Earth Observation Payload Data Long Term Archiving
The ESA's Multi-Mission Facility Infrastructure

Gian Maria Pinna
EO Ground Segment Department, ESA Directorate of Earth Observation Programmes
ESA/ESRIN Via Galileo Galilei,
00044 Frascati – Italy
e-mail: GianMaria.Pinna@esa.int

Eberhard Mikusch, Manfred Bollner
German Aerospace Center, German Remote Sensing Data Center, DLR/DFD
82234 Weßling – Germany
e-mail: Eberhard.Mikusch@dlr.de, Manfred.Bollner@dlr.de

Bernard Pruin
Werum Software & Systems AG
Wulf-Werum-Strasse 3
D-21337 Lueneburg – Germany
e-mail: Bernard.Pruin@werum.de

Abstract

In 2003 the European Space Agency has launched, in line with the ESA's Oxygen initiative, a strategy for the evolution of the several Earth Observation (EO) missions ground segments (handled and/or to be developed) into an open multi-mission architecture, which includes as main goals:

- Adoption of a common architecture for all missions
- Decomposition of the facility architecture into functional block elements
- Harmonization and standardization of interfaces
- Evolution of current missions payload data segments into the common architecture
- Re-utilization of already available and tested elements
- Development of a generic multi-mission infrastructure where the elements specific to each EO mission can be plugged in

For this purpose, the Agency has already or is in the process of harmonizing within one coherent frame some of the facilities basic functional elements as long-term and on line archive, order handling, systematic processing, product distribution including online delivery and inter-facility product exchange. Support features like monitoring and control and software management complement the framework. The long-term goal of this effort is to define a harmonized European infrastructure.

The resulting architecture, based on the ISO 14721:2003 OAIS (Open Archival Information System) standard [1] [2], is named Multi-Mission Facility Infrastructure (MMFI) and forms the common infrastructure over which ESA intends to build its future payload data ground segments.

The MMFI, via its Multi-Mission Facility infrastructure Elements (MMFE), provides data producer oriented services for data archiving, data processor oriented services for data retrieval and processing management and consumer oriented services in support of the ESA central infrastructure services.

The first instance of the MMFI has been designed in the framework of the ADAR project [3] and is utilized by ESA for several of its present operational missions and used as framework for the implementation of future ground segments. Starting from this infrastructure, ESA is in the process of further rationalizing and enhancing the MMFI with the FEOMI (Facilities Evolution into and Open Multi-mission Infrastructure) project, which aims at porting all present operational ESA's ground segments (i.e. Envisat, ERS and Third Party Missions) into the MMFI. The FEOMI project is the main step towards the consolidation of the facilities infrastructure and the evolution of the current specific missions to the common open architecture. It provides advanced features for long-term preservation and value adding.

1. Rationale

The European Space Agency today manages the payload data operation of a number of Earth Observation satellites since 1975. The activity includes acquisition, archive, processing and products distribution of data from ESA and Third Parties missions, for which more than 1 PetaByte of data is presently archived. The activity is performed via a network of facilities distributed in Europe and in Canada (for ERS only) mostly belonging to National and private entities, operating on behalf of the Agency via contractual agreements. The management centre of this network of facilities is located in the ESRIN ESA centre of Frascati, near Rome – Italy.

The ESA EO activity started in 1975 and is currently progressing with the operation of various ESA and Third Party satellites: ERS-2, Envisat, Landsat 5/7, NOAA 16/17, SeaWiFS, Spot 2/4, Terra, Aqua, Proba. Future missions are also planned to fly, which will be managed by the ESRIN EO network of centers: Cryosat, GOCE, SMOS, ADM, etc.

2. ESA's Strategic Objectives for Multi-mission EO Data Exploitation (Oxygen Principles)

In 2003, the ESA's Directorate of Earth Observation embarked in a new activity aimed at facilitating the access to EO data from ESA's owned and other missions. The initiative, nicknamed Oxygen, has as its main goals to:

- increase the sustainability of data provision by widening the range of data source
- demonstrate the immediate benefits of a cost-efficient Ground Segment infrastructure for future European and member states missions
- respond to user needs for access to 'maximum number' of mission data
- prepare the 'Infrastructure Component' for the ESA's GMES (Global Monitoring for Environment and Security) proposal along with the mission and service proposals to assure coherent, planable, sustainable EO input to GMES
- establish the (mission independent) single ground segment infrastructure and operations concept for the new missions

