Long-Term Data Preservation for ESA Earth Observation satellite data approach: an example for ENVISAT (A)ATSR data series

Sabrina Pinori *(Serco), Sergio Folco (Reha), Massimo Cardaci (Serco), Pauline Cocevar (Telespazio), Fay Done (Telespazio), Marco Cavicchioni (Serco), Mirko Albani (ESA-ESRIN)*

Introduction

- The aim of this work is to present the application of the ISO standard "*ISO 19165-2:2020 Geographic information — Preservation of digital data and metadata" for Earth Observation data* to a real case, and how this has been afforded

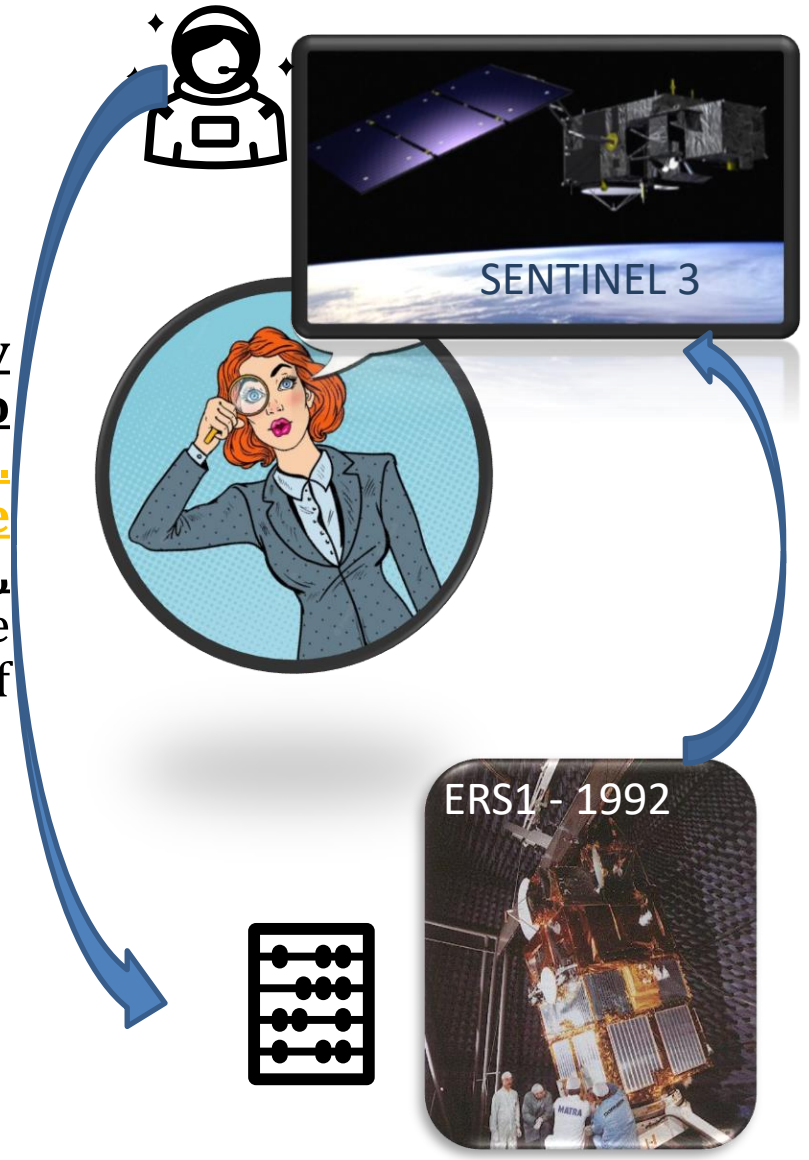
- ISO standard are sometime difficult to translate in practice, here is presented the steps approach used, the difficulties, and the future method learned from this activty.

- ESA Heritage missions, considering ESA historic mission, like ERS and Envisat but also Third-Party Mission (missions from other agencies or industries), are more than 50 missions (and even more instruments!).

Aim of this exercise

- Main ISO aim:
  - ✓ "For the longer term, **when the focus of the research community shifts toward new missions** and observations**, it is essential to preserve the previous mission data and associated information.** This **will enable a new user in the future to understand how the data were used (or generated, n.d.r.) for deriving information, knowledge and policy recommendations** and to "repeat the experiment" to ascertain the validity and possible limitations of conclusions reached in the past. …."
- and requirements:
  - ✓ Mission stages
  - ✓ Documentation for each stage,
  - ✓ But not only: data, processor, ancillary and auxiliary data!

