



The ESA CASPAR Scientific Testbed and the combined approach with GENESI-DR

S. ALBANI

(ACS c/o ESA-ESRIN)

Sergio.Albani@esa.int

PV2009, 1-3/12/2009, Madrid



Information Society
and Media





SUMMARY

- ESA, CASPAR and Long Term Data Preservation
- ESA CASPAR testbed
- CASPAR & GENESI-DR combined approach





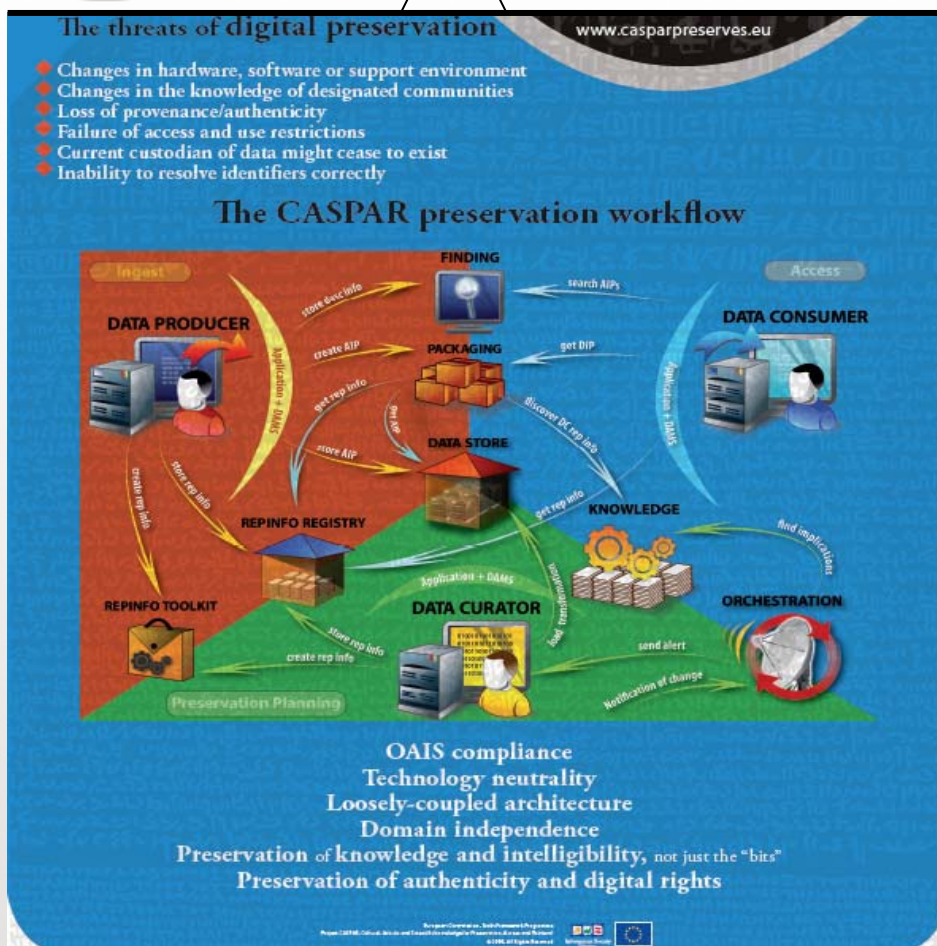
ESA and EO introduction

- ESA users worldwide have access to ~4 PB of EO data
 - EO data provide global coverage of the Earth
 - Data volumes are increasing dramatically
 - Large requirements for accessing historical archives
- This unique dataset has to be preserved!
 - ESA is promoting a European EO LTDP Strategy
 - ESA is involved in several international preservation activities





CASPAR



- *Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval*
- CASPAR is an Integrated Project co-financed by EU within the Sixth Framework Programme (Priority IST-2005-2.5.10, "Access to and preservation of cultural and scientific resources").
- CASPAR has built a framework to support the end-to-end preservation lifecycle for digital information, based on the OAIS reference model, with a strong focus on the preservation of the knowledge associated with the data.



Duration: April 2006 – November 2009





ESA role in CASPAR

- ESA participation to CASPAR was mainly driven by the interest in:
 - consolidating and extending the validity of the OAIS reference model, already adopted in several internal initiatives (e.g. SAFE);
 - developing preservation techniques/tools covering not only the data but also the knowledge associated with them in order to maintain the scientific capabilities of ES data users.
- CASPAR Scientific Testbed
 - ESA user and data/infrastructure provider
 - ACS technical side of the testbed implementation





Testbed focus

- Testbed scenarios have been implemented taking into account the current ESA archives and the European EO LTDP Common Guidelines
- Strong focus on:
 - knowledge management and preservation
 - data accessibility/usability
 - preservation of higher level data, processing capabilities and science applications

- *Archived data (in the form of **AIPs**) shall contain all the elements necessary to be accessed, used, **understood and processed** to obtain mission products to be delivered to users (in the form of DIPs)*
- ***Provide and maintain mission products generation capability** (systematic or through ordering) from AIPs to DIPs including the **processing chains***
- *Allow **information extraction** from low-level EO products and information preservation through supporting chainable information based services.*
- *Adopt a common standard reference model for the archives (ISO 14721 - **OAIS standard**)*





Testbed goals

Development of a complete 100% CASPAR components based preservation system (ESA CASPAR System)

- supporting data providers in the **preservation of the users capabilities to process data** using appropriate knowledge
- providing basic archiving features as Ingest, Access, Retrieve AND:
 - Knowledge preservation
 - RepInfo creation and appropriate browsing
 - User Communities profiling
 - OAIS compliance
 - On demand generation of data





Testbed activities

The ESA testbed has covered:

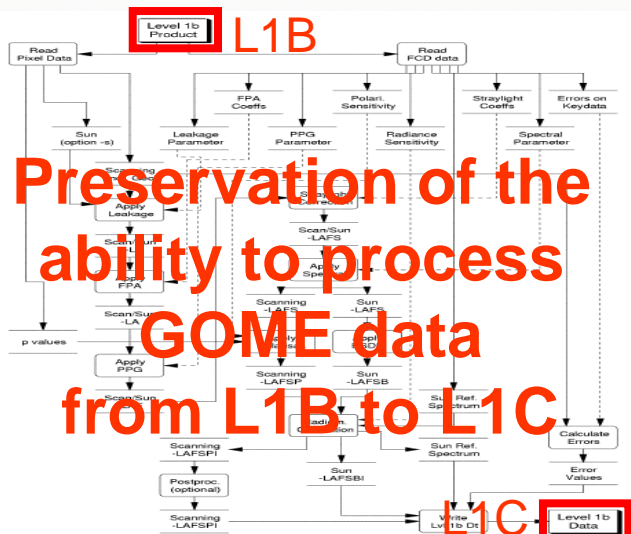
- the setup of the framework in ESA-ESRIN;
- the definition and collection of a significant sample of a whole processing chain dataset;
- the conversion of data from the native format to a OAIS compliant format;
- the analysis of ontologies to describe and preserve scientific workflows (e.g. the applicability of CIDOC CRM on scientific data);
- the generation of appropriate Representation Information, Descriptive Information, Knowledge Modules and Scientific Community profiles;
- the implementation of a 100% CASPAR-based archiving system;
- the ingestion and the retrieval (through a profile-based access) of data and related RepInfo;
- the coping with some long term data preservation problems by using only CASPAR components, methodology and tools.





Testbed Dataset

The ESA selected dataset for the CASPAR scientific testbed consists of data from **GOME (Global Ozone Monitoring Experiment)**, a sensor on board the ESA ERS-2 (European Remote Sensing) satellite



Preservation of the ability to process GOME data from L1B to L1C

GOME L1 products

L1B→L1C processor

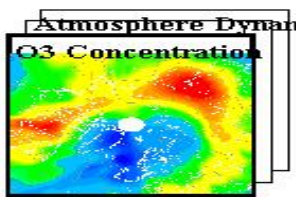
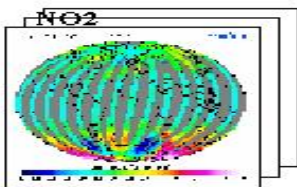
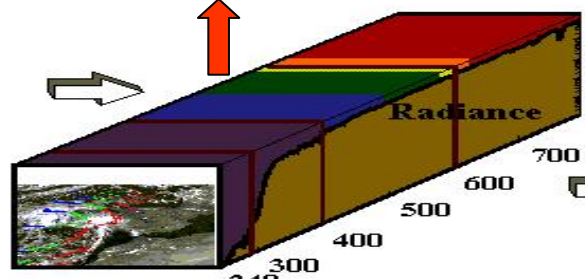
L1B→L1C source code

