

GRANT AGREEMENT: 601138 | SCHEME FP7 ICT 2011.4.3

Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics [Digital Preservation]



Guide lines for observational big data  
preservation and reuse from the PERICLES  
project.

Christian Muller, B.USOC

# PERICLES Motivation

- Observation that digital objects and related metadata are generated as a

**continuum within a changing environment**



**Change might lead not only to loss of access or functionality, but to loss of meaning and understanding of information**

# What is PERICLES?

Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics

- 4 year FP7 EU-funded project addressing the challenge of long-term access to digital content in continually evolving environments (1 Feb 2013 –1 April 2017)
- With a consortium of 11 Partners



# Focus on change

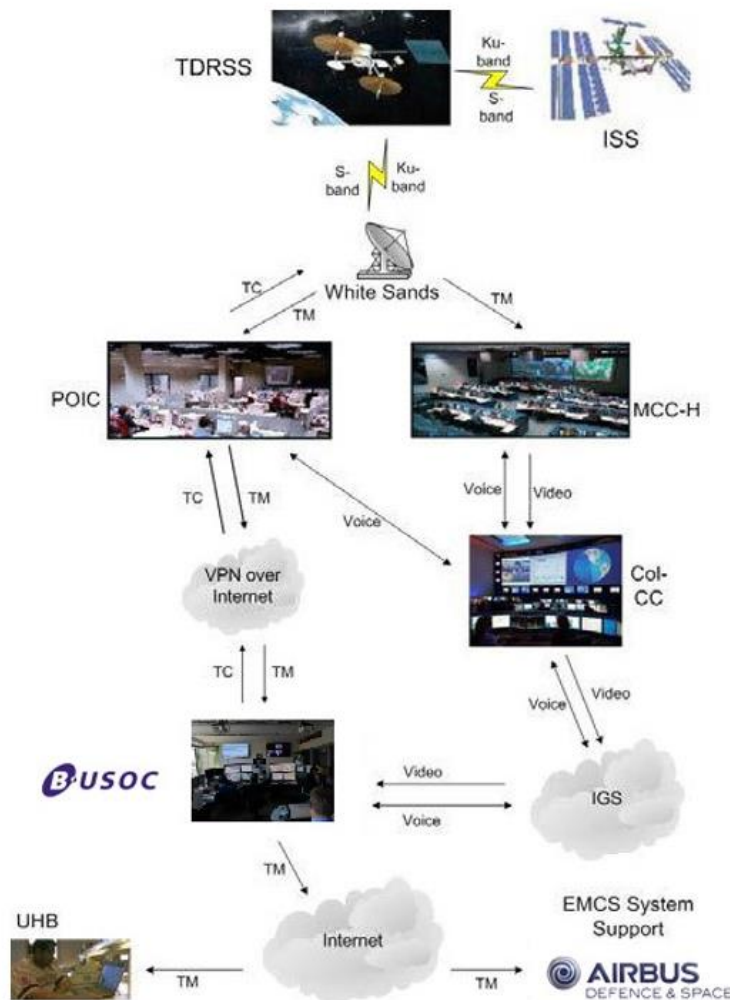
- **Long-term sustainability** requires us to address change more broadly



- Our basic hypothesis to be investigated during the project as the conceptual solution to dealing with change:

Preservation-by-design

# A space experiment: SOLAR on the ISS



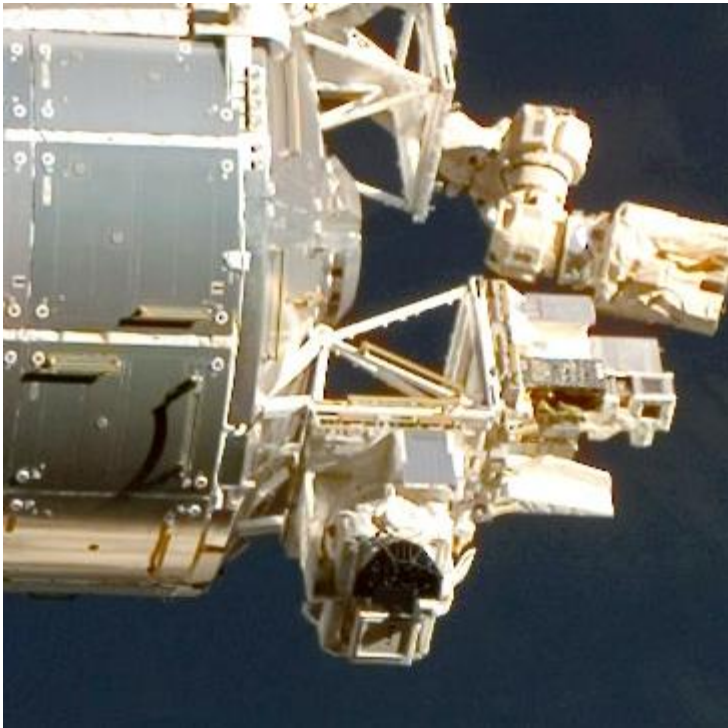
## What is the role of B.USOC in ISS data flow?

B.USOC manages the experiments, transmits the data requested by the scientist to the UHB and keeps a data repository according to ESA policy.

The data transmitted by ColCC to the USOC's is in the CCDS format (consultative committee for space data standards), it is in packages regrouping all experiments assigned to a USOC.

