**Proposed Terminology to be used in Information Curation Process Standard**

\* = existing definition in CCSDS terms registry

**Data**: A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.\*

*Note: Better use the term “Information” in ICP (generic):* Any type of knowledge that can be exchanged. In an exchange, it is represented by data. An example is a string of bits (the data) accompanied by a description of how to interpret the string of bits as numbers representing temperature observations

measured in degrees Celsius (the Representation Information).

**As per discussion on 22/09/2014:**

**Preserved Data Set Content [[1]](#footnote-1)(PDSC):** All the information associated with a data record to be preserved for the long term. It consists of a consistent and complete set of data enabling current and possible future utilization of the Content Information and knowledge preservation. It consists of the Content Data and all associated Preservation Description Information, Packaging Information and Representation Information. It includes also Metadata for discovery and browse images when available.

**Contains only:**

**Data Records**[[2]](#footnote-2): A single, consistent, consolidated [[3]](#footnote-3)and validated set of data tailored to a particular project these include raw data, Level 0 data and higher-level products, browses, auxiliary and ancillary data, calibration and validation data sets, and metadata.

**Associated Knowledge**: mission, sensor, format, calibration documentation, software ....

Which contains also:

**Representation Information:** The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a **Data Object**) is mapped into a symbol.\*

3 following terms from August 23, 2011 by Butch Lazorchak

http://blogs.loc.gov/digitalpreservation/2011/08/digital-preservation-digital-curation-digital-stewardship-what%E2%80%99s-in-some-names/

**As per discussion on 22/09/2014:**

**Contains also:**

**Digital Curation[[4]](#footnote-5)**: The selection, organization, maintenance, collection, preservation or archiving and presentation of digital data or digital information typically using professional or expert knowledge.

Contains also:

**Digital Preservation[[5]](#footnote-6)**: The actions that maintain (including consolidate) digital data or digital information in its original or existing form, as well as actions that maintain the users’ ability to access and use it

Contains also:

Data consolidation: ….

**Information**: Any type of knowledge that can be exchanged. In an exchange, it is represented by data. An example is a string of bits (the data) accompanied by a description of how to interpret a string of bits as numbers representing temperature observations measured in degrees Celsius (the Representation Information).\*

**Metadata[[6]](#footnote-7)**: Data about other data\*

**Representation Information:** The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a **Data Object**) is mapped into a symbol.\*

*I suggest that we use the normal understanding and not provide a definition in this standard for the following term:*

**Adding Value[[7]](#footnote-8)**:

Note: CCSDS WG is not convinced by the terms “Data valorisation or optimisation”

**Documentation**:

**Knowledge[[8]](#footnote-9)**

**Processing Software:** all the software used in data product generation, quality control and data product visualization as well as any value adding tools.

# Discussion

**Generally speaking, when using the generic term with a specialized meaning, proposal to use the prefix “LTDP”, to avoid confusion in definitions**.

**LTDP preservation**

The LTDP activity, taken into the context of CCSDS, needs to be consistent with OAIS concepts and terminology. This activity should not try to redefine ‘preservation’, but rather map the LTDP concepts into OAIS and other related CCSDS standards concepts. For example, the LTDP ‘preservation’ involves:

- generation of a data set (*a part of a research or application endeavor outside OAIS archive*) that is consistent, consolidated and validated.

Generation of a data set is generally outside an OAIS archive. (check on exactly how we considered generation of updated data set from data in the archive. We certainly considered that generation as acting as a Producer, although it may still be within the OAIS Archive. I don’t remember.) I also agree that there are possibilities to record more information during generation that can help to ensure authenticity and to provide provenance.

“generation” is not part of the OAIS, but part of a process along the data lifecycle, isn’t it? YES Perhaps this should be integrated in the definition of a data lifecycle workflow? OK

An important thing to remember while discussing all this is that the focus of the OAIS Reference Model is what takes place within an OAIS Archive and at the interfaces to the OAIS Archives. Certainly however there are preservation concerns and affiliated actions that take place outside the OAIS Archives.

PAIMAS covers some of the actions and activities that take place around the Producer-Archive interface. Many of those actions take place outside the OAIS Archives (they are handled by the Producer) but those actions are still important for preservation.

- ensuring its long term integrity, discovery, accessibility and usability (*fully consistent with the functions of an OAIS archive*)

Discovery, accessibility and usability issues are also issues that might be addressed before reaching an OAIS Archives.

- focused on an individual mission/sensor or multi-mission data set and tailored to specific preservation/curation requirements.(*a mission or community objective, driven by research or application needs, decided outside of an OAIS archival function*)

The identification of these requirements is a first step in the workflow, isn’t it? YES (and not part of the definition)

Should curation requirements be considered as part of the LTDP preservation process? (The curation requirements should be considered part of the LTDP Preservation Workflow in the Initialization Phase) Seem wrong if LTDP is trying to separate preservation and curation concerns.

Agreement that it is useful to try to identify these concerns as early as possible. But they also change over time especially as technology changes and knowledge is generated.

Development of the original requirements is generally begun outside the OAIS Archives. But our metrics require that we have this time of information for information held by an OAIS Archives.

- consists of all activities needed to ensure data set bit integrity over time and to optimize (in terms of format and coverage) its (re)use in the long term (e.g. through metadata and catalogue improvement (*this much can be fully consistent with OAIS internal activities in terms of maintaining information understandability and usability over the long term. Updating formats may be needed to maintain understandability and usability. Adding to coverage is decided outside the OAIS, but acceptable to the OAIS in creating a new edition of the data set. Metadata and OAIS catalogue improvements are within the OAIS prevue to the extent they are initiated to maintain information understandability and usability).*

In the OAIS: metadata and catalogue improvements: are extracted from the SIPs via the Ingest function, and are populated via the Data Management function.

