

# ESA PLANS – A PATHFINDER FOR LONG TERM DATA PRESERVATION

**Vincenzo Beruti**

ESA-ESRIN,  
Via G. Galilei,  
CP 64,00044 Frascati,  
Italy

[Vincenzo.Beruti@esa.int](mailto:Vincenzo.Beruti@esa.int)

**M.Eugenia Forcada**

ESA-ESRIN,  
Via G. Galilei,  
CP 64,00044 Frascati,  
Italy

[Eugenia.Forcada@esa.int](mailto:Eugenia.Forcada@esa.int)

**Mirko Albani**

ESA-ESRIN,  
Via G. Galilei,  
CP 64,00044 Frascati,  
Italy

[Mirko.Albani@esa.int](mailto:Mirko.Albani@esa.int)

**Esther Conway**

STFC  
Rutherford Appleton  
Laboratory  
Didcot, Oxon, UK

[esther.conway@stfc.ac.uk](mailto:esther.conway@stfc.ac.uk)

**David Giaretta**

STFC  
Rutherford Appleton  
Laboratory  
Didcot, Oxon, UK

[david.giaretta@stfc.ac.uk](mailto:david.giaretta@stfc.ac.uk)

## ABSTRACT

Digital preservation is difficult. The technical difficulties are the cause of much research. Other types of difficulty are those to do with organisational commitment, funding and context.

With the increasing interest on global change monitoring, also the use and exploitation of long time series of Earth Observation (EO) data has been increasing systematically, calling for a need to preserve the EO data without time constrains.

On the other hand:

- Data archiving and preservation strategies are still mostly limited to the satellite lifetime and few years after.
- The data volumes are increasing dramatically.
- Archiving and data access technology are evolving rapidly.
- EO data archiving strategies, if existing at all, are different for each EO mission, each operator or agency.

In the meantime the issue grows more urgent since more and more EO missions' data can be called 'historic' and more and more operators are faced with the decision of whether and how to preserve their data.

This paper describes the European Space Agency's (ESA) plans for long term commitment to preserving EO data concerning Europe. We believe this shows that ESA provides a pathfinder example of the way in which all these difficulties can be tackled.

## 1. INTRODUCTION

The need for accessing historical Earth Observation (EO) data series has significantly increased over the last ten years, mainly for long term science and environmental monitoring applications. This trend is likely to increase even more in the future in particular because of the growing interest on global change monitoring which is driving users to request time-series of data spanning 20 years and more, and due also to the need to support the United Nations Framework Convention on Climate Change (UNFCCC).

There are therefore strong drivers to preserve EO space data, keeping them accessible and exploitable for the long term. The preservation of EO space data can be also

considered as a moral responsibility of the Space Agencies and other data owners as they constitute an asset for all mankind. In the next decade, the wealth of information currently locked inside the global data archives must be fully exploited and re-analyzed on a global scale (Figure 1 and 2 show examples of information extraction from long-term data series). This challenge relies on full accessible and exploitable archives (Figure 3).

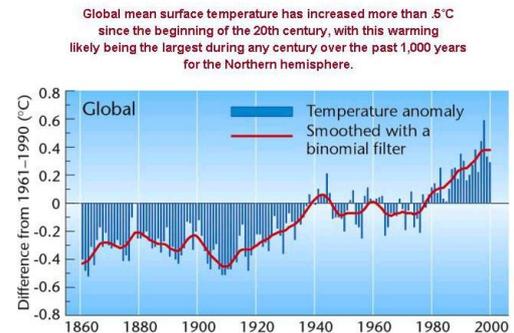


Figure 1 Surface Temperature increase

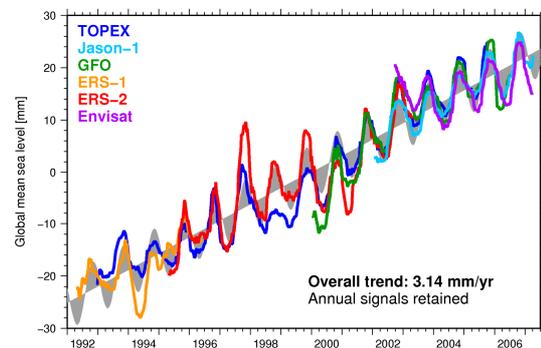


Figure 2 Global sea level raise (courtesy of Remko Scharroo)

