the framework of a Consortium Agreement.

Funding to RIs in Spain comes from multiple sources. RIs included in both the ICTS and ESFRI Roadmaps are supported by the Government (Central and/or Regional Administration) while stakeholders (RPOs, University or Public Partnership) manage the RIs in the framework of a Consortium Agreement. It is important to note that these RI are mainly funded by the stakeholders (own budget from public and private funds). Nevertheless, these stakeholders can apply for public national, regional and local funding for scientific equipment for the RI and ERDF funds to promote access to their services, network operation, to stimulate the participation of the business sector as users of its scientific-technological services, as suppliers of technology and services ("science industry").

The user support, development of service layers is also funded through research projects, mainly through the calls of the HORIZON 2020 - 1.4. (Excellent Science - European Research Infrastructures)[14]. Up to March 2020, 6675 projects[15] under this call have Spanish partners (24% of all the projects funded), although this figure does not include only partners which are part of a RI. At the national level, the National Subprogram of Scientific Infrastructures and Equipment of the National Plan for Scientific and Technical Research of Innovation 2017-2020[16] had a budget of 257M€. Additionally, regional governments have also developed lower-scale funding programs in this regard.

---

[13] http://www.ciencia.gob.es/stfls/MICINN/Investigacion/FICHEROS/ICTS_EN_ED2019.pdf
[14] https://cordis.europa.eu/programme/id/H2020-EU.1.4.
[15] https://cordis.europa.eu/search/en?q=contenttype%3D%27project%27%20AND%20(programme%2Fcode%3D%27EU.1.4.%27%20OR%20programme%2Fcode%3D%27H2020%27)%20AND%20relatedRegion%2Fregion%2FeuCode%3D%27ES%27&p=1&num=100&srt=Relevance:decreasing
[16] http://www.ciencia.gob.es/stfls/MICINN/Prensa/FICHEROS/2018/PlanEstatalIDI.pdf

2.3. Present status with regard to Commission Recommendation (EU) 2018/790 on access to and preservation of scientific information

Spain is actively involved with the European Commission in promoting the design of a European Scientific Data Cloud (EOSC) as a federated system to store, share and reuse in an interdisciplinary way the data generated in scientific and innovation projects financed with public funds, using quality standards, certification and common rules of participation.

The **Spanish Strategy on Science, Technology and Innovation 2013-2020[17]** is the RDI policy for the State Administration and the Regions sharing vision with Europe2020 Strategy, Innovation Union Flagship and H2020, addressing key actions in close connection to the goal of EOSC:

- To strengthen strategic planning and coordination of the Spanish Landscape of Research Infrastructures (RIs), included in the ESFRI Roadmap and in the Map of Unique Scientific and Technological Infrastructures (ICTS) through the the Council of Scientific, Technological and Innovation Policy and its Executive Commission.
- To promote e-Infrastructures and data sharing.
- To strengthen R&D Institutions for developing or improving their data infrastructures.
- To promote open access to publication and public funded research results, adopting measures, and shared standards at all levels of administrations and research institutions.

In order to boost access to research data, funded R&D projects may include, optionally, a research Data Management Plan that will be uploaded to institutional, national and/or international repositories after the completion of the project and after the deadline defined in the corresponding calls.

However, all situations in which they must be protected for reasons of confidentiality, security, protection, etc. will be respected, as well as when they are necessary for the commercial exploitation of the results obtained.

Finally, Spain has set up the National e-Science Network which aims at promoting and discussing policies for supporting researchers with computing and data services. Expected to be established by early 2020. This network would be similar to a "national EOSC".

---

[17] http://www.ciencia.gob.es/stfls/MICINN/Investigacion/FICHEROS/Politicas_I+D+i/Resumen_Ejecutivo_Estrategia_Espanola_01022013.pdf

3. EOSC compliant resources

3.1. Characteristics of services/resources

Spain is already contributing to 10 EOSC services in the EOSC marketplace. This figure is abnormally low (4%) as many EOSC-candidate services are not listed in such a marketplace. Another reason is that some aggregated services such as EGI Compute cloud include Spanish centres in the federation. According to the activity of the Spanish institutions in the RI activities, it is likely to expect contributions in:

- ICT: Spain is highly active in Distributed computing and Supercomputing, both at the level of the provisioning of research infrastructures and as the platform level, with services for efficiently exploiting such resources. Examples such as Deep Learning as a Service or Cloud Orchestration can be already found in the EOSC-Marketplace. It is envisaged that in the coming months new services for software quality, data analytics, distributed programming models (among others) will arise.
- Data Services: Spain is strongly organizing the research repositories through network activities and aggregators that facilitate the searching and access to digital objects. It is envisaged that this activity will be improved with efficient linkage of research articles with data and processing services and long-time data preservations services. Institutional repositoires in Spain are mostly managed by University/research libraries. Two are the main channels whereby cooperation and harmonization take place at national level: on the one hand, REBIUN Working Group of Repositories, with a number of action lines ranging from quality, preservation, to metrics and FECYT through dedicated and ad hoc working groups, mostly related to National guidelines for repositories (currently under review) and to monitoring of Science Law compliance (art. 39 requests the deposit and OA availability of state funded publications in open access repositories)

- In addition, FECYT, CSIC and ISCII have been working together in INEOS since 2018, a pilot project to promote research data management and publishing in institutional repositories and raise awareness about good practices and value added functionalities. DIGITAL.CSIC participates in INEOS by sharing good practices in research data management and integration of value added services; semantic enrichment of deposited outputs coming from State funded projects, and integration of ORCIDs and public CVN on repositories infrastructures

- Thematic services. Scientific communities are the key innovators for boosting the EOSC concept. They gather data, processing algorithms and knowledge, which is key for building useful services. The presence of such services in the EOSC-marketplace is still very low, and we expect that services linked to ESFRIs, ICTS and reference centres in biomedicine, earth observation,

climate and environment, energy and engineering will become EOSC-compliant in the short future.

The direct or indirect use of ERDFs for the acquisition of infrastructure resources normally impossibilities obtaining revenues for the access to the resources by third parties. Industrial contributions should go through other modalities, such as advisory, training, patronage and targeted research contracts, which normally focus on the added value, although they may include some kind of access to the RI. Therefore, EOSC-related Research Infrastructures do not normally obtain direct additional revenues.

All service providers in the Research Infrastructures have to go sign some kind of Service Level Agreements (SLAs). For example, EGI Cloud compute EOSC service, service providers negotiate and sign SLA contracts with the users who made the requests in which service providers commit to provide the resources agreed. Moreover, being part of the EGI Compute federated infrastructure also requires to achieve some quality metrics in terms of availability, which are monthly revised and published. Breaching institutions are temporarily suspended. Reiterative breaches on availability lead to the definitive exclusion of the defaulting providers.

One of the requirements for the publication of a service into the EOSC marketplace is to publish a Service Level Agreement. Therefore, all services that want to be exposed in the EOSC portal must elaborate one.

The main limitation in SLAs for research services is to reach the levels of response and availability that production-level services have, especially after business hours.

The expectatives of Spain in EOSC are mainly focused on three directions:
- To align the national initiatives related to EOSC by adopting European Best Practices and contribute to the EOSC design with the perspective of Spain.
- To contribute to the implementation of EOSC by providing services, data and resources through the Spanish RIs.
- To boost the international positioning of Spanish researchers by exposing their services to a wider audience and to leverage other institution's services, data and resources to improve their research.

Spanish institutions do not expect to find funding for covering international users or the implementation of exclusion policies based on the geographic location.

3.2. Data Services

3.2.1. Data management, curation and (long-term) preservation

Research infrastructures in Spain deal with different data and therefore there are different metadata management procedures.

The IBERGRID Distributed Computing infrastructure hosts data on the usage statistics. This data is very relevant for ICT developers and business models as it provides crucial information on access patterns,

workload characterization and usage of resources. This information is collected by the EGI accounting portal[18], which permits different filtering based on geographic criteria, resource providers, research disciplines and timeline. The information can be nicely browsed on the web or can be downloaded as JSON or CSV files which are properly annotated. Access to the accounting data does not require authentication.

Recolecta provides access to 137 resources, and provides an advanced search that enables filtering by repository, document type, language, production date, author (name or ORCID), title or editor, among other fields, redirecting to the to the open access repository or journal that hosts the resource. It does not redirect to "catalogues".