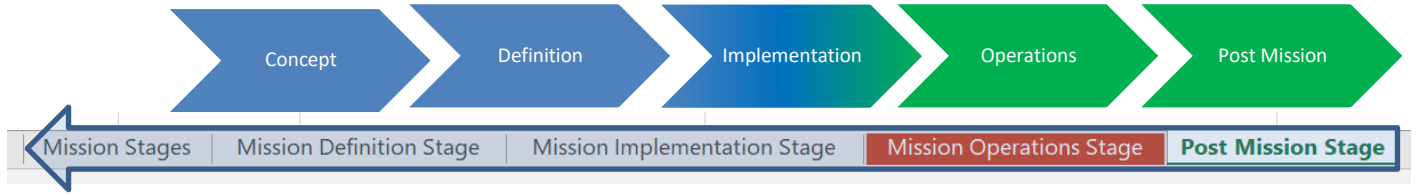
Stage
Mission Concept
Mission Definition
Mission Implementation
Mission Operation
Post Mission



How to translate them?

Stage
Mission Concept
Mission Definition
Mission Implementation
Mission Operation
Post Mission

ISO requirements



Working with Historical dataset, filfilling the ISO requirements, started from the last stage going backward.
For Third Party Missions it was not simple (or even impossible) to retrieve documentation on the first mission implementation and defintion stages.

Subset from Post Mission stage

Excel file translation

ID	Ty			Deliverables
PM_1.0	Do		eserved at this stage to satisfy content s specified below. Updated version of MO_1.0	
PM_1.1A	Doc	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. Processing	Description of final processing and/or calibration changes including provenance and context.	Description of final processing and/or calibration change including provenance and context. (metadata) Documentation with the metadata Product Format Specifications
PM_1.1B	Doc/ Data Records	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. Ancillary, Auxiliary	Updated Ancillary, Auxiliary data and their description (metadata) Name and location of the ancillary/auxiliary data archive facility if ancillary/auxiliary data will not be stored with the products	Ancillary, Auxiliary data records Name and location of the ancillary/auxiliary data archive facility if ancillary/auxiliary data will not be stored with th products with the metadata Documentation with the metadata Software
PM_1.1C	Doc/ Data Records	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. QA	Quality information updated as part of reprocessing	Quality information updated as part of reprocessing Documentation: reports (test, verification, validation, etc. with metadata Data records with metadata (Reprocessed Dataset) Test Data sets
PM_1.2	Data Records (Reproces sed data set)	Data consolidation & reprocessing strategy, implementation plans, and consolidated/ reprocessed data. L0, L1, L2+	Reprocessed data & products The final version of a derived product should be the version archived. If results reported in peer reviewed publications were based on earlier versions of the product, those versions or at least representative subsets of those versions should also be archived. At a minimum, the algorithm and software that generated such earlier versions should be archived.	Reprocessed data & products Documentation (EO-SIP speciications, ATBD) with metadata Data collection with metadata showing the provennace

