From HARM to safe
The ESA's Proposal For A Standard Archive Format For Europe

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

Stephane Mbaye, Mathias Moucha
GAEL Consultant
Parc Nobel Descartes
25, rue Alfred Nobel
77420 Champs-sur-Marne – France
e-mail: Stephane.Mbaye@gael.fr, Mathias.Moucha@gael.fr

Valter Spaventa, Davide Castellazzi
Advanced Computer Systems S.p.A., Space Division
ACS, Via della Bufalotta, 378,
00139 Rome – Italy
e-mail: vspaventa@acsys.it, dcastellazzi@acsys.it

Abstract

The volume of EO data archived by ESA exceeds 1.5 PB and is in constant increase with the on-going operations of ERS, ENVISAT and Third Party Missions. Prediction is that this volume will exceed by far 2 PB in few years. New datasets, coming from the future Earth Explorer and GMES missions will contribute to increase the volume in the years to come. All these datasets are archived for the long term in several centres around Europe and are used to generate systematic and on demand end user products. The ESA's mandate is to preserve the datasets it owns for many years, normally 10, after the end of the mission active life. On the other hand, it is well known that the data to be preserved for the long term require a special attention, which reflects in costly operations for their exploitation and maintenance, one of main reasons being the excessive proliferation of diverse and heterogeneous data formats, caused by mainly three reasons:

- *The lack of an agreed standard in the Earth Observation community, reason for which the formats tended to be specific for the sensor(s) each mission carried on board.*
- *Legacies from old ground segments architectures, which tended not to reuse elements previously developed.*
- *The non-mature status until recently of the information technologies and standards used to describe and package the data, preventing the creation of a unique format able to satisfy at the same time the requirements for the long-term data preservation and their handling in the processing centres.*

The ESA's Earth Observation Department has since time recognized the need to standardise and harmonise its ground segments architectures to reach scale economies during development, operations and maintenance, including the need to standardise the format in which the datasets are preserved. By achieving this goal, the road would be also paved for simplifying the exchange and interoperability of data between ESA and external operators. In early 2004 ESA has setup a project called HARM (Historical Archives Rationalization and Management), which aimed mainly at converting its historical datasets into a new modern format, based on the latest technologies and standards and able to ensure the long term preservation of its holdings.

The SAFE (Standard Archive Format for Europe) has been designed to act as a common format for archiving and conveying data within ESA Earth Observation archiving facilities. As such, SAFE benefits from the experience gathered while developing standards related to data formats. SAFE intends to resolve the major challenges coming from the packaging and the long-term preservation of Earth Observation data. A special attention has been put to ensure that SAFE conforms to the ISO 14721:2003 OAIS (Open Archival Information System) reference model and related standards like the emerging CCSDS/ISO XFDU (XML Formatted Data Units) packaging format. Although the primarily goal of SAFE, in the framework of the HARM project, is to handle EO data with processing levels close to the usually called "level 0", no limitation exists regarding the packaging of higher level products as well as other technical and scientific information. Actually, experience has demonstrated that packaging and archiving higher processing levels or auxiliary data in a common format may be effective in many situations. SAFE undergo this concept by offering a single framework for packaging a large variety of information.

1. ESA's Long-Term Archive Strategy And Rationale

The European Space Agency 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 Party missions, for which more than 1.5 PetaByte of data are 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 archives activity is currently progressing with the operation of various ESA and Third Party satellites from past and present missions: ERS-1/2, ENVISAT, Proba, Landsat, NOAA, Nimbus, MOS1-1b, JERS-1, IRS-P3, SeaWiFS, Spot, Terra, Aqua, Proba, Kompsat, Scisat. Future missions are also planned to fly, which will be managed by the ESRIN EO network of centres: GOCE, SMOS, ADM, SWARM, GMES, etc., plus a number of Third Party Missions, which will substantially increase the size of the long-term archive.

The normal mandate for ESA's EO missions is to maintain the archive for at least 10 years after the end of the mission. The management of this vast amount of heterogeneous datasets poses problems for their long-term preservation, and this is why the EO Ground Segment Department has recognized since many years the need to establish a clear strategy for the management of the long-term archives.