DIGITAL.CSIC is the institutional Open Science repository of the CSIC that gathers more than 205.000 research outputs by existing and past CSIC institutes and centers. 62% of such resources are available in open access at time of writing. Despite this repository mainly includes scientific publications, it is remarkable that it provides access to nearly 12.000 scientific datasets from a wide range of scientific research disciplines. DIGITAL.CSIC published its Research data policy in 2013[19].

Most datasets that are uploaded into DIGITAL.CSIC are associated with publications however in other instances the repository hosts collections of datasets that are the outcome of long term research projects where the data stand as main results. CSIC Open Access Mandate[20], which entered into force in April 2019, requests CSIC researchers to deposit and make their data associated with publications FAIR in DIGITAL.CSIC.

Data in the repository has to fulfil the publication policy[21], which includes quality guidelines. The datasets include ICTS in Spain also act as Scientific Data Research Infrastructures. There are examples such as:

- The SeaDataNet (SND) www.seadatanet.org, a distributed Marine Data Infrastructure for the management of large and diverse sets of data deriving from in situ of the seas and oceans in which the Spanish Institute of Oceanography (IEO) and CSIC contribute to.
- The Thematic Repository of Research Data (REDIA) from the National Institute for Agricultural and Food Research and Technology (INIA).
- The datas services from the Biological Station of Doñana, available in http://icts.ebd.csic.es/servicios-tic.
- The Canary Islands Astrophysics Institute (IAC), which provides his researchers and collaborators with fast storage and processing services.
- The NANBIOSIS platform, which provides support for the design and validation of biomolecules, with the support of the supercomputing node of the University of Zaragoza.

---

[18] https://accounting.egi.eu/

[19] http://digital.csic.es/dc/politicas/politicaDatos.jsp

[20] https://digital.csic.es/handle/10261/179077

[21] http://digital.csic.es/dc/politicas/ >

- The Infrastructure of Omics Technologies (ICTS OmicsTech - https://www.omicstech.es/), which offers researchers access to facilities, dedicated to omics technologies, including large computational facilities.

All the Open Science data harvesters provide searching services that enable filtering by metadata, which is in some cases automatically retrieved from the external repositories. The metadata is typically coded following the Dublin Core metadata schema and can be retrieved through OAI-PMH (Open Archive Initiative-Protocol for Metadata Harvesting). Authorship is identified mainly through ORCID. Metadata is stored in the original language of the scientific item, which is normally English or Spanish. Machine-readable access is possible through the web end-points, although it requires web crawling in most cases. Data from research infrastructure usage is downloadable from a web URI in XML or JSON. There is still a need for an effective and simple API mechanism to query the databases, which surely will become widely available in the near future. The Guidelines for Repositories, an initiative by FECYT and a few institutional repositories including DIGITAL.CSIC, is currently under review and addresses new recommended ways to enhance interoperability, including usage of additional APIs and protocols

With respect to certification, only the DIGITAL.CSIC repository has been certified with the Data Seal of Approval since 2015.

Open Science harvesters provide some best practices to foster the FAIRness of the data. Among the information requested, we outline the need for a PID; the availability of documentation for the understanding of the data; access to metadata; clear usage license that permits reutilization; usage of standard and clear format models; linkage to other references or research results.

There is no systematic study in our knowledge about the reticence of Spanish Researchers for publishing data openly. However, there are some articles that debate on this point according to the knowledge they have on the discipline. Also CSIC Abierto Bulletin by DIGITAL.CSIC gathers a number of interviews to CSIC researchers about research data management and open data[22]. In those studies[23, 24, 25,26], the main issues listed are:

- The considerable effort required to publish data following FAIR principles.
- The difficulty of the recognitions on the sharing data.
- Fears of losing control on the intellectual property.
- Perception of a lower quality of digital objects in open repositories.

The data collection available in the Spanish RIs cover different types of data, depending on the discipline, such as large textual information from genomic databases, numerical information from high-energy

---

[22] https://digital.csic.es/handle/10261/150210

[23] https://datos.gob.es/es/noticia/directrices-para-la-apertura-de-los-datos-de-investigacion-desde-la-perspectiva-juridica

[24] https://www.consorciomadrono.es/noticias_eventos/2015/JornadaPGD/teresaVictoria.pdf

[25] https://www.redalyc.org/pdf/5355/535554765001.pdf

[26] https://digital.csic.es/handle/10261/92404

physics, still images and videos from medical imaging databanks, geospatial repositories and digitalised information from e-humanities.

