

PV 2015

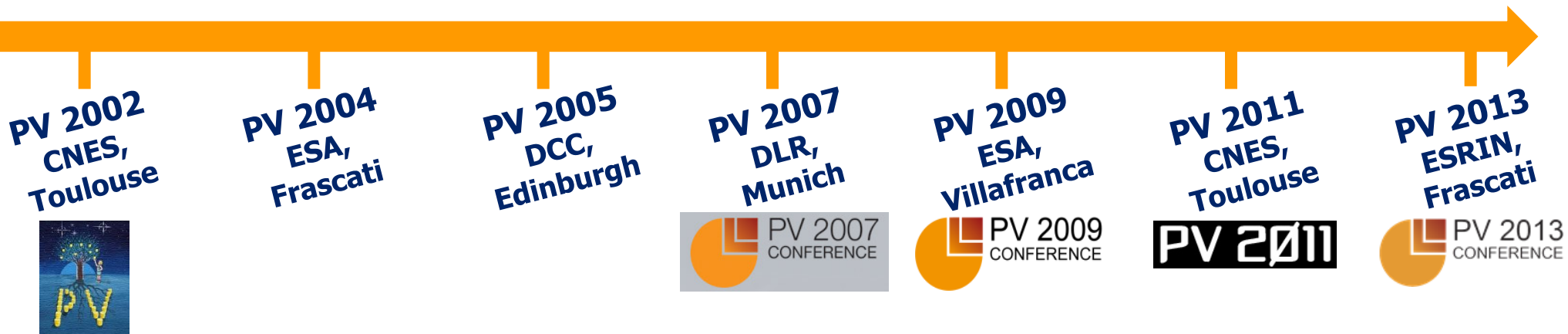
ENSURING THE LONG-TERM PRESERVATION AND
VALUE ADDING TO SCIENTIFIC AND TECHNICAL DATA

Summary and Conclusions

3-5 NOVEMBER 2015, DARMSTADT, GERMANY

The PV Conferences

The PV conference series is about ensuring **Long-term Preservation and Adding Value to Scientific and Technical Data**



Eighth edition

*Standards for Archives and Interoperability
enabling new added value services*

PV2015 - Objectives

- Share knowledge, experiences, lessons learnt and best practices as well as foster Cooperation in the area of Data Preservation and archived data Stewardship.
- Address prospects in the domain of high volume, high-velocity, high-variety scientific and technical data preservation together with value adding to these data.
- Provide a forum for organisations dealing with preservation of own data and value adding to present the status of their activities, plans and expectations.



PV2015 - Participation

- PV2015 participation is on a high level compared to previous events
- Some numbers:
 - 92 attending participants
 - 40 Oral presentations
 - 25 Posters
 - 25% overseas presentations



<https://www.flickr.com/photos/eumetsat/22161315083/in/dateposted/>

PV2015 - Key Points raised during conference

The following slides present key points raised during the conference in oral and poster presentations in the different sessions (Percentage of overall presentations):



- Session 1: Data Stewardship approaches to ensure long-term data and knowledge preservation (22%)
- Session 2: Adding value to data and facilitation of data use **(44%)**
- Session 3: Metadata, formats, standards and interoperability (17%)
- Session 4: Data preservation lessons learnt and future prospects (17%)

Data Stewardship Approaches to ensure Long-Term Data and Knowledge Presentation - Key Points (1)

Session: Harald Rothfuss

- Most Archives work towards providing faster access to data (e.g. Hadoop) and “users to the data” szenarios.
- Similar approaches – standard based - are generally used to ensure safely archived and re-useable data.
- Exponential data growth for most Archives is observed.
- Most archives now offer a large online component.
- Enhancements in metadata include
 - United collections metadata and file level metadata (CMR)
 - Improvements in accessibility (e.g. via Persistent Identifiers such as DOI)
- Use of indices such as Data Maturity Matrix and Data Stewardship Matrix - derived from Climate data processing – allow an improved assessment of the preservation quality level.
- Need for international coordination of standards in Data preservation

Data Stewardship Approaches to ensure Long-Term Data and Knowledge Presentation - Key Points (2)

Session Mirko Albani

- Several activities have been presented where archive data is combined with linked data to provide users with data access that is encompassing more related data sources. OAIS standard will be updated in 2017, minimal changes are expected to Core Part.
- With massively growing data, aspects of data corruption in Archives (Bit rot – termed by CERN) becomes an issue which needs to be addressed.
- Preservation of data analysis software and tools is essential for data use.
- Rescue and consolidation of legacy data is crucial and urgent.

