IAS
Institut d'Astrophysique Spatiale
Orsay

IDOC Guidelines for Pipeline integration

IDOC-OD-005

Préparation

	Nom et Fonction	Date
Rédacteurs	Gilles Poulleau	Avril 2016
Vérificateur		
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	23/01/2020	Entirely rewritten

TABLE DES MATIERES

1 SCOPE OF THE DOCUMENT

This document is a guideline to IDOC staff in order to create a coherent data set for the control of the instrument, the validation of its operation and the data produced, and the formatting of this information in the appropriate format.
The sum of the different operations required for this construction is called a pipeline, but the analogy with a data refinery would probably be more telling.

This document is derived from the IDOC Guidelines Data production, integration and archive document that describes IDOC (Integrated Data & Operation Center, https://idoc.ias.u-psud.fr/), which combines the missions of a satellite operations center and a space data center.

This document also describes, in addition to the steering, strategy and implementation aspects within IDOC, the overall approach for taking into account any new demand.

2 APPLICABLE DOCUMENTS

	Référence	Titre
AD1	IDOC-LI-000	IDOC item list

• REFERENCE DOCUMENTS

	Référence	Titre
RD1	IDOC-EX-001	IDOC description executive summary
RD2	IDOC-OD-002	IDOC Risk analysis and management
RD3	IDOC-OD-003	IDOC General principles applicable to project design
RD4	IDOC-OD-004	IDOC Guidelines for Data integration
RD5	IDOC-OD-005	IDOC Guidelines for Pipeline Data Production
RD6	IDOC-OD-006	IDOC Guidelines for Archive long term preservation
RD7	IDOC-OD-007	IDOC Guidelines for Instrument operations
RD8	IDOC-OD-008	IDOC Guidelines for new services
RD9	IDOC-INF-009	IDOC General guidelines for Dataset dissemination
RD10	IDOC-INF-010	IDOC Organigramme
RD11		REGARDS – A generic CATALOG ACCESS SYSTEM AND data VALORIZATION tool

