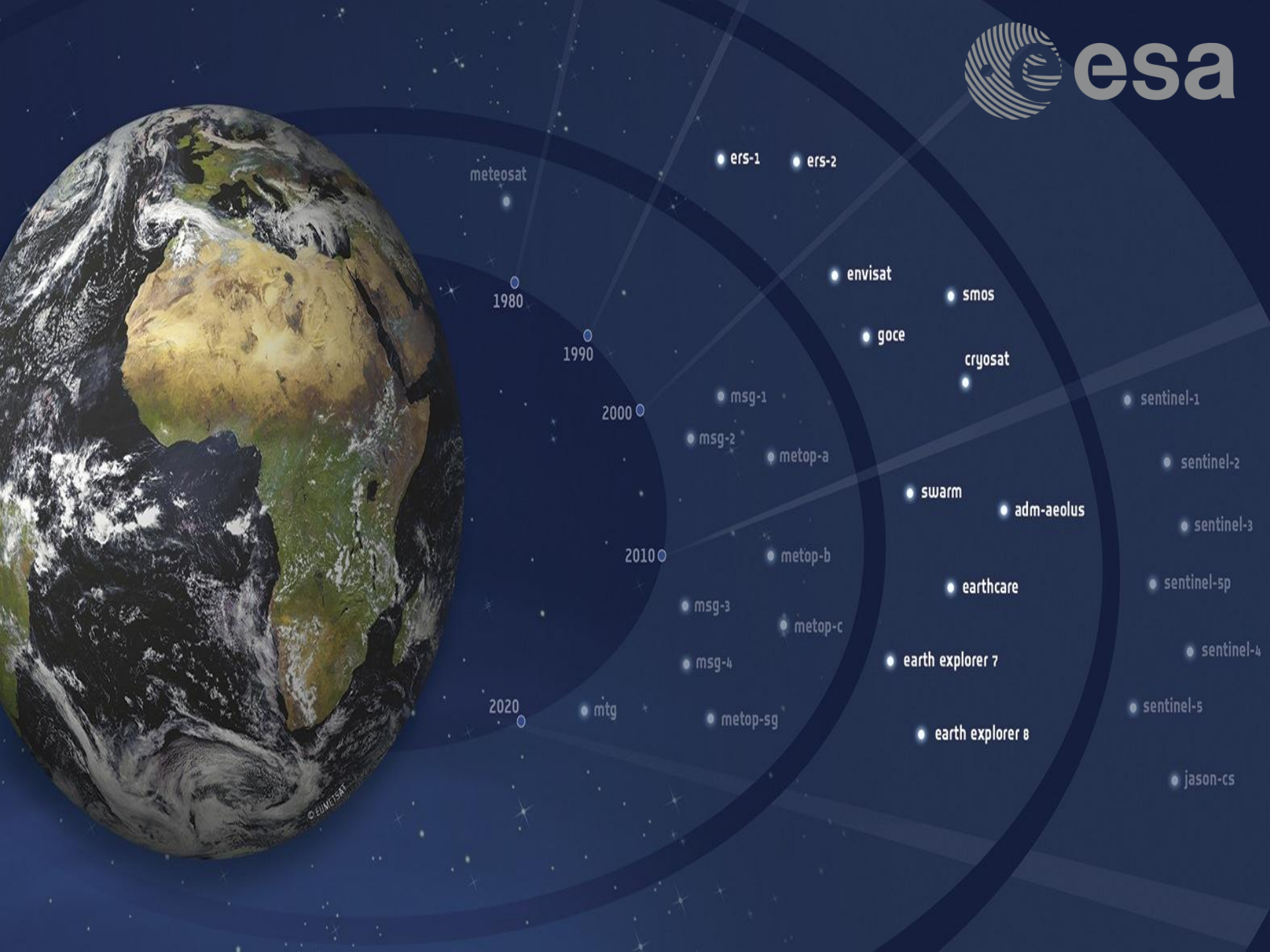
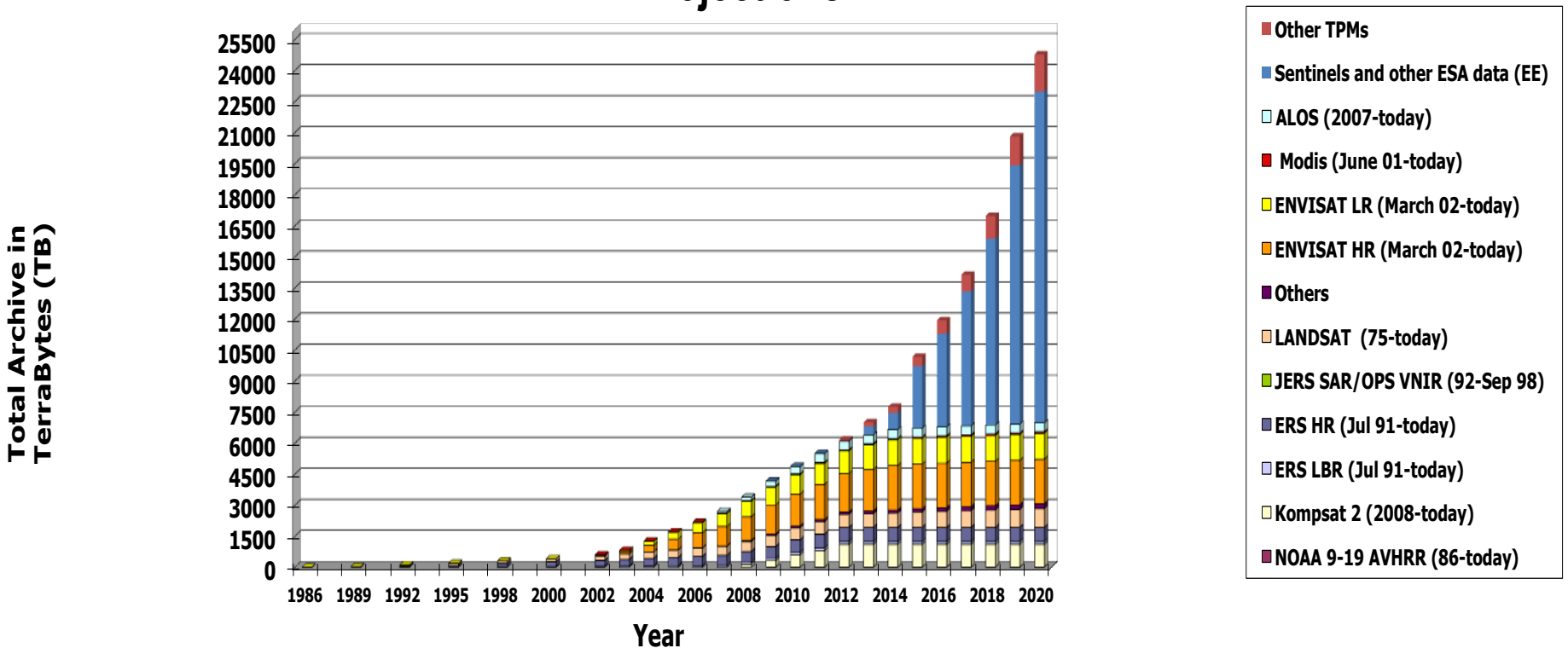
EO LTDP
And
The SCIDIP-ES project

