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]

pericles
FP7 Digital Preservation

Data management plans and standards for space data: how to train future space data providers and users.

BigSkyEarth, Sorrento, 2016, Christian Muller, B.USOC

SEVENTH FRAMEWORK PROGRAMME

Plan of the talk

- A word about PERICLES and the way it considers data preservation.
- History of space data.
- Space data lifecycle, the example of the SOLAR ISS payload as managed by B.USOC
- How data mismanagement impacted science: Apollo, ozone hole, SKYLAB.
- One success of long term data management: LANDSAT.
- Data Management Plan guidelines.
- Advanced Data Management.

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

Pericles
FP7 Digital Preservation

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 – 31 Jan 2017)
- With a consortium of 11 Partners

pericles
FP7 Digital Preservation

Focus on change

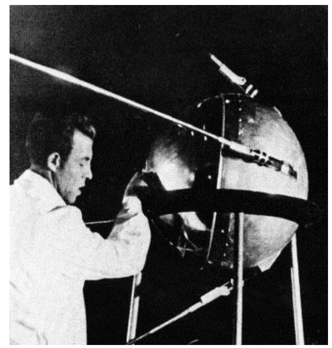
- **Long-term sustainability** requires us to address change more broadly

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

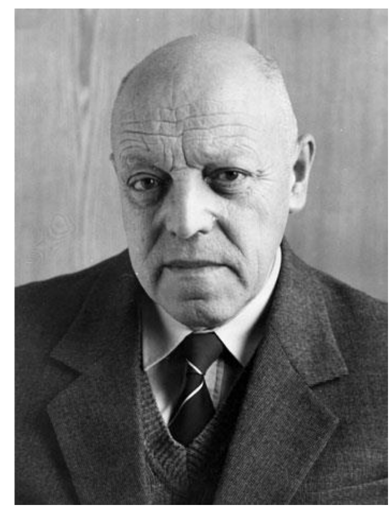
Preservation-by-design

History: 1957, launch of the first artificial satellites, first scientific results on the ionosphere and space plasma layers.

James Van Allen came from sounding rockets.



Konstantin Gringauz choose the transmitter frequency of Sputnik 1 and switched it on the launch pad. Both Explorer 1 and Sputnik 1 met their science objectives.



Originally: space data are generated and published by the scientists who designed the instruments.

1969: Apollo: 6000 engineers simultaneously on console

APOLLO mission control used at the same time primitive digital communication, voice links, telex, pneumatic tubes and runners to communicate.(very little flexibility).

Fantastic management success.
The main data were samples and films (images and spectra on argentic film).

Apollo was however the begin of the digital age with the Apollo Guidance Computer. The AGC was programmed by a human who invented the term software engineering to describe her job.

“Partially piloted **manually** by Armstrong, the Eagle landed in the Sea of Tranquillity in Site 2 at 0 degrees, 41 minutes, 15 seconds north latitude and 23 degrees, 26 minutes east longitude. This was about four miles downrange from the predicted touchdown point and occurred almost one-and-a-half minutes earlier than scheduled. It included a powered descent that ran a mere nominal 40 seconds longer than preflight planning due to translation manoeuvres to avoid a crater during the final phase of landing.”

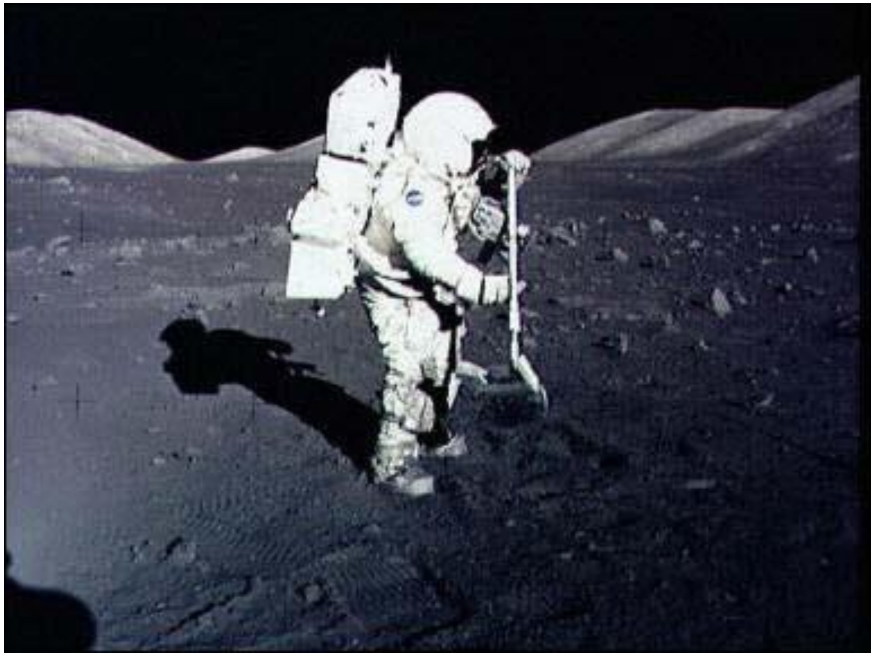
Margaret Hamilton’s software gave the right indication to Armstrong so that he could land manually. (Early data appraisal)

