IDOC-OPE Instructions for Ground Segments

IDOC-OD-005

Préparation

	Nom et Fonction	Date
Rédacteurs	Gilles Poulleau	Aout 2022
Vérificateur	Marian Douspis	
Approbateur	Prénom Nom, *fonction*	05/04/2016

Liste de diffusion

Nom	Fonction	Société

Evolutions

Edition	Date	Modifications
1.0	23/01/2016	1st draft in french
2.0	01/02/2017	translated in english
2.1	23/07/2022	Entirely rewritten

Table des matières

CONTEXT OF THIS DOCUMENT AND RECOMMENDATIONS FOR USE

This document is related to the "IDOC-OD-008 IDOC Instructions for new services" which needs to be handled first.

This document also describes, in addition to the management, strategy and implementation aspects within IDOC, the overall process for taking into account any new request.

1 REFERENCE DOCUMENTS

Acronym	Reference of the document	Document full name
RD1	IDOC-EX-001	IDOC-EX-001 IDOC executive summary
RD2	IDOC-OD-002	IDOC-OD-002 IDOC Risk analysis and management
RD3	IDOC-INS-003	IDOC-INS-003 IDOC Instructions applicable to project design
RD4	IDOC-INS-004	IDOC-INS-004 IDOC-DATA Instructions for Data Ingestion and Curation
RD5	IDOC-INS-005	IDOC-INS-005 IDOC-OPE Instructions for Ground Segments
RD6	IDOC-INS-006	IDOC-INS-006 IDOC-DATA Instructions for Data Preservation
RD7	IDOC-INS-007	IDOC-INS-007 IDOC-OPE Instructions for Instrument Operations
RD8	IDOC-INS-008	IDOC-INS-008 IDOC Instructions for Services
RD9	IDOC-INS-009	IDOC-INS-009 IDOC-DATA Instructions for Data Provision
RD10	IDOC-INF-010	IDOC-INF-010 IDOC Organigrammes
RD11	IDOC-DW-011	IDOC-DW-011 Diverses schemas for documentation
RD12	IDOC-INS-012	IDOC-INS-012 IDOC instructions for architecture and coding practices
RD16	IDOC-EX-016	IDOC-EX-016 OSUPS Schéma Stratégique Numérique
RD17	IDOC-OD-017	IDOC-OD-017 Services offerts par IDOC
RD30	IDOC-HO-030	IDOC-HO-030 Presentation IDOC-public-english
RD31	IDOC-HO-031	IDOC-HO-031 Presentation IDOC Français

2   PROCESS FOR INTEGRATING A NEW PIPELINE INTO THE INFRASTRUCTURE

A new pipeline is integrated into IDOC's infrastructure according to a set of rules that aim to ensure the best efficiency of the efforts deployed to build and maintain the pipeline.

These pipelines, which are part of the "ground segment" of an instrument, may involve a large number of organisations, data sources and interfaces.

It is therefore essential that the architecture of ground segments and in particular pipelines is flexible and respects the following implementation rules:

2.1   ORGANISATION

Allocation of tasks according to skills: allows higher levels of expertise to be achieved

Redundancy of skills and knowledge in the project: avoids building up too many unavoidable individual profiles

2.2   PIPELINE

Automate all operational actions in the pipeline

Plan for automation from the beginning of the development design

Minimise data transfers across the network.

Allow quick and easy access to the different sub-products and products generated by the different pipeline levels to allow their verification and validation.

Each stage of the pipeline can be restarted in isolation in the event of a problem or for enhancement purposes, without disrupting the existing data structure.

Any data enhancement or completion from any of the pipeline sources must be able to be taken into account by the pipeline and allow for updates to the delivered products

All software parts must handle operational exceptions in a predictable manner with the least possible disruption

Each software part must be able to monitor and evaluate its own data output. Checks must be made at each sub-step and alerts must be triggered which can be easily retrieved by an alarm management tool, sent by email, etc.

Similarly, each piece of software must be able to monitor and evaluate its own operation.

A dashboard must reflect the operational status of the pipeline and the information needed to understand the operation must be accessible from this dashboard.

Traces of the execution of the different modules of the pipeline should be kept

2.3 INFORMATION PACKAGES, INFORMATION OBJECTS

The information packages delivered (OAIS Submission Information Package SIP) may contain a variable set of information objects whose processing is the purpose of the pipeline; they may also include objects that are not relevant or that must be handled by another pipeline.

The pipeline will produce stored information packages (OAIS Archival Information Package AIP) which will contain information objects

The information packages must store groups of information objects that are consistent with each other.