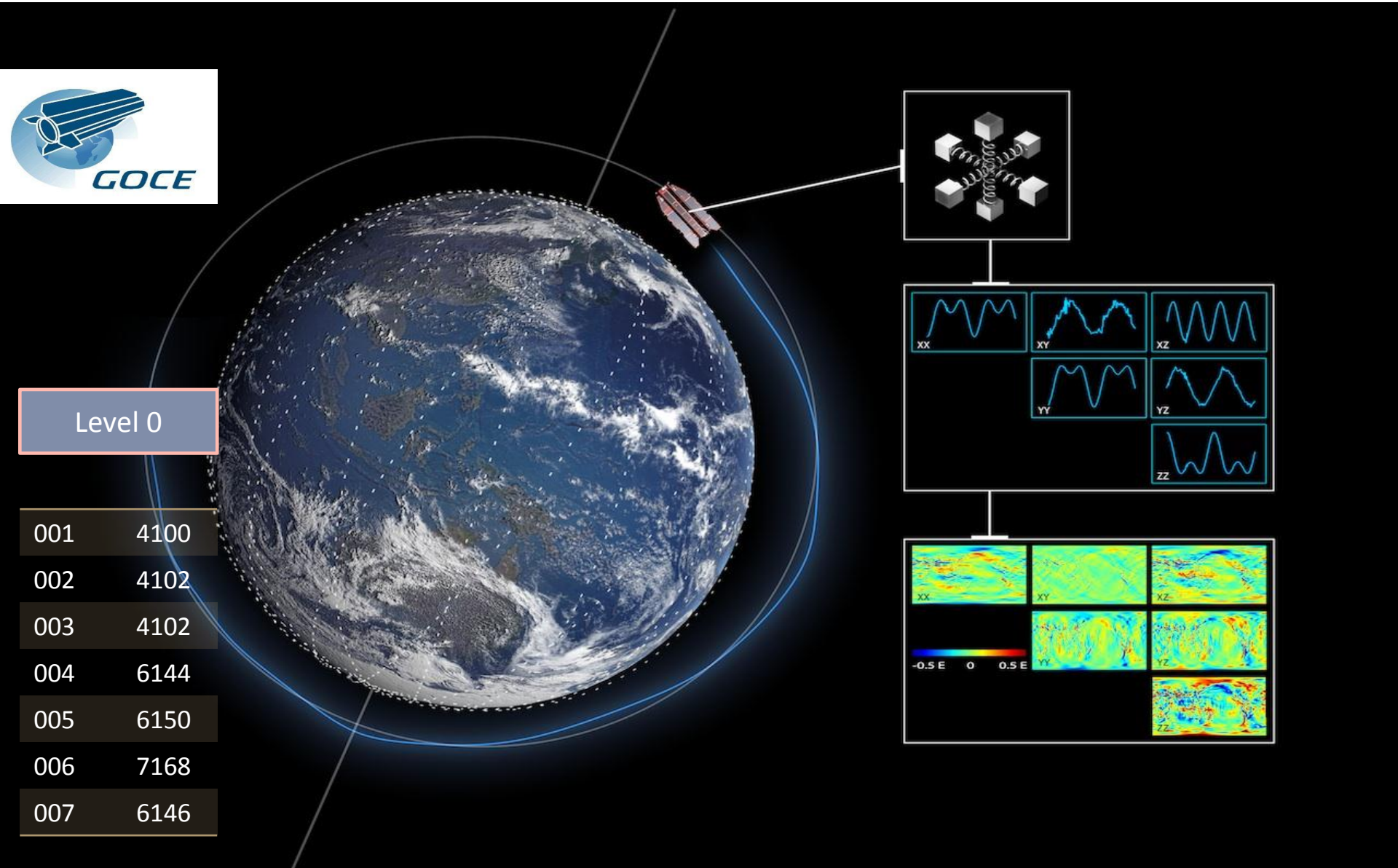
Fulvio Marelli - ESA


esa
meteosat
ers-1
ers-2
envisat
smos
goce
cryosat
1980
1990
2000
2010
2020
msg-1
msg-2
msg-3
msg-4
metop-a
metop-b
metop-c
metop-sg
mtg
swarm
adm-aeolus
earthcare
