EUMETSAT LONG TERM DATA PRESERVATION ROADMAP

David Berry , Michael Schick

EUMETSAT

ABSTRACT

Managing large amount of data, preserving the data for the future and ensuring it can be accessed and used at any time, requires to follow Long Term Data Preservation (LTDP) best practises. Within this presentation we will show the EUMETSAT approach to safely manage and cope with the expected data volumes for the next 10 years.

EO DATA CONFIGURATION MANAGEMENT: PARALLEL DEVELOPMENT OF LIKE SYSTEMS BY INDEPENDENT AUTHORITIES

richard campbell [1] **, Simon Oliver** [2] **, Michel Douzal** [3] **, Domenico Castrovillari** [4] **, Alessandra Paciucci** [1] **, Francesco ferrente** [1] **, Peggy Fischer** [3]

(1) Serco , (2) GEOSCIENCE AUSTRALIA , (3) ESA , (4) Intecs

ABSTRACT

Abstract: ESRIN, The European Space Agency's centre for Earth observation, and Australia's Federal Geoscience Agency, Geoscience Australia (GA), are undertaking parallel development of methodologies to enable management of Earth observation data collection, configuration and management.
ESA has developed a common approach through a systematic collect, consolidate, configure and report workflow. GA are developing their solution with a view to manage rapid, frequent, iterative collection production. Both approaches, though quite differently implemented, show a great deal of convergence on aspects of managing large EO archives.
This paper investigates the two projects, the drivers that led to their initiation and the scope of each project. We review the reasons that each organisation has embarked on their project and consider how these requirements have shaped the implemented systems and examine the benefits that the chosen solution provides.
The underlying IT infrastructure have facilitated these ambitious projects, we consider the technologies used and how their rapid evolution will impact on these activities.

The paper concludes by reviewing similarities and overlap that these two projects may have and considers if each project may benefit from adopting some aspect of the other project.

CERN\'S 100PB SCIENTIFIC DATA ARCHIVE AND ITS BIT PRESERVATION CHALLENGES

German Cancio

CERN - European Organization for Nuclear Research

ABSTRACT

CERN manages the largest scientific data archive in the High Energy Physics (HEP) domain. The archive currently holds over 100 Petabytes of custodial data from past and present HEP experiments, with some of its data being over 30 years old, and most of it to be preserved ad aeternum. Over the years, the amount of data has been - and will continue - growing exponentially. Influx rates from CERN's Large Hadron Collider experiment are expected to augment from currently 40 Petabytes / year to around 600 Petabytes / year in a few years time, therefore reaching archive volumes in the Exabyte-scale. Long-lived and custodial storage for such volumes requires a realistic estimation of reliability, as bit-rot –the tendency of content in storage to become corrupt over time– is unavoidable at this scale. In this contribution, we will present our experiences and activities regarding storage (or bit-stream) level preservation of such a large archive; including the challenges such as hardware and media failures, bit flip corruption, environmental threats, and technological obsolescence. We will also explain the approaches followed for continuously improving the reliability and availability of our archived data, such as performing integrity verification at multiple storage levels, data replication, and archive migration across technology generations.

USING BUSINESS RULE SYSTEM FOR A SATELLITE DATA PROCESSING MONITORING.

Krzysztof Drypczewski , Marek Moszynski , Waldemar Korlub

Gdansk University of Technology

ABSTRACT

The data flow coming from data producers to Payload Data Ground Segment of European Space Agency missions influences its operations in a crucial way. The model of PDGS data processing and the resulting infrastructure is composed of several components including data ingestion, its management, storage etc. Relying on thoroughly defined requirements for each component the data providers involved in the realization of the process often endeavor to develop reliable tools for simulation of individual subsystems. It becomes important to create the monitoring software that is able to 1) validate and verify component before it is integrated into the final system and 2) monitor its behavior during system usage. Drools® business rule management platform provides a core business rules engine, web authoring and rules management tools ready to be utilized for a variety of specific tasks and can be used for large systems monitoring. In the paper the Drools® application, so called Monitoring Layer (ML) that monitors satellite data products processing is proposed. Aforementioned software is an add-on component for a SWARM mission environment developed by GMV in the frame of the ESA project. The ML is a solution that oversee base EO processing system (which is it independent from) execution, learns its behavior patterns and infers problematic situations. Authors discuss the ML system integration, architecture and Drools inferring rules. For validation of implemented software several scenarios based on actual data processing cases were simulated to prove considerable benefits of employing rule management system.

THE ARCHIVE OF THE EUMETSAT SATELLITE APPLICATION FACILITY ON CLIMATE MONITORING

Petra Fuchs , Nathalie Selbach , Karsten Fennig , Britta Thies

DWD Deutscher Wetterdienst

ABSTRACT

The EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) generates, archives and distributes widely recognized high-quality satellite-derived products and services relevant for climate monitoring. Several Thematic and Fundamental Climate Data Records (FCDR and TCDR) have been released by CM SAF during the last years and corresponding Interim Climate Data Records (ICDR) are provided continuously based on the related TCDR. These data are stored in a central CM SAF archive at DWD. Thus, users have access to many parameters of the water and energy cycle based on operational satellite instruments. The time series of the climate data records range from 8 to about 30 years with a global coverage for data based on polar orbiting satellites, while those based on geostationary satellite data cover the region of the METEOSAT disk.