**The scientist is relegated to his « User Home Base »**

# SOLAR



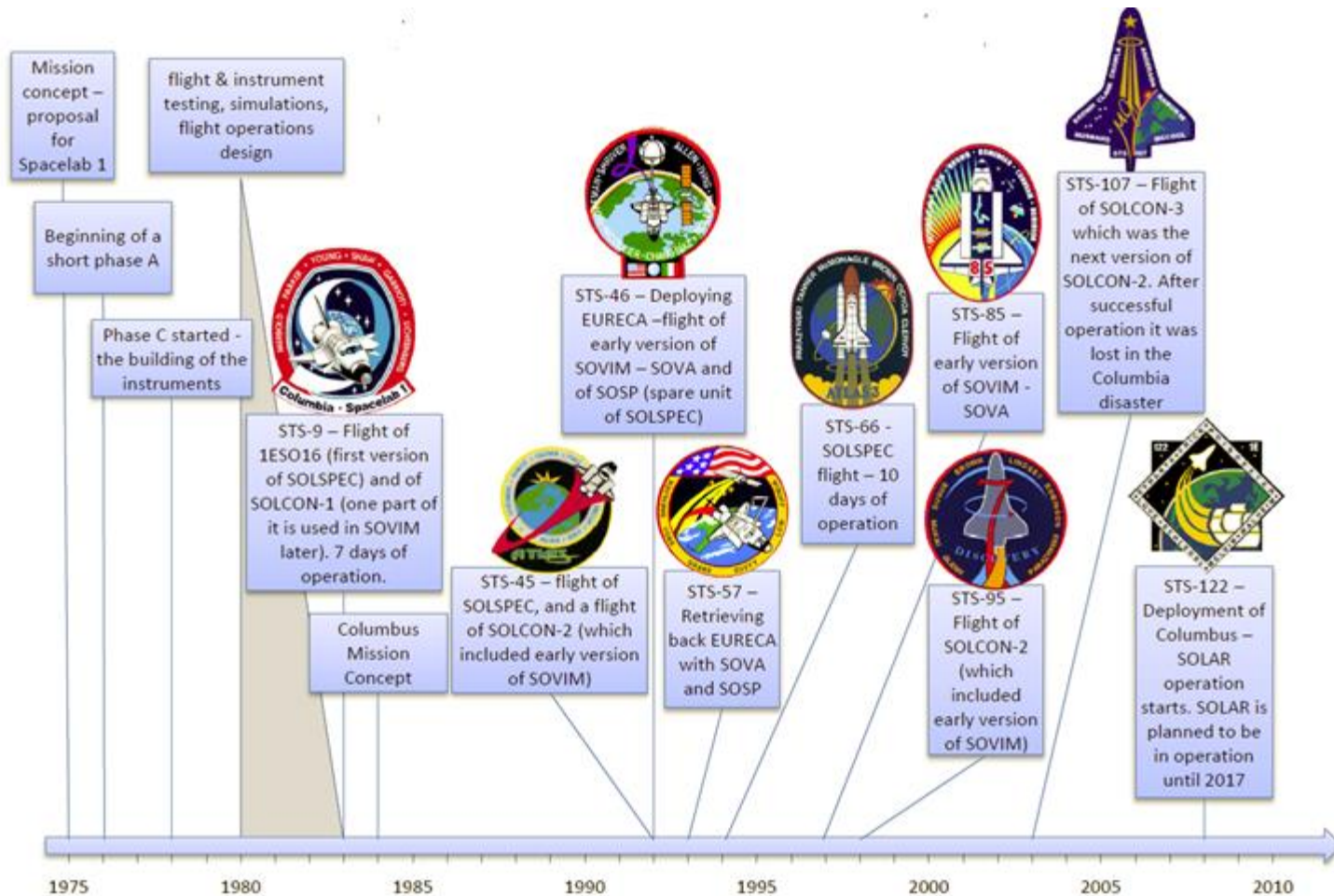
- ▶ The SOLAR package on the COLUMBUS module of the ISS.
- ▶ Three instruments: SOLSPEC, SOLACES and SOVIM .
- ▶ Operated by B.USOC from 2008 to January 2017 at the benefits of ESA and the Pi's.

# SOLAR as an example of change of semantics.

- ▶ Until around 1976: total energy input to the surface of the earth: solar constant.
- ▶ Space age: discoveries of possible variations: total solar irradiance.
- ▶ Since 1943, first through sporadic balloon and rocket observations, than from space, discovery of short term variabilities in the UV.
- ▶ Comparison with proxies: long term trends, at the same time: short time scale: link to space weather.
- ▶ Different paradigms dominate the current study of the solar data.