Pericles
FP7 Digital Preservation

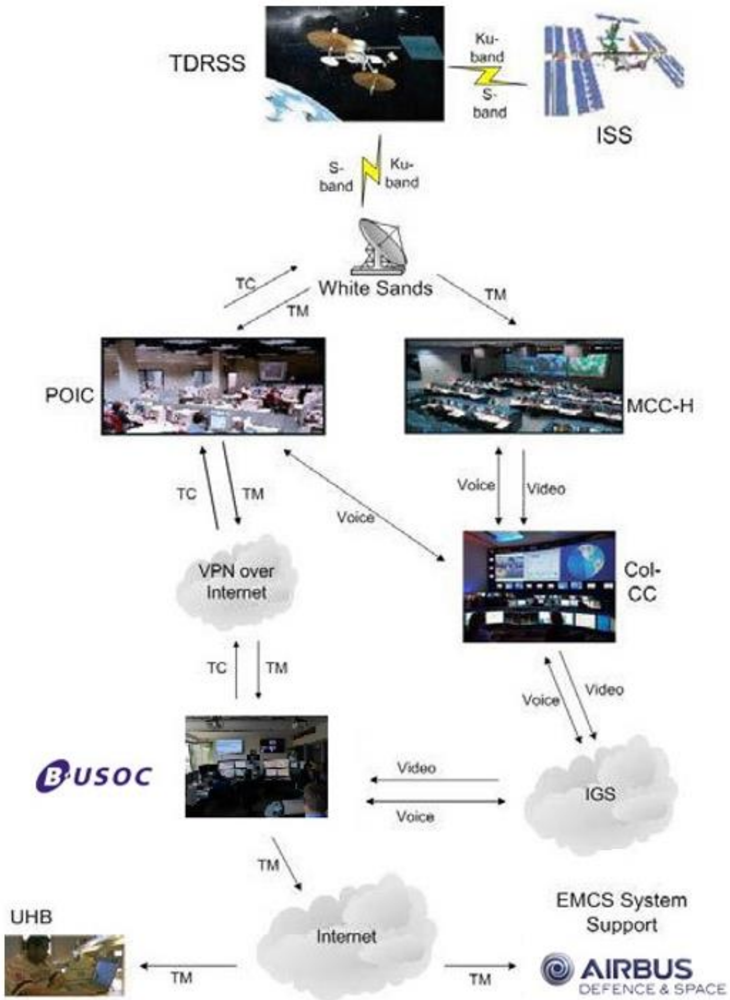
What were Apollo operations and data?

- Apollo was not designed for science in the first place, there was no specific science operation centre, the operations were led by flight controllers.
- However, science plans could be designed for Apollo 17,18 and 19 using the experience gained by operators since Apollo 11.
- Astronaut Harrison Schmitt was actually the scientist in charge of Apollo 17.

- **Science data flow was not the priority.**



A current space experiment: SOLAR on the ISS



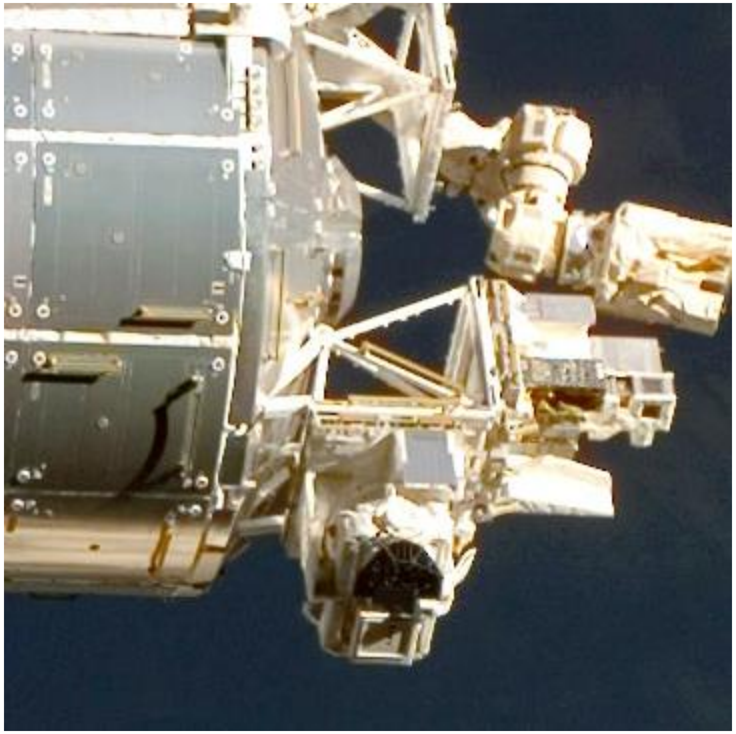
What is the role of B.USOC in ISS data flow?