The Ozone
The ERS-2 satellite
The GOME sensor

ERS-Products.pdf
ProductSpecification.pdf
PSD.pdf
license.doc
disclaimer.pdf

readme_1st.doc
readme.doc
lease_I01.doc
er_manual.pdf
vhouse_I01.doc

The C Bible
The OS Bible



Level 0
Raw Data

Level 1
Calibrated Radiances

Level 2
Atmospheric Trace Gas

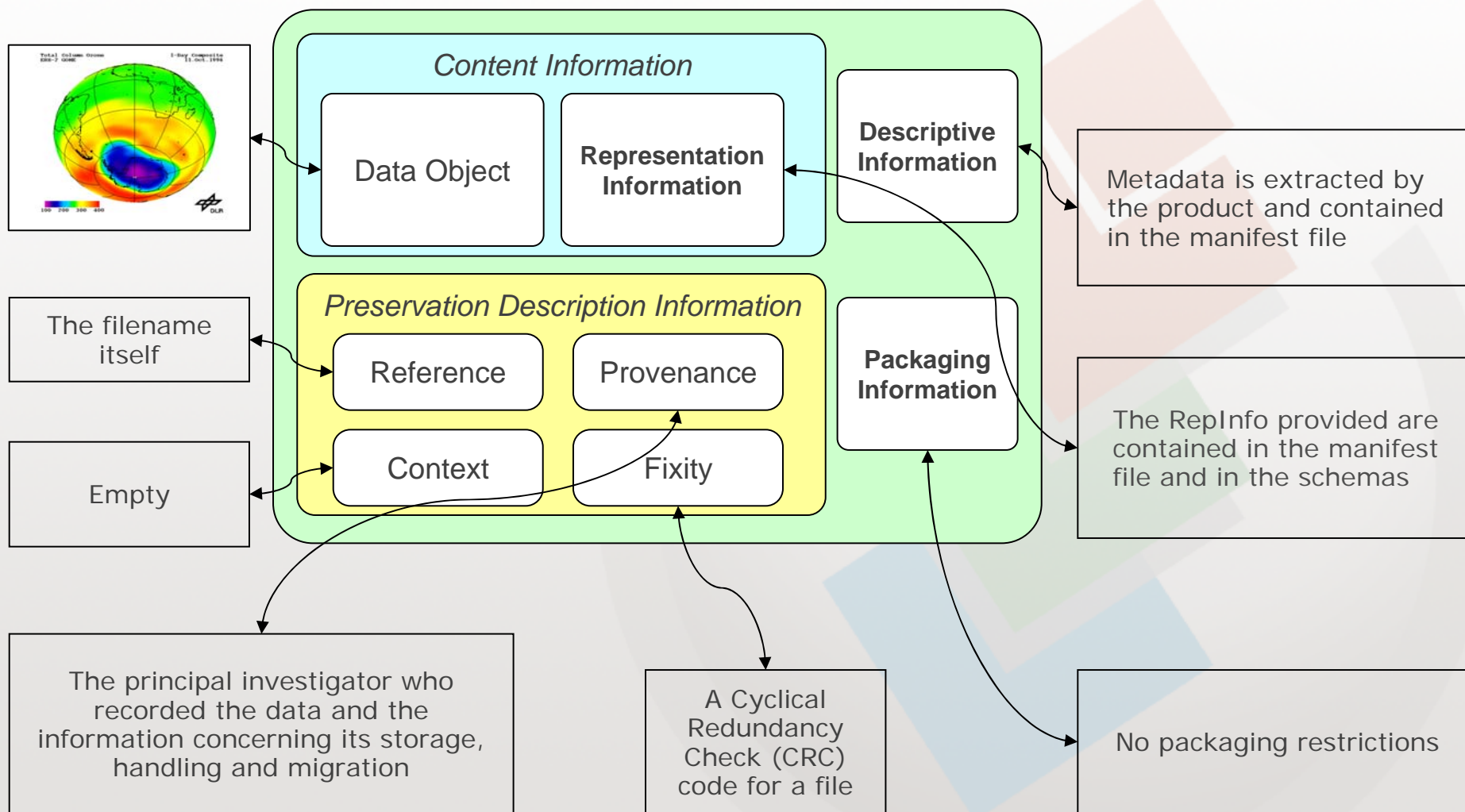
Level 3
Global Maps

The steps of GOME data processing



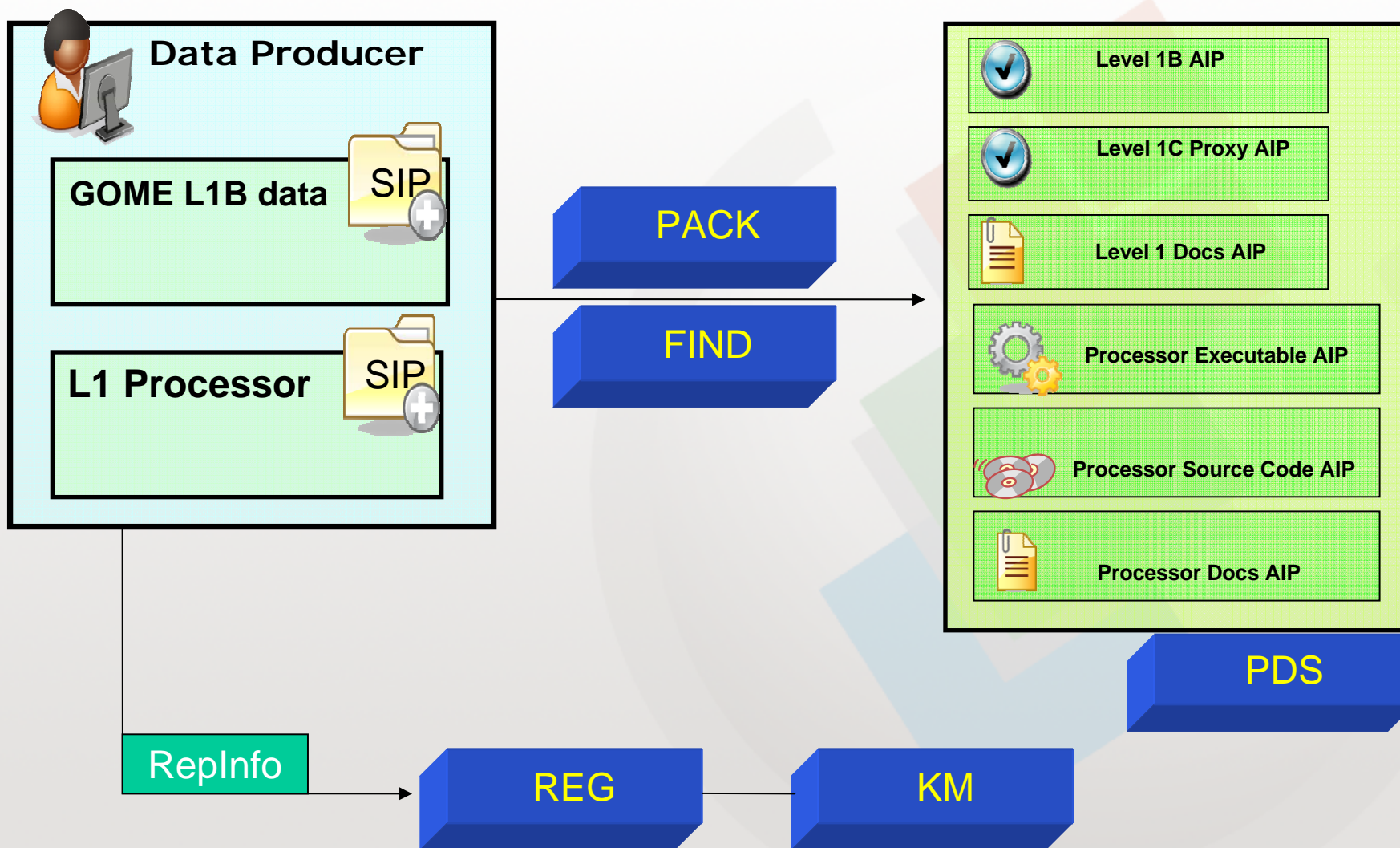


Ingested AIP (OAIS compliant)