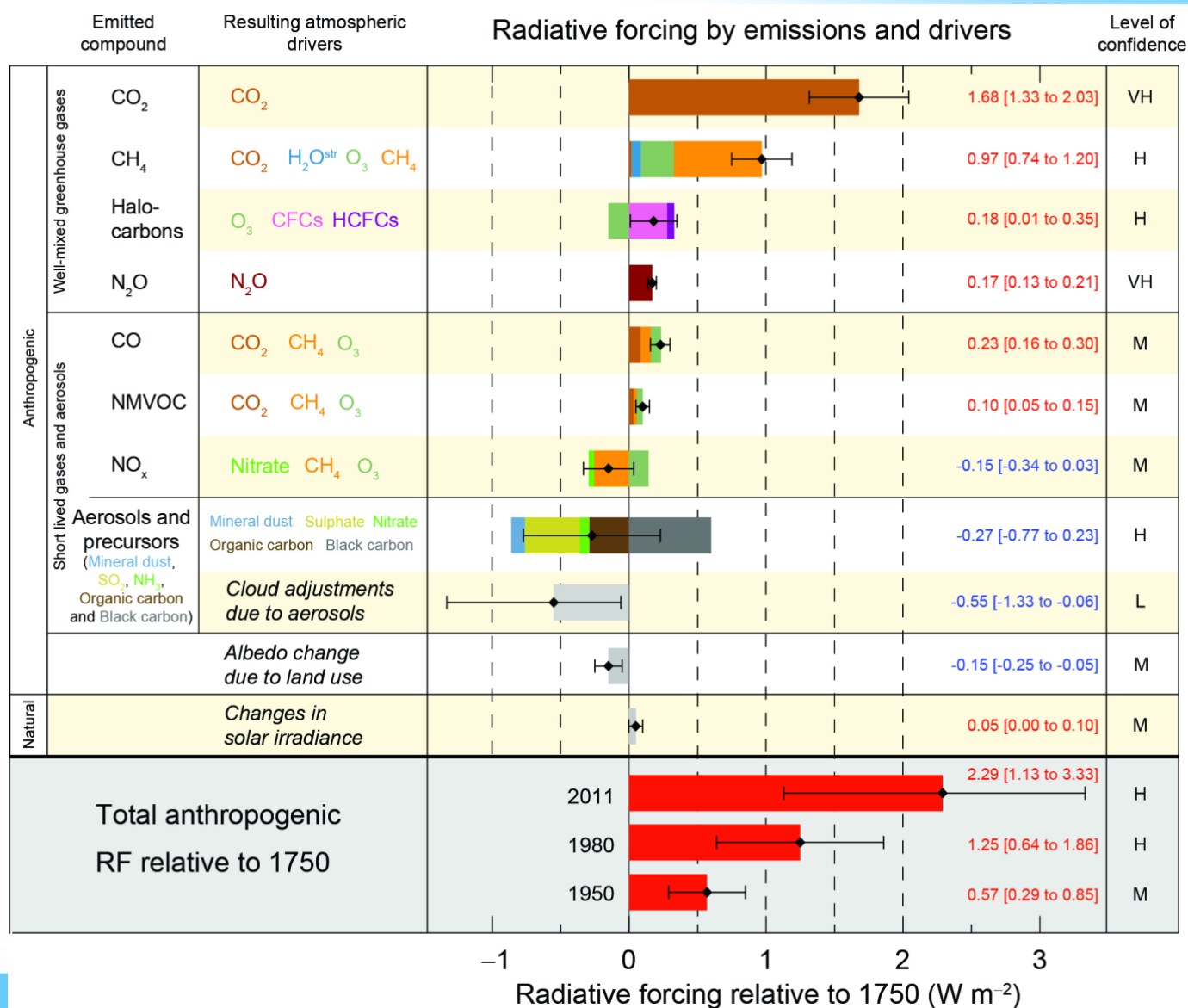


# History of the solar package



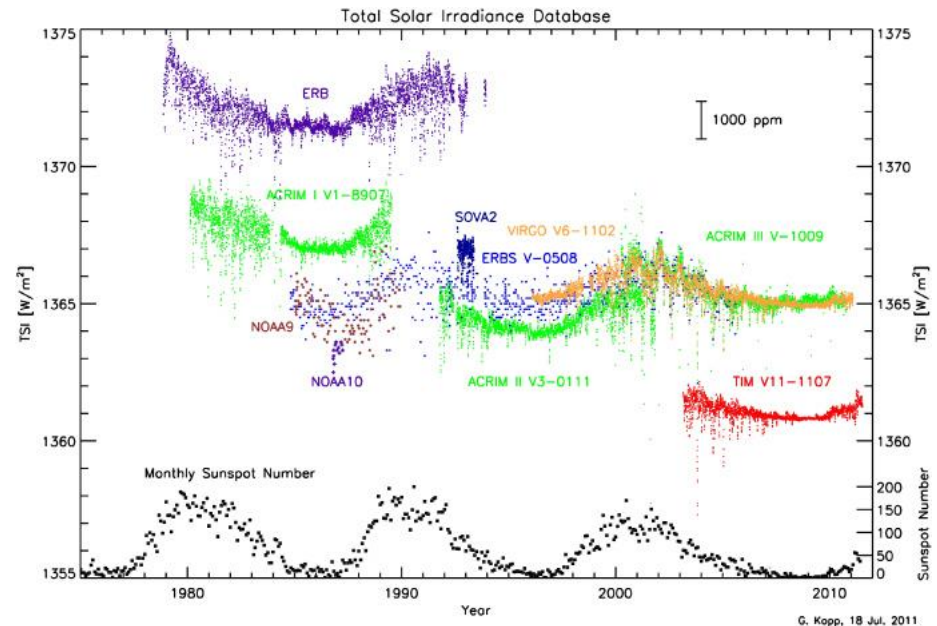


# SOLAR data have societal importance.



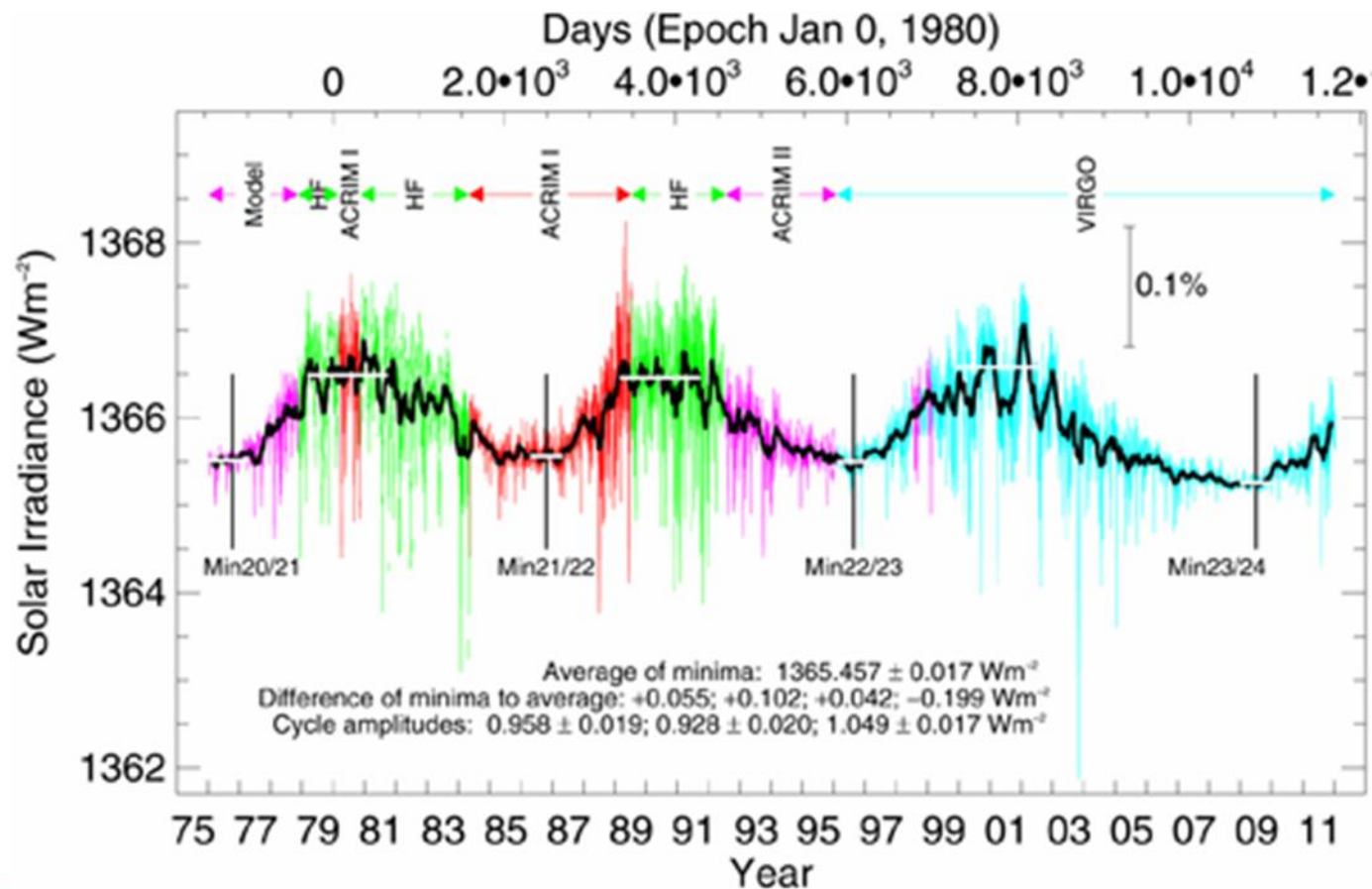