The criteria for grouping objects in the packages must be uniform over the dataset

If there are multiple criteria for grouping objects, the smallest common factor should be kept, so as not to penalise the path according to one of these criteria.

The information packages produced must be as coherent as possible with the subsequent use of the dataset:

Allow an efficient and easy search in the dataset,

Allow access to the data in accordance with the tools that will be used (e.g. tiling of surfaces for rapid exploration)

The structure of the packages must allow easy insertion or modification of the objects contained.

The definition of these packages is a key step that conditions the accessibility and durability of the dataset produced.

2.4 GENERIC RULES OF CONCEPTION AND DESIGN

Always look for the simplest solution (KISS Keep It Simple Stupid)

Use the most common adapted standards

Always and above all, look for and study the answers already given to similar problems.

Reduce coupling and dependencies between software modules to the strict minimum.

Reduce temporal dependencies to the strict minimum

Develop the software components by favouring their parameterisation

Anticipate the redundant operation of each component of a ground segment and integrate this behaviour into the entire design: architecture,
 Procedures for continuous operation of the pipeline using redundancy are tested and described

Recovery procedures detail the actions to be taken for each incident. See document "writing a response form".
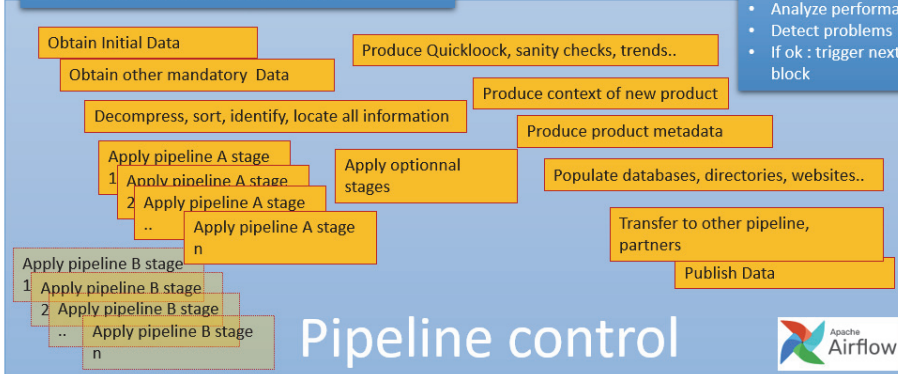
A pipeline consist in a combinaison of chained blocks
- Independent and autonomous
- The block and its followers can be rerun anytime
  (new version, new incoming packets, ..)

After each block :
- Check product
- Analyze performance
- Detect problems
- If ok : trigger next
  block

Obtain Initial Data

Obtain other mandatory Data

Produce Quickloock, sanity checks, trends..

Produce context of new product

Decompress, sort, identify, locate all information

Produce product metadata

Apply pipeline A stage 1

Apply pipeline A stage 2

Apply optionnal stages

Populate databases, directories, websites..

Apply pipeline A stage n

Apply pipeline B stage 1

Transfer to other pipeline, partners

Apply pipeline B stage 2

Publish Data

Apply pipeline B stage n

Pipeline control

Apache Airflow

3 SECURING DATA AND OTHER COMPONENTS

It is essential that all pipeline components ensure that they do not distort the quality of the data. Thus, the transfer or replication functions must include checksum mechanisms to guarantee the identity of the data before and after transfer.

The test data sets of the pipeline sub-components must also take this aspect into account in their design.

4 ANNEX

4.1 PROCEDURE TO PREPARE : NEW PIPELINE INTEGRATION

4.1.1 Input data
– Origin (location, responsibility,..)
– Format
– Volume, frequency, duration
– Location and means of access
– Validity check
– Information packages delivered (OAIS Submission Information Package SIP): content, structure

4.1.2 Output data
– Format
– Volume, frequency, duration
– Redistribution
– Clients
– Interfaces
– Validity check
– Stored information packages (OAIS Archival Information Package AIP): content, structure

4.1.3 Algorithms
– Existing
– To be developed
– Production of the corresponding code
– Consumption (cpu, memory)
– Production priority of products
– Execution status checkpoints
– Incident recovery points

4.1.4 Other elements
– Number of possible software versions
– Number of versions of data produced to be retained

5   REFERENCES USED FOR THIS DOCUMENT

Description of the IDOC infrastructure means and resources

Description of the tools used in the transfer, creation and backup of data sets. Characteristics of these tools in terms of control and security
Cf document risk analysis and IDOC responses
Description of the means of monitoring the software and hardware infrastructure
See document risk analysis and IDOC answers
Description of the documentation describing and operating the infrastructure
Cf document risk analysis and IDOC answers
Description of the software development monitoring tools
Cf document risk analysis and IDOC answers