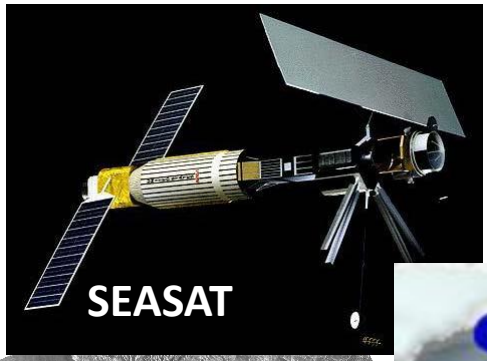
Subset from Missions Operations stage

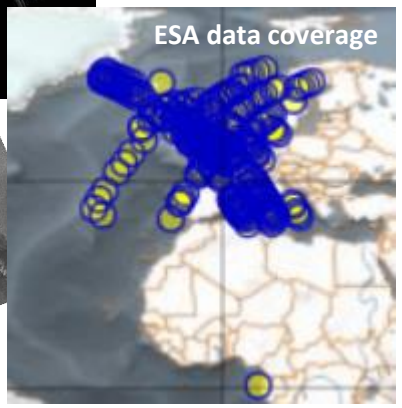
ID	Need For	Type	Identification	
MO_1.13	SAT, AIR	Doc/ SW Code	Data Processing Software	Instrument processing algori Description of lineage – i.e., input data and attributes covering all input data used by the algorithm - primary sensor data, ancillary data, forward models and look-up tables. Lineage information at granule level – i.e., all inputs (including ancillary or other data granules, calibration files, look-up tables, ground control, climatology etc.) used to generate the product. All information needed to verify what output data was created by a run, including data volume and file sizes. Documentation of expected exceptions, and how they are identified, trapped, and handled. Source for values of constants and look-up tables used in the algorithm, or explanation of how they were calculated. Documentation of processing history and production version history (which versions were used when, why different versions came about, and what the improvements and changes were from version to version). Descriptions of data sets used for software verification and validation. Test reports or summary of test results in sufficient detail to show that products meet requirements. Software release notes, including references to versions of operating systems, compilers, commercial or other software libraries used in the code. Description of potential future enhancements to the algorithm, the limitations they will mitigate. For all products held in the archive, the versions of source code used to produce the products. Where different versions of ancillary, input data, or calibration were used, the history of those changes should be available as part of the processing history.
MO_1.14	SAT	SW Code	Quality Control Software	Software used for quality assessment and developing quality indicators and/or quality flags.
MO_1.15A	SAT	SW Code	Science Data Tools	Product access (reader) and analysis tools. Source code to facilitate use of the calibration data, ancillary data and the data products at all levels. Source code useful for creating programs to read and display the calibration data, ancillary data and product data and metadata values. References to applicable dependency tools and libraries, and version numbers. Release notes, sample inputs and corresponding output results.
MO_1.15B	SAT	SW Code	Visualization Tools	Processing and visualizing tools

Implementation: first step

Single mission, single instrument **SEASAT → SAR in band-L**

the first ever civilian spaceborne imaging radar instrument (SAR) was flown on SeaSat in 1978 for 105 day (98 days of acquisition). Raw SAR data released in 2013 – **ESA L0 acquisitions consolidation presented in PV2013!**





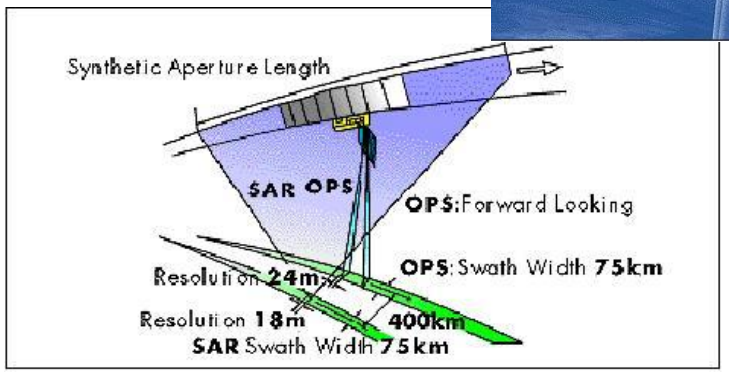
Mission Stages | Mission Definition Stage | Mission Implementation Stage | Mission Operations Stage | Post Mission Stage