In 2001 the ESA's Programme Board on Earth Observation (PB-EO) endorsed a strategy for the "Management of Historical Archives" (ref. PB-EO(2001)4).
In 2003 DOSTAG (the technical advisory board of the ESA's PB-EO) endorsed a document ESA/PB-EO/DOSTAG(2003)6 related to the promotion of products and services across missions and exploitation of historical archives.
Furthermore, in 2003 the Oxygen implementation plan (ESA/PB-EO(2003)51) has reinforced and enhanced in a wider context the overall issue of data archive and improved data sources user access and availability.
Finally, in 2004 the ad-hoc nominated PB-EO Ground Segment Tasks Force has concluded his report with a set of recommendations (ESA/PB-EO(2004)53) related to the facilities and archives infrastructure aiming at:

- maximized competitive approach
- enhanced infrastructure exploitation
- facilities and operations rationalization
- technology exploitation
- cost reduction
- possible standardization and re-utilization at European level of the Agency investment.

The ESA's Earth Observation Department has since time recognized that the main process to be undertaken is the standardization and harmonization of its ground segments architectures to reach scale economies during development, operations and maintenance. This includes the need to achieve the following goals:

- archive maintenance in order to ensure data integrity
- archive and data management rationalization
- data conversion to new technologies in order to reduce cost of operations
- enhancement of data access
- standardization of the format in which the datasets are preserved.

By achieving these goals, the road would be also paved for simplifying the exchange and interoperability of data between ESA and external operators.

It is well known that the data to be preserved for the long term require a special attention, which reflects in costly operations for their exploitation and maintenance. Among these challenges:

- the datasets have to be regularly converted into new media technology, to prevent the problems created by their obsolescence
- being the long-term archive normally based on datasets archived up to a very low processing level, the so called L0, higher level products have to be generated by processing systems
- also in the case of distribution of the data holdings directly in the archived format, it is normal practice that they are converted or reformatted into a format more oriented to the end-user utilization
- it is a common requirement to have to extract from the long-term archive a portion of the single data file (subsetting) to create a "child" product, optionally to be pre-processed
- finally, more and more the data are distributed to the end-users, and exchanged among data holders, in electronic format over network infrastructure (private Intranets, public Internet, academic networks, etc.)

One of the reasons that contribute to the high operations and maintenance costs of the long-term archives is the excessive proliferation of diverse and heterogeneous data formats, caused by mainly three reasons:

- the lack of an agreed standard in the Earth Observation community, reason for which the formats tended to be specific for the sensor(s) each missions carried on board
- legacies from old ground segments architectures, which tended not to reuse previously developed elements
- the non-mature status until recently of the information technologies and standards used to describe and package the data, preventing the creation of a unique format able to satisfy at the same time the requirements for the long-term data preservation and their handling in the processing centres.

Taking all this into account, in early 2004 ESA has setup a project called HARM (Historical Archives Rationalization and Management), which aimed mainly at converting its historical datasets into a new modern format, based on the latest technologies and standards and able to ensure the long term preservation of its holdings.

2. The HARM Project

The HARM project aims at achieving the following main objectives:

- analysis and definition of a new ESA standard long-term archive data format
- conversion of archived data from the original format into the new format and transcription onto the ESA's automated long-term archive, based on the ESA's MMFI (Multi-Mission Facility Infrastructure) architecture
- rationalization of the archives by centralization of historical data (non-active mission) and archive of on-going missions into specialized centres
- optimization of the archives, reducing operator intervention for system maintenance and improvement of archive capacity through removal of redundant data (whenever applicable and due to orbit overlap, duplicated transcriptions etc.)
- generation of the related metadata and inventory information, to be ingested in the centralized ESA catalogue.

The rationale behind the HARM project is the prevention (and sometimes the recovery) of the problems created by the archiving media obsolescence, achieved by transferring the archives on a state-of-the-art architecture and technology. At the end of this transfer all the data now archived on various media (DLT, Sony D1, Exabytes, IBM 3590, etc.) will be available on line or near-line on a robotized long-term archive module of the MMFI, interfaced as a data server by the client applications. The archive tape technology will be the presently adopted ESA standard, which is StorageTek T9940B.