3.2.2. Data Sharing and Access

Access policies to the resources of the Research Infrastructures typically follow different criteria, according to its nature.

- Supercomputing infrastructures from the Spanish Network of Supercomputing provide access to the resources based projects which go through competitive calls, reviewed by a scientific panel. Projects selected grant access for free to a fraction of the resources that depend on the requirements, previous achievements and the scientific evaluation. Requests should normally be supported by a Spanish institution, with the potential collaboration of other institutions. A fraction of the resources in the regional centres is dedicated to users from regional institutions. Users grant individual access to the resources and data.
- The IBERGRID Distributed and cloud computing facility provides access to scientific users based on the enrollment to thematic VOs. The on-boarding process is coordinated with the EGI cloud compute and HTC on-boarding procedures, which go through a negotiation with the providers. There is no restriction based on geographic criteria. Access to the resources is normally obtained as a combination of group membership and individual access, according to the service rules of the providers.
- The Spanish Academic Network REDIRIS is RedIRIS has an open peering policy and will generally accept public peering requests from any organization, subject to certain technical, commercial and legal requirements[27]. Private peering requests are also considered but managed under a selective peering policy. RedIRIS has over 500 affiliated institutions, mainly universities and public research centres. User access is based on the membership to the affiliated institutions. Only Spanish institutions can participate, although indirectly any user connected to the International Academic networks (or the whole Internet) can access resources in REDIRIS.
- The Open Science Harvester Recolecta RECOLECTA harvests open access reposiotries and open access journals in Spain, mostly from Universities and research performing organizations. RECOLECTA has its guidelines for data contributors to be harvested via OAI-PMH. DIGITAL.CSIC is the largest repository amongst RECOLECTA content providers.
- ICTS and ESFRI define their own access policies and depend on the operation cost of their infrastructures. Normally, they do not impose any geographic restriction.

The access to the infrastructures or data from the sources investigated is free of charge. It does not preclude service providers to restrict access to users that reach a specific level of scientific excellence and fulfil ethical requirements.

---

[27] http://www.rediris.es/rediris/instituciones/afiliacion.html.en

The Spanish Strategy of Science, Technology and Innovation 2013-2020 defines the policy for Open Access to results and Research Data of research activities subsidized with public resources. The works published in scientific journals funded by the State Plan will be deposited in repositories (institutional and / or international thematic or international. FECYT has a working group to monitor to what extent Spanish repositories comply with Science Law open access clause), in open access considering the specific characteristics of the different subjects, in compliance with the provisions of Article 37 of Law 14/2011, from June 1, on Science[28], Technology and Innovation and the recommendations linked to the European agenda on open access and open science[29].

In order to boost access to research data, funded R&D projects may include, optionally, a research Data Management Plan that will be uploaded to institutional, national and/or international repositories after the completion of the project and after the deadline defined in the corresponding calls. Moreover, the repository of the Spanish Research Council (DIGITAL.CSIC) has a specific access policy[30] (it is the repository policies as regards types of contents and collections, data and metadata use, privacy, statistics, content retention, services to institutional users etc) that implements the CSIC Institutional Open Access Mandate[31]. Institutions will surely define their access policy in the coming years, as DMPs will become compulsory in further research programs. However, all situations in which they must be protected for reasons of confidentiality, security, protection, etc. will be respected, as well as when they are necessary for the commercial exploitation of the results obtained.   There is no specific mention to the FAIR principles in the strategy although most the initiatives are currently considering the adoption of FAIR principles in the design.

The access to personal data requires special regulations. This mainly happens with the research data from health systems. Spain has 6 ICTs and participates in 12 ESFRIs related to health. The management of personal information requests the approval of ethical committees appointed and the approval of the contributor entity in case of federated resources. For example, The Spanish node of ELIXIR manages an access instance of the European Genome-phenome Archive[32], which requests the approval of the individual contributors. The Spanish nodes of the EuroBioImaging ESFRI dedicated to imaging databanks also follow the same approach. Data is anonymised in most cases (except for medical data which typically is dissociated or it cannot be totally anonymous by definition, like in the case of genomic data).

