European
Commission

Study to achieve an increase in the scientific exploitation of data from European space missions

"Science Data"

Final report

Written by Strategy&
September 2015

strategy&
Part of the PwC Network

EUROPEAN COMMISSION

Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (DG GROW)

Directorate Space Policy, Copernicus and Defence

Space policy and research Unit I-1

Contact: Nicolas Peter

E-mail: Nicolas.PETER@ec.europa.eu

European Commission

B-1049 Brussels

EUROPEAN COMMISSION

Study to achieve an increase in the scientific exploitation of data from European space missions

"Science Data"

Final report

Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (DG GROW)
Science Data
2015 EN

LEGAL NOTICE

Luxembourg: Publications Office of the European Union, 2015

ISBN:978-92-79-52398-4
doi:10.2873/839169

Author: Paolo Ariaudo and the team at PwC Strategy& (Netherlands) B.V. Acting as sub-contractor to PwC Strategy& (Germany) GmbH.

Contributions: SpaceTec Partners was the main contributor to the analysis of the previous data exploitation initiative documented in Chapter 4.

Study analysis was completed at the end of June 2015

Date: 20 September 2015

Table of Contents

1 INTRODUCTION

This document is the final report for the "Study to achieve an increase in scientific exploitation of data from European space missions," performed primarily in the first half of 2015 under the framework contract number ENTR/2009/050/ Lot1.. This study has been performed by Strategy& (Part of the PwC Network) with the support of SpaceTec Partners for the analysis of the datasets concerning on-going data exploitation initiatives documented in chapter 4.

1.1 Study objectives and scope

The objective of the study is to provide the European Commission with elements in support of the development of future research activities and approaches aimed at maximizing the exploitation of scientific data produced by space missions.

Building upon the efforts of FP7, Horizon 2020 considerably intensifies its efforts towards increased data exploitation from space sources. As established in the Horizon 2020 Work Programme LEIT – Space 2014-2015: "Exploitation of space science data is being addressed across H2020 on a recurring basis, ensuring a more extensive utilisation of scientific data originated from European missions and missions with European participation ."

The study analyses are focused on how data from European space missions is scientifically exploited and how that exploitation can be monitored and increased. The primary use of the study output is intended to:

- Feed the formulation of future activities where proactive stimulus could be introduced within Horizon 2020 Space;
- Establish concepts for effective monitoring mechanisms for space data exploitation in Europe; and
- Feed the assessment of current science data exploitation in FP7 and under Horizon 2020.

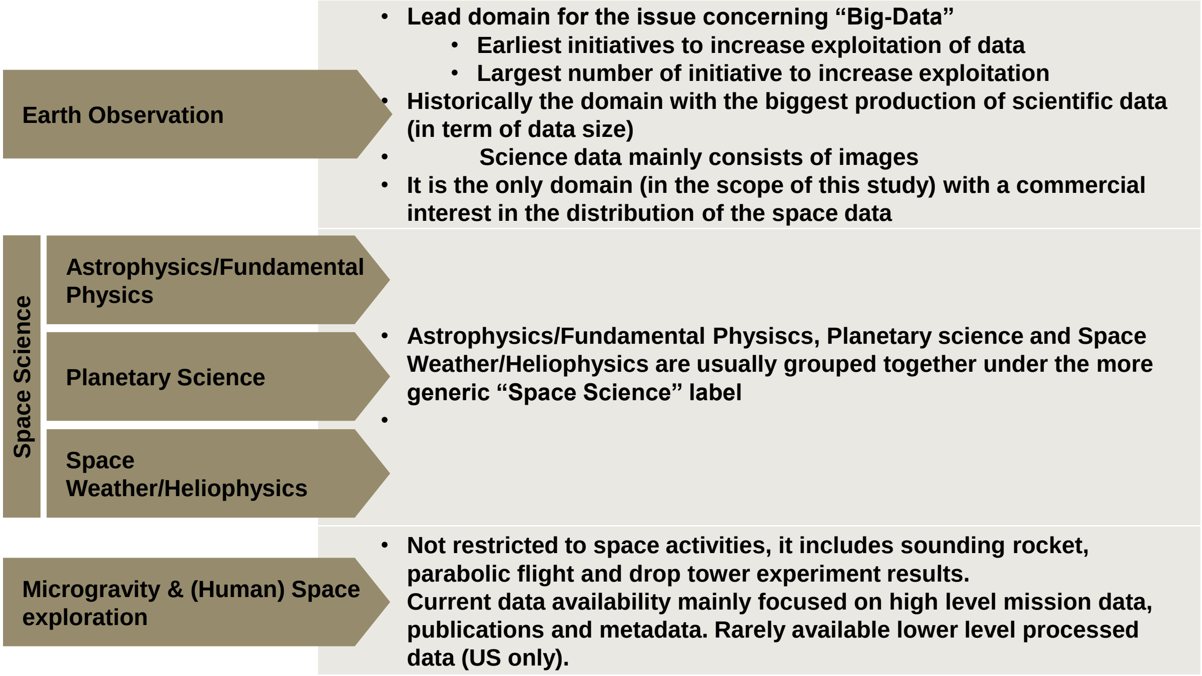
1.1.1 Study scope

Space missions are very complex endeavours with schemes of governance, funding and utilized technologies varying dramatically according to the typology of missions (e.g. civil, dual-use, military) and to the targeted application domain (science, commercial, defence). Civil missions can be further classified as institutional missions (i.e. owned and funded by public, governmental or international institutions) and commercial missions (i.e. owned and funded by private satellites operators).

Furthermore, implementation of a space mission includes the management and operations of the systems components (e.g. satellite platforms, International Space Station (ISS) modules) and the mission component (e.g. EO payloads for satellite missions, scientific laboratory and payloads on the ISS). Many different mixed ownership and funding schemes (e.g. institutional and/or commercial) for the system and mission components of the mission can be identified.

The study analyses are focused on the European institutional missions targeting applications in the science domains, and in particular in the fields of:

- Earth Observation;
- Astrophysics/Fundamental Physics;
- Planetary Sciences;
- Heliophysics (and Space Weather); and
- ISS and Space Exploration (human)



Earth Observation

- Lead domain for the issue concerning “Big-Data”
  - Earliest initiatives to increase exploitation of data
  - Largest number of initiative to increase exploitation
- Historically the domain with the biggest production of scientific data (in term of data size)
- Science data mainly consists of images
- It is the only domain (in the scope of this study) with a commercial interest in the distribution of the space data

Space Science

Astrophysics/Fundamental Physics

Planetary Science

Space Weather/Heliophysics

- Astrophysics/Fundamental Physiscs, Planetary science and Space Weather/Heliophysics are usually grouped together under the more generic “Space Science” label
-

Microgravity & (Human) Space exploration

- Not restricted to space activities, it includes sounding rocket, parabolic flight and drop tower experiment results.
  Current data availability mainly focused on high level mission data, publications and metadata. Rarely available lower level processed data (US only).

Figure 1: grouping of space research fields

In particular as described in **Figure 1**, above, in the course of the study these five fields of research have been grouped in three different domains: Earth Observation, Space Science (grouping Astrophysics/Fundamental Physics, Planetary Sciences, Heliophysics and Space Weather), and Microgravity (Human) Space exploration.

Earth Observation is the domain with the biggest production (in volume) of scientific data where the earliest activities to foster data exploitation have been historically implemented. Today Earth observation is the domain with the most advanced solutions for the data preservation and scientific exploitation.

In addition Earth Observation it is the only domain with the strongest commercial interests in the utilisation of the mission data. The duality aspects with the scientific utilisation have also been considered in this study, limited to commercial endeavours such as Airbus-Geo and e-Geos based on the utilisation of data generated by Institutional EO missions.

Astrophysics/Fundamental Physics, Planetary Sciences, Heliophysics and Space Weather are generally grouped together under the unique label of Space Science. The close connections of many scientific investigations and the fact that often the same space platform hosts payloads investigating in the various fields are the main drivers for the grouping. The grouping is found both in Europe and US, and is adopted also in the study.

Microgravity and Human Space exploration are grouped together as they both imply the presence of astronauts. Besides the space mission, microgravity missions include also sounding rockets, parabolic flight and drop towers experiments.

1.2 Study logic

To achieve these objectives, the study has characterized European space scientific data exploitation from acquisition to open distribution through the data exploitation chain, highlighting how data is made available for scientific exploitation for each of the following five space research domains:

- Earth Observation;
- Astrophysics/Fundamental Physics;
- Planetary Sciences;

- Heliophysics (and Space Weather); and
- Additional domains, such as Space Exploration (human).

In addition the study has identified best practices and best results obtained in past European efforts where increased scientific exploitation of space data has been targeted. This has been done especially in terms of:

- Effects that mission governance and funding structures can have upon the data exploitation chain
- Improvements in the specific mechanisms and approaches in the data preservation
- Activities to increase the awareness of the scientific community

The results of these analyses plus relevant cases from the global space sector have been considered to develop recommendations for planning and aligning specific stimulus actions with current policy for Horizon 2020 Space in order to increase future scientific exploitation of European space data.

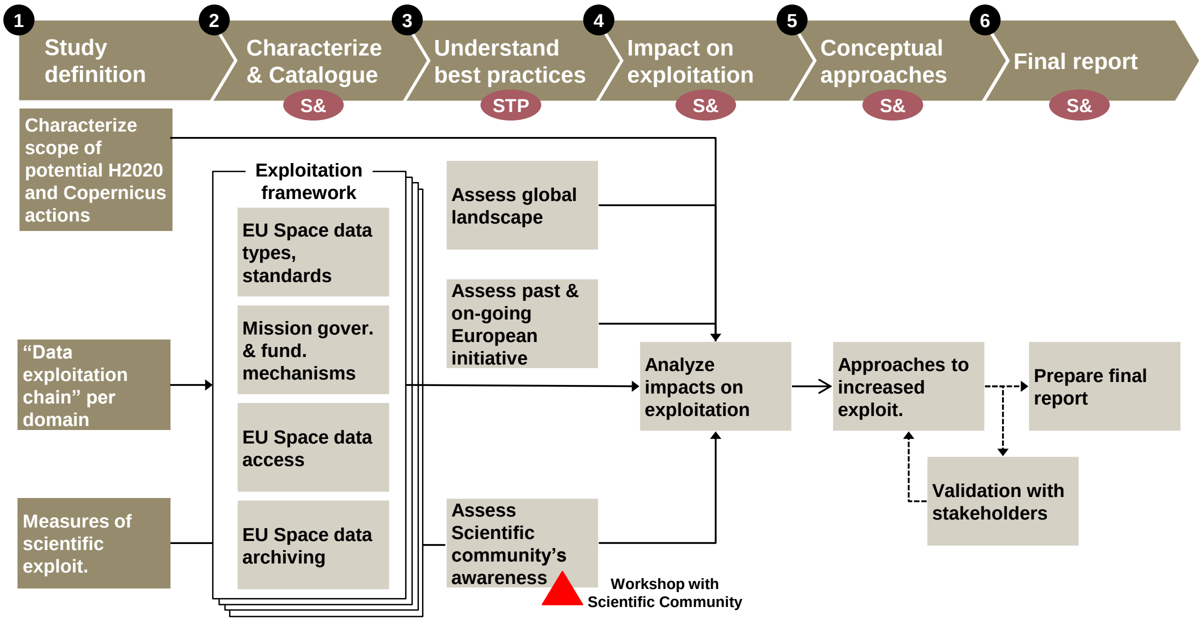


Figure 2: study logic

The study has been carried out using the six broadly defined work-streams, depicted above. The first three work-streams have been designed to be generally independent, with a few minor exceptions. They represent extensive research efforts to characterize the entire problem-solution landscape and synthesize the analyses in a structured set of hypotheses.

The hypotheses have been then validated with the support of relevant stakeholders and consolidated into a synthesizing analysis in work-stream 4. Activities within this work-stream are on-going at the time of development of this progress report.

Work-stream 5 is where potential actions have been identified, and work-stream 6 is used to consolidate the entire report, including synthesized results and recommendations.

1.3 Data sourcing and stakeholder consultation

As explained above, the definition of the study hypotheses has been mainly driven by desktop research, where a large collection of articles and documents on the topic of scientific data exploitation have been reviewed, characterised and analysed.

Following the desktop research, a crucial step supporting the analysis process for the hypothesis testing and refinement required extensive stakeholder consultation. Several different categories of stakeholders were engaged to ensure a sufficiently broad set of perspectives would be brought to bear. Three main methodologies have been implemented for the engagement and data collection from these stakeholders:

- Face-to face interviews, based on pre-defined and pre-circulated interview guidelines tailored to the stakeholder category
- Questionnaires targeted via email and tailored to the stakeholder category
- Definition and publication of a web-survey tailored to the stakeholder category

Direct interviews primarily focused on relevant European Institutional players involved with the implementation of scientific space programmes, conducted either via teleconference or face-to-face meetings. Stakeholders belonging to the following categories were addressed:

- European Space Agency: relevant POC for scientific data preservation and exploitation for Earth observation, Space Science and Microgravity
- National Space Agencies
- Lead of the FP7 Ulysses and Circe projects

Following each meeting or teleconference and interview summary was circulated to the stakeholder to confirm the understanding of the points discussed and validate stakeholder views of the problem.

A tailored questionnaire was circulated via email to the leads of other FP7 and Horizon 2020 projects covering different aspects of scientific data preservation and exploitation. The Research Executive Agency (REA) supported dissemination of the web-survey to ensure involvement of the project leaders.

The web-survey was addressed to broader the scientific community that makes use of data generated by space missions. Different channels were used to disseminate the web-survey. Importantly, the European Science Foundation (ESF) disseminated the survey to its large group of associated scientists. In addition, following interviews with relevant national stakeholders, the web-survey was further distributed to relevant and accessible scientific communities in Italy, France and the United Kingdom.

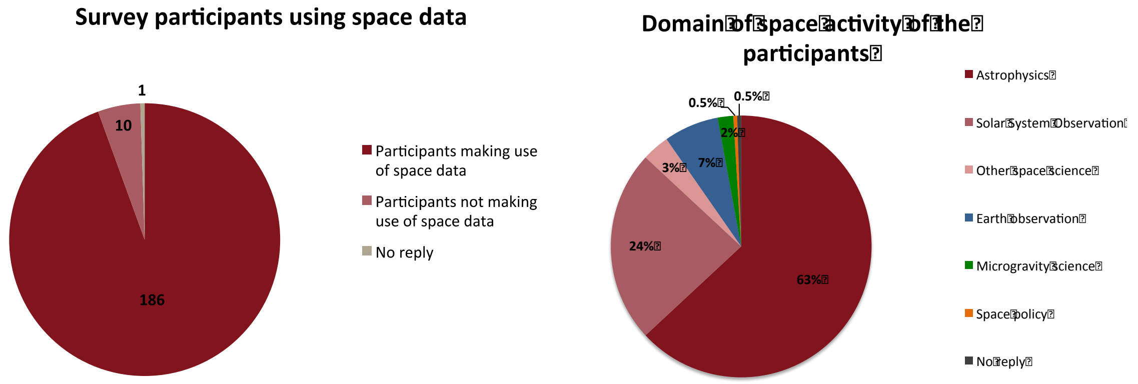


Figure 3: Web-survey participation

Participation in the web-survey has been quite important to the study. However a large majority of the participants are mainly active in astrophysics and this unbalances the global results towards this domain. Overall 90% of the participants are active in some form of space science, 7% in Earth observation and 2% in microgravity.

For more than 60% of the participants, the majority of the research activities they carry out are based on space data.

1.4 Background and space science mission context

The objective of this chapter is to describe the main European actors involved in scientific exploitation of space and introduce the principal models of governance observed in the past space missions.

Scientific missions in space are rarely implemented by a single funding entity; they are usually the result of a collaborative effort between different organisations. There is no single model of governance for space missions and, consequently, there is no single model to define the responsibilities for the scientific exploitation of the mission data. In order to understand the sharing of responsibilities, it is important to identify the primary actors and the mission models.

The way in which the data are made available to the scientific community may vary by mission. It is therefore important to understand the main models of data access observed in the different missions. This includes both models applicable to data accessed in real-time during the missions and to data accessed from the archives of data repositories.

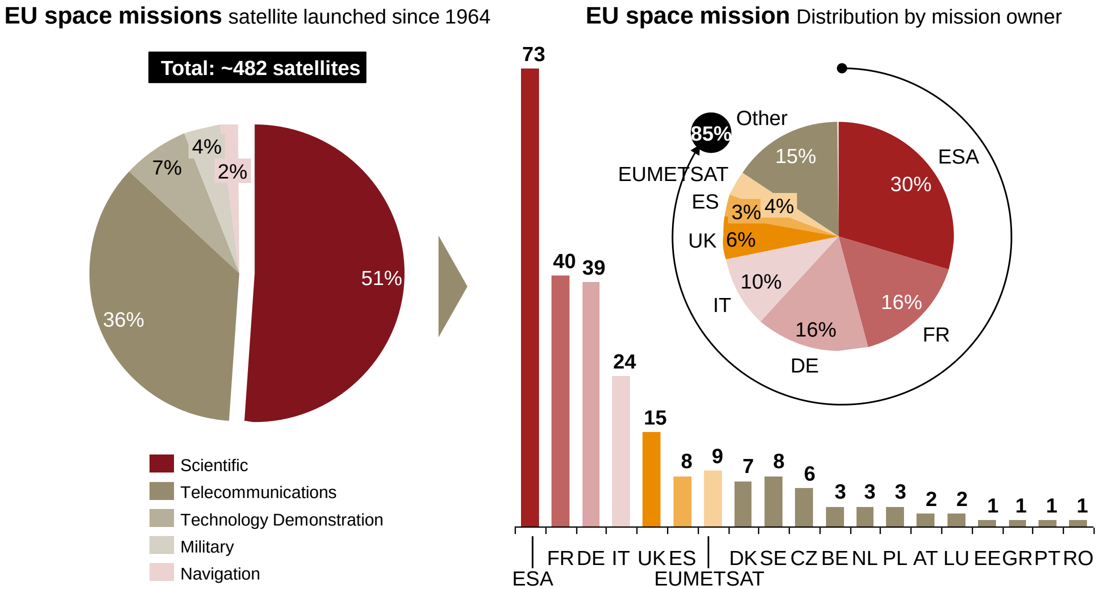
As we will see in the upcoming chapters, the understanding of these models is the basis for the understanding of other concepts related to the exploitation of the scientific data such as the data exploitation chain, the data infrastructures for the access to real-time and archived data, and the data exploitation roadmaps.

The present chapter will provide a brief historical overview of the space scientific missions in Europe and then explore the main mission and data access model.

1.5 History of European space missions

The fostering of scientific excellence has been the main driver for implementation of scientific space missions by European Countries and the European Space Agency since the beginning of the space era.

As summarized in Figure 4 below, starting with Italy’s Marco Polo satellite mission in 1964, European countries and ESA have executed many scientific space missions traversing many different scientific fields.


EU space missions satellite launched since 1964

EU space mission Distribution by mission owner

Source: UN satellite database, Strategy&.

Figure 4: European satellite mission launched since 1964[1]

European institutional and commercial satellite operators have launched a total of 482 satellites, of which roughly ~50% of the total were primarily scientific missions.

As of now, the European Space Agency has launched around 73 planetary and Earth orbiting missions, including the European Union Copernicus (Sentinel-1 and Sentinel-2 satellite). Together with scientific missions spearheaded by EU Member States France, Germany Italy, UK, Spain and the intergovernmental EUMETSAT agency, these represent 85% of the total number of European scientific satellite missions launched since the beginning of the European space era. [1]

The Commission, with its GMES/Copernicus programme, can now be considered a major institutional producer of space data. The exploitation of the data produced by all these missions is one of the priorities for EU Space Policy consistently implemented through FP7 and the current Horizon 2020 work programmes such as FP7's Cooperative ISS Research data Conservation and Exploitation (CIRCE) project.

ESA has been conducting scientific missions on behalf of its European Member States that have produced petabytes of scientific data from space. Many of those projects produced and continue to produce large amounts of data from their space-based assets. The investment made in space science through ESA over the years is substantial.

In 2014 alone, ESA has allocated roughly **EUR 1.9 Bln**, **47%** of its total **EUR 4 Bln** 2014 budget, to programmes aiming at the scientific exploitation of space (Science, Earth Observation, Human Spaceflight and Robotic Exploration)[2], yet this is only a portion of the total public funds invested in European scientific endeavours.

Analysis of space sector spending by Member States indicates that their investments in ESA represent about 56% of the total European investment in the space sector, and as a result, a number of other national missions are also contributing substantial volumes of space data to the scientific community.[3]

Well known European national missions include: CHAMP (Germany), GRACE (Germany), AGILE (Italy), BeppoSAX (Italy), Picard (France), Odin (Sweden), just to name a few. In addition, Members States also developed payload and scientific equipment for NASA's Mars Exploration missions (e.g. the Spanish Rover Environmental Monitoring Station and the French CHEMistry CAMeraon the Curiosity Rover). The scientific community has been successful at producing discoveries from these data as well (e.g. regarding numerous exoplanets, galaxy clusters, Earth's magnetic and gravitation fields, black holes, supernovae, gamma ray bursts, etc.).

When the EU's contributions with ESA for GMES/Copernicus are included, another approximately EUR 1.4 Bln[4] have been invested in space over the last seven years that will produce extraordinarily high volumes of new Earth Observation data. As summarized in **Figure 5** below, the number of scientific missions launched every year has increased by a factor of 10 over the last 40 years. At the same time, the production of data grew by a factor of 109, going from the Kb[5] produced by the first San Marco satellite (1964) to the 2.5TB per day produced by the Copernicus/Sentinal-1 satellites.[6]

It is estimated that the Sentinel-1 X-band SAR instrument will produce about 16 petabytes of scientific data during its operational life.[6]

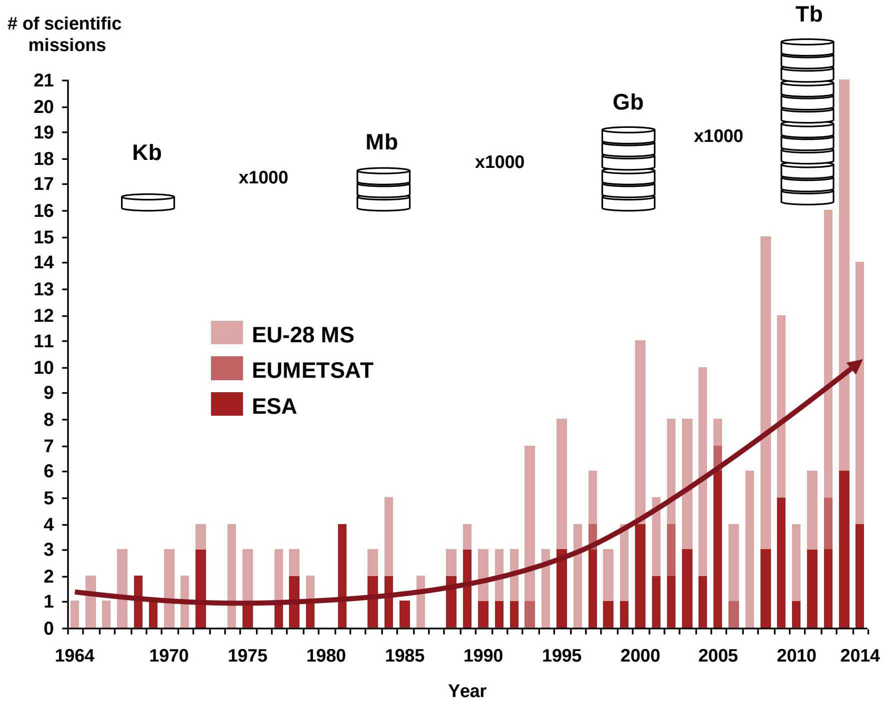


Figure 5: increase in number of yearly mission and in the production of data

The number of scientific space missions illustrated in **Figure 5** is based on satellite missions registered in the UN satellite database, which clearly excludes microgravity experiments.

There are different platforms available to perform experiments in microgravity, these are not necessarily space related. The International Space Station (ISS) is the platform providing long-term microgravity conditions, around three months. Every three months, indeed, the ISS is re-boosted to the nominal altitude (about 400 Km), during this manoeuvre microgravity conditions are perturbed, so not all experiment can be performed.

Beside the ISS there are other platforms providing shorter periods of microgravity conditions: sounding rockets (about 15 minutes), parabolic flights (about 5 minutes), and drop towers (few seconds).

Platforms other than the ISS can be accessed via different routes and organisations, while for the ISS ESA is the main channel available to European Scientists.

In ESA, microgravity experiments are supported via the European Programme for Life and Physical Sciences (ELIPS), which began in 2001. ELIPS provides support to scientific advances through experiments mainly carried out inside the ISS Columbus laboratory, but also through ground-based facilities in Europe, baseline data collection, bed rest studies, drop towers and parabolic flights. The programme engages some 1500 scientists involved in hundreds of experiments, as well as a large and diverse group of industrial research and development users.

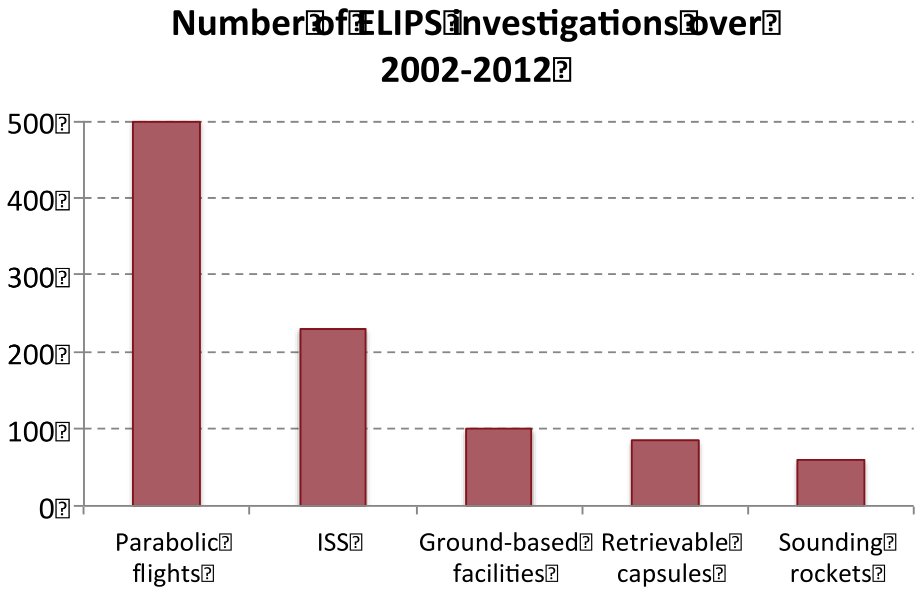
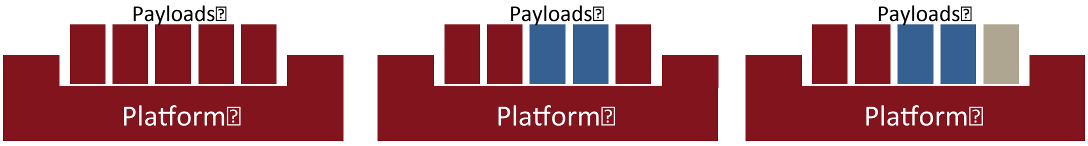


Figure 6: ELIPS microgravity science investigations over 2002-2012

The number of scientific investigations carried out per year under the ELIPS programme is averages 45 on the ISS, 35 in ground-based facilities, 35 in parabolic flights and 5 aboard sounding rockets. It is clearly a massive programme touching on a great variety of scientific domains, and aiming at improving our life on Earth while enabling humankind's long-term presence in space.

1.6 Mission governance and responsibilities

As introduced in the previous sections, the European Union, the European Space agency and Member States are the main entities implementing scientific space missions in Europe. Three principal mission models are used in the implementation of these missions: national missions, cooperation with other countries (intra and/or extra EU) and cooperation with international Institutions (intra and/or extra EU).



National mission	**International cooperation with other countries**	**Cooperation with international institutions**
One country implements the entire space segment (platform and payloads)	Different countries share the development of the platform and payloads	International institutions generally provide the platform and contributing countries provide payloads
Examples: AGILE mission (ASI) CHAMP mission (DLR)	**Examples:** IPDA and LIDAR (DLR) onboard MERLIN (CNES) Pamela (ASI/INFN) onboard the Russian Resurs-DK1 satellite	**Examples:** Tropomi (Dutch National Space Office) onboard Sentinel-5 precursor (ESA) SEIS (CNES) onboard InSight (NASA)

Figure 7: three main mission models

Models differentiate by the type of parties involved in the cooperation, i.e. National Space Agencies (e.g. CNES, DLR, ASI), International Institutions (e.g. EU) or Organisations (e.g. ESA, EUMETSAT, NOAA) implementing space activities; not by their geographical location (i.e. intra or extra-EU).

1.6.1 National Missions

National missions are implemented primarily by a single country with domestic funds budgeted for space. With this model, the responsibility for all aspects of the mission are of the implementing country and are often delegated to the national space agency or national space offices- where present - or to other institutional entities with mandate for R&D activities in space (e.g. national research centre, development agencies). Responsibilities include, among other things, the design and operations of the ground segment to operate the satellites and its payload(s), the definition of the data policy for the distribution of the mission generated data, and the preservation and long-term access to the data. In some case, e.g. France and Italy, the national space agencies also issue grants for the exploitation of the data. Other countries may contribute to specific aspects or infrastructure of the missions (e.g. provision of satellite components, provision of ground stations for data reception, scientific expertise in support to the mission design, support to satellite design and testing), but such participation is too minor to be considered a full collaboration between nations. A few historical examples of this implementation model are:

- The CHAllenging Minisatellite Payload (CHAMP), a German small satellite mission for geoscientific and atmospheric research and applications[10]
- The Astrorivelatore Gamma a Immagini Leggero (AGILE), an Italian mission dedicated to the observation of the gamma-ray Universe[11]
- The PICARD satellite, a French mission for solar observation [12].
- COnvection, Rotation & planetary Transits (CoRoT) space telescope is a mission of astronomy led by CNES in association with French laboratories (CNRS) and with several international partners (Austria, Belgium Germany, Brazil and Spain, plus the European Space Agency). Austria, Belgium, Germany and ESA contributed to different components of the spacecraft. With the exception of ESA, all other international partners provided for ground stations, ground observation capabilities and data analysis [13]

In addition to the missions described above, other national missions are implemented in the Earth Observation domain under the form of Public Private Partnerships (PPP). These are, in the majority of the cases, dual-use civil/military systems with the civil mission dedicated to institutional or commercial use. A few examples of this implementation are:

- The Pleiades satellites, a French high resolution optical system
- The Cosmo-SkyMED, an Italian high resolution Synthetic Aperture Radar (SAR) system
- The TerraSAR, a German high resolution SAR system
- The Ingenio, a Spanish High resolution optical system.

For these missions exclusive data distribution agreements are usually established with commercial companies for the distribution of the data and their preservation. As described in the following section, national space agencies in France, Germany and Italy, also foster scientific exploitation of data.

1.6.2 International cooperation with other countries

International cooperation with other countries is a mission implementation model where different countries share the responsibility for the development of the space and ground segment of the mission. Many variations of the international cooperation model are possible. In most cases one country provides for the space platform (e.g. satellite platform) and perhaps some of the payloads, while the other countries provide the remaining payloads. A few examples of this kind of mission implementation are:

- Odin, funded jointly by the space agencies of Swedish National Space Board (SNSB), National Technology Agency of Finland (TEKES), CNES, Canadian Space Agency (CSA) and the Canadian Natural Sciences and Engineering Research Council (NSERC). SNSB provided for the satellite platform design and operations, while the two payloads, the Sub-millimetre wave Radiometer (SRM) and Optical Spectrograph and Infrared Imaging System (OSIRIS), were international cooperation. SRM was designed and manufactured as result of a SNSB, TEKES and CNES cooperation, while OSIRIS was a CSA and NERSC joint activity.[14]

- The Payload for Antimatter Exploration and Light-Nuclei Astrophysics (PAMELA), and payload developed by the Italian National Institute of Nuclear Physics (INFN), with the support of ASI, hosted on the Russian Resurs-DK1 satellite. The PAMELA mission is implemented within the framework of the Russian-Italian Mission (RIM) space research program.[15]
- Meghatropique mission, a collaboration between CNES and the Indian Space Research Organisation (ISRO), studies the water cycle and the energy exchanges in the tropical atmosphere which contains a large part of water vapour of the planet. CNES provides two payloads (SAPHIR and SCARAB) and cooperated with India on the development of a third payload (MADRAS), all hosted on the ISRO satellite platform.[16]
- The Chinese-French Oceanic SATellite (CFOSAT) mission is a Chinese and France collaboration devoted to the monitoring of the ocean surface winds and waves, and related ocean and atmospheric science and applications. CNES is developing one of the two payloads, the Surface Waves Investigation and Monitoring (SWIM) while the Scatterometer (SCAT) payload and the satellite platform are supplied by the Chinese National Space Administration (CNSA).[17]
- The MEthane Remote sensing LIdar missioN (MERLIN) mission is a joint effort between France and Germany to investigate climate changes and monitor greenhouse gases. CNES provides the satellite platform (Myriad Evolution) while the two payloads, the Integrated Path Differential Absorption (IPDA) and the Light Detecting And Ranging (LIDAR), are the responsibility of DLR. [18]

The design of the ground segment for these missions is strongly influenced by the cooperation agreements between the different parties. In general, the country providing the satellite platform is also responsible for the satellite operations, the related ground segment and for the distribution of the scientific data to the mission partners.

Scientific data processing is often distributed through the various partners according to the ownership of the payload. Each partner processes and is responsible for the preservation of the data generated by its own payload.

The approach to the scientific data processing, distribution and preservation varies by country. A more detailed description of the approach adopted by the main space players in Europe is reported in **section 3.2**.

1.6.3 Cooperation with international institutions

The cooperation with international institution implementation model features cooperation with European Institutions (i.e. European Commission, ESA and EUMETSAT) as well as with non-European institutions (e.g. NASA).

This model also features different implementation possibilities. In the majority of the cases, the collaborating Institution(s) provides the satellite platform, the platform and payload integration services and the core ground segment, while the contributing countries provide payloads and/or instruments, dedicated ground segment components and scientific expertise to analyse the data. A few examples of this model are:

- The EU Sentinel 5 precursor mission is one of the Copernicus sentinels and will perform atmospheric measurements, with high spatial-temporal resolution, relating to air quality, climate forcing, ozone, and UV radiation. The TROPOspheric Monitoring Instrument (TROPOMI) is the main scientific instrument on board the Sentinel 5 precursor satellite. It is a Dutch National Space Office contribution in kind to the Copernicus programme and is developed in cooperation with ESA. [19]
- The Solar Orbiter mission is part of the ESA’s Cosmic Vision programme. The mission is devoted to the observation of the sun to examine how the Sun creates and controls the heliosphere. ESA provides for the satellite platform and for the integration of scientific

instruments. Solar Orbiter hosts 10 scientific instruments, each of which is developed under international cooperation between different countries and/or international institutions: [20]

- Energetic Particle Detector (EPD) is developed as collaboration between Spain, Germany, US and ESA.
- Magnetometer (MAG) is a United Kingdom contribution.
- Radio and Plasma Waves (RPW) is a collaboration between France, Sweden, Czech Republic and Austria
- Solar Wind Plasma Analyser (SWA) is developed as collaboration between United Kingdom, Italy, France and US
- Extreme Ultraviolet Imager (EUI) is a Belgian, UK, French, German and Swiss collaboration
- Coronagraph METIS, is a collaboration between Italy, Germany and Czech Republic
- The Polarimetric and Helioseismic Imager (PHI) is developed as collaboration between Germany, Spain and France
- The Heliospheric Imager (SoloHI) is an US contribution
- The Spectral Imaging pf the Coronal Environemtn (SPICE) is a UK, Germany, France, Switzerland and US collaboration
- The X-ray Spectrometer/Telescope (STIX) is a Swiss, Polish, German, Czech and French cooperation.
- The Interior Exploration using Seismic Investigations, Geodesy and Heat Transport (InSight) is a NASA Discovery Program mission that will place a single geophysical lander on Mars to study its deep interior.

The science payload is comprised of two instruments: the Seismic Experiment for Interior Structure (SEIS), provided by the CNES, with the participation of the Institut de Physique du Globe de Paris (IPGP), the Swiss Federal Institute of Technology (ETH), the Max Planck Institute for Solar System Research (MPS), Imperial College and the Jet Propulsion Laboratory (JPL); and the Heat Flow and Physical Properties Package (HP3), provided by the DLR.

Similar to what is described in the previous section, the design of the ground segment is strongly influenced by the nature of the cooperation. In general, missions are led by the international institutions which provide the satellite or mission platform. These institutions also provide the mission operations, the related ground segment and for the distribution of the scientific data to the mission partners.