earth explorer 7
earth explorer 8
sentinel-1
sentinel-2
sentinel-3
sentinel-5p
sentinel-4
sentinel-5
jason-cs
© EUMETSAT

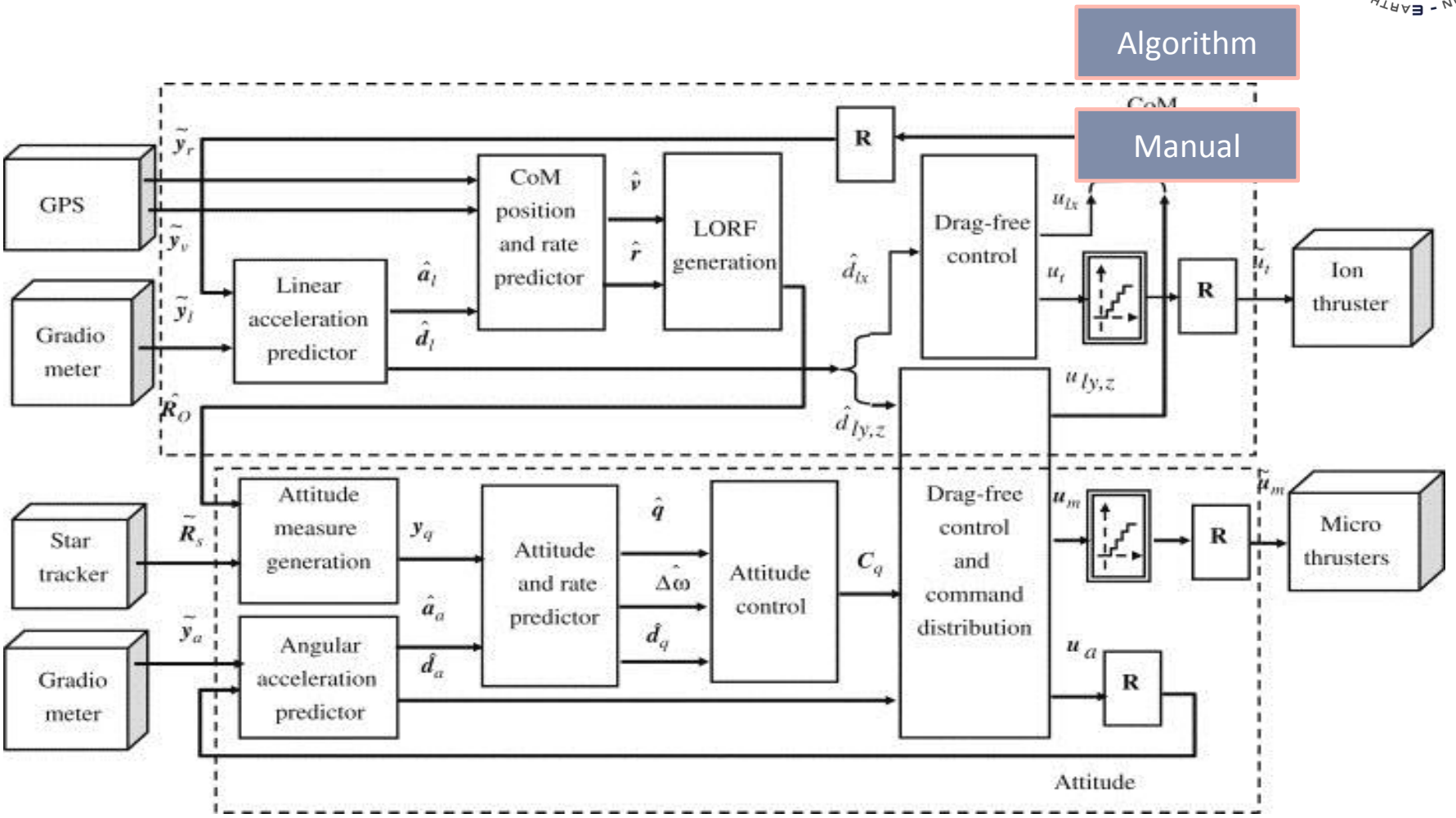
Evolution of ESA's EO Data Archives between 1986-2012 and future Projections