One of the main HARM goals is to optimize the ESA historical archives applying an Orbit Overlapping Compression (OOC) (stitching). This processing, applied on the data after SAFE conversion, can significantly reduce the size of some critical missions and improve the overall archive quality by a selection of the best data on a line-per-line basis.

The following table gives, as an example, the present (at 1st of March 2005) archives status and its potential reduction (obtained concentrating all data in one site) for ERS-1&2 SAR data:

ARCHIVED AT ESA CENTERS	
TOTAL (Terabytes)	347.66
OVERLAP (Terabytes)	43.50
OOC REDUCTION FACTOR (potential)	12.51%

Figure 1: ERS 1/2 SAR archive status

The following tables instead provide a present and future scenario of Third-Party missions data archived at the ESA's archiving centres:

Mission	LANDSAT	MODIS	JERS OPS	JERS SAR	SPOT	TOTAL
ARCHIVED	248.41	9.55	5.07	10.82	11.51	285.36
OVERLAP	24.70	1.26	0.12	0.17		26.26
OOC REDUCTION FACTOR (%)	9.94%	13.20%	2.39%	1.61%		9.20%

Figure 2: Third Party Missions archive (numbers are in TBytes)

Mission	NOAA AVHRR	SeaStar SeaWiFS (DLT)	NIMBUS CZCS	IRS P3 MOS (DLT)	TOTAL
ARCHIVED	27.83	2.40	3.00	2.00	35.23
OVERLAP	4.42				4.42
OOC REDUCTION FACTOR (%)	15.89%				11.57%

Figure 3: Other Third Party Missions archive (numbers are in TBytes)

The operational plan for the HARM project is based on four logical phases, some of them overlapping:

1. loading of the data from loose media into the MMFI long-term archive element (AMS – Archive Management System) in the original data format, in order to avoid waiting for the SAFE format consolidation, minimize the processing applied and maximize the usage of existing hardware (media drivers and stackers) in order to reduce as much as possible the operator's intervention
2. conversion of the datasets into the new SAFE format, exploiting the usage of the robotized library (fully automated operations). This phase can be contemporary and concurrent with phase 1, even if conversion tasks are retrieving and putting data in the centralized storage (example: the site could convert a mission while loading another one from the original media)
3. concentration of homogeneous data (same mission, sensor and mode) into specific locations, in SAFE format
4. orbits stitching at the target archives of the datasets already fully converted in SAFE format, with regeneration of browse and metadata for the ESA's central catalogue.

The high level of automatism in phases 2, 3 and 4 is possible thanks to the AMS robotics and the project architecture, based on an SQL database, recording and centralizing the AMS catalogue content. Distinct processing queues are automatically populated, according to the insertion of new records into the HARM catalogue table, by stored procedures or by the operators using simple HMIs.

A Monitor&Control can activate/monitor/stop the processing jobs (conversion, stitching or browse & metadata generation) that can run as daemons, concurrently and unattended, until orders are available.

3. SAFE Main Characteristics

The format developed by the HARM project has been named SAFE (Standard Archive Format for Europe) and uses the latest available technologies to achieve its goals of preserving the archived data for a long-term, facilitating the conversion into different formats, simplifying the extraction from the archive and enhancing their utilization by end-users and/or processing systems.

During the development of SAFE, particular attention has of course been put to the long-term preservation aspect. To this end, the information model of the generic Archival Information Package (AIP), introduced in the ISO 14721:2003 OAIS (Open Archival Information System) Reference Model [1] [2], has been used.

Furthermore, SAFE is based on the XFDU (XML Formatted Data Units) standard under development by the CCSDS (Consultative Committee for Space Data Systems). In its essence, SAFE is a profile of XFDU, and it restricts the XFDU specifications for the specific utilization in the EO domain.