# What happens in an observational data series?

- ▶ Horizontal Gaps: there are missing observation periods
- ▶ Vertical gaps: two data sets differ more than the observational error.

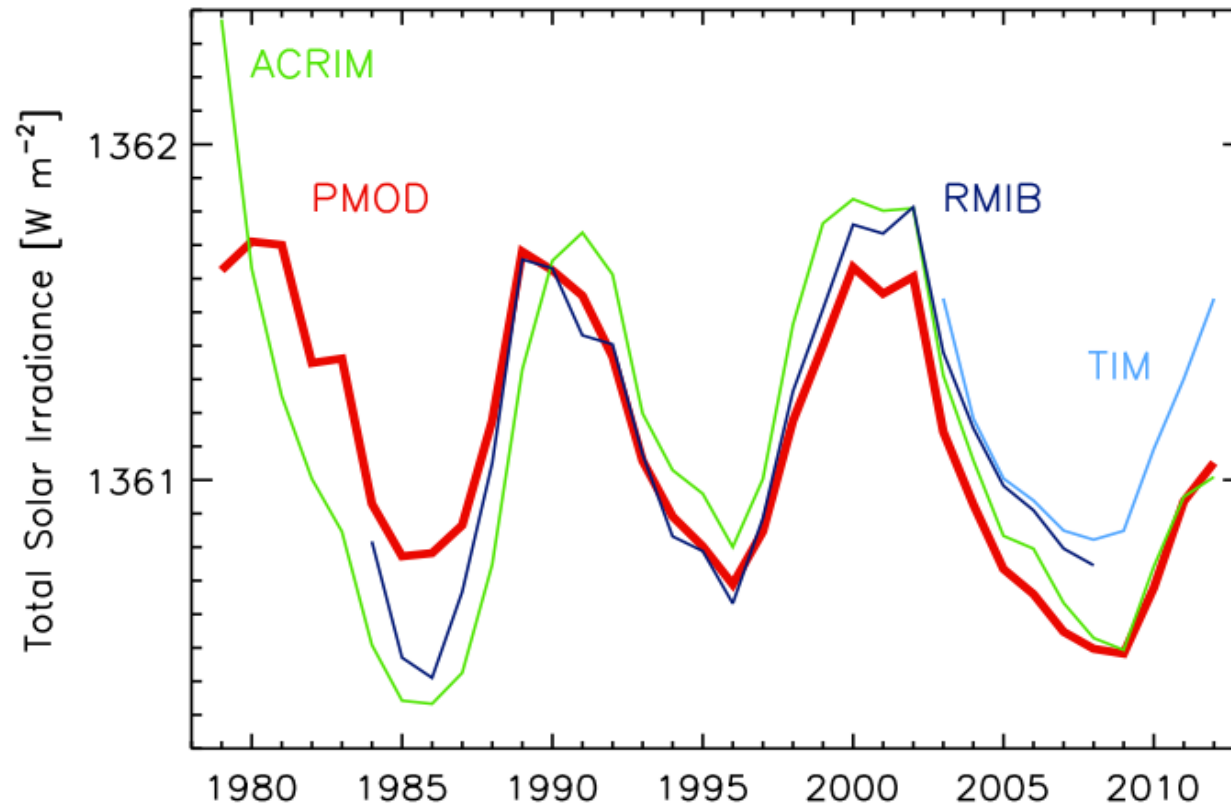


Total Solar Irradiance between 1970 and 2011.