CM SAF is offering CDRs generated from ATOVS, AVHRR SSM/I and SSMIS on different polar orbiting satellites as well as from the MVIRI, SEVIRI and GERB instruments onboard the METEOSAT series. These CDRs are made available via a web user interface which also allows applying post-processing procedures, such as the extraction of sub-areas or re-projection.

This contribution will present the architecture of the CM SAF archive and will give a general overview of the CM SAF metadata, formats and identification methods. It will describe the user access to all CM SAF products and services.

PERICLES: A KNOWLEDGE MANAGEMENT PROGRAMME APPLIED TO SOLAR DATA FROM COLUMBUS, APPLICATION TO LONG TERM CLIMATE FORCING DATA.

Christian Muller

B.USOC and the PERICLES consortium

ABSTRACT

PERICLES (Promoting and Enhancing the Reuse of Information throughout the Content Lifecycle exploiting Evolving Semantics) is an FP7 project started on February 2013. It aims at preserving by design large and complex data sets. PERICLES is coordinated by King's College London, UK and its partners are University of Borås (Sweden), CERTH- ITI (Greece), DotSoft (Greece), Georg-August-Universität Göttingen (Germany), University of Liverpool (UK), Space Application Services (Belgium), XEROX France and University of Edinburgh (UK). Two additional partners provide the two case studies: Tate Gallery (UK) brings the digital art and media case study and B.USOC (Belgian Users Support and Operations Centre) brings the space science case study.

PERICLES addresses the life-cycle of large and complex data sets in order to cater for the evolution of context of data sets and user communities, including groups unanticipated when the data was created. Semantics of data sets are thus also expected to evolve and the project includes elements which could address the reuse of data sets at periods where the data providers and even their institutions are not available any more.

B.USOC supports experiments on the International Space Station and is the curator of the collected data and operation history. The B.USOC operation team includes B.USOC and SpaceApps personnel and is thus ideally configured to participate in this project. As a first test of the concept, B.USOC has chosen to analyse the SOLAR payload flying since 2008 on the ESA COLUMBUS module of the ISS. Solar observation data are prime candidates for long term data preservation as variabilities of the solar spectral irradiance have an influence on earth climate. The paradigm of these observations has already changed a lot in the last fifty years from a time where scientists were aiming at determining with high accuracy the "solar constant" which was the total solar energy per surface unit received at the top of the earth's atmosphere to the present situation where the same quantity is known as the total solar irradiance and has been shown by thirty years of space observations to vary of about one tenth of a per cent in synchronism with the solar cycle. Right now, larger variations have been detected at UV wavelengths but their effects on climate and atmospheric chemistry are still a matter of scientific discussion. In this paper, the present stage of PERICLES will be described in the context of the solar case study .

The different tools in development in PERICLES will be shown and their relevance to the SOLAR case study will be discussed.

SOLAR data is exemplary of large data sets relating to the climate forcing, other similar long term data will be discussed in the frame of the tools developed in PERICLES, in particular the long term series beginning with GOME-SCIAMACHY on ERS-2 and ENVISAT and continuing on the METOP series of satellites will be discussed.

SENTINEL-3 SLSTR AND SYN LAND PAC: DISK-BASED ARCHIVE: BEST SOLUTION FOR FURTHER USE AND EXPLOITATION ?

Gilbert Barrot [1] **, Odile Fanton d'Andon** [1] **, Nigel Houghton** [2] **, Pascal Gilles** [2]

(1) ACRI-ST , (2) ESA/ESRIN

ABSTRACT

The Long Term Archive of the Sentinel-3 SLSTR and SYN Land PAC, designed and implemented by ACRI-ST, has been reviewed in April 2015. Only disk-based archive in the Sentinel PAC world, our LTA offers additional functionalities that can be enabled in the future, among others: multiple access points and protocols, generic tasks orchestrator for easy implementation of new functionalities, reprocessing-ready configuration (the reprocessing activity can be performed inside the LTA environment). Unlike tape-based systems, its most important benefit is its equivalent performance in sequential and random data access modes. As soon as the archive is used as the data source for exploitation, this is an unavoidable advantage toward the unwelcome copy of the whole huge archive to a new disk-based system. The random access, the excellent performance of the internal storage elements (1.2 GB/s per NAS, 54 TB/day global retrieval throughput in its initial configuration with two storage elements) and the use of a generic tasks orchestrator (ACRI-ST DPMC system), allow the simultaneous use of the LTA as an archive system and a data source for new applications and collaborative working environments.

SOLAR SPACE SCIENCE PORTAL

David De Weerdt

Space Applications Services

ABSTRACT

PERICLES (Promoting and Enhancing Reuse of Information throughout the Content Lifecycle exploiting Evolving Semantics) is an EU FP7 project that targets long-term preservation and reuse of digital content. One of the case studies in the project focuses on space science experiments and in particular on the SOLAR payload on ISS that uses several instruments to measure the solar spectral irradiance. The B.USOC (Belgian User Support and Operations Centre), one of the project partners, is both operator of the experiment and curator of the collected data.

Long term archival of space science data is critical as the measurements are often cost-prohibitive or, for observational data, impossible to reproduce. In addition, space science experiments are typically very complex, generate a huge amount of raw outputs and require a large amount of metadata that is to be preserved to allow future understanding and reuse of the collected raw data: calibration curves and scripts used, applied rules and regulations, operations plans, schedules and logfiles.

