Information Lifecycle Framework (ILF) Standard Project

1 PURPOSE AND SCOPE STATUS

The purpose of this document is to provide a high-level overview of the plans for developing a Consultative Committee for Space Data Systems (CCSDS) recommendation and its associated ISO Standard. This document will be used to gather additional support and participation in the CCSDS effort. It will also provide one of the initial sources for the CCSDS recommendation text.

This recommendation will define a framework for the information lifecycle, from the proposal to the disposition of the resulting information objects. It will describe the stages of the information lifecycle and the objectives, high-level activities and typical deliverables of each stage. Within each stage, this recommendation will identify concerns that must be addressed in order to preserve and utilize information objects for the long-term. It will identify standards, best practices and software tools that could be applied to address these concerns. This first Recommendation will focus on the preservation activities that should be undertaken at each stage. Subsequent Recommendations will address additional aspects that should be addressed at each stage of the Information Lifecycle Framework. For example, future Recommendations or issues of this first Recommendation could address Data Management Plans, Risk Management issues, etc.

This framework considers curation and preservation not as a separate activity to be considered at the end of an information production project, but as a set of actions that must be conducted throughout the information lifecycle. A number of high-level steps and associated activities will be identified that support curation and preservation at each stage of the lifecycle. Other concerns, such as data management plans, costing, risk management, metadata management, data formats, policies and workflow, value-adding and service architectures, are addressed at a high-level especially where they impact the curation and preservation. It is expected that these others concerns are identified in this recommendation but full treatment of the issues raised will require additional, more focused standards.

While this process is originating in the space community, it is being designed in a generic way and should be applicable to any science domain and to the wider library and archival communities.

This Recommendation accomplishes the following:

– identifies the different stages in the data/information lifecycle ;

– defines the objective of each of these stages, a high-level set of concerns that should be addressed during each of these stages, some possible actions to address those concerns and a typical set of expected deliverables (e.g., administrative, technical, contractual) at the end of a phase;

– forms a general methodological framework, which should be able to be applied and reused in any information stewardship, curation or preservation context (this general framework should provide sufficient flexibility to be applied to individual user's situations);

- – forms a basis for the identification and/or development of additional standards and implementation guides including those that address particular concerns in more detail;

- – forms a basis for identification and/or development of a set of software tools that will assist the development, operation and checking of the different stages of the lifecycle.

2 MAIN INPUT REFERENCES

1   OAIS Reference Model   http://public.ccsds.org/
2   OAIS Auditing and Certification Standards   http://public.ccsds.org/
3   Long Term Preservation of Earth Observation Space Data, Preservation Workflow, CEOS/WGISS/DSIG/PW, December 2014 (under review)
4   CCSDS Glossary   http://www.sanaregistry.org/
5   LTDP EO Data Stewardship Definitions
6   Data Lifecycles
    - o Data Lifecycle Models and Concepts, September 2011, http://ceos.org/ourwork/workinggroups/wgiss/documents/
    - o http://blogs.loc.gov/digitalpreservation/2012/02/life-cycle-models-for-digital-stewardship/
    - o https://dl.dropboxusercontent.com/u/6959356/ICP/Many%20models.pptx
7   PAIMAS Standard   http://public.ccsds.org/
8   PAIS Standard   http://public.ccsds.org/

- • Several other input materials are available and will likely contribute to additional models and mapping of processes to models as the standard.

3 EXPECTED BENEFITS

- • A consistent approach to the capture of preservation related information for existing and future projects/missions during the entire project lifecycle
- • A checklist of concerns that should be considered at each step of the data lifecycle
- • Decreased costs for archiving due to capture of needed metadata at the most efficient time
- • Decreased costs for archiving due to more consistently documented structures and the capture of that documentation more efficiently
- • Increased quality of digital information due to a more consistent capture of preservation information
- • Opportunities to support easier sharing of and access to digital information and development of interoperability tools and techniques
- • More authentic and reliable digital information which can more easily be made available to a wider audience if desired
- • Identification of areas where additional standardization is needed and achievable

4 KEY TECHNICAL FEATURES

- Will be applicable to all space missions (and ideally to other scientific, library and archival domains)
- Will address needed action from early planning to disposition of data
- Will provide a checklist of concerns and actions at each stage of the lifecycle from early planning to disposition of data
- Will provide links to other existing standards that should be applied at various points
- Will propose needed standards

5 TERMINOLOGY

An important part of this standard will be coming to an agreement on terminology that will be acceptable to the current space, science, library and archival communities. The OAIS Reference Model provides a starting point. Inputs from the EU LTDP project glossary provides proposals for additional terms although a number of them will need to be adjusted to use current OAIS terminology.

At this time, it seems as if we need to get agreement on the definition of the following terms (or another equivalent term for the concept):

Digital

Stewardship – Curation – Preservation

- Preservation (does not include Consolidation)
- Curation (wider concept that includes Preservation and Consolidation)
- Stewardship (wider concept that includes Curation)


- Metadata

- Data Record
- Preserved Data Set Content (PDSC)

- Data Management Plan (Wikipedia): A **data management plan** or **DMP** is a formal document that outlines how you will handle your data both during your research, and after the project is completed. The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation, and analysis before the project begins. This ensures that data are well-managed in the present, and prepared for preservation in the future.

Note: In common usage there is confusion/overlap between many of these terms. To avoid confusion and to allow precision, we will need specific definitions that will apply within the recommendation. Hopefully these definitions will resonate with others and will gain wider usage. However we expect that existing domains will continue with their own usage, but we expect that they will be easily able to map their usages to the usage within the recommendation.

6   THE LIFECYCLE FRAMEWORK: THE MAIN STAGES, ACTIVITIES, AND DELIVERABLES