Relationship with XFDU

It's worth repeating here the main rationale for the emergence of the "XML Formatting Data Units" – XFDU (credit CCSDS):

The current CCSDS Standards for Data Packaging have not undergone a major revision in 15 years. The computing environment and the understanding of metadata have changed radically:

- *Physical media → Electronic Transfer*
    - *The primary form of access to, and delivery of, both archived and recently produced data products has shifted from hard media to include substantial network delivery*
- *No standard language for metadata → XML*
    - *After 'bits' and 'ASCII', the language 'XML' can be viewed as the next universal data standard, as it has grown exponentially*
- *Homogeneous Remote Procedure Call → CORBA, SOAP*

- *Communicating heterogeneous systems are increasingly using standard remote procedure calls or messaging protocols. The primary RPC and messaging protocol for the WWW is SOAP, an XML based protocol*

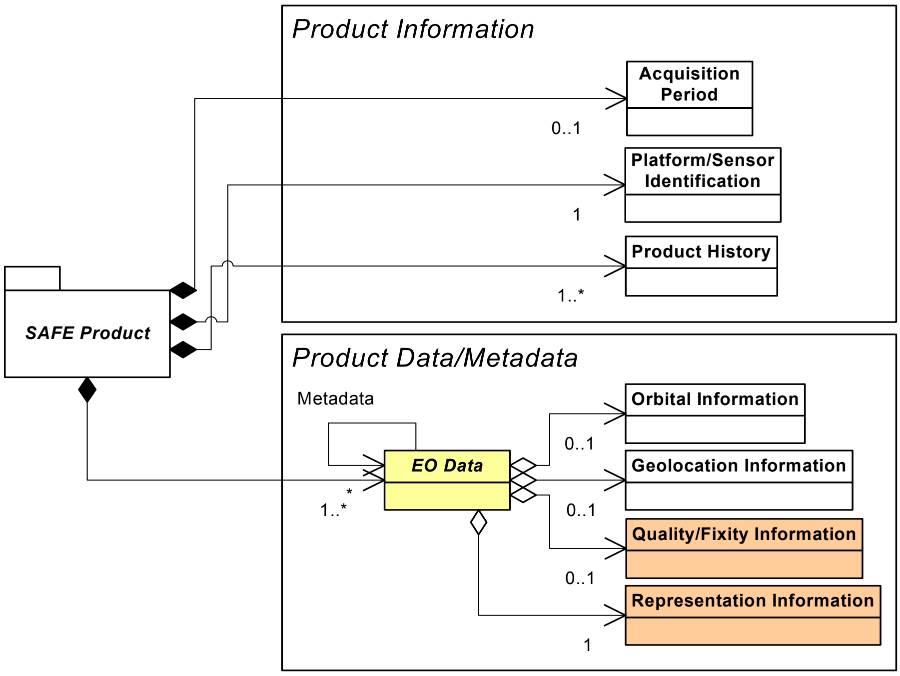
- *Little understanding of long-term preservation → OAIS RM*
  - *The OAIS Reference Model has become a widely adopted starting point for standardization addressing the preservation of digital information. The OAIS defines and situates within functional and conceptual frameworks the concepts of Information Packages for archiving (Archival Information Packages, or AIPs), producer submission to an archive (Submission Information Packages, or SIPs), and archives dissemination to consumers (Dissemination Information Packages, or DIPs).*
- *Record formats → Self describing data formats*
  - *Commensurate with XML, and rapidly growing computing power and storage capabilities has been an increasing tendency to use data formats that are more self-describing.*

Further, there are a number of new requirements that are needed in the Space domain to facilitate such functions as being able to describe multiple encodings of a data object, and to better describe the relationships among a set of data objects. Therefore it is necessary to define a new set of packaging standards while maintaining the existing functionality.

Although XFDU is still in the pipeline of the CCSDS development, the maturity of the current working draft suffices for the primary needs identified for SAFE. SAFE has, therefore, been designed to fully comply with the current definition of this emerging standard, by the way offering several advantages and opportunities, among which:

- inheriting from experiences gathered from the several international agencies, laboratories and companies composing the CCSDS
- sharing at least the same compliance level with the OAIS reference model
- making SAFE immediately compatible with the software due to be developed within or outside ESA scope
- facilitating interchanges between several archiving management systems maintained by agencies following the CCSDS recommendations
- supporting the development of XFDU by providing the CCSDS working groups with returns from experience of using and implementing the intermediate working drafts
- supporting the development of XFDU by providing software material implementing the standard.