The licenses of the data publicly shared depend on the institution. Institutions have elaborated DMPs which include the data licensing models. The EGI accounting portal, for example, licenses data under the Creative Commons. The data from the institutional repositories, indexed in the Open Science Harvester

---

[28] https://www.boe.es/buscar/pdf/2011/BOE-A-2011-9617-consolidado.pdf
[29] https://ec.europa.eu/research/openscience/index.cfm
[30] http://digital.csic.es/dc/politicas/
[31] https://digital.csic.es/bitstream/10261/179077/5/CSIC_institutional_open_access_mandate.pdf
[32] https://www.ebi.ac.uk/ega/submission/phenotypes

also depend on the individual open science policies of the institutions. In most cases the publications inherit the conditions imposed by the editors for Open Access or author's versions.

The access to the resources also depends on the infrastructures. REDIRIS provides support to EDUGAIN[33], a pan-european identity confederation for the research and education environment. This IdP is supported by EGI Check-in, which is the preferred method in IBERGRID, together with other AARC compliant methods such as IAM developed in INDIGO-Datacloud. The Spanish Network for Supercomputing uses its own local username and password mechanism, and other RI use either their own authentication mechanisms or EDUGAIN. REDIRIS Identify federation Service (SIR2) uses REFEDS Research and Scholarships.

Thematic services which based the authentication on external AAIs (such as EGI Check-in) manage the authorisation either on individual basis or through Virtual Organisations (VOs). The VO-based authorisation typically consists on a group membership managed by an external recognized manager that authorises or revokes the membership according to a verification process and the signing of the User Access Policy. This is the case of the Distributed Computing infrastructure of IBERGRID (both cloud and High-Throughput compute services) and all the scientific services relying on it.

---

[33] https://www.rediris.es/sir/edugain/

4. Procurement of and transnational access to services and resources compatible with EOSC

4.1. Transnational access to national resources/services

All the services identified in this landscaping activity allow transnational access

4.2. Potential for harmonization of national policies for:

4.2.1. Joint Procurement

In the specific case of supercomputing, Spain participates through BSC in the The European High Performance Computing Joint Undertaking (EuroHPC JU) which has launched today a call for tender for the procurement of three world-class precursors to exascale supercomputers, one of them for Spain.

The legal complexity and the administrative procedures makes collaborative service or hardware procurement complex.

Procurements take place at the level of the institution for recurrent services although some efforts are being taken place for joint procurement of services related to publications.

4.2.2. Coordinated service provisioning

Spain has organized both distributed computing and supercomputing service provisioning in two different forms. On one hand, the Spanish Supercomputing Network (RES)[34] is a distributed virtual infrastructure of 11 supercomputers that provides a single access to a set of federated resources (reserving a part of the resources of the datacentres for regional use). On the other side, Spain created a Joint Research Unit in the area of Distributed Computing RIs to set up the National Grid Initiative in 2007, with 14 signatory institutions. This JRU (ES-GRID) has been used to organize the participation in some RI projects, such as the EGI-InSPIRE or EGI-ENGAGE. The participation of the institutions is coordinated under the the Iberian Grid Infrastructure (IBERGRID)[35] which federates and operates distributed computing and data resources across the Iberian area, tightly coordinated with Portugal.

4.3. Procuring services/resources

Procurement of services and resources is mainly constrained to the obligations of the Spanish Law of Contracts for the Public Sector[36] which implements the Directives of the European Parliament and

---

[34] https://www.bsc.es/marenostrum/access-to-supercomputing-resources/access-to-res
[35] https://www.ibergrid.eu

European Council 2014/23/UE and 2014/24/UE, modified in the Decree 3/2020 of February the 4th, as well as the Regulations on the use of European Regional Development Funds[37].

In this sense, services or resources below 15.000 € (taxes excluded) can be exceptionally contracted without the need of a tender. This limit rises up to 50.000 € in the case of services or resources for a specific research project. If funding comes from ERDF or European Funds, procedures require at least three different competitive offers. In the case of recurrent contracts (e.g. pay-per-use cloud resources on public cloud platforms), it is strongly advisable that the institutions go through pre-negotiated contracts through public tenders.

For any service or resource contract above such limit, a public tender is required. Public tenders normally base on an objective evaluation based on quantitative criteria plus an expert evaluation by a committee. Public tenders normally take 4-5 months to complete.