The Oxygen initiative will be the basis for an implementation plan, to be exploited in the period 2003-2005, based on three pillars:

C-Pillar: Providing access to current data sources

- Identification of elements that Oxygen could benefit from and inventory of existing infrastructure imply:
  - Technically, the creation of EO Web Portal for awareness of missions, user services, products, information and service providers; the creation of an infrastructure for the provision of online products and the improvement of the online exchange of data, the rationalization of the present archives, ensuring their integrity and access for ESA and national missions support; the harmonization of the ground segments for current and future missions.
  - Policy-wise and programmatically, the enlargement of data sources through 3rd parties cooperation; the evolution of the present Data Policy, adjusting it to the GMES service needs and to internal cooperation requirements; the setting up of a Ground Segment Task Force to define the goals for European EO Infrastructure.

S-pillar: Developing Services

- Prototype Service development
- Build up a new "Service Development Environment"

N-pillar: Integrating upcoming EO National projects in Europe

- Feasibility assessment at political, technical and programmatic level

The following chapters will describe the approach followed by ESA for the harmonization of its ground segments, aimed at satisfying one of the most demanding requirements of the Oxygen's C-pillar.

3. Ground Segments Harmonization

The harmonization of the Payload Data Ground Segments (PDGS) starts from the analysis of the commonalties that can be exploited among them. As schematically shown in the next Figure 1, the analysis of the elements composing a typical EO ground segment shows that many services are in common among all missions and can thus be operated by common elements.

The archiving system, the data management system, the network, generally the systems that perform the data packaging for the users, and the user services, including the catalogues and the user access functions, are all elements that can be used as part of a common infrastructure. Present and future missions can benefit from a common approach for the development and deployment of these elements and achieve savings in the development during the project phases B/C and the following operations during phase E. Into this infrastructure, mission specific elements like processors, acquisition systems, quality control systems and other elements can be plugged in, provided that they conform or adapt to standard interfaces.

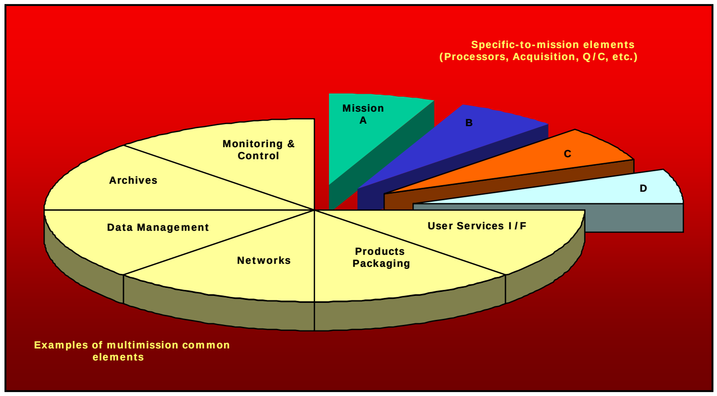

Figure 1 PDGS Decomposition in multimission and mission-specific elements

The approach for the achievement of the Oxygen goals was based on several initiatives, one of which being the development and adoption of a common architecture for all missions and applicable to all centers part of the ESA's network of facilities. The main activities and tasks to be realized in order to develop and implement this common architecture were:

- Decomposition of the facility architecture into functional block elements
- Harmonization and standardization of interfaces
- Standardization of products and formats across missions
- Evolution of current missions payload data ground segments into the common architecture
- Re-utilization of already available and tested elements
- Priority-based integration and operational phase in
- Develop a generic multi-mission infrastructure where the elements specific to each EO mission can be plugged in
- Harmonization and rationalization of archives

The basis for this implementation was during 2004 an ESA's R&T project, called ADAR (Advanced Data ARchive) [3]. The ADAR project aimed at defining on one side a conceptual model for the ESA EO Payload Data Ground Segment, and on the other at identifying the multi-mission elements that were potential candidates to be reused in the development of future ground segments. At the same time the phase 1 of the ADAR project tried to identify multi-mission infrastructure elements that were potential candidates for future development or procurement. The project was completed by the development of a prototype environment where the main concepts and systems are demonstrated.