Session Tom Stein

- Capturing knowledge underlying the data is as important as preserving the data.
- It is essential to qualify and quantify data uncertainties - especially key in data restoration projects.
- Model-driven preservation can help capture digital ecosystem for future use.

Adding Value and Facilitation of Data Use

Key Points (1)

Session Hannes Thiemann

- different approaches to add value to data or facilitate ease of use:
 - Enhanced data services not only provide access to selected datasets but also generate data sets itself (ICDC).
 - systems offering data mining and knowledge discovery functionalities as well as machine learning techniques to semantic annotation (DLR EOLib).
 - Improved data access and analysis when data is organised in data cubes, moving from data stewardship to data service stewardship (EarthServer).
- Storage Modelling is important to optimally handle data chains and make use of the tremendous data amounts (ICRAR).

Adding Value and Facilitation of Data Use

Key Points (2)

Session Aurelie Bellucci

- Tapes are still largely in use in data preservation with pro's and con's
- Archives generally want to address the needs of large and various communities and not only the experts :
 - Requires easy access in terms of data discovery and transformation
 - User are very interested in tailoring of data prior to delivery which requires relevant services and tools
- Complexity in infrastructures addressing multiple scientific domains when built without involving the main user communities

Session Jamie Shiers

- Increasing use of / believe in virtualisation as a promising and used tool for “software preservation”, perhaps over timescales of around 1 decade.
- Increasing demand from Funding Agencies for Data Management Plans and response from communities that these should a) be useful b) be dynamic and not static, written once to pass the grant phase.

Adding Value and Facilitation of Data Use

Key Points (3)

Session Christophe Arviset

- Providing added value products (image mosaics from mars rover, all sky maps from astronomy missions) enhance archive usage and respond to "customers" wishes.
- Visualization facilitates data exploitation but also science data exploration.
- Science coming from combination of data from professional and amateur radio astronomy, using VO (Virtual Observatory) protocols.
- New paradigm for big data archive (ie Gaia): bring the user and code to the data instead of ship the data to the user.
- Long Term Preservation of archive systems requires improvement through regular technology migration (eg CNES REGARDS project).

Metadata, Formats, Standards and Interoperability

Key Points

Session Tom Stein

- Linked open data and RDF models can facilitate complex search and discovery.
- Metadata models can support interoperability in search of complimentary data with different metadata structures.
- Adding browse and download functions to existing catalogue and order interfaces can improve access to data.

Session Nancy Ritchey

Main themes were supporting the increase in user requests for data and services and the exponential growth of data for specific scientific disciplines:

- EUDAT as an example of a cross-discipline metadata portal supporting the users through common vocabularies and a comprehensive metadata catalogue and faceted search.
- Systems evolve to include lessons learned, need for user tools, services and application and are hardened in design to provide improved data preservation and usability, as demonstrated by next generation PDS4.

Data Preservation Lessons learnt and Future Prospects

Key Points

Session Eberhard Mikusch

- *For session III: Metadata, formats, standards and Interoperability:*
Community best practices (in the EO domain) are essential for the use of Persistent Identifiers (e.g. DOI) in order to establish homogeneity and acceptability for (EO) data users
- Data curation and cyclic migration necessary for understandability and usability over decades
- Success and effort of LTDP depends on the start of the activities early in the projects

Session David Giaretta

- Interoperable services should be easy to use, hiding the technical complexity involved.
- Interoperable Research e-Infrastructure can facilitate data management processes of smaller scientific organisations to share research information.

PV2015: Provided a forum for exchange



More pictures will be uploaded on flickr (<http://bit.ly/1RB1kCJ>)

PV2015 feedback

We would like to ask for your feedback on various aspects of the PV2015 Conference.

An online feedback questionnaire has been made available:

<https://www.surveymonkey.com/r/GHNYKGZ>

Thank you for your timely comments!

Next PV Conference