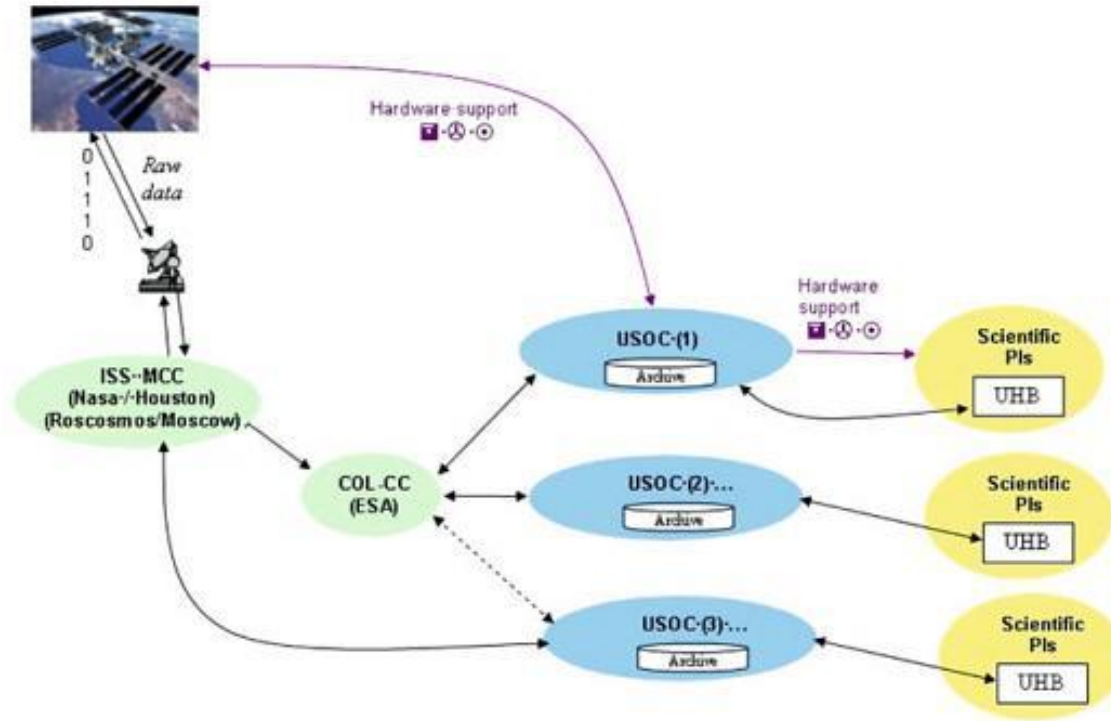
# Lean and Fröhlich reconstruction



# IPCC (2014): RMIB and PMOD correspond to SOLAR



# Again, what did we with SOLAR data.

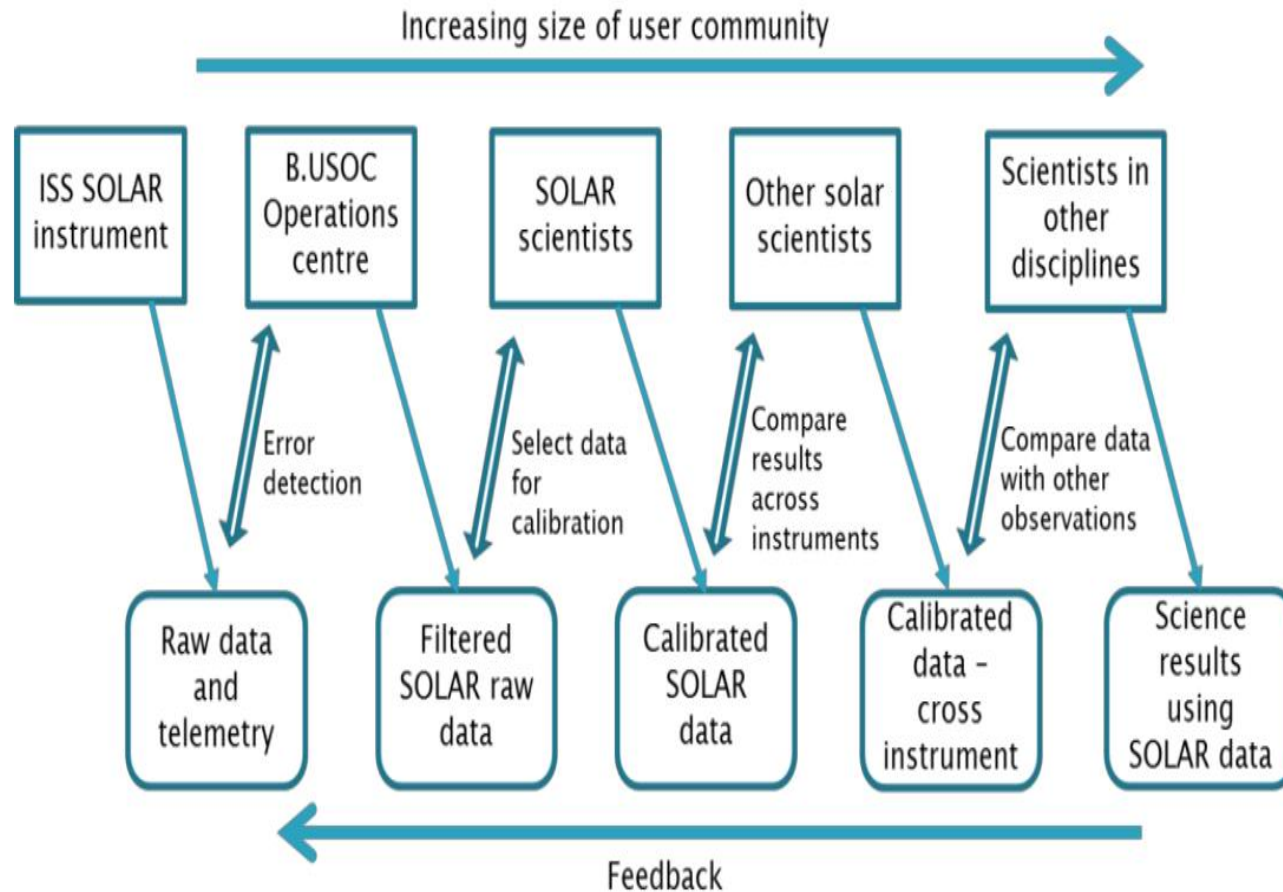


B.USOC supplies the data requested by the scientist using a dedicated software, the YAMCS which parses the data flow so that each experimenter receive what he requests and the essential monitoring parameters are transmitted to the screens of the operators.

The YAMCS is a proprietary B.USOC software developed by SpaceAps, it is also used for ERA (European Robotic Arm)



# Data flow: a scientist's point of view





# HRDP and ESA data policy

- HRDP is a rack of servers and network equipment, it is owned by ESA and installed at the respective USOC's, it is highly secure and can only be accessed by certified personnel. The data stored on HRDP are in the original CCSDS format. In our case, only a YAMCS application could retrieve SOLAR data. HRDP has its own backup hardware and procedures.
- ESA present data policy: absolutely no outside access to the HRDP, data distribution limited to PI, up to now, B.USOC has not been allowed to supplement HRDP storage by an accessible storage.
- B.USOC does not archive the data it sent to the PI's nor the scientific products generated by the PI's.
- **Data storage at USOC's is limited to 10 years.**
- All the negative aspects of this list can be solved by a revised data policy. (and a corresponding adaptation of means for the USOC's)