Scientific data processing is often distributed through the various partners according to the ownership of the payload. Data can be either re-routed as raw data from the main mission operations centre, or directly acquired by the participating partners by means of dedicated ground stations.

The approach to the scientific data processing, distribution to scientific community and preservation changes from country to country, nevertheless, in this model the leading institution often takes the responsibility to archive and preserve the mission scientific data. The distribution is then implemented according to the lead institution data policy. A more detailed description of the approach adopted by the main space players in Europe is reported in **section 3.2**.

1.7 Data access models

The previous section has introduced the main models observed in the implementation of space scientific missions while highlighting the aspects related to the provision of the mission space and ground assets. Another important aspect to be considered is the way the scientific community accesses the data generated by the missions.

All space missions have an initial operational phase in which the scientific data produced by the on-board instruments are validated on the ground. This phase is necessary to ensure the observations are valid and can be reliably be used to execute good quality scientific research. The validation phase is executed with the support of the scientists, from now on called mission Principal Investigators (PI), involved in the design and development of the scientific instruments.

Once validated, the data are made available to the scientific community mainly in two ways:

Near-real time, the data are generated on board, received on the ground, processed and immediately made available via the mission ground segment. This data access model is only possible during the mission operational phase.

Long-term data access, the data are made available providing access to the mission archive database. This data access model is possible both during the mission operational phase and after the end of the mission, provided that the mission archive is made available and the data are preserved.

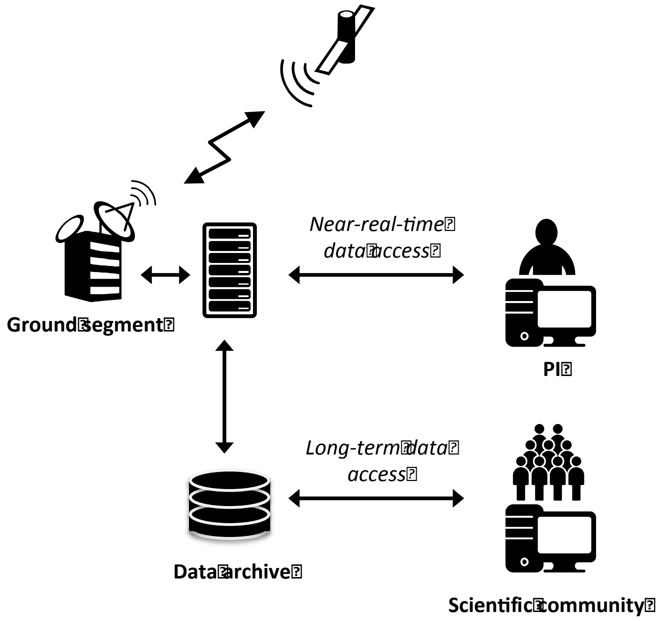


Figure 8: Near-real-time and long-term data access

Near-real-time access is implemented in almost all space missions. In some cases, the data can only be received on the ground at the end of the mission (e.g. sample return missions). In the scope of the present study, these cases are still grouped under the near-real-time access. This is justified as the first opportunity for a mission scientist to access the data and it is part of the planned mission achievements.

Conversely, long-term data access is not necessarily a mission requirement. As further discussed in the next chapter, access to archived data requires the implementation of a dedicated data infrastructure that is rarely planned within the mission activities. Mission budget, in most cases, covers the preservation of the generated data for a limited amount of time (about 5 years for ESA missions) after which other financial tools need to be utilized to ensure the long-term access to the data. [21]

1.7.1 Near-real-time data access

There are two main models to implement the near real-time access to the data the multi-PI model and the single PI model. The multi PI model is mostly implemented in the Earth Observation and Space Science missions, while the single PI model is mostly observed in the Microgravity missions.

1.7.1.1 Multi-PI data access

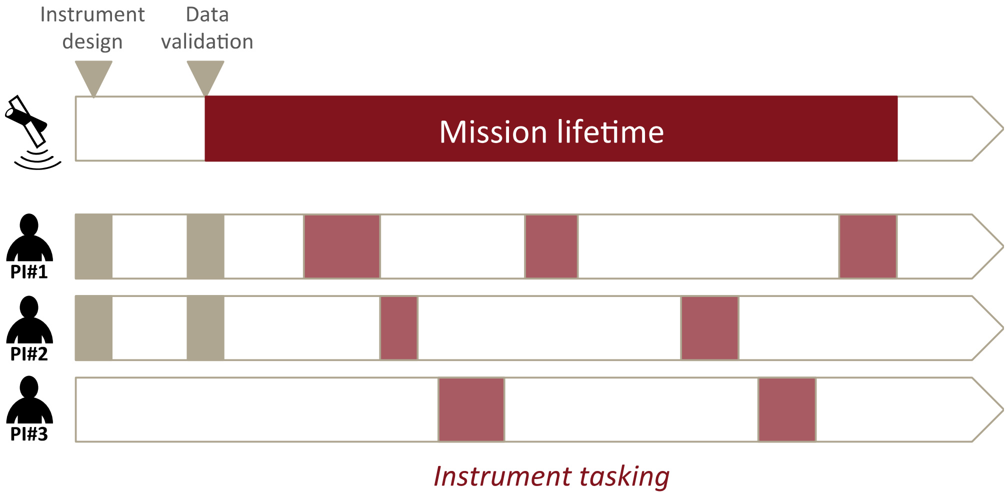


Figure 9: Multi-PI data access

The multi-PI model grants access to the PIs involved in the missions design phase and also to other PIs who did not support this phase. As seen in the previous section, the design of the scientific instruments for a scientific mission is typically the result of a cooperative effort among different countries (e.g. Solar Orbiter instruments). PIs from different national research centre (e.g. IPGP, the Max Planck Institute, the Italian National Institute for Nuclear Physics, *et cetera*) can be involved in these activities and in the consequent data validation activities at the beginning of the missions.

These PIs are usually granted access to slots of observation time during the mission. They may request specific observations to be made at specific times of the mission. The process of requesting specific observations at specific times is known as *tasking* the instrument or satellite.

The observation time reserved to the mission PIs is usually less than the entire observation time possible in the mission. Therefore the possibility exists for PIs not involved in the early phases of the mission to request satellite tasking and to access the available observation time.

Mission funding entities (e.g. ESA, National Space Agencies, and NASA) usually issue calls for observation time exploitation proposals. These are usually peer reviewed by a scientific committee to evaluate their feasibility and their scientific value. Once the proposal is accepted, near-real-time access to the data is granted to the requesting PI. The peer-review process ensures that requests for observation time are fair and proportional to the scientific objective of the proposals.

Not all mission funding agencies are mandated to directly fund of data exploitation (e.g. ESA, DLR) and therefore the PI usually covers the exploitation costs. When the funding agency covers exploitation (e.g. NASA, CNES, ASI) the PI is usually requested to contribute to the exploitation costs in proportion to the requested observation time for the space science missions or to the dimension of the observation areas for the Earth observation missions. This is another measure implemented to ensure the fairness of the requests. For example, NASA requires a contribution of 1 USD per second of observation [22] for its space science missions, while CNES charges 1 EUR per squared kilometre and per acquisition for its EO missions [24].

1.7.1.2 Single PI model

Figure 10: Single PI data access

The single PI model is typical of microgravity or sample return missions. In these cases the mission PI, or group of mission PIs, conceiving the experiment is the only one accessing the data in near-real time.

Other PIs access the data exclusively following the long-term data access scheme, provided that this scheme is implemented.

1.7.2 Long-term data access

As anticipated in the introduction, the long-term data access provides access to the data to the wider science community. It is not always possible to implement a long-term data access as it requires three conditions be satisfied:

After data processing during the mission, the scientific data are preserved and made accessible by means of an appropriate data infrastructure

The data policy applicable to the mission allows for the data to be made available to PIs that are distinct from the mission PIs.

There is a budget allocated to the implementation of the long-term data access.

Preserving the data does not merely mean to keep an archive of the data. Beside the scientific data it implies preserving the associated knowledge, i.e. all the data required to properly interpret the scientific data, and the possibility to re-process the data. Preserving the processing capability is part of the data valorisation as it allows for the same scientific data sets to support different kind of investigation.

The preservation of the data is implemented by means of dedicated data infrastructure which are not necessarily the same ground segment infrastructure providing for the near-real time data processing and distribution. There are different approaches to the development of data infrastructure and these are often influenced by the history of the data preservation activities in a specific domain or country, and by the actual volume of the data to be preserved.

ESA has implemented a distributed infrastructure to ensure the preservation of Earth Observation data generated by all its missions. Many redundant data repositories are distributed all over Europe and a unique data access point is provided to the users via the ESA Earth On-line Data Portal. [21] A different approach is implemented for the Space Science data. The ESAC centre in Madrid hosts the centralised data repositories for all Space Science mission implemented by ESA in collaboration with Member States and other European and extra-European partners. [22] As for the Microgravity missions, a network of User Support and Operations Centres (USOCS), distributed in different European countries, is in charge of the preservation of the data. Nevertheless, so far no mechanism has been established to grant access to non-mission PIs for the preserved data. [23]

A more detailed discussion on the ESA and major European Member States approaches to data preservation is reported in **chapter 4**.

Besides the availability of a data infrastructure, the possibility for the scientific community to access preserved data also needs to be agreed. In most cases the scientists accessing near-real time data benefit from an exclusive access to the data for a limited time (usually 1 year). This is normally done to give the Mission PIs a head start to perform their research and publish the data.

After this exclusive access period (also known as proprietary data access period), access to the data is open to other scientists. Duration of the exclusive access period and the possibility for scientist other than the (mission)-PI are documented in the data policy applicable to the missions.

As we will discuss in details in **section 4.6**, different data policies are applied to the data generated in the three domains of interest for this study. The exclusive data access model discussed above is still very fashionable, especially in the Space Science where a completely open access policy is applied following the end of the exclusive access period. [22]

Similar schemes are very frequent also for Earth Observation missions, but the advent of the Copernicus programme is dramatically changing the general approach in the domain. Free access to the scientific Earth Observation data is seen more and more as a "service" to the community, and therefore there is an increasing tendency to line up data policies to that of the Copernicus programme; granting in near-real time access to the widest possible audience of scientists and blurring the line with long-term data access.

This, of course, applies to scientific Earth Observation missions and excludes the commercial distribution of Earth Observation images generated by specialised, market focused Very High Resolution systems.

As for the Microgravity domain, general data policies allow for the open access to data, but in practice the absence of a proper data infrastructure blocks any potential access.

Last but not least, access to long-term data necessitates financial resources to sustain the access. Mission budgets usually cover the preservation of the data for a limited period of time (around 5 years), after which long-term data preservation is unfunded by the mission. Mission funding entities implement different mechanisms to ensure the preservation of the data beyond mission budget coverage.

ESA is trying to establish a budget line within its general budget for the preservation of the Earth Observation data beyond the coverage of the mission budget. The volume of data in the ESA EO archive is substantial – estimated in the order of magnitude of the of the Petabytes (PB) - and, as a consequences, so are the maintenance costs thus justifying the need for a dedicated budget line.

As for the Space Science, the volume of data from the entire ESAC data repository is quantifiable to be in the order of magnitude of the Terabytes (TB). The maintenance costs are not as big as the one discussed above for the EO data and they can still be managed with the general expense budget line of the ESA science budget.

Similar to Space Science, ESA Microgravity experiments so far generated a modest volume of data quantifiable in the order of magnitude of TB. The expenses for the data preservation are generally covered by the support contact for the USOCs.

1.8 Conclusions

The previous sections have provided an overview of the main European actors in the implementation of the scientific missions, the principal mission models used to implement the missions and the principal scientific data access models observed in various missions.

There are different players in Europe historically involved in the implementation of space scientific missions. Base on the numbers of space objects registered in the UN Space Object Database, the European Space Agency, together with EUMETSAT and the major European Contributors to the space activities (France, Germany, Italy, UK and Spain) have implemented the vast majority –

about 85% - of the total European Scientific satellite missions. Similar percentages are observed in other kind of missions such as the experiments in microgravity executed on the International Space Station.

Space scientific mission are rarely implemented by a single funding entity and are, more and more often, the result of collaboration between different players. These missions are implemented following three main models:

- National Missions
- International cooperation with other countries,
- Cooperation with international institutions (e.g. ESA, NASA)

Each model implies a different share of responsibilities between the different partners for the processing and distribution of the real-time mission data, the preservation of the data and the associated knowledge, and the provision of access to the archived data.

Similar to the mission model, the data access model differs from mission to mission. As discussed above different models are applied for the near-real-time access and for the long-term data access:

- The Multi-PI is mainly implemented for the real-time access to the data in the Space Science and Earth Observation missions.
- The Single-PI model is mainly implemented for microgravity and sample return missions.
- The Long-term data access is not always implemented as it requires a data infrastructure, an agreement on the data access and a budget to cover the costs not covered by the mission budget.

The mission and data access model play an important role in the definition of the elements which build up the data exploitation chain and the data infrastructure implemented to grant access to the real-time and archived data. These will be discussed in details in the next chapter.

1.9 Notes and references

[1] UN satellite database, Strategy&.

[2] http://www.esa.int/For_Media/Highlights/ESA_budget_2014

[3] Strategy&, 2015, Analysis of the European space institutional markets including future launch procurements and launchers

[4] European Union

[5] Stakeholder interview with ASI, Barbara Negri, Alessandro Coletta, Osvaldo Piperno, Cristiana Cirina, Roma, 25/06/2015

[6] https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/copernicus-sentinel-1; Strategy& analysis

[7] Includes CP, CP-FP and CP-FP-SICA

[8] Includes CSA, CSA-CA and CSA-SA

[9] Horizon 2020 Work Programme 2014-2015 – Leadership in enabling and industrial technologies (Space)

[10] http://op.gfz-potsdam.de/champ/

[11] http://agile.asdc.asi.it/overview.html

[12] http://smsc.cnes.fr/PICARD/

[13] http://smsc.cnes.fr/COROT/

[14] https://eoportal.org/web/eoportal/satellite-missions/o/odin

[15] http://pamela.roma2.infn.it/index.php?option=com_mjfrontpage&Itemid=159

[16] http://smsc.cnes.fr/MEGHAT/

[17] http://smsc.cnes.fr/CFOSAT/

[18] http://smsc.cnes.fr/MERLIN/

[19] https://sentinel.esa.int/web/sentinel/missions/sentinel-5p

[20] http://sci.esa.int/solar-orbiter/51168-summary/

[21] Stakeholder Interview: ESA EO, G. Kohlhammer, Paris, 22-5-2015.

[22] Stakeholder Interview: ESA SCIENCE, M. Kessler, F. Favata, Noordwijk, 06-06-2015

[23] Stakeholder Interview: Telespazio, G.Carotenuto, Napoli, 28-05-2015

[24] Stakeholder Interview: CNES, G. Gargir, T. Levoir, Steven Hosford, Toulouse, 28-05-2015

2 DATA EXPLOITATION CHAIN AND SPACE DATA INFRASTRUCTURES

The objective of this chapter is to introduce and explain the concepts of data exploitation chain and data infrastructures. A theoretical model of these concepts is introduced in the first part of the chapter, followed by a detailed description of the main data infrastructures implemented in Europe for the exploitation of space scientific data.

This chapter also illustrates how the previously introduced mission and data access models affect the definition of the role and responsibilities of each actor in the various links of the exploitation chain.

2.1 Data exploitation chain and infrastructure

There is a substantial set of processes that must function together in order for scientists to eventually have the opportunity to utilise data produced by European space missions. These processes can be conceptualised as a supply chain with six links. This concept will be referred to as the "Data Exploitation Chain" from this point forward.

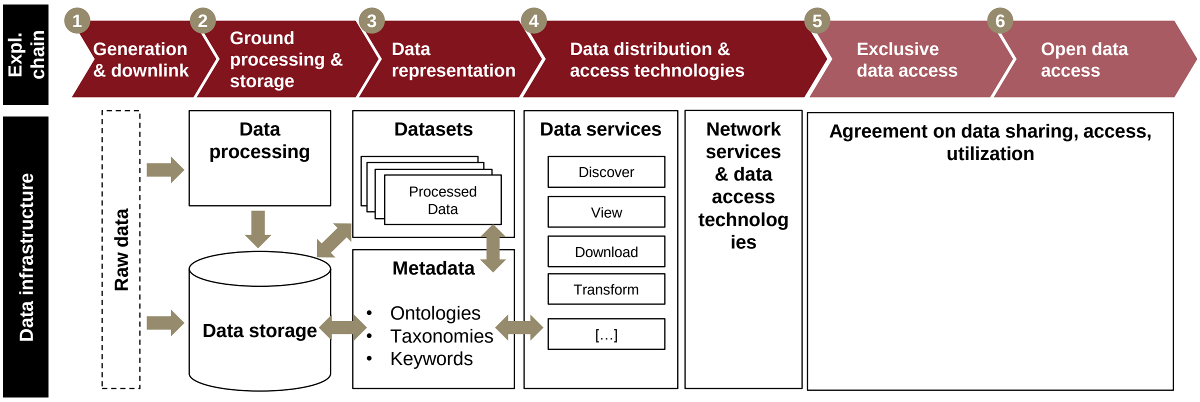


Figure 11: Data exploitation chain

As described in **Figure 11**, the functioning of the processes in the exploitation chain is supported by the data infrastructure, which is an ensemble of hardware, software, associated knowledge, services and legal agreement necessary to manage and distribute data.

Figure 11 presents a theoretical model of a data infrastructure. This model will be used in the rest of the study to describe the existing European data infrastructures providing access to science data both in near-real time and to long-term archives. As we will see later, the structure is completely modular and each module, encompassing one link of the exploitation chain, can be implemented as a single centralised instance or as a multiple instances distributed over geographically separate locations.

As they progress across the six links of the exploitation chain, the scientific data are transformed from raw data to scientific products (i.e. publications and other publishable material). The six links of the Data Exploitation Chain can be defined as:

- Generation and Downlink
- Ground Processing & Storage
- Data Representation
- Data Distribution & Access Technologies
- Exclusive Data Access
- Open Data Access

The first four links portray where the data in its various forms are processed, characterised, and permanently archived, and thus consume all of the funding allocated for the preservation and providing public access to it. The mechanisms of the processes of the first four links are largely automated.

In contrast, the processes of the final two links are controlled by policy and human decisions and interaction. Those processes include all the scientific analysis performed on the processed data by the scientists, thus they also produce analysed data and scientific products that are unlikely to have downstream consumers.

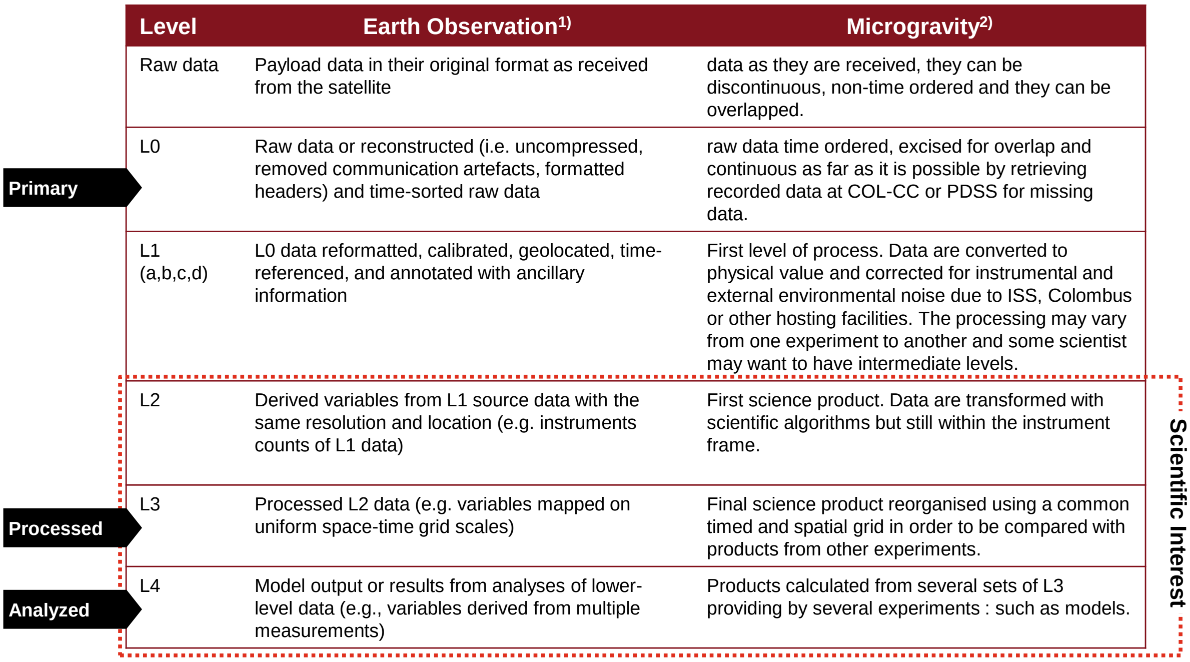
The transformation of the data from its raw values to analysed derivative products is organised in a stack of six processing levels, which are described in the next section.

2.1.1 Space Data Processing Levels

More specific to the domain of Earth Observation and Microgravity data, there is widespread concept known as Data Processing Levels*.* As summarized in **Figure 12**, there are six distinct levels defined, raw data through L4, that correlate with the Data Exploitation Chain links.

The characteristics of the data have a processing range from raw data at full instrument resolution (L0) to the output of models and simulations (L4). The terminology is internationally recognized and used by many space agencies, e.g. NASA, JAXA, ESA, CNES, etc.

Within the Space Science domain, there is no consolidated definition of data processing levels. For each mission, an ad-hoc definition tailored to the mission needs is documented in the dedicated Science data Management Plan. However, similarly to EO and microgravity domains, a rough classification of Space Science data in Primary, Processed and Analysed data is still applicable.



Level	Earth Observation$^{1)}$	Microgravity$^{2)}$
Raw data	Payload data in their original format as received from the satellite	data as they are received, they can be discontinuous, non-time ordered and they can be overlapped.
L0	Raw data or reconstructed (i.e. uncompressed, removed communication artefacts, formatted headers) and time-sorted raw data	raw data time ordered, excised for overlap and continuous as far as it is possible by retrieving recorded data at COL-CC or PDSS for missing data.
L1 (a,b,c,d)	L0 data reformatted, calibrated, geolocated, time-referenced, and annotated with ancillary information	First level of process. Data are converted to physical value and corrected for instrumental and external environmental noise due to ISS, Colombus or other hosting facilities. The processing may vary from one experiment to another and some scientist may want to have intermediate levels.
L2	Derived variables from L1 source data with the same resolution and location (e.g. instruments counts of L1 data)	First science product. Data are transformed with scientific algorithms but still within the instrument frame.
L3	Processed L2 data (e.g. variables mapped on uniform space-time grid scales)	Final science product reorganised using a common timed and spatial grid in order to be compared with products from other experiments.
L4	Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements)	Products calculated from several sets of L3 providing by several experiments : such as models.

Figure 12: data processing level definitions

The L0 to L2 data levels are grouped as "primary" data. These levels are contribute to instrument measurements and are not directly usable as a scientific product.

The data at the L3 level is considered "processed". These are aligned against uniform reference frames and thus are comparable to other data generated by dissimilar instruments. L3 data sets are considered scientific products that can be suitable for publishing.

The final product of the exploitation chain are the L4 data sets which are produced by PIs as result of analysis of lower level data.

As mentioned above the processing levels correspond to the different links of the exploitation chain. This relationship and a more detailed definition of the different links are described in the next sections.

2.1.2 Link 1: Generation and Downlink

For the overwhelming majority of scientific space data, the origins are individual sensors located within satellite payloads. These sensors typically scan continuously and store binary data in memory buffers. When the satellite that is hosting the payload is within the line of sight of one of its ground Earth stations, the satellite will dump the cached payload data during the downlink.

The data are sent in a format specific to the payload, but they are typically encapsulated in an International Standard Organisation (ISO) standardised format, such as the Consultative Committee for Space Data Systems (CCSDS) packets.[1] With the encapsulation packets, these data are classified as "raw data", a precursor to Level 0 of the Date Processing Levels scheme. Once the packets are stripped away, they are classified as Level 0 data, which occurs within Link 2.

ISO Standards

ISO (International Organisation for Standardisation) is a non-governmental organisation networking more than 162 national standards bodies around the world and developing International Standards.

ISO establishes technical committees and sub-committees responsible for the development and maintenance of the International standards covering aspects of technology and business in different domain of application.

Raw Data	Payload data in their original format as received from the satellite

Table 1: Data Processing Level of Link 1

2.1.3 Link 2: Ground Processing & Storage

After the raw data are fully acquired, they are processed to transform them to Level 0 standards. This data processing, which is performed by the ground segment, defines the start of the *Ground Processing and Storage* link.

The actual space data are not modified at this stage, but rather the transmission encapsulation is removed. The data were downlinked in individual packets, so Level 0 processing is discarding the packaging and reintegrating the data in the form in which they were stored on-board the satellite. For satellite-based data, this is entirely an automated process that occurs upon receipt of the raw data.

Level 0	Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artefacts removed. This includes, but is not limited to, duplicate data, synchronisation frames and communications headers.

Table 2: Data Processing Level of Link 2

---

[1] ISO 17355:2007 Space data and information transfer systems -- CCSDS file delivery protocol

The Level 0 classification of microgravity data is achieved by manually processing them. The responsible control centre reorders the raw data, eliminates duplicates, and fills in any missing data from archived sources as necessary.

The Level 0 data are archived by a database service after processing is complete, although the length of preservation varies. It is possible that the Level 0 data may be deleted after transforming them to a higher processing level instead rather than preserving both forms of them. The decision on how many data transformations to preserve varies per mission and per agency policy. Even if the initial decision to preserve all forms of the data is made, some forms may still be deleted once the mission ends and funds for preservation run out.

2.1.4 Link 3: Data Representation

The data representation is where the data are organized in sets of processed data meaningful for the scientific analysis and the content of the datasets is described and documented.

As presented in section **2.1.1**, the processing of the data follows different levels above the raw data level. Level 0 data are not really useful unless the Level 1 data has been lost and needs to be recreated, or perhaps an error was discovered in the processing definitions and the Level 1 data must be regenerated with updated calibration parameters. Once the Level 0 data has been transformed to its final Level 1 form, they may be further processed to produce normalised derivative data. The increasingly derived and combined data are classified at processing levels 2, 3 or 4.

Once the data has been transformed to its final automatically processed form, it is integrated into a dataset normally specified during the design of the mission. Several related sets of data could be time-synchronised or otherwise correlated and saved together as a set. During this automated process, metadata are identified, extracted and scribed into the format used to house the data. The following illustration indicates the processes and data that are produced during this section of the Data Exploitation chain.

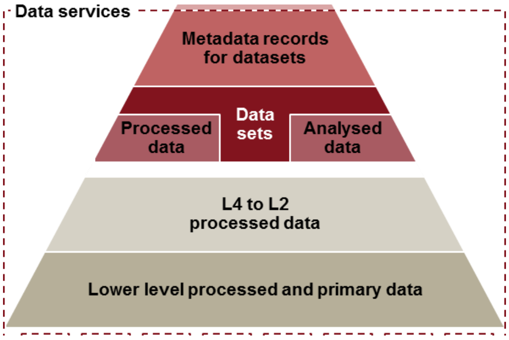


Figure 13: automated data processing

The description and documentation of the datasets includes the definition of the metadata and the provision of all the information necessary to correctly interpret the data.

A correct and complete representation of the data is paramount for the exploitation phase.

When correctly implemented it allows for data to be identified, understood and utilized.

Utilisation of common methodologies, taxonomies and ontologies allows for the correct interpretation of the data and for the interoperability of different data infrastructures.

Metadata are stored alongside the datasets and sometimes are bigger in volume than the scientific data. In the microgravity domain it is estimated that each byte of scientific data needs 1Kb of metadata to be correctly interpreted and utilized.

The general context in which the data are collected needs to be documented in the metadata so that correct scientific analysis can be implemented, for instance:

- Type of sensor acquiring the data
- Geographical location
- Ambient temperature on-board the ISS and level of µ-gravity
- General Astronaut health status and diet at time of blood sampling

All the knowledge associated with the scientific data has to be stored and alongside the raw data in the data archive. The method in which the data are actually archived may vary. Depending on the design, the database may store dataset components separately and another process would assemble the dataset upon demand. Alternatively, the dataset could be constructed once and just saved as a file to a directory, in which case the L1-L4 components may not be available separately. In any case, the metadata must be indexed separately in order to assist the search and discovery of the data. The selection of the metadata used is dictated by the data format selected for the mission.

2.1.5 Link 4: Data Distribution and Access Technologies

At this point in the chain, all of the automated data processing is complete. The pre-defined datasets have all been generated and archived, and they are ready for scientists to analyse them. Before that can happen, the scientists need to be aware of their existence and they need to be able to retrieve the data.

The data retrieval, also referred to as data distribution, is accomplished using standard internet technologies. These include using the File Transfer Protocol (FTP), the HyperText Transfer Protocol (HTTP), their secure versions (SFTP/HTTPS), Session Control Protocol (SCP), Secure Shell (SSH), CD-ROM discs, DVD-ROM discs, or even portable hard-discs or thumb drives. Most data centres in Europe distribute their data via web portals, where data can be discovered by basic searches of themes, missions or keywords; interesting data sets can be directly downloaded in electronic format via the web portal or ftp.

Similarly, the analysis of the data may lead to the generation of new derivative data that the scientists that created them wish to share with the community. In these cases, he or she will need to upload or otherwise deliver data back to a central repository. They will need to provide metadata in some form that the repository can index in order that others can discover their existence and determine their usefulness. The mechanisms for delivering the data are exactly the same as retrieving them, although delivering it surely requires authenticated credentials before the data are incorporated into a central database.

One frequent obstacle to the sharing of data is the protection by scientific teams of their research efforts and academic credentials. In practice, this leads to certain delay allowing for the preparation of the first publications before data are uploaded to open repositories. This is legitimate behaviour that also guaranties the continuation of the scientific interest and motivation to work with space data. Scientific teams working on Earth observation data have the additional aspect of the commercial potential, which also contributes to the endurance of scientific interest. It has also been observed that the high formatting and annotating requirements for submissions to the central repositories can discourage sharing due to the amount of effort required to prepare the data to those standards.

2.1.5.1 Specialised Processing and Distribution Technologies

Scientific need has driven the development of various technologies designed to provide users with the ability to discover, access, process and retrieve data within their fields of interest. Two notable examples are the emergence of Virtual Observatories and the use of cloud technologies to drive dedicate data processing platforms.

2.1.5.1.1 Virtual Observatories

The concept of the Virtual Observatory (VO) originated with the collective desire of astronomers to share acquired data with colleagues. It is a network of individual nodes that conform to the same interoperability standards which combine to form a gigantic, searchable database of astronomy data. The concept is beginning to be adopted by communities of other scientific fields as well. Each country typically has its own Virtual Observatory node.

In 2010 the TELIOS consortium[2] demonstrated that application-specific virtual observatories can be created from select datasets and processing with successes real-time wildfire monitoring and the sample generation of TerraSAR-X catalogues.

The creation of the Virtual Observatory concept was driven by the exponential increase of data being collected. Scientists want to correlate their data with related sources, but the vast quantities of data make discovering the existence of suitable data challenging, much less their retrieval. Users of the VO can perform global searches, specifying parameters such electromagnetic wavelength scanned and name/identifier of observed area, to discover candidate datasets that may augment their own research. This can be done using a standard browser over the internet, and the data is presented as files which can be transferred and saved locally. Another common method is to use catalogue services (e.g. VizieR, SIMBAD, cone searches of surveys) directly within a tool such as TOPCAT, which can manipulate and view data as well as discover, retrieve and manipulate it.

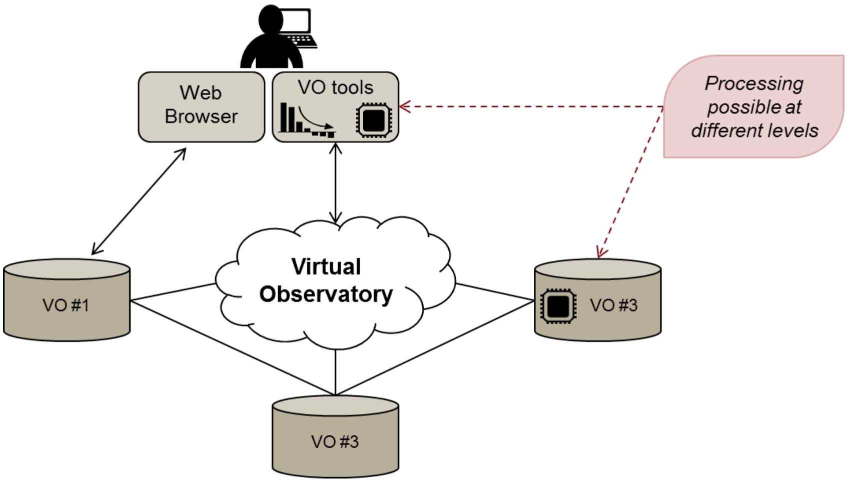


Figure 14: Virtual Observatory

The standards body defining interoperability and facilitating worldwide collaboration is the International Virtual Observatory Alliance (IVOA). The alliance comprises of 19 Virtual Observatories from every continent except for Africa. The executive committee meets up to four times per year, and individual working groups are designated for applications, data access layer, data modelling, grid and web services, resource registry, semantics, VO query language, theory interest, and VOTable format.

Europe has a unique role in the Virtual Observatory. In addition to several national nodes (UK, Spain, France, Germany, Hungary, Italy), it has a group known as EURO-VO, which is organisationally located between the IVOA and the national virtual observatories. The membership of EURO-VO includes ESA along with members of national VO initiatives, and it has been funded by the European Commission continuously since 2001 under the Framework Programmes (5, 6, and 7) [25]. This unique arrangement has proven to be efficient, allowing the expertise to be pooled and allowing national and European initiatives to be linked. The main responsibilities of EURO-VO are to support scientific users, support the data providers, and to coordinate technical activities. Thus far, EURO-VO has been successful in building a European-wide VO community and its

---

[2] Consortium led by the National and Kapodistrian University of Athens responsible for the Virtual Observatory Infrastructure for Earth Observation Data (TELEIOS) project funded by the EU under the FP7/ Intelligent Information Management research area.

infrastructure, which is used as an example of successful cooperation and collaboration of European nations within global scientific endeavours.

2.1.5.1.2 Processing Services

Prior to the time when personal computers were prevalent, powerful, and equipped with relatively large amounts of RAM and storage capacity, scientists relied on external data processing, which was often performed by the agency responsible for the scientific mission. As scientists became better equipped and more versed in programming and data manipulation, they began to prefer to have data delivered at lower processing levels in order to generate derivative data in-house. The responsibility for L3 and higher classifications of data shifted from the agencies to the scientists. As a side effect, the derivative data was less likely to be preserved and globally accessible as the responsibility for ensuring long-term preservation shifted to the scientists.

The data acquired and delivered by modern satellites has been growing exponentially, a trend that is expected to continue. The datasets are now measured in TB. The sheer amount of data that needs to be delivered has begun to present logistical problems. A large amount of bandwidth is needed to deliver the data (not available in many locations), and the scientists need significant IT infrastructure to save, organise, and process the data – tasks and knowledge that are increasingly out of scope of the scientist’s knowledge and responsibilities. As a result, the processing trend has begun to reverse with the appearance of hosted processing.

For nearly a decade, ESA has provided a “cloud” service to stimulate the generation of scientific added value products known as Grid Processing on Demand (G-POD). The motivations of G-POD include:

- stimulate the exploitation of global EO mission archives;
- promote the development of Earth science applications, particularly those with high data and processing requirements
- foster partnerships in Earth Science research, algorithm development and operational deployment of applications
- demonstrate and promote electronic collaboration within Earth Science
- encourage synergistic use of EO data
- enable the combination of EO data with other space-borne and ground data
- facilitate modelling and multidisciplinary applications
- enable the creation of products derivative of ESA mission data that would not otherwise exist.

There are several obvious benefits to a service like G-POD. The scientist is free to concentrate on the development of scientific modelling and algorithms rather than have concerns about system administration and purchasing / maintaining IT infrastructure (which is likely to be done poorly if they manage to succeed at all). The amount of processing power available in a service like G-POD is almost certainly far more than what an individual organisation could afford to construct (over 350 CPU, over 330 TB of local storage, access to 180 TB of EO data). Additionally, the sizes of the generated products are often far less than the sum of the data used to generate them, which significantly reduces the bandwidth requirements between the scientists and data providers.