This poster presents the developments carried out by the project partner Space Applications Services to augment the operational workflow at B.USOC with Topic Maps based technologies. This enhances its preservation capabilities and ensures that the experiment outputs, and their metadata, can be easily accessed and exploited by an as wide as possible scientific community. At the same time it mitigates risks that arise due to technology and policy evolutions

VESPA: DEVELOPING THE PLANETARY SCIENCE VIRTUAL OBSERVATORY IN H2020

Stéphane Erard [1] **, Baptiste Cecconi** [1] **, Pierre Le Sidaner** [1] **, Teresa Capria** [2] **, Angelo Pio Rossi** [3]

(1) Observatoire de Paris , (2) IAPS/INAF , (3) Jacobs University

ABSTRACT

The Europlanet H2020 programme will develop a research infrastructure in Horizon 2020. The programme includes a follow-on to the FP7 activity aimed at developing the Planetary Science Virtual Observatory (VO). This activity is called VESPA, which stands for Virtual European Solar and Planetary Access. Building on the IDIS activity of Europlanet FP7, VESPA will distribute more data, will improve the connected tools and infrastructure, and will help developing a community of both users and data providers. One goal of the Europlanet FP7 programme was to set the basis for a European Virtual Observatory in Planetary Science. A prototype has been set up during FP7, most of the activity being dedicated to the definition of standards to handle data in this field. The aim was to facilitate searches in big archives as well as sparse databases, to make on-line data access and visualization possible, and to allow small data providers to make their data available in an interoperable environment with minimum effort. This system makes intensive use of studies and developments led in Astronomy (IVOA), Solar Science (HELIO), plasma physics (SPASE), and space archive services (IPDA). It remains consistent with extensions of IVOA standards.

FACILITATING THE USE OF HIGH-RESOLUTION EO DATA TO SUPPORT THE DEVELOPMENT OF MAPPING PRODUCTS FOR NATURAL DISASTERS.

Daniela FAUR [1] **, Mihai DATCU** [2]

(1) Politehnica University of Bucharest , (2) DLR Deutsches Zentrum für Luft- und Raumfahrt

ABSTRACT

Earth observation capabilities are used to respond to major disasters around the world, for humanitarian aid and security. Satellite derived information needs to be used in combination with additional data to be presented in a proper geospatial context for the work of civil protection agencies and relief organizations. This paper aims to reveal a methodology developed to quantitatively evaluate the impact of a natural disaster over a region. The proposed approach was initiated in the frame of GEODIM Project (http://geodim.meteoromania.ro) whose goal is to develop a Romanian downstream emergency response service in order to contribute to current disaster and risk management approach based on Earth observation data. The project is focused on developing experimental processing algorithms and mapping products for natural disasters (floods, earthquakes, landslides) damage assessment in urban areas based on very high resolution optical and SAR satellite imagery acquired worldwide. The usefulness of remote sensing data for natural disasters damage assessment clearly rely on the number of available images, their type and quality and last, but not least, the timeliness of the data sets, or how delayed are the available post disaster images relative to the damaging event. Previous work demonstrated the use of a semi-automated data processing method in order to reveal and determine the area affected by the disaster, considering both qualitative and quantitative approaches. The proposed scenarios consider knowledge discovery from pre and post event EO images by mapping the extracted data features into semantic classes and symbolic representations like ”buildings”, ”vegetation”, ”streets”, ”bare land”, and ”damaged buildings”, etc. In order to fully exploit the high-resolution EO data a method based on patches is proposed to extract relevant contextual information to be further used to build the situation maps.

SERAD DATA REFERENCING AND ARCHIVE

Martine Larroque

CNES Centre National d'Etudes Spatiales

ABSTRACT

SERAD (CNES Service for Data Referencing and Archiving) Martine Larroque(1), Philippe Gros(1), Pierre Bourrousse(1), (1) CNES 18 av E. Belin, 31401 Toulouse Cedex 9, France EMail: martine.larroquen@cnes.fr ABSTRACT The project SERAD in CNES propose the following objectives - Keep in time, information from instruments and space systems sometimes complex - Meeting the research objectives and use of spatial data for scientific communities over the long term(largetime series, episodic events or non-repeatable) - Ensuring long-term quantity and quality of data produced during missions - Keep information accessible and understandable to sustainably in order to extract the maximum value of scientific or technical - Ensuring sustainable access for reuse without loss of data from scientific missions - Reduce long term costs of preservation through control of all data and a dedicated organization SERAD (Service for Data Referencing and Archiving) is a project that will allow CNES to dispose of a centralized repository structure that collects, stores and disseminates information and metadata on Data that are under CNES responsibility. The objective for CNES is to better handle and to improve the access to this data patrimony. In order to achieve this goal, it is mandatory to identify all the missions which are under CNES responsibility and verify whether these data are properly archived; if not, then, to proceed to the archiving of these data. The SERAD missions are the following: - Constitute and maintain a centralized repository of all data that are under CNES responsibility, - Make long term archive, if necessary, - Make surveys in the data production centers in order to prepare and guaranty the long term preservation of these data in particular in the case of end or stop of data production in the data centers. The objective of this paper is to present the objectives of SERAD and he work that has already been done on this project and an overview of the tools and in particular in referencing data part . In particular the following points will be addressed: - Data Management adding value do data and facilitation of data use - Referencing catalog characteristics and demonstration - General Overview of tools used in the project and need for future tools