# How to avoid failure in the preservation of space data?

- ▶ Standardisation: CCSDS (consultative committee on space data systems), agreement between 11 space agencies on common formats for the space segment.
- ▶ “Inspire” directive of the European Union (for geolocated data). This directive is public and includes ISO standards.
- ▶ **Drafting of a Data Management Plan even if not required by the funding agency.**

# First step: data catalogue

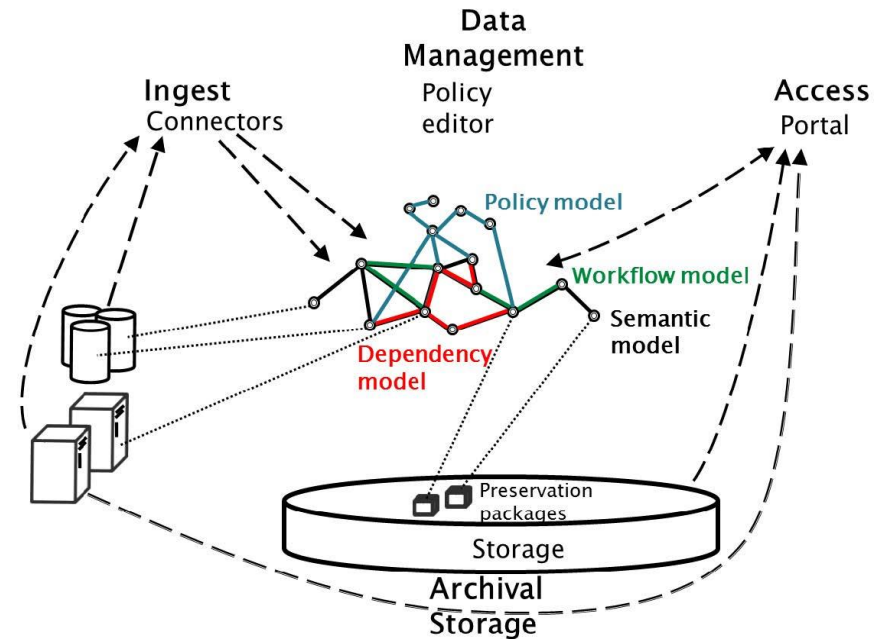
- Engineering Documentation.
- Operations documents
- Documents generated during operations.
- Data generated during operations
- Science data
- 29 items categories have been identified at B.USOC including the science data in HRDP.
- **Scientific products and their metadata: managed by the PI's.**

## Second step: data preservation

- Regroup all the data in a single data base.
- Convert HRDP data to a format accessible by qualified users.
- **Scientific products and their metadata: should also be part of the preserved data. This aspect is even central to the new (Nov 2014) NASA preservation policy.**
- **Then, the processes developed by the PERICLES IT partners can be implemented.**

# What should have been done in PERICLES?

- ▶ The scientific data approved by the PI's, publications and all B.USOC inventories should have been ingested in the same data base.
- ▶ **Then, the PERICLES models and tools should have been applied to it.**