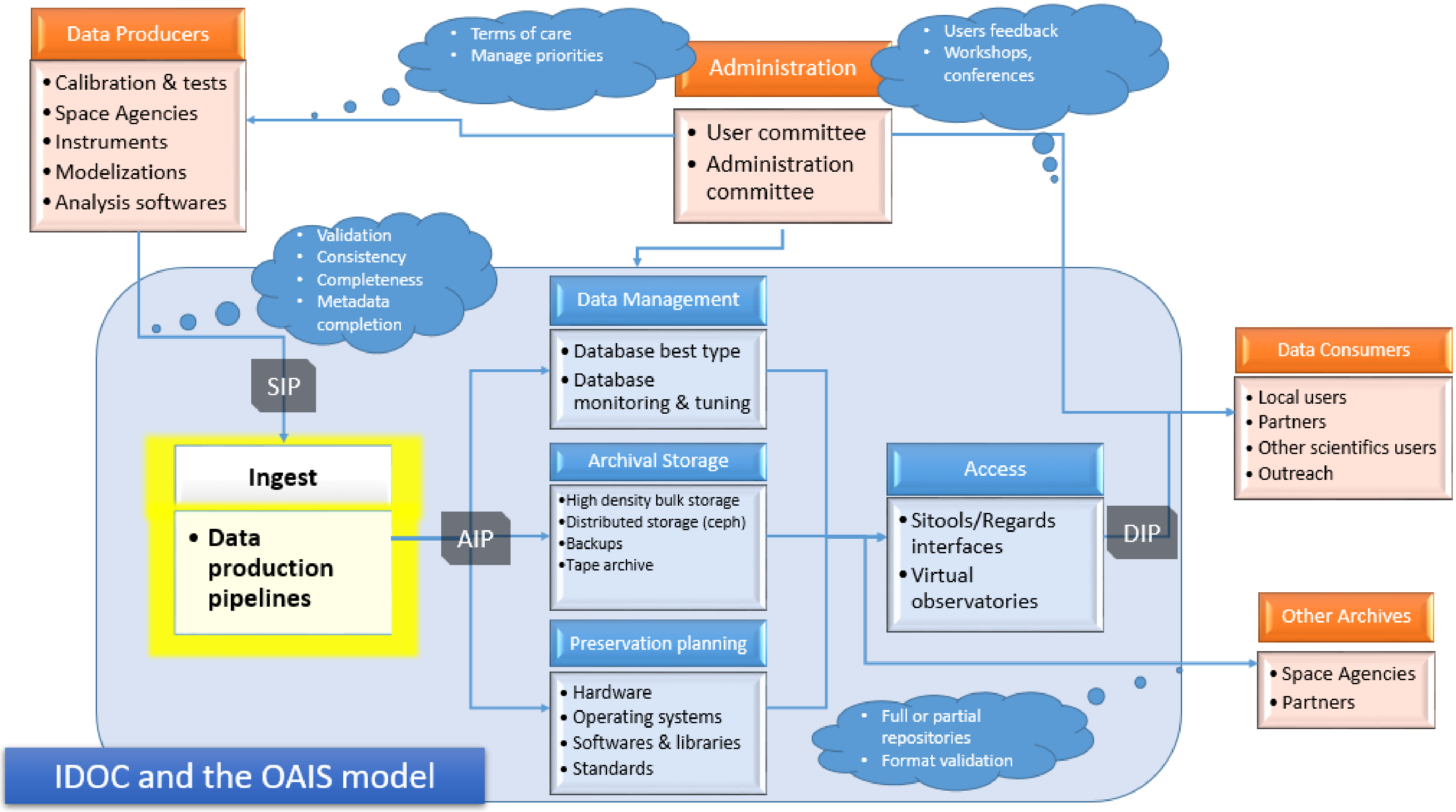
3 APPROACH TO INTEGRATING A NEW PIPELINE INTO THE INFRASTRUCTURE

A new pipeline is part of IDOC's infrastructure, following a set of rules that are designed to ensure the most efficient effort is made to build and maintain the pipeline.

These pipelines that are part of the "ground segment" of an instrument can involve a large number of organizations, data sources, interfaces.

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

This document describes the yellow underlined part of the IDOC application of the OAIS model in the figure below.



3.1 ORGANISATION

- Allocation of tasks according to competences: enables higher levels of expertise to be achieved
- Redundancy of skills and knowledge in the project: avoids building too inevitable profiles of individuals

3.2 PIPELINE

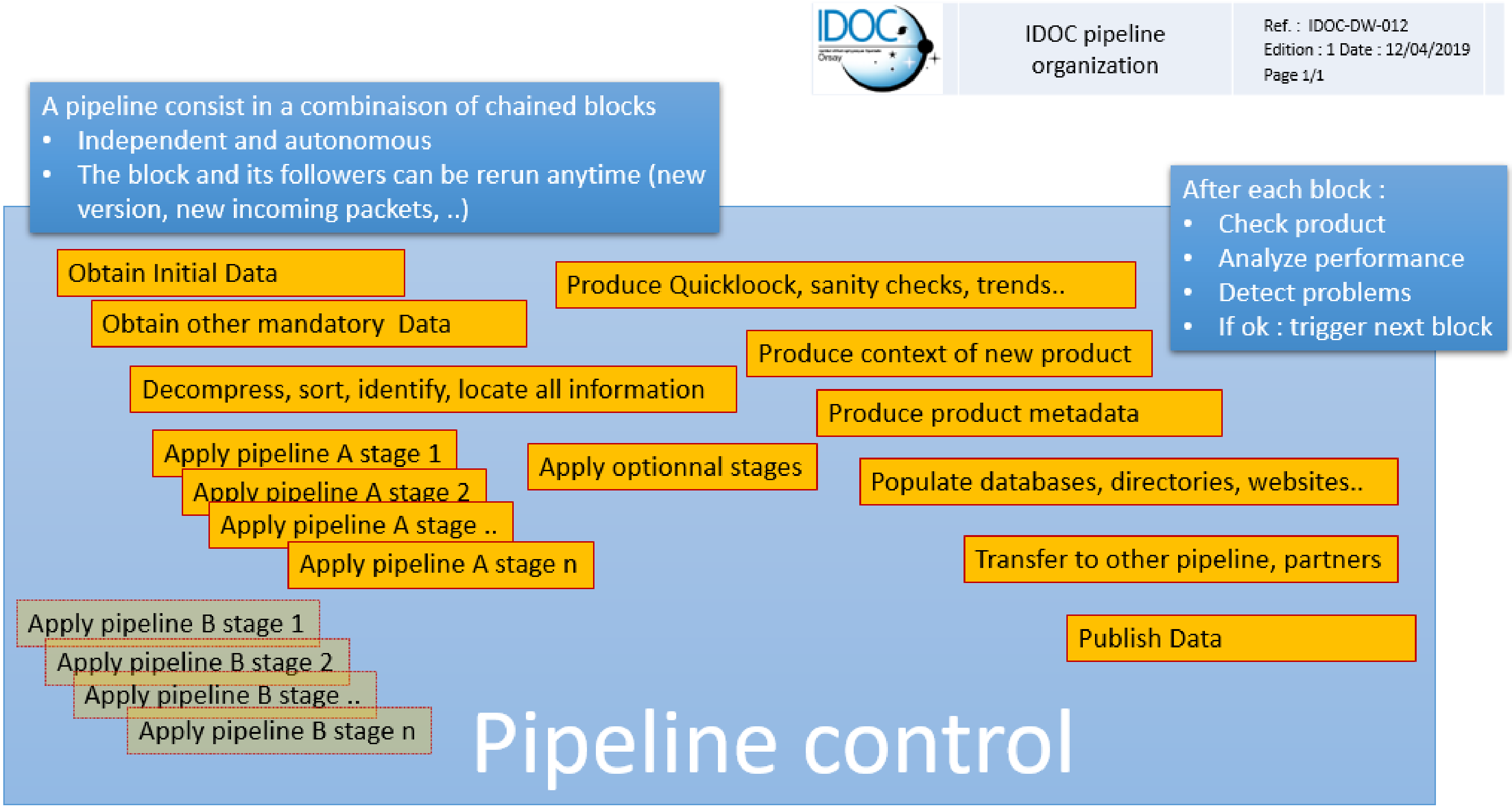
- Automate all pipeline operational actions
- Plan for automation from the very beginning of development design
- Minimize data transfers across the network.