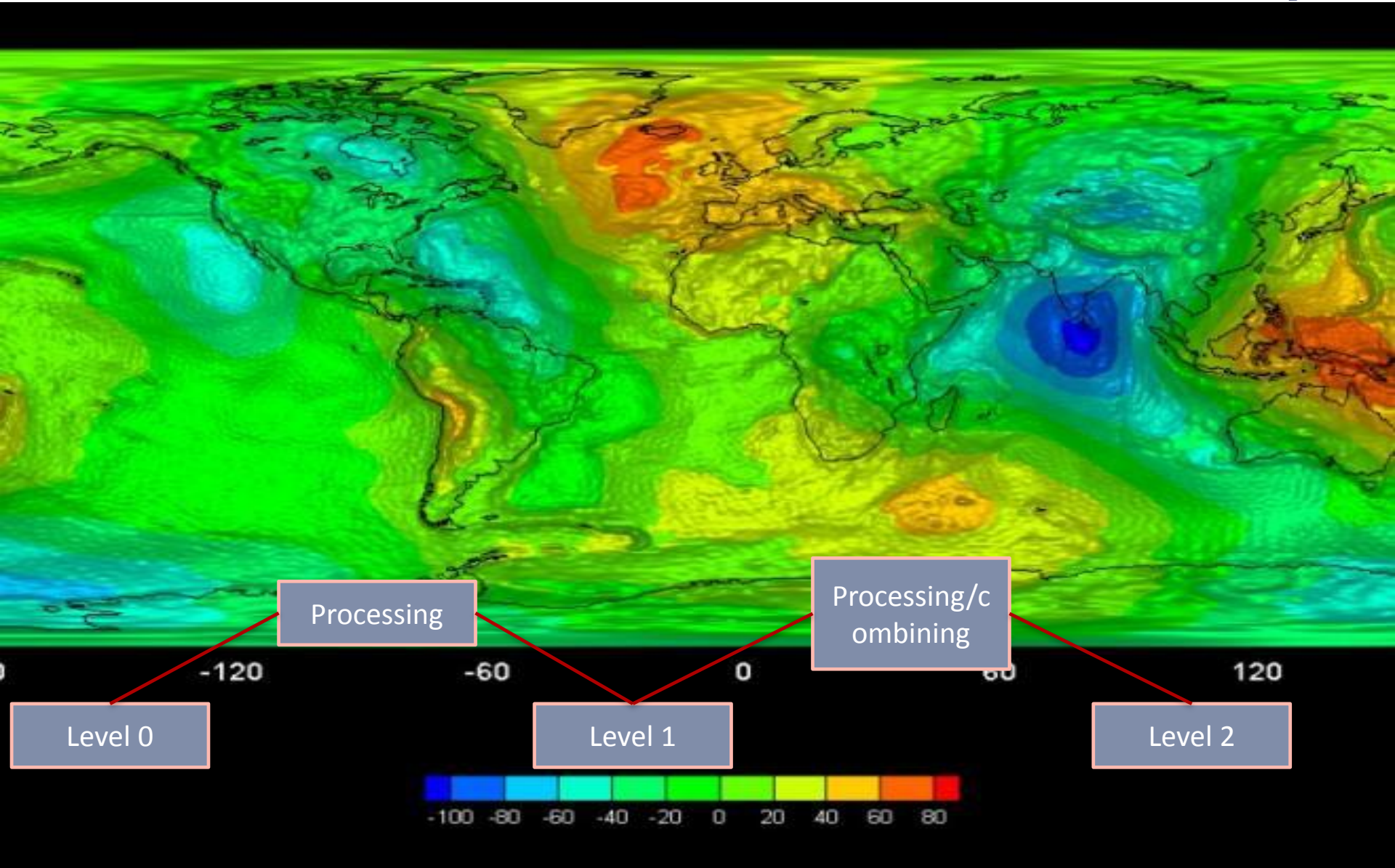
An example of data lifecycle: sensed data need to be acquired…



...through complex processing schemes



...to be combined and processed to get this



Example of Earth Science data use: Arctic sea ice melting

International Context

- "A fundamental characteristic of our age is the raising tide of data – global, diverse, valuable and complex. In the realm of science, this is both an opportunity and a challenge."