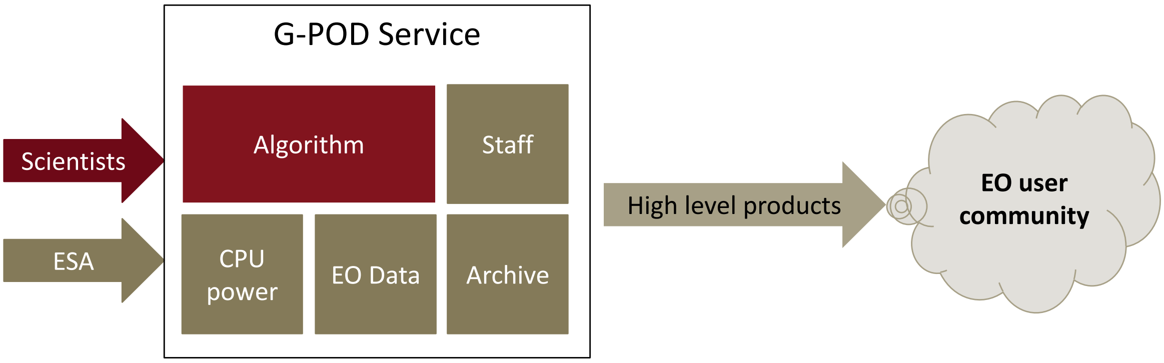


Figure 15: ESA G-POD processing service

A major benefit for the long-term preservation of data is that products generated by a service like G-POD can be automatically archived with generated metadata as a by-product of a job process. When publishing a paper involving the data, it can simply be referenced so other scientists could retrieve the dataset in a simple and reliable manner.

2.1.5.2 Scalable cloud computing

If a processing service like G-POD is available and the EO mission data is sufficient, it is a great option for the scientists that can leverage it. However, this is not always possible. Perhaps the required mission data is not an EO satellite, or is otherwise not available within G-POD. Perhaps the scientists need greater control over the algorithms or how the data is archived (or not).

If an existing processing service is insufficient, the scientist can turn to commercial cloud computing options such as Amazon's EC2, Rackspace Managed Cloud, or the Google Compute Engine. This is known as Infrastructure as a Service (IaaS). These are commercial services that charge based on the amount of work done – a properly designed system can quickly scale if required – but also offer a great deal of flexibility.

Like G-POD, the infrastructure is maintained by expert staff and the scientist does not have to be concerned with it at all. However, while the G-POD Research and Service Support (RSS) handles the implementation of the algorithms and data combinations, a scientist using commercial cloud computing is responsible for implementing all the calculation. True High Performance Computing involves parallel processing which often requires an expertise to implement correctly. If the software for the algorithms already exists and is ready for cloud scaling, it could be a relatively simple process. However, unique algorithms could be so challenging to implement, that scalable cloud computing is not an option unless consultants / contractors are brought to implement the requirements.

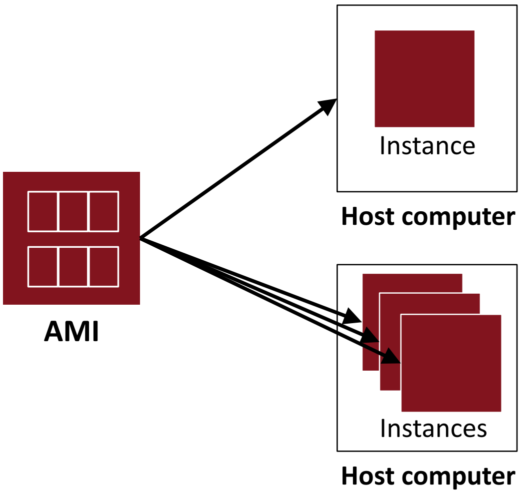


Figure 16: Amazon EC2 - Amazon Machine Image and instances

Using the Amazon EC2 service as an example, the Scientist would have to obtain an Amazon Machine Image (AMI) first. It is possible that a generic one has been donated to the community and configuring it for a specific set of tasks is relatively easy. However, should the study require advanced parallel computing concepts or unique calculations, the AMI could be a huge challenge to create and maintain. This obstacle can be overcome by the hiring of an expert, however, so it mainly remains a matter of expense.

2.1.5.3 Data Standards for Access and Distribution

There are several data standards historically and currently in use by EO and other space science missions. Today, the approach taken by NASA-sponsored missions is different from ESA's approach. NASA has been trending towards preferring the use of a flexible generic format, such as HDF5 or modern NetCDF, sometimes using versions tailored to the discipline (e.g. HDF-EOF5 for Earth Observation or NetCDF –CF for Climate & Forecast metadata). ESA has embraced the SAFE scheme (Standard Archive Format for Europe) which is an extensible XML-based format based on an ISO standard (OAIS[26]) that has been widely adopted for data archival needs outside of space.

Each approach has its advantages. The NASA standards are not approved by international standards bodies, although their contents may be [27]. They tend to be developed by a committee formed by researchers and industry members. The binary formats are effectively packed, and through the discipline-specific definitions, standard tools have been developed that work across many missions. These standards have evolved over decades and multiple missions, so new missions can select an existing standard that fully supports its needs and one that that has tools already developed allowing researchers to effectively analyse data as soon as the mission starts. For formats that have been effectively frozen, the lack of approval by an international standards body has little to no impact, but for formats in flux (such as PDS), this may be a shortcoming – although it wouldn't be possible to submit for standardisation if the specification isn't frozen anyway.

ESA tends to be drawn to internationally approved standards. Even the NeXus-HDF format, based on NASA's HDF5, is intended to be submitted to ISO. ESA is focusing on the use of the ISO-approved SAFE format. This is not limited to future missions or even European missions. ESA has begun archiving data from historical missions by first converting into tailored SAFE format. So far they have archived ENVISAT (EO, 10 separate products), ERS (EO, 1 product), JERS (EO, JP, and 1 product), Landsat (EO, 1 product), MOS (EO, JP, and 1 product), NOAA (EO, US, 2 products), SEASTAR (EO, US, 2 products), SPOT (EO, FR, and 1 product), TERRA/AQUA (EO, US, 1 product).

The SAFE formats are XML-based, which is to say text-based, so they will have to encode binary (probably with UUENCODE64). This provides the ability for standard archiving (media independent) at the expense of file size. Also, metadata would have to be extracted automatically before converting to safe so it can be embedded into the SAFE format for discovery purposes. Each mission has its own specification. Standard tools can generically read XML, and extract information, but other tools may be needed after extraction. As a result, ESA has been basically obligated to develop toolsets for researchers, such as NEST [28], which was superseded by the Sentinel-1 toolbox [29].

2.1.5.3.1 Data Services

The data services include the ability to discover the data, view it, download it, and transform it.

2.1.5.3.2 Discoverability

All SAFE formats have 4 components:

- Acquisition Period
- Platform/Instrument identification
- Product History
- Product Data/Metadata

If the platform and Acquisition Period desired are already known, then searching for the data is straight forward. Otherwise, a search of the metadata is required. The metadata contains the following information:

- Orbital information
- Grid Reference
- Geolocation information
- Quality/Fixity information
- Fixity Information
- Representation information

It appears that SAFE formats are strictly limited to Earth Observation products based on the available metadata. The most important data are geo-location and orbital information. Searches like “show me list of records including X geolocation acquired between Y and Z times” should work, and they could be narrowed by instrument type and orbital parameters. It is not known if searches could apply to the data themselves, but chances are they do not.

For other needs, such as heliophysics, astronomy, and microgravity, it appears that SAFE will not address those needs. Indeed, all SAFE products thus far have been limited to Earth Observation missions. That is not to say that the same approach could not be taken, but a new schema based on OAIS would have been developed that would support the characteristics of data for non-EO space disciplines.

Regardless, not all data are in SAFE format, not even all European data. There is no “one stop shop” where a user can put in search parameters and receive results from separate databases all over the global. Thus, for a user to discover data, he or she currently has to get access to every portal of these databases and search each one individually. Ideally a global portal could be developed that could search all databases simultaneously similar to how someone would search for available air fare, but such an service would have to a separate “driver” for each database it interfaces with, one that was familiar with the available data that could be searched.

2.1.5.3.3 View

The nature of the archives means that the data sets have to be downloaded and the information extracted in order to view it with either the extract tool (e.g. Sentinel-1 toolbox) or an external tool that can read and view the contents. In theory the search interface could show a preview of the data, but this would increase its complexity enormously.

In practical terms, the user needs to know what format the archive presents, and ensure that he/she has access to a reader that understands that format.

2.1.5.3.4 Download

The download of search results is independent of data standards. The search interface will allow a direct download (e.g. HTTP) or perhaps via FTP or SSH protocols, but it will come in the advertised format.

2.1.5.3.5 Transformation

To leverage existing tools, converters can be developed although they probably already exist. For example, a tool that only reads EOS-HDF5 files could be used if a SAFE-ENVISAT/EOS-HDF5 converter was used on the archive’s contents. However, data transformation is the responsibility of the user and not the concern of data exploitation.

2.1.5.4 Network Services and Data Access Technologies

There are no centralized archival facilities for the world. The HARM[30] (Historical Archives Rationalisation and Management) project of ESA, aimed mainly at converting its historical datasets into a new modern format, based on the latest technologies and standards and able to ensure the long-term preservation of its holdings, is consolidating data, which reduces the number of interfaces globally. However, there are always going to be multiple portals used to access an even greater number of databases.

Global seekers of space data use standard internet technologies both to discover and access the data. Clearly a web-based interface is popular, especially when the data doesn't require the user to be authenticated.

If user authentication is required, then the user has to be previously registered with the data archival system. Credentials can come in the form of username/password (http or ftp), SSH public keys, 2-step authentication (tokens or one-time pads), etc. In the vast majority of the cases, the data are going to be made available to the public, and registration requirements will be minimal or non-existent.

The biggest issues are:

How does a prospective user know where all the portals are? Is there a centralized directory? If so, how well is that known? Is it simply institutional knowledge?

How does the user know how to search each portal? Presumably the search interfaces are unique, partially by independent desire and partially dictated by the contents of the archived data (and the metadata)

How burdensome is access to each portal?

An obvious solution would be a single, well-publicized portal that could access essentially all important space data databases using a standard search interface, and also providing retrieval services. The user would be aware of all available data sets and select the ones most relevant for their studies.

2.1.6 Link 5 and 6: Exclusive Data Access / Open Data Access

The final links of the Data Exploitation chain are where the actual exploitation occurs. During the Exclusive Data Access period (Link 5), the distribution of data is intentionally limited to the PI team that was largely responsible for the design of the mission. During a relatively short period (normally ranging from 6 months to a year) the PI team will analyse the data and potentially present their findings through papers or some other scientific forum. When this exclusive period ends, the rest of the scientific community (or anyone in the world) will have the right to receive the data and use it for their own purposes. This starts the Open Data Access period, which is the final link in the Data Exploitation chain.

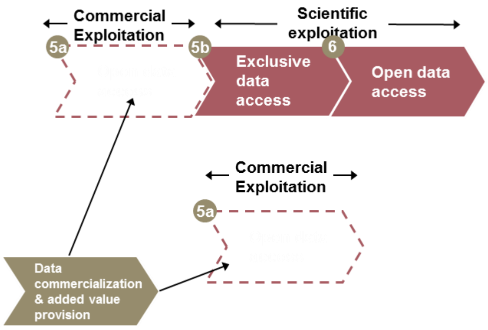


Figure 17: scientific and commercial data exploitation

It is certainly possible that new derivative data products are generated in both periods, but there is a question of intellectual ownership and there is no obligation to share these products. Any entity exploiting space mission data products in order to generate new data voluntarily shares them, and they can do so through whatever means they prefer. That said, there are repositories that specialise in products developed in these periods that the creators contribute back to the community, e.g. Planetary Data System (PDS).

Until this point, indeed, the analysis has centred on data that originated with publically funded space missions. When Earth Observation data are concerned, another source of data for the scientific community involves commercial missions that may have received public funding at some point. It is not uncommon for data from these commercial endeavours to be donated to the public after they no longer generate significant revenue. In these cases, the "Exclusive Data Access" period may be omitted as the original purpose of the data was commercial rather than spearheaded by a primary investigator. It is also possible for a commercial mission to be provided public funds under the provision that the data must be released after some condition is met (e.g. passage of time), and that more than one Exclusive Data Access periods are established so that access is tiered with wider pool of users at each tear.

2.1.6.1 Data exploitation indicators

The most common ways for space agencies to measure the exploitation of their scientific data by the scientific communities are to measure data downloads tracking the nature of data users and the tracking of the number of scientific publication generated on the basis of the data.

Data downloads are mainly reflecting the amount of data extracted from the archives but are not a good indicator to evaluate the actual usage of the data. This is even more accurate with the evolution of processing platforms where only a smaller but more useful amount of processed data is downloaded from the repositories.

Tracking the nature of the users who download data through their registration profile is of somewhat greater interest as it can allow identifying the various institutes, disciplines, or geographical repartition of users. However, it does not allow for a quantification of the level of data exploitation.

In addition the possibility to host processes to analyse the data and generate new products, is changing the way the exploitation of the data is measured reducing the need to track data downloads and increasing the need to track the number of scientific publications.

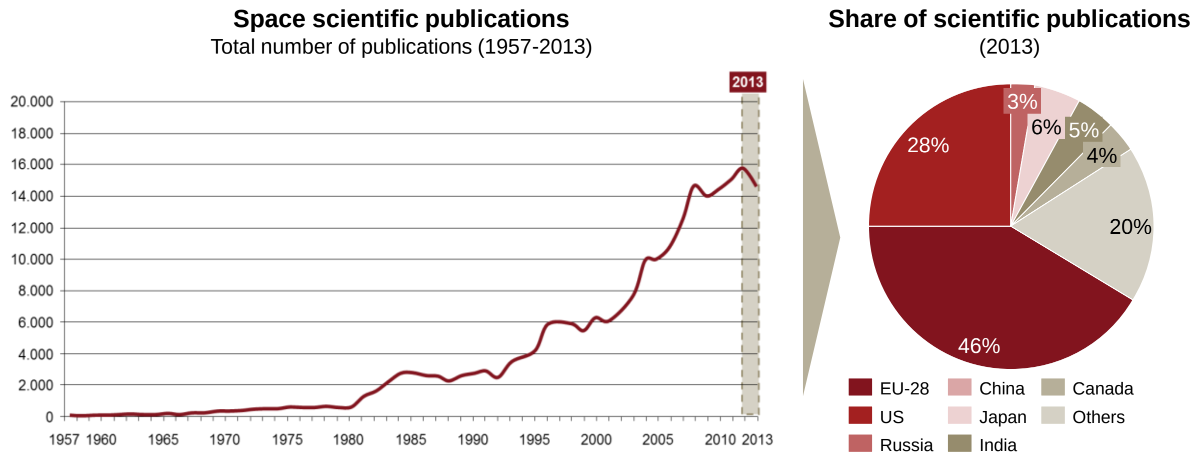


Figure 18: Space scientific publications [42]

As shown in **Figure 18**, the number of global scientific publication based on space data increased from few products in 1957 to an excess of 14.000 products in 2013. Since 2005 the global production of space scientific publication is constantly above 10.000 products per year.

The analysis and tracking of the publications is not an obvious activity. In addition to the growing number of publications, an additional difficulty is that articles are often the result international research team activities; this leads to the same products being published in different locations and being traceable back to different research institutions in different countries. As shown in the pie on the right of **Figure 18**, US research institution were mentioned as authors and co-authors in about

28% of the total number of scientific publications in 2013 while research institutions from EU-28 MS in about 46% of the publications.

Bibliometry is the science of monitoring these publications. It is an advanced science where a lot of research is being pursued.

For space science, NASA lists all astrophysics publications in the ATS (Astrophysics Data System). ESA and some national agencies have developed software that scan through this archive in order to generate statistical results. However, because of the absence of standards, discrepancies can be observed between the different software scanning the ATS since different algorithms are implemented. Although this is a painful process, it remains however the most relevant method to evaluate the exploitation of scientific data.

The DOI (Digital Object Identifiers) system is an international standard and a nomenclature system used to uniquely identify a digital object. The DOI system is implemented through a federation of registration agencies coordinated by the International DOI Foundation. This service comes with a cost. The benefit of the DOI system is to provide a persistent identification of the source of the data. It can therefore be used as a way to cite the originators of the datasets themselves in publications. This method allows giving credit to the scientists in charge of the measurements and not only to the authors of a paper. This has come as an issue in some cases when data becomes publically available and the PI teams have not yet been able to fully exploit their data and prepare their own publications. Using this system could therefore be considered as an option to allow an earlier release of space data to the wider scientific community while still providing a form of credentials to the PI team in charge of the observation.

2.2 European space data infrastructures

This section describes the different data infrastructures implemented in Europe for the distribution and access to data generated by space missions. As anticipated in the previous sections, there are different implementation models of data infrastructures spanning between the distributed and the centralised models depicted in **Figure 19**

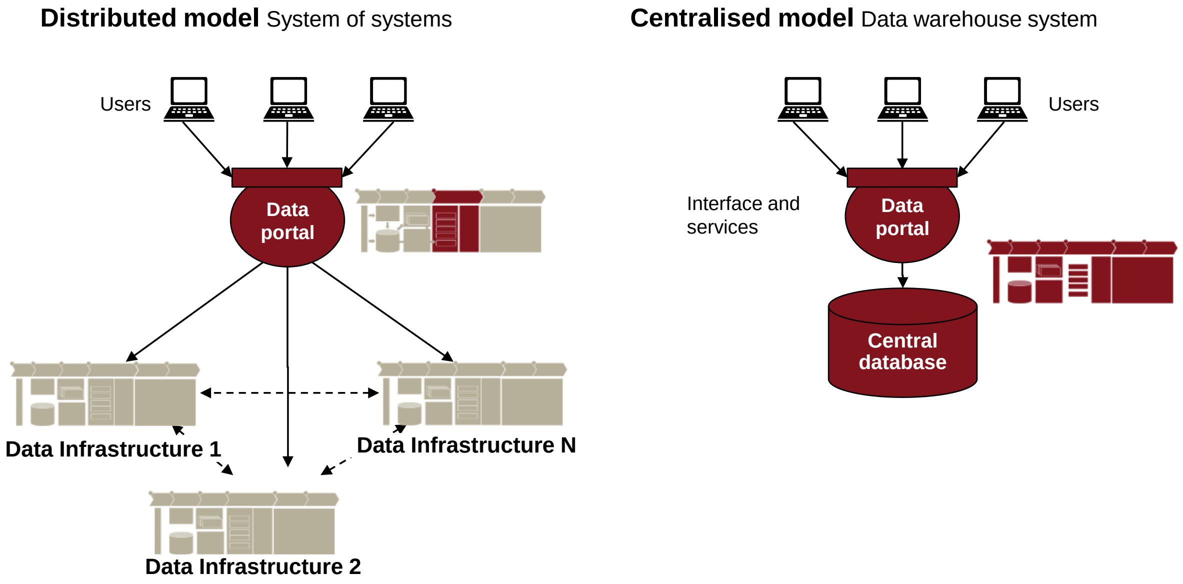


Figure 19: data infrastructure implementation models

Each model implements all the modules of the data infrastructure described in the previous section. The distributed model is a system of systems, where each module, or groups of modules, of the infrastructure has multiple instances physically distributed over different locations. The distributed instances might also be self-standing data infrastructures dedicated to data generated by a single space mission and providing local access to the PI (e.g. University or Research Centre servers hosting data and providing local access to them). Typically a centralised data portal provides access to all the data distributed within the instances of the infrastructures.

Conversely, the centralised model has a unique implementation of each module of the infrastructure with a single repository for all the data, and a single access point.

As discussed in the next sections, there are different data infrastructures in Europe providing access to the space scientific data. Both models are implemented and often hybrid models are adopted with the coexistence of a centralised and many distributed data repositories.

In some cases the same data infrastructures can be used to provide data in near-real time data and as long-term data access. In other cases, different infrastructures or groups/modules are implemented for the near-real time and long-term data access.

2.2.1 ESA-EO [31]

For ESA Earth observation missions, a distributed and shared approach has been implemented for more than 20 years. This network of repositories is distributed all over Europe and shares the infrastructure and operations with those of public-national missions and some commercial missions. The distributed network serves to access to data both in near-real time and to long-term archives.

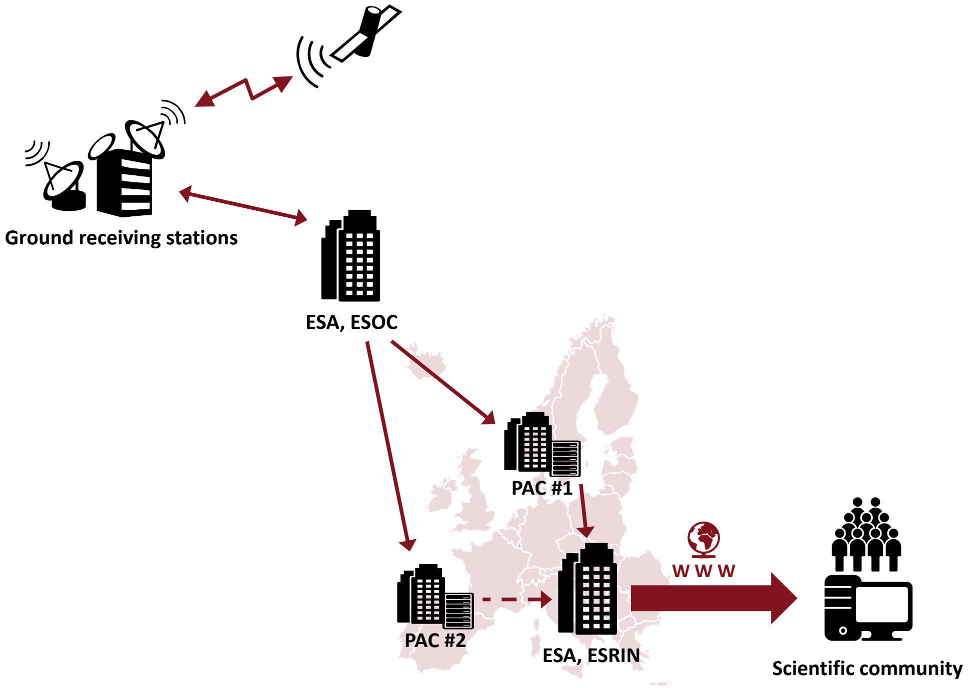


Figure 20: ESA EO ground segment

For each repository, a contract is signed with the responsible entity to implement data processing and archiving services through Processing and Archiving Centres (PAC). These scientific expert centres are selected for quality management, e.g. on the basis of their previous experience in handling the specific mission data (e.g. optical, SAR, etc.). The distributed infrastructure is transparent to the data users through single data portals, end-to-end flight operations at ESOC, data flow management infrastructure, and user management in ESRIN.

Fast data access

For specific missions, the ESA distributed ground segment includes ground-receiving stations with fast delivery capability (the EO data are made available in less than 3h from sensing). The processing of these data is much faster but "less-accurate" than the one performed at the PACs. The data are temporarily stored in a buffer archive and are overwritten when the buffer is full.

ESOC is the centre responsible for the satellite control and housekeeping. Data generated in space (system/platform plus payload/mission) are received on the ground by the ESA ground stations. ESOC receives all the data acquired by the ground stations and, after separating the system/platform data from the payload/mission data, distributes the latter to the processing centres.

For Earth observation, the raw data are sent to the PAC centres. The PACs are part of the Sentinel Payload Data Ground Segment (PDGS), and are responsible for the exploitation of the instrument data. They consolidate and reintegrate level-0 data and perform systematic and request-driven processing of data to higher-level products. Archiving and long-term preservation of data is ensured for all Level-0 data and for a set of configurable systematic higher level products. In order to ensure redundancy and reduce risks, all data are stored at least at two separate sites and in predefined storage formats in line with the long-term data preservation (LTDP) principles elaborated by the ESA coordinated Ground Segment Control Body (GSCB). For additional details on the GSCB activities and LTDP guidelines, refer to **Section 4.5.1.1** in the next chapter.

Large reprocessing campaigns of Earth observation data (dedicated reprocessing service contracts) are issued by ESA. Online data product archives always contain the data processed using the most recent versions of algorithms.

The same European infrastructure is also used to manage the Earth observation data from international partners of ESA in order to offer this data conveniently to European users.

2.2.2 ESA-Space Science [32]

Unlike for Earth observation where the same data infrastructure is shared for all data access, ESA adopted a hybrid model for the Space Science data.

Within ESA, Space Science is different from other domains (e.g. EO, microgravity), namely because Space Science is one of the mandatory programmes of ESA. This means that all ESA Member States contribute to its budget (Science Budget) proportionally to their GDP at each Ministerial Conference. This implies that funding available for Science Missions is highly stable.

The mandatory nature of the programme provides ESA with a much higher control over the implemented activities. Mission selection is not discussed in a top-down fashion at the Ministerial conference, but rather by a bottom-up approach. Principal investigators generate proposal for activities that are evaluated by means of a peer-review process. The ESA Science Programme Committee (SPC) reviews and endorses the result of the peer review process and green-lights the activity implementation. On average, only 1 of 6 proposals is accepted for implementation with this mechanism.

This arrangement of the Space Science programme provides ESA with a high level of control on the definition of the space data utilisation. For each mission, ESA generates a Science Management Plan that must be approved by the SPC and by the Member States. The Science Management Plan is specific to each mission and includes the definition of the approach to the data processing and distribution within the ground segment.

Space Science missions implemented at the national level are very rare; they are usually the result of cooperation with other Members States or contributing partners (e.g. NASA). This is often reflected in the ground segment with distributed nodes for the processing and transmission of near-real time data. For long-term data access, ESA established a centralised data repository in the ESAC centre in Madrid.

ESOC centre in Darmstadt is responsible for the satellite control and housekeeping. ESOC is responsible for the processing and/or re-distribution in raw format of data generated in space (system/platform plus payload/mission) and received on the ground by the ESA ground stations.

The leg of the ground segment dedicated to the payload/mission is usually distributed between ESAC and Expert Centres dislocated in the different Member States supporting the mission. ESA plays a major role in the

Private ground station for data collection

Sometimes the ground segment configuration includes utilisation of private/ commercial ground stations. These are mostly used for orbital missions as they also serve other commercial/satcom projects. Deep-space mission data are generally collected by ESA's dedicated station. The utilisation of private/commercial ground stations might increase in future.

coordination of the activities for the set-up and operations of the distributed ground segment. The final configuration varies by mission and includes different levels of participation of the Member States. For most space science missions, Member States cover up to 80% of the implementation costs, which is a major difference to the approach to Earth Observation missions where ESA controls most of the ground segment.

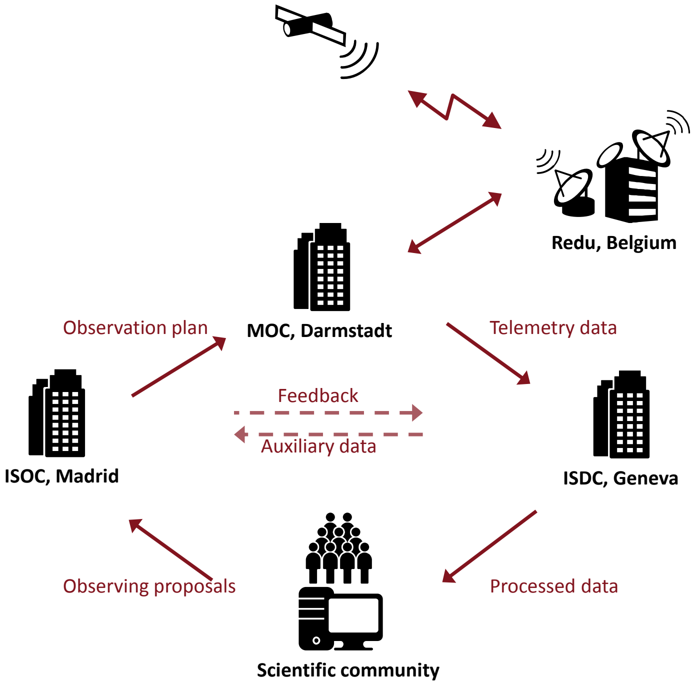


Figure 21: ESA Space Science ground segment (Integral Mission)

In general, ESAC responsible for the integration of the utilisation requirements and the definition of the satellites utilisation plan (e.g. tasking of the satellite for specific data acquisition). The actual commanding of the payloads according to the plan is then implemented in ESOC. ESAC might or might not have a direct responsibility to process the scientific data acquired on-board, this depends on the mission.

In any case, the majority of the raw scientific data is received and processed by the different processing centres distributed within the contributing Member States. Each centre has different infrastructures providing different data processing capabilities. They are responsible for providing and keeping the software maintained for the data processing and for the generation of the metadata.

As for the access to the data, provisions of the Science Management Plan might vary dramatically from mission to mission. In general exclusive access to the acquired data is granted for a period of 1 year to the PI/Science Team requesting the specific data acquisition. In these cases, the request for the access to the data is evaluated by means of a peer review process.

Some missions do not grant exclusive data access. As an example, the entire set of scientific data that the Gaia mission will acquire will be accessible from day 1 of acquisition by the entire scientific community.

Once processed, all data are then stored in the ESA repository centralized in ESAC for the long-term data access, and an open data policy applies. After a process of validation of the data processing activities and of some of the products, data are open for utilisation with no specific restrictions. This is very different from the approach adopted at NASA with the Planetary Data System (PDS), where the actual scientific content of the data is validated.

The ESAC repository includes currently 15 different data repositories dedicated to different ESA missions[3]. Six additional repositories are currently under development[4]. The data are stored according to the Flexible Image Transport System (FITS) format, which is widely adopted by the international Space Science community.

2.2.3 ESA-Microgravity

To characterize the data infrastructure dedicated to microgravity experimentations, we need to differentiate between the data generated on the ISS and the data generated on other platforms such as sounding rockets and parabolic flights.

The experiments executed outside the ISS are implemented as national missions or in cooperation with and overall ESA's microgravity experimental campaign.

No data infrastructure exists for the generated data as the experimental platforms do not offer extensive data management and transmission capabilities. Experiment equipment, in the majority of the cases, consists of a self-standing facility with local data storage capability (e.g. hard drive or any other mass memory platform). In other cases the experiments are implemented as return samples with no production of data in microgravity.

Access to the data is implemented according to the single-PI model (described earlier in the document in section **1.7.1.2**). The data are usually owned and stored by the PI who has also the responsibility to preserve them.

Providing long-term data access is also a PI responsibility, but there is no centralised web portal to access the data. Scientist interested in accessing the data must contact the PI to request a copy of the experiment results.

For the missions implemented as cooperation between a MS and ESA, a centralised archive called Erasmus Experiment Archive (EEA) hosted in ESTEC is used. The EEA provides access to a high-level description of the experiment objectives, main achievements, a list of related publications, and the contact information of the PI personnel. No access to the data in any format is provided.

As for the experiments executed on the ISS, these are always implemented in cooperation with ESA or with other international institutions (e.g. NASA and Roscosmos). At a minimum, ESA participates in the integration of the experiment in the resource budgets available to ESA and therefore in the allocation of critical resources such as up-and download masses, power, data bandwidth, astronaut time, *et cetera*. In other cases, ESA provides access to experimental facilities on-board the ISS, or is involved in the design and manufacturing of new facilities.

When the experiment is designated to take place inside their ISS facilities, NASA and Roscosmos handle the integration activities. Roscosmos implements these activities in-house with internal experts, while NASA has recently delegated this responsibility to the Centre for the Advancement of Science in Space (CASIS) project, an external non-profit organisation.

---

[3] ISO, Exosat, Ulysses, Herschel, XMM-Newton, HST, Planck, Rosetta, Mars Express, Venus Express, Huygens, SMART-1, Giotto, SOHO, Cluster.

[4] Gaia, Euclid, BepiColombo, Exomars, Lisa PathFinder, Solar Orbiter.

Independently from the International institution, the data generated on-board the ISS are managed by the network of European User Support and Operations Centres (USOC) established by ESA. USOCs are distributed over several European countries and support experiments according to their scientific expertise (e.g. biology, material science, fluid science at cetera). Real-time data are processed and distributed to the PIs following a single PI data access model.

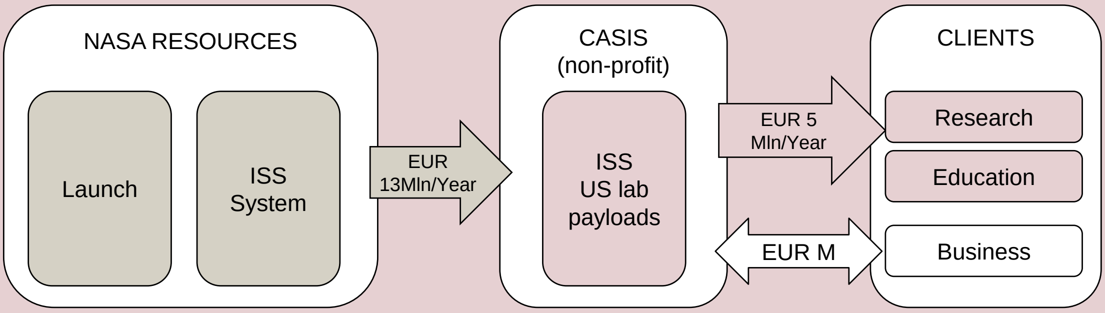
CASE STUDY: CASIS ISS science exploitation

CASIS is a non-profit organisation that manages entirely the exploitation of the International Space Station U.S. National Laboratory.

Its mandate for research is to support and assist researchers and principal investigators in transitioning their science and research experiments into manifested payloads that will be launched and delivered directly to the ISS US National Lab. The scope of the experiments is multipurpose and includes: biological research, physical science and material research, Earth and space observation.

To accomplish this task CASIS receives an annual budget from NASA of about EUR 13 Mln/Year.

CASIS undertook public outreach by promoting the value of the ISS NL to the American nation. CASIS supports activities related to research and education but also developed a robust financial model to supplement government funding by reaching the private sector for product development and innovation.



Post-flight, CASIS facilitates post-processing activities, data and reports submission to NASA and Principal Investigators. There is however no clear initiative from CASIS to facilitate post-mission data sharing amongst the different scientific communities. [36]

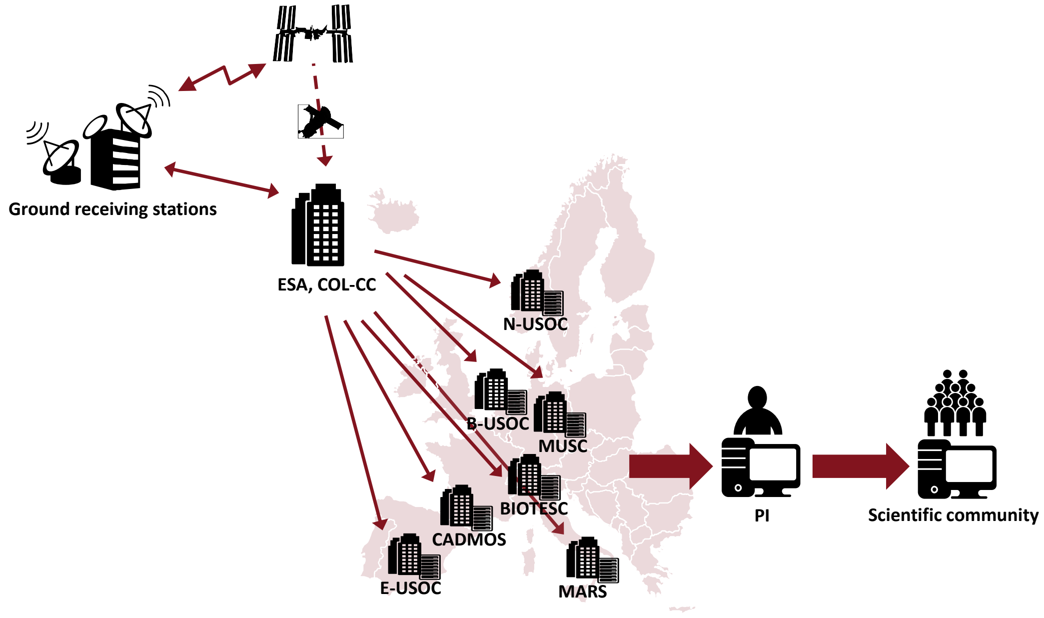


Figure 22: ESA Microgravity ground segment

In addition, USOCs have the mandate to preserve the data according to the PI and ESA's instructions. However, there is no mandate to grant access to the data for third parties PIs nor do the USOCs possess an infrastructure capable of doing so. ESA owns the raw data but the access to the data has to be agreed by the Principal Investigator of the experiment that generates them.

Raw data are stored encapsulated in CCSDS packet and the USOCs are the only entities maintaining the capability to re-process the data and distribute the products. The processing tools are part of an industry proprietary environment (CD-MCS). Furthermore, there is uncertainty about the level of metadata accompanying the data; it may be inadequate or non-existent.

As for the data access policy, in addition to the PI and ESA's disposition, there are different models applicable to the data according to the nationality of the PI. This might represent an additional issue to be addressed in order to grant access to the data and foster their exploitation.

2.2.4 EUMETSAT

The EUMETSAT in an intergovernmental organisation composed of 28 Member States[5] implementing earth observation (meteorological, climate change, ocean and atmospheric composition monitoring) programmes.