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

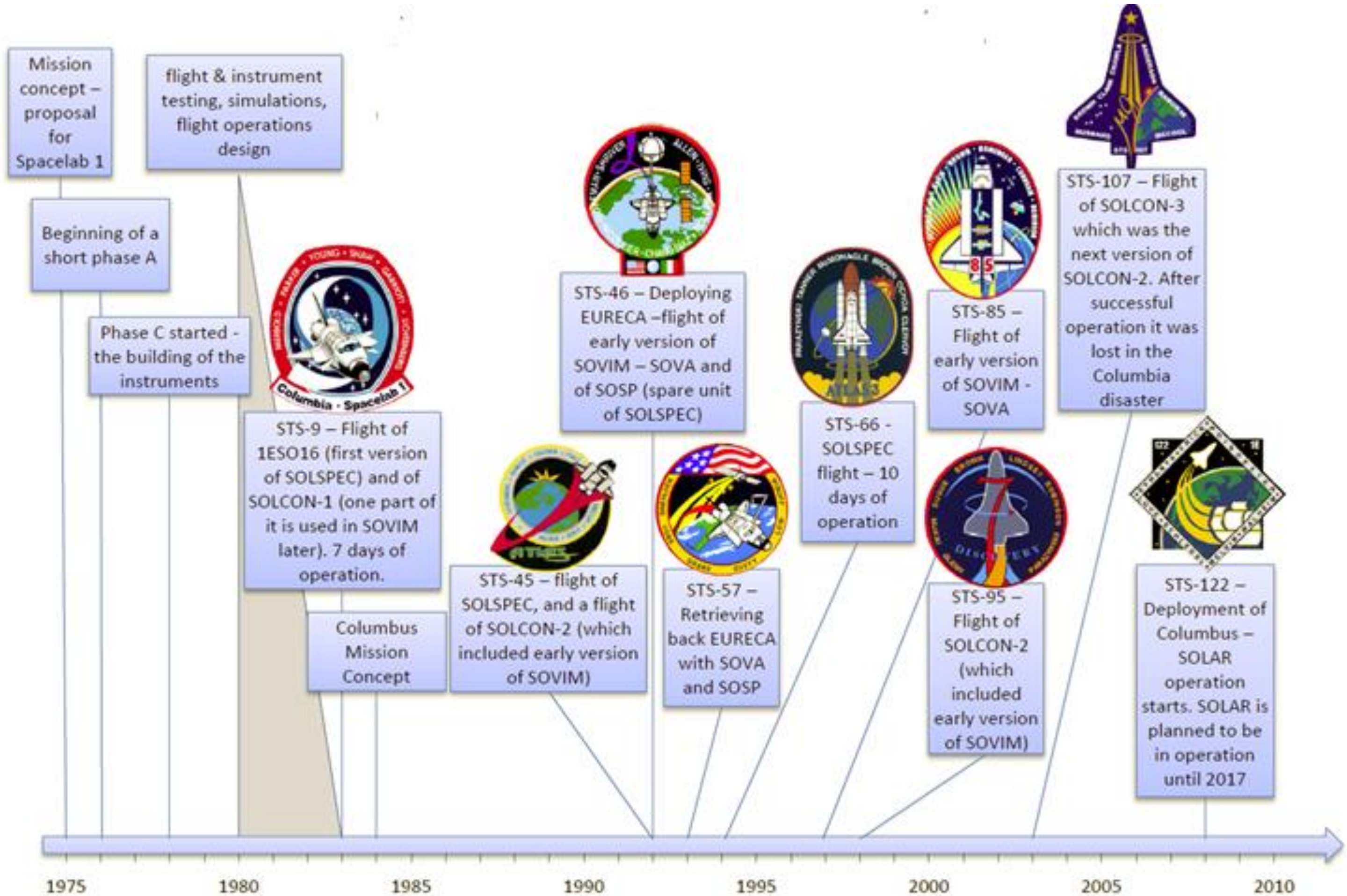


- The SOLAR package on the COLUMBUS module of the ISS.
- Three intruments: SOLSPEC, SOLACES and SOVIM .
- Operated by B.USOC since 2008 at the benefits of ESA and the Pi’s.

SOLAR as an example of change of semantics.