SAFE Information Model



SAFE Information Model

A SAFE product, which can of course also be considered an "XFDU package", wraps or references EO data and associate them with information expressed in EO vocabulary. The primary objective of SAFE is to hold the so called

“Level-0” (or “L0”) data which is close to the telemetry level but it has, moreover, been qualified for the packaging of higher levels products, e.g. the ENVISAT GOMOS Level 1 and 2 products have already been successfully implemented in SAFE. All SAFE products contain the following metadata:

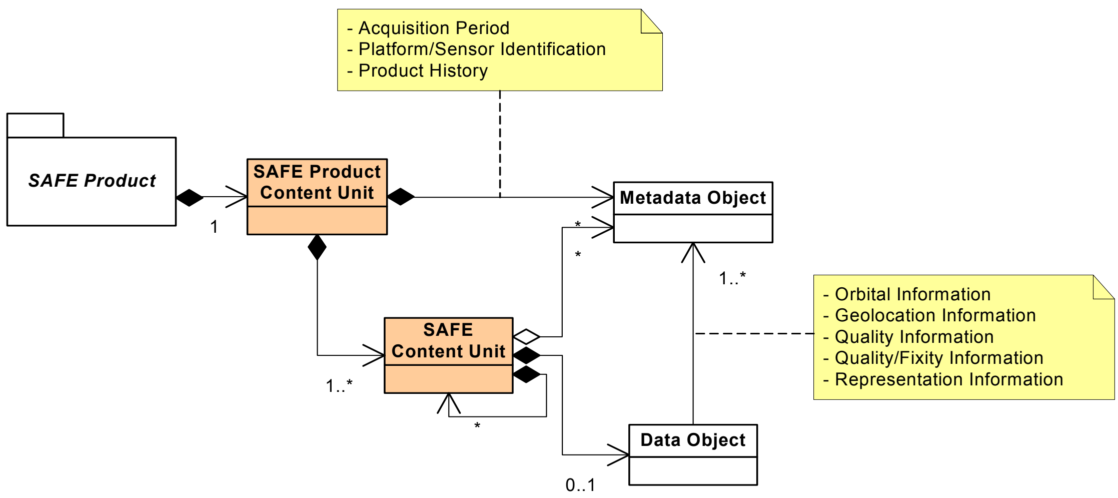
- **Acquisition period**: the acquisition period metadata that provide the time extents of all the data contained in the SAFE product. It is mainly dedicated to allow a fast time ordering and framing of the overall contained data.
- **Platform/Sensor identification**: the platform identifies the system (satellite/aircraft) that acquired the EO data wrapped by the SAFE product. It has sub-elements that unequivocally identify the platform as well as the specific sensor that acquired the data.
- **Product History**: a processing log collecting the historical information dedicated to the maintenance and the traceability of the product. The main feature of this logging system is its capability to store several processing threads regarding all the components that affected the product. As an example, if the product originates from the concatenation of several data objects, all logs of the involved objects will be kept, identifying their precise role of each in the production of the described SAFE product.

For each wrapped or referenced EO dataset, a collection of metadata information may be attached:

- **Orbital information**: the reference to the trajectory of the platform that acquired the data. This information may locate one or several orbit paths, the corresponding cycle, track, etc.
- **Geolocation information**: the information locating the product footprint on the Earth’s surface, either as a series of tie points or by reference to a world reference system of the acquiring platform. The Geolocation information, may also attach additional information to each localized element, including cloud coverage vote notation, meteorological information, etc.
- **Quality/Fixity information**: information about the quality of an EO dataset. SAFE makes use of techniques (i.e. XPath) that allows the precise location of the corrupted or missing elements up to the bit level.
- **Representation information**: any data contained in a SAFE product shall be accompanied with its representation information formally and numerically exploitable. Although the semantic information is partially implemented (i.e. all elements composing the dataset are named but the semantical links to standardized vocabulary is not embedded in the product), the SAFE aims at complying with the OAIS reference model in that area for assuring the maximal theoretical long-term preservation.