The unit cost of the services that rely on service contracts can be estimated more easily than those that rely on the Total Ownership Costs of acquired resources. However, it is fairly evident that in many cases it is more efficient to rely on own infrastructures, so effort on the evaluation of unit cost is needed. Moreover, initiatives of the European Commission such as the Virtual Access to Research Infrastructures will foster this type of analysis.

---

[36] https://www.boe.es/buscar/pdf/2017/BOE-A-2017-12902-consolidado.pdf

[37] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/825719/ESIF-GN-1-001_ESIF_National_Procurement_Requirements_v6.pdf

5. Conclusions

This document has briefly analysed the situation with respect to open science data, research infrastructures to support them and open science policies in Spain. The document has been prepared by collecting the information from previous existing landscape activities, mapping this information to the sections and questionnaire items of the agreed template and completing the missing gaps with strategic documents from the Ministry of Science and Innovation. Finally, the document has been shared with the EOSC GB Spanish representatives.

The document has been produced using mainly these three sources of information:

- A landscape analysis from the Spanish General Directorate of Internationalization of Science and Innovation, from the Ministry of Science and Innovation, based on a questionnaire, which was distributed along the major stakeholders related to EOSC in Spain.
- The contribution to the EOSC-WG Landscape analysis from Spain.
- Official documents from the Ministry of Science and Innovation (Singular Scientific and Technological Infrastructures Roadmap, Spanish Law for Science, Spanish Strategy for Science, Technology and Innovation 2013-2020).

The landscape in Spain shows two main categories of Research Infrastructures and e-Infrastructures: RIs as producers and consumers of Computing and Data processing services, which are mainly related to ESFRIs and Singular Scientific and Technological Infrastructures (ICTS); and RIs as Computing and Data services providers, such as the Spanish NREN, the Spanish Supercomputing Network, IBERGRID or the repository of DIGITAL.CSIC.

As major findings we outline the strong international character of Spanish RIs, which is reflected in a strong positioning on international initiatives. Spain has been a member of EGI through a JRU since 2007, participates in PRACE since its start where BSC-CNS is a key member of the EuroHPC Joint Undertaking, and the Spanish ICTS and other RIs participate in 48 ESFRIs. Spanish RIs are open to international access with generally no differences in cost or SLAs between Spanish and foreigner users.

Additionally, Spanish policies in Open Science and Research Infrastructures are aligned with EOSC EU Policies, with special emphasis on strengthening planning and coordination of the Spanish ESFRIs and ICTS and considering EOSC as an opportunity to strengthen development or improvement of data infrastructures. Spain plans to align national initiatives related to EOSC and contribute to the EOSC design and implementation by providing services, data and resources.

The main sources for funding of Research Infrastructures are through competitive processes involving external evaluation agencies. At the regional dimension, the governments of the regions opened research

grants mainly leveraging European Regional Development Funds. At the national level, the State Plan of Scientific and Technical Research (2017-2020) and Spanish Roadmap for the ERA Development 2016-2020 - had managed a budget of 257M€. At European level, there is a Strong participation in calls of the HORIZON 2020 - 1.4. (Excellent Science - European Research Infrastructures) with 6675 projects including Spanish partners (24%) up to March 2020. There is however an important challenge to vertebrate cross-connection between HPC, Distributed Computing, Data Repositories and Networks.

Infrastructure grant calls only fund equipment and installation costs, excluding staff for operation and funding for operation & user support must be obtained through other research calls or institutional funding.

However, there is a low visibility of Spanish RI services in EOSC. This is reflected in a small presence of services from Spanish institutions in the EOSC Marketplace, with only 10 services listed (4%). It is important to understand if there is low interest (a low cost-benefit perception) or a lack of information.

Appendix A – Acronyms

AAI	Authentication and Authorization Infrastructure
AEI	Joint Research Unit
AES	State Agency for Research
ANAEE	Infrastructure for Analysis and Experimentation on Ecosystems
BAES	Spanish Antarctic Facilities
BBMRI	Biobanking and BioMolecular Resources Research Infrastructure
BSC-CNS	Barcelona Supercomputing Centre - Centro Nacional de Supercomputación.
CDI	Spanish Institute of Oceanography
CESSDA	Consortium of European Social Science Data
CLARIN	Common Language Resources and Technology Infrastructure
SCReN	Spanish Clinical Trials Research Network
CSIC	Spanish Research Council
CTA	Cherenkov Telescope Array
DARIAH	Digital Research Infrastructure for the Arts and Humanities
AAI	Authentication and Authorization Infrastructure
E-ELT	Extremely Large Telescope
EATRIS	European Advanced Translational Research Infrastructure in Medicine
ECCSEL	European Carbon Dioxide Capture and Storage Laboratory Infrastructure
ECRIN	European Clinical Research Infrastructure Network
EDUGAIN	REDIRIS Identify federation Service
EGI	European Grid Initiative