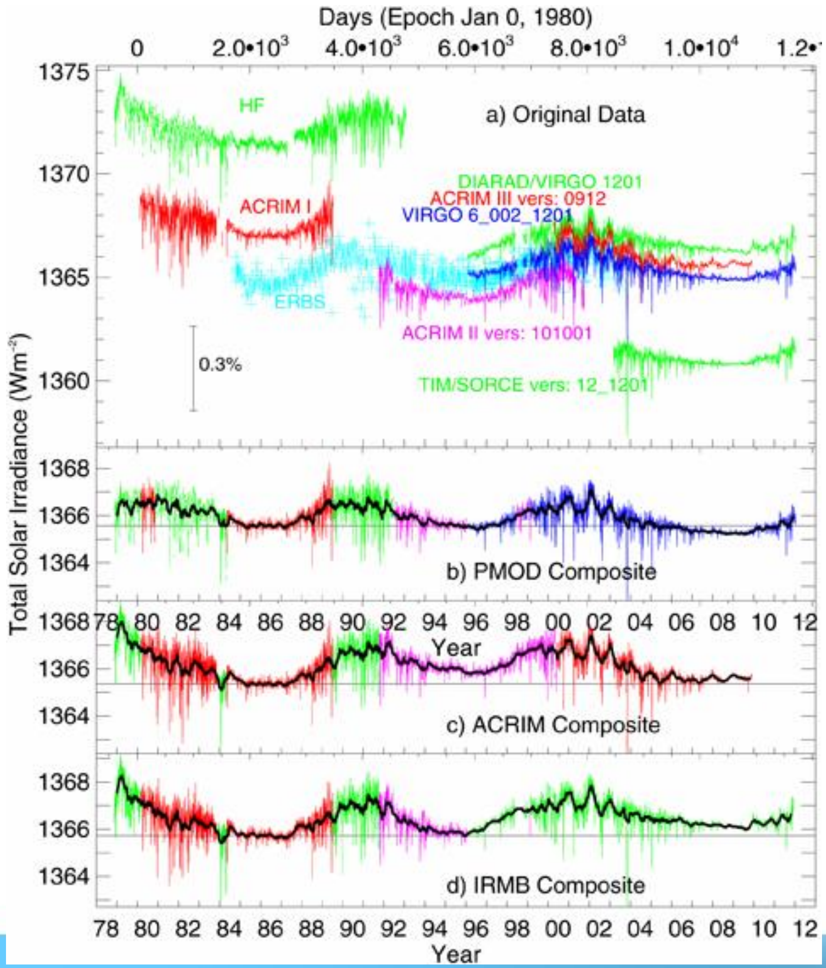
- Until around 1976: total energy imput to the surface of the earth: solar constant.
- Space age: discoveries of possible variations: total solar irradiance.
- Since 1943, first through sporadic balloon and rocket observations, than from space, discovery of short term variabilities in the UV.
- Comparison 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