EXTENDING THE VALUE CHAIN OF KNOWLEDGE ASSETS AND INTEROPERABILITY WITH FUTURE

Maurizio Lunghi

DEDAGROUP Spa

ABSTRACT

In the Scientific, Cultural heritage and public Administration (SCA) sectors it is fundamental to maximise accessibility and usability of all the data produced, to exchange and compare it with other institutions, to combine together data from various sources and create some 'assets' to be preserved with significant properties very long term or forever in some case. Many projects and initiatives have tried to create conditions for interoperability among different user sectors both to increase usability of data and to enlarge the coverage of the services based on that data. The Linked Open Data is the most explicit example confirming this trend, but to make 'open' your data is not enough if other users find difficulties in understanding formats, attributes, relations, properties, the raw data and context information can be merged together to form some first basic assets reusable by experts not belonging to the same sector, and in order to enlarge the usability of data, the assets could be normalised and certified and made compliant with a semantic model agreed/shared by other communities of researchers. A special attention must be devoted to data with legal value that, respecting privacy and law, can help develop very important services. The systematic use of Persistent Identifiers (PI) for people, institutions, digital and physical objects, in addition to be instrumental for preservation tasks, PI can also act as aggregators to get relations and connections among different contents making possible to enrich the knowledge base of the global user communities. Up to now the SCA communities have faced the digital preservation working together at sectoral level, archives, libraries, hospital, regions, ... what can be the next step? Digital preservation can be considered as 'interoperability with future'. So in order to 'harmonise and reuse' information produced by different user communities it is necessary to have a common vision and some common terminology and definitions, a stable understanding of entities format, properties, relations. So we must face two challenges: i. to maximise the usability of data from different user communities defining some common assets; ii. to preserve these assets and significant properties for time requested by users. We would like to extend the value chain of knowledge assets aggregating data from different sources and user communities (thorugh common semantic and vocabulaies and PI), these added value assets will be understandable usable by researchers in other sectors like for example medicine, biology, ecology, doing so the value will be extended in time and coverage.

EO DATA PRODUCTS CONSOLIDATION APPROACH: METHODOLOGY USED IN X-PRESS CONSORTIUM TO GENERATE A CONSOLIDATED MASTER DATASET

Alessandra Paciucci [1] , Michel Douzal [2] , Peggy Fischer [2] , Gareth Davies [1] , Marco Meloni [1] , Bojan Bojkov [2] , Sebastien Saunier [3] , Jean-Francois Piolle [4] , Stefano Natali [5] , Lidia Saavedra de Miguel [1]

(1) Serco , (2) ESA , (3) Magellium , (4) IFREMER Institut Français de Recherche pour l'Exploitation de la Mer , (5) Sistema

ABSTRACT

As part of the ESA (European Space Agency) Data Service Initiative managed by ESA's Ground Segment Operations Division (EOP-GT), the X-PReSS consortium led by Serco is performing the final data consolidation on ESA missions (ERS1, ERS2, ENVISAT and Third Party Missions) in order to generate master datasets for both future ESA reprocessing campaigns and for data preservation activities. The X-PReSS team, under the guidance of ESA's Sensor Performance Processor and Algorithm section (EOP-GMQ) and its supporting consortium IDEAS+, are defining the approach to proper consolidation. The consolidation activity is typically comprised of two parts: generic activities (like e.g. filename convention alignment, data gap analysis, removal of exact duplicates, etc…) and activities requiring tailoring for different instruments and missions (e.g. analysis and handling of overlaps, identification and removal of corrupted products). X-PReSS is implementing themethodology in several ongoing projects. The X-PReSS team will internally verify the results using a defined verification process and will pass these on to the ESA EOP-GMQ section (and IDEAS+ team) for validation. The main achievements obtained so far will be presented at the congress as well as the on-going activities.

RECONCILING CLIMATE DATA RECORD GENERATION WITH AN OPERATIONAL MULTI-MISSION SATELLITE ARCHIVE AT EUMETSAT

Kristina Petraityte , Harald Rothfuss

EUMETSAT

ABSTRACT

As part of its Climate activities and in the framework of different European projects, EUMETSAT has committed to reprocess various data sets from its Geostationary and Low Earth Orbiting Satellites. The reprocessing of satellite data using the latest operational algorithms for long times series, often involving various satellites and several TB in size, leads to stable and homogeneous Climate Data Records (CDR) which can be used in reanalyses and for climate studies.

Key resource in the CDR generation process is the EUMETSAT Data Centre, a multi-mission, operational Archive, holding all of the past, current and future data from EUMETSAT satellites. The available repository – going back over 30 years in the case of Meteosat - encompasses the raw mission data as well as calibrated radiances and many relevant collections of geophysical data products in support to climate monitoring.

The technical CDR generation process can be divided in four main phases:
1) Obtaining the input data from the EUMETSAT Data Centre;
2) Processing of the data;
3) Ingestion of the processed and validated Climate Data Record into the EUMETSAT Data Centre;
4) Preparing the user data access to the Climate Data Record (Catalogue and Archive);

In all these phases it is essential that despite the large volumes and the strong dependency of the CDR generation process on operational resources, in particular the multi-mission Archive, EUMETSAT Operations of the satellite missions are not affected.