Finally, SAFE does not limit the information to the content listed above but supports extensions as far as they preserve the integrity of the mandatory items.

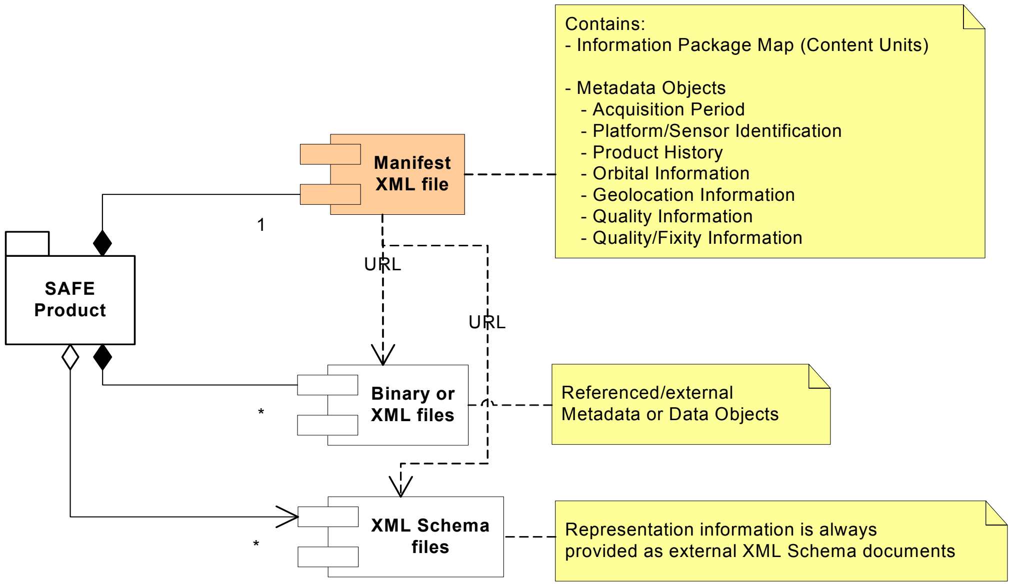
SAFE Logical Model



SAFE Logical Model

As specified in the XFDU model, a SAFE product is a logical tree of “Content Units” [2] forming the so-called “Information Package Map”. Conversely to XFDU, only one map is expected per SAFE product. The root Content Unit has predefined associations to the information applicable to the overall product, i.e. at least the “Acquisition Period”, the “Platform/Sensor Identification” and the “Product History”. The structure of the children Content Units is less constrained and depends mainly of the logical view of the wrapped data. In most cases, one Content Unit matches one EO dataset and its accompanying metadata. Several Content Units may, however, share the same metadata.

SAFE Physical Model



SAFE Physical Model

A SAFE product physically contains the following components:

- **Manifest file**: an XML document conforming the XFDU Manifest file. It contains the definition of the Information Package Map, the wrapped Metadata Objects (i.e. in general all Metadata Objects are embedded in the SAFE Manifest file), the wrapped Data Objects (i.e. Data Objects are rarely embedded in the SAFE Manifest file) and references to the external files containing the Metadata and Data Objects[1].
- **Binary or XML files**: the data or metadata object contents. Currently, only two types of files have been identified, i.e. binary matching MIME `octetstream` definition and XML documents. Each of these files shall be accompanied with one or more XML Schema document controlling its content.
- **XML Schema files**: the representation information of the data held by a SAFE product. In comparison to XFDU, SAFE does not allow multiple notations for storing the representation information of its objects. This restriction is mainly imposed because SAFE does not only reference a representation information technique but intends to define it. In order to represent the binary information, SAFE also defines specific markups that annotate the XML Schema documents to provide information on the physical structure, i.e. the so-called SDF markups. Thanks to these specific annotations, the contents of the binary files are described up to the bit level with a common technique as for XML documents.

SAFE Specializations

The SAFE specifications provide abstract definitions and ruling for handling a product. Similarly to XFDU, SAFE offers a framework that assures the consistency between all products but may not suffice for the complete definition of a product. Actually, EO products are generally accompanied with metadata specific to the mission or the sensor that acquired the referenced data. Moreover, a specific product may control more precisely the content of the generic information such as the Information Package Map structure, the platform name, etc.