Pericles
FP7 Digital Preservation

Current status of solar data

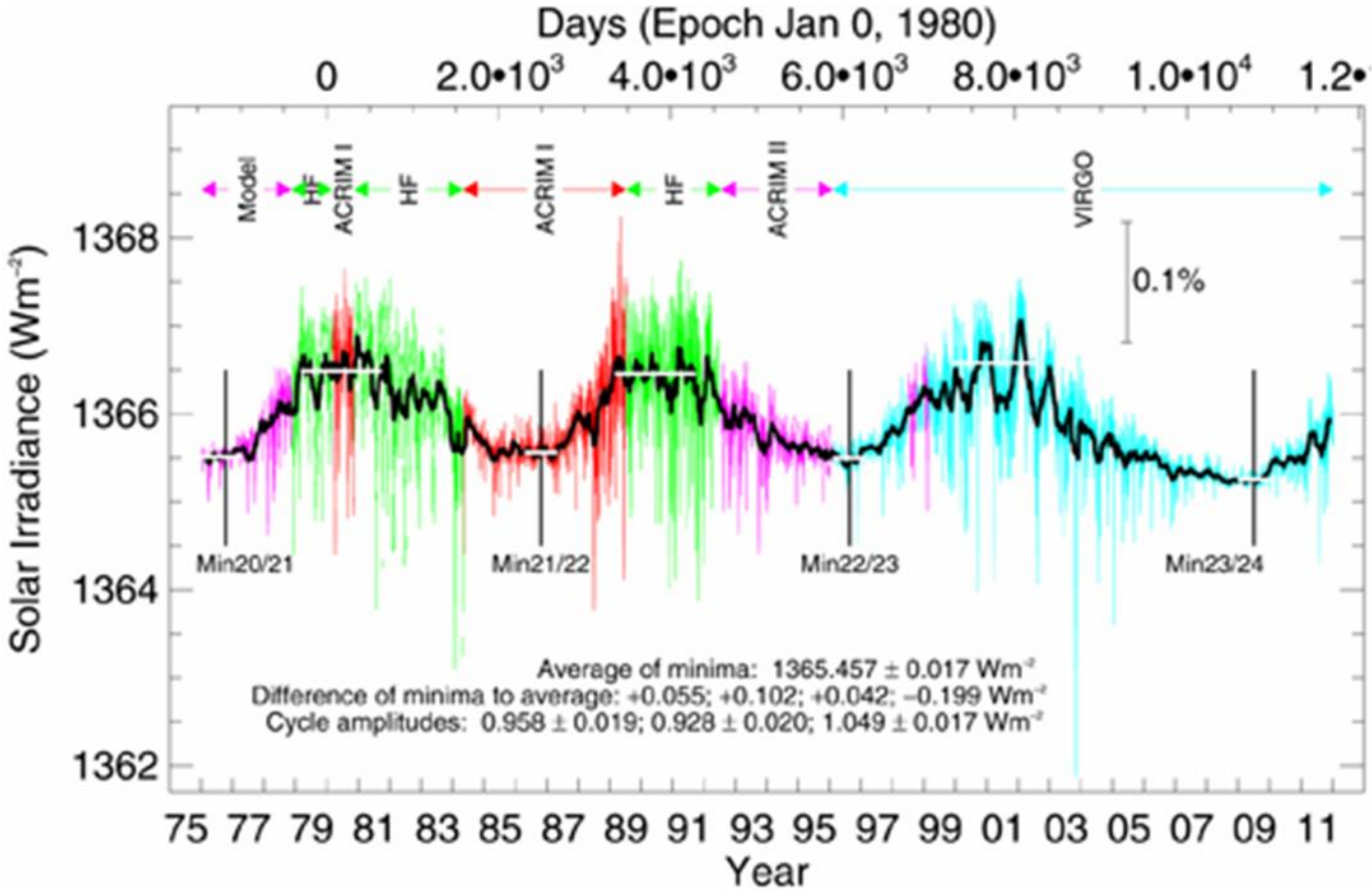


Three groups measure Total solar irradiance: horizontal gaps, vertical gaps.

The figure includes also the new TIM instrument on the SORCE satellite; the SORCE payload will probably become the reference of the 21th century.

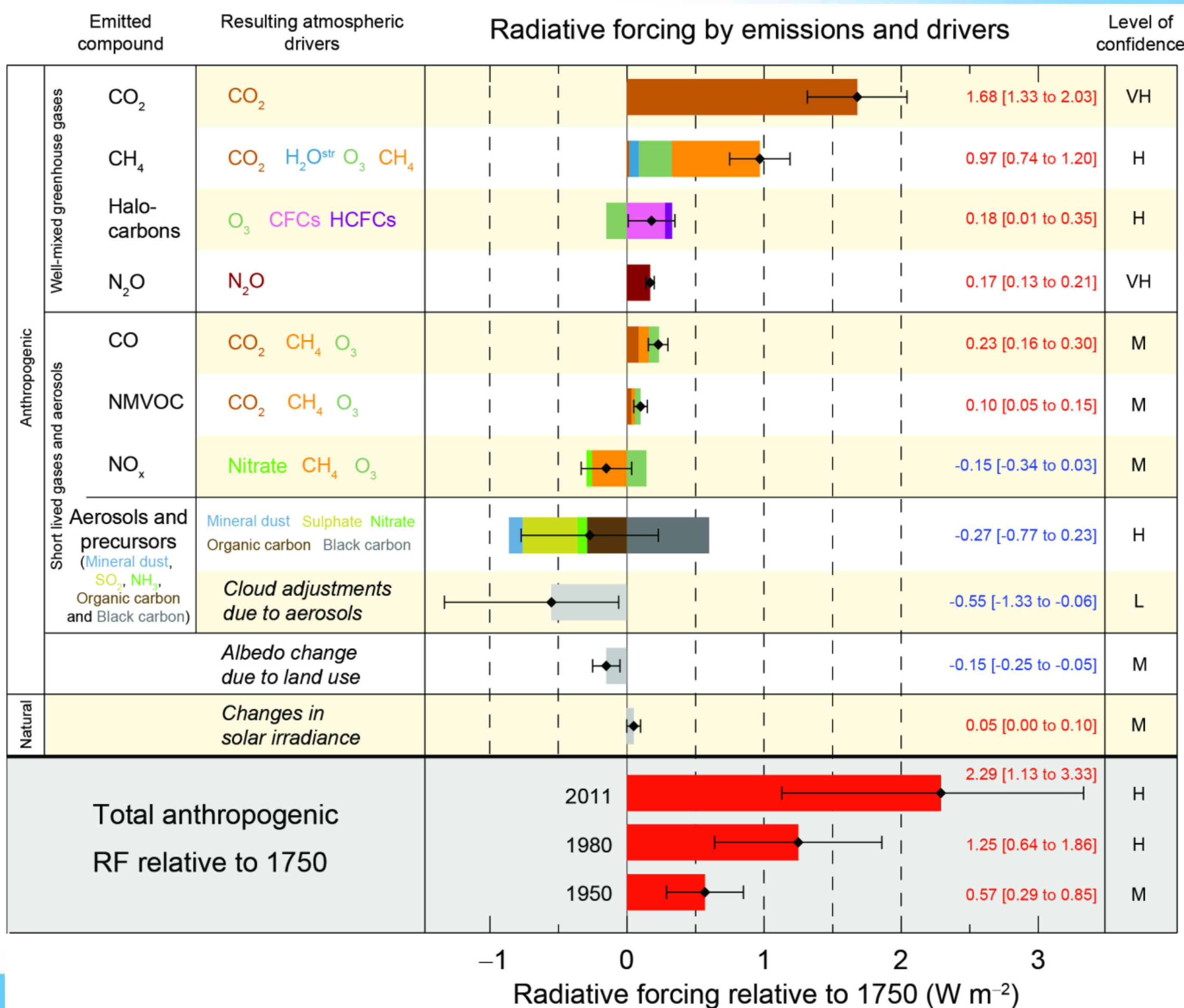
From here it is clear that we should be ready for reinterpretation.