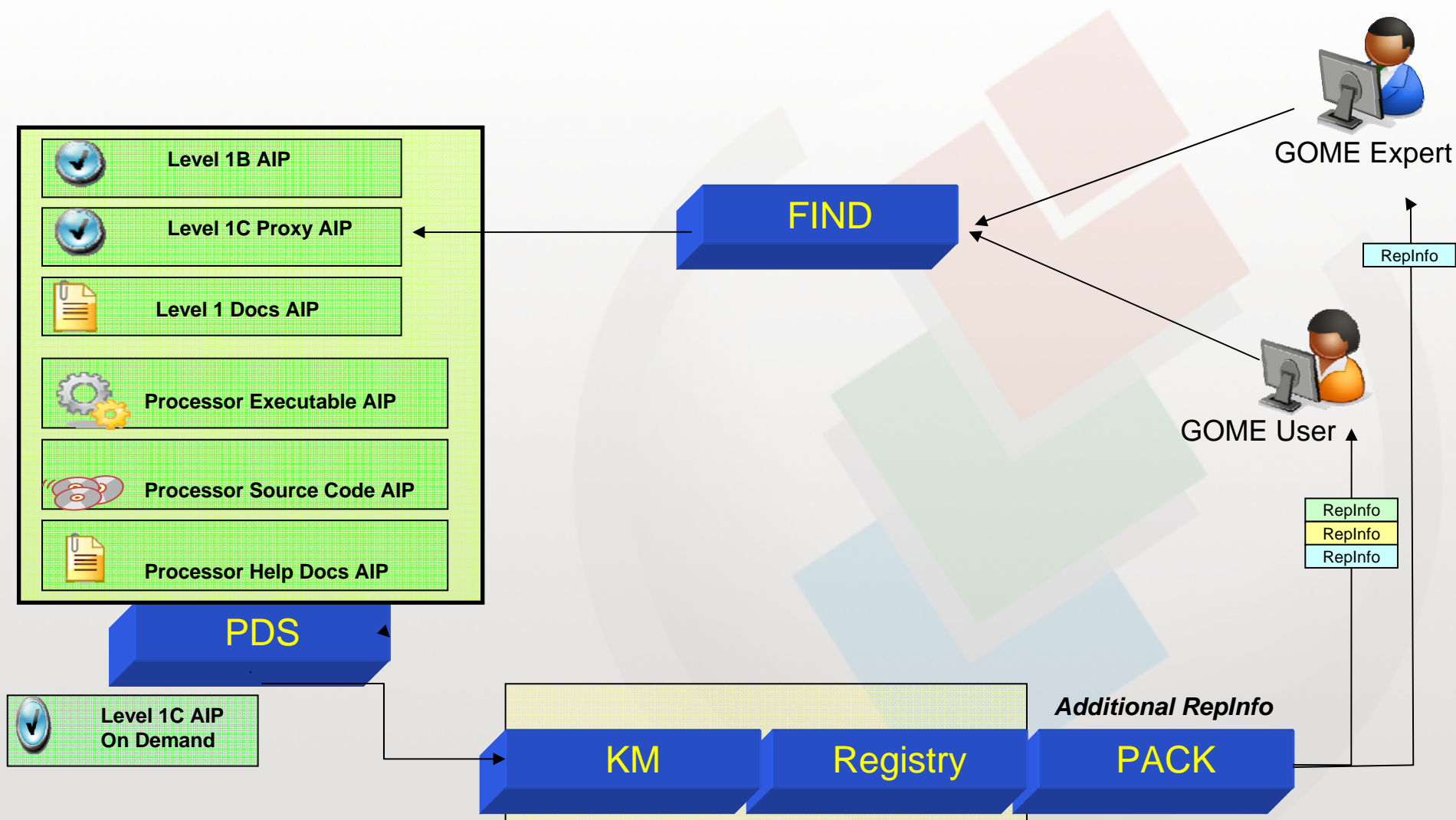


Ingest phase



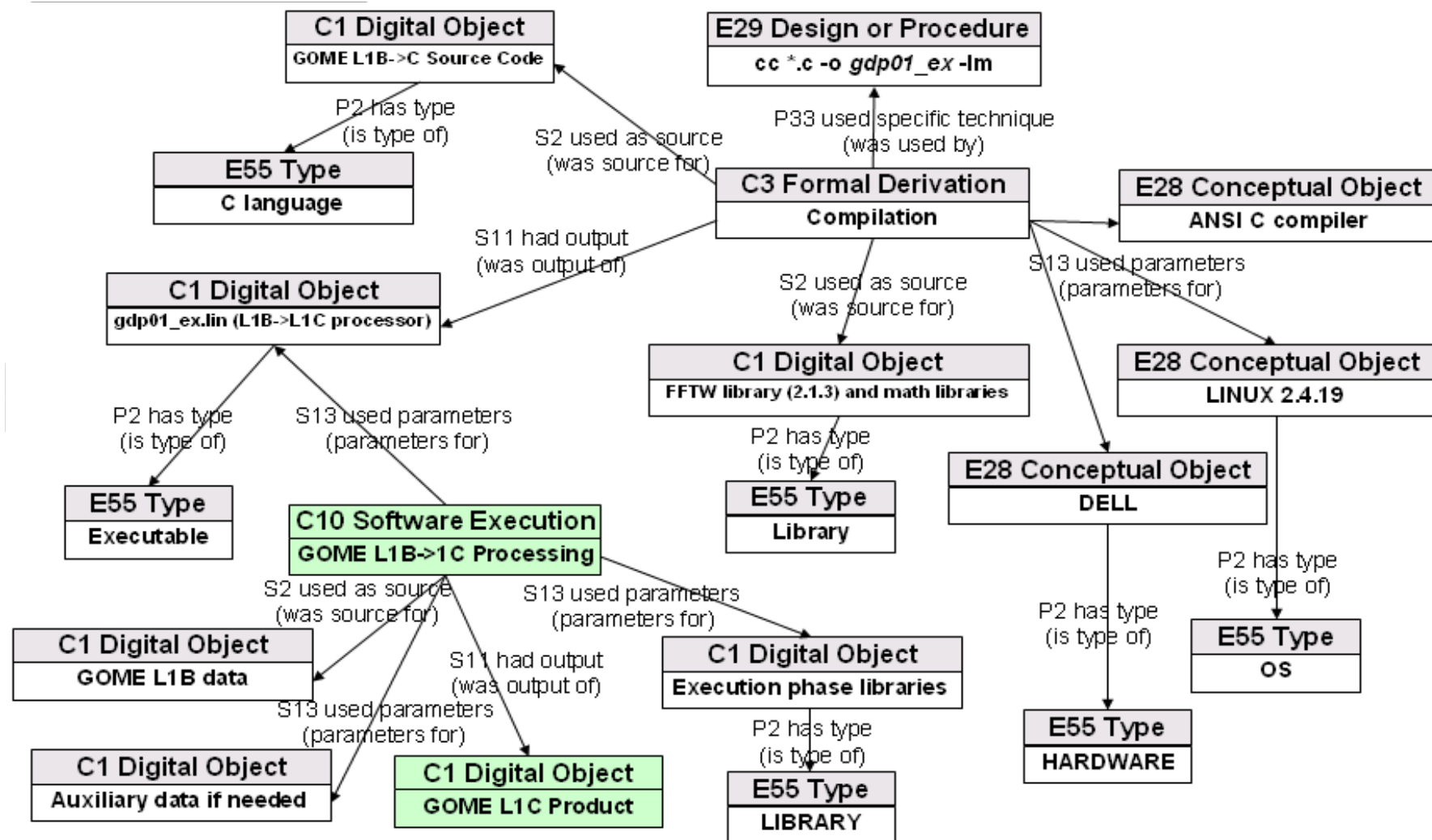


Search and Retrieve phase



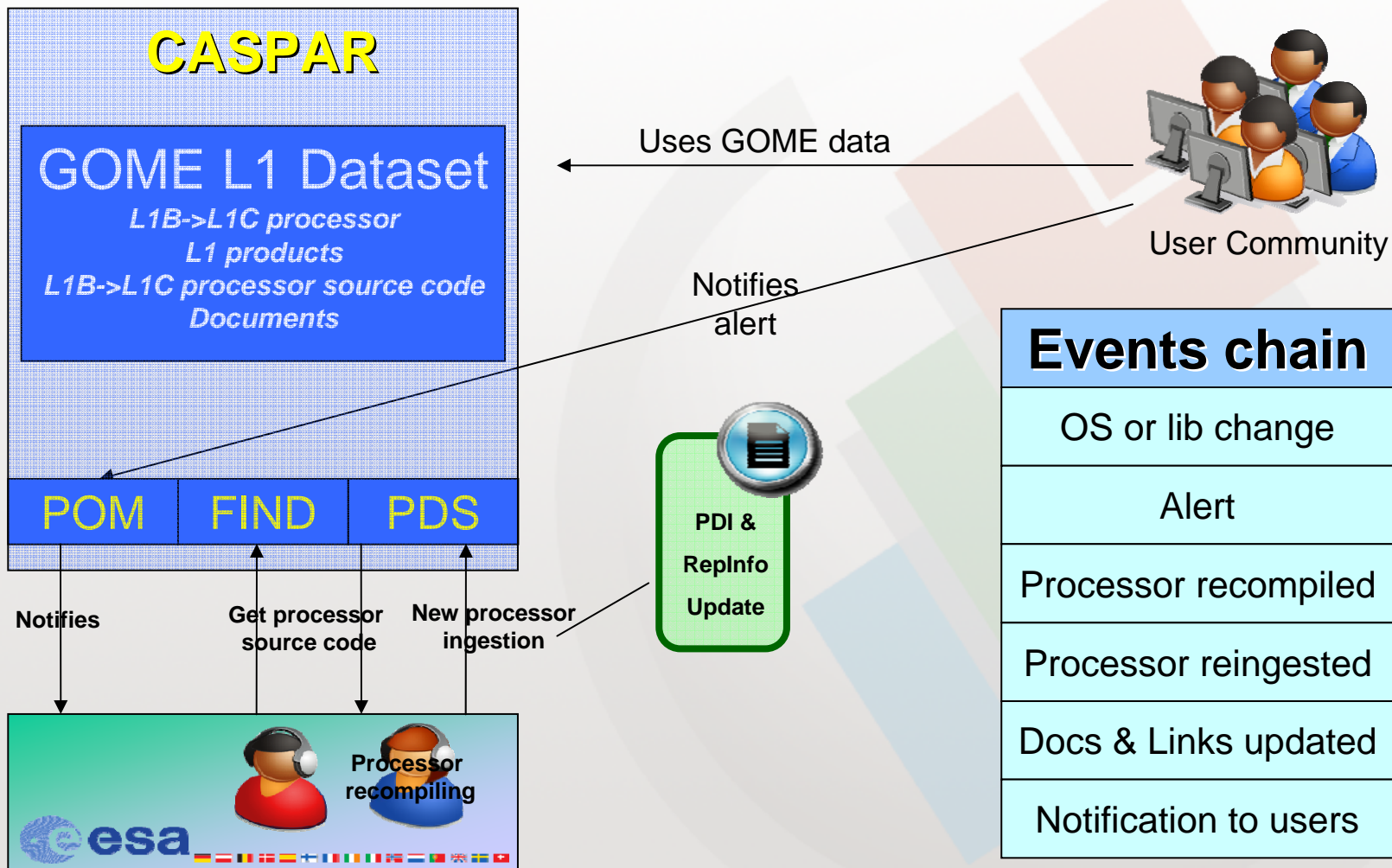