The paper presents the methodologies and strategies followed to ensure a performant and strongly automated CDR generation process integrated in the routine operational activities of EUMETSAT.

MAXIMUM LIKELIHOOD REPRESENTATIONS OF MIPAS PROFILES

Johannes Plieninger , Thomas von Clarmann , Norbert Glatthor

Karlsruher Institut für Technologie

ABSTRACT

In order to avoid problems connected with the content of a priori information in volume mixing ratio vertical profiles measured with the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), a user-friendly representation of the data has been developed which will be made available in addition to the regular data product. In this representation, the data will be provided on a fixed pressure grid coarse enough to allow a virtually unconstrained retrieval. As to avoid data interpolation, the grid is chosen to be a subset of the pressure grids used by the Chemistry Climate Model Initiative and the Data Initiative within the Stratosphere-troposphere Processes And their Role in Climate (SPARC) project as well as the Intergovernmental Panel of Climate Change climatologies and model calculations. For representation, the profiles have been transformed to boxcar base functions, which means that volume mixing ratios are constant within a layer. This representation is thought to be more adequate for comparison with model data. While this method is applicable also to vertical profiles of other species, the method is discussed using ozone as an example.

HIGH-PERFORMANCE ONLINE DATA ACCESS FOR COLLABORATIVE INFRASTRUCTURES

Christoph Reck , Gina Campuzano , Klaus Dengler , Torsten Heinen

DLR Deutsches Zentrum für Luft- und Raumfahrt

ABSTRACT

The value of Earth Observation data is increased by accessibility. While catalogues improve finding datasets, retrieval of the physical data products is still cumbersome and often slow. DLR's German Remote Sensing Data Center (DFD) archives petabytes of data, and it interfaces internal processing chains to the EO Product Library, enabling semi-parallel and performant data driven processing or re-processing to generate value added products. For huge datasets, as those accumulated over years or from the new era of Sentinel satellites, new data access and exploitation mechanisms need to be devised.

We propose to place huge series of data products on online disks in parallel and distributed file systems and provide novel access mechanisms based on existing standards. The system shall additionally enable retrieval of historical data from the archive.

The access mechanisms include, such as:

- Simple catalogue interface, like OGC OpenSearch EO profile
- Full featured catalogue using OGC Catalogue Service Web (CSW)
- Improved dataset browsing with OGC Web Map Service (WMS)
- High performance data access using HTTP, SFTP and GridFTP
- Improved data access using OGC Web Coverage Service (WCS)

These services enable performant local and remote processing in collaborative infrastructures with very high parallelism. These also provide all interfaces from "Discovery" over "Visualization" to "Download" for users with novel clients.

This presentation will give an overview of the architecture, the integration of standards and some performance measurements of the demonstrator system implemented in the OPUS project.

REPORTING ON ARCHIVE PERFORMANCE AND USER ACCESS IN THE EUMETSAT DATA CENTRE

Joaquín Rodríguez-Guerra , Jose Antonio Barba , Harald Rothfuss

EUMETSAT

ABSTRACT

The Data Centre is EUMETSAT's Long Term Archive holding all mission data and products from current and previous European weather satellites. As a multi-mission facility it will also serve as the Archive for the planned future missions, such as Sentinel-3, Meteosat Third Generation and EPS-Second Generation. Users can access the Data Centre free of charge to order the historic data.

Monitoring user access and Archive performance is essential in providing a good user service. Many variables come into play relevant in assessing the user access and tuning the system delivery performance: What is the development of ordered amounts and user activity? Which products are the most requested? How can bottlenecks be detected? Which delivery method are the most popular? These are typical questions the Archive Operations Team tries to answer when deciding how to deal with the resources in the system and where a detailed reporting system provides real benefits.

A Web Reporting tool for internal use, based on JavaScript, has been developed which provides a fully portable user-friendly and configurable HTML site displaying the statistics of the archive. Offline data extractions are conducted overnight to prepare fresh static data for feeding every morning the more than 50 charts provided. Among them:

- Products ordered: popular products to locate in the fastest access areas of the archive.
- User Access: help to understand the needs of the users and how to manage user priorities.
- Performance: provide estimations of delivery times for future requests.
- System Stats: help Operations in the optimal authorisation of large orders.

The basic architecture and the functioning of this tool are illustrated in this paper together with examples of relevant charts.

THE D-SDA STATISTICS SYSTEM: REPORTING AND USER ACCESS AT DLR-EOC

Johanna Senft , Cristian Chereji , Katrin Molch , Eberhard Mikusch

DLR-DFD Deutsches Zentrum für Luft- und Raumfahrt/ Deutsches Fernerkundungsdatenzentrum

ABSTRACT

The Earth Observation Center (EOC) of the German Aerospace Center (DLR) operates the German Satellite Data Archive (D-SDA), which provides archiving and access services to national and international spaceborne and airborne Earth Observation (EO) missions, campaigns, and scientific projects. The technological basis of D-SDA is the Data and Information Management System (DIMS), developed in-house in cooperation with an industrial partner. Making use of flexible, generic components in a multi-mission environment, DIMS manages ingestion, cataloging, archiving, and delivery of the EO data and products. The current data volume held in the D-SDA exceeds 5 petabytes, with an annual increase nearing 1 petabyte.