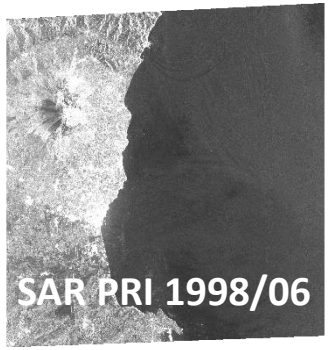
ID	Need For	Type	Identification	Description	Deliverables	Notes on deliverables	Common to the MIssion
MO_1.13	SAT, AIR	Doc/ SW Code	Data Processing Software	Instrument processing algorithms, context and source codes, testing context. Description of lineage – i.e., input data and attributes covering all input data used by the algorithm - primary sensor data, ancillary data, forward models and look-up tables. Lineage information at granule level – i.e., all inputs (including ancillary or other data granules, calibration files, look-up tables, ground control, climatology etc.) used to generate the product. All information needed to verify what output data was created by a run, including data volume and file sizes. Documentation of expected exceptions, and how they are identified, trapped, and handled. Source for values of constants and look-up tables used in the algorithm, or explanation of how they were calculated. Documentation of processing history and production version history (which versions were used when, why different versions came about, and what the improvements and changes were from version to version). Descriptions of data sets used for software verification and validation. Test reports or summary of test results in sufficient detail to show that products meet requirements. Software release notes, including references to versions of operating systems, compilers, commercial or other software libraries used in the code. Description of potential future enhancements to the algorithm, the limitations they will mitigate. For all products held in the archive, the versions of source code used to produce the products. Where different versions of ancillary, input data, or calibration were used, the history of those changes should be available as part of the processing history.	Processing baseline package: - Documents - Software Design Document -- Algorithm Specification, including input/output products and required auxiliary/ancillary data types -- Software Release Notes - IPF Technical Notes -- Test Plan, Test Report - Software binaries (mandatory where available, ideally virtualized and executable), source code and required configuration files Software test harness (including any files, data required for test)	binaries and source code for all versions corresponding to products held in the archive	SEASAT_IPF_V1.37.CentOS5.7.i386.install.tar.gz SEASAT_IPF_V1.37.CentOS6.4.x86.install.tar.gz TDS1.tar (for 1.37) SEASAT-SDD-v1.1 SEASAT-SRN-v1.37 SEASAT-IPF-SUM-v1.1.pdf SEASAT-IPF-STP-v1.0 (for ver 1.11) SEASAT IPF AR Report - SEASAT-V1.3.4-IDEAS+-SER-IPF-REP-2038 IDEAS+-SER-IPF-REP-2277_VerificationReport_Seasat1.37_1.1.doc

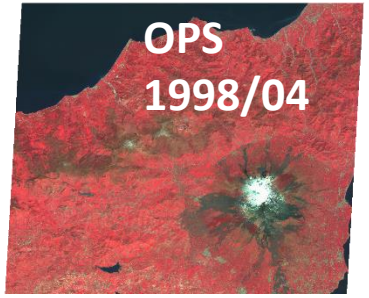
Implementation: second step

Single mission, multiple instruments **JERS-1 → SAR (band-L), OPS**

First Japanese imaging radar instrument (SAR) and optical radiometer (OPS) flown on JERS-1 from 1992 to 1998.







Each instrument has one dedicated column.
A cross-mission common column is used to share the digital object with the instruments and the platform.

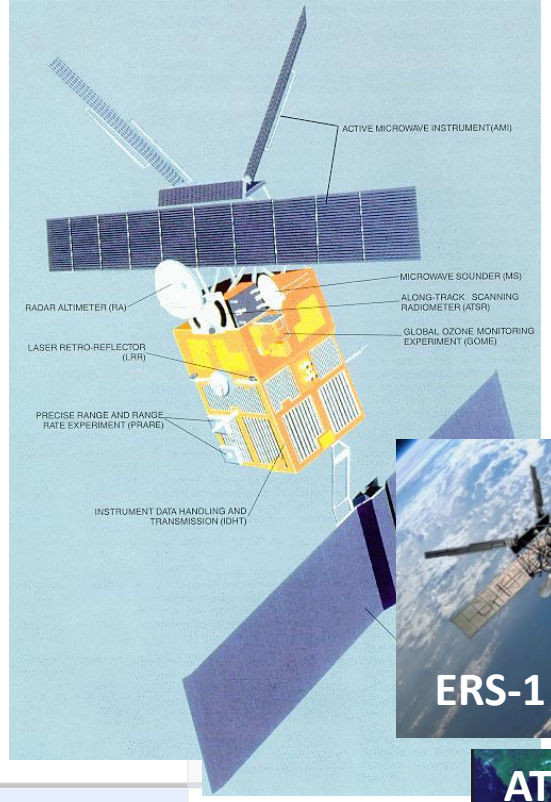
ID	Need For	Type	Description	Deliverables	Common to the Mission	JERS-SAR	JERS-OPS
PM_1.2	ALL	Data Records (Reprocessed data set)	Reprocessed data & products The final version of a derived product should be the version archived. If results reported in peer reviewed publications were based on earlier versions of the product, those versions or at least representative subsets of those versions should also be archived. At a minimum, the algorithm and software that generated such earlier versions should be archived.	Reprocessed data & products Documentation (EO-SIP speciications, ATBD) with metadata Data collection with metadata showing the provennace		SF: EO SIP - Scene Specification tailoring for JERS 1.8 JERS-1 SAR Level 1 Precision Image (JSA_PRI_1P) JERS-1 SAR Level 1 Single Look Complex Image (JSA_SLC_1P)	SF:JERS EO SIP-Quality Report Definition 11.0 JERS-1 Optical VNIR Level 1 System Corrected Products (OPS_SYC_1P)