## Models

LRM with semantic versioning

DEM

Domain ontologies

## PERICLES Approaches

ERMR

Model Impact Change

Explorer (MICE)

PeriCat

PET

Policy Editor

Process Compiler (PC)

Technical Appraisal Tool

## Tools for models

EcoBuilder

LRM Service

PET2LRM

PROPheT

## Semantic tools

SemaDrift

Somoclu

Ncpol2SDPA

## Other (Helper components)

PeriCoDe

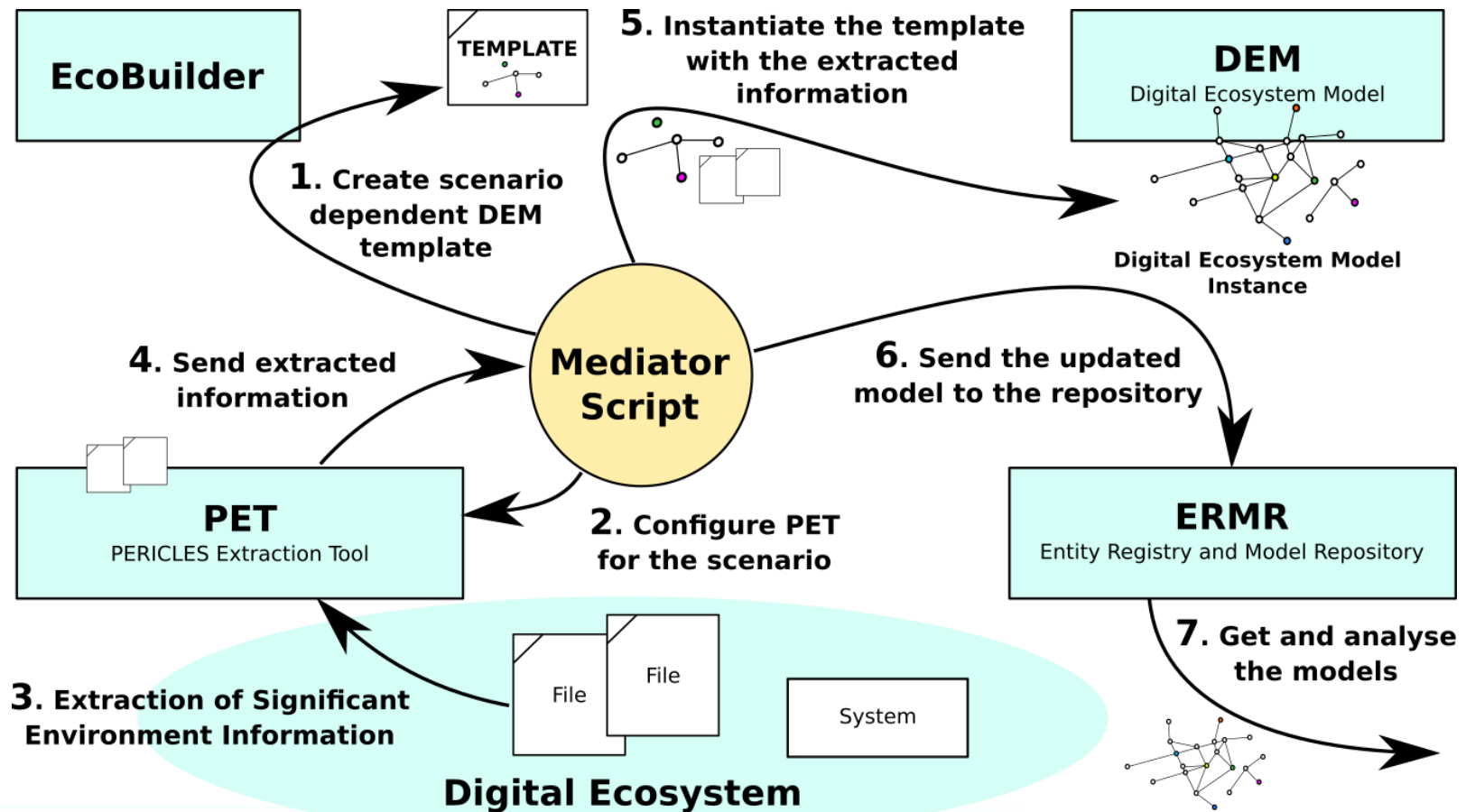
PERICLES Semantic

Interpretation API (PERSIsT)

Testbed infrastructure



# Create an ecosystem model using EcoBuilder



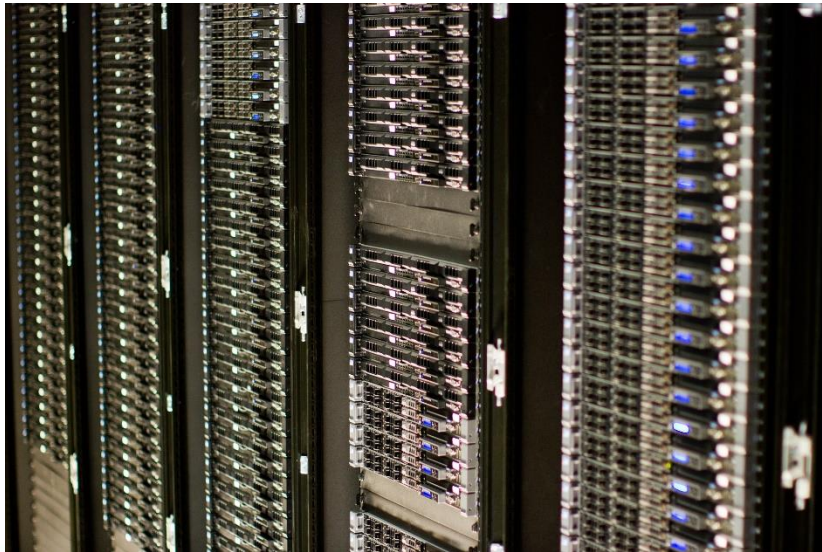
# What was actually done?