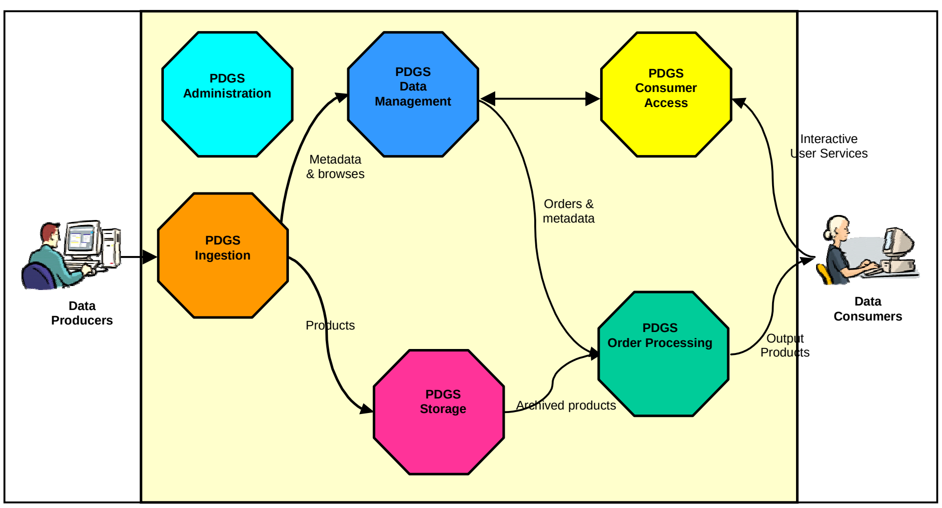

Figure 2 Payload Data Ground Segment Logical Model (OAIS-based)

The Figure 2 on the left represents the logical model of the generic ESA PDGS, as defined during the ADAR study and based on the OAIS Reference Model.

Due to the distributed nature of the ESA's ground segments, composed of many acquisition and archiving centers, a PDGS for a generic mission can be considered as is composed of (see Figure 3):

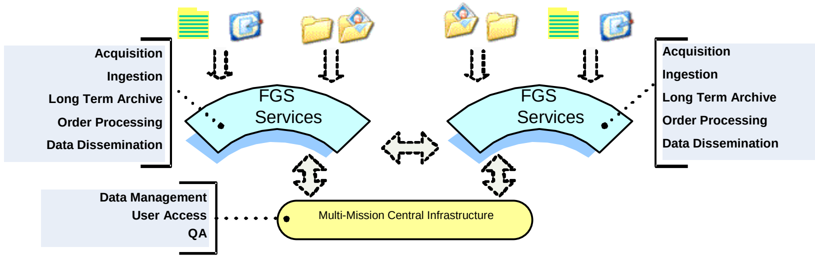

Figure 3 PDGS Generic Decomposition

- a Multi-Mission Central Infrastructure component, consisting of all elements required to provide User Services (cataloguing, user access, data ordering, etc.), and Quality Assurance services (payload data quality control, sensor performance assessment, etc.)
- a distributed Multi-Mission Facility Ground Segment (FGS) component, consisting of all elements necessary for the acquisition, ingestion, long-term archive, order processing and data disseminations to end users of a specific mission. As highlighted before, a generic FGS generically requires elements that are mission-specific like processors and quality control systems, but most of its services can be provided by means of common multi-mission elements. Multi-mission centers typically have one “logical” FGS for each mission operated, but the sub-systems used can be shared if they have built-in multi-mission characteristics.

4. MMFI Architecture

As discussed in the previous sections, the services offered by the Facility Ground Segment, can be provided by a mixture of common elements shared among multiple FGSs, and few mission-specific elements.
In designing a harmonized solution, the specific characteristics of the ESA PDGSs had to be taken into account. ESA operates its ground segments in distributed centers within Europe and additional acquisition stations around the globe. While data archiving and processing is done at the centers, mission planning as well as search and ordering facilities are located in centralized facilities in ESRIN. For higher level data processing and in specific cases data consolidation, the centers are required to exchange some of their data assets.

The common system elements deployed in each centre constitute the ESA’s Multi-Mission Facility Infrastructure (MMFI). In line with the principles highlighted above, the MMFI should consist as much as possible of COTS and MOTS to reduce costs and to profit from a potentially increased functionality and higher level of stability of these components.