Lean and Fröhlich reconstruction



Pericles
FP7 Digital Preservation

SOLAR data have societal importance.



Pericles FP7 Digital Preservation

Apollo case: The Search for the Apollo 11 SSTV Tapes

- Apollo images were acquired by slow scan TV.(10 images per second, 120 or 320 lines).
- Converted to broadcasting format.
- The original format tape were sent to NASA after the mission.

Two Australian stations became the main receiving sites, amateur films and photographs of the original screens were of better quality than the broadcasted data.

Present image processing would restore them to a much better quality, the original tapes are still to be located.

Apollo 9 tapes found in non-standard sized boxes in the National Archives, Washington

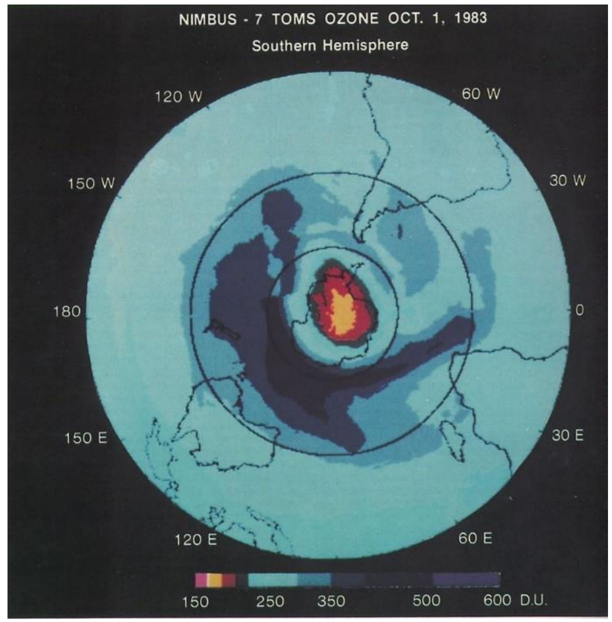
What went wrong?

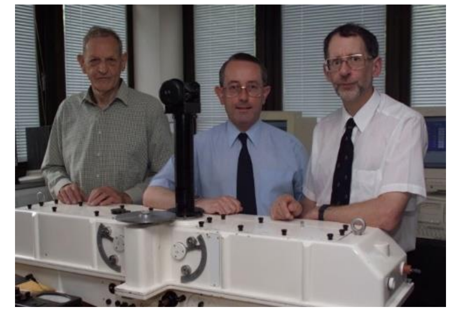
- The camera was not designed as a science experiment, it was added quite late when the data flow had been already defined, some parameters were reduced in rate to have a sufficient bandwidth for the camera which had to be designed for the purpose.
- **Nobody even thought that the images of the moon surface and limb would have scientific value.**
- No plan existed for the tape disposition, NASA transferred them to the National Archives and even recovered them during a tape shortage in the 80's. and then reused at Goddard Space Flight Center for Landsat archiving.
- Tapes remaining at the National Archives are unlabelled but the boxes(non standard) contain accession documents.
- The camera tapes have not yet been found.