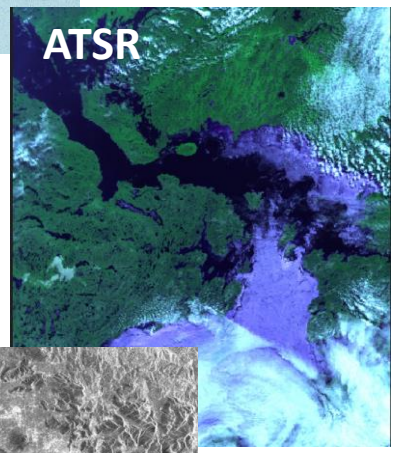
Implementation: third step

Multiple mission, multiple instruments
ERS-1/ERS-2 → SAR, ATSR, RA, MWR, SCATT (GOME)

ERS-1 and ERS-2 are the first two ESA EO satellites.
First launched in July 1991 and completed their tasks in 2010.

Each mission has a proper story and a proper documentation. The same document is replicated in each mission tailoring document.





ID	Identification	Description	Deliverables	SAR	SCATT	RA
MO_1.1	Data Processing Software	Instrument processing algorithms, context and source codes, testing context. Description of lineage – i.e., input data and attributes covering all input data used by the algorithm - primary sensor data, ancillary data, forward models and look-up tables. Lineage information at granule level – i.e., all inputs (including ancillary or other data granules, calibration files, look-up tables, ground control, climatology etc.) used to generate the product. All information needed to verify what output data was created by a run, including data volume and file sizes. Documentation of expected exceptions, and how they are identified, trapped, and handled. Source for values of constants and look-up tables used in the algorithm, or explanation of how they were calculated. Documentation of processing history and production version history (which versions were used when, why different versions came about, and what the improvements and changes were from version to version). Descriptions of data sets used for software verification and validation. Test reports or summary of test results in sufficient detail to show that products meet requirements. Software release notes, including references to versions of operating systems, compilers, commercial or other software libraries used in the code. Description of potential future enhancements to the algorithm, the limitations they will mitigate. For all products held in the archive, the versions of source code used to produce the products. Where different versions of ancillary, input data, or calibration were used, the history of those changes should be available as part of the processing history.	Processing baseline package: - Documents - Software Design Document -- Algorithm Specification, including input/output products and required auxiliary/ancillary data types -- Software Release Notes - IPF Technical Notes -- Test Plan, Test Report - Software binaries (mandatory where available, ideally virtualized and executable), source code and required configuration files (example the task table) Software test harness (including any files, data required for test)	MC: 6.01 - CAE-RN-52-7563_IDEAS-PF-ERS V6.01 S MC: 6.01 - CAE-RN-52-7564_IDEAS-PF-ERS V6.01 MC 6.00 - 524693_ATPP_v2_1.pdf MC 6.00 - 524826_ASAR_ERS_ICD_2_0_draft.doc MC 6.00 - 524827_ASAR_ADD_2_0_draft.doc MC 6.00 - 524827_v2_0.pdf MC 6.00 - PF_ERS V6.00_SRN.pdf MC 6.00 - PF_ERS_V6.00_SMR.pdf MC 6.00 - PF-ERS_V6.00_installation_guide.txt MC 6.00 - PF-ERS-ICD-20[CAE-1G52-4826].pdf MC: ersRT.v.600_to_v6.01_patch.zip MC: 6.01 - IDEAS+-SER-IPF-REP-2475_IDEAS_PF-E MC: 524826_PFASAR_ICD_1_1 MC: 6.01 - TDS21.zip MC: 6.01 - IDEAS+-SER-IPF-REP-2475_IDEAS_PF-E MC: OTF_Task tables MC: ewac_task_tables_corrected MC: 6.00 - DEAS-SER-IPF-REP-1191_PF-ERS-600_V	MC: ASPS_10_04_SRN_1_1.doc MC: ASPS_Def_TN_2_4.pdf MC: ASPS_IG_1_3.doc MC: ASPS_SUM_2_4.pdf MC: WS_ICD_3_2.doc MC: 10.04_patch.zip MC: ASPS-prod-10.04.tar MC: ASPS-AnomalyCorrectionReport MC: asps_cots.tar MC: ASPS_ddd_v10_04.tar MC: asps_aux_db.tar MC: ASPS_src_v10_04.tar MC: gribex_library.tar MC: IDEAS+-SER-IPF-REP-DDE_Technical MC: patch_installation_instructions MC: WS_ADD_1_1 MC: IDEAS+-SER-TOO-TSP-2470-Configur MC: tables_1001.zip MC: TDS2.zip MC: addmain-030519.pdf MC: dddmain-030521.pdf MC: icdmain-030521.pdf MC: ingestmain-040617.pdf MC: soamain-030521.pdf MC: tp2main-040715.pdf MC: tpmain-030520 .pdf MC: tpmain-cmod5.pdf MC: 9.doc (ASPS Product format) MC: tosca-soamain-061204.pdf MC: tosca-tpmain-061204.pdf MC: tosca_install_main-061204.pdf MC: user-provided-grid-product-110719.pdf MC: tosca-deliveries_20120413.zip MC: asps2netcdfconverter-531.zip MC: SCI-TNO-16-0033-v01_SCiRoCCo_AS	MC: REAPER_Delivery_Note_RP01.pdf MC: Reaper_RP01_delivery_check_12.1