In terms of the OAIS reference model, such a distributed system could be called a federated, co-operating archive system. In order to fully cover and highlight the complex aspects of value adding through processing, the OAIS model was extended by a model element "Processing" that, in essence, is both a data consumer and data producer in terms of OAIS (see Figure 4). Combining the two elements into new model elements highlights the control loops that are introduced when systematic processing or re-processing is done and allows distinguishing it from standard access functionality for on-request value-adding processing.

The dynamics of the distributed archive with federation and co-operation also triggered the introduction of explicit interoperability model elements for data and metadata exchange that cover the synchronization with the central catalogue for metadata and browse images as well as data exchange between sites. The resulting logical model for the requirements process

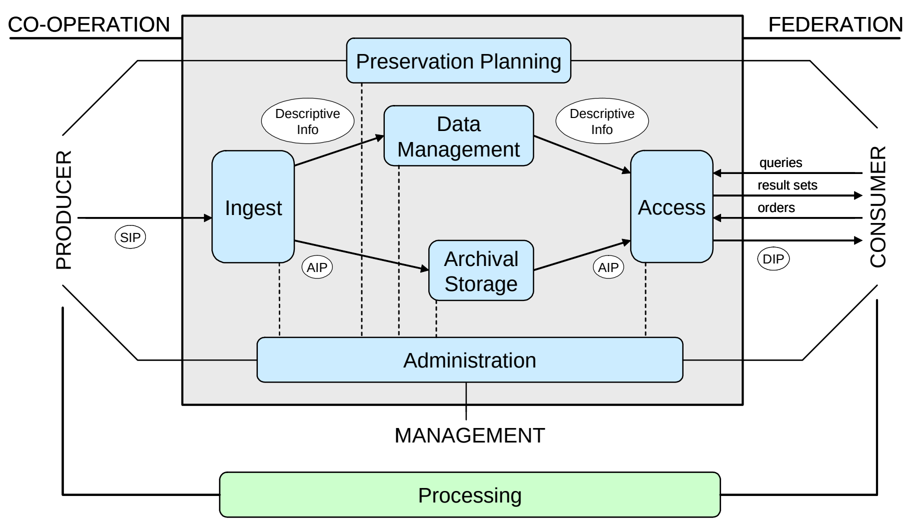

Figure 4 OAIS Reference Model Extended by a Processing Element

is shown in the Figure 5 on the left.

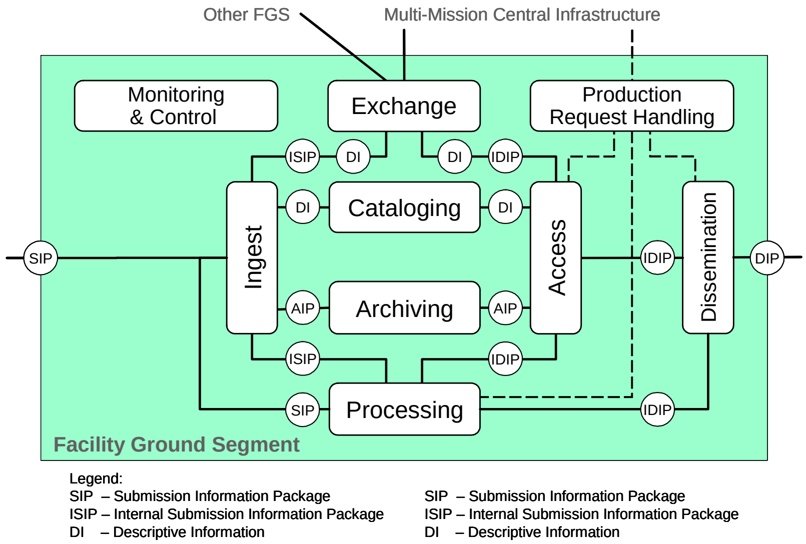

Figure 5 Facility Ground Segment Logical Model

The model elements were further mapped onto components that were either already in operations in the ESA FGSs, could be created by configuration of COTS or required to be custom build.

The project FEOMI (Facilities Evolution into an Open Multi-mission Infrastructure) was the second project, after ADAR, that consolidated the design of the ESA's operational FGSs and of the MMFI that composes them.

The final generic architecture of the MMFI, as consolidated during the FEOMI project, is shown in Figure 6, where the MMFI elements are shown in blue and the mission-specific elements in purple.