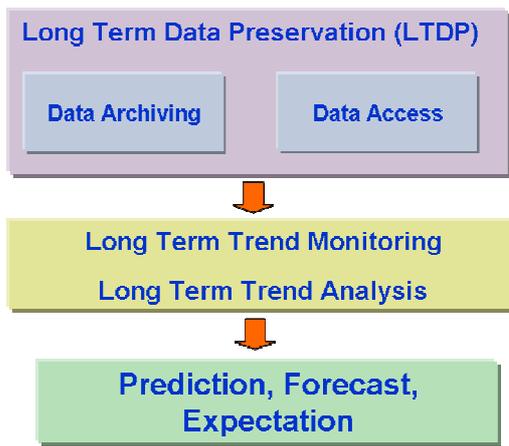


Figure 3 Long term trend monitoring

The application areas heavily benefiting from the EO long term data archiving exploitation are wide and can be summarized as:

- EC Policies with long-term perspective.
- European and Global Environment (e.g. Forest Monitoring, Soil Monitoring, Urban Development, Air Quality Monitoring, Ecosystems Monitoring and management for protection of terrestrial, coastal and marine resources).
- Management of energy resources (e.g. solar, etc.).
- Development and Humanitarian Aid Health including the understanding of environmental factors affecting human health and well-being.
- Food security including sustainable agriculture and combating desertification.
- Water resource management through better understanding of the water cycle.
- Civil Protection and disasters monitoring (e.g. Flood Prediction and Mitigation, Landslides, Subsidence, Volcanoes Monitoring).
- Global Climate Change (e.g. Systematic Climate Observations, Drought Monitoring, Monitoring of the Atmosphere, etc.).
- Climate understanding for assessing, predicting, mitigating and adapting to climate changes, as well as the improvement of weather information, forecasting and warning.
- Biodiversity enhanced understanding, monitoring and conserving.
- Global Security and Sustainable Development.

### 1.1. European EO archive challenges

The large number of new Earth Observation missions planned to come into operation in the next years (Figure 4) will lead to a major increase in the volume of EO space data. This fact, together with increased demands from the user community, marks a challenge for Earth Observation satellite operators, Space Agencies and EO space data providers regarding coherent data preservation and optimum availability and accessibility of the different data products.



Figure 4 Current and future European and Canadian EO missions (excerpt)

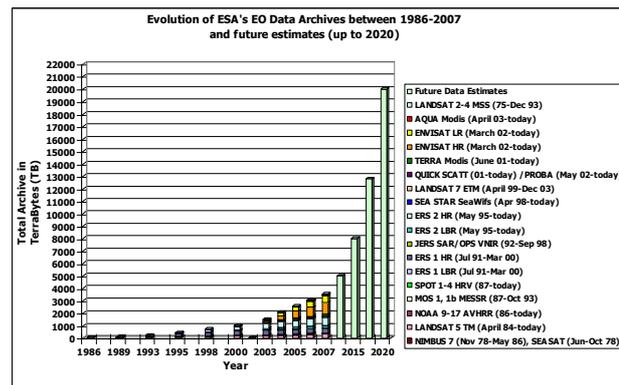


Figure 5 – ESA expected volume of archives

As an example, Figure 5 shows the approximate current and predicted data volumes from a number of missions. The rate of growth of data volumes increased since the launch of ERS and in particular of the Envisat mission and will be increasing even further with the contribution of additional Third Party Missions (TPM) and with the GMES program. The plans of new ESA missions indicate 5-10 times more data to be archived in next 10-15 years. Similar trend is also monitored at all National archives.

Traditionally in Europe, there has been poor cooperation in this field with no common approach for long term preservation and access to EO space data despite the need for cooperation and sharing for the benefit of the user community. Preserving today's science records (e.g. data, publications) as well as their context is fundamental in order to preserve the future of science but single organizations have difficulties to afford data preservation in the long term that calls for the need of optimising costs and efforts, identifying commonalities.

A cooperative and harmonized collective approach on Long Term Data Preservation (LTDP) in Europe (i.e. a European EO LTDP Framework) is needed to coordinate and optimize European efforts in the LTDP field and to ultimately result in the preservation of the complete European EO space data set for the benefit of all European countries and users and with a reduction of overall costs.

## 2. ESA LTDP APPROACH

The European Space Agency now has a commitment to the long term preservation of its holdings and is putting in place

the technical, financial and organisational wherewithal to accomplish this. Although at the time of writing not all the details have been decided, this paper describes some of the key principles which have been adopted. We believe that this approach can act as a pathfinder example for other disciplines.