  Report of the High-Level Group on Scientific Data, October 2010, "Riding the Wave: how Europe can gain from the raising tide of scientific data"

- **Data Intensive Science (4th Paradigm):** data at the centre of the scientific process.

- **EC Recommendation 2012/417/EU, July 2012:** "Access to and preservation of scientific information".

- **Data is the new gold. "We have a huge goldmine ….. Let's start mining it."**

  Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda

Data & Knowledge Preservation

Level 0	Level 1	File specs.	Desc. Info
Processor	Algorithm	User Manual	Publications

- ***The preservation of data (the "bytes") is useless without the preservation of the knowledge associated with the data (e.g. the "quality", the process to generate them).***
- We must:
  - Ensure and secure the preservation of archived data and associated knowledge ideally for an unlimited time span.
  - Ensure, enhance and facilitate archived data accessibility.
    - Allowing to combine data from different sources and to perform more complex analyses (data unfamiliar become familiar)
  - Ensure coherency of approaches among different Earth Science providers.

- Preservation of EO data but also of all context information (documentation, CAL/VAL databases, algorithms, etc).
- Completeness and coherency among all elements to be preserved to ensure present and future exploitability.
- Data Quality, Context and Provenance (documented).
- Maintenance of capabilities to (re-)generate data products.
- Harmonized data and information accessibility.
- Coordinated and coherent approach in Europe.

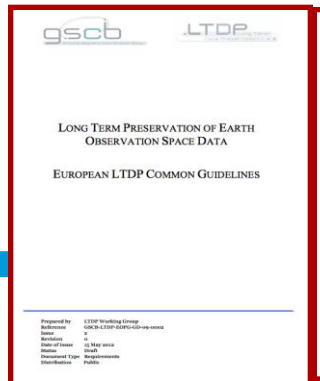
Preservation has to be addressed in all phases of a mission starting from its design. Incremental approach for past and current missions, systematic planning and approach for future missions.

EO LTDP WG

- ESA is coordinating the LTDP cooperation activities in the Earth Observation domain with European partners through the LTDP WG formed within the Ground Segment Coordination Body (GSCB).

- Implemented the basic rules of the European LTDP Framework in Earth Observation: "LTDP Common Guidelines and Preserved Data Set Content"
  - ✓ Reflecting the consensus of the European EO Data providers.
  - ✓ Reviewed at GEO, CEOS and with NASA.
  - ✓ Being reviewed with QA4EO.

New Issues available

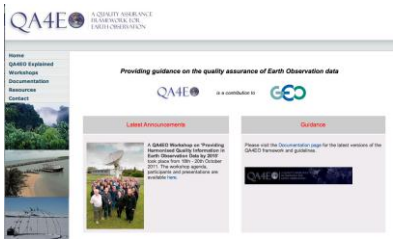


EO LTDP WG

- Defined the initial data set to be preserved, including the related glossary.

- Started several technical activities:
  - ✓ LTDP User Requirements Study (FIRST)
  - ✓ Archive Technology Study (LAST)
  - ✓ LTDP Initiatives and Standards Survey
  - ✓ LTDP/QA4EO Study
  - ✓ LTDP Architecture Definition Project

- Guaranteed the information flow through workshops, web sites, participation to conferences and LTDP related events.



LTDP Guidelines

1. Preserved data set content
2. Archive operations and organization
3. Archive security
4. Data ingestion
5. Archive maintenance
6. Data access and interoperability
7. Data exploitation and re-processing
8. Data appraisal and purge prevention

Mission phases

- Mission Concept (MC)
- Mission Definition (MD)
- Mission Implementation (MI)
- Mission Operation (MO)
- Post Mission (PM)

Content

- For an **Earth Observation** mission (i.e. project, campaign, experiment), the key data set target of preservation shall be composed by:
- **Data Records** (ranging from raw data to higher level products, browses, auxiliary, ancillary and calibration and validation data set) including quality indicators..
- **Processing Software**: this includes the processors to generate mission products, visualization, quality control and value adding software and tools.
- **Mission Documentation:**

Web Sites

- Institutional web site:
  - ✓ http://earth.esa.int/gscb/ltdp/
  - ✓ Contains basic LTDP documents (e.g. Guidelines).
  - ✓ LTDP events and related material (e.g. Workshops).