- algorithms evolutions and related (re)processing, linking and improvement of context/provenance information (*these are decided and developed outside the OAIS framework and would result in new submissions of information to the OAIS either for updates to an existing data set or for a new data set*)

Agreement that algorithms evolution and reprocessing are outside the OAIS Archives ; see linking and improvement of context/provenance sometimes happening within the OAIS Archives. Especially when trying to maintain authenticity and understanding over the long-term.

WG: this LTDP definition aims at giving on one hand some requirements, and on the other hand activities of a datalife cycle workflow for data creation or transformation in order than they could be used on the long term.

NOTE: The OAIS standard defines a conceptual entity. A given organization may incorporate an OAIS conforming function as a part of its broader set of activities.

**LTDP:**

**Curation:** aims at establishing and increasing the value of ***“EO Missions/Sensors Datasets”*** over their lifecycle, at favouring their exploitation through the combination with other Datasets and at extending their user base. It includes the activities for the definition of the preservation objectives, for the coordination and management of Data Time Series and Collections (e.g. from similar sensor family) in support to specific applications. It includes international cooperation activities.

JGG: I note that curation defines preservation requirements (but not curation requirements) The curation requirements should be defined by the LTDP Stewardship (after this discussion). However preservation which precedes curation is tailored to the preservation and curation requirements.

**Analysis:**

Neither OAIS nor CCSDS has a definition of ‘Curation’. Curation: derivative of Curate

Curate: 1. Select, organize and look after the items in (a collection or exhibition)

* 1. Select the performers or performances that will feature in (an arts event or program)
  2. Select, organize, and present (online content, merchandise, information, etc.), typically using professional or expert knowledge

*(from Oxford Dictionaries)*

Digital curation is the selection, preservation, maintenance, collection and archiving of digital assets. Digital curation establishes, maintains and adds value to repositories of digital data for present and future use. This is often accomplished by archivists, librarians, scientists, historians, and scholars. Enterprises are starting to utilize digital curation to improve the quality of information and data within their operational and strategic processes. Successful digital curation will mitigate digital obsolescence, keeping the information accessible to users indefinitely.

*(from wikipedia)*

WG: notes that generation of items is not included in this definition. Does creation/generation of new data (or new formats for the data) from existing data fall within the realm of curation? The creation/generation of new data (or new formats for the data) is a requirement defined during the initial phase of the Preservation Workflow. It depends on the initial analysis which involve also the cost, budget and resources availability.

The UK Digital Curation Centre has defined ‘Digital Curation’ for its own use (see Wikipedia). However the steps outlined in the Wikipedia article are basically what an OAIS does.)

The activities under the LTDP ‘curation’ appear focused on adding value by combining data sets and on the promotion of their use. This would be outside of an OAIS function. It also addresses a definition of preservation objectives (not clear on the scope, but probably outside OAIS) and the management of data time series and collections (which are within the OAIS framework) in support of specific applications (decided outside OAIS). It includes international cooperation activities which are probably (if broader than maintaining understandability and usability) outside the OAIS.

1. One or more **Archival Information Packages** (**AIPs**): an AIP is an **Information Package**, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS. *(from OAIS column of European LTDP Glossary and Technical Definitions)* [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. tailored for each Mission/Sensor according to its specific preservation/curation requirements, and consists of all the activities needed for Data Record Analysis, Cleaning, Gap Filling, Pre-processing, Processing/Reprocessing and Ingestion. A “Consolidated Data Record” is the basic input for any further higher level (re)processing [↑](#footnote-ref-3)
4. LTDP Curation: Aims at establishing and increasing the value of “EO Missions/Sensors Datasets” over their lifecycle, at favouring their exploitation through the combination with other Datasets and at extending their user base. It includes the activities for the definition of the preservation objectives, for the coordination and management of Data Time Series and Collections (e.g. from similar sensor family) in support to specific applications. It includes international cooperation activities. [↑](#footnote-ref-5)
5. LTDP: Aims at the generation of a single, consistent, consolidated and validated “EO Missions/Sensors Dataset” and at ensuring its long term integrity, discovery, accessibility and usability. It is focused on an individual Mission/Sensor or on a multi-mission Dataset (when one Master Dataset is made up of data coming from different missions/sensors) and tailored according to its specific preservation/curation requirements. It consists of all activities needed to ensure “EO Missions/Sensors Dataset” bit integrity over time and to optimize (in terms of format and coverage) its (re)use in the long term (e.g. through metadata and catalogue improvement, algorithms evolutions and related (re)processing, linking and improvement of context/provenance information).

   CCSDS **Long Term Preservation**: The act of maintaining information, Independently

   Understandable by a Designated Community, and with evidence supporting its Authenticity,

   over the Long Term. [↑](#footnote-ref-6)
6. LTDP “Metadata” corresponds to CCSDS “Descriptive Information” (the term “Metadata” is defined but not used in the OAIS, considered as too wide) [↑](#footnote-ref-7)
7. LTDP : aims at establishing and increasing the value of one or more *“Preserved Datasets”* over their lifecycles, at favouring their exploitation also through the combination with other Data Records and at extending their user base. [↑](#footnote-ref-8)
8. LTDP : needed to make the Data Records understandable to the Designated Community. This includes among others mission architecture, products specifications, instruments characteristics, algorithms description, Cal/Val procedures, mission/instruments performances reports, quality related information, etc. It is necessary to ensure data understandability and usability

   **CCSDS Knowledge Base**: A set of information, incorporated by a person or system, that allows that

   person or system to understand received information [↑](#footnote-ref-9)