The specifications of SAFE have therefore been broken down in two main layers:

- the Core level that controls the relationship with XFDU and the general structure that shall be followed by all SAFE products;
- and the Specialization level that implements the Core level up to obtain a complete and accurate definition of a specific product type.

SAFE Format Control Volumes

The specification of the SAFE format for the ESA handled datasets has been broken down in several control volumes:

[1] The current definition of SAFE does not support the XFDU "Behaviour Objects". Studies for their support are already on-going, for example to host translators to/from original format.

Identification	Volume	Title
PGSI-GSEG-EOPG-FS-05-0001	Volume 1	Core Specifications
PGSI-GSEG-EOPG-FS-05-0002	Volume 2	Recommendation for specializations
PGSI-GSEG-EOPG-FS-05-0003	Volume 3 Book 1	ENVISAT ASAR Products
PGSI-GSEG-EOPG-FS-05-0004	Volume 3 Book 2	ENVISAT MERIS Products
PGSI-GSEG-EOPG-FS-05-0005	Volume 3 Book 3	ENVISAT AATSR Products
PGSI-GSEG-EOPG-FS-05-0006	Volume 3 Book 4	ENVISAT DORIS Products
PGSI-GSEG-EOPG-FS-05-0007	Volume 3 Book 5	ENVISAT GOMOS Products
PGSI-GSEG-EOPG-FS-05-0008	Volume 3 Book 6	ENVISAT MIPAS Products
PGSI-GSEG-EOPG-FS-05-0009	Volume 3 Book 7	ENVISAT RA2 Products
PGSI-GSEG-EOPG-FS-05-0010	Volume 3 Book 8	ENVISAT MWR Products
PGSI-GSEG-EOPG-FS-05-0011	Volume 3 Book 9	ENVISAT SCIAMACHY Products
PGSI-GSEG-EOPG-FS-05-0012	Volume 3 Book 10	ENVISAT Auxiliary Data Products
PGSI-GSEG-EOPG-FS-05-0013	Volume 4	ERS Products
PGSI-GSEG-EOPG-FS-05-0014	Volume 5	Landsat Products
PGSI-GSEG-EOPG-FS-05-0015	Volume 6	Terra/Aqua MODIS Products
PGSI-GSEG-EOPG-FS-05-0016	Volume 7	NOAA AVHRR Products
PGSI-GSEG-EOPG-FS-05-0017	Volume 8	SeaStar SeaWiFS Products
PGSI-GSEG-EOPG-FS-05-0018	Volume 9	Nimbus CZCS Products
PGSI-GSEG-EOPG-FS-05-0019	Volume 10	SPOT HRV(IR) Products
PGSI-GSEG-EOPG-FS-05-0020	Volume 11	JERS SAR/OPS Products
PGSI-GSEG-EOPG-FS-05-0021	Volume 12	MOS MESSR/VTIR Products
PGSI-GSEG-EOPG-FS-05-0022	Volume 13	IRS-P3 MOS Products

SAFE Format Control Volumes

where the Volume 1 – "Core Specifications" – controls all common definitions applicable to all SAFE products.
The Volume 2 – "Recommendation for specializations" – provides guidance for developing a specialization of SAFE. The other volumes are specializations tailored to data originating from a specific mission or instrument already acquired and archived by ESA. All format volumes are generated automatically from the XML Schema defining the specialization.

SAFE I/O Library

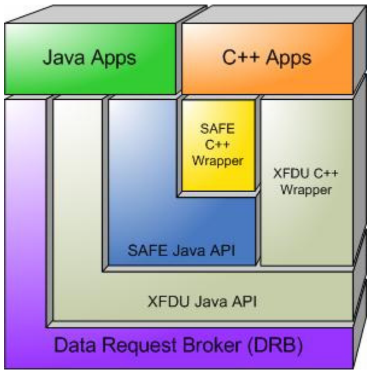
In addition to the SAFE specifications of the Core level and the Specialization level, the HARM project is implementing a set of software APIs dedicated to SAFE product users. The main use cases of these software components are:

- Create a SAFE product
- Open a SAFE product
- Remove a SAFE product
- Move a SAFE product considering or not the referenced objects
- Validate a SAFE product including or not the external object contents
- Add/Remove/Browse Content Units of the Information Package Map
- Add/Remove/Retrieve the Metadata Objects
- Add/Remove/Retrieve the Data Objects
- Browse/Select/Extract information from either Binary or XML objects
- Identify/Sort Quality Information to a specific part of the objects.