The main building blocks of the MMFI are:

Ingestion

The ingestion workflows are controlled by the "Generic Front End" (GFE) that provides a configurable workflow engine and a standard set of re-usable plug-ins relevant to ingestion. For data product dependent functionality for metadata extraction and browse generation the ingestion assembly relies on the functionality provided by the Data Request Server (DRS) that allows accessing data (including binary data) in a configurable manner.

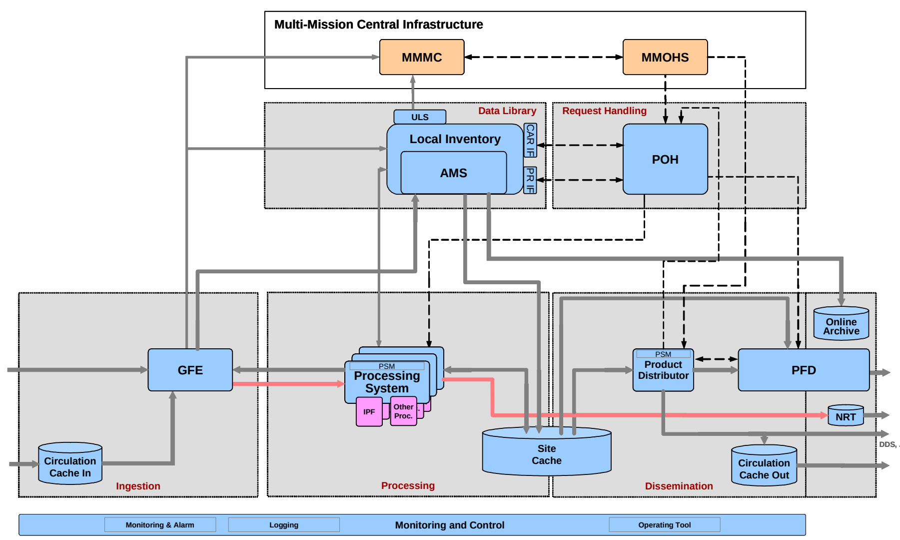

Figure 6 Generic MMFI Architecture at an ESA centre

Data Library

The data and metadata holdings of a centre are stored in a data library assembly that consists of the data storage component "Archive Management System" (AMS) and an inventory component "Local Inventory" (LI). The AMS manages the actual archiving of the data product and as an abstraction layer allows to change the underlying storage technology without affecting the architecture of the overall system. It is worth noting that, in line with the goal of achieving the maximum harmonization and standardization of the FGSs, a unique storage technology has been adopted by ESA for its EO missions. The migration of ESA's holdings to the new archives is presently on-going.

Request Handling

Interfacing ESA's Multi-Mission Order Handling System (MMOHS) is a request handling subsystem, the POH (Product Ordering System) that handles production and dissemination requests and organizes the required workflow based on the product type and output medium requested in the order. The POH is supported by a set of auxiliary components that interface other MMFI elements and provide specific functionality for workflow management.

Dissemination

ESA centers need to support a variety of dissemination scenarios and dissemination channels. The key component for dissemination is the "Product Formatting and Delivery" (PFD), a dissemination workflow management component that can handle different dissemination channels in a configurable manner, and "Product Distributor", a component that manages systematic delivery procedures for data circulation between centers, subscriptions and standing orders.

Processing

The MMFI support processing by providing a generic "Processing System Management" (PSM) framework. The framework allows integrating mission and sensor specific processing facilities with minimal effort and offering a choice of protocols for the definition of the MMFI - processing facility interface.

Additional there are utility components for service monitoring, logging and automated monitoring and alarm.

The MMFI is used to implement Facility Ground Segments for all current ESA missions including ENVISAT, ERS and past and current Third Party Missions. The MMFI based ground segments are presently deployed at the ESRIN centre in Frascati (Italy), Matera (Italy) , Oberpfaffenhofen (Germany), Maspalomas (Spain), Farnborough (UK) and Kiruna (Sweden).

5. MMFI Advanced Features for Long-Term Preservation and Value Adding

The MMFI provides comprehensive functionality for processing and archiving centers by the use of modular components as described in the previous section. Within this section we will especially focus on the features of the MMFI covering preservation and value adding concepts.