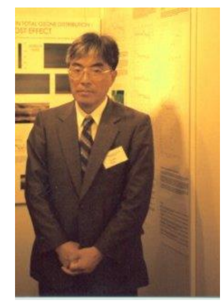
Pericles FP7 Digital Preservation

The Antarctic ozone hole: undetected by TOMS/SBUV

- The TOMS ozone mapper was flying since 1977 on Nimbus 7.
- The ozone hole was published by ground based observers in 1984 and 1985.
- If the satellite data had been better processed, it would have been published in 1983 and the Montréal protocol could have happened two years before.







Farman et al, 1985, Shubashi, 1984.

Syowa station where ground observers discovered the ozone hole.

What went wrong?

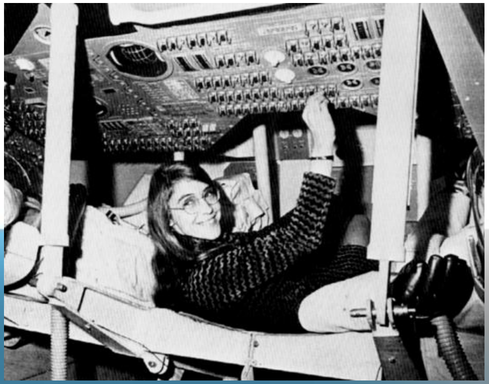
- "Part of the answer lies in the sheer volume of data generated by the TOMS/SBUV instruments. About two hundred thousand measurements were made each day, placing a great strain on the computer capacity of the day; it took a year or two to process the data and get them into the hands of scientists. Then there was a long list of interesting patterns to investigate, and the researchers knew that not all of them would turn out to be real. The low ozone amounts seen in the Antarctic spring were below what was considered to be the lowest possible amount of ozone and were flagged as potential 'bad' data." (Stolarski, 1985).
- TOMS was not designed nor to measure trends nor to produce operational data, there was no delay for validation of data.
- "When I started working on ozone in 1977, there was not a known issue with the ozone layer and we weren't worried about it thinning out. Instruments I was working on weren't designed for observing that. Starting with Nimbus-4 in 1970, we simply wanted to measure the atmosphere. It was curiosity-driven research." (P. Bhartia., current NASA GSFC site).
- **The TOMS-SBUV team was not prepared to a change in scientific paradigm.**

Pericles
FP7 Digital Preservation

The SKYLAB data.

- SKYLAB was the first space station from 1973 to 1979.
- Data format was specified by mechanical switches which were omnipresent in Apollo hardware.
- At some point, these switches went into a random position and nobody could read the data anymore. (Garriott, private communication).

Owen Garriott aboard SKYLAB



Margaret Hamilton finding these switches funny in an Apollo simulator.

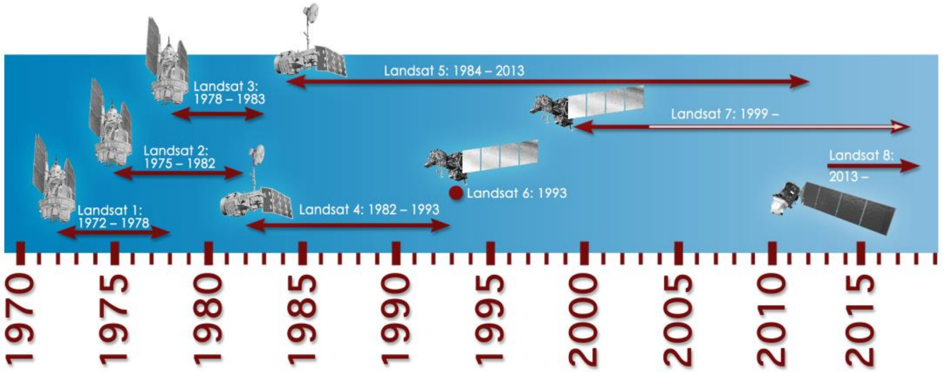
Pericles FP7 Digital Preservation

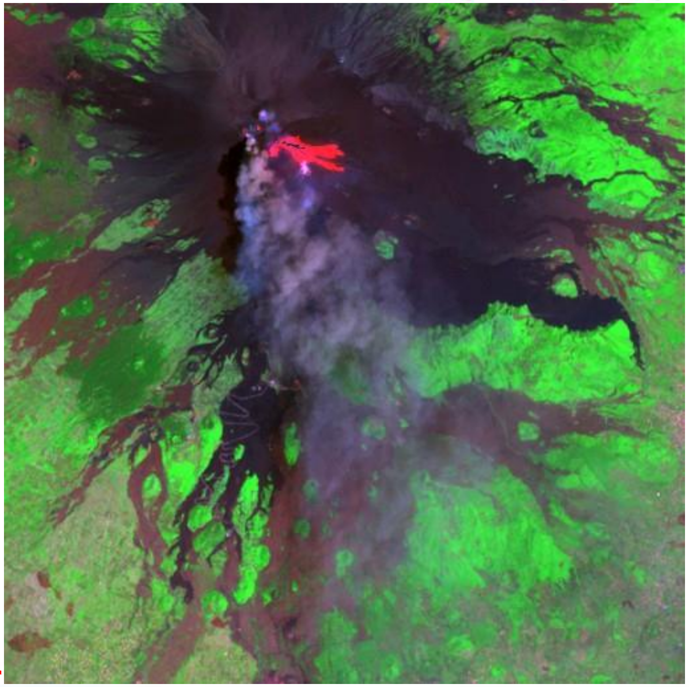
What went wrong?