Challenges

Baseline: define a concrete concept of interoperability baseline between SW, Docs, Products, AUX, etc.

No duplication: Manage effectively the situation of information common…

To several instruments of the same mission (e.g. ERS, ENVISAT, JERS, ALOS,…)

To multiple instances of the same mission (e.g. ERS / LANDSAT)

Simplicity for complexity: avoid an escalation of complexity in the structure

Mapping to stages: many items evolve over mission life.
Where to map the ones which change across phases ?

Flavours: some elements are instantiated differently depending on purpose.
How to manage this complexity ?

Filling former stages (before "QA4EO" visibility): still to be addressed.

Digging into historical documents, SW,… going back even 15 years or more

Main lessons learned

1

The Auxiliary files released in the Mission Operation Stage, and used also in the Post Mission Stage. **What is the ISO requirement to assign the Aux file?**

→

The Auxiliary files is assigned to the **mission phase in which they are released** (Mission Operation Stage). This is **applicable to all input items**.

2

If the data baseline is generated in the Post Mission, together with SW and documentation, and some documents were not expected in this phase, but in another one (for example in the Operation Phase), where to put them?

→

The document follows always the SW mission phase. If the document/SW has been generated in the Post Mission Phase it goes there and no in the Operation Phase.

3

The historical documentation was released in the past. The product baseline fits the **Post Mission Phase**. Where historical documents will fits?