In summary:

- A number of important principles have been identified and agreed with key Earth Observation space data holders and stakeholders in Europe and Canada.
- The key datasets to be preserved are being identified. This collection includes many types of digital objects, both data and documents.
- Common guidelines for key stakeholders have been defined in cooperation with European and Canadian EO space data stakeholders and are being refined.
- A technical and organisation framework is being prepared (European LTDP Framework).
- The way in which ESA LTDP fits into the broader international efforts in domains different from the Earth Observation one is being consolidated in order to maximise the usefulness of what ESA is doing and minimises duplication and waste of effort.

### 3. KEY PRINCIPLES

Main goals of the European EO Long Term Data Preservation Framework are to:

- Preserve the European, and Canadian, EO space data sets for an unlimited time-span.
- Ensure and facilitate the accessibility and usability of the preserved data sets respecting the individual entities' applicable data policies.
- Adopt a cooperative and harmonized collective approach among the data owners (LTDP Framework) based on the application of European LTDP Common Guidelines and sustained through cooperative (multi-source) long term funding schemes.
- Ensure, to the maximum extent possible, the coherency with the preservation of other non-space based environmental data and international policies.

The European LTDP Framework is open and is intended as a collaborative framework consisting of distributed and heterogeneous components and entities cooperating in several areas to reach a harmonized preservation of the European EO space data set. The framework is based on the contribution of European EO space data owners through their ideas and possibly their infrastructure in accordance to the commonly agreed LTDP Guidelines and should follow a progressive implementation based on a stepwise approach (short, mid, long-term activities).

A common approach in the field of Long Term Data Preservation should aim at the progressive application of the European LTDP Common Guidelines but also at cooperation of the archive owners in several areas for a progressive development and implementation of technology, methodology, standardization, operational solutions and data exploitation methodologies as key aspects for the set-up of the framework. The European LTDP framework is outlined in Figure 6.

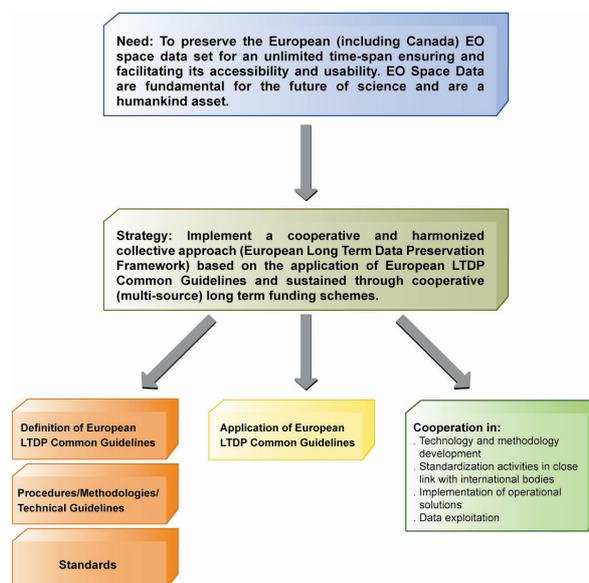


Figure 6 – Long Term Data Preservation Outline Framework

A cooperative framework can facilitate for EO space data owners and archive holders the achievement of the common goal of preserving and guaranteeing access to the own data through benefiting from proven technologies, procedures and approaches and through the possibility to reuse and share infrastructure elements in the long term. The adoption of standards (e.g. for data access interfaces and formats, procedures, etc.) and common technical solutions can also allow to significantly reduce preservation costs.

The European LTDP Framework should be sustained through a cooperative programmatic and long term funding framework based on multilateral cooperation with multiple funding sources from at least the European EO space data owners.

The existence of a European LTDP Framework will also increase the awareness on data preservation issues favouring the start of internal processes at private or public European EO space data owners and providers. A European framework could also trigger the availability in the long term of additional permanent funding sources (e.g. European Commission) and can increase the possibility for any European (including Canada) EO space data owner to preserve missions data beyond their funding schemes into the cooperative and distributed framework.

### 4. THE LTDP COMMON GUIDELINES

In 2006, the European Space Agency (ESA) initiated a coordination action to share among all the European (and Canadian) stakeholders a common approach to the long term preservation of Earth Observation space data. During 2007, the Agency started consultations with its Member States presenting an EO Long Term Data Preservation strategy [1] targeting the preservation of all European (including Canada) EO space data for an unlimited time-span ensuring and facilitating their accessibility and usability through the implementation of a cooperative and harmonized collective approach among the EO space data owners.