The MMFI architecture follows basic principles within the fields of preservation and value adding:

- It supports operational scenarii for preservation strategies, like the periodic migration of digital products to new information technology.
- It provides encapsulation by self-describing items as defined by the information packages in the OAIS model. The maintenance of metadata is performed and means for data consistency are supplied.
- It supports automated production by means of sophisticated data access and processing management.
- It follows a modular design, has an open architecture and streamlined interfaces that permit an easier substitution of one or more of its elements, if the need will arise in the future to ensure the long-term preservation of its data holdings and of its services.

Special attention was turned on the architecture for a well balanced assignment of functionality-to-components which covers both concepts in an effective way. The result is a number of advanced system design and functional features which are described below.

System design features

Archiving technology migration

A major challenge of long term data preservation is to cope with changing storage technology in an efficient manner. The MMFI provides concepts and layers of abstractions to support technology changes without undue impact on the system components. The physical data-sets are shielded from the applications by several software layers, that can be all used to handle lower level technology changes. In its simplest form, changes in tape technology are covered by the hierarchical storage management (HSM) solution incorporated in the AMS. A potential change of the HSM technology would be fully resolved in the AMS. There is only a limited number of components that directly interface the AMS, thus also a change of the AMS or AMS interfaces could be handled with minimal archive impact.

Technological evolution and scalability through modularity

The MMFI is largely built from autonomous, networked components that can be combined by configuration allowing to exchange some components by other implementations easily or to handle increased load by instantiating more components in optimised configurations. An example is to improve the throughput of a processing system by just cloning it onto additional hardware.

Advanced algorithm (processor) integration

Long term preservation considerations are not restricted to the archived data itself, also the data processors used for value adding to the data are subject to long term changes either of their implementations by scientific or technological advances or by changes of the operational environment when HW platforms and operating systems evolve or become obsolete. The MMFI uses a Processing System Management (PSM) framework that supports the integration of processor by natively supporting a variety of processor interfaces and by allowing integrating processor adapters. The flexibility of this approach makes it possible to substitute processors and processor interfaces without undue effort and without affecting other parts of the participating workflows.

Product data model migration

The evolution of processing algorithms requires changes in the data models of the products to be archived. To cope with this type of change, the MMFI provides configurable product object models within the data library that can be extended.

Functional features

In the functional area a comprehensive set of features or services is needed for preservation and value adding. Several MMFI components contribute to a specific service but at the same time retaining there individual autonomy. Table 1 tries to give a complete list of the functional features with a short description.

Cataloguing and archiving	Cataloguing and archiving is the basic feature to consistently manage data and metadata for long-term. In addition advanced access capabilities are available to search and retrieve products for automated production.
Automated Product Ingestion	During automated product ingestion data products are checked for consistency before archival. Metadata are extracted and where applicable browse images are generated for catalogue applications.
Order driven processing and delivery	Order driven delivery is the classic dissemination workflow initiated by a user order. Optionally a value adding processing step may occur before delivery.
Systematic data driven processing	Systematic processing describes the capability to initiate automatically processing workflows for higher level products upon the reception of a lower level product.
Systematic re-processing	Systematic re-processing is used to generate a new revision of a product collection due to processing algorithm or configuration update. It’s a processing schema with data from/to archive.
Systematic dissemination	Subscription type systematic dissemination is similar to systematic data driven processing with the difference that the newly arrived products are not processed but delivered to one or more customers. The dissemination process may include optional post-processing or product re-formatting steps.
Online archive access	Online archive access allows to directly retrieving the product data in a file based transfer protocol.
Data circulation	Data circulation is concerned with the distribution of data between centres to serve data migration purposes incl. auxiliary products for remote processing.

Table 1 Summary of functional features

Systematic workflows and relevant mechanisms

The systematic workflows – as can be seen from the table above - are getting more and more important in an automated environment. Those workflows are considered to be quite advanced in the MMFI. Therefore their description is detailed depicting also the relevant mechanisms of the components involved. For this purpose Figure 7 in the next page is used to illustrate the systematic data driven processing, systematic re-processing and systematic dissemination. The data library is expanded for raw (Level 0) and value added (Level x) data areas and the necessary data and control flows are shown.