EGI-ENGAGE	Engaging the EGI Community towards an Open Science Commons
EGI-InsPIRE	Integrated Sustainable Pan-European Infrastructure for Researchers in Europe
EISCAT	European Incoherent Scatter Radar System
ELIXIR	European life-sciences Infrastructure for biological Information
EMBRC	European Marine Biological Resource Centre
EMFL	European Magnetic Field Laboratory
EMSO	European Multidisciplinary Seafloor and water column Observatory
EOSC	European Open Science Cloud
EPOS	European Plate Observing System
ERIC	European Research Infrastructure Consortium
ERINHA	European Research Infrastructure on Highly Pathogenic Agents
ESFRI	European Strategy Forum on Research Infrastructures
ESS	European Social Survey
EU-OPENSCREEN	European Infrastructure of Open Screening Platforms for Chemical Biology
EU-SOLARIS	European Research Infrastructure for Concentrated Solar Power
EURO-ARGO	Global Ocean Observing Infrastructur
EUROBIOIMAGING	European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences
EuroHPC JU	European Open Science Cloud
FAIR	Findable Accessible Interoperable Reusable
FAIR	Facility for Antiproton and Ion Research
RISSS	Research and Innovation Smart Specialisation Strategy
FLOTA	Spanish Oceanographic Research Fleet

HIPER	High Power Experimental Research Facility
HPC	Collaborative Data Infrastructure
I2PC	Spanish Centre of Biotechnology
OAI-PMH	Open Archive Initiative-Protocol for Metadata Harvesting
IAGOS	In-service Aircraft for a Global Observing System
IBERGRID	Iberian Grid Infrastructure
ICOS	Integrated Carbon Observation System
ICT	Service Level Agreements
ICTS	Singular Scientific Technological Research Infrastructures (Infraestructuras Singulares Científico-Recnológicas).
IdP	Virtual Organization
IEO	Partnership for Observation of the Global Oceans
IAC	Canary Islands Astrophysics Institute
IFMIF	International Fusion Materials Irradiation Facility
INFRAFRONTIER	European Research Infrastructure for the generation, phenotyping, archiving and distribution of mouse disease models
IEO	Spanish Institute of Oceanography
INSTRUCT	Integrated Structural Biology Infrastructure
ISBE	Infrastructure for System Biology Europe
ITEMAS	Proteomics, Genomics and Cellular Lines Platform
JHR	Jules Horowitz Reactor
JRU	Joint Research Unit
KM3NET	KM3 Neutrino Telescope

LIFEWATCH	Science and Technology Infrastructure for Biodiversity Research
MIRRI	Microbial Resource Research Infrastructure
MYRRHA	Multi-purpose hYbrid Research Reactor for High-tech Applications
NGI	National Grid Initiative
DMP	Data Management Plan
POGO	Instruct Image Processing Center
PRACE	Partnership for Advanced Computing in Europe
FECYT	Spanish Foundation of Science and Technology
REDIA	National Institute for Agricultural and Food Research and Technology
REDIRIS	Spanish Academic & Research Network
RES	Spanish Supercomputing Network
RI	Resarch Infrastructures
RIA	Astronomy Infrastructures Network
RIS3	Research and Innovation Smart Specialisation Strategy
RPOs	Information and Communication Technologies
SCREN	Spanish Medical Innovation Platform
SHARE	Survey of Health, Ageing and Retirement in Europe
SIOS	Svalbard Integrated Arctic Earth Observing System
SIR2	Identity Provider
SKA	Square Kilometer Array
SLA	Thematic Repository of Research Data

SOCIB	Balearic Islands Observing and Forecasting System
SPIRAL2	Système de Production d'Ions Radioactifs en Ligne de 2e génération
VO	European High Performance Computing Joint Undertaking
XFEL	European X-Ray Free-Electron Laser