EUMETSAT retains the operational and data distribution activities for all its programmes while the procurement of the systems and the development of technologies is delegated to the European Space Agency. No Missions PI are identified for EUMETSAT missions, data are utilised for generation of meteorological products and distributed following an open data policy.

Near-real time data for all EUMETSAT programmes (METEOSAT, METOP and international cooperation Jason-2 and Jason-3) are processed to Level 2 at EUMETSAT Central Facility located in Darmstadt (DE). All data and products generated within the Central Facility are archived in the EUMETSAT Data Centre and can be retrieved, on demand, by users.

EUMETSAT set up in the data centre in 1995 specifically for the Meteosat programme. Today the data centre serves all EUMETSAT satellite programmes, it collection of data comprises different sets of satellite products, including:

---

[5] AT, BE, KR, CZ, DK, EE, FI, FR, DE, EL, HU, IE, IT, LH, LV, LU, NL, PL, PT, RO, SK, SL, ES, SE, UK, plus NO, CH and TK; and 3 cooperating states BG, IC, SB

- Images and derived meteorological products from geostationary Meteosat satellites since 1981,
- Metop-A and Metop-B data since 2007,
- Jason-2 data since 2008, including various sea surface altimetry products.
- Data from international partners (e.g. NOAA) since 2006.

The data centre guarantees a long-term preservation of data and generated products and provides registered users with different data access services (e.g. browse data, make automated orders, and retrieve data from EUMETSAT's catalogue of products). Registered user access the data services via the Earth Observation Portal or via the EUMETCast service. EUMETCast is a multi-service dissemination system based on standard Digital Video Broadcast (DVB) technology. It uses commercial telecommunication geostationary satellites to multi-cast files (data and products).

EUMETSAT retains the copyright on all satellite images. Users can access the data free of charge, but must credit the copyright.

EUMETSAT Direct dissemination

In addition to the access via the Earth Observation Portal, EUMETSAT provides a Direct Dissemination service for Meteosat and Metop missions. Data, products and services, limited to the instantaneous sub-satellite observations, are delivered directly to user reception stations installed at user premises.

2.2.5 EU Member States

As discussed above, with the exception of the Earth Observation missions, most of EU Member States space missions are currently implemented as cooperation with ESA or other international institutions.

Unlike past missions such as CoRoT (France), PICARD (France), CHAMP (Germany) and AGILE (Italy), today space science missions are very rarely implemented nationally by any Member State.

For the mission in cooperation with international institutions, national centres (research Centres, University departments, Astronomy observatories, *et cetera*) are usually involved in the processing and distribution of the near-real time data. The cooperating institution is usually responsible for the long-term preservation of the data (e.g. refer to **section 2.2.2**). In some cases the national centres mirror part of the processed data and make them available to national users, but do not bear responsibility for their long-term preservation.

As for the microgravity experiments, the ones executed on the ISS are always supported by ESA ground segment, while there is not really an active distribution of the data for the remainder. As discussed above the ISS data are preserved in the national USOCs that are partially funded by ESA and partially funded by national space budget.

Member States implement many **Earth Observation** scientific missions as national mission or cooperation with other countries. **Section 2.2** has already introduced a number of these missions (e.g. Odin, Meghatropique, CFOSAT, MERLIN, et cetera) for which the model of the national centres responsible for the near-real time data distribution still applies. Depending on the county, the national centre might be responsible for the long-term data preservation. Germany, Spain and UK follow this approach. In other countries, such as France and Italy, the National Space Agency funding the mission takes responsibility for the long-term data preservation.

A special case in the Earth Observation domain is the national dual-use mission. As already described in **section 2.2**, these are usually implemented as PPP with commercial companies. The latter have the exclusive data distribution agreement for the commercial exploitation of the data and for the preservation of the archive.

The generated data can also be scientifically exploited. It is usually a responsibility of the involved Public entity (e.g. national space agency, specific ministries) to establish a mechanism to ensure the scientific exploitation of the data without interfering with the commercial interest of the private company.

The following sections provide a more detailed description of the data infrastructure and approach to the data access implemented in the European Countries with major space activities at national level.

2.2.5.1 France/CNES [37]

Regardless of the level of contribution of CNES of a mission, CNES analyses what happens with the data over the long-term and addresses the aspect of data preservation in the early stage of the mission/system design.

SERAD (Service for data referencing and archiving) is a process that CNES established to implement the preservation of the data from the early stage of mission design for all missions where CNES invested somehow. The early mission reviews, indeed, already include discussion points where long-term preservation must be prepared. There is also a budget line in the contract to cover long-term preservation. This is mandatory in the design of the mission according to CNES guidelines.

For the Earth Observation missions, thematic data centres are involved in the distribution of the mission data:

- The Icare data centre takes care of cloud and aerosol related data.
- A data centre called Ether and located in IPSL (Paris) focuses on atmospheric chemistry.
- A land surface data centre has also been re-administrated recently.
- Two other data centres for oceans and solid Earth are under construction.

These data centres receive Level 2 data and process them to higher levels using in-situ or other space data sources.

As for the EO missions, Space science data are usually preserved in its thematic centre. This is addressed within the SERAD process that addresses the preservation of orphan data that does not fit into a specific science domain or thematic centre. There are thematic centres assigned to preserve the data for plasma physics, solar physics, and astronomy.

- The Centre de Données Physique des Plasmas (CDPP) thematic data centre for plasma physics was created before 2000 with the initial objective to recover all data stored on tapes.
- The MEDOC data centre handles all solar physics data
- The CDS thematic centre in Strasbourg is responsible for astronomy data.

Not all data are preserved in France under CNES responsibility. Data preservation is usually the responsibility of the entity leading mission implementation and design (e.g. ESA/ESAC, NASA/PDS). Not all mission data are preserved by these entities and for each mission CNES executes an evaluation of what data should be preserved as a complement and/or duplication.

CERES and TOSCA are two scientific committees representing major France research entities and leading this evaluation on behalf of CNES respectively for the Space Science and Earth Observation missions. The process suggests the data to be preserved and gives an indication of the preservation time. Mixed models are possible where part of the data is stored by an international partner (e.g. NASA or ESA) and other parts are stored by CNES. CNES also support preservation of data for partner entities (e.g. Megatropique).

In addition to the preservation of the all CNES mission data, CNES cooperates with other research institutes in France, notably the Centre de Données astronomiques de Strasbourg (CDS) and the Centre National de la Recherche Scientifique (CNRS)/Insitut National des Sciences de lÚnivers (INSU), for the development and maintenance of the French Virtual Observatory, part of the International Virtual Observatory Alliance (IVOA).

2.2.5.1.1 Dual-Use EO missions

The optical EO system Pleaides, implemented by CNES as a French national mission, is equipped with one dual-use optical payload. Its exploitation and distribution is civilly/commercially shared between by AIRBUS GEO (former SpotIamge) and military under the French ministry of Defence. There is a quota of the data reserved for military. The Spot/Pleiades ground segment is compliant with Long Term Data Preservation guidelines.

Data are owned by CNES, therefore it is a CNES responsibility to preserve the data. However CNES cannot distribute the data as AIRBUS has exclusive distribution rights. This is only true up through the Spot 5 mission as the Spot successors do not belong to CNES.

Things are slowly changing with the implementation of new programmes for the access to Pleiades/SPOT data for scientific exploitation, the ISIS programme for the Pleiades system and the Spot world heritage programme for the SPOT archived data.

The idea behind the SPOT world heritage programme is to provide a free SPOT satellite archive imagery over five years old for non-commercial uses. Under the Spot world heritage programme, several hundred thousand SPOT scenes with a 10m resolution have been defined by science users and should become openly available to European users.

ISIS is a programme financed by CNES providing subsidies for accessing EO Data (defined for SPOT in the 90’s for French users only). It is now in the process of being extended for Pleaides and open to all European Science Users, in negotiation with Airbus Defence & Space. The precursor OASIS FP4 programme was for European users. The programme has ended, but the access to the data has been maintained.

Within ISIS scientists submit a request for data exploitation of data (fresh or archived data). Proposals are approved after a peer-review process. If approved, the project team has to contribute to the project costs paying 1€ per square kilometre and per acquisition, CNES funds the rest. The contribution from these institutes is usually quite low and thus accessible to many. However, as the CNES budget is limited, the scientific institutes are limited in how many images can be ordered.

Once the data is acquired and purchased for scientific use, it is accessible to all French Institutional users for science exploitation. There is therefore mutualisation of data usage within the French science user community. There is also negotiation on-going with Airbus Defence & Space to mutualise all Pleiades science data. Users registered in the systems and validated as scientific users can access the images in the archives acquired for scientific exploitation. French “scientific” users might also request for a specific acquisition and tasking of the satellite.

2.2.5.2 Germany [38]

Currently there are no national German Space Science missions. All missions are implemented as international cooperation with ESA or with other counties. Few examples of missions are:

- DLR cooperates with NASA on the Sofia mission (20% of the data is from DLR).
- DLR collaborated with CNES on the CoRoT mission.
- DLR also collaborated on the ESA Microscope mission.
- DLR also had a small contribution on Hayabusa-2 with JAXA.

All space science data follows an open access policy after an exclusive access period. From the legal point of view, the data belongs to the PI and they are responsible for their distribution. There is not a legally binding DLR policy to share the data, but this is what DLR usually does, and this is usually reflected within a Memorandum of Understanding stipulated between DLR and the mission-PI.

There are different centres in Germany with expertise in different research domain, each with the responsibility for the data processing and preservation. Their involvement in the mission depends on the mission objectives. After the end of the exclusive access period, data is transferred to ESAC. The priority period can be up to one year, depending on the technical work needed.

The preservation of the data is currently under the responsibility of the Institutes, and they decide how long the data shall be preserved.

In term of funding, DLR mostly gives grants for the implementation of the scientific part of the mission. These grants usually cover up to 50% of the design and operations costs, but do not designate anything towards exploitation. Funds for the exploitation costs are usually borne by the institutes analysing the data and thus the federal R&D programme.

The German Federal Ministry of Education and Research, funds the activities of the German Astrophysical Virtual Observatory (GAVO), part of the IVOA.

EO missions are implemented both as fully public and as PPP. The PPP usually include a data distribution agreement with a private company but do not exclude the possibility to scientifically exploit the data. Examples of missions are:

- The national Terrasar-X and Tandem-X missions that are PPPs.
- The EnMAP and Merlin missions are fully public missions. Merlin is an international collaboration with CNES.

For EO most of the processing is done in the DLR Satellite operation centre in Oberpfaffenhofen. All data for EO missions are also archived in the data centre in Oberpfaffenhofen.

For EnMAP data policy would be as close to Copernicus as possible since there is no commercial aspect. It is possible to request satellite tasking for scientific purposes, the tasking is open to international community. Conflicts can happen (e.g. during global DEM acquisition). It is the role of the science coordinator to mitigate these conflicts.

The access to the satellite data is currently only for scientific purpose. The science is not exclusive to universities but could also be for company research (data policy based on purpose and not on user).

There are processes to evaluate if a tasking request is really scientific or commercial All data access requests are received by DLR that, on the basis of the user registration information and data request details, evaluates the nature of the request and authorise or deny data access.

A study is undergoing to evaluate how data could be attributed for commercial purposes but there is no clear policy for that at the moment. The general strategy of the national EO missions from DLR is to have a certain applicative and operational use of the data and not just pure science, which is done more with the ESA missions.

2.2.5.2.1 Dual-Use EO missions

For the PPPs in EO (e.g. TerraSar-X and Tandem-X) the tasking is shared by 50% between commercial and scientific exploitation, but the archives are accessed by all collaborators. The priority on data access would be given to crisis management first, then data calibration and then scientific use, especially for science where the acquired data is the most critical.

There is a review process to evaluate the purpose of the data request. If the purpose is scientific the data are generally almost free. If the purpose is commercial, the request then goes through Airbus D&S and the data are priced according to the official TerraSar-X and Tandem-X product and price catalogue. Once in the long-term archive, data are available for all, nevertheless terms and conditions of the PPP might change in the next years.

The funding to preserve the data usually goes up to 5 years after the end of a mission. Therefore the LTDP will become an issue when TerraSar-X and Tandem-X missions end, because the budget to preserve their data (also mutualised for other missions) will end. The oldest data archived might be from the ERS mission. No data has been deleted but there is no long-term preservation mechanism in place yet. Discussions to implement a long-term preservation policy for TerraSar-X, Tandem-X and EnMAP are on-going.

Discussions are also on-going with ESA to align everything with the OSCB-LTDP guidelines and with the INSPIRE directives. The available budget might limit ESAs ability to fully do so.

2.2.5.3 Italy [39]

The **Astrorivelatore Gamma a Immagini LEggero (AGILE)** is currently the only Italian national mission in the Space Science domain. All other missions are implemented as international cooperation with ESA or with other countries. Few examples of missions are:

- **SWIFT,** NASA mission with contribution from Italy (XRT telescope mirrors, Malindi's ground station) and UK
- **LISA-Pathfinder**, ESA missions with Italian contribution for the definition of the overall mission architecture.
- **EUCLID,** ESA mission with Italian contribution for the management of ground segment dedicated to the management of scientific data

Data of all missions with an ASI contribution to the definition and design of the space segment or to the mission data management on the ground (e.g. utilisation of the Malindi ground station or ASI-net for data distribution) are preserved at the ASI Science Data Centre (ASDC) located in the premises of the Italian Institute of Nuclear Physics in Rome.

In 2000 ASI established the ASDC as the data management centre for the BeppoSAX national mission. Today the ASDC is a multi-mission science operations, data processing and data archiving centre providing support to several scientific space missions. At the moment the ASDC has significant responsibilities for a number of high-energy astronomy/astro-particle satellites (e.g. Swift, AGILE, Fermi, NuSTAR and AMS) and supports at different levels other missions, such as Herschel and Planck. In addition it offers mirrors of different repositories from other international missions of interest for the Italian scientific community.

Mission-PIs remain the owner of data and retains an exclusive access and utilisation right according to the data policy applicable to the specific satellite mission. The ASDC implement mission applicable data policies provisions also to define which mission data at which processing level can be made available at the end of the exclusive access period. Data can then be accessed openly with no privileged access rights. Also no utilisation license is requested to download and use any data management software developed at ASDC.

In addition to the direct access via the ASDC website, ASDC data are also accessible via the Italian Virtual Observatory (VOBS) hosted by the Italian Institute of Astrophysics (INAF). ASDC cooperate with the VOBS, part of the IVOA, and makes space science data systematically available through the Italian Virtual Observatory node.

As for the Earth Observation missions, ASI is currently establishing a data centre in Matera to manage all EO mission data. The centre currently has an archive with past NASA EO mission and will archive all future ASI future international cooperation projects. The latter include also a number of bi-lateral agreements to exchange EO data in different bands generated by national missions of the two countries. Two examples are:

- Italo-Argentine Satellite System for Emergency Management (SIAGE) to exchange X-band date taken by the Italian COSMO-SkyMed with L-Band data taken by the Argentinian SAOCOM missions
- ASI agreement with the Japanese Space Agency (JAXA) for the exchange of COSMO-SkyMed X-Band data with Japanese ALOS-2 SAR L-Band data

Besides preserving the EO mission data, the ASI centre in Matera will also serve as Italian contribution to the Copernicus collaborative ground segment. No web-site or data portal has been established so far to access the data

2.2.5.3.1 Dual-Use EO missions

The Italian Space Agency has also implemented the dual-use COSMO-SkyMed mission on behalf of the Italian Ministry of Defence. Tasking for military utilisation of the satellite data has always priority, followed by civil protection utilisation and commercial exploitation of the images. When not tasked for a specific purpose the satellites acquire imaged for the so called "background mission" (e.g. regular mapping of the Italian and Mediterranean area to measure consequences of seismic and volcanic activities).Military and civil data are managed in two dedicated data centre. The data centre for military data has been established within the Air Force base of Pratica di Mare, while the ASI centre in Matera manages all data supporting civilian utilisation and background mission. ASI retains the intellectual property of COSMO-Sky-Med data, while the eGeos company is the commercial partner authorised to exploit the data.

Civil data are made available at different costs according to the entity requesting them. Specific catalogues and prices are established for Commercial (higher cost) and Institutional (lower cost) entities. Scientists are usually identified as Institutional users and therefore access the data at lower possible cost. Once accessed via one of the two mechanisms data cannot be redistributed.

In addition to the two price catalogues described above, ASI publishes open call for near-real time and/or archived data utilisation. More than 167 project between 2008 and 2012 benefitted of the open call mechanisms and accessed COSMO-SkyMed data for free. ASI opened a new open call in April 2015 and more than 30 projects have been implemented so far.

2.2.5.4 Spain [40]

So far all Earth Observation missions have been implemented as an international cooperation. Therefore data policies and data preservation responsibilities were aligned with the ones established by the main implementing country or entity (mainly ESA).

Ingenio and PAZ are the first two 100% Spanish missions that are, from the point of view of the data policy, trail blazers for Spain. The data policy has not been established at national level. The scientific data of Spanish national missions are generally delivered as raw data to the responsible centres for processing (e.g. CNIS, INTA and University of Sevilla for the scientific payloads on-board the Ingenio mission).

As for science missions, these have been implemented also in cooperation with countries outside of Europe, mainly the USA. As an example, the Centro de Astrobiologia (CAB) provided the Rover Environmental Monitoring Station (REMS), a weather monitoring station, as the Spanish contribution to the Mars Science Laboratory rover.

The CAB Centre is also the centre responsible for the data preservation of all the Spanish space science missions implemented outside of the ESA's frameworks. This includes also the REMS missions.

The CAB receives government funds also to develop and maintain the Spanish Virtual Observatory and make data accessible to the scientific community. The Spanish VO is also part of the Euro VO project and IVOA.

2.2.5.5 United Kingdom [41]

In the UK, the Rutherford Appleton Laboratory (RAL) facility supports the UK scientific community for the access to Space Science (mainly Astronomy) and Earth Observation space data.

The RAL is responsible for preserving and distributing scientific and Earth observation data to national users. The RAL mainly provides data from ESA, EUMETSAT and some NASA missions. Raw data are not processed at RAL; they are mirrored from the ESA, EUMETSAT and NASA repositories and are then made available to the UK scientific community.

Preservation of these data is not a STFC responsibility, but of the data provider (ESA in most cases). UK users can download data from the RAL facility to further process them in their own processing facilities, but since connection is always a problem they prefer to work on the data hosted by RAL. In fact, STFC also provides data analysis and processing facilities (hosted process) allowing the user to generate different products without downloading the data sets. All the products generated are then stored and STFC/RAL takes responsibility for their preservation.

STFC is also one of the major contributors to the AstroGrid project, the UK's Virtual Observatory per of the EuroVO and IVOA projects.

2.3 Conclusions

This chapter introduced the concept of the exploitation chain and of the data infrastructure supporting the exploitation of scientific data, both as near-real time and as long-term preserved data.

The analysis shows that, with the exception of some missions in the Earth Observation domain, the vast majority of the space scientific missions with the involvement of a Member State are implemented as a cooperative effort with other countries or with international institutions such as ESA, NASA and Roscosmos.

ESA plays a major role in the implementation of European Missions and, consequently, in the implementation of the data infrastructure per the near-real time and long-term access to the data:

- ESA's Earth observation data infrastructure is implemented following the distributed model and is basically the same for the near-real time and long-term preserved data.
- For the Space science mission, ESA adopted a hybrid model with a distributed infrastructure for the near-real time access to the data and a centralised infrastructure for the long-term preserved data. The distributed infrastructure sees and heavy involvement of the Member States supporting the missions, while the centralised one ensures an open access to the data.
- As for microgravity, the distributed model is implemented for the near-real time data leveraging the network of European USOCs. This provides access to the data following the single-PI data access model. USOCs are also responsible for preserving the data, but no infrastructure or process is implemented for accessing them.

At national level, Member States adopt different approaches for the real-time and long-term data preservation. National centre with specific expertise are usually involved in the real-time processing and distribution of the data.

Long-term preservation is a responsibility of the National Space Agencies in France and Italy, while it relies on the national centres in other countries (Germany, Spain, and UK). In other cases it is delegated to ESA.

More interesting is the case of the dual-use EO missions implemented at national level as PPP. The private partner usually has exclusive data distribution rights for the commercial exploitation of the data. Public entities have the responsibility for the data preservation and for establishing mechanism to ensure also the scientific exploitation of the data without interfering with the commercial activities.

2.4 Notes and references

[25] Euro-VO – Coordination of virtual observatory activities in Europe (16 March 2015, Elsevier)

[26] Open Archival Information System (ISO 14721:2003 => ISO 14721:2012)

[27] For example, the Planetary Data System references several ISO, IEEE, RFC and CCSDS standards to describe the allowed contents.

[28] Next ESA SAR Toolbox (NEST) is an ESA open source toolbox under the GNU GPL licence for reading, processing, analysing and visualising ESA (ERS-1/2, ENVISAT, SENTINEL-1) and other spaceborne* (TerraSAR-X, RADARSAT 1-2, COSMO-SkyMed, JERS-1, ALOS PALSAR) SAR data processed to Level-1 or higher.

[29] The Sentinel-1 Toolbox (S1TBX) consists of a collection of processing tools, data product readers and writers and a display and analysis application to support the large archive of data from ESA SAR missions including Sentinel-1, ERS-1 & 2 and Envisat, as well as third party SAR data from ALOS PALSAR, TerraSAR-X, COSMO-SkyMed and RADARSAT-2

[30] http://earth.esa.int/SAFE/background.html

[31] Stakeholder interview with ESA Earth Observation, G. Kohlhammer, ESA-HQ, Paris, 22/04/2015

[32] Stakeholder interview with ESA Science, M. Kessler, F. Favata, ESA-ESTEC, Noordwijk, 06/05/2015

[33] http://www.iss-casis.org

[34] CASIS Hardware and facilities fact sheet

[35] CASIS Strategic plan

[36] CASIS 2014 annual report

[37] Stakeholder interview with CNES, Geneviève Gargir, Thierry Levoir, Steven Hosford, CNES, Toulouse, 28/05/2015

[38] Stakeholder interview with DLR, Claudia Lindeberg, Wolfgang Frings, Godela Rosner, Bonn, 02/06/2015

[39] Stakeholder interview with ASI, Barbara Negri, Alessandro Coletta, Osvaldo Piperno, Cristiana Cirina, Roma, 25/06/2015

[40] Stakeholder interview with CDTI, Cristina Garrido Gonzalo, Pillar Román, Emilio Vez Rodríguez, Madrid, 26/05/2015

[41] Stakeholder interview with UK Space Agency: Lee Boland and UK Science and Technology Facilities Council: Peter Allan, Victoria Bennett, Chris Pearson, RAL Harwell Oxford, 08 May 2015

[42] OECD (2014), The Space Economy at a Glance 2014, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264217294-en

3 DATA EXPLOITATION ROADMAP AND FRAMEWORK

The objective of this chapter is to define the general data exploitation roadmap applicable in all three domain of interest for this study (EO, Space Science and Microgravity) and understand which are the activities influencing the most the roadmap implementation in each domain.

In the first part of the chapter the roadmap to the maximised exploitation of the data in introduced.

Sections 3.1 shows how all the elements introduced in the previous chapters (e.g. mission model, data access model, data infrastructure) play a role in the implementation of the roadmap.

In section 3.2 the main elements influencing the achievement of the maximised data exploitation are consolidate in the data exploitation framework. Additionally the *status-quo* of the elements of the framework in Europe is provided.

3.1 Data exploitation roadmap

The path to the maximised exploitation of scientific data passes through two intermediate milestones: the near-real time exploitation of the data generated during the missions, and the long term preservation of the data.

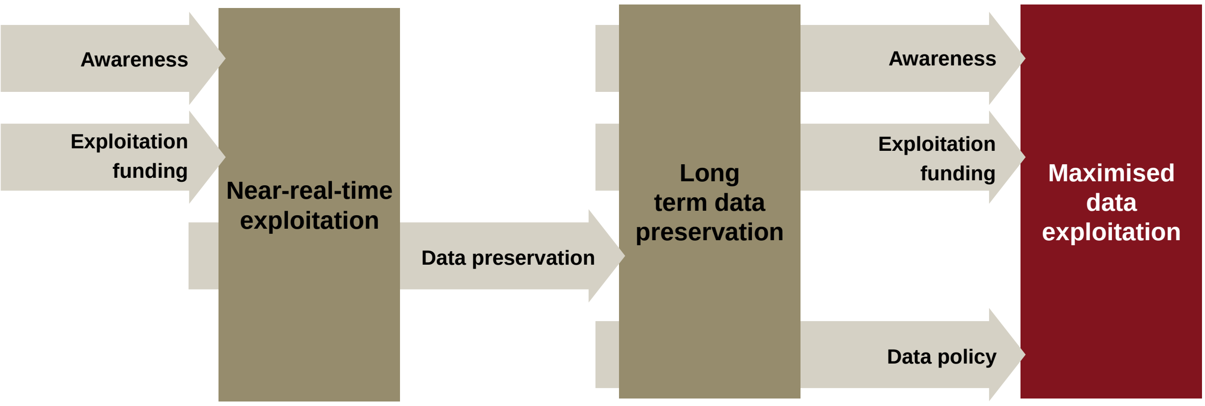


Figure 23: maximised data exploitation roadmap

3.1.1 The near-real time exploitation

The near-real time exploitation of the data generated during the mission is implemented during the operational phases of the missions. As described in **section 2.2**, there are different mission models applied in the three domains of interest for this study. Each model grants access to the data to one or more Principal Investigators.

The PIs accessing the mission data in near-real time are not necessarily the ones involved in the design and validation phases of the missions. According to the mission model, the PI involved in the design phases of the mission typically benefits from an exclusive or privileged access to the data. Nevertheless, this does not exclude other members of the scientific community from using the data in near-real time; scientists can submit proposals for data acquisition and utilisation.

Study analyses and stakeholders interview show two factors as the main influencers for the maximisation of the mission data exploitation: the awareness of the scientific community of opportunities to use the data, and the exploitation funding to implement the research activities.

As we will discuss in **section 4.2**, when observation time is made available, it turns out that awareness within the scientific community is only a minor issue. At least, it is not an issue within the scientific community implementing research activities in the specific mission domain (e.g. Astronomy, Oceanography). The awareness outside the specific mission research field may however be an issue for the goal of increasing cross-fertilisation between different research domains.

As also reported in **section 4.4**, the funding of data exploitation is not uniformly implemented for all Space missions. Each mission is unique and depends on the mandate of the entities involved in its governance.

In particular, ESA's mandate does not include funding exploitation of the data generated by its missions. ESA's mandate stops with the provision of the infrastructure, operational services – including data processing - and access to the data. In Germany, DLR implements a similar approach while other agencies, such as CNES in France and ASI in Italy, have initiatives to directly fund the exploitation of the data.

3.1.2 The long-term data preservation

The long-term preservation of the data is the second milestone of the roadmap. Once the data are acquired, processed, and exploited in real-time, they are usually archived in dedicated data repositories.

The objective of the long-term preservation is to make the data available and exploitable to the maximum extent. To achieve this objective, all the associated knowledge (e.g. metadata and all the information necessary to properly interpret the scientific data) and processing capabilities (e.g. processing software, processing software environment, calibration data) must be preserved together with the raw data.

The three domains of interest for the study each have their own approach to the preservation of the data, but the maturity level of each implementation varies:

- Data standards and data preservation guidelines are established in the Earth Observation domain at European Level. All the major European space agencies (ESA, CNES, DLR, ASI *et cetera*) participated in the definition of the standard and guidelines and are in (non-binding) agreement on their application to the design of the agency's Earth Observation missions.
- Consolidated data standards are established at international level within the Space Science community. In Europe ESA leads almost the totality of the missions applying the "contribution in kind" model for the participation of the Member States. For these missions ESA is responsible for the data preservation and makes use of a centralised archive (ESAC). EU Member States also contribute with payloads and instruments to extra-European (mainly NASA) missions. In these cases, data preservation is a shared responsibility of primary agency and the participating Member States.
- As for the Microgravity domain, no consistent preservation activities have been implemented for the data generated by the experiment platforms. Scientific data generated on the ISS are currently archived in the network of the national USOCs. Data storage standard and the processing environment are proprietary and not open.

As discussed in **section 4.5**, preservation activities also include the "recovery" of data archives from old missions. This activity is not straightforward and could be very time and resource consuming. Data are often archived on old and sometimes non-digital media (e.g. magnetic tapes), and the associated knowledge and processing capabilities might be missing. Therefore preserving old archives often entails digitalisation of the data and the reconstruction of the associated knowledge and processing capabilities.

The achievement of the long-term preservation is the *sine qua non* for the maximisation of the archived data exploitation. This, together with the maximisation of the near-real time exploitation, represents the final milestone of the data exploitation roadmap.

Three factors have been identified as the main influencers for the maximisation of the archived data exploitation: the data policy, the awareness of the scientific community, and the availability of exploitation funding for the data exploitation.

Data policy influences the possibility for the scientific community to access preserved data. The three domains have implemented very different approaches to the definition of the data policy and, in particular, for the access to the data:

- The advent of the Copernicus era is dramatically changing the profile of the institutional Earth Observation missions. The scheme of exclusive access to real time observation and public access to the archived data is still implemented. Nevertheless this model is expected to decrease, as an effect of the Copernicus programme implementation, in favour of data provision as a service to the user and near-real-time availability via web-portals.
- As for the Institutional commercial EO missions, where data distribution agreement with industries are established (e.g. Pleaides/Airbus Defense & Space, Cosmo-SKYmed/e-Geos, TerraSAR/Airbus Defence & Space), National space agencies have also established a mechanism to grant scientific exploitation of part of the archived data.
- An open access policy is consistently implemented in the Space Science domain. Once the data are archived, they are made available to the wider scientific community without restriction, other than perhaps a registration procedure in order to identify scientific users.
- As for the **Microgravity** domain, no general data policy could be identified. For the non-ISS missions, the PI for the experiment owns the data and grants access to them at his/her discretion. For experiments performed on the ISS, ESA owns the raw data but third party access to the data has to be agreed, on a case-by-case basis, by the PI of the experiment that generated them.

For the awareness and availability of exploitation funding, considerations are similar to the ones made for the near-real time data exploitation. The space scientific community is aware of the availability of the data repositories and is making a good use of it. The limitations are more related to the awareness of the non-space scientific community and to the availability of initiatives to fund the scientific exploitation of the data.

3.1.3 IT technology

The absence of an IT technology factor influencing the implementation of the roadmap is not an error or an oversight. The implementation of space data preservation and access does not require development of disruptive IT technologies. On the contrary, innovation of IT technologies and performances of integrated solutions develop faster than the space data exploitation requirements.

The management of information created by in the growing internet community (email, web-pages, search engines), scientific research in medical (e.g. cancer), genetic (e.g. genomic sequences, protein sequences, protein structure and function, bimolecular interactions) domains, as well as management of financial transactions in the bank system, drive much more stringent requirements on IT infrastructure than space data preservation.[43]

Despite the growing amount of data (volume), the increasing rate of data production (velocity), and the increasing number of observation systems (variety) made available to space scientists, the management of the scientific data does not represent a challenging Big-Data issues for current and fast developing IT technologies.

The big-data related issues for the space data are more related to the issues of identifying the right information in the growing volume of data and across the growing variety of sensors have generated and continue to generate. Therefore big-data issues revolve around the definition of common standards for content description and methodologies for data mining, rather than underlying IT technologies.

3.2 Data exploitation framework

The discussion in the previous section shows that there are mainly four elements that influence the exploitation of the scientific data generated by space missions: scientific community awareness, exploitation funding, data policy, and data preservation.

The four elements influence the data exploitation roadmap and play different roles in the implementation of its phases.

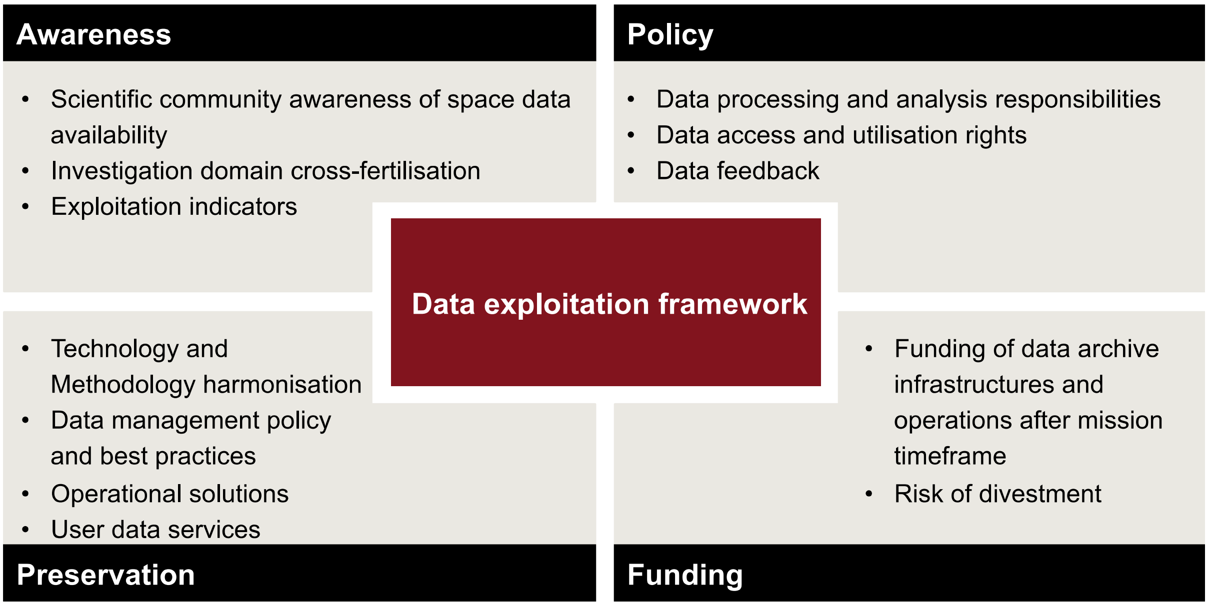


Figure 24: Data exploitation framework

Each one of these elements can be a limiting or enhancing factor to the exploitation of data by the scientific communities. By mapping the current status of these elements, comparing them to best practices outside of Europe, and identifying potential improvement opportunities adapted to the European specific environment, recommendations can be drawn and confronted to major stakeholders' experience.

The following sections look into the details of these four driving elements based on desk research, surveys and stakeholder interviews.

3.2.1 Awareness

In general, European scientific community awareness about the possibility to exploit space scientific data is not a limiting factor, at least not for the near-real-time exploitation of the data.

As discussed in **section 2.4**, Space Science missions and many Earth Observation missions implement a multi-PI data access model. Slots of observation time are reserved by the PI involved in the mission design and in the scientific data validation. The remaining observation time is available for exploitation by other PIs on a competitive basis.

The entities funding the mission issue announce the availability of observation time against which PIs submit proposals for utilisation. Mechanisms are established to evaluate the proposals scientifically in order to grant data acquisition and access to the best proposals.

When proper mechanisms are implemented, a large number of proposals are received. On average, the requests for utilisation exceed the available observation time by a factor of 10. [45], [50]

The high level of awareness is also confirmed by the results of the web survey taken by the scientific community. As described in **section 1.3** the web survey has been one of the pillars of the study data collection activity; it aimed at investigated the broader scientific community awareness of the availability of space data and their utilisation experience.

As shown in **Figure 25**, 95% of the respondents indicate that they are making or have made use of data generated by space mission. Also the 95% of the respondents indicated that they are aware of the existence of science data repositories to access archived data.

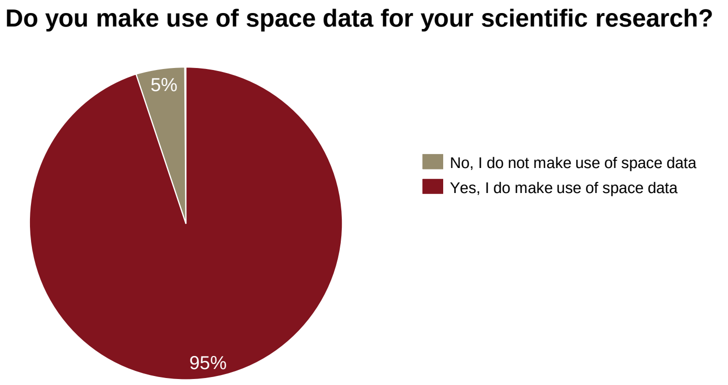


Figure 25: use of space generated data [47]

It is also important to note that not all respondents are or have been members of a mission team, e.g. involved in the mission and data validation activities and therefore granted with reserved observation time. Indeed, only 60% declared that they are or have been member of a mission science team. The remaining 40% have accessed the data via the call for observation time utilisation described above, or by accessing public archives.

The survey confirmed that the high level of awareness also applies to long-term preserved data. As shown in **Figure 26**, 92% of the respondents declared that they have accessed archived data via a data repository. Cross-referencing this data with the previous on the mission science membership, we can conclude that a minimum of 40% of the interviewed scientists have accessed space data exclusively via a data repository.