- Allow quick and easy access to the various by-products and products generated by the different levels of the pipeline to allow for their verification and validation.
- Each stage of the pipeline can be restarted individually in case of problems or for improvement purposes, without disrupting the existing data structure.
- Any improvement or completion of data from one of the pipeline's constituent sources must be able to be taken into account by the pipeline and allow the products delivered to be updated.
- All software parts must handle operating exceptions in a predictable manner and with as little disruption as possible.
- Each piece of software must be able to control and evaluate its own data output. Checks must be carried out at each sub-step and alerts must be triggered that can be easily retrieved by an alarm management tool, sent by e-mail,...
- Likewise, each software part must be able to control 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.
- Records of the execution of the various pipeline modules must be kept.

3.3 Information packages, information objects

The OAIS Submission Information Package (SIP) may contain a variable set of information objects whose processing is the purpose of the pipeline; it may also include objects that are irrelevant or that need to be supported by another pipeline.
The pipeline will produce the OAIS Archival Information Package (AIP) which will contain the following information objects
- Information packages must store groups of information objects that are consistent with each other.
- The criteria for grouping objects in packages must be uniform across the dataset.
- If the criteria for grouping objects are multiple it is advisable to keep the smallest common factor, so as not to penalize the route according to one of these criteria.
- The information packages produced must be as consistent as possible with the use that will be made of the data set:
    - Enable an efficient and easy search in the dataset,
    - Allow access to the data in accordance with the tools to be used (e.g. surface tiling for rapid exploration).
- The structure of the packages must allow easy insertion or modification of the contained objects.
- The definition of these packages is a key step that conditions the accessibility and durability of the product dataset.

Generic design rules and design
- Always look for the simplest solution (KISS Keep It Simple Stupid)
- Use the most common adapted standards
- Always and above all, research and study the answers already given to similar problems.
- Reduce coupling and dependencies between software modules to the bare essentials.
- Reducing temporal dependencies to the strict essential.
- Develop software components by promoting their parametrization
- Plan the redundant operation of each component of a ground segment and integrate this behaviour in the entire design: architecture, modules, etc.
- The procedures for allowing continuous operation of the pipeline using redundancy are tested and described.
- The recovery procedures detail the actions to be taken for each incident. See the document "drafting a response sheet".



IDOC pipeline organization

Ref. : IDOC-DW-012
Edition : 1 Date : 12/04/2019
Page 1/1

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

Decompress, sort, identify, locate all information

Apply pipeline A stage 1

Apply pipeline A stage 2

Apply pipeline A stage ..

Apply pipeline A stage n

Apply pipeline B stage 1

Apply pipeline B stage 2

Apply pipeline B stage ..

Apply pipeline B stage n

Apply optionnal stages

Produce Quickloock, sanity checks, trends..

Produce context of new product

Produce product metadata

Populate databases, directories, websites..

Transfer to other pipeline, partners

Publish Data

Pipeline control

4 PROCEDURE TO PREPARE

The creation of a new pipeline service must be based on initial knowledge of the following elements that must be provided by data producers who must therefore provide responses to this questionnaire :

4.1 INPUT DATA

- Origin (location, responsibility,...)
- Format
- Volume, frequency, duration
- Location and means of access
- Validity check
- OAIS Submission Information Package (SIP): content, structure

4.2 OUTPUT DATA

- Format
- Volume, frequency, duration
- Redistribution
- Clients
- Interfaces
- Validity check
- Stored Information Packages (OAIS Archival Information Package AIP): content, structure

4.3 ALGORITHMS

- Existing
- To be developed
- Putting the corresponding code into production
- Consumption (cpu, memory)
- Product production priority
- Execution status checkpoints
- Incident recovery points

4.4 OTHER ITEMS

- Number of possible software versions
- Number of versions of data produced to be retained

5 ANNEXES

5.1 REFERENCES USED FOR THIS DOCUMENT

5.2 DESCRIPTION OF IDOC INFRASTRUCTURE MEANS AND RESOURCES

5.3 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 IDOC Risk Analysis and Responses document.

5.4 DESCRIPTION OF THE MEANS OF MONITORING THE SOFTWARE AND HARDWARE INFRASTRUCTURE

Cf IDOC Risk Analysis and Responses document.

5.5 DESCRIPTION OF THE DOCUMENTATION DESCRIBING AND OPERATING THE INFRASTRUCTURE

Cf IDOC Risk Analysis and Responses document.

5.6 DESCRIPTION OF TOOLS FOR MONITORING SOFTWARE DEVELOPMENT

Cf IDOC Risk Analysis and Responses document.