Underlying ontology





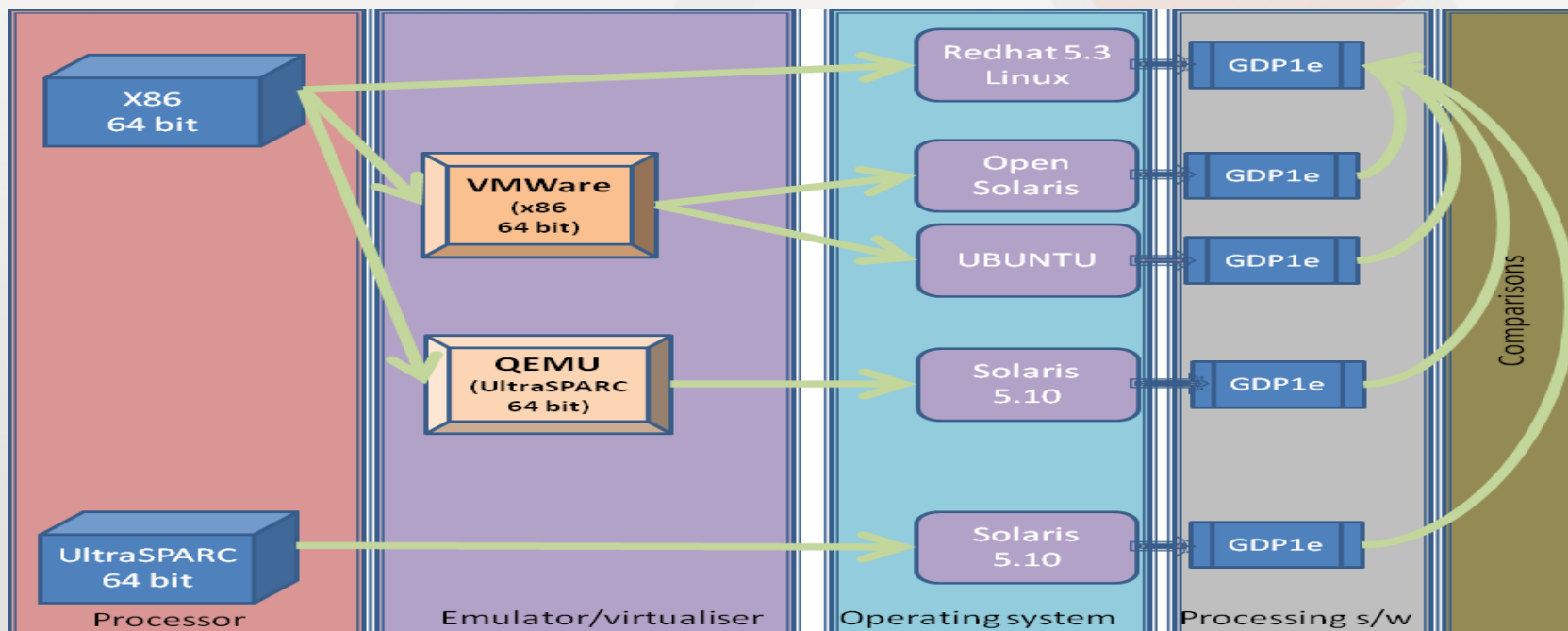
Preservation process (*update phase*)





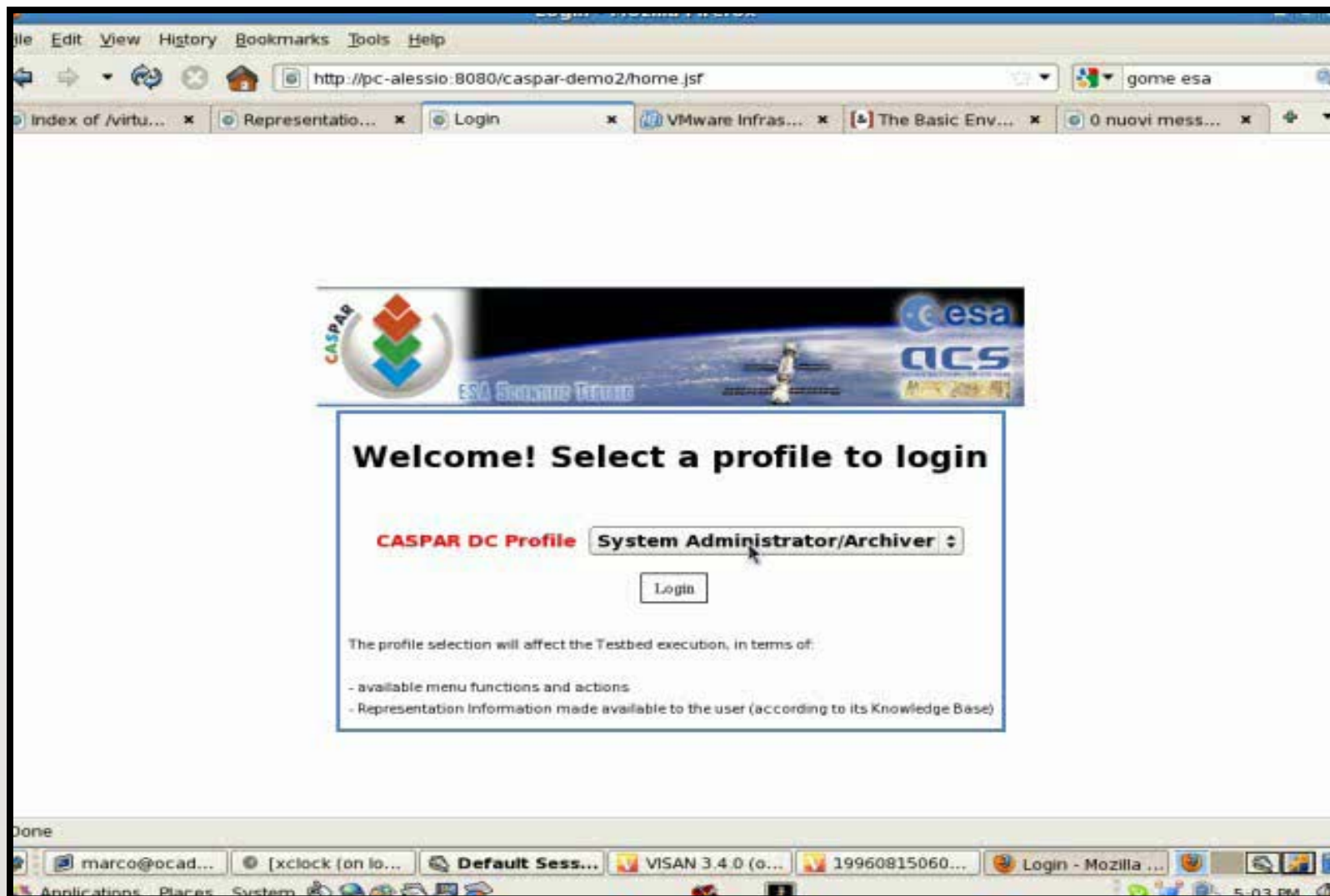
Testbed Validation

- Change in Software (new release of FFTW library needed to compile the processor)
- Change in Environment (migration from obsolete LINUX operating system to the more used SUN SOLARIS)