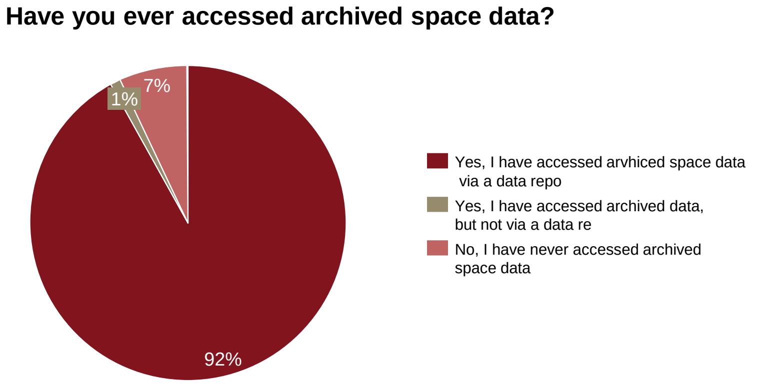


Figure 26: access to archived data

The situation is different for Microgravity missions. As explained in **Section 2.4**, these missions implement a single-PI data access model for the near-real time data. Therefore the concern of a lack of awareness of near-real time data utilisation opportunities is not applicable as the data are accessed exclusively by the mission PI designing the experiment.

As for the access to the long-term preserved data, we have seen that no infrastructure or mechanism is available to grant scientists with a permanent access to preserved data. Indeed access to archived microgravity data when implemented is usually via direct exchange of data with the PI.

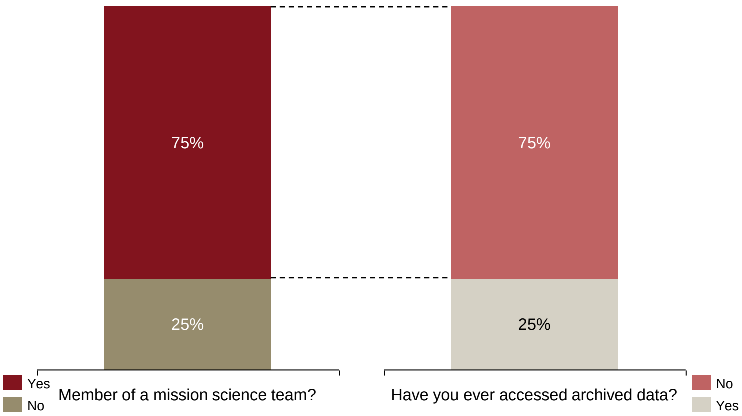


Figure 27: microgravity mission science team membership Vs access to data

As shown in the picture above, this is also supported by the survey results. 75% of the respondents, active in the microgravity field, declared that they are or have been a member of a mission science team and that they have never accessed archived data.

The remaining 25% are not or have never been a mission science team member and have accessed archived data. Their research is mainly based on archived data, but they have never accessed the data via a data repository. The main mechanism to access archived data is via direct exchange with a mission PI or his/her team.

3.2.2 Funding

As discussed in the previous section, there is a good level of awareness of the possibilities to exploit data generated in space for scientific research. Nevertheless, stakeholders interviewed during the study activities consistently reported that there is an underutilisation of the data by European scientists.

The main reason for this is identifies with a lack of funds available for the exploitation of the data, especially when compared with the situation in the US.

NASA, indeed, funds and coordinates all US space scientific missions for Space Science and Microgravity domains, and the vast majority of the EO missions[6]

As such NASA is responsible for the design and manufacturing of the space and ground infrastructure, for their operations and for the distribution and preservation of the data to the US scientific community. Besides these activities, the mandate of NASA includes also the funding of the scientific exploitation of the data.

NASA has established different NASA Science Centres (NAS), dedicated to the exploitation of the scientific data produced by NASA and international space missions. Most active NAS are the Space Telescope Science Institute (STSI), the Infrared Processing and Analysis Center (IPAC) and the Chandra X-Ray Observatory (CXC). [48]

---

[6] The National Oceanic and Atmospheric Administration (NOAA) and the US Geological Survey (USGS) have their own satellites missions and distributes, among other things, satellites data based analysis and products on their websites. These two organisations facilitate the distribution of the data, but do not fund the exploitation of the data.

The US National Science Fund (NSF) has its own research funding programme based on CubeSat technology. The overarching goal of the program is to support the development, construction, launch, operation, and data analysis of small satellite science missions to advance geospace and atmospheric research...

Through the different NAS cyclical calls for exploitation of the data generated by US and International Space Missions are issued. The NAS also have availability of financial tools to finance the accepted proposal and therefore support the data exploitation. The mechanism usually includes a contribution of the proposal submitting entity proportional to the requested observation time (it is usually tuned around 1USD per second of observation). [44], [45], [63]

The Calls address the US scientific community with the objective to stimulate the exploitation of the "open observation" time made available by the space assets. As an example the Cycle 23 proposal for the exploitation of the Hubble Telescope is currently open under the coordination of the STSC. [49]

European scientists are allowed to submit proposal when teamed-up with US scientists, but the funding are issued exclusively for the US components of the international team. [44]

In Europe there are different entities funding scientific mission in space, but there is not a consistent approach to the direct funding of the exploitation.

ESA funds the most of the European missions, but has no mandate to directly fund exploitation of the data. The mandate of ESA stops with the provision of the space and ground assets and the distribution and preservation of the data. [45]

At national level CNES in France and ASI in Italy have mechanisms in place to fund the exploitation of the data from national and international missions. The processes are similar to the one described for NASA. Scientist submits proposals for utilisation of space data that undergo a peer-review and the evaluation of a scientific committee. Once accepted the proposal are funded by the national agencies with a small contribution requested to the scientist.[63][44]

Other European national agencies, such as DLR, follow an approach similar to the ESA's one with no direct funding of the exploitation. In these cases the research teams involved in the exploitation receive funds through other mechanisms, often within national Research & Development programme no strictly related to the research in any of the space disciplines.

Evidences show that when implemented these mechanism for the direct funding of the exploitation generate good results and line up the scientific production with the one achieved in the US thanks to the NASA funding (ref to case study on the Cooperation between the Italian Space Agency and the Italian Institute for of Astrophysics)

CASE STUDY: Cooperation between the Italian Space Agency and the Italian Institute for of Astrophysics

The **Italian National Institute of Astrophysics (INAF)** is the organisation coordinating all Italian observatories, research centres and cooperates with universities. INAF cooperates with the Italian Space Agency (ASI) for the exploitation of the scientific data generated by the Italian space missions (past and current). The latter includes also all cooperation with other countries (e.g. US/NASA, other EU-member States and ESA).

A specific agreement cooperation ASI-INAF was established to address the underutilisation of the fresh and archived data generated during the missions. The agreement aimed at funding the research activities and not at creating awareness about data availability.

The fact that proposal for data exploitation are accepted by national space agency does not necessarily means that the research is then executed at its best. Often unavailability of necessary funding reduces the quality of the research activities (with limitation of means and/or time).

This was the main reason why the ASI-INAF cooperation was established: to fund these research projects and increase the exploitation of the data. Scientists submitted their projects to a joint ASI-INAF for a peer review scientific and technical evaluation.

It funded about **40 research projects** of (**12/18 month duration** and a **funding of rough order of magnitude of 2 to 300 KEuros**) in the Cosmology and Astrophysics domains, addressing both fresh and archived data. As reported in **Error! Reference ource not found.** these activities generated a good number of publications already after one year and half from the kick-off of the cooperation.

Type of publication	Number of publications
Articles on magazine with referees or in press	158
Additional submitted articles	60
Articles in preparation	52
Other publications (e.g. proceedings)	150

Table 3: number of publication produced by the ASI/INAF cooperation after one year and half from kick-off [50]

The number of publication are lined-up with the general profile recorded in the US for similar missions [51]

Figures of the Chandra mission publication statistics confirm these rates. The first publications are normally made about 2 years after the data acquisition, and generally significantly involve the PI in charge of the observation. The number of subsequent publications based on the data reaches its maximum release rate around 5 years after the data has been archived and released publicly. These publications come from the exploitation of the released datasets by other scientific teams. Typically, the maximum exploitation is reached 3 years after the data observation. Very few publications are produced 12 years after the observation. Therefore, over a typical 5-year mission, the maximum data exploitation will be reached on year 5 and datasets will be used at least until year 17.

3.2.3 Data preservation

Data preservation is a fundamental element of the exploitation framework. As described in section 3.1, availability of long term preserved makes scientific data accessible by PIs other than the ones involved in the definition of the mission and allows for the execution of relevant scientific investigations. Preservation means preserving the data, the associated knowledge and the possibility to reprocess the data at any time.

The original processing of data is performed by using algorithms specified by the original mission scientists. Every new mission might implement newer and better processing algorithms. In order to create “Time-Series” and compare data of new mission with data of old missions, the latter must be regularly reprocessed comparable to the new algorithm and made compatible with the former in terms of quality and format. Therefore it is fundamental to maintain the ability to reprocess archived data which is a challenge given the rapid evolution of processing hardware and software.

CASE STUDY: ENVISAT Advanced Along-Track Scanning Radiometer (AATSR) [10]

The AATSR was an instrument on board the European Space Agency (ESA) satellite ENVISAT. It was designed primarily to measure Sea Surface Temperature (SST), following on from ATSR-1 and ATSR-2 on board ERS-1 and ERS-2.

The algorithms designed to process the data achieve the scientific mission objectives filtered out the “disturbance” in the signal generated during each passage of the ENVISAT over the South Atlantic Anomaly i.e. the area where satellites enters the Van Allen belts. Satellites and other spacecraft passing through this region of space actually enter the Van Allen radiation belt and are bombarded by protons exceeding energies of 10 million electron volts at a rate of 3000 'hits' per square centimetre per second. This can produce 'glitches' in astronomical data, problems with the operation of on-board electronic systems, and premature aging of computer, detector and other spacecraft components.[52]

After the end of AATSR mission a group of PI investigating the dynamic of the SAA generated a new algorithm to re-process the preserved AATSR data without filtering out, in fact amplifying, and the disturbance in the acquisition signal. The products generated by the reprocessing were then used to investigate the variation in time of the SAA position with respect to the Earth.

European Institutions and Organisations implementing scientific space mission have among their scientific objectives the preservation of space science data for years after the end of a mission. Yet long-term preservation is not always considered as a necessary part of the post-mission phase or budget constraints may limit the possibilities to do so. Typical mission-project budgets cover the cost of data management for at least five years beyond the end of the mission operations. After that new mechanism and new funds need to be identified for the long term preservation.

3.2.3.1 Earth Observation

The figure below, illustrate this concept for the ESA Earth Observation missions. To ensure data preservation beyond these initial five years, ESA has been proposing a Long Term Data Preservation (LTDP) budget as part of its ESA General/mandatory budget. After the first five years covered by the mission budget, the preservation of the data would under the LTDP budget.

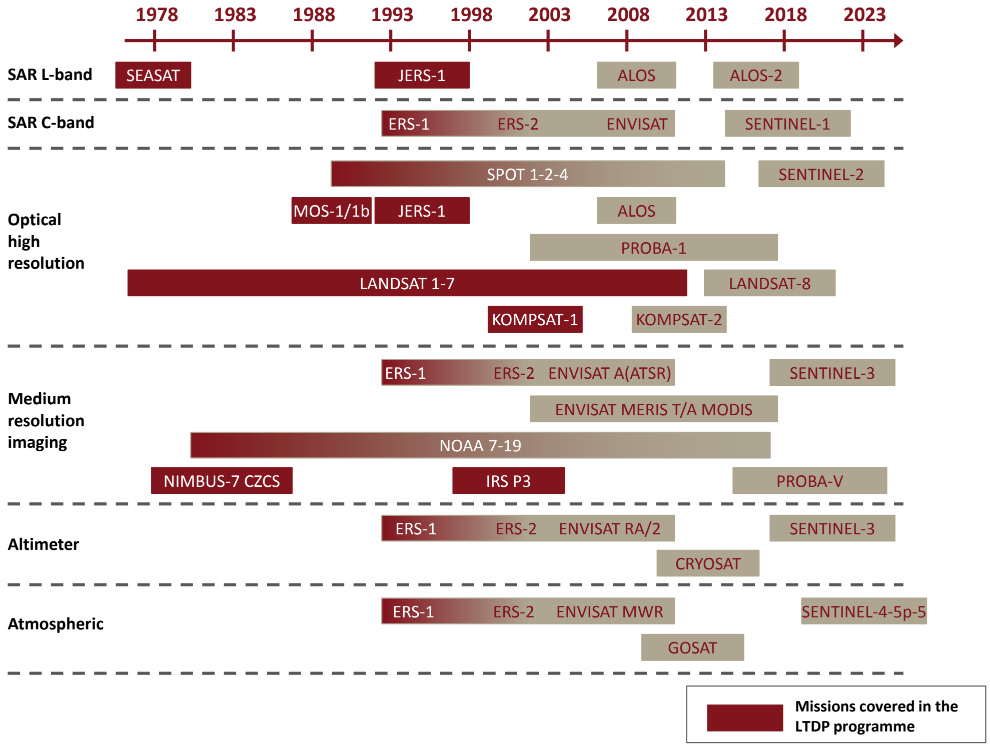


Figure 28: ESA mission under the LTDP budget

ESA, in cooperation with major Institutional Space Players in Europe (CNES, DLR, ASI and others) has also developed LTDP guidelines for the design of the elements of the data infrastructure dedicated to the preservation and access to the EO data. The idea is to line-up the design of the future ground segments supporting new EO mission across all the European countries. This will facilitate the interoperability of the data repositories and the access to the data for the scientists.

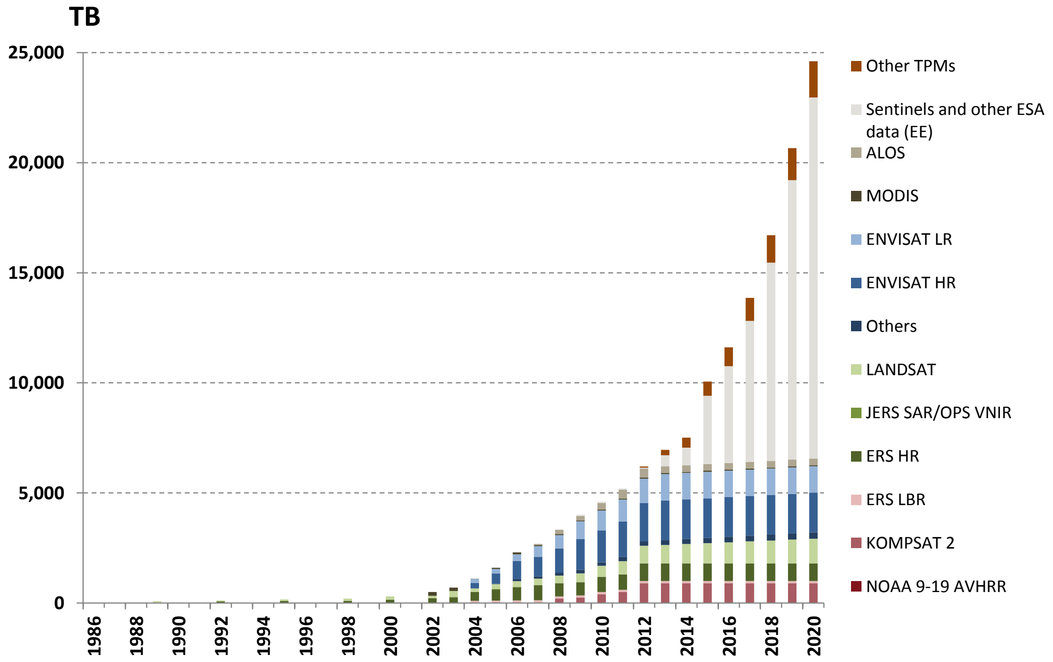


Figure 29: ESA Earth Observation Data Archive size [46]

As we will see in **section 4.5.1.1** dedicated to the LTDP guidelines, ESA Member States generally support this approach but face financial constraints and discussions on priorities which often affect the level of adherence. This often results in the reduction of "data preservation" to sheer "data archiving" leaving out reprocessing or data accessibility improvements. Nevertheless the LTDP guidelines will ensure a certain level of interoperability between future missions.

CASE STUDY: Long Term Data Preservation Guidelines

The LTDP is one of the activities of the **Ground Segment Coordination Body (GSCB)**. The GSCB has established a series of detailed guidelines for Long Term Data Preservation of Earth Observation space data. [59], [60], [61], [62]. The GSCB was established in 2005 and is composed of Member State agencies managing EO data ground segments. The group coordinates and shares its findings with other coordination and standardisation entities outside of Europe such as CEOS (Committee on Earth Observation Satellites), OGC (Open Geospatial Consortium) and CCSDS (Consultative Committee for Space Data Systems), and it plans for regular consultation with industry and commercial missions.

The GSCB was established to coordinate the activities of all the entities in Europe implementing ground segments for EO space mission (e.g. ESA, CNES, DLR, ASI, CSA, UKSA, and EUMETSAT). GSCB activities over the last 10 years continue to include:

- the definition of common LTDP guidelines for the implementation of the ground segments. LTDP guidelines define design requirements and interface requirements among the different blocks of the data infrastructure and, as a consequence, also interface between different data infrastructures. The main objective is to ensure interoperability and support the preservation. The LTDP guidelines also establish the SAFE standards as the reference standards for the archiving of data.
- Common GS building blocks to allow re-use of technologies across different missions
- Common data quality parameters and standards
- Common position in international for a such as GEO and CEOSS
- Sharing of networks (e.g. GEANT)
- Cross-calibration of instruments
- Complementary mission/acquisition planning

Copernicus core ground segment and the collaboration interfaces and processes are benefitting from these the GSCB activities. A data infrastructure accommodating the Copernicus dedicated data and providing access to the Contributing Missions data. GSCB's activities were kicked-off, Copernicus was not yet planned.

The GSCB established general principles for long-term data preservation of Earth Observation data:

- **How long** should the data be preserved: A minimum of 50 years
- **When** should the preservation process be executed: mainly during the post-mission phase (phase F)

- **What** data should be preserved:
    - Data records archiving: raw data to higher level products, browses, auxiliary, ancillary and calibration and validation data set
    - Processing software: processors to generate mission products, visualisation, quality control and value adding software and tools
    - Mission documentation: mission architecture, product specification, instrument characteristics…
- **How** should data be preserved: by following LTDP guidelines. For instance by generating inventory, assessing capability of preservation and accessibility, defining preservation strategy and approach, implementing preservation actions and monitoring associated risks

Because of the variety in governance of data centres in Europe, the GSCB cannot define mandates but can only propose to follow guidelines. A total of 69 key guidelines have been established. For each one of the long-term preservation guidelines 3 levels of priority have been defined.

Level A concerns basic data security, integrity and access.

It includes for instance:

- guideline 1.1: preserve data records, processing software and mission documentation
- guideline 1.2: generate and maintain a complete inventory of the archived Preserved Data Set Content

Level B concerns medium data security, integrity, access and interoperability.

It includes for instance:

- guideline 1.4: assess and harmonize the format of all the "preserved data set content" elements
- guideline 1.5-1.7: adopt a common standard archive, exchange and documentation format
- guideline 6.12: Apply policies and procedures that enable the dissemination of EO products that are traceable to the source data
- **Level C** concerns high level of data security, integrity, access and interoperability.
- It includes for instance:
- guideline 2.11: put in place and maintain mechanisms for monitoring the understandability and usability of the archive content.
- guideline 5.3: perform archived data repackaging and/or reformatting to comply with new standard formats and/or exchange formats.
- guideline 6.7: pursue common approach for the Earth Observation data set content access systems to improve compatibility of different systems

Some organisations in Europe are already compliant with level A and partially with level B. Most of the national agencies and data centres (described in **section 3.2.4** of the infrastructure chapter) generally try to comply with these first levels of preservation.

For instance SERAD (service for data referencing and archiving) is a process that CNES established to implement the preservation of data from the early stage of mission design for all missions where CNES invested somehow. All early mission reviews at CNES, already include discussion points where long-term preservation has to be prepared. A budget line is also drawn contractually to cover long-term preservation and is mandatory in the design of the mission according to CNES guidelines. The SERAD process covers also for the preservation of orphan data that doesn't fit into a specific science domain or thematic data centre [63].

National agencies are often well aware of the GSCB long-term data preservation guidelines but their implementation is often subject to limited fund availability. The funding to preserve the data typically goes up to 5 years after the end of a mission. Beyond this period, if no dedicated budget is allocated for the long-term preservation, agencies try to mutualise mission budgets (e.g. TerraSAR-X and Tandem-X support the preservation of other EO missions) or have to make trade-offs. For instance only lower level data may be preserved with the associated knowledge in order to preserve the ability to reprocess all data or on the contrary only higher-level processed data of most scientific interest might be preserved. The latter would be the case for very old data that was not preserved in a digital format. The cost to digitalise the low level data could then be too high and the associated knowledge may be lost. [64] [65]

Preservation of past missions, with supporting data infrastructures not aligned to the LTDP guidelines is a major issue. In some cases the data produced by these missions are not stored on a digital support, but still on tape or other kind of magnetic support. In order to ensure preservation of these old archives it is necessary to:

- Access the data (data are sometimes stored on tape or on other platforms that might be difficult to access if the original hardware system is not available or not in operational conditions)
- Access the associated knowledge and other support information: if present and accessible (similar problem as for raw data), this information will support the understanding of the scientific data and their processing. In case metadata are not available or not accessible they need to be rebuilt from scratch, and this can be a high resource consuming activity.
- Process the data: this implementation requires the original software and/or the original data processing system. Also in case the software is not available it needs to be re-built on the basis of the available data and metadata.
- The processing of the data with the original software also implies the availability of the right software environment and/or hardware to run the software and/or the processing system.

Once the GS are lined-up with the guidelines, ensuring data preservation is an “easy” job (from a technical point of view and excluding all possible data policy issues). Data, indeed, are already in digital format and stored in the same format (mostly the ESA’s defined Standard Archive Format for Europe SAFE). This facilitates the migration to new platform and standards, following evolution of the IT and its related technologies (one migration every 5 to 7 years is considered normal).

In addition to the EO mission implemented by ESA, or in which ESA participated at different levels, ESA is also considering the recovery and preservation of third parties EO missions that generated data of interest for Europe. ESA interest is mainly focused on the data covering the European region and whose preservation is not ensured. In most of these cases, indeed, the third party implementing agency is not ensuring long term preservation of the data, while in other cases ensures the long term data preservation exclusively for the a subset of data of interest.

As an example, NASA doesn’t ensure the preservation of the complete Landsat 1 to 7 data archive. The original missions provided global coverage of Earth but, due to the limited on-board storage capability available in the 70’s, a network of ground stations distributed between the American and European continents was used to acquire the data on the ground. Today the data are still available in different distributed repositories, nevertheless the U.S. are only interest in the preservation of the dataset covering the American continent and the former Soviet Union area of influence (including the territory of the Soviet Union itself). Preservation of dataset covering European and Mediterranean region is out of NASA scope. [66]

The ESA Earth Online portal currently lists about 40 third parties mission that generated subsets of data of interest for Europe, the preservation of which is not ensured. Relevant examples of historical missions are:

- Landsat 1 to 7 missions operated by NASA, National Oceanic and Atmospheric Administration (NOAA) and U.S. Geological Survey (USGS).

- SeaSat mission operated by NASA/JPL Earth Observation division
- QuikScat mission operated by NASA/JPL
- Nimbus-7 mission operated by NASA and NOAA
- Kompsat-1 mission operated by the Korea Aerospace Research Institute (KARI)
- JERS-1 mission operated by Japan Aerospace eXploration Agency JAXA
- IRS-P3 mission operated by Indian Space Research Organisation (ISRO)
- ALOS a Japanese Earth-Observation satellite, developed by JAXA

ESA is currently evaluating the relevance of the 40 third party mission of interest with the support of the scientific community, to prioritize their recovery. [66]

Ensuring the preservation of data from past ESA and third party missions is a time consuming and expensive activity. A total of 15 years are estimated as the time necessary to acquire and preserve the data sets of interest for Europe, with a cost of about EUR 20 Mln/ per year to ensure also the maintenance of the total EO data archives. The estimated cost of some 20M/year is rather constant as the cost of the hardware and IT implementation is expected to decrease counter balancing the need for bigger storage and faster processing capabilities. [66]

3.2.3.2 Space Science

A different situation applies for the ESA Space Science Missions, the main difference being that there is a much smaller volume of data to be preserved. As described in **Figure 29**, ESA EO data archive size is estimated in around 10 PB of data, while the Space Science total archive is estimated below 400 TB in total. [45]

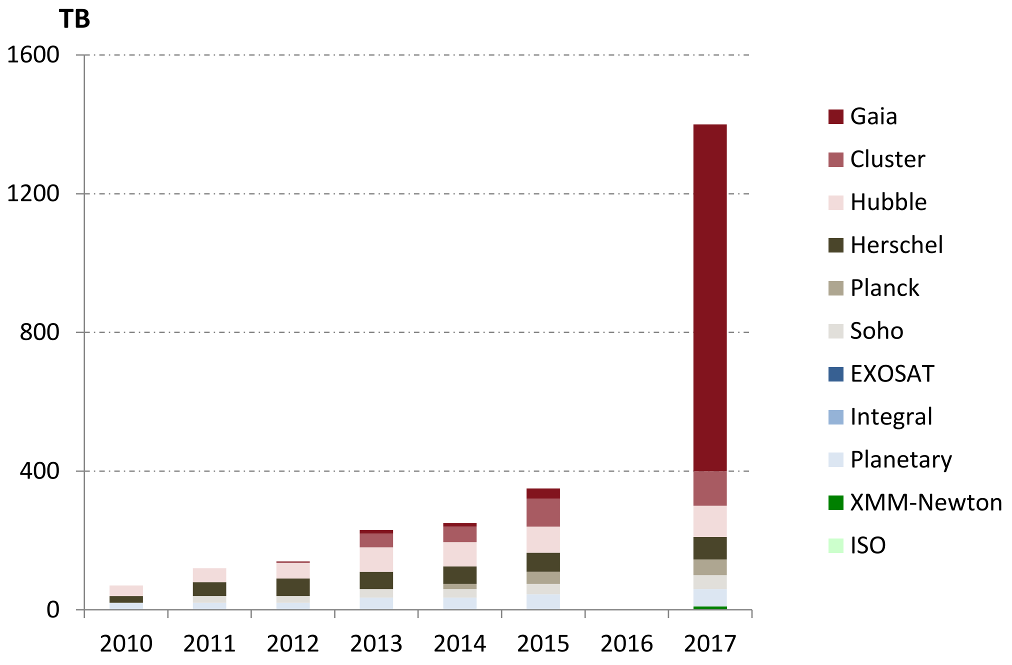


Figure 30: ESA Space Science Data Archive [45]

This is going to change in the next years as new mission as Gaia and Euclid expected to produce respectively 1PB by 2017 and 170 PB by 2025 bringing the total archive size beyond the threshold of the Petabytes. Because of the volume of data, current costs to run the archives are pretty low (estimated to EUR 0.3 Mln per year per mission archive), but these costs are expected to increase as the volume of data increases. The costs of maintenance and operation of the archives are included in the 10-15% of the total Science budget allocated to "General Activities". [45]

The data are stored according to the Flexible Image Transport System (FITS) format, which is widely adopted by the international Space Science community. It should be noted that there is no legally binding requirement to store data in the FITS format.

CASE STUDY: Comparison between the SAFE and the FITS data standards.

There have been many data formats used for Earth Observation missions. In the past, the criteria for selecting the format focused on the satellite's mission. As the missions ended and time passed, it became apparent that preserving data from these missions was becoming increasingly more difficult.

One of the early issues involved the degradation of physical storage media. Magnetic tape can degrade as the magnetic charge weakens, CD-ROMs and laser discs can become unreadable as the dye they were manufactured with breaks down, and hard disks can fail at an instant. Sometimes the only hardware that can read the mothballed media is discarded when no one realizes archives exist than have no other way to be read.

There is a false sense of security that anything that has been digitised can be copied indefinitely, which is true, but either this data is properly stored with backups on dedicated and permanently funded archival systems, or someone must periodically transfer the data to new media. This requires both planning and funding, and because increasingly impractical as the sheer quantities of data increase with each new mission.

Even if the problem of archiving the data is solved, sometimes the ability to use them is lost. Perhaps a unique program was written to decode the data, but that program only runs on obsolete computer systems that are no longer produced. In cases like that, an emulator of the original hardware can be used if the program itself is still available. With experience, the ability for space data to be preserved over the long term became an increasingly more important requirement. The formats used on Earth Observation missions tended to be driven by the sensor used, and ground segments were reinventing processes and components for each mission. Those issues, coupled with some shortcomings on the formats, inspired ESA to develop the Standard Archive Format for Europe (SAFE) which is compliant with ISO-14721:2003, the Open Archival Information System (OAIS). The Sentinel missions are the first to package their original data in the SAFE format with the express goal of facilitating the preservation of the data for a long time. ESA is converting completed missions, even missions of other agencies such as JAXA and NASA, to the SAFE format to preserve their data. It is reasonable to expect that further Earth Observation missions from Europe will adopt SAFE.

The format situation for Astronomy has historically been more uniform as compared to Earth Observation. The Flexible Image Transport System (FITS) has been used continuously in this domain since version 1.0 was standardised in 1981. The current version of FITS is 3.0; it was standardised in 2008. By design, FITS is always backwards-compatible, so any tool that can read FITS 3.0 files can also read FITS 1.0 and 2.0 files. The designers of FITS have always made the suitability of the format for long term storage a priority.

While FITS files typically contain images, they are not limited to this data type. Files can also contain lists, records, and even databases. The metadata scheme in FITS is flexible. There are standard keywords and information, but individual missions can customize metadata as necessary. The flexibility of the format explains why it has endured within the Astronomy community. It even has applications for digital preservation. Rather than use an OAIS-based format or TIFF, the Vatican selected FITS v3.0 as the format to preserve 80,000 ancient manuscripts, some almost 500 years old. Some of the reasons cited for the selection of FITS were the open nature of the format (unlike TIFF), the ability to have files greater than 4.2 gigabytes (the limit of TIFF), its dedication to backwards compatibility, and the availability of tools and converters that can already read the format.

	Standard Archive Format for Europe	Flexible Image Transport System
Abbreviation	SAFE	FITS
Domain	Earth Observation	Astronomy
ISO standard	Yes - ISO-14721:2003 Open Archival Information System (OAIS)	No
Developed by	European Space Agency	IAU FITS Working Group
Current version	Control Book 2 - Recommendation for Specialisations v1.11 (2010) SAFE Basic Schema Set v1.3 (2011)	3.0 (2008)
Backwards compatible	Yes (inherent to XML, unknown attributes can be ignored)	Yes – by design
Mission specific	Yes – each mission has dedicated SAFE format	No – format confirms to specification
Available tools	Tool set provided by ESA. Tool needs to be updated for each new mission	FITS Liberator (co-developed by ESA), 30+ years of compatible tools GIMP, Photoshop, ImageJ, etc. can read image portion of FITS files
API available	All XML tools	Yes, more than a dozen
Metadata	Specified by mission specific SAFE format	Standard keywords and supported by specification (mission definable)
Conducive to long term archiving	Yes (primary goal)	Yes (primary goal)
Efficiently packed	No (text/xml based)	Yes (binary)
Size limit of file	?? (should only be limited by OS)	Unlimited (limited by Operating System?)
Data types	No limitation (general format)	No limitation (general format)
Mixing data possible	Yes (no limit)	Yes (no limit)
Self-describing	Yes (inherent in format)	Yes, through human readable metadata
Open/Free format	Yes	Yes
Adoption	Europe only (so far)	Universal for domain
Indexable / searchable	Yes	yes

Table 4: comparison between the SAFE and FITS standards

With regard to data exploitation, neither format impedes use by interested scientists. The FITS format is widespread and there are more tools and libraries available for FITS files, but ESA provides free and open tools to read and convert SAFE-formatted files. In theory, a SAFE format could be developed for future European astronomy/astrophysics missions but there may not be any appreciable benefit over FITS in this case, other than perhaps the sharing of archival and indexing infrastructure which would drive down maintenance costs in the long term.

.

3.2.3.3 Microgravity

The amount of data generated by the European microgravity experiments on the ISS is estimated in the order of magnitude of the hundreds of TB. This is the sum of the raw data and the associated knowledge. In the microgravity domain it is estimated that each byte of raw data requires 10 bytes of associated knowledge to be correctly interpreted.[68]

Example of these would the general health condition of the astronauts at the time of the experiment (e.g. blood sampling), information on his/her diet in the days before the experiment and so on. Other example would be the general conditions of microgravity and temperature at the time of crystal formation for material science and fluid dynamics experiments.

As described in the previous chapter, no active preservation activity has been implemented so far by ESA. Data are owned by ESA and are stored in the processing centres and their distribution is under the authority of the PI and ESA.

Within ESA the microgravity activities are organised in two major branches, the Life Science (e.g. biology, human physiology) and Physical science (e.g. material science, fluid dynamics, fundamental physics, astronomy). While the scientific communities in both Life and Physical science would welcome the possibility to preserve and exploit data generated in space, it seems that there is not harmonized position in ESA. ESA 's Life science organisation is mostly in favour of investing resources in the data preservation, while different opinions are registered within the ESA Physical science organisation. Discussion are on-going within ESA to sort out this issue, but so far this had led to the lack of a coordinated effort to preserve scientific data generated during microgravity missions of all kinds. [68]

The main results in these fields have been obtained with the ULISSE and the CIRCE projects funded respectively by the European Commission FP7. These two projects represent a success story as they set the first building blocks to preserve data from space experiments.

In 2008 ULISSE was a real pathfinder. It was the first brick of the data preservation for the microgravity and space weather ISS experiments. The project defined standards for the preservation and storage of the data based on the tailoring of the applicable international standards (as ISO19115 for metadata schema, ISO/IEC 13250 for semantic technologies and OAIS-ISO14721 for long-term preservation guidelines). ULISSE approached legal issues (related to data property and use rights) and technological solution, implementing also a demonstrator that collected and exploited the data generated by more than 30 experiments from ESA (data use was authorized by the higher management of the ESA's Human Space Flight Directorate at that time). [68]

The demonstrator worked fine, proving the feasibility and usefulness of a data e-infrastructure supporting preservation and exploitation of scientific space data. The design solutions identified for the ULISSE demonstrator are applicable for implementing a larger scale data e-infrastructure, able to record a measurable impact on European research, reaching a critical mass of experimental data. It is suitable to preserve whatever data is not already covered by EO and Space Science data preservation activities.

CIRCE has taken the work further developing a roadmap for the implementation of the system on a large scale. It also addressed other issues specific to the data exploitation in the microgravity domain, such as the handling of different country based data policies and the definition of easy to use front-end systems to facilitate data discovery and interpretation. Very innovative solutions have been proposed to address both issues.

3.2.4 Data policies

Data policy plays a fundamental role in the definition of the access to the mission generated data. In general the scope of mission data policy can be very wide, but not all elements of a generic data policy are of interest for this study.

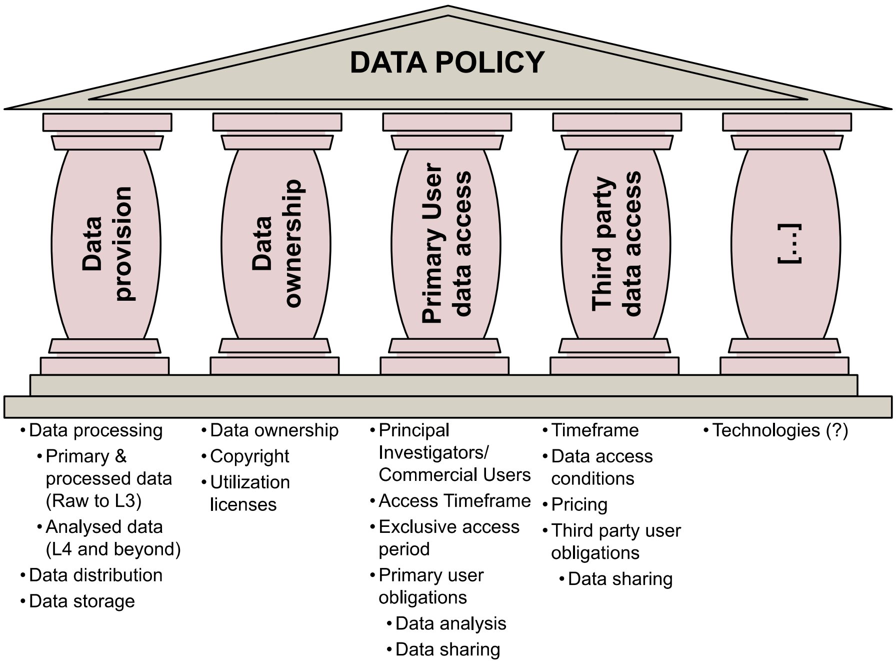


Figure 31: Data policy pillars

As summarised in **Figure 31**, below, the analysis in this study is focused on four main pillars of a generic data policy:

- Data provision establishes the role and responsibility for the processing, distribution and storage of the data. According to the time of mission and to the specific scientific objectives, the share of responsibility between the mission implementing/funding entity and the PI might be different. In particular the responsibility for the processing of the data might vary with the mission implementing entity processing and storing raw data up to a different processing level.
- Data ownership, establishes who owns the data, who has the copyright on the data and who has the utilisation rights.
- Primary user data access is where the access of the mission PI or the PI requesting near-real-time observation time is defined. This pillar defines the observation time dedicated to the mission PI and establishes the legal principles for the access to the satellite tasking and observation time for other PI. The applicability and the length of the exclusive data access are also defined here.

