


Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval 

The EO community vision for Long Term Data Preservation: Needs & technological, infrastructure, operational and political plans

Luigi.Fusco@esa.int

CASPAR Training Day for the Scientific Domain
Rome, 12-13 January 2009

CASPAR Training event - Jan 2009

 **Summary**

- EO community perspective
 - Example of needs
- Operational and political plans
- Technologies and infrastructure aspects

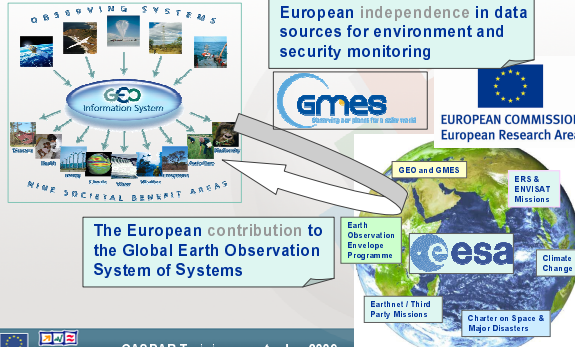
Presentation goals

Provide an overview of the needs of the Earth Science specific community for data long term archiving and preservation

CASPAR Training event - Jan 2009

The EO community big ambitions

GEOSS, GMES & ESA Living Planet



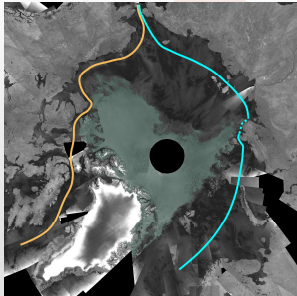
European independence in data sources for environment and security monitoring

The European contribution to the Global Earth Observation System of Systems

CASPAR Training event - Jan 2009

Examples of needs (1)

Lowest Arctic ice coverage in "history":
News published on ESA website on 14 September 2007



Northwest Passage open (orange line) and Northeast passage only partially blocked (blue line). The dark grey colour represents the ice-free areas, while green represents areas with sea ice.

CASPAR Training event - Jan 2009

Examples of needs (2)

Google search "Arctic ice coverage Envisat"
11 June 2008, most results are "copies" of ESA published data.



A philosophical question: which image is the original one? How to decide it? How to preserve its identity? (Some have also added information)

CASPAR Training event - Jan 2009

Managing Authenticity: critical issues

Integrity

- A resource has *integrity* when it is complete and uncorrupted in all its essential respects.
- The maintenance of the bit sequences is not always necessary, but the completeness of the 'intellectual form' is required, e.g. maintenance of colours in a map... the essential components must remain the same.

Identity

- A crucial point is that *identity* must be intended in a very wide meaning...uniquely identify it and distinguish it from any other resource ...

Authenticity Management Tools

- Provide *content and contextual information* relevant to authenticity, i.e. to the identity and integrity profile, along the whole preservation process.

CASPAR Training event - Jan 2009

Community challenge: global change

Space-based EO contributes significantly to global change monitoring

The IPCC Report 2007

- Model: Global temperature increase between +2.4 and 6.4 degrees until 2100
- Arctic: ice-free as of 2nd half of the century
- Permafrost: up to 90% melting until 2100, freeing high amounts of Methane gas
- Precipitation: decrease in arid regions and increase in wet areas
- Storms and surges: less in number but significantly stronger in intensity
- Gulf Stream: significantly weakened
- Sea level rise: up to 48cm until 2100 due to thermal expansion of water only

CASPAR Training event - Jan 2009

Recent periodical datasets

GlobCover: New portrait of Earth shows land cover as never before

300 m resolution!

Curator ??
preservation information ??

CASPAR Training event - Jan 2009

Sea Surface Temperature trends

Measurement of Residual Trends in Global Sea Surface Temperature

ERS / Envisat
Trend: +0.13°K / decade

AVHRR data
trend: 0.09°K / decade

University of Leicester

CASPAR Training event - Jan 2009

Sea level rise: improved models

The ENVISAT altimeter provides continuity to the measurements initiated in the early 1990

Sea level rise
Trend: +3 mm/yr

TOPEX:
Linear trend = 3.13 mm/yr
Annual (4.22 mm) removed

ERS-1:
Linear trend = 3.29 mm/yr
Annual (2.85 mm) removed

ERS-2:
Linear trend = 2.84 mm/yr
Annual (8.37 mm) removed

Courtesy of Remko Scharroo, NOAA/US

CASPAR Training event - Jan 2009

Keep track of the past for modelling the future

Sea level rise due to global warming

Sea level rise over the last century

Sea level rise scenarios for 2100

Preservation of data is not enough!