- ▶ The scientific data were analysed by the PI's and published in peer review literature.
- ▶ SOLAR data was presented to the relevant COSPAR commission and given a quality stamp.
- ▶ **ESA (HSO, human and space operations) transferred the science data to ESAC (European Space Astronomy Centre) where it is preserved in the heliophysics archive.**
- ▶ **Replaying the mission to reinterpret it is thus for ever impossible but the “final” data is preserved.**
- ▶ **A composite data set is only possible from the data sets preserved by the scientists and with their support.**

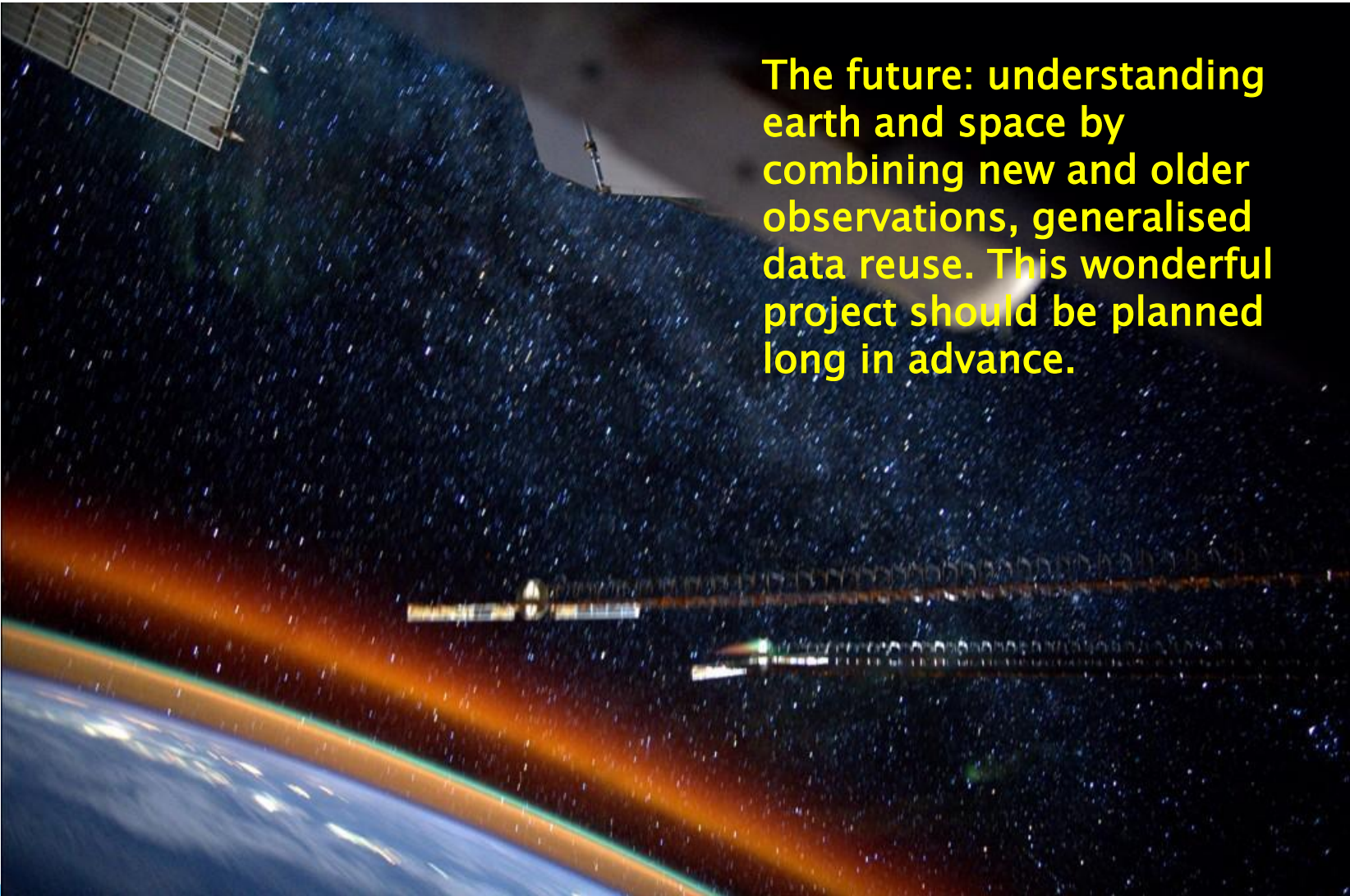
# Why was the PERICLES process not initiated?

- ▶ The space segment of SOLAR and the PERICLES project stopped almost simultaneously, the financing through both ESA SOLAR and EU FP-7 contract could not be renewed.
- ▶ The new ESA data policy allowing USOC's to generate products was not yet implemented.
- ▶ The restrictions on use of ISS data were such that the final PERICLES review used EUMETSAT data from the SEVIRI instrument on the METEOSAT second generation satellite in order to demonstrate the various PERICLES tools.
- ▶ **Finally, the SOLAR scientists would probably never have released their calibration data base nor their retrieval algorithms.**

Preservation in ESAC is already a big progress compared to what usually happens to other data sets obtained in ESA manned space missions.







The future: understanding earth and space by combining new and older observations, generalised data reuse. This wonderful project should be planned long in advance.