  In case of commercially exploited mission, this pillar also defines the duality between the commercial and scientific access to the data. It also defines the distribution rights for the data commercial exploitation.
- The third parties data access defines the principles for the access to the long term preserved data, i.e. the access to the data for the wider scientific community.

The next sections provide a summary of the main principle established by the data policies in the three domains of interest for this study.

3.2.4.1 Earth Observation

In the Earth Observation domain the advent of the Copernicus programme is radically changing the approach to the data access. So far the scheme of the exclusive data access for the mission PI has been, and still is, widely applied to the Institutional scientific missions. Nevertheless, the upcoming free availability of an unprecedented volume of near-real-time EO data generated by the Copernicus system is orienting the institutional missions toward a more service oriented approach for the access to the data. In this approach the data are made available to the wider scientific community in near-real-time within hours from acquisition in space.[63][65][66]

The exclusive data access scheme applied to the most of current Institutional scientific EO missions is summarised in Figure 32, below, and reflects the current ESA's approach to the data access. Data generated on board upon a PI request are processed by ESA up to level 2 or 3. ESA owns the processed data and has obligation to store them and make them available, under an utilisation licence agreement to the PI requesting the specific observation time.

The PI covers for the cost of analysing the data and generating scientific products. He/she owns the analysed data and the scientific products and grants ESA with utilisation rights. These apply exclusively in the case ESA implements a new scientific mission or programme requiring the analysed data as starting point for the achievement of the new mission scientific objectives.

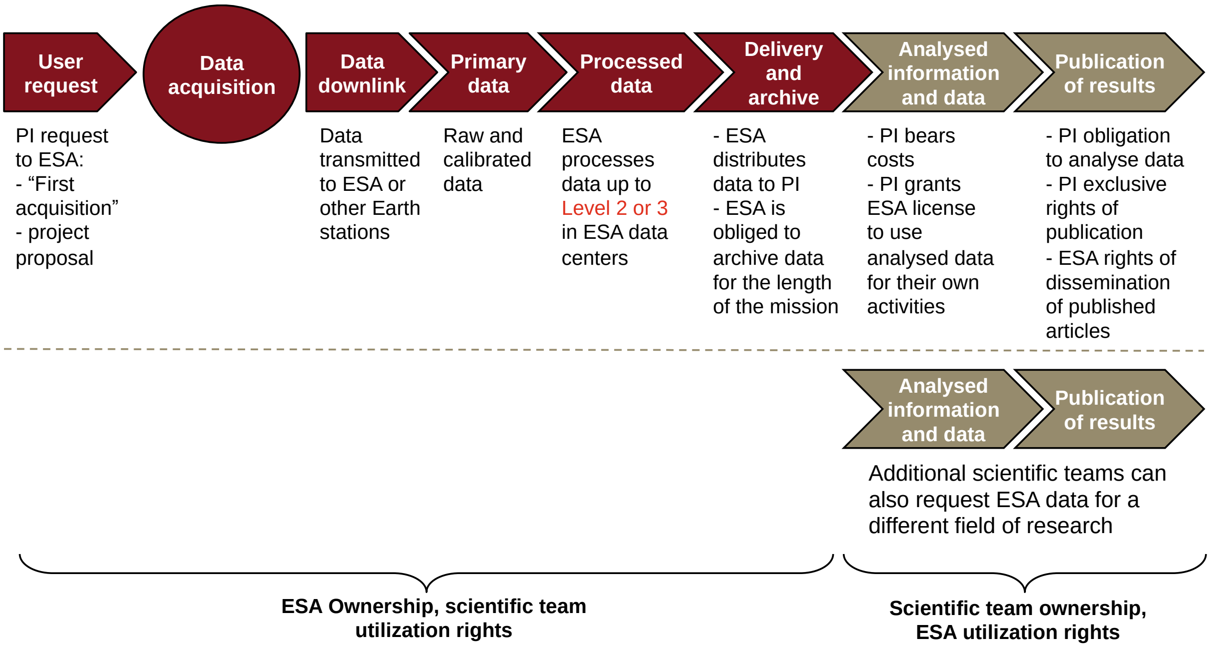


Figure 32: Data ownership[66]

In any case ESA has the copyrights on the scientific publications generated by the PI; i.e., after the PI publishes the mission results for the first time, ESA can re-distribute the publications to third parties with no limitations.

The PIs accessing near-real-time data are granted with an exclusive data access to the data for a period of time of 1 to 2 years. This is usually deemed sufficient to perform the data analysis and produce scientific publications.

For past mission the access was literally exclusive, i.e. the PI requesting the observation time was the only one accessing the data. As described in the picture above, ESA is now adopting a new approach providing more flexibility for third parties PI to access the same processed data. Once processed the data are stored into the PAC archive and are accessible by third parties PI provided that the scope of the intended scientific investigation is different from the one of the PI requesting the original tasking of the satellites.

At the end of the exclusive data access period, the data are made available to the wider scientific community via the PAC repository and the centralised ESA Earth Observation web-portal. According to the specific mission, different types of access are usually implemented:

- **Open:** No restrictions, data is available anonymously.
- **Open-SR:** Open-Simple Registration. Simple registration (access granted in less than 2 days) is required to access data on the Internet.

- **Open-AP:** Open-Advanced Protocol. No specific restrictions on data access, but the approval process may require human approval and justification may be requested. Data kept in non-public archives which can be requested free of charge.
- **Restricted:** Data tightly controlled and with specific conditions for access. This includes all commercial data that requires payment for services or data that is limited to specific user groups.

Different type of access might apply to the same EO missions. More specifically the access to the data generated by the different instruments integrated in a single mission might be different according to the mission characteristics and objective. As summarised in the **Figure 33**, below, ESA implements the open access with simple registration of the user for the totality of its EO instruments.

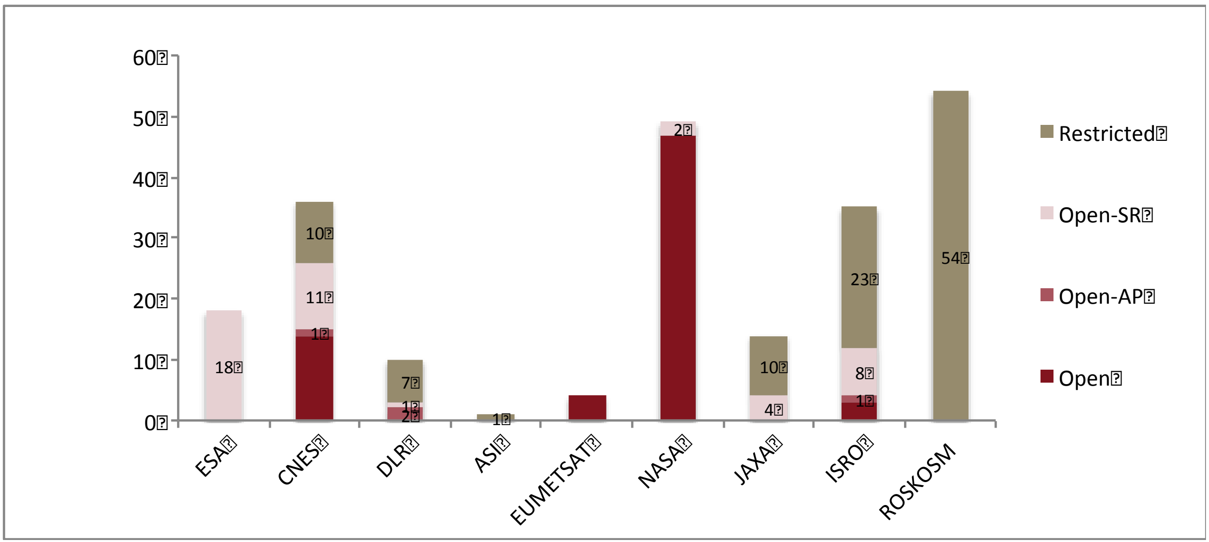


Figure 33: Number of EO instruments with different data access policies from the top 9 agencies in the world [66]

EU Member States tend to use different data access. This is mostly due to the presence of Very High resolution sensor with civil and military dual-use applications. Data for military application are always restricted, while for high resolution data for civil application the advanced protocol is usually requested.

The latter, indeed, are often distributed by commercial companies under and exclusive distribution license. As described in **section 3.2.4**, the institutions implementing the mission often establish mechanisms to grant access to part of these data for scientific exploitation. This is often realised with the implementation of the advanced protocol access scheme.

3.2.4.2 Space Science

The approach adopted by Space science mission is very similar to the one described in the previous section for the Earth Observation. PI requesting observation times are usually granted with an exclusive data aces period of about 1 to 2 years. After the end of this period the data are made available with an open access policy to all requesting users.

In the majority of the cases access to archived data requires a simple user registration process. This is mainly to avoid misuse of the data and unauthorised mirroring of entire mission archives.

3.2.4.3 Microgravity

Microgravity presents a different situation. During the study and the stakeholder consultation, different version of the data policy principle applicable to the microgravity mission have been recorded and it difficult to consolidate them into a single, coherent story.

The near-real-time data are usually accessed by a single PI, i.e. the PI defining the mission and experiment scientific objective. After the end of the mission, data are archived in the USOC and theoretically available to the scientific community for a period of 10 year.

In reality this is not applied as no mechanism has been established to provide data access to the wider scientific community. In most cases a direct request is made to the mission PI that, with the agreement of ESA, might agree to it. Nevertheless the design of the ESA ground segment supporting microgravity mission includes industry proprietary systems and the access to archived data can only be realised by means of specific user terminals.

In the majority of the cases data are distributed directly between scientists by means of email exchange, ftp download from university and research centres server or even via usb memory sticks.

3.3 Conclusions

This chapter has introduced the last two concepts needed to define globally define and analyse the issue of the exploitation of the scientific data: the data exploitation roadmap and the data exploitation framework.

The data exploitation roadmap shows that maximised exploitation of scientific data passes through two intermediate milestones: the near-real time exploitation of the data generated during the missions, and the long term preservation of the data. While the data exploitation frameworks consolidate the awareness, exploitation funding, preservation and data policy as the elements with the greatest influence on the implementation of the roadmap.

The *status-quo* in Europe of the four elements of the roadmap shows three different situations for the three domains of interest for the study:

- Earth observation is the domain with the domain were the problems related to the data exploitation and the data preservation have been faced first, were the biggest volume of data is generated and, with the new upcoming missions the highest data production rate is observed. The major issues are registered in the data preservation and, more specifically, in the recovery of old mission archive and in the lining-up to the LTDP guidelines.
- For Space Science the total amount of data generated by the space missions is much lower than the one generated by EO missions and the most of the old mission archives are already available at ESA and an open access policy is consistently applied to the Space Science mission data. Nevertheless stakeholder interviews highlighted that, despite the open data access approach, the underutilisation of the data collected by Space Science mission is still an issue, mainly related to the fact that the scientific exploitation of the data is not in the mandate of many space mission funding institutions.
- Microgravity domain presents the less advanced status of development of the data exploitation roadmap. Despite an interest of the scientific community in accessing experimental data generated on all types of available platforms. No active preservation activities have been implemented and also different aspects of the data policy require clarifications.

3.4 Notes and references

[43] http://www.nature.com/nature/journal/v455/n7209/full/455001a.html

[44] Stakeholder interview, Italian National Institute of Astrophysics, G. Micela, Teleconference, 26/05/2015

[45] Stakeholder interview, ESA-ESTEC/Science, M. Kessler, F. Favata, Noordwijk, 06/05/2015

[46] Stakeholder interview, ESA-ESRIN, M. Albani, Webinar, 09/06/2015

[47] Strategy&, Web-Survey addressing European Scientific Community

[48] 15 George Helou IPAC 2013 Lessons Learned! From Past and Current ESA-NASA Partnerships November 2013, IPAC, Astronomy & Astrophysics Advisory Committee, Caltech 1

[49] 38 STSC Hubble Space Telescope Call for Proposals for Cycle 23 - Policies, Procedures & Phase I Proposal Instructions Call for proposal January 2015 1

[50] ASI final presentation, "Supporto all'analisi dati per l'astrofisica delle alte energie e la cosmologia delle strutture formate", 17/01/2012.

[51] Arnold H. Rots, Sherry L. Winkelman, Glenn E. Becker; 2012; Chandra Publication Statistics, CXC / Smithsonian Astrophysical Observatory

[52] http://image.gsfc.nasa.gov/poetry/ask/q525.html

[53] ESA/ESRIN, "Monitoring the South Atlantic Anomaly using ATSR historical data sets" or ... why all data should be preserved! Stefano Casadio / Olivier Arino. 07/11/2013

[54] There is no generic ESA data policy for scientific missions. So far data policies have been specifically defined for each ESA mission.

[55] APA/APARSEN Sustainability Brochure – Securing the value of our digital asset. Dr David Giaretta.

[56] http://vds.cnes.fr/sipad_ng/index.html

[57] http://www.scidip-es.eu

[58] 2014 Conference on Big Data from Space BiDS'14, 12-14/11/2014

[59] European framework for the long-term preservation of space data.

[60] European LTDP Common Guidelines V2.0. Long Term Preservation of Earth Observation Space Data. GSCB. 30/06/2012

[61] European Earth Observation Space Data Set V2.0. Long Term Preservation of Earth Observation Space Data. GSCB. 31/01/2013

[62] Earth Observation preserved Data Set Content. LTDP/PDSC. V4.0. GSCB. 25/07/2012

[63] Stakeholder interview with CNES, Geneviève Gargir, Thierry Levoir, Steven Hosford, CNES, Toulouse, 28/05/2015

[64] Stakeholder interview with CDTI, Cristina Garrido Gonzalo, Pillar Román, Emilio Vez Rodríguez, Madrid, 26/05/2015

[65] Stakeholder interview with DLR, Claudia Lindeberg, Wolfgang Frings, Godela Rosner, Bonn, 02/06/2015

[66] Stakeholder interview with ESA Earth Observation Organisation, G Kohlhammer, ESA-HQ, Paris, 22/05/2015

[67] Long-term strategy for ESAC science archives. Christophe Arviset, Danny Lennon, Fred Jansen, Martin, Kessler, Nick Hanowski, Pedro Osuna. Ensuring the long-term preservation and value adding to scientific and technical data 4-6 November 2013, ESA-ESRIN, Frascati (Rome), Italy.

[68] Stakeholder interview with Telespazio, G Carotenuto, Telespazio-Napoli, 28/04/2015

[69] Committee on Earth Observation Satellites (CEOS)

4 DATA EXPLOITATION INITIATIVES

This section will provide an assessment of a targeted set of case studies regarding European and non-European initiatives aimed to increase or enhance scientific exploitation of space data and will analyse lessons learnt through targeted case studies.

4.1 International benchmarking

With regards to international initiatives, the **benchmarking** has included activities undertaken by space power fairing countries like USA, People Republic of China, Russia, Japan, India, and Canada. As summarised in Figure 34, below, these countries plus the European Union, provide for about 98% of the global space expenditure. Therefore an analysis of the initiatives aimed to increase or enhance scientific exploitation of space data in these countries provides a solid global picture.

The present section focuses on the analysis of activities in countries outside the European Union, while section 4.2 focuses on initiatives within the European Union.

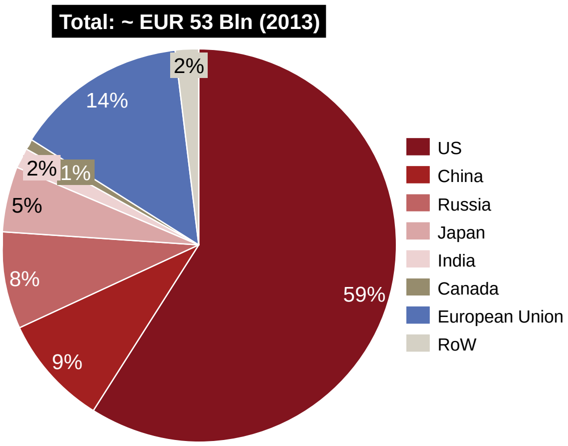


Figure 34: Civil space budget distribution (2013) [70]

As discussed in previous chapter, in section 3.2.2, direct funding of data exploitation is one of the four pillars of the data exploitation framework and has relevant influence on the exploitation of the near-real-time and long term preserved data.

The situation in Europe with respect to the direct funding of data exploitation has been extensively discussed in chapter 3. As for the main space faring countries outside Europe, the different space agencies implement different approaches summarised in the table below.

In the **United States**, NASA directly fund space data exploitation with the Research Opportunities in Space and Earth Science programme. ROSES solicits basic and applied research in support of NASA's Science Mission Directorate (SMD). Awards range from under EUR 90K per year for focused, limited efforts (e.g., data analysis) to more than EUR 1 Mln per year for extensive activities (e.g., development of science experiment hardware).

In **Japan**, the national space agency JAXA has established two different research institute dedicated to space research and exploitation of data. The **Institute of space and astronautical research (ISAS)**, host two main data archives:

- the Data Archives and Transmission System(DARTS) collects data retrieved by Japanese scientific satellites such as SUZAKU and HINODE covering a variety of investigation fields including astronomy, solar physics and solar-terrestrial physics.
- The HAYABUSA Project Science Data Archive, collects asteroids exploration data collected by the HAYABUSA space programme.

In addition to executing research activities with internal staff, the ISAS also offers external investigators with grants and fellowship to access the data and execute research activities.

The **Earth observation research centre (EORC)** processes and preserves Earth Observation data generated by Japanese and international missions. EORC also promotes research and application of satellite data in the fields of meteorology, control of forestry and fisheries resources, disaster prevention and national land use, and global environmental changes.

Country	Space Agency	Acronym	Exploitation funding (Y/N)	Funding mechanism	Funding (EUR/Year)
US	National Aeronautics and Space Administration	NASA	Y	ROSES : research opportunities in space and Earth science	EUR 450 Mln/Year
China	China National Space Administration	CNSA	N	N/A	N/A
Russia	Russian Federal Space Agency	RFSA	N	N/A	N/A
Japan	Japan Aerospace Exploration Agency	JAXA	Y	ISAS : Institute of space and astronautical research EORC : Earth observation research centre	Not available
India	Indian Space Research Organisation	ISRO	Y	RESPOND : sponsored research	EUR 3,7 Mln/Year
Canada	Canadian Space Agency	CSA	Y	Class Grant and Contribution Program	EUR 6,8 Mln/Year

Table 5: data exploitation funding outside the European Union [71]

In **India,** the Indian Space Research Organisation (ISRO) promotes the utilisation of space data with the **sponsored research (RESPOND)** programme. The programme provides for research grants in the fields of space science, space technologies and space applications.

In **Canada** the Canadian Space Agency (CSA), implements the **Class Grant and Contribution Program** to support research, awareness and learning in Space Science and Technology. The programme covers both Research, and Awareness & Learning initiative and aims to support knowledge development and innovation in CSA areas of priority while increasing the awareness and participation of Canadians in space-related disciplines and activities. The agency issues announcement of opportunities covering different topics on a yearly basis.

Space agencies in **China** and **Russia**, respectively the China National Space Administration (CNSA) and the Russian Federal Space Agency (RFSA), implements an approach similar to the ESA's one. The Agencies are responsible for the establishment of the space infrastructure and the provision of the data, but not for the direct funding of data exploitation.

The CNSA supports the **Asian Pacific Space Cooperation (APSCO)** platform. This is an international platform to share data between countries of the Asian-Pacific region[7] and make them available for scientific and application driven exploitation.

4.1.1 Dataset analysis

When analysing the potential impact of space data exploitation projects, flagship initiatives and projects implemented or supported in these countries form a valuable source of input. By characterizing their type of efforts and giving an overview of selected case studies, this section will generate a set of overall 'lessons learnt' from activities in the international landscape.

Five domains have been benchmarked: Earth Observation, Astrophysics/Fundamental Physics, Planetary Sciences, Heliophysics/Space Weather and ISS/Space Exploration. A dataset of **38 global initiatives and projects** has been analysed. As the figure below shows, out of these initiatives the majority (38%) comes from the US, followed by Japan (19%) and China (14%)...

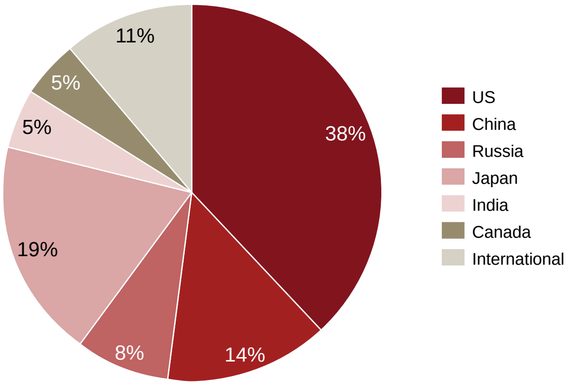


Figure 35: Geographic distribution of the dataset of dataset global initiatives aimed to stimulate space data exploitation

Within this sample, three overall categories of global initiatives and projects have been identified: a) data dissemination portals/archives & standardisation activities, b) international cooperation projects and c) awareness raising activities.

---

[7] Bangladesh, China, Iran, Mongolia, Pakistan, Peru, Thailand, Turkey

4.1.2 A. Data Portals/Archives and Standardisation Activities

A wide range of data dissemination portals or online archives are available on an international level. These data portals provide online access to scientific data sets, either directly or through providing access to multiple other data portals. Overall, this category includes the following type of projects:

- **Thematic** data archive/portals
- **Mission-specific** data archives/portals
- **Overarching** data archives/portals, providing single-point access to multiple databases
- Interactive databases with virtual labs and analytical tools
- Standardisation initiatives to promote interoperability

CASE STUDY: **Data Archives & Transmission System (DARTS)**

Type: Database, Virtual Lab	**Domain**: Space
Status: Active (2002-2015)	Country: Japan
Organisation: JAXA/Center for Science-satellite Operation and Data Archive (C-SODA	
Topic(s): Astrophysics, Heliophysics, Space Weather, Planetary Sciences, ISS/Human Exploration	

Project Description

DARTS primarily archives high-level data products obtained by JAXA's space science missions in Astrophysics (X-rays, radio, infrared), Heliophysics, Solar-terrestrial physics, and Lunar and Planetary Sciences. In addition, it archives related space science data products obtained by other domestic or foreign institutes, and provides data services to facilitate use of these data. Furthermore, DARTS provides 'Virtual Labs' to display collected science data in an orderly and interactive fashion. DARTS services are free of charge for scientific and/or educational purposes and personal use. Its main objectives include:

- (i) Enable access to space science missions' data of JAXA
- (ii) Stimulate interdisciplinary research
- (iii) Provide services to facilitate the use of space science data

Project outcome

Overall, DARTS shows that an interdisciplinary database can be effective at achieving the following goals:

Foster **cross-fertilisation** through a multidisciplinary database to

- Provide **interactive presentations** of science data to offer comprehensive overviews
- Stimulate collaboration with international partners (e.g. NASA)
- Provide **single-point access** to multiple databases

4.1.3 B. International Cooperation Projects

International cooperation projects characterise those projects where international partners cooperate with the goal of facilitating space data sharing or optimizing joint research output from existing space data. This includes projects and initiatives such as:

- **Joint research activities** to improve scientific output of existing scientific data
- **Coordination activities** between specialised institutions
- **Working groups** to foster international discussion between scientific communities
- Cross-domain cooperation to optimise thematic data output

CASE STUDY: Sentinel Asia	
Type: International Cooperation	**Domain**: Space
Status: Active (2005-2015)	**Country**: International (Asia-Pacific)
Organisation: Asia-Pacific space agencies & disaster management agencies	
Topic(s): Earth Observation	

Description

The Sentinel Asia initiative is an international collaboration among space agencies, disaster management agencies, and international agencies for applying remote sensing and Web-GIS technologies to support disaster management in the Asia-Pacific region. Sentinel Asia is promoted by cooperating partners in the space community such as APRSAF, and the international community—including UN ESCAP, UN OOSA, ASEAN, and the Asian Institute of Technology (AIT). A step-by-step approach was adopted for the implementation of Sentinel Asia as follows:

- **Step 1 (2006-2007):** Implementation of the data dissemination system as a pilot project, to form the backbone of the Sentinel Asia and showcase the value and impact of the technology using standard Internet dissemination systems.
- **Step 2 (2008-2012):** the expansion of the system with additional member countries and their agencies, and the expansion of the dissemination backbone with new satellite communication systems, such as the wideband Internetworking Engineering Test and Demonstration Satellite (WINDS; JAXA).
- **Step 3 (2013-present):** Establishment of a comprehensive, operational and enduring disaster management support system in the Asia-Pacific region

Project outcome

- Improved safety in society through **enhanced EO-data dissemination**
- **Increased cooperation** between relevant institutions
- Successful **integration of existing space infrastructure** to provide data in thematic project (utilizing many and varied satellites, such as earth observation, communication and navigation satellites)
- **Optimised data dissemination** and **faster delivery** of disaster information products to end-users through **pilot projects**
- Working groups to discuss and understand data dissemination needs

4.1.4 C. Awareness raising & stakeholder engagement initiatives

In addition to data dissemination portals and international cooperation projects, a set of varied international initiatives to improve space data exploitation has been found. Examples include:

- Crowd sourcing platforms

- International competitions to promote accessibility to existing space data
- Workshops to foster cooperation between researchers and end-users
- Engagement activities with user community to translate user requirements

CASE STUDY: Zooniverse

Type: Stakeholder engagement	**Domain**: Space
Status: Active (2007-2015)	**Country**: International (HQ’s in UK & US)
Organisation: Citizen Science Alliance (CSA)	
Topic(s): Astrophysics/fundamental physics, Heliophysics/Space Weather, Planetary Sciences	

Description

Zooinverse is a citizen science web portal operated by the Citizen Science Alliance (CSA). The portal collects a large number of so-called ‘citizen science’ projects, which aim at involving the general public in using raw scientific data. The project grew from Galaxy Zoo, which successfully engaged citizens to assist in the morphological classification of a large number of galaxies through a web-based portal. Galaxy Zoo was important because not only was it incredibly popular, but it produced many unique scientific results, ranging from individual, serendipitous discoveries to those using classifications that depend on the input of everyone who has visited the site.

Today, the Zooinverse community counts over 1 million registered users and over twenty projects, many of which are located in the field of Space Sciences. Examples include successful projects such as Moon Zoo, Solar Stormwatch, Planet Hunters and Asteroid Zoo.

Project outcome

- Active involvement of general public in translating raw science data
- Generating **continuous stream of new data** for scientific research papers
- Awareness raising on available datasets through citizen science
- Innovative contribution to interactive big data handling

Data dissemination trends in the global landscape denote:

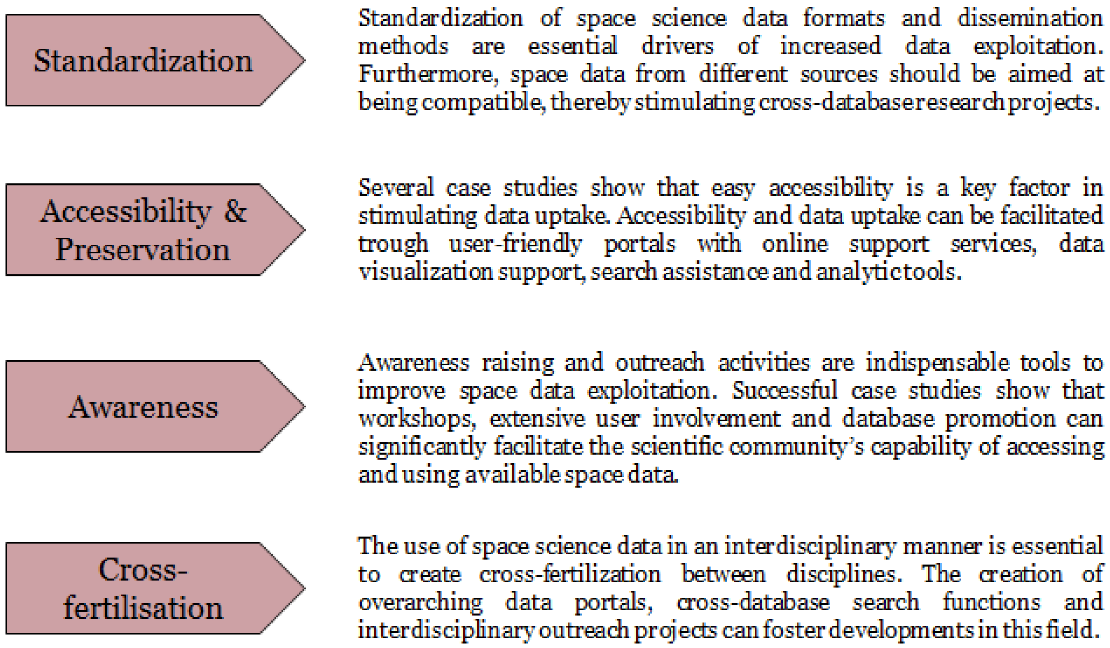
Increased efforts of upcoming spacefaring nations to facilitate space data dissemination among stakeholder community

Introduction of **interactive online platforms** and ‘virtual labs’, allowing the scientific community to interact and collectively foster new research

Increasingly active involvement of scientific and user community trough workshops, crowd-sourcing projects and competition formats

Use of **overarching portals** to facilitate access to multiple datasets and interdisciplinary search functions and associated **standardisation initiatives** to promote interoperability

Conclusions from global initiatives to improve (space) data dissemination can be summarised in four main buckets:



Standardization	Standardization of space science data formats and dissemination methods are essential drivers of increased data exploitation. Furthermore, space data from different sources should be aimed at being compatible, thereby stimulating cross-database research projects.
Accessibility & Preservation	Several case studies show that easy accessibility is a key factor in stimulating data uptake. Accessibility and data uptake can be facilitated trough user-friendly portals with online support services, data visualization support, search assistance and analytic tools.
Awareness	Awareness raising and outreach activities are indispensable tools to improve space data exploitation. Successful case studies show that workshops, extensive user involvement and database promotion can significantly facilitate the scientific community's capability of accessing and using available space data.
Cross-fertilisation	The use of space science data in an interdisciplinary manner is essential to create cross-fertilization between disciplines. The creation of overarching data portals, cross-database search functions and interdisciplinary outreach projects can foster developments in this field.

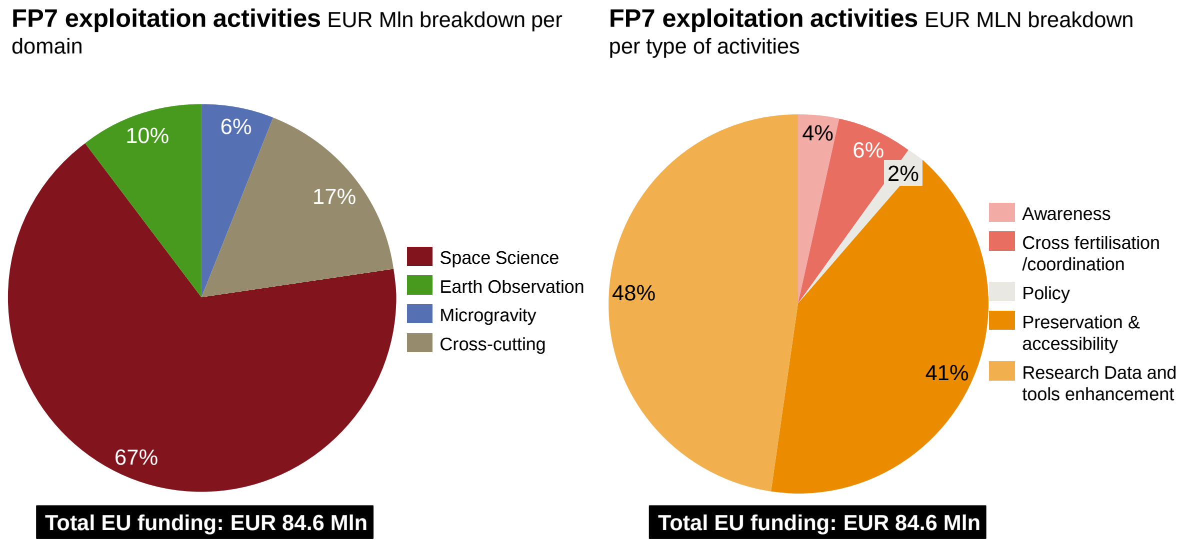
4.2 European initiatives

This section aims to assess how the objective of increasing scientific exploitation of space data has been addressed via past and current initiatives in Europe. In addition, it presents a comprehensive catalogue of relevant EU initiatives, characterized in the terms and definitions of this study. Examples of *success stories* attributable to these initiatives have been selected as case studies.

4.2.1 FP7, space initiatives

Throughout the Seventh Framework Programme (FP7), several initiatives focused on improving space data exploitation have been executed. These include a broad set of projects, ranging over different domains and different focus of the exploitation activities, as shown in the chart below. In total, under FP7 (over the 2007-2013 period), the European Commission has invested approximately EUR 84 Mln across 35 projects to increase space data exploitation

- Average project duration: **35 months**
- Average total costs per project: **EUR 3.0 Mln**
- Average total EU funding per project**: EUR 2.3 Mln**
- Funding schemes:
- **Collaborative Project (CP)**[7] - 88.6%
- **Coordination & Support Action (CSA)**[8]– 11.4%



FP7 exploitation activities EUR Mln breakdown per domain

Total EU funding: EUR 84.6 Mln

FP7 exploitation activities EUR MLN breakdown per type of activities

Total EU funding: EUR 84.6 Mln

Figure 36: Thematic distribution of FP7-Space projects targeting space data exploitation

The vast majority of the FP7 space data exploitation activities concentrated in the space science domain and related to preservation and accessibility tasks (such as systematic collection and categorisation of data of past space missions/ creation of databases, tools to create open access to scientific data ,etc.), and to a great extent also aimed at enhancing existing research tools (e.g. combining space data with ground-based measurements into dynamic models to more sophisticated scientific observations/predictions, development of simulators, up to broader facilities like in ULISSE aimed at supporting the operations for scientific experiments on board the International Space Station).

Overall, most projects are aimed to achieve objectives belonging to one or more of the following categories:

- Creation of a **new database** with science data from past missions
- Creation of an **overarching database/portal**, providing single-point access
- **Cross-fertilisation databases** with data from different scientific domains
- Provision of **virtual labs and analytic tools** to foster scientific return of existing data
- **Awareness raising** activities and engagement with scientific user community
- Standardisation and interoperability improvement initiatives
- Creation of **networks of stakeholders** to optimize cooperation amongst them

Domain	Avg. duration (months)	EU Contribution (EUR Mln)	Estimated total cost (EUR Mln)	Total n. of projects per domain	% EU funding by domain	Activity ratio based on the nb of projects
Earth Observation	36	8.7	10.5	2	10%	100% preservation and accessibility
Space science	41	56.5	74.2	26	67%	61% research data and tools enhancement, 31% preservation and accessibility, 4% awareness, 4% cross-fertilisation
ISS and microgravity	27	5.2	7.1	2	6%	50% research data and tools enhancement, 50% awareness
Cross domain	35	14.1	17.3	5	17%	40% preservation and accessibility, 40% research data and tools enhancement, 20% policy
TOTAL	35	84.6	109.1	35	100%	

Table 6: European Commission investments in FP7 space data exploitation projects per domain

In particular (as shown in the table below), out of the analysed sample, 50% are mainly focused on preservation and accessibility tasks and 45% on research data and tools enhancement (mainly associated to space science domain e.g. to provide data and develop new tools to further enhance our understanding of the solar system, etc.), while only a limited percentage has been allocated to awareness and cross fertilisation.

Activity	Avg. duration (months)	EU Contribution (EUR Mln)	Estimated total cost (EUR Mln)	Total nb of projects per domain	% EU funding by domain	Domain ratio based on the nb of projects
Awareness	27	2.2	2.9	2	3%	50% ISS, 50% space science
Preservation and accessibility	37	42.1	53.0	14	50%	65% ISS, 21% cross, 14% EO
Research data and tools enhancement	45	38.3	50.1	16	45%	94% space science, 6% ISS and microgravity
Policy and cross-fertilisation	32	1.9	2.5	3	2%	67% cross, 33% space science
TOTAL	35	84.6	109.1	35	100%	

Table 7: European Commission investments in FP7 space data exploitation projects per activity

In terms of consortia running the above projects, all of them are led by either universities or research centres (about 46% each of the total sample), while other primes include either industry players or consultancy companies. In terms of geographic distribution, the consortia's leading partners are mainly based in the UK (29%), Italy (20%), France and Germany (9% each), and Austria, Greece, Spain (each 6%) along with ESA (6%) leading two projects as well and the remaining other 11% spread in different countries.