The foremost value adding process of the MMFI is the systematic data driven processing that, for every newly acquired input or set of inputs, initiates a processing workflow to compute higher level products for archiving and dissemination. Within the MMFI the domains acquisition, processing and dissemination are decoupled for operational stability and clear assigned responsibilities. For example a processing chain in maintenance should not require changes in an acquisition chain to avoid cache overflow. The decoupling of the sub-workflows is achieved through the subscriber functionality of the local inventory. Successor workflows register for changes in their input data collections and react on notifications by initiating their processing. This functionality can be chained to provide reliable value added processing up to any level.

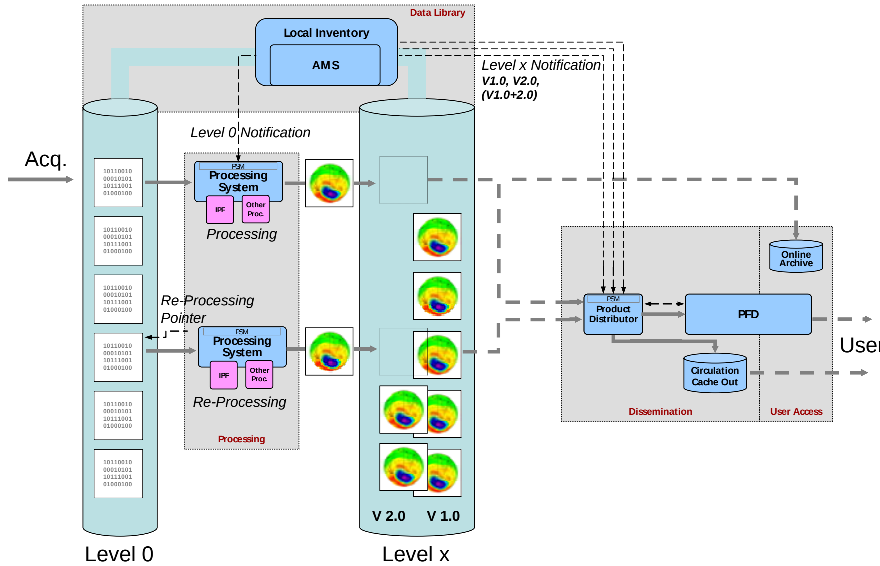

Figure 7 Major systematic workflows in the MMFI

Systematic data-driven dissemination, like processing, utilizes the Local Inventory subscriber functionality to drive dissemination by notification. A subscription specification can be parameterized by arbitrary metadata filters to ensure that only the desired set of products is delivered or checked. As depicted in Figure 7, dissemination workflows can be set up to deliver all products carrying the version number V1.0, only those of version number V2.0 or both. Additional metadata can be used to separate the datasets resulting from systematic data-driven from those resulting from re-processing.

Systematic re-processing is a standard processing facility functionality that is needed to cope with algorithm advances. Most datasets are only useful for long term analysis when the complete set is processed with the same processor version. Processor version changes within the scope of an analysis often leads to unwanted artifacts in the higher level data. To avoid this, the introduction of a new processor version often requires the re-processing of all historical data. The re-processing exercises may be long running activities that are performed with a limited load to ensure that other nominal facility workflows are not blocked or slowed down. When implementing re-processing workflows, the order of the processing (chronological, inverse chronological, sorted by some metadata value, see “Re-Processing Pointer” in Figure 7) as well as the scope of activity can be controlled in a configurable manner based on an object oriented metadata query language (OQL) provided by the local inventory [4].

References

[1] ISO 14721:2003, “Space Data and Information Transfer Systems - Open Archival Information System - Reference Model”, Edition 1, February 2003

[2] CCSDS 650.0-B-1., “Reference Model for an Open Archival Information System (OAIS)” – CCSDS Blue Book, Issue 1, January 2002

[3] G.M. Pinna, P. Mougnaud, G. Petitjean, C. Demange – “ADAR, a Study for an Advanced Data ARchiving System for Earth Observation” – SpaceOps2004, Montreal, 17-21 May 2004

[4] Kiemle, S., Schroeder-Lanz, A.-K., Reck, C., Object Query Language – “Enabling Service for Earth Observation Product Processing, Access and Dissemination” – Ensuring the Long Term Preservation and Adding Value to the Scientific and Technical Data, PV 2004, Frascati, 5-7 October 2004, ESRIN, ESA, WPP-232, p. 17-24, (2004)