CASPAR Training event - Jan 2009

Earth Science next mission

ESA's Gravity Mission mission : GOCE

Gravity field map and improved global geoid models

Improved understanding of ocean circulation and energy distribution

Global unification of height systems

Future projects shall follow preservation standards

CASPAR Training event - Jan 2009



Community needs for preservation

In short:

- Understand historical data
- Make data and products heritage usable / comparable
- Maintain the available knowledge for future access
 - Transfer methodology and science achievements to other communities for interoperability
- EO data are increasing with a high rate...
 - Reduce costs for generation of preservation information

But today's ESA mandate for "data archiving" is limited to 10 years after mission operations completed



CASPAR Training event - Jan 2009 13




LTDP

European Earth Observation Initiative on Data Preservation – <http://earth.esa.int/gscb/ltdp/>

- Wanted by ESA and GSCB
 - ground Segment Coordination Body, representing Space Agencies in Europe (ASI, CNES, DLR, ESA, EUMETSAT) and Canada Space Agencies (CSA)
 - Very large no. of EO mission data from all involved agencies (ERS, ENVISAT, SPOT...)
- Building political momentum for long term funding
 - Proposal to be submitted to ESA Ministerial Conference (Nov 2008)
- First LTDP workshop organized by ESA in May 2008






CASPAR Training event - Jan 2009 14




Workshop Objectives

involving all European EO data owners/providers and archive holders

- Exchange information on LTDP policies and technical approaches;
- Collect feed-back and derive recommendations for a European LTDP common approach/policy;
- Identify sharing opportunities;
- assess the impact (benefits, drawbacks) of the proposed European LTDP common policy to each archive owner.






CASPAR Training event - Jan 2009 15




Cooperation scheme – conclusions (1)

- **LTDP must be seen in a very long term prospect**
 - should be intended as a permanent initiative
 - EO mission independent activity
 - with a (funding) timescale around 15-20 years.
- **Progressively cooperate to jointly perform:**
 - technology development, standardization activities;
 - operational solutions, data exploitation for cost-effective.
- **ESA to trigger and coordinate following steps**
 - toward the progressive LTDP System implementation






CASPAR Training event - Jan 2009 16




Cooperation scheme – conclusions (2)

- **All European EO Space data owners**
 - contribute to the European LTDP System through their infrastructure.
- **The LTDP Common Policy**
 - is to be intended as a Guideline,
 - full adherence to be pursued in the mid-long term.
- **Sustained through a cooperative (multi-source) programmatic and long term funding framework**
 - Multilateral Cooperation with multiple funding sources (e.g. European EO data owners).






CASPAR Training event - Jan 2009 17



Which space related data / information


- Archived data shall contain:
 - Raw data
 - Global or higher level products generated systematically
 - Metadata and browse images when generated
 - Spacecraft ephemeris information and Auxiliary data required to process the telemetry payload data
 - CAL/VAL databases (including processing/reference validation data sets).
 - Mission related documentation.
 - Processing algorithms description and processors needed to obtain mission

CASPAR Training event - Jan 2009 18

Operation plans

- Adoption of OAIS standard reference model
- Archives maintenance and data integrity
 - Periodical data migration
 - Data repackaging
 - multiple copies of the same data
 - common certification / security / procedures
- Reprocessing
 - Data reprocessing as needed with maintenance of history
- Data access and interoperability
 - inventory, metadata and browse maintenance
 - processing chains maintenance
 - Ensure on-line access and on-line delivery.
 - Evolution of systems
 - Standard interfaces
 - Harmonization of EO products specifications
 - Common development of technology




Long Term
Data Preservation


CASPAR Training event - Jan 2009

Relation with other key actors


• The EO world is an interesting, but narrow science community ...



Shared
Infrastructures





Alternative and
complementary
technologies for
preservation



Other communities
requirements

Many other projects / initiatives are to be compared

CASPAR Training event - Jan 2009



Ground European Network for Earth Science Interoperations

Digital Repositories



EC Grant Agreement no. 212073

INFRA-2007-1.2.1 • Scientific Digital Repositories
Duration • January 1, 2008 – December 31, 2009
Total EC funding • 4.4 M€

*For scientists to get access
and share seamlessly
observations of the Earth
System and derived
information and knowledge*

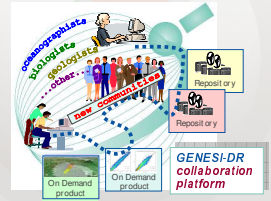


CASPAR Training event - Jan 2009

GENESI-DR projects goals

- To provide a base for (establishing) a world-wide e-infrastructure for Earth Science repositories
- To provide reliable, easy, effective, and operational access to a variety of data sources (space and non-space)
- To harmonise operations at key Earth Science data repositories

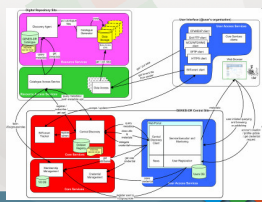
- To demonstrate effective **curation** and prepare the frame for **long term preservation**
- To validate capabilities to access distributed repositories for involving **new communities, including education...**
- To integrate new scientific and technological derived paradigms



CASPAR Training event - Jan 2009

Architecture Highlights


- Exploit and strengthen best practice in distributed data archiving, discovery, access and processing
- A set of services for indexing, searching, sharing and storing very large spatial data sets.
- Publication of data through their of metadata, their syndication between peers automatically
- Discovery of data relevant to an application, establishment of data usage / access rights
- Processing on user demand in open Grid environment
 - Pre-processing (avoid bulk data transfer)
 - Move algorithm to data



CASPAR Training event - Jan 2009

GENESI-DR “preservation” aspects

- Need to document “Preservation Description Information”
 - Response to system changes (in hardware/storage, operating system, ...)
- Semantics
 - Documentation of meanings and inter-relationships
- Need to be able to share the load
- Exploit OAIS (Open Archive Information System) concepts to the fullest extent and **INSPIRE Directive**
- Need to preserve information and knowledge
 - Data and Documents...
 - more Representation Information



CASPAR Training event - Jan 2009 <http://www.casparpreserves.eu>

Curation aspects in INSPIRE – GeoSpatial community

INSPIRE DIRECTIVE
Decision 2007/177/EC of the European Parliament and of the Council of 14 March 2007 establishing an infrastructure for spatial information in the European Community (INSPIRE) was published in the official Journal on the 25th April 2007. The INSPIRE Directive entered into force on the 16th May 2007.

1.3	Resource title	dc:title
1.2	Abstract	dc:abstract
1.3	Resource type	md:topic
1.4	Resource locator	md:link
1.5	Unique resource identifier	md:uuid
1.6	Coupled resource	md:link
1.7	Resource language	md:lang
2.1	Topic category	md:topic
3.1	Keyword value	dc:subject
3.2	Organized controlled vocabulary	md:vocabulary
4.1	Geographic bounding box	geo:box
5.1	Temporal extent	time:span
5.2	Date of publication	md:pubdate
5.3	Date of last revision	md:modified
5.4	Date of creation	md:created
5.5	Alternate reference	md:ref
6.1	Lineage	md:isppp
6.2	Spatial resolution	md:resolution
7.1	Specification	md:specification
7.2	Degree	md:certainty
8	Conditions applying to access and use	md:access
9	Limitations on public access	md:restrictions
10.1	Responsible party	md:contactname, md:contactbox
10.2	Responsible party role	md:role
11.1	Responsible point of contact	md:contactname
11.2	Responsible data	md:modified
11.3	Responsible language	md:modified

3.2 Originating controlled vocabulary = Semantic RepInfo

6.1 Lineage = Provenance

7.1 Specification = RepInfo

8 Conditions applying to access and use = DRM, Access Control

2009 25

In summary

- EO community perspective – Example of needs
- Operational and political plans
- Technologies and infrastructure aspects

Thank you

26

CASPAR Training event - Jan 2009