In order to illustrate the type of projects and their impact, a **catalogue with the most relevant case studies** across the different scientific domains has been constructed. To this end, a variation in scientific domain and type of projects has been taken into account in the following sections.

4.2.1.1 A. Earth Observation

CASE STUDY: MEDEO (FP7)	
Full title: Methods and Tools for Dual Access to the EO Databases in the EU and Russia	
Starting date: 01-01-2011	**Duration**: 24 months
Funding scheme: CP-FP-SICA	**Call**: FP7-SPACE-2010-1
Estimated total costs: EUR 619,468	EU contribution: EUR 499,436

Project Description

The market of EO services is rapidly growing and becoming one of the main driving forces of innovative development both in Europe and Russia. At the same time, still the exchange of EO data available in both regions is a technical challenge. MEDEO aimed at reducing technical barriers for the joint use of EO data available both in the EU and Russia. In particularly, the project aimed to:

- structure and make publically available internal data format used in the Resurs satellite system,
- develop mathematical methods for data conversion and several software toolboxes enabling independent EO service providers and GIS developers to access to and use both European and Russian EO datasets from a single application,
- develop the web interface for the collections of the Resurs satellite data, thus enabling access to and efficient search for EO images.

The major technical outcomes have been tested and evaluated by integration into the real European and Russian applications and their validation in the context of provisioning real services to potential final users. The consortium has also undertaken steps to make the project results available to the prime dissemination audience - SMEs working in the field of EO services and/or GIS software development

Project Outcome:

- **A structured and web-accessible description of data format** allowing any third party to develop the own software tools for the Resurs data access and transformation;
- A set of **data converters** allowing easy transformation of the satellite images into the **commonly accepted standards** (Geotiff and Image) and back. The project also provides the guidance on how **easily integrate data conversion** in new Copernicus applications;
- **An easy-to-use web-based interface** to the dataset allowing any potential user willing to acquire Resurs satellite images to browse the collection catalogue, make orders and receive the needed data with or **without conversion**
- **Increased dialogue and international cooperation** between researchers and industry working in this field

4.2.1.2 B. Astrophysics/Fundamental Physics

CASE STUDY: GENIUS (FP7)	
Full title: Gaia European Network for Improved data User Services	
Starting date: 20-09-2013	**Duration**: 42 months (on-going)
Funding scheme: CP-FP	**Call**: FP7-SPACE-2013-1
Estimated total costs: EUR 3,217,237	EU contribution: EUR 2,493,463

Project Description

The GENIUS project aims to the implement the Gaia data processing, of which the final result will be a catalogue and data archive containing more than one billion objects. The archive system containing the data products will be located at the European Space Astronomy Centre (ESAC) and will serve as the basis for the scientific exploitation of the Gaia data. The design, implementation, and operation of this archive are a task that ESA has opened up to participation from the European scientific community.

GENIUS is aimed at significantly contributing to this development based on the following principles:

- Develop an archive design driven by the needs of the user community;

- Provide exploitation tools to maximize the scientific return;
- Ensure the quality of the archive contents and the interoperability with existing and future astronomical archives (ESAC, ESO, ...);
- Stimulate cooperation with the two other astrometric missions, nanoJASMINE and JASMINE (Japan);
- Facilitate outreach and academic activities to foster the public interest in science in general and astronomy in particular.

Project (Preliminary) Outcome

- **Add value to space missions and earth based observations** by significantly contributing to the effective scientific exploitation of collected data
- Enable space researchers to take full advantage of the potential value of data sets
- Contribute to dissemination of Gaia mission data on a global scale
- **Facilitate access** to, and appropriate use of high-quality Gaia data for those scientists who were/are not part of the team having obtained the space mission data
- Raise the awareness of coordination and synergy efforts among Gaia stakeholders
- Increase international cooperation on Gaia data sets

4.2.1.3 C. Planetary Sciences

CASE STUDY: IMPEx (FP7)	
Full title: Integrated Medium for Planetary Exploration	
Starting date: 01-06-2011	**Duration**: 48 months (on-going)
Funding scheme: CP	**Call**: FP7-SPACE-2010-1
Estimated total costs: EUR 2,564,606	EU contribution: EUR 1,998,719

Project Description

IMPEx aims to create an infrastructure which bridges the gap between spacecraft data bases and the scientific modelling tools, enabling their joint interconnected operation and serving, therefore, better understanding of related physical phenomena. It aims at the creation of an interactive framework where data from planetary missions will be interconnected with numerical models providing a possibility to:

Simulate planetary phenomena and interpret space missions measurements;

- Test models versus experimental data and perform further improvement of models;
- Fill gaps in the measurements by appropriate modelling runs;
- Perform preparation of specific mission operations and solve various technological tasks, including preparation of new missions.

Project (Preliminary) Outcome

- Creation of a **simulation data model**, compatible with the Space Physics Archive Search and Extract (SPASE) effort, which facilitates information retrieval across the space and solar physics data environment

- Design of a **website to promote information exchange**, initiate remote procedures and exploit services within the infrastructure
- Boost of collaboration between modelling experts and space mission data experts, making great strides in interpreting space mission data
- **Encourage joint analysis of space mission data** that will further our understanding of space significantly and lay the groundwork for future missions

The influx of data and resulting analyses will boost the European knowledge economy and European Research Area (ERA)

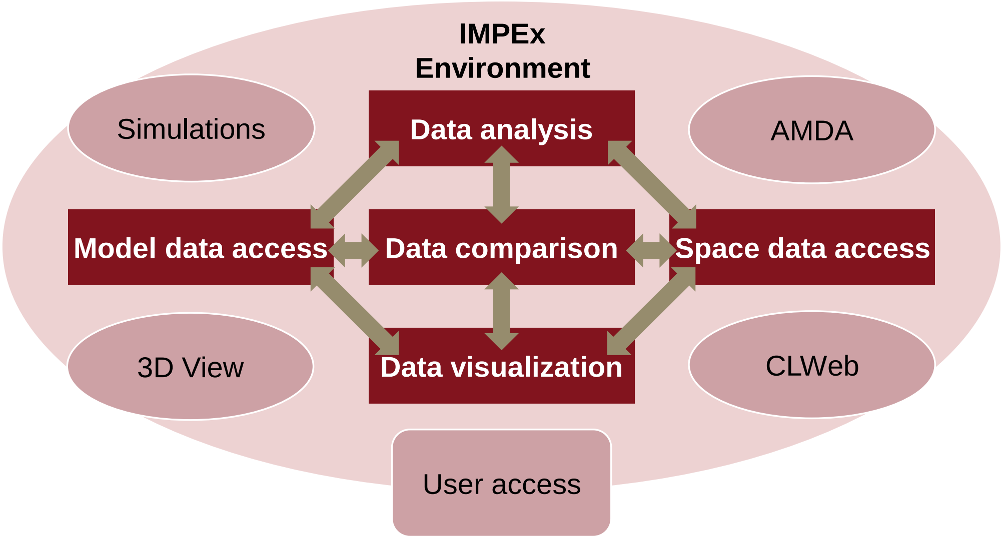


Figure 37: IMPEx: Operational Environment Concept

4.2.1.4 D. Heliophysics and Space Weather

CASE STUDY: SEPServer (FP7)	
Full title: Data Services and Analysis Tools for Solar Energetic Particle Events and Related Electromagnetic Emissions	
Starting date: 01-12-2010	**Duration**: 36 months
Funding scheme: CP	**Call**: FP7-SPACE-2010-1
Estimated total costs: EUR 2,484,126	EU contribution: EUR 1,932,173

Project Description

The main objective of the SEPServer project was to produce a new tool, which greatly facilitates the investigation of solar energetic particles (SEPs) and their origin: a server providing SEP data, related electromagnetic (EM) observations and analysis methods, a comprehensive catalogue of the observed SEP events, and educational/outreach material on solar eruptions.

SEPServer aimed to add value to several space missions and earth-based observations by facilitating the coordinated exploitation of and open access to SEP data and related EM observations, and promoting correct use of these data for the entire space research community. This was aimed to lead to new knowledge on the production and transport of SEPs during solar eruptions and facilitated the development of models for predicting solar radiation storms and calculation of expected fluxes/fluencies of SEPs encountered by spacecraft in the interplanetary medium.

Project Outcome

- Production of **versatile web-based tool for the analysis of SEP events**, the SEPServer, containing a comprehensive set of observations of SEPs in the interplanetary medium and of the related solar EM emissions stored in a relational database
- Creation of an **overarching server with broad functionality** - allows user browsing the metadata, browsing and downloading the event lists, plotting event data with or without a plot template, browsing event catalogues, browsing and downloading simulation datasets, applying/downloading simulation based analysis tools, and browsing the users own environment.
- **Access to state-of-the-art modelling tools** to infer the characteristics of particle emission from the Sun
- **Quality assessments of all the delivered data**, observed and simulated, have been performed during the project, available to the SEPServer user
- Contribution to the **coordination of the exploitation of existing and future data collection** and thereby enhances the possibility to base research on data sets providing comprehensive or full coverage
- **Facilitated access to and appropriate use of data** for those scientists who are not part of the team having obtained it
- **Raising the awareness** of coordination and synergy efforts among stakeholders

4.2.1.5 E. ISS and Human Space Exploration

CASE STUDY: ULISSE (FP7)	
Full title: USOCs Knowledge Integration and dissemination for Space Science Experimentation	
Starting date: 01-01-2009	**Duration**: 36 months
Funding scheme: CP	**Call**: FP7-SPACE-2007-1
Estimated total costs: EUR 6,678,549	EU contribution: EUR 4,858,223

Project Description

A network of centres already operative in space experimentation conceived the ULISSE project. This network was mainly based on the European USOCs including a number of research centres and companies. The USOCs (User Support and Operation Centres) are a network of centres engaged by the European Space Agency (ESA) in various EU countries to support the operations for scientific experiments on board the International Space Station (ISS).

For this purpose, ULISSE intended to pursue the valorisation and exploitation of ISS scientific data and of the already available data from previous space experiments as well as data from other space platforms, increasing the involvement of specialized community and the awareness of general public. ULISSE provided scientific and technical data concerning most scientific disciplines, as Life Sciences including Space Medicine and Exobiology, Biotechnology, Material and Fluid Sciences. The data was integrated with specific services and tools for their exploitation through a middleware platform. The project included dissemination activities: scientific as well as more general publications, public events, educational activities on space research.

In addition, the project team surveyed users and analysed their needs to understand how to improve the use of emerging data. It aimed at documenting and harmonising legal constraints for disseminating ISS data, identifying high-tech tools to manage knowledge efficiently and exploit data effectively.

Project Outcome

- Creation of a **portal** to promote the project and involve the user community
- Development of a set of online services through this portal that supports searching, browsing and managing knowledge bases
- **Promotion of cross-fertilisation** through providing scientific and technical data concerning most scientific disciplines, from space medicine and exobiology to biotechnology and materials sciences
- Creation of a **research project catalogue**, with the definition of a metadata standard to help describe experiments in the catalogue.
- Overall: construction of a successful e-infrastructure for scientific data preservation and exploitation

4.2.1.6 F. Other data exploitation projects

CASE STUDY: SPACE-DATA ROUTERS (FP7)	
Full title: Space-Data Routers For Exploiting Space Data	
Starting date: 01-11-2010	**Duration**: 42 months
Funding scheme: CP	**Call**: FP7-SPACE-2010-1
Estimated total costs: EUR 2,253,972	EU contribution: EUR 1,686,477

Project description & objectives:

According to the project background considerations, currently, Space-Data exploitation faces two major obstacles: Firstly, Space Centres and Academic Institutions have limited access to scientific data since their limited connectivity time via satellites directly confines their scientific capacity. Secondly, Space-Data Collection Centres lack sufficient mechanisms for communicating with interested end-users let alone the lack of mechanisms for data dissemination. The result is frequently quite disappointing: Space data remains stored and unexploited, until it becomes obsolete or useless and consequently is being removed.

Along these lines, the ultimate goal of “Space-Data Routers” was to boost collaboration and competitiveness of European Space Agency, European Space Industry and European Academic Institutions towards an efficient architecture for exploiting space data. The proposed approach relied on space internetworking – and in particular in Delay-Tolerant Networking (DTN), which marks the new era in space communications, unifies space and earth communication infrastructures and delivers a set of tools and protocols for space-data exploitation within a single device: The Space-Data Router.

Project outcome:

- Extension of end-user’s access to data
- Increase of data acquisition frequency
- Elimination of data loss and increase of data volume received
- Increase of **access speed** to deep space data
- Acquisition, management and efficient dissemination of large volumes of data

- Real-time access to data from multiple missions
- **Improved interconnection** between ground segment and space-based assets
-

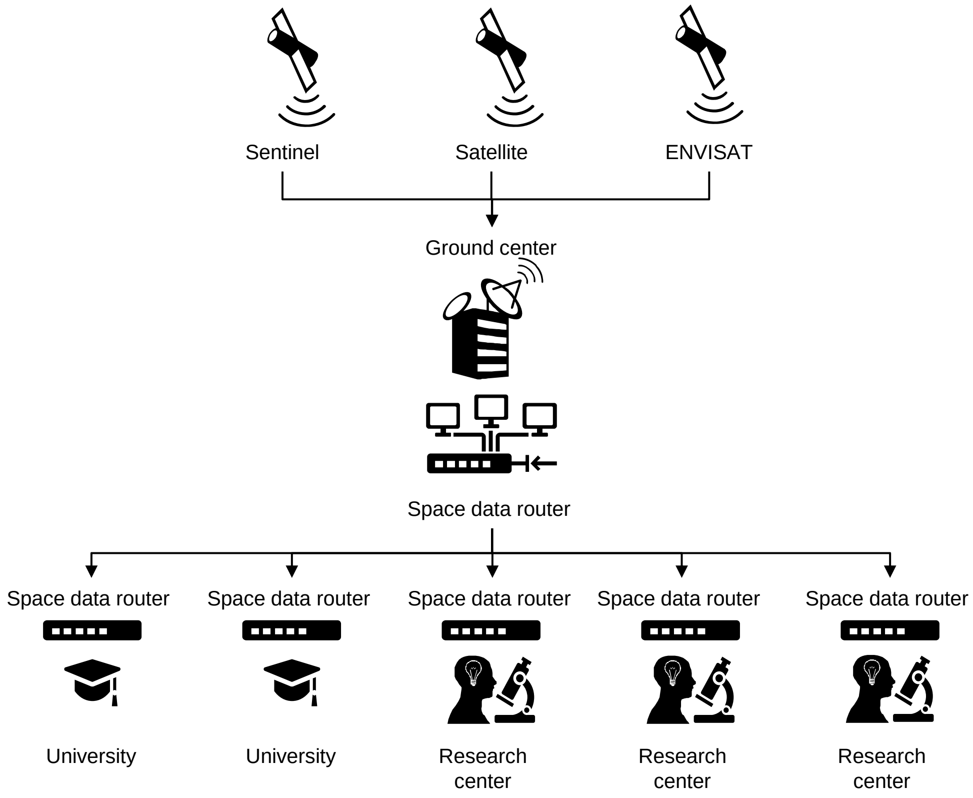


Figure 38: Space Data Routers: Operational Concept

4.2.2 Dedicated survey to gather feedback from past FP7/H2020 space programmes

As described in **section 1.3** a focussed survey has been launched, in the context of the present study in order to gather some primary feedback and information from the past and on-going FP7/H2020 space projects. So far the survey has received 8 responses from the following projects.

#	Project name	Start date	End date	Project budget (EUR Mln)
1	ESPaCE	01/06/2011	31/05/2015	2.0
2	ETAEARTH	01/01/2013	31/12/2017	2.0
3	BIO+SOS, GA 263435	01/12/2010	30/11/2013	2.5
4	EuroVenus	01/10/2013	01/10/2016	2.2
5	STORM	01/01/2013	31/12/2015	26.6
6	GENIUS	01/10/2013	31/03/2017	3.2
7	SENSYF	01/11/2012	31/10/2015	2.5
8	PRoViDE	01/01/2013	31/12/2015	2.5

Table 8: Survey responses from FP7 projects

The majority of the above projects falls into the Space Science domain (representing 23% of the total FP7 projects in the domain) and two from EO (100% representation of the FP7 EO data exploitation projects).

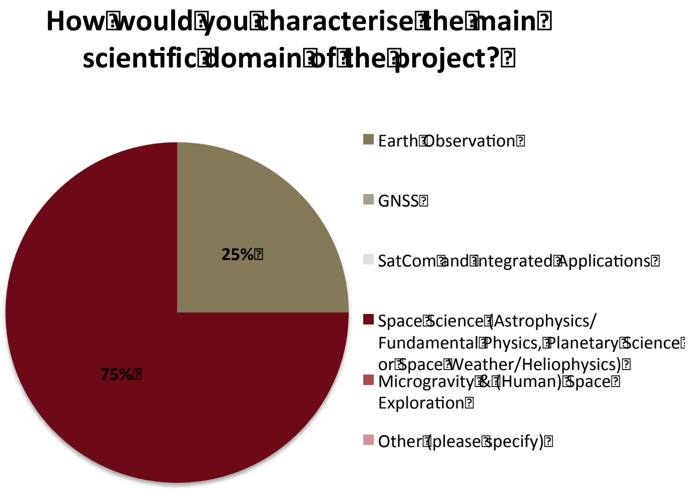


Figure 39: FP7 Survey responses, main scientific domains

Even if these data do not cover the full spectrum of launched EU FP7 projects aimed at exploiting space data, some emerging findings are worthwhile to report at this stage.

4.2.2.1 Key findings from the survey response

As illustrated in the following figure there is a fairly equal distribution among the projects' main objectives, ranging from foster standardisation, to support the long term preservation, increase the awareness, stimulate cross-fertilisation, and, in case of EO, integrate remote sensing/in-situ data).

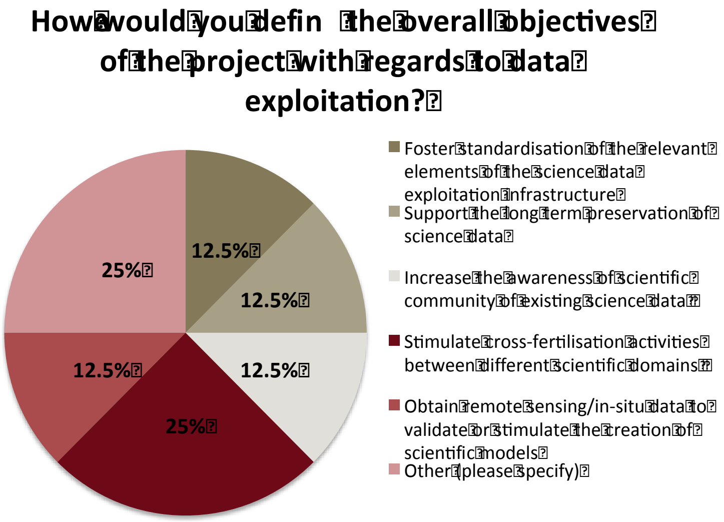


Figure 40: FP7 survey responses, project objectives

Where the specified "other" objectives include (as indicated by the responders):

- Exploiting existing datasets, but also co-ordinating new observations
- Provide improved/advanced tools for data exploitation

Most of the projects in the sample aim at enhancing and better leveraging existing infrastructure.

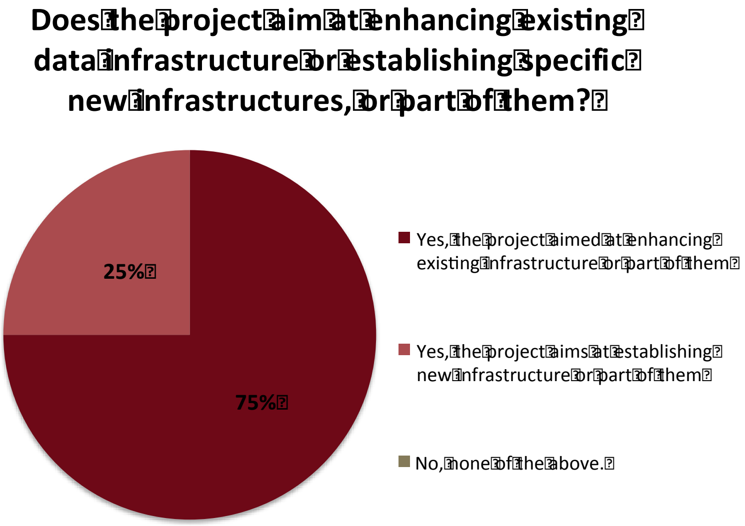


Figure 41: FP7 Survey responses, data infrastructure

In terms of main target scientific communities of the project, these mainly include the research institutes and scientific organisation (over 52% together) and universities. Note: "Other" specified response named "all of them"- referring to the options provided in the legend).

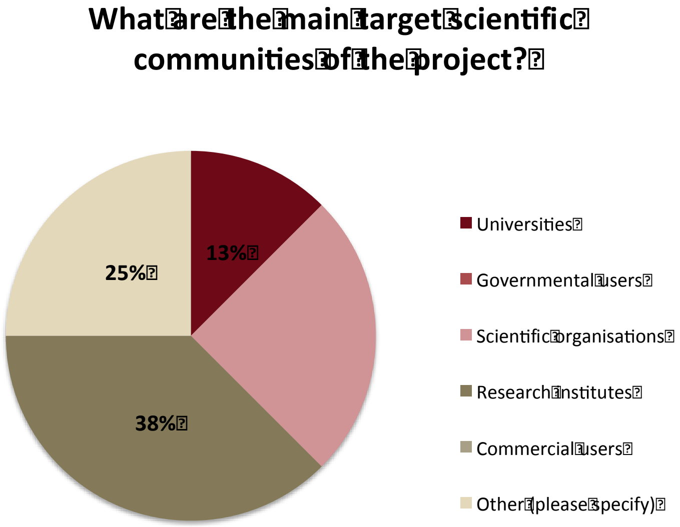


Figure 42: FP7 survey responses, target scientific communities

It is worth noticing that most of the projects actually make us of science data from sources outside the EU, and especially from the USA.

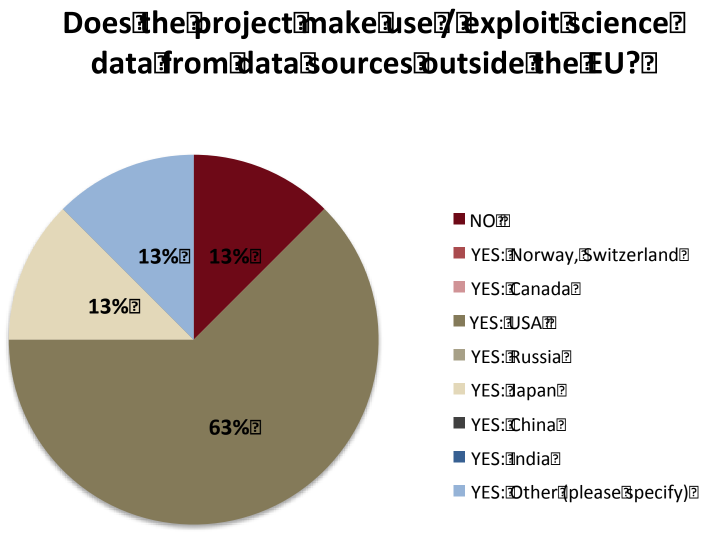


Figure 43: FP7 Survey responses, non-EU data

As far as the applied standards are concerned, as shown in the table below quite a diverse range of standards was used.

Project Ref. nb.	Response text (used standard)
1	SPICE Kernels + VOTable
2	Our project utilizes domain specific standards that is FITS, that apply to both the space and ground-based data.
3	"ESA PSA (PDS-compliant); netCDF;"
4	In general we used data provided by the European Space Agency. The use of the CDF format was very helpful as we could adapt and made versatile our data analysis routines.
5	We adhere to the standards of the International Virtual Observatory, mainly to provide data through its TAP protocol. We also use Java as a standard for software development. Otherwise, the distribution of data will take place through web interfaces that allow download in some common formats TBD.
6	"OGC and OpenSearch for Data Discovery and Data Access NetCDF and SAFE for Data Access as they were the basis for Sentinel data formatting."
7	PDS; OGC

Table 9: FP7 survey responses, standards used

Most of the projects stated that there was no need to develop additional specific standards.

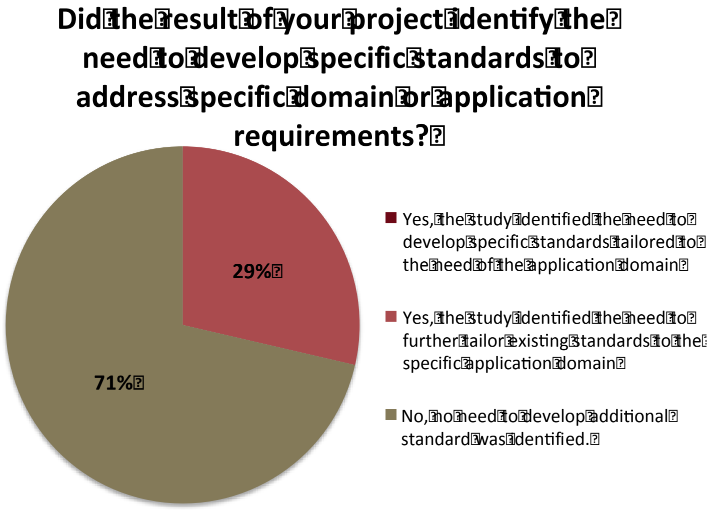


Figure 44: FP7 survey responses, development of specific standards

4.2.3 FP7, non-space initiatives

In addition to the set of initiatives dedicated to Space, FP7 has also generated a considerable amount of projects aimed at improving exploitation for scientific data in general. This includes the creation of new e-infrastructure, awareness raising activities, the creation of new databases and online portals, and so-called of networks of excellence. Two illustrative case studies have been selected to illustrate the type of projects that have been executed in this light and their relevance for space data exploitation pilot projects. Examples of relevant topics within FP7 non-space calls include:

- FP7-INFRASTRUCTURES-2010-2: "Virtual Research Communities"(INFRA-2010-1.2.3.) – Total budget of Euro 23 millions
- FP7-INFRASTRUCTURES-2011-2: "Data infrastructures for e-Science"(INFRA-2011-1.2.2.) – Total budget of Euro 43 millions
- FP7-ICT-2009-6: "Digital libraries and digital preservation" (ICT-2009.4.1) – Total budget of Euro 69 millions

CASE STUDY: APARSEN (FP7)	
Full title: Alliance Permanent Access to the Records of Science in Europe Network	
Starting date: 01-11-2011	**Duration**: 48 months
Funding scheme: NOE	**Call**: FP7-ICT-2009-6
Estimated total costs: EUR 8,212,966	EU contribution: EUR 6,840,000

Project description

APARSEN was a Network of Excellence that brought together an extremely diverse set of practitioner organisations and researchers in order to bring coherence, cohesion and continuity to research into barriers to the long-term accessibility and usability of digital information and data, exploiting diversity by facilitate the building of a long-lived Virtual Centre of Digital Preservation Excellence. The Joint Programme of Activities aimed to cover

- Technical methods for preservation, access and most importantly re-use of data holdings over the whole lifecycle;
- Legal and economic issues including costs and governance issues as well as digital rights
- Outreach within and outside the consortium to help to create a discipline of data curators with appropriate qualifications

Project outcome:

- Development of a Virtual Centre of Digital Preservation Excellence
- Creation of a long-term roadmap to understand issues of interoperability, intelligibility and scalability and determine how data and digital objects can remain useable and understandable in the long-term
- Development of a **methodolog**y for capturing, modelling, managing and exploiting various **interoperability dependencies**
- Development of recommendations and guidelines for ensuring interoperable digital preservation services
- An integrated overview of the options for storage solutions for preservation of digital resources
- **Exemplary business cases**, of economically-sustainable digital preservation initiatives

- Create an overview of standardisation activities and progress
- Creation of training and course materials and raise awareness via dedicated workshops

CASE STUDY: SCIDIP-ES (FP7)	
Full title: Science Data Infrastructure for Preservation – Earth Science	
Starting date: 01-09-2011	**Duration**: 40 months
Funding scheme: CPCSA	**Call**: FP7-INFRASTRUCTURES-2011-2
Estimated total costs: EUR 7,721,601	EU contribution: EUR 6,599,992

Project description

The objectives of SCIDIP-ES were to:

- Help to preserve digitally encoded data
- Make data easier to re-use now and into the future

The overall aim of the initiative was to deliver generic infrastructure services for science data preservation and to build on the experience of the ESA Earth Observation Long Term Data Preservation (LTDP) programme to favour the set-up of a European Framework for the long term preservation of Earth Science (ES) data through the definition of common preservation policies, the harmonisation of metadata and semantics and the deployment of the generic infrastructure services in ES domain.

Preservation systems are mostly hidden to users, but are crucial for their work. Properly preserved data and associated knowledge allows what is “unfamiliar” and “unusable” to become familiar and usable. This is crucial to enhance data interoperability between scientists belonging to different disciplines. Therefore, SCIDIP-ES delivered Data Preservation Services and Toolkits and guidelines for data managers who are willing to create/enhance their preservation systems. After analysing the state of the art, SCIDIP-ES delivered preservation best practices agreed among some of the most important Earth Science European repositories.

Project outcome:

- Development of **e-infrastructure services**, which are generic enough to be used in every preservation environment but at the same specific enough to satisfy Earth Science repositories’ evolution needs
- Definition of a **common preservation policies and the harmonisation** in the ES domain, boosting the development of the Earth Science LTDP framework
- **Facilitating interoperability** among the different actors and **addressing the long term preservation** of data in this challenging and sensitive domain
- Dedicated workshop and awareness raising activities for the Earth Science stakeholder community
- **Interaction with user community** through in-depth surveys and translation into user needs

4.2.4 Horizon 2020, space initiatives

Building upon the efforts of FP7, Horizon 2020 considerably intensifies its efforts towards increased data exploitation from space sources. As established in the Horizon 2020 Work Programme LEIT – Space 2014-2015:

"Exploitation of space science data is being addressed across H2020 on a **recurring basis**, ensuring a **more extensive utilisation of scientific data** originated from European missions and missions with European participation."

Indeed, in order to address the specific challenges of ensuring that Europe's investments made in space infrastructure are exploited to the benefit of citizens, as well as supporting European space science and enhancing Europe's standing as attractive partner for international partnerships in space science, increased efforts in (international) space data exploitation are necessary. Efficient and widespread exploitation of the existing and planned operational European space infrastructure is only possible if further efforts are made for the **processing, archiving, standardised access and dissemination** of space science data. Sustainable availability has also to be coupled with generic **search, data-mining and visualisation techniques** inviting wide data use, also allowing for **standardised and automated approaches**. Equally, space data obtained for specific purposes can subsequently **reveal novel scientific insights** which were not specifically intended or expected at the time of space sensor launch. [9] In this light, selected calls from Horizon 2020 in Space sciences & Microgravity, and Earth Observation are briefly discussed.

4.2.4.1 A. Space science

A number of calls within the 'Competitiveness of the European Space Sector: Technology & Science – 2014-2015' have been addressing space data exploitation in specific areas:

4.2.4.1.1 Space Exploration and Science

- COMPET-08-2014: Science in context: sample curation facility and scientific exploitation of data from Mars missions (total budget of Euro 4 millions)
- COMPET-05-2015: Scientific exploitation of astrophysics, comets, and planetary data (total budget of Euro 6 millions)

Selected projects within this field have been funded either under the RIA (Research & Innovation Action) or the CSA (Coordination & Support Actions) schemes.

As only a limited number of projects within these calls have already been started/ are on-going, make a viable assessment of their average size and impact is not yet feasible but some considerations will be presented in the section related to the conducted survey towards the project representatives. For only illustrative reasons, though, two exemplary case studies, selected among the COMPET-08-2014 call, are shown below.

CASE STUDY: UPWARDS (Horizon 2020)	
Full title: Understanding Planet Mars With Advanced Remote-sensing Datasets and Synergistic Studies	
Starting date: 01-03-2015	**Duration**: 36 months (on-going)
Funding scheme: RIA	**Call**: COMPET-08-2014
Estimated total costs: EUR 2,103 ,594	EU contribution: EUR 2,103,593

Project description

The goals of the UPWARDS project match the topics, challenges and scope of the Compet-8-2014 call. UPWARDS has as its overarching objective the revision and exploitation of data from the European Mars Express (MEx) mission as well as other Martian missions using a synergistic combination of state-of-the-art atmospheric/subsurface models and novel retrieval tools. UPWARDS will (i) address major open science questions which require an integrated understanding of the Mars coupled system from the subsurface to the upper atmosphere; (ii) prepare for ExoMars 2016 Trace Gas Orbiter data analysis and exploitation; and (iii) deliver enhanced scientific context and datasets for ExoMars 2018 Rover operations and future missions.

All topics are addressed by experts in the field, exchanging results and knowledge in a truly synergistic and interdisciplinary collaboration. All topics share a common methodology and work flow: (i) **compilation of new or unexploited data** from MEx; (ii) generation of **added-value products with new/validated tools** developed in the Consortium; (iii) **analysis and combination of the results with state-of-the-art models**. Included is a novel data-assimilation devoted to supply as an end product, the **first of its-kind 4-D (x, y, z, and t) database for ExoMars** and beyond.

CASE STUDY: EURO-CARES (Horizon 2020)

Full title: European Curation of Astromaterials Returned from the Exploration of Space	
Starting date: 01-01-2015	**Duration**: 36 months (ongoing)
Funding scheme: RIA	**Call**: COMPET-08-2014
Estimated total costs: EUR 2,103 ,594	EU contribution: EUR 2,103,593

Project description

Europe has a very strong legacy in the curation and research of precious extra-terrestrial materials. To maintain European leadership and ensure high-level involvement in future SRMs, Euro-CARES will develop a European Sample Curation Facility (ESCF) and aims to:

- Evaluate and critically **assess the state of the art** within Europe and internationally to identify critical requirements for the ESCF
- **Determine and verify European readiness levels** to identify where investment is required and opportunities for European leadership fields related to curating extra-terrestrial samples
- **Engage with scientific, industrial, governmental and public stakeholders** through community workshops, conferences, publications and educational opportunities
- Deliver **recommendations and roadmaps** defining the steps necessary to deliver a ESCF to ensure high-level involvement in future ESA and international SRMs

4.2.4.2 B. Earth Observation

For Earth Observation (Copernicus), it is highlighted in the H2020 Work Programme 2014-2015 that **wide use of collected EO data** has to be achieved at European and global levels, and coordination with mechanisms promoted in the context of the Global Earth Observation System of Systems (**GEOSS**) and the Committee on Earth Observing Satellites (**CEOS**) is to be achieved. For successful exploitation of space borne sensors to take place, it is furthermore necessary to provide access to **easy-to-use, calibrated and validated data products**, taking into account the latest and emerging remote sensing capabilities and the most recent online data manipulation, collaboration, visualisation and sharing technologies. Validation efforts have to provide researchers and users with well-defined uncertainty ranges of space data to make the subsequent usage verifiable and to allow for **cross-sensor or cross-satellite use of data**.

In order to further **enhance scientific, operational and commercial exploitation** of collected EO data, new upstream data products and analysis methods suitable for subsequent integration into applications should be generated within the Horizon 2020 Programme. In this context, a **combination of EO data and data from other sources** (e.g. in-situ sensors, gravity data, magnetic data, and navigation signals) could be broaden the data scope beyond conventional EO images.

Most proposals within the EO calls are expected to have a **significant impact in stimulating wide and further exploitation of the used data**, be it in scientific or commercial use, or operational services. The **results shall be actively disseminated** in the relevant scientific publications, as well as towards potential user communities as appropriate. For operational exploitation, the **needs of the user community** are expected to have been validated in order to ensure a positive impact. [9]

For Earth Observation, projects within the following calls within 'Earth Observation – 2014-2015' have been addressing increased EO data exploitation:

Space enabled applications

- EO-1-2014: New ideas for Earth-relevant space applications (total budget of Euro 4 millions)
- EO-1-2015: Brining EO applications to the market (total budget of Euro 9 millions)

Tools for access to space data

- EO-2-2015: Stimulating wider research use of Copernicus Sentinel Data (total budget of Euro 11 millions)

4.2.5 Horizon 2020, non-space initiatives

Finally, as in FP7, there are also a number of calls within Horizon 2020 which address the general need for scientific data exploitation. Examples of calls and indicative budgets include[8]:

- EINFRA-1-2014: Managing, preserving and computing with big research data (Total budget of Euro 55 millions)
- EINFRA-2-2014: e-Infrastructure for Open Access (Total budget of Euro 13 millions)
- EINFRA-3-2014: Towards global data e-infrastructures – Research Data Alliance (Total budget of Euro 4 millions)
- GARRI-4-2015: Innovative approach to release and disseminate research results and measure their impact (Total budget of Euro 1.5 million)

As with the space-based initiatives, only a limited number of projects within these calls have already commenced, and therefore a conclusive summary of their characterisation is not yet possible. An exemplary case study has been selected.