→

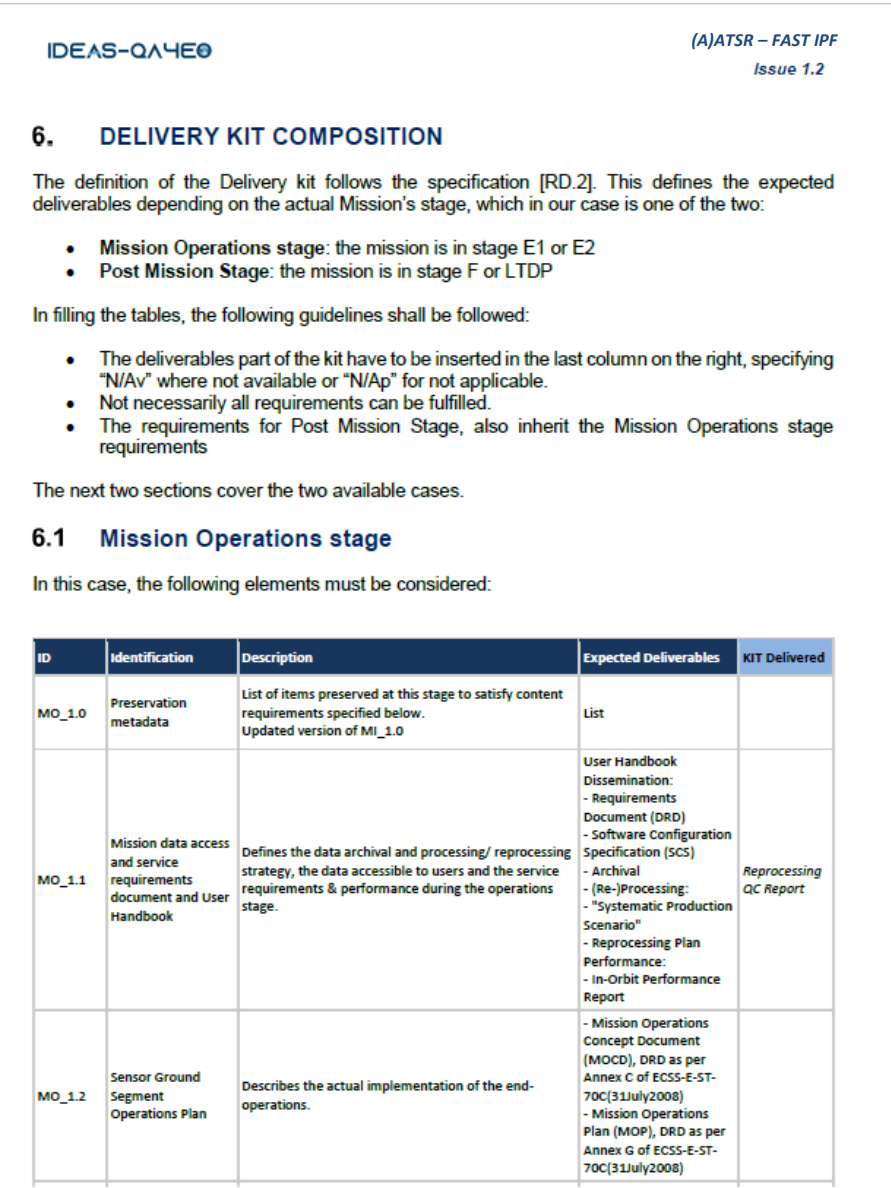
The Auxiliary files is assigned to the **mission phase in which they are released** (Mission Operation Stage).
EX: Space To Ground Interface ICD. The document describes the satellite raw data format that could be an input to the product baseline.

Conclusions

The work was not easy and it was a tryal and error exercise, several times the work implemented for a step was not feasible for the next step and we had to adjust the appoach.

New approach

ESA-GMQ section is in charge of processor development and data QC, the ISO requirements have been introduced in the verification approach



IDEAS-QA4EO

(A)ATSR – FAST IPF
Issue 1.2

6. DELIVERY KIT COMPOSITION

The definition of the Delivery kit follows the specification [RD.2]. This defines the expected deliverables depending on the actual Mission's stage, which in our case is one of the two:

- **Mission Operations stage**: the mission is in stage E1 or E2
- **Post Mission Stage**: the mission is in stage F or LTDP

In filling the tables, the following guidelines shall be followed:

- The deliverables part of the kit have to be inserted in the last column on the right, specifying "N/Av" where not available or "N/Ap" for not applicable.
- Not necessarily all requirements can be fulfilled.
- The requirements for Post Mission Stage, also inherit the Mission Operations stage requirements

The next two sections cover the two available cases.

6.1 Mission Operations stage

In this case, the following elements must be considered:

ID	Identification	Description	Expected Deliverables	KIT Delivered
MO_1.0	Preservation metadata	List of items preserved at this stage to satisfy content requirements specified below. Updated version of MI_1.0	List	
MO_1.1	Mission data access and service requirements document and User Handbook	Defines the data archival and processing/ reprocessing strategy, the data accessible to users and the service requirements & performance during the operations stage.	User Handbook Dissemination: - Requirements Document (DRD) - Software Configuration Specification (SCS) - Archival - (Re-)Processing: - "Systematic Production Scenario" - Reprocessing Plan Performance: - In-Orbit Performance Report	*Reprocessing QC Report*
MO_1.2	Sensor Ground Segment Operations Plan	Describes the actual implementation of the end-operations.	- Mission Operations Concept Document (MOCD), DRD as per Annex C of ECSS-E-ST-70C(31July2008) - Mission Operations Plan (MOP), DRD as per Annex G of ECSS-E-ST-70C(31July2008)	