The Long Term Data Preservation Working Group with representatives from ASI, CNES, CSA, DLR and ESA was

formed at the end of 2007 within the Ground Segment Coordination Body (GSCB, [2]) with the goal to define and promote, with the involvement of all the European EO space data and archive owners, the LTDP Common Guidelines and also to increase awareness on LTDP. The resulting draft LTDP guidelines were reviewed by all ESA member states in the DOSTAG. The scope of Long Term Data Preservation as intended in the guidelines is not limited to the preservation of the data in the archives but also of the capabilities to generate products from the archived data and includes therefore also processing aspects. The insurance and facilitation of access, respecting the individual entities applicable data policies, and exploitation of the archived data are also part of the guidelines. Data access policies are on the other hand not part of the European LTDP Common Guidelines.

During the 1st Earth Observation Long Term Data Preservation workshop in May 2008 [3], the draft guidelines and the framework were presented and debated by all European and Canadian EO data owners, data providers and archive holders. The participants discussed and developed a joint strategy to move ahead technically and programmatically concerning the Long Term Data Preservation of EO Data and recognized the need and benefits of a common approach. Furthermore all the participants identified and agreed the draft LTDP Common Guidelines presented at the workshop as a first concrete and fundamental step to move ahead in creating the Long Term Data Preservation Framework. The guidelines should be adopted for old missions with a step-wise approach and straightforward for new missions and projects. ESA was given the task to trigger and coordinate the following steps toward the progressive European LTDP Framework implementation.

A consolidated LTDP Common Guidelines document has been produced on the basis of the comments and feed-backs received during the LTDP workshop. The document addresses the following nine main themes defining for each the "Guiding Principle" and the "Key Guidelines":

- Preserved data set composition
- Archives maintenance and data integrity
- Archives operations
- Data security
- Data ingestion
- Data access and interoperability
- Data exploitation and re-processing
- Standardization
- Data Purging/Appraisal

An extensive public review process of the guidelines document was undertaken which collected additional comments and feed-back from the EO space data owners and archive holders. The review process has been completed and the guidelines document can be found at [5], but they are under continued review.

The LTDP guidelines constitute a basic reference for the long term preservation of EO space data. Their application by European EO space data owners and archive holders is fundamental in order to preserve the European EO space data set and to create a European LTDP Framework. The application of the identified guidelines is not mandatory for European EO space data owners and archive holders but is strongly recommended following a step-wise approach starting with a partial adherence. To this end different priorities have been associated to each guideline and three different levels of adherence to the LTDP Common Guidelines as a whole have been defined. Adherence to the

guidelines should start from the basic level ones to reach full adherence in the long term. The LTDP guidelines document is intended to be a living document and can be also considered as a starting point to support the establishment, and aid the implementation, of more detailed technical procedures/methodologies when missing, favouring active cooperation in Europe in the LTDP field.

## 5. COOPERATIVE ACTIVITIES

The initial areas of cooperation related to LTDP to be addressed are on:

- Policies for the consolidation and issue of the European LTDP Common Guidelines, and adherence to them, and for the definition and application of a purge alert / appraisal procedure to EO space data.
- Technology, methodology and developments. The aim is to jointly evolve archive and data access technology through studies/pilots sharing the acquired know-how and infrastructure and to share knowledge/experience exchanging information to favour technical cooperation (cross participation into reviews, share of solutions, products, developments, etc). An additional fundamental activity is the continuation of development of harmonized access mechanisms (like HMA) and the definition of common operational procedures.
- Standardisation activities in close link with international bodies (e.g. CCSDS, CEOS, OGC, INSPIRE, EU initiatives, GEO).
- Operational solutions setting-up the principles for a common European distributed archiving concept aiming at the creation of an interoperable network of archive centres possibly reusing infrastructure of the different entities (as a single archive) in the long term. This would pave the way to future cooperation programs starting from standardized and certified services (e.g. share of data archives, archive transfer on demand or in case of a purge alert, coordination of re-processing schemes, format adoption/conversion, etc.).
- Data exploitation through the definition and implementation of joint EO historical data exploitation programmes.