CASE STUDY: OpenAIRE2020 (Horizon 2020)

Full title: Open Access Infrastructure for Research in Europe 2020

---

[8] Source: H2020 Work Programme 2014-2015

Starting date: 01-01-2015	**Duration**: 42 months
Funding scheme: RIA	**Call**: EINFRA-2-2014
Estimated total costs: EUR 13,132,500	EU contribution: EUR 13,000,000

Project description

OpenAIRE2020 represents a pivotal phase in the long-term effort to implement and strengthen the impact of the Open Access (OA) policies of the European Commission (EC), building on the achievements of the OpenAIRE projects. OpenAIRE2020 will expand and leverage its focus from:

- the agents and resources of scholarly communication to workflows and processes,
- from publications to data, software, and other research outputs, and the links between them,
- strengthen the relationship of European OA infrastructures with other regions of the world, in particular Latin America and the U.S.

Through these efforts OpenAIRE2020 will truly support and accelerate Open Science and Scholarship, of which Open Access is of fundamental importance.

OpenAIRE2020 continues and extends OpenAIRE's scholarly communication infrastructure to manage and monitor the outcomes of the European Commission-funded research. It combines its substantial networking capacities and technical capabilities to deliver a robust infrastructure offering support for the Open Access policies in Horizon 2020, via a range of pan-European outreach activities and a suite of services for key stakeholders.

It provides researcher support and services for the Open Data Pilot and investigates its legal ramifications. The project offers to national funders the ability to implement OpenAIRE services to monitor research output, whilst new impact measures for research are investigated. OpenAIRE2020 engages with innovative publishing and data initiatives via studies and pilots. By liaising with global infrastructures, it ensures international interoperability of repositories and their valuable OA contents.

4.3 Notes and references

[70] Based on data from "OECD (2014), The Space Economy at a Glance 2014, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264217294-en", accessed on September 2015

[71] NASA, CNSA, ISRO, RFSA, CSA, JAXA, Strategy& analysis

5 RECOMMENDATIONS

In this study, we have seen the three domains of interest have unique situations with respect to the exploitation of the scientific data produced by their space missions.

This chapter provides recommendations concerning activities that the European Commission could potentially take to improve the exploitation of the European space missions.

The situation of each domain will first be mapped against the data exploitation roadmap to identify upon which elements of the roadmap and of the data exploitation framework the European Commission could intervene. Recommendations consisting of potential European Commission actions will follow, including rough order of magnitude estimates on project cost and where possible.

The study team considers the Horizon 2020 Space Theme, under the Industrial Leadership pillar, as the main tool available to the European Commission to implement these activities. Nevertheless other elements, under different pillars of the H2020, may also be appropriate implementation vehicles. In the scope of this study the European Commission Copernicus programme is not considered as a tool for the implementation of the European Commission activities for the preservation of scientific space data generated by European Earth Observation missions.

5.1 Earth Observation

The roadmap for the exploitation of the Earth Observation data is the most advanced. This domain has been the first to deal with problems related to the data exploitation and data preservation. The biggest volume of data has been generated by Earth Observation missions, and the data production rates will significantly increase with upcoming missions.

Earth Observation missions are implemented both at international and national levels using all the mission models described in **section 1.6**. The on-going European Commission Copernicus programme will generate an unprecedented volume of data that will be released freely to the users' community – including scientists – a few hours after acquisition. Therefore we can confidently state that the status of implementation of the EO data exploitation roadmap is very advanced.

Nevertheless the roadmap has issues that, if properly addressed, should increase the level of exploitation of the data.

The major issues are concentrated in the area of data preservation, specifically with the recovery of old mission archives and the adherence to the LTDP guidelines. As described earlier in the document, recovery, preserving and valorising data from old archives increase the possibility of exploitation, but this is just due to their new digital availability. The process also allows for data reprocessing to create coherent time series with new algorithms to increase the chances of stimulating new research and exploitation of the data.

With the establishment of the LTDP budget, ESA intends to recover and ensure preservation data stored within old archives for all EO missions in which it had an active role. Nevertheless, a great volume of EO data relevant to Europe, remain stored in old archives of **third party** mission, waiting to be recovered and preserved.

The old archives are at significant risk of being lost for different causes: physical damage of deteriorating storage supports, unavailability of appropriate platforms (e.g. specific tape recorders) to read the data, unavailability of appropriate associated knowledge to correctly interpret the data, *et cetera*. For this last point, it is also important to consider that some data requires the memory of the PI involved in the mission, which is a precious asset that unfortunately deteriorates and ultimately disappears with time.

ESA estimated the total cost for the execution of the all Earth Observation data preservation activities, including third party missions, at EUR 20 Mln annually. This cost includes both the infrastructural costs (CAPEX) for the acquisition of the necessary infrastructure and the operational costs (OPEX) for their operations. These figures are expected to stay constant over the years. The

CAPEX are not expected to grow proportionally to the increasing volume of data to be preserved. However, the OPEX associated with the recovery of old archives are expected to decrease with the progress of the recovery activities and the consequent decrease in number of archives to be recovered.

5.1.1 Support the recovery of old third party EO mission archives

The European Commission could support ESA in the prioritisation of third party mission data archive and in the subsequent the recovery of data with projects targeting a single mission. The final objective of these projects would be to support the establishment of the elements of a data infrastructure that is interoperable with existing data infrastructures and/or integrated with them.

The cost of recovery of data from a single mission depends on the size of the archive and on the status of preservation of the data at the beginning of the recovery activities. It can be estimated in rough order of magnitude at **EUR 1 Mln**, for a total duration of 1 year of activity. [72]

This cost represent a “worst case” assumption where each data recovery project is an isolated activity with no synergies with similar activities. These figures are expected to reduce in time as hardware, software and expertise can be utilised across multiple projects. Costs are therefore expected to reduce and stabilise around **EUR 0,5 Mln** with implementation of multiple data recovery activities.

Similar activities have already been implemented as *Small or medium-scale focused research projects (CP-FP)* or *Small or medium-scale focused research projects for specific cooperation action dedicated to international cooperation (CP-FP-SICA)* under the FP7 programme. An applicable example is the Collaborative Project Methods and Tools for dual access to the EO databases of the EU and Russia (MEDEO) discussed in **chapter 5**.

5.1.2 Support integration of Space and In-Situ data

In addition to the recovery of old archives, and with a lower priority, the European Commission could also support the integration of in-situ sensor data and other Earth observation data into European space data archives.

This action would be aligned with the provisions of the Infrastructure for Spatial Information in the European Community (INSPIRE) regulations supporting the sharing of environmental spatial information among public sector organisations and facilitating public access to spatial information across Europe.

The costs of these actions would depend on the size of the archive and on the status of preservation of the data. They should be similar in rough order of magnitude to the cost for the recovery of old space generated data described in the previous section.

5.2 Space science

As seen in the previous chapter, the priorities of the Space Science domain differ to that of Earth Observation. The total amount of data generated by the Space Science mission is much lower in volume than the amount generated by EO missions. Most of the old mission archives are already available in an accessible form at the ESA/ESAC facility in Madrid.

The ESAC archive covers the vast majority of data generated by European Space Science missions. The data generated by missions of Member States which do not involve ESA are either preserved and made accessible by the cooperating international Institution (mostly NASA) or preserved locally in “thematic” research centres.

An open access policy is consistently applied to the Space Science mission data and other initiatives, such as the European Commission/EU cofounded European Virtual Observatory EURO-VO, facilitate the access to the data and foster the general awareness about the availability of the data.

We have also seen that awareness of the availability of near-real-time data is not a problem. Requests for satellite observation time are usually exceeds the time available by 10 times.

Also for the Space Science we can conclude that the status of implementation of the data exploitation roadmap is very advanced. All elements have already addressed by ESA and by the Member States involved in the mission implementation.

Nevertheless the underutilisation of the data is still perceived as an issue, mainly related to the fact that the scientific exploitation of the data is not in the mandate of many space mission funding institutions, including ESA. As discussed in **section 3.2.2**, where these Institutions have a mandate to fund the exploitation of the data, e.g. NASA, it is accepted that a higher number of scientific publications is produced on the basis of the available data than would occur otherwise.

5.2.1 Fund the exploitation of Space Science data

To address the issue described in the previous section, the European Commission could establish mechanisms to fund the exploitation of the available space science data. The scope of such activity should include both the near-real-time and the long term preserved data. As discussed in **section 3.2.1**, indeed, long term preserved data are largely used to produce scientific publications.

Previous experience, such as the ASI-INAF cooperation described in **section 4.4**, has proven that the funding of multiple research projects with a **12 – 18 month** duration and a maximum cost around **EUR 300 K,** is very effective in aligning the production of scientific publications to levels set by the US.

The ASI-INAF funds were restricted to Italian scientific team, European Commission could consider funding similar activities also to stimulate international cooperation between science team from different Member States. In order to allow for international cooperation the cost for each project should be a minimum of EUR 1 Mln for 12 month activity, increasing according to the project length.

The ERA-Net co-fund and the Frontier Research funded by ERC are seen as potential H2020 tools to implement this type of action. The actions should not be restricted to a single field of research (e.g. Solar observation, Planetary observation, Astronomy), but should rather be open to all space science disciplines and should specifically address the topics of distributing satellite observation time and the long-term plan for archiving data

Support the consolidation of the Virtual Observatory

As described in **section 2.2.5**, representatives of Member States organisation responsible for the scientific exploitation of space data have indicated a national interest in the development and support to national virtual observatory nodes as part of the International Virtual Observatory Association (IVOA).

Virtual observatory are indeed considered a very powerful tool to facilitate access to Space Science data for a very vast community of users going from Research Centre and University departments to individual students.

The European Commission has always been very active in supporting the development of a network of European virtual observatories across its member states. Already under the FP5 programme in 2002 more than **EUR 5 Mln** were allocated in cooperation with other European organisations[9] to demonstrate the feasibility of VO for European Astronomy. Today the European Commission co-funded European Virtual Observatory (EURO-VO) initiative has the objective to develop and operational virtual observatory in Europe providing web-based access to the available astronomical data archives of space and ground-based observatories and other sky survey databases.

The European Commission is suggested to continue supporting the EORO-VO project consolidating and facilitating access to Space Science data to all European users.

---

[9] ESA, the European Southern Observatory (DE), AstroGrid (UK), the CNRS (FR), the University Louis Pasteur in Strasbourg (FR), the Jodrell Bank Observatory of the Victoria University of Manchester (UK).

Support the integration of space and ground based observation data

In addition to the actions described in the previous section, and with a lower priority, it is also suggested to investigate mechanisms to increase the level of integration between observations generated by space sensors and observation generated by ground observatories and telescopes.

Astronomical data are often consolidated with an augmented by other ground data that are not necessarily preserved or accessible in the same repositories. Integration of the data would provide additional material for scientific analysis and also stimulate the cross-fertilisation between different sub-domains of the Space Science and Astronomy domains.

Indicative costs for these activities are expected to be lower than the cost of integrating space and in-situ Earth Observation data. Most of ground based astronomical data are already stored in the FITS standard format and made available to the users' community. Therefore the rough order of magnitude of **EUR 1 Mln** can be considered as the higher end of the cost estimate for these activities.

5.3 Microgravity

The Microgravity domain presents the least advanced status of development of the data exploitation roadmap, despite an interest by the scientific community in accessing experimental data generated on all types of available platforms. The European Space Agency is the main channel available to European scientist to perform microgravity experiments. Stakeholder interview indicated that different opinions on the need of data preservation are registered within ESA organisations responsible for the implementation of microgravity experiments. Discussion are on-going and no activities have been initiated so far.

For experiments executed on non-ISS platforms, no active preservation initiative has been implemented so far.

Data access for real-time experiments follows a single PI scheme. The data are owned by the PIs that individually take responsibility to preserve them as they deem necessary. There is no pre-defined standard to store the data; data from each experiment are usually stored as generated by the experimental facility in their native format and on all kind of platforms (e.g. magnetic tapes, CD, DVD). Access to the data over the long-term requires a direct request to the PI, which may or may not be granted.

The exploitation of data generated by experiments executed on ISS is accomplished slightly differently where near-real-time data access follows a single-PI scheme. Generated data are owned by ESA and are processed in the national USOCs centres. Data processing is implemented using an industry proprietary system (CD-MCS) and the data are then stored in their experiment native format encapsulated in CCSDS packets.

There are no formal processes established to request access to the preserved data; this usually requires direct contact with the mission PI. The USOC can provide ad-hoc access to the preserved to third party PIs provided that:

- there is the agreement of ESA and of the mission PI, and;
- The requesting PI installs a specific terminal called UHB to connect to the USOC repository and access the data.

ESA has built a centralised archive called Erasmus Experiment Archive (EEA) hosted in ESTEC. EEA provides access to a very high level description of the experiments objectives, main achievements, a list of related publications and the contact information of the PI personnel. No access to the data in any format is provided. The total volume of data generated by European microgravity experiment is unknown. The range can be estimated up to hundreds of TB, considering both the raw data and the associated knowledge.

The main results for increasing the exploitation of the microgravity data have been obtained with the ULISSE and the CIRCE projects funded under the FP7 programme,.

Therefore, as described in the next sections, the European Commission is suggested to continue the activities started under FP7 and to cooperate with ESA on the establishment of a mechanism for accessing data preserved in the USOCs. .

5.3.1 Implementation of a data infrastructure for the access to the long term preserved microgravity data.

In 2008 ULISSE represented the first successful attempt to define standards for the preservation and storage of the data to demonstrate the feasibility of the data preservation for European microgravity experiments.

CIRCE was the continuation of the ULISSE project. It developed implementation roadmap for a data preservation system based on the ULISSE's demonstrator and expanded its scope addressing issues related to the management and compliancy to the different country based data policies and facilitating data mining for the user.

Cost estimated to implement the system proposed by ULISSE and Circe studies is evaluated in about EUR 10/12 Mln over 3 years. After that, the operational phase could be kicked-off, including the execution of awareness creation activities. The operation and the awareness creation are not covered by the above sum.

The European Commission could give continuity to the ULISSE and CIRCE projects evaluating the implementation of the proposed system under the H2020 Space Theme managed by DG GROW.

5.3.2 Cooperate with ESA on the consolidation of a data policy for microgravity data

This suggested action is the conditio sine qua non to the operations of the data infrastructure discussed in the previous section. ESA owns the data preserved in the USOCs and therefore ESA's agreement on the data accessibility is mandatory.

A similar cooperation has already been established within the framework of the project ULISSE. ESA management already agreed to provide access to the data generated by 30 ISS experiment to support the realisation of the system demonstrator. This agreement could be the starting point of new discussion focused on the following points:

- Agreement on the access to the data generated by experiment executed on the ISS;
- Support to the collection of the data generated by non-ISS experiment;
- Integration of the European Commission's provided data infrastructure and of the ESA's Erasmus Experiment Archive.

The EEA could represent an entry point for the scientific community similar to what is provided by NASA for the Physical Science experiment with the Physical Science Informatics (PSI) system. [73]

The European Commission's project would establish the underlying data infrastructure providing access to the experimental data.

5.4 Cross-cutting activities

In addition to the domain specific actions suggested in the previous sections, there are other actions that can be applicable to all domains. Despite the different status of implementation of the roadmap, and once all data infrastructure are established, all domain will face similar issue related to:

- the growing amount of data, and;
- the increase in the awareness of the scientific community (also outside of it) concerning the potential utilisation of space data.

As discussed in the previous chapters, a large amount of bandwidth is needed to deliver the data (not available in many locations), and the scientists need significant IT infrastructure to save, organise, and process the data – tasks and knowledge that are increasingly out of scope of the scientist’s knowledge and responsibilities. As a result, the processing trend has begun to reverse with the appearance of hosted processing and the development of exploitation platform.

The exploitation platforms are located between the scientists and the data, providing the scientists with all tool necessary to access the data, find and select the desired data set (including building of time or frequency series), process them according to a scientist designed algorithm hosted on the platform, analyse the data, and extract results. These hosted activities erase the need to download large amount of data, and the need to acquire dedicated processing infrastructure.

Both the European Commission and ESA have been active in the development of exploitation platforms in the past. The FP7 Gaia European Network for Improved data User Services (GENIUS) project and the ESA G-POD programme are two such examples. The GENIUS project is also an excellent example of cooperation between the European Commission and ESA where the European Commission supports the development of the infrastructure and ESA takes care of the operational aspects.

Further examples of activities similar to the GENIUS projects could be envisaged for the future where data service and data analysis tools are developed and integrated in existing, or under development, exploitation platforms

5.4.1 Development of data services and analysis tools

The European Commission could support the development of platforms dedicated to the exploitation of space scientific data by developing specific data services and data analysis tools (e.g. data reduction algorithm, knowledge representation ontologies and methodologies, advanced data mining services based on semantic maps, *et cetera*).

In the era of the big-data, the development of these tools and service could also build expertise within European industries that can easily be re-used in domains different from space.

The development of the tool and services should be targeted to the integration in existing or under development exploitation platforms such as the ESA G-POD[10].

Typical costs for such initiative would be in the range of EUR 1 Mln per year, with variable duration in the range of 1 to 3 years depending on the specific project [74]

5.4.2 Awareness building

Once the data infrastructures are up and running, the European Commission should implement generic awareness building activities within the scientific community to increase the exploitation of the data. The awareness raising initiative should target both the space science community and the wider scientific community.

There should be a reach-out to the space science community to increase the awareness of the new data availability, increased exploitation possibilities and services, and to stimulate the development of cross-domain scientific analysis.

The reach-out to the wider scientific community should focus on increasing the general awareness of the availability of science data generated by space missions, and fostering cross-fertilisation with other domain different from space.

5.5 Final recommendations

In this chapter we have discussed different activities the European Commission could implement in support of the increased exploitation of the scientific data generated by European Space missions. An overview of the actions is summarised in **Table 10** below. As indicated in the table actions have been identified specifically for each domain and as domain cross-cutting actions.

---

[10] Similar activities have already been implemented in the framework of the Copernicus programme for the development of specific data services using the Copernicus Grant funding scheme.

Domain	Action	Action priority	Indicative implementation costs
Earth Observation	Support the recovery of old third party EO mission archive	Priority 1	EUR 1 Mln per project, for a total duration of 1 year
Earth Observation	Support integration of Space and In-Situ data	Priority 2	EUR 0,5 to 1 Mln, for a maximum duration of 1 year
Space Science	Fund the exploitation of Space Science data	Priority 1	EUR 0,3 Mln per project, for a duration of 12 to 18 months
Space Science	Support the consolidation of the Virtual Observatory	Priority 1	N/A
Space Science	Support the integration of space and ground based observation data	Priority 2	EUR 0,5 to 1 Mln, for a maximum duration of 1 year
Microgravity	Implementation of a data infrastructure for the access to the long term preserved microgravity data.	Priority 1	EUR 4 Mln year for a total project duration of 3 years
Microgravity	Cooperate with ESA on the consolidation of a data policy for microgravity data	Priority 1	N/A
Cross-cutting	Development of data services and analysis tools	Priority 2	EUR 1 Mln per year, for a duration of 1 to 3 years
Cross cutting	Awareness building	Priority 2	

Table 10: summary of suggested actions

We suggest to treat the three domain of interest independently as each domain is at a different level of implementation of the data exploitation and each domain faces different issues concerning data volume, and data access.

Implementation of action aiming at aligning infrastructure, standards and methodologies between different domains are not suggested at this stage. This could be envisaged as being appropriate later when the infrastructure is more mature and comparable after further development.

Within each domain, the activities identified as the highest priority are strongly suggested for implementation, while the lower priority actions are not on the critical path of exploitation.

The activities cutting across all domains are all categorised as secondary priority as their implementation is for the longer term and not on the current critical path for data exploitation maximisation.

5.6 Notes and references

[72] Stakeholder interview, ESA-ESRIN, M. Albani, 19-05-2015

[73] http://psi.nasa.gov/home.aspx

[74] http://www.ecmwf.int/sites/default/files/COP-CDS-WS-Pechorro.pdf

6 BIBLIOGRAPHY

Articles

[1] Jean-Christophe Malapert, Luigi Carotenuto, Olivier Belbis, Patrick Hambloch, Rainer Willnecker, Sebastien Barde 2013 A Data e-Infrastructure for the International Space Station 1

[2] Jeff Foust 2014 Making the most of the ISS The Space Review, March 2014 0

[3] Nick Skytland 2012 What is NASA doing with Big Data today? Open.NASA, 4th October 2014, Last accessed in March 2014 1

[4] Gunther Kohlhammer 2001 The ENVISAT exploitation policy ESA Bullettin 106, June 2001 1

[5] "Christophe Arviset, Danny Lennon, Fred Jansen, Martin Kessler, Nick Hanowski, Pedro Osuna" 2013 "Long Term Strategy for ESAC Science Archives" Ensuring the long-term preservation and value adding to scientific and technical data 4-6 November 2013, ESA-ESRIN, Frascati (Rome), Italy 2

[6] Fulvio Marelli, Massimo Argenti, Mirko Albani 2013 On Harmonizing Earth Science Policies, Semantics, Metadata and Ontologies Ensuring the long-term preservation and value adding to scientific and technical data 4-6 November 2013, ESA-ESRIN, Frascati (Rome), Italy 2

[7] Arnold H. Rots, Glenn E. Becker, Sherry L. Winkelman 2012 Chandra Publication Statistics 0

Documents

[1] EC 1205/2008 COMMISSION REGULATION (EC) No 1205/2008 of 3 December 2008 Implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata Regulation 3/12/2008

[2] GSCB Ground Segment Coordination Body European Framework for the long term preservation of Earth Observation space data

[3] GSCB Ground Segment Coordination Body "Long Term Preservation of Earth Observation Space Data European LTDP Common Guidelines V2.0, 30/06/2012" Guidelines

[4] GSCB Ground Segment Coordination Body "Long Term Preservation of Earth Observation Space Data European Earth Observation Space Data Set V2.0, 31/01/2013" Data Sets

[5] EC 2014 Conference on Big Data from Space BiDS’14, 12-14/11/2014 Executive Summary

[6] EC Proceedings on the 2014 Conference on Big Data from Space BiDS’14, 12-14/11/2014 Conference Proceeding

[7] Telespazio CIRCE. Project leaflet Project leaflet

[8] NASA CDF User’s guide V3.5. Space Physics Data Facility NASA / Goddard Space Flight Center. 04/09/2013 User's guide 4/9/2013 4

[9] CASIS CASIS Hardware and facilities fact sheet

[10] CASIS CASIS Support services fact sheet

[11] CASIS CASIS Strategic plan

[12] CASIS CASIS 2014 annual report

[13] EUMETSAT EUMETSAT Data Policy Data Policy 01/2015

[14] European Union 2007/2/EC "DIRECTIVE 2007/2/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)" Directive 14/03/2007

[15] DLR "DLR Earth Observation Data Policy Presentation for GENESI-DR Workshop" Data Policy

[16] DLR COFUR Prices for TerraSAR-X – products (scientific use) Data Price

[17] DLR "User License for the Utilisation of TerraSAR-X / TanDEM-X Data and Products for Scientific Use" Data Policy

[18] European Space Agency Revised ESA Data Policy for ERS, Envisat and Earth Explorer missions Data Policy 2010

[19] European Space Agency V1 11/05/11 Terms and conditions for the utilisation of ESA's Earth Observation Data Data Policy 2011

[20] European Space Agency V14/02/26 Term and conditions for the utilisation of data under ESA's third party missions scheme Data Policy 2014

[21] LandSat LandSat data distribution policy Data Policy

[22] NASA Landsat 7 Science Data Users Handbook Data Access Handbook

[23] European Space Agency "ENVISAT AND ERS MISSIONS Data Access Guide" Data Access Handbook

[24] European Union 98/34/EC "DIRECTIVE 98/34/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 June 1998 ►M1 aying down a procedure for the provision of information in the field of technical standards and regulations and of rules on Information Society services;" Directive 22/06/1998

[25] CNES CAD-NT-223-31-CNS "USOC ARCHIVE AND DATA STORAGE REPORT" Tehcnical Note 14/06/2002 0 1

[26] United Nations A/RES/41/65 Principles relating to remote sensing of the Earth from space Principles 3/12/1986

[27] EC The Role of e-Infrastructures in the Creation of Global Virtual Research Communities 2010

[28] European Space Agency PV 2013 Ensuring the long-term preservation and value adding to scientific and technical data 4-6 November 2013, ESA-ESRIN, Frascati (Rome), Italy Conference Proceeding 11/2013

[29] European Space Agency SCIDIP-ES-DEL-WP33-D33.1 "Earth Science common preservation policies and data access policies analysis" FP7 Project deliverable 31/07/2013 1.0

[30] European Space Agency SCIDIP-ES-DEL-WP33-D33.2 Earth Science metadata, semantics and ontology harmonisation report FP7 Project deliverable 10/06/2013 3.0

[31] APA - Alliance for Permanent Access Sustainability - Securing the value of our digital assets FP7 Project brochure

[32] GSCB Ground Segment Coordination Body LTDP-GSEG-EOPG-RD-11-0003 "Long Term Data Preservation Earth Observation Preserved Data Set Content LTDP/PDSC" Requirements 25/07/2012 4 0

[33] European Union GENESIS Project brochure 01/2009

[34] European Space Agency Guidelines for the submission of registrations Technical note 02/2012 5 1

[35] STSC Hubble Space Telescope Call for Proposals for Cycle 23 - Policies, Procedures & Phase I Proposal Instructions Call for proposal January 2015

Presentations

[36] Luigi Carotenuto Telespazio 2014 A data e-infrastructure for the International Space Station "H2020 Information days, February 13-14 2014 European Commission - Brussels"

[37] Luigi Carotenuto Telespazio 2014 Exploitation of ISS scientific data "Copernicus Big Data Workshop March 13-14 2014 European Commission – Brussels – Belgium"

[38] European Space Agency 2000 (?) ESA Data Policy

[39] European Space Agency 2000 The ESA EO Data Policy for ERS and ENVISAT September 2000

[40] Mirko Albani European Space Agency 2007 "European Strategy for Long Term EO Data Preservation and Access" "FRINGE 2007 Workshop 26th November 2007 – ESRIN"

[41] Aurelie Bellucci CNES 2013 SIPAD-NG: a multi-field system for archiving and accessing data. Feedback on 10 years of development and operational use PV 2013 , ESRIN, 6th November 2013

[42] Martin Kessler European Space Agency 2015 "Assorted slides in support of discussions with strategy&" ETEC, 6/05/2015

[43] Giusi Micela INAF- Istituto Nazionale Astrofisica 2012 "Accordo attuativo della Convezione Quadro per ""Supporto all'analisi dati per l'astrofisica delle alte energie e la cosmologia delle strutture formate""I/009/10/0 Riunione finale Relazione sulle attivita' svolte" Roma, 17/01/2012

[44] Andrea Della Vecchia, Julio Carreira European Space Agency 2013 "Earth Observation Data and Applications" "Sustainability: Securing the Value of Digital Data Den Haag, Netherlands, 10th and 11th December 2013"

[45] Luigi Carotenuto Telespazio 2013 Preservation and exploitation of scientific data produced on the ISS November 26-27 2013, Royla Observatory of Belgium, Brussles

[46] George Helou IPAC 2013 Lessons Learned! From Past and Current ESA-NASA Partnerships November 2013, IPAC, Astronomy & Astrophysics Advisory Committee, Caltech

[47] Olivier Arino European Space Agency 2013 “Monitoring the South Atlantic Anomaly using ATSR historical data sets” or … why all data should be preserved! PV 2013 , ESRIN, 6th November 2013

[48]

Web-pages

[1] European UnionCopenricus http://www.copernicus.eu

[2] European UnionLEIT https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/2443-compet-05-2015.html

[3] European UnionINSPIRE Geoportal http://inspire-geoportal.ec.europa.eu

[4] European Space Agency https://eoportal.org/web/eoportal

[5] European Space Agency http://earth.esa.int

[6] European Space Agency http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus

[7] European Space Agency http://copernicusdata.esa.int

[8] European Space Agency https://sentinel.esa.int/web/sentinel/home

[9] European Space Agency https://scihub.esa.int/dhus/

[10] European Space Agency http://www.sciops.esa.int

[11] European Space Agency Rosetta PIs proprietary period http://blogs.esa.int/rosetta/2014/07/16/access-to-rosetta-data

[12] European Space Agency Herschel data processing levels http://herschel.esac.esa.int/Docs/PACS/html/ch07s02.html

[13] European Space Agency http://eea.spaceflight.esa.int/portal

[14] ESAEuroEuropean Space Agency long term data preservation http://earth.esa.int/gscb/ltdp

[15] NASA http://earthdata.nasa.gov

[16] NASA http://gcmd.gsfc.nasa.gov

[17] NASA http://nssdc.gsfc.nasa.gov

[18] NASA http://cdaweb.gsfc.nasa.gov/istp_public/

[19] NASA http://lsda.jsc.nasa.gov

[20] NASA Microgravity/General experiments info) http://www.nasa.gov/mission_pages/station/research/experiments/experiments_hardware.html

[21] CNES http://www.cnes.fr/web/CNES-fr/11153-st-centres-de-donnees.php

[22] JAXA Global Change Observation Mission http://gcom-w1.jaxa.jp

[23] JAXA Globe Portal https://www.gportal.jaxa.jp

[24] JAXA Data Archives and Transmission System http://www.darts.isas.jaxa.jp

[25] ISRO http://bhuvan.nrsc.gov.in/bhuvan_links.php

[26] ESF - European Space Foundation http://www.esf.org/media-centre/ext-single-news/article/the-european-space-sciences-committee-reacts-after-the-european-space-agency-council-meeting-at-mini.html

[27] European UnionCirce http://cordis.europa.eu/project/rcn/104205_en.html

[28] European UnionCirce http://www.circeproject.eu

[29] E-USOCUlisse http://www.eusoc.upm.es/en/projects/ulisse.html

[30] European UnionUlisse http://www.ulisse-space.eu

[31] European Space Agency ELIPS http://www.esa.int/Our_Activities/Human_Spaceflight/International_Space_Station/Taking_the_ISS_to_the_next_level_ISS_exploitation_and_ELIPS

[32] ISO Standards: ISO 191 http://www.iso.org/iso/home/store/catalogue_ics/catalogue_ics_browse.htm?ICS1=35&ICS2=240&ICS3=70&

[33] European Space Agency http://www.esa.int/About_Us/Industry/Space_Related_Standards

[34] NASA NASA EO standards https://earthdata.nasa.gov/data/standards-and-references

[35] NASA NASA science data standards http://nssdc.gsfc.nasa.gov/nost/curation.html

[36] NASA Common data formats http://cdf.gsfc.nasa.gov

[37] EUMETSAT EUMETSAT data formats http://www.eumetsat.int/website/home/Data/Products/Formats/index.html

[38] European Space Agency Sentinel-1 SAR data processing levels https://sentinel.esa.int/web/sentinel/missions/sentinel-1/data-products

[39] ESF - European Space Foundation SMOS payload data processing levels https://earth.esa.int/web/guest/-/data-types-levels-formats-7631

[40] NASA NASA EO data processing levels http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/

[41] data processing levels http://www.wmo.int/pages/prog/sat/dataandproducts_en.php

[42] European Space Agency ESA/EO Data policy https://earth.esa.int/web/guest/data-access/how-to-access-eo-data/earth-observation-data-distributed-by-esa;jsessionid=966C6E2A61427C78340D6C8266FA1577.eodisp-prod4040

[43] Science Direct Utilisation Indicators http://www.sciencedirect.com

[44] Hedden Information Management From Taxonomies to Ontologies: Customized and Semantic Relationships http://accidental-taxonomist.blogspot.co.uk/2012/11/from-taxonomies-to-ontologies.html

[45] European Space Agency Sentinel 2 processing data https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/processing-levels

[46] European Space Agency ESA PV 2013 http://www.congrexprojects.com/2013-events/pv2013/programme

[47] European Space Agency Heterogeneous Missions Accessibility(HMA) http://earth.esa.int/hma/

[48] CNES SIPAD-NG http://vds.cnes.fr/sipad_ng/index.html

[49] European UnionScience Data Infrasctructure for Preservation - Earth Science http://www.scidip-es.eu/scidip-es/

[50] APA - Alliance for Permanent Access Alliance for Permanent Access http://www.alliancepermanentaccess.org

[51] European Space Agency G-POD, Grid Processing On Demand http://gpod.eo.esa.int

[52] European Space Agency G-POD User Manual http://wiki.services.eoportal.org/tiki-index.php?page=GPOD+User+Manual

[53] European UnionGENESIS - GENeric European Sustainable Information Space for environment http://cordis.europa.eu/project/rcn/87874_en.html

[54] European UnionENVRI - Common Operations of Environmental Research Infrastructures http://cordis.europa.eu/project/rcn/101244_en.html

European Union GEOWOW (GEOSS interoperability for Weather, Ocean and Water) http://cordis.europa.eu/result/rcn/58708_en.html

7 LIST OF ACRONYMS

Acronym	Definiton
ALR	Austrian Space Agency
ARTES	Advanced Research in Telecommunication Systems
ASI	Agenzia Spaziale Italiana
BELSPO	Belgian Science Policy Office
CDTI	Centro para el Desarrollo Tecnológico y Industrial
CFOSAT	Chinese-French Oceanic SATellite
CNES	Centre National d'Etudes Spatiales
CNSA	Chinese National Space Administration
CoRoT	COnvection, Rotation & planetary Transits
CTP	Science Core Technology Programme
CUST	Cryogenic Upper Stage Technologies
DLR	Deutschen Zentrums für Luft- und Raumfahrt
EDA	European Defence Agency
EGEP	European GNSS Evolution Programme
EGNSS	European GNSS
ENPI	European Neighbourhood and Partnership Instrument
EOEP	Earth Observation Envelope Programme
EOPA	Earth Observation Preparation Activities
ESF	European Science Foundation
ESA	European Space Agency
ESTER	European Space Technology Requirements database
ESTMP	European Space Technology Master Plan
ETEHP	European Transportation and Human Exploration Preparatory activity
EU	European Union
EU-28 MS	EU-28 Member States
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EWD	Earth Watch Definition
FAA	Federal Aviation Authority
FLPP	Future Launchers Preparatory Programme
FP	Framework Programme
GDP	Gross Domestic Product
GERD	Gross domestic expenditure on R&D
GIS	Geographical Information System
GMES	Global Monitoring for Environment and Security
GNSS	Global Navigation Satellite System
GSA	European GNSS Agency
GSTP	General Support Technology Programme
IPD	Instrument Pre-Development
ISRO	Indian Space Research Organisation
ISS	International Space Station
LBS	Location Based Systems
LSI	Large Scale Integrators
MERLIN	MEthane Remote sensing LIdar missioN
MFF	Multiannual Financial Framework
MoD	Ministry of Defence
MoE	Ministry of Economy/Enterprise
MoSE	Ministry of Scientifi Education
MoT	Ministry of Transport
MREP	Mars Robotic Exploration Preparation
NSO	Netherland Space Office
PPP	Public Private Partnership
PRS	Public Regulated Services
R&D	Research & Development
R&I	Research & Innovation
REA	Research Executive Agency

SABRE	Synergetic Air-Breathing Rocket Engine
SAR	Synthetic Aperture Radar
SME	Small and Medium Enterprises
SNBS	Swedish National Space Board
SSA	Space Situational Awareness
SST	Space Surveillance & Tracking
THAG	Technology Harmonisation Advisory Group
THE	High Thrust Engine
TRL	Technology Readiness Level
TRP	Basic Technology Research Programme
UKSA	UK Space Agency

8 LIST OF INTERVIEWED STAKEHOLDERS

ID	Organisation	Principal Point of contact	Interview date	Type of interaction
1	ESA/Earth Observation	G. Kohlhammer	22 April 2015	Face to face
2	ESA/Space Science	M. Kessler, F. Favata	22 April 2015	Face to face
3	Telespazio	L. Carotenuto	28 April 2015	Face to face
4	Italian National Institute of Astrophysics	G. Micela	26 May 2015	
5	UK Space Agency	L. Boland	8 May 2015	Face to face
6	UK Science and Technology Facilities Council	Peter Allan, Victoria Bennett, Chris Pearson	8 May 2015	Face to face
7	Centro para el Desarrollo Tecnologico Industrial	Cristina Garrido Gonzalo, Pillar Román, Emilio Vez Rodríguez	26 May 2015	Face to face
8	CNES	Geneviève Gargir, Thierry Levoir, Steven Hosford	28 May 2015	Face to face
9	DLR	Claudia Lindeberg, Wolfgang Frings, Godela Rosner	02 June 2015	Face to face
10	ESA/ESRIN	M. Albani	19 May 2015	Webinar

ET-04-15-793-EN-N

doi:10.2873/839169