- Technical Web Site:
  - ✓ http://ltdpts.eo.esa.int/
  - ✓ Contains technical documentation and results of LTDP Cooperation activities.
  - ✓ Forum for technical discussions.

International EO LTDP Context

SCIDIP-ES

SCIENCE DATA INFRASTRUCTURE FOR PRESERVATION - EARTH SCIENCE

Project Overview

Introduction & Participants

Project: SCIence Data Infrastructure for Preservation – Earth Science (SCIDIP-ES)

INFRA-2011-1.2.2 Data infrastructures for e-Science

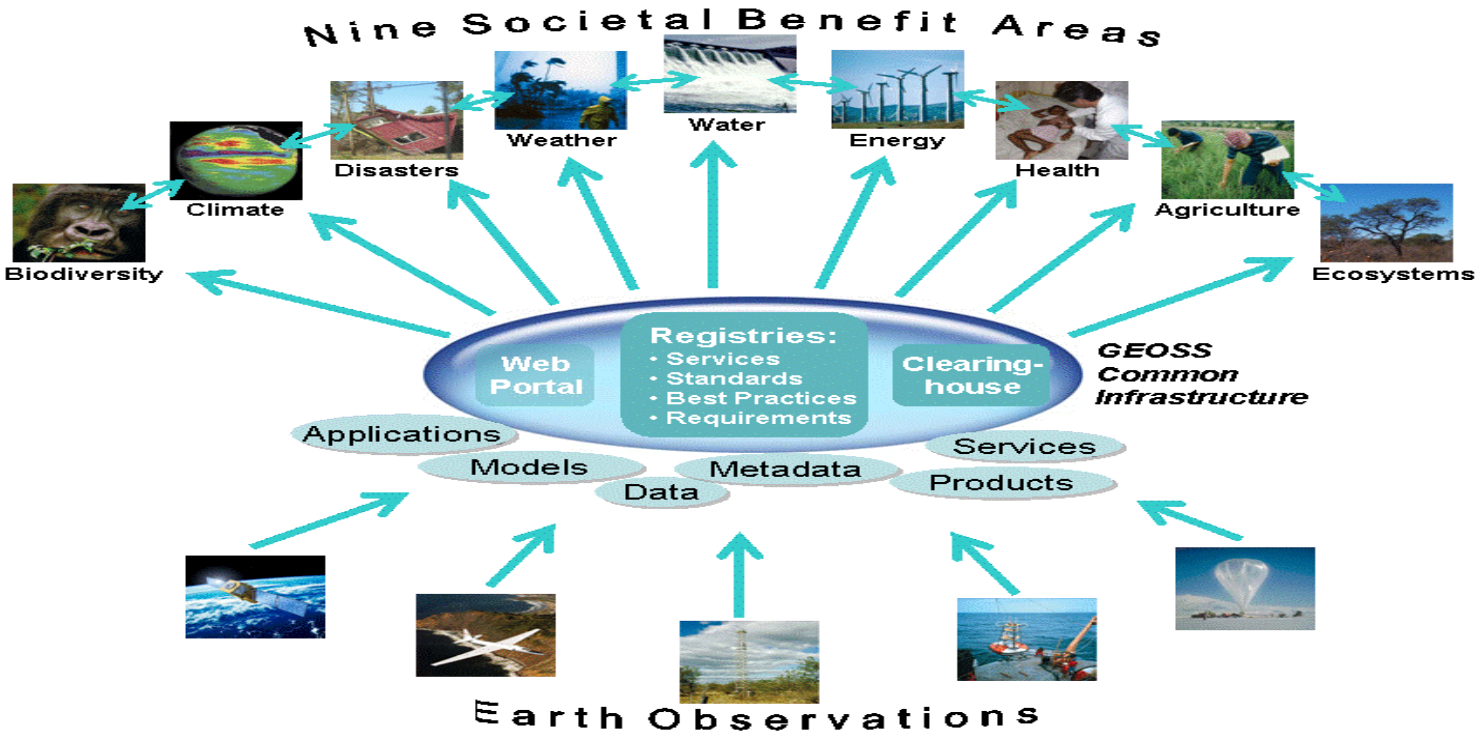
- **Project ID:** 283401
- **Project Type**: CP-CSA
- **Start Date**: 01.09.2011
- **Duration**: 36 Months
- **Website**: www.scidip-es.eu

- **Total Budget**: 7,721,082 €
- **EC Funding:** 6,599,992 €
- **Total funded effort in person/months**: 605
- **Coordinator: European Space Agency**
- **Contact Person**: Mirko Albani (ESA)

Project Consortium (17 partners):

Earth Science **Industrial** **Researchers**

Our context


Nine Societal Benefit Areas
Weather
Water
Energy
Disasters
Health
Climate
Agriculture
Biodiversity
Ecosystems
Registries:
• Services
• Standards
• Best Practices
• Requirements
Web Portal
Clearing-house
GEOSS Common Infrastructure
Applications
Models
Data
Metadata
Services
Products
Earth Observations

Earth Science can count 9 different data categories.