An initial set of activities has been started in some of the areas mentioned above (e.g. studies on next generation archive technology and for the definition of LTDP users' requirements and composition of the data set to be archived to guarantee knowledge preservation). Details on the activities and their results will be provided published on the LTDP area of the GSCB web site [4]. Participation to the cooperation activities is open to European EO space data owners and archive holders and based on voluntary contribution.

## 6. TECHNICAL ROADMAP

ESA has developed the Standard Archive Format for Europe (SAFE) [6] an extension of the XFDF standard [7]. SAFE has been designed to act as a common format for archiving and conveying data within ESA Earth Observation archiving facilities. Many of the most important datasets have been converted to this format. The important point is that XFDF, and therefore SAFE, is designed to implement the OAIS Archival Information Package [8], which in principle

has everything needed for long term preservation of a piece of digitally encoded information.

Some of the other components under consideration for the ESA LTDP technical implementation are the hardware needed to store the large volumes expected. Detailed studies of available and near-term storage solutions will inform the decisions which need to be made as a matter of urgency.

ESA has been involved with several EU part-funded projects concerned with digital preservation. Of these the PARSE.Insight project provides a roadmap [9] which is built on a broad survey of researchers, publishers and data managers [10].

The CASPAR project [11, in which ESA played an important role, has identified a number of key preservation components which should form part of a shared infrastructure. These components will be evaluated as part of the ESA LTDP Framework.

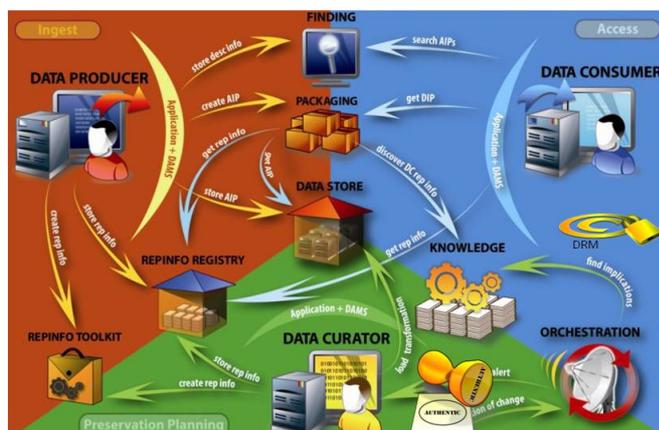


Figure 7 CASPAR workflows

For example when a data producer which to archive data, an AIP is created using the **Packager**. The AIP contains **Representation Information**, possibly from a **Registry of Representation Information (RepInfo)** or created with the **RepInfo Toolkit**. The amount of RepInfo depends upon the chosen **Designated Community** and the **Knowledge Manager** helps with this. Information supporting **Digital Rights (DRM)** and **Authenticity** are also needed. The AIP is deposited in the **Data Store**.

More broadly one can view projects funded within the EU which form the basis of a Science Data Infrastructure (SDI) (taken from Mario Compalargo).

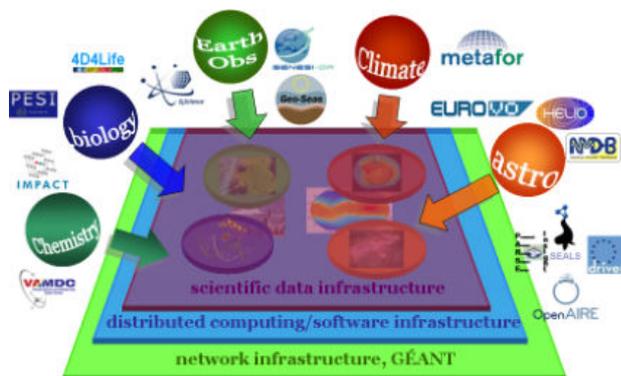


Figure 8 Landscape of e-infrastructure projects in the SDI area

Figure 8 illustrates the view that there should be a number of aggregators, including Earth Observation, Astronomy and Chemistry. These will be brought together into an overall science data infrastructure, which must include infrastructure components which assist in the preservation of digitally encoded information.

## 7. ORGANISATIONAL ROADMAP

The roadmap for the set-up of the European LTDP Framework can be articulated in three main phases to follow the best practices defined in the LTDP Common Guidelines and to progressively cooperate in the areas defined in Section 5. The initial situation is characterized by weak standardization, no clear and common methodology, poor cooperation with each entity dealing with the preservation of its own data often in a one by one mission basis.