The D-SDA IT infrastructure carefully tracks all data management and archive operations. This information is a valuable source for generating statistics and reports. Therefore, it is of critical importance to collect, prepare, present, and preserve this information for different needs and customers.

Within the DLR-EOC each mission has its own requirements and their implementation is based on distributed sources. Thus - from a reporting perspective - achieving the necessary level of harmonization in terms of data structure is of paramount importance. On the other hand, the demand for statistical information in such a big system covers a wide range of reports for various purposes: pure technical information, system performance reports, usage statistics, prediction reports and comparisons between systems or missions. All of these are very important, with some being key factors in management decisions.

The D-SDA Statistics System has been developed at DLR-EOC in order to answer these challenges. With its modular architecture, the D-SDA Statistics System collects data from different sources and integrates them in a harmonized data warehouse which is compatible with all of the missions in the multi-mission environment. As a consequence it features a high degree of flexibility in data analysis and aggregation, opening the way to a large variety of statistics and reports. These are presented to the customer via a modern user-friendly interface, which not only generates, disseminates, and archives the reports, but also ensures the security and confidentiality of the information.

This presentation will give an overview of the architecture and capabilities of the D-SDA Statistics System. Examples of important reports underline the system's flexibility and usage.

THE HD(CP)² DATA WAREHOUSE: MANAGED SERVICES FOR A METEOROLOGICAL DATA ARCHIVE

Erasmia Stamnas [1] , **Volker Winkelmann** [1] , **Andrea Lammert** [2] , **Ulrich Lang** [1]

(1) Universität Köln , (2) Universität Hamburg

ABSTRACT

The project 'High definition clouds and precipitation for advancing climate prediction' HD(CP)² is a research initiative, funded by the German Federal Ministry of Education and Research, to improve our understanding of cloud and precipitation processes and their implication for climate prediction.

The observation and simulation domain of HD(CP)² is concentrated on Germany and the Netherlands. In this area, a new observation network of a multitude of instrument types containing ceilometers, supersites, GPS stations, and a radar network is synthesized.

One of the project goals is to organize the observation data derived from this network in a data base. The heterogeneous project data are ranging from ground based long-term observations to short-term area-wide remote sensing observations to high resolution satellite data.

To encourage data sharing and data based research we take a lot of effort in making the data understandable and usable for future re-use. The data is quality checked and accompanied by metadata and supporting documentation.

To ensure that the data can be used in a uniform manner, we have developed the HD(CP)² Observation Data Product Standard. The standard describes in detail the strong file-naming conventions, versioning, and classification of the data according to measuring instruments, variables, and processing state. Furthermore requirements on the metadata, which have to comply with the HD(CP)² - XML Schema, are specified.

An in-house developed online editor facilitates creation and modification of the metadata. In addition, the editor supervises the correct form and adherence to the HD(CP)² (meta) data standard.

The core administration of the observation data is based on Unidata's hierarchical data management system THREDDS (Thematic Realtime Environmental Distributed Data Services).

Supplementary software was self-developed for data processing and monitoring. This software allows users to perform data management and quality control tasks, as well as basic reporting on data access. Additionally, data publishing by the THREDDS Server is provided.

Keywords: data management, data standard, data quality, adding value, user-friendliness, metadata editor

AN INTRODUCTION OF DARTS WEB APPLICATIONS

Hideaki TONOOKA , Ken EBISAWA , Keiichi MATSUZAKI , Yukio YAMAMOTO , Akira YOSHINO , Koji IMAI

JAXA Japan Aerospace Exploration Agency

ABSTRACT

DARTS (Data ARchives and Transmission System, http://darts.isas.jaxa.jp/) is a multi-disciplinary space science data archive of JAXA (Japan Aerospace Exploration Agency) for, e.g., astrophysics, solar physics, solar-terrestrial physics, lunar and planetary science, and microgravity science. We have been developping nice applications which provide quick look images of our archive data, such like "JUDO2" and "UDON" for astrophysics, "C3" for meteorology, "Kaguya 3D Moon Navi" and "Apollo Moon Seismic Monitor" for lunar science, and also a space craft FOV visualizer, "FLOW on the Web". We are going to introduce them at the talk.

STANDARDIZATION OF OBSERVATORIES, INSTRUMENTS AND REFERENCE FRAMES FOR PLANETARY SCIENCES

Baptiste Cecconi , Stéphane Erard , Pierre Le Sidaner

Observatoire de Paris

ABSTRACT

The recent developments on planetary science interoperability showed that a standardization of naming conventions was required for observatories (including ground based facilities and space mission), instruments (types and names) as well as reference frames used to describe planetary observations. A review of existing catalogs and naming for those entities is presented. We also report on the discussions that occurred within the IVOA (International Virtual Observatory Alliance), IPDA (International Planetary Data Alliance) and VESPA (Virtual European Solar and Planetary Access) working groups. A proposal for standard lists, possibly to be endorsed by IAU, is presented and discussed.