Each of them has its own:

- Data preservation policies and approaches (if available)
- Metadata and data formats
- Data description and semantic
- Different also inside each category depending on the owner

Data are used by different categories of users with different tools and techniques.

Products generated by users are often not duly preserved.

SCIDIP-ES Objectives

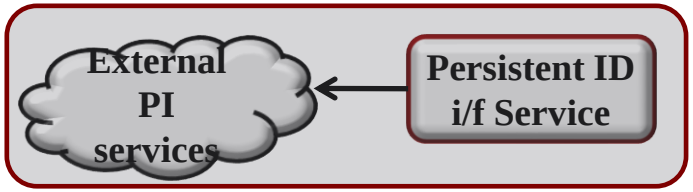
1) To develop and deploy generic and sustainable digital data preservation services and toolkits.
   - Validate and use them in the Earth Science domain as a start.

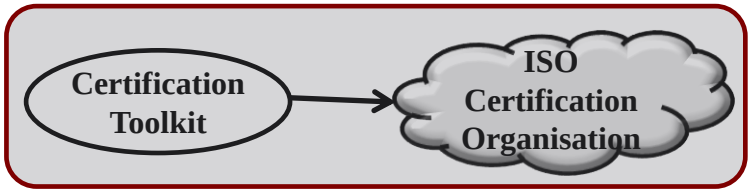
2) To harmonise data preservation policies and approaches, metadata and ontologies in the Earth Science domain:
   - Paving the way for the set-up of an harmonized and common approach for the Long Term Preservation of Earth Science Data.

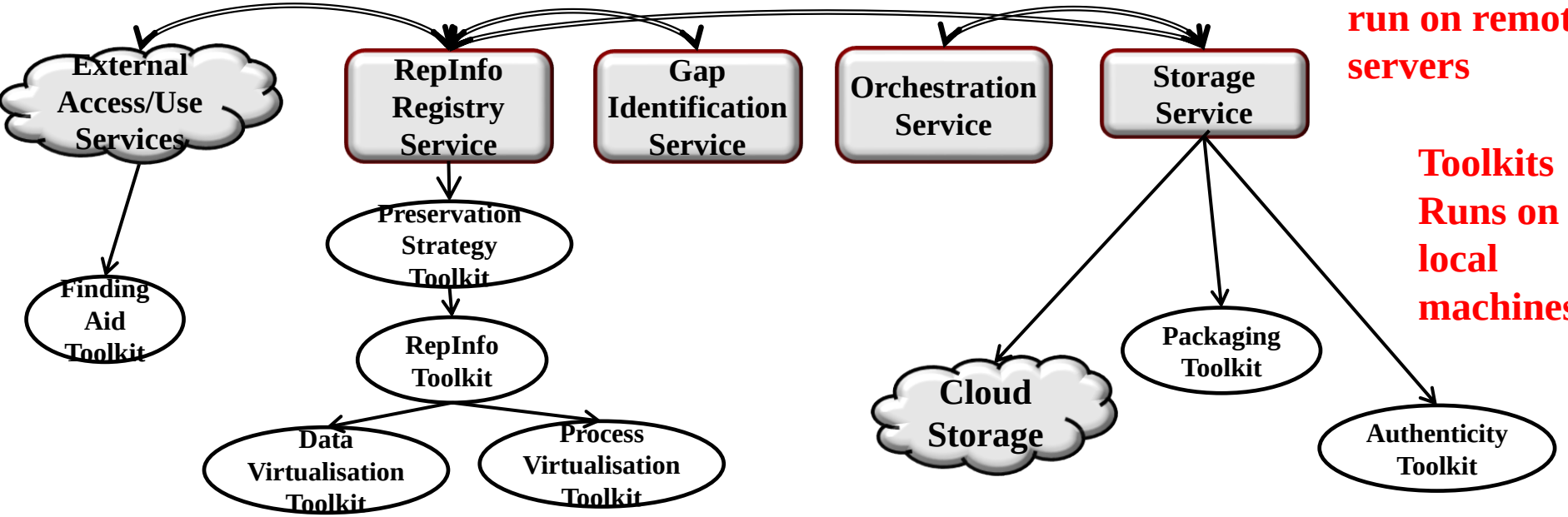
Earth Science Data Preservation status and needs

- Earth-Science data managers have individually dealt for decades with preservation and access to heterogeneous and complex scientific data.
- BUT need to:
  - **Improve effectiveness** of data preservation through the application of standard approaches and tools and **exploitation of latest Digital Preservation developments and achievements** (i.e. SCIDIP-ES services and toolkits). Reduce preservation threats.
  - **Ensure preservation of** data but also of all context/provenance/quality information: **Data Records, Processing Software, Documentation**.
  - **Harmonize preservation approaches** and policies and coordinate efforts to better support pressing needs from very sensitive applications.
  - **Enhance interoperability** and facilitate data usability and access by the scientific user community.