- “Switch Guarding – As a corollary to the discovery by the Skylab crews that virtually any piece of equipment had its uses as a restraint, toes and fingers penetrated more universally than had been planned. Thus the mechanical protection provided for switches, while adequate to prevent accidental tripping by hands working a console, was not adequate against toes blindly seeking support.”(MSFC, 1974).
- At the time, no international data standards for space existed.

A success case: LANDSAT: thematic mappers.

- Began as an experimental technology earth digital observation programme in 1972
- Now, LANDSAT is a prime source for the study of global change on the earth and since 2013 operates its eighth version: the LANDSAT continuity mission with joint data management by NASA and the USGS (US Geological Service). One of its satellite (LANDSAT 5) holds the record of the longest operating earth observation satellite, launched in 1984 and decommissioned in 2013.
- All the archive is available and the LANDSAT site presents even a tool for users so they can draft a data management plan for their new products.





LANDSAT image of the Etna 2001 eruption, The wavelength ranges used by Landsat are these: deep blue, blue, green, red, four near and shortwave infrared bands, a panchromatic (grayscale) band, and two thermal Infrared bands.

Why are all examples from NASA?

- Russian and European space agencies usually do not publicly report on failures, especially if the space segment is not the cause. A lot of information should have the been quoted as oral communication without supporting documents.
- NASA always analysed and made public the lessons learnt from its failures so that following mission designers would be informed, **this post mission data debrief and analysis should also be a part of the data management plan.**

Pericles
FP7 Digital Preservation

How to avoid failure?

- 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.**

CCSDS

- Fixed format with a header.
- Compromise between readability and data transmission.
- Needs a dedicated software to read it.
- In the ISS case, several experiments are present in the same data packet.
- **This format is unfit for preservation and as it has to be converted to be distributed to the scientists, an other format would better address the preservation needs.**
- **OAIS (Open Access Information System) began as a CCSDS standard for metadata.**

INSPIRE

- The INSPIRE Directive (EC, 2007, EC 2008) aims at shaping a common framework for the communication of information with geographic reference throughout Europe. The Directive explicitly aims at establishing a 2-way beneficial relationship with Galileo and the Copernicus services. Earth observation and modelling data on atmospheric composition and associated parameters typically lie within the INSPIRE scope. INSPIRE is fully compliant with the ISO categories referring to geographical data.
- **For atmosphere and ocean data, INSPIRE has a dimensionality problem: altitude, depth, time.**
- **INSPIRE does not request a quality marker.**

Data Mznzgement Plan: advice from Colorado University (Boulder)

- Make a plan!
- 1. Go to DMPTool (http://dmptool.org)
- 2. Log in with institutional credentials (or create an account)
- 3. Find appropriate funding agency template
- 4. Fill out each section of the DMP
- 5. Export file for grant application or other use

Pericles
FP7 Digital Preservation

NSF general DMP requirements

- 1. What types of data will you produce?
- 2. What (if any) standards will you use?
- http://www.dcc.ac.uk/resources/metadata-standards
- 3. When and how will you share data?
- http://figshare.com/
- 4. What can people do with your data?
- 5. How will you archive and preserve data?

Data Archiving and Preservation

- Proper archiving for long-term preservation of data is critical
- What should be archived?:
- Data
- Metadata
- Research products
- Scripts
- Anything required to reproduce the results of research

PERICLES aspect.

- PERICLES uses a different model than OAIS: linked resource model.
- A PERICLES based archive links all the different categories of data.
- The PERICLES data survey goes from proposal level to post-mission including recovered flight hardware.
- **The final result is the possibility to replay and reanalyse the mission long after the original teams and their institutions have disappeared.**

PERICLES and Advanced Data Management plan.

- ADMP builds on a dynamic DMP, PERICLES departs from OAIS precisely to take into account semantic change.
- PERICLES proposes tools for process monitoring, data appraisal, data encapsulation.
- These tools lead to a better reorganisation of the Data Management Plan in project lifetime.
- **Again, the final result addresses data preservation long after the original teams and their institutions have disappeared.**

The future: understanding earth and space by combining new and older observations, generalised data reuse.