The top level APIs are provided in **Java** and **C++**, the C++ version being a wrapper that requires the Java components. They have been built on top of several other components of interest:

- **XFDU Java API**: this API provides the general features common to all XFDU packages. This API has been developed jointly with the SAFE Java API. This layer is presently used in the framework of the activity of the CCSDS Information Packaging and Registries (IPR) Working Group in support to the validation of the XFDU recommendation developed by this group. It may also support users that intend to develop applications compliant with all XFDU packages, whether complying SAFE or not.
- **XFDU C++ Wrapper**: a C++ wrapper on top of the XFDU Java API. Most of the functionality is preserved from the Java API, apart the capability of browsing the binary contents.
- **Data Request Broker (DRB)**: from which SAFE inherits the capability to access data independently from their formats. DRB supports in particular the SDF markup language used for annotating the XML Schema documents acting as representation information of the SAFE product objects.

The following figure represents schematically the architecture of the SAFE API suite.



Top-level architecture of the SAFE I/O library

ESA intends to freely distribute the SAFE API.

SAFE Translators as XFDU behaviours

As introduced earlier, the mechanism specified by XFDU as "Behaviour Object" is not yet allowed in SAFE products. Although Behaviour Objects are very attractive, their definitions are still under investigation within the CCSDS working group in charge of its development. They may be subject to important changes until the next versions of the XFDU working drafts.
Several experiences and study are however in route for hosting translators "to/from" original formats as "Behaviour Objects".

4. SAFE Benefits

The adoption of a common format like SAFE gives major benefits for a cost-effective long-term preservation and exploitation of these data for several reasons:

- SAFE is mainly devoted to the long-term preservation of data, as its full compliance with the CCSDS/ISO OAIS RM and XFDU standards reveal.
- SAFE permits an easier and more effective migration of data to other future standards.
- SAFE permits and easier reformatting to other formats, including end-user formats for products distribution.
- SAFE will easy the maintenance of the SW that use the data, even historical, thus decreasing the chances of obsolescence and improving their usability.
- SAFE being self describing, it greatly facilitates the format maintenance.
- SAFE format documents are built automatically from the XML-Schemas, so the maintenance of the SAFE format is more robust with respect to "old-fashion" formats.

SW that access SAFE-formatted datasets will benefit from the usage of the same SAFE XML Schemas for reading and writing the datasets. This greatly improves the overall cycle of creation/ingestion/quality control/transformation/distribution of the datasets.

5. Future prospects

The SAFE format provides a structure for long-term archiving of the ESA's historical and future datasets. The main operational activity presently carried out is the archiving of the datasets from the present media into the automated archive, in their original format. The HARM project will deliver the SW to perform the selection of the overlapping datasets, their conversion into SAFE and their stitching whenever an overlap reduction is required.
By doing so ESA expects to greatly reduce in the future the effort required to maintain its data holdings and improve the long-term preservation of the data.

The adoption of SAFE for the ESA's historical datasets in order to safeguard them from the possible loss is just the first step of the ESA's strategy for the long-term preservation of its EO missions' data archive. The future adoption of SAFE as archiving format since the design phase of the future missions will enhance even more the capability of ESA to preserve these important data, thus guaranteeing their accessibility and usability by the future generations.
With this perspective, ESA welcomes and will support activities outside of ESA for the adoption of SAFE as standard exchange archive data format and/or exchange format.

ESA is in course of establishing a SAFE dedicated web site where the documents, APIs and other information will be made available to external users. At the time of writing this paper, the SAFE web site is under construction and reachable at the following URL: http://earth.esa.int/SAFE/

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