THE PLANETARY DATA SYSTEM GEOMETRY METADATA MODEL

Edward Guinness [1] **, Mitch Gordon** [2]

(1) Washington University in St Louis , (2) SETI Institute

ABSTRACT

The NASA Planetary Data System (PDS) recently developed a new set of archiving standards, known as PDS4, based on a rigorously defined information model. An important part of the new PDS information model is the model for observational geometry metadata, which includes, for example, attributes describing the lighting and viewing angles, position and velocity vectors of a spacecraft relative to Sun and to the observing body at the time of observation and the location and orientation of an observation projected onto the target.

Prior to PDS4 there were no standards on what geometry metadata to include in PDS labels. The result is that data sets varied in terms of the geometry metadata in labels from none to fully describing the geometry of an observation. The new PDS4 geometry model provides standardization in the definitions of the geometry attributes and provides consistency of geometry metadata across planetary science disciplines. This standardization will enhance the analysis and interpretation of observational data by the science community and will enable harvesting of the geometry information to support discipline level searches by users to discover data of interest to them.

The PDS4 geometry metadata model is based on requirements gathered from the planetary research community, data producers, software engineers who build search tools and on a survey of geometry data contained in existing PDS data sets. An overall requirement for the model is that it fully support the breadth of PDS archives including a wide range of data types collected by instruments observing many types of solar system bodies such as planets, ring systems, moons, comets, and asteroids. Specific geometry model requirements include: (1) Separate geometry classes to support different mission types, e.g., orbiters and flybys, landers and rovers, and Earth-based observations; (2) Flexible classes that require a minimum set of attributes, but define optional attributes to fit the wide range of planetary observations; (3) Metadata pointers geometry source data and relevant coordinate/reference; (4) Metadata for footprints of observations projected onto a planet's; and (5) Updates to geometry data for improved instrument pointing or spacecraft position.

The PDS4 geometry model is implemented in XML and uses XML schema for validation. Using XML greatly enhances the ability to build a standardized structure for PDS labels in that parameters appear in a specified order in the label, and required and optional parameters are clearly indicated. XML also makes it easier to read the PDS labels using software that can parse an XML document, and label validation is straight forward by testing the label against the model schema.

The geometry model is structured such that high-level components are focused on a specific class of missions, such as orbital/flyby or landed/rover missions. Future implementations of the model will include observations made from earth-based telescopic instruments. The high-level components use lower-level classes that define fundamental objects such as generic vectors and quaternions. The model can be extended for a particular mission by using the generic classes from the lower-level component to extend the higher level model. The high-level model for orbital and flyby missions contains classes for specific distance and velocity vectors (e.g., spacecraft to target and target to sun), lighting and viewing angles, and the projected field-of-view onto the target for both an individual point (e.g., pixel in an image) or the full footprint of the observation. Geometry information can be provided for more than one body in the same PDS label if multiple targets are observed, such as a planet and one or more of its moons. The landed and rover mission high-level component includes classes to define the vehicle position and orientation. It contains classes to describe a camera model for image data. There are also classes that specify the position and orientation during an observation of a robotic arm and its tools.

The XML schema for the geometry model, along with all other PDS4 XML schema can be obtained at http://pds.nasa.gov/pds/schema.

LESSONS LEARNED FROM REEXAMINING COLLECTION-LEVEL METADATA RECORDS

Nancy Ritchey

NOAA/NESDIS/NCEI

ABSTRACT

NOAA's National Data Centers (i.e. National Climatic Data Center, National Geophysical Data Center, National Oceanographic Data Center and National Coastal Development Data Center) recently merged into NOAA's National Centers for Environmental Information (NCEI). The original institutions were established in the early 1950's and 1960's when all data holdings were in physical format. Three decades later digital data became the preferred form of data for preservation. However digital data and metadata management were in its infancy and data documentation was not easily refreshed. Today NCEI follows established data and metadata standards that enable discovery and documents the data for the designated community. Recently NCEI reexamined all weather and climate data collections to ensure they meet current metadata and data preservation standards; the information is consistent across all data management systems; and a data steward is identified for each collection. This presentation will describe the project goals, progress and lessons learned.

ONE-CLICK PRESERVATION OF DATA ANALYSIS SOFTWARE

Luis Dias [1] , Alexandre Constantino [1] , Emanuel Barbosa [1] , Vicente Navarro [2] , Andy Pollock [2] , Carlos Gabriel [2] , Aitor Ibarra [2] , Bruno Merin [2] , Jorge Bakker [2] , Christophe Arviset [2] , Ruben Alvarez [2] , Virginia Martin [2]

(1) EDISOFT , (2) ESA/ESAC

ABSTRACT

One of the important activities of ESA Science Operations Centre is to convert raw data received from Mission spacecraft into science data. This is usually done through pipeline software that systematically processes data in batch mode. Furthermore, the SOC also provides data analysis software to enable users and scientists interactively to process data further to higher levels. In both types of work, calibration data are required to describe how instruments perform. Calibration data are sometimes supplied separately.

Mission Data Analysis Software (DAS) is crucial to enable data to be understood and scientific results and refereed papers to be produced. DAS is developed by the SOC and others during the Mission and usually run on a specific technology stack (OS, compilers, libs, custom-made code ...). Users download the Mission DAS from the web, install it on local computers together with all necessary additional libraries and calibration data, if supplied separately.

During Mission operations and post-operational phases, the DAS is fully maintained and is updated for new OS and library releases. Nonetheless, once a Mission goes into the “legacy” phase, there are very limited funds to maintain DAS and through the years as the IT world evolves, DASW slowly but inevitably also becomes more and more difficult to install and run on modern computers.