ESA CASPAR System DEMO



<http://caspar-nas.esrin.esa.int:9999/caspar-demo2>





Benefits and major outcomes

- Framework validation (CASPAR components are suitable for preservation of ES data)
- Lesson learnt
 - preservation of knowledge associated to data
 - preservation not only of data but also of data processing
 - best practices to cope with long term data preservation problems by using OAIS model real applications
- Main outcomes
 - development of a 100% CASPAR components based framework (ESA CASPAR System is available for further enhancement/testing and for users and data owners/providers willing to see a practical approach to preservation using CASPAR solutions)
 - demonstration of the suitability of CASPAR solutions for applications in the Earth Science field (in the ESA EO Ground Segments infrastructure)
 - integration with GENESI-DR





GENESI-DR



GENESI-DR

INFRA-2007-1.2.1 : Scientific Digital Repositories

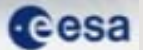
Duration : January 1, 2008 – April 30, 2010

Total EC funding : 4.4 M€



EC Grant Agreement no. 21 2073

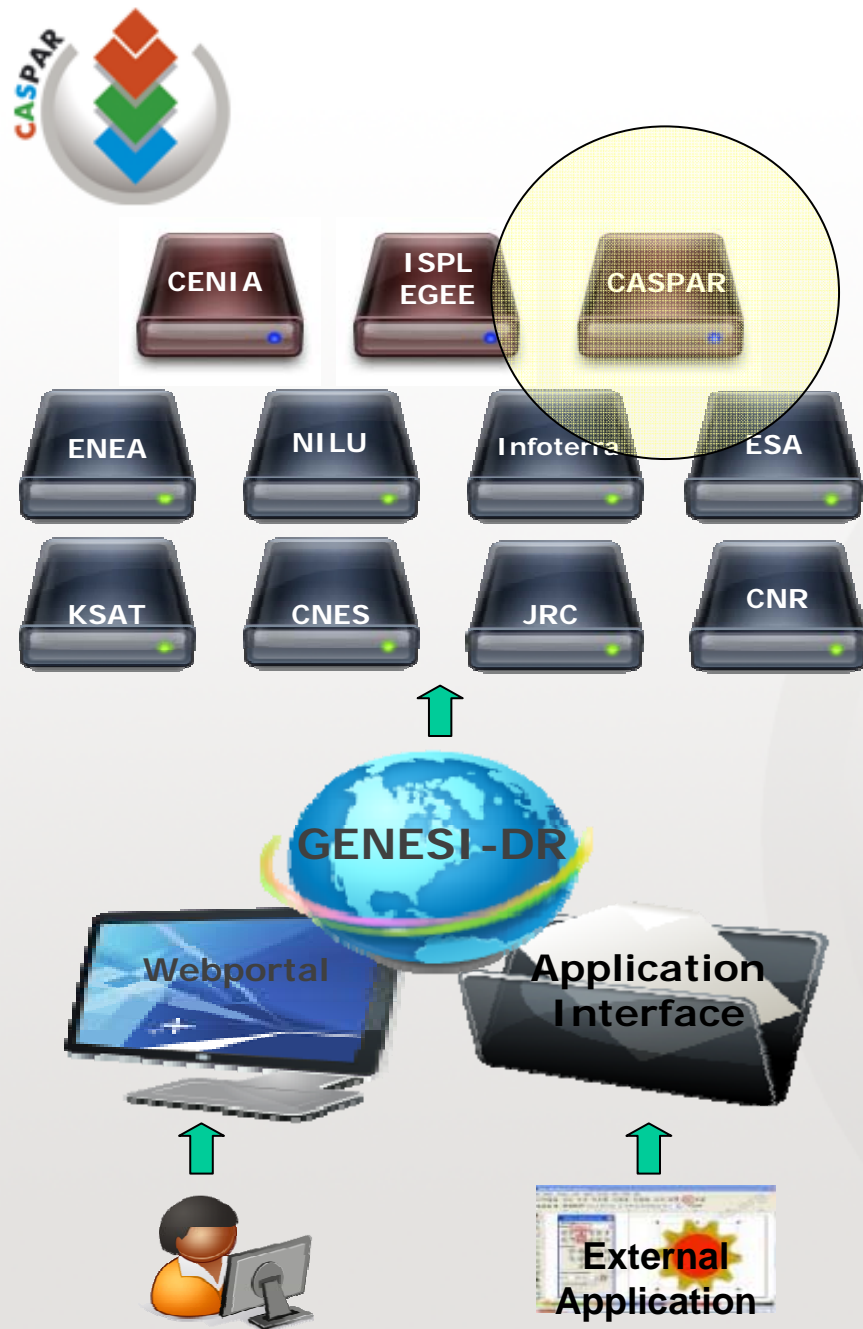
- *Ground European Network for Earth Science Interoperations – Digital Repositories*
- GENESI-DR is a federation of Digital Repositories (DR) dedicated to Earth Science
- GENESI-DR provides to users/applications open access to different European Earth Science Digital Repositories through the same interface.