SCIDIP-ES Services & Toolkits







Services: run on remote servers

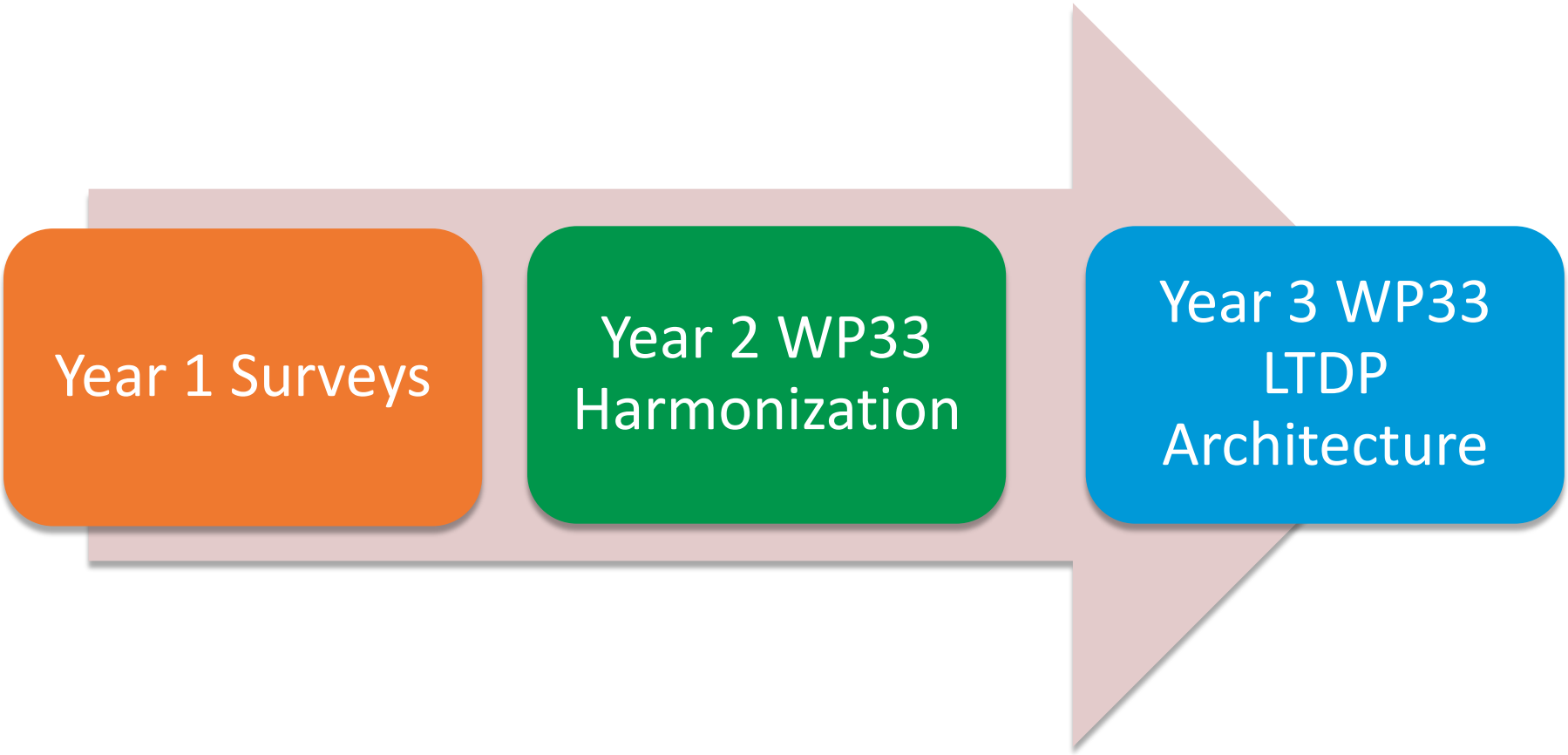
Toolkits Runs on local machines

- These SUPPLEMENT what repositories do (customised for repositories)
- Make it easier for repositories to do preservation – share the effort

What will SCIDIP-ES do for Earth Science?

- Provide state of the art data preservation services and toolkits.
- Raise awareness on data preservation issue and enhance dialogue in this context between ES data repositories:
  - Paving the way for the set-up of an Earth Science LTDP Framework
- Define common preservation policies and approaches.
- Propose path for harmonizing metadata, semantics, ontologies, and data access policies:
  - Facilitating data discovery and use.
- Define an Earth Science LTDP framework governance model and architecture.

Earth Science Activities Workflow



www.scidip-es.eu