The objective of the first phase is to reinforce the LTDP approach at each entity, to start the cooperation among agencies and EO space data owners (e.g. in methodology, standardization, sharing of information, etc...) and to define the future European organisation of LTDP with a very long term perspective. At the end of the first phase, standardization and methodology should be greatly defined on the main areas, EO space data owners should deal with LTDP as a transversal activity not closely tied to single missions and good cooperation among a significant number of EO space data owners should be in place.

In the mid term perspective (second phase) cooperation should be strengthened through the implementation of common activities among European EO space data owners with the goal to achieve an interoperable network of archives (e.g. share of solutions and systems, coordination of common technology developments, adoption of standards...), but also improving operational services according to user needs (i.e. Climate changes monitoring operational systems, etc.). Additional entities in Europe will be attracted and become part of the European LTDP Framework that is characterized at the end of the second phase by common technical views and solutions and standardized services, coordinated LTDP approaches and schedules between members and high interoperability between archives. LTDP is at this stage a common process which may be offered as a service able to manage and preserve the huge amount of EO space data owned by the different entities with a cost effective approach.

In the long term perspective cooperation should be further extended through the sharing of infrastructure (e.g. common and shared access points, interoperable and transparent data access and infrastructure), allowing to have a unique network of data, shared resources for data reprocessing and products generation and a common and harmonised security levels and layers. At this stage the EO space data archives could be extended also to other types of data according to user needs and alternative scenarios for the management of the exabyte/zetabyte archives era could be analyzed.

## 8. CONCLUSIONS

The European Framework for LTDP has been initiated. All European EO space data owners and archive holders are becoming part of a progressive process to guarantee the EO

space data preservation in the long term. The European LTDP Common Guidelines have been issued and published on the GSCB web site [5] and their promotion within CEOS or GEO international communities will be performed in the near future. The first identified technical cooperation activities have been started (e.g. studies on next generation archive technology and for the definition of LTDP users requirements and composition of the data set to be archived to guarantee knowledge preservation) and methodologies and standards available in international committees (e.g. CCSDS) or generated in the framework of ongoing projects (e.g. EC funded) are being revised for possible recommendation for adoption within the LTDP Common Guidelines. The international context of the European LTDP Framework is shown in Figure 7. In November 2008 at the ESA Ministerial Council, an LTDP programme for the period 2009-2011 was approved and ESA is now planning to apply the LTDP Common Guidelines to its own missions. ESA will implement the high priority activities in the next three years focussing on data preservation and enhancement of data access; an LTDP programme proposal will be prepared for the period beyond 2011.

## 9. REFERENCES

- [1] European Strategy for Long term EO data preservation and access, ESA/PB-EO/DOSTAG(2007)2, 8 October 2007.
- [2] GSCB Web Site, <http://earth.esa.int/gscb>
- [3] 1st LTDP Workshop, May 2008, [http://earth.esa.int/gscb/ldp/LTDP\\_Agenda.html](http://earth.esa.int/gscb/ldp/LTDP_Agenda.html)
- [4] LTDP area on GSCB Web Site, <http://earth.esa.int/gscb/ldp>
- [5] European LTDP Common Guidelines, Draft Version 2, [http://earth.esa.int/gscb/ldp/EuropeanLTDPCommonGuidelines\\_DraftV2.pdf](http://earth.esa.int/gscb/ldp/EuropeanLTDPCommonGuidelines_DraftV2.pdf)
- [6] SAFE web site <http://earth.esa.int/SAFE/>
- [7] XFDU standard available from <http://public.ccsds.org/publications/archive/661x0b1.pdf>
- [8] Reference Model for an Open Archival System (ISO 14721:2002), <http://public.ccsds.org/publications/archive/650x0b1.pdf> or later version. At the time of writing the revised version is available at <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf> or elsewhere on the CCSDS web site <http://www.ccsds.org>
- [9] PARSE.Insight Roadmap [http://www.parse-insight.eu/downloads/PARSE-Insight\\_D2-1\\_DraftRoadmap\\_v1-1\\_final.pdf](http://www.parse-insight.eu/downloads/PARSE-Insight_D2-1_DraftRoadmap_v1-1_final.pdf)
- [10] PARSE.Insight general survey report [http://www.parse-insight.eu/downloads/PARSE-Insight\\_D3-4\\_SurveyReport\\_final\\_hq.pdf](http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf)
- [11] CASPAR project web site <http://www.casparpreserves.eu/>