While the long-term preservation of data is addressed through the Science Archives, the issue of long-term preservation of Data Analysis Software is still open. Recent evolution of the IT world, and in particular of the Virtual Machine (VM), Cloud computing and Software as a Service (SaaS) paradigms, might provide the answer, ensuring availability to the science community for a period that goes well beyond the end of the active mission phases.

Therefore, building on these technologies, the activity has aimed at providing long-term preservation of Data Analysis Software for the following missions:

- PIA for ISO (1995)
- SAS for XMM-Newton (1999)
- Hipe for Herschel (2009)
- EXIA for EXOSAT (1983)

Moreover, the proposed architecture has been shaped around the following goals / principles:

- Support for missions in operations, post-operations and archive/legacy phases.
- Support for execution of DAS on local and cloud (ESAC & Amazon) environments.
- Support for expert users, who want full access and know how to use the complex DAS, and standard users, who are only familiar with the software and just want to run tailored-made processing pipelines under batch mode.
- Provision of a simple web-based interface abstracting the user from the complexities of the required software stack and cloud computing platforms.

This paper describes the architecture, results and lessons learnt gathered as part of the implementation of the prototype during the execution of the project.

TOSSING OUT THE BABY WITH THE BATH-WATER: WHAT OUR SCIENCE STANDS TO LOSE

Elizabeth Griffin

Dominion Astrophysical Observatory

ABSTRACT

"Data" in science come in all forms and formats, even though most of today's Data organizations (such as the RDA, the WDS, and the PV series itself) are based on the assumption that science's data are all born-electronic. The ONLY science which those modern data support is therefore what can be derived for the past 30 years, at most. But the natural world has undergone changes for aeons, and for well over a century observations that can help us understand better some of those changes have been recorded somewhere, in some format. The planned talk will illustrate the kinds of hot science that risk being misunderstood simply because of an inability (or a lack of leadership or funding) to access our heritage observations, and will outline what is being attempted to redeem the situation before it is too late.

ANALYSIS AND DESIGN OF A PORTAL FOR IONOSPHERIC DATA

Klaus-Dieter Missling , Henrike Barkmann , Christian Krafft

DLR-DFD Deutsches Zentrum für Luft- und Raumfahrt/ Deutsches Fernerkundungsdatenzentrum

ABSTRACT

Since 2004 DLR Neustrelitz operates the Space Weather Application Center – Ionosphere (SWACI). This Center is involved in several projects of the Space Situational Awareness (SSA) program of ESA and of EU FP7. It is largely based on services and tools of DLR's Data and Information Management System (DIMS) and its long-standing User interface EoWEB. SWACI provides access to near real time products and ensures long-term preservation. The lecture will analyze the existing solution and various project-specific applications and will derive requirements for a recent user interface. In the second part the talk will draw a design of an user interface for the Ionospheric Monitoring and Prediction Center (IMPC), which continues utilizing existing services of DLR, offers OGC compliant interfaces and which is so generalized that future projects can be easily plugged in.

FENGYUN SATELLITES DATA SERVICE USER ANALYSIS: A WEB-BASED SURVEY

Di Xian [1] **, Li Xue** [2]

(1) CMA/NSMC National Satellite Meteorological Center , (2) CMA/NMSC National Meteorological Satellite Center

ABSTRACT

Fengyun satellites are developed by China Meteorologica Administration (CMA). With the open policy, data and products from Fengyun satellites and other related satellites are sharing via internet for about one decade. 23 satellites and 95 products can be found and downloaded on National Satellite Meteorological Center (NSMC) website. To increasing influence of Fengyun satellites data sharing service, a web-based survey held in the end of 2014 for two months. The object of this survey was all data users who have downloaded and processed data from NSMC website. Users can find a notification in their email boxes and on the visible corner of the website. Up to the end of this survey, 417 valid respondents have finished all questions on line and submitted correctly. This article analyzes the statistical results from respondents and annual reports of Fengyun satellite data services, describes following aspects: user description, data utility, data service opinion, and social analysis. The results confirm the increasing influence of Fengyun satellite data in scientific researches, meteorological services and other social activities. People are paying more attention to Fengyun satellites than other satellites, because they are easy to get and its dataset are becoming more and more continues, and rising data qualities. However, there still are some not ideal and inconvenience part in the whole system. All those suggestions will be taken in the next system architecture.

THE LEGACY DATA ARCHIVES: APOLLO AND VIKING SEISMIC EXPERIMENTS

Yukio Yamamoto [1] **, Yamada Ryuhei** [2] **, Yosio Nakamura** [3]

(1) JAXA Japan Aerospace Exploration Agency , (2) National Astronomical Observatory of Japan , (3) The University of Texas at Austin

ABSTRACT

About forty years ago, Apollo and Viking missions brought important data for human beings. In both missions, the data obtained by seismometers were transferred to and archived on reel-to-reel tapes. DARTS now provides both the Apollo and Viking data with all the avaiable information. Although these missions were performed by NASA, this recent data restoration and archiving activity were performed by tapes stored at ISAS/JAXA. These tapes were obtained from the University of Texas at Austin twenty five years ago as a result of collaborative effort to migrate the data from 7- and 9-track tapes to 8-mm video cassette tapes. In this presentation, we report the details of this activity.