Information Society
and Media



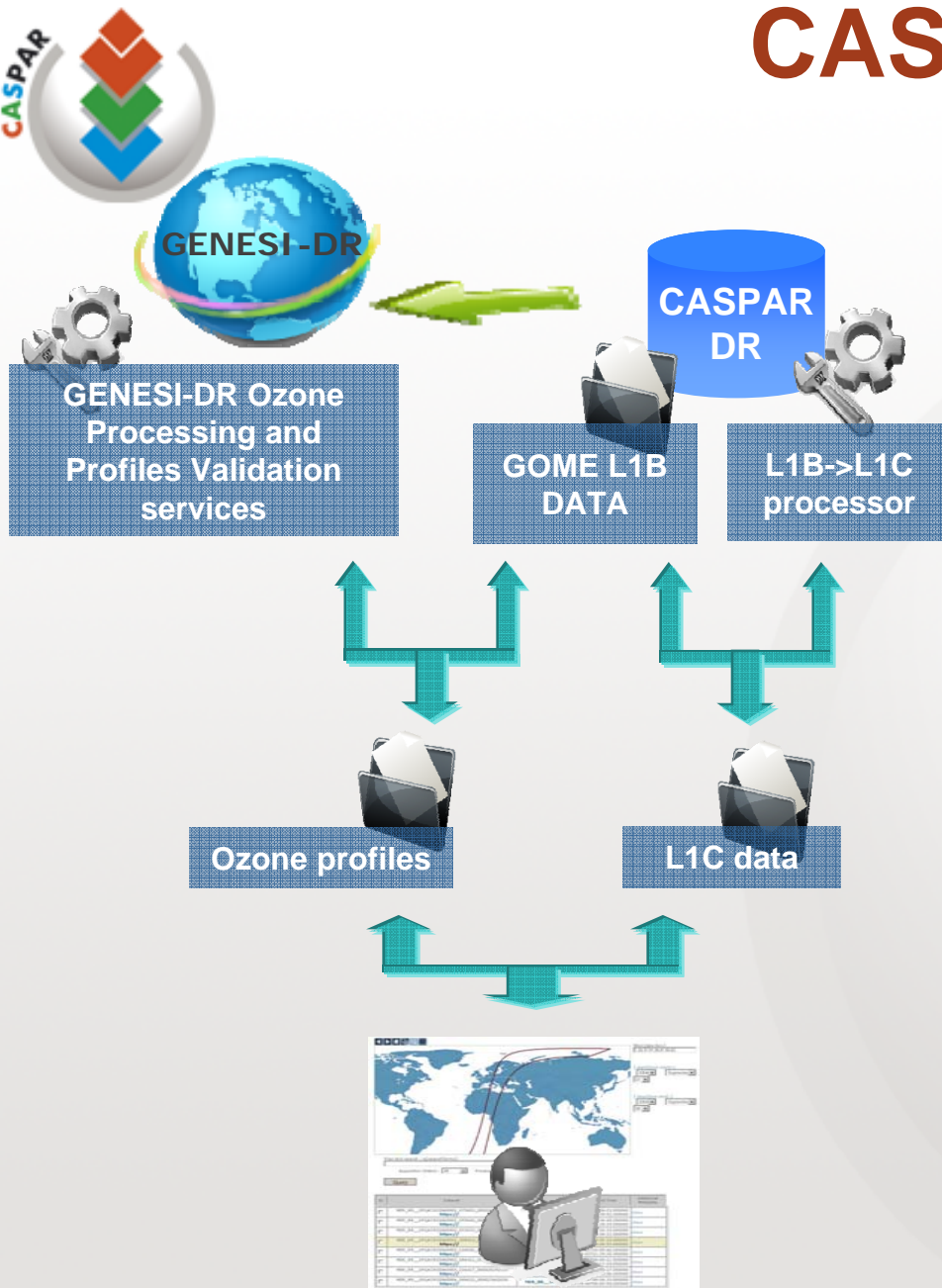
CASPAR & GENESI-DR



- CASPAR will benefit from the GENESI-DR services to validate in a more complete form its data preservation framework in the Earth Science domain
- GENESI-DR Research Infrastructure will demonstrate its ability to adopt data preservation and curation mechanisms defined in CASPAR.
- The integration of the ESA CASPAR System in the GENESI-DR infrastructure will promote the “CASPAR preservation model” in a wide community sharing the ESA CASPAR experience with other ES stakeholders
- We are evaluating how to evolve CASPAR and GENESI-DR to respond to new requirements in the ES community



CASPAR & GENESI-DR approach



1. GENESIfication of a CASPAR-based DR (ESA CASPAR System)
2. Development of services accessible through GENESI-DR to estimate vertical profiles of ozone or generate L1C data using processing software and data both preserved in CASPAR
3. Allow users to preserve their processing results in CASPAR
4. To return profile-based Representation Information to GENESI users
5. To define a strategy for propagating CASPAR features to other interested GENESI DRs.





CASPAR & GENESI-DR

DEMO


A single access point to Earth Science data Eliana Li Santi (Proxy Certificate)

→ GENESI-DR

Home | Logout | Blog | F.A.Q. | Partners | Services | My folder | My Profile | My Virtual Organization

We're querying one access point :
✓ caspar_gome_lv1 at <http://dr-site.genesi-dr.eu>
Info taken from http://dr-site.genesi-dr.eu/genesi/ers2_gome/caspar_gome_lv1/description/

These queries have the following common elements:



Free text search :

Search Digital Repositories using OpenSearch

bbox{geo:box} | -5.14,15.43,51.43,59.14

| start{time:start} | | 1996 | January | 01

| stop{time:end} | | 1997 |

- November
- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December

 | 16

Query

ID	Metadata
----	----------





THANK YOU!!!

ESA CASPAR TEAM:

- Luigi Fusco,
- Sergio Albani,
- Pasquale Renna

ACS CASPAR TEAM:

- Ugo Di Giammatteo,
- Fulvio Marelli,
- Marco Fulcoli,
- Alessio d'Innocenti

ESA GENESI-DR TEAM:

- Roberto Cossu,
- Eliana Li Santi

www.esa.int



www.acsys.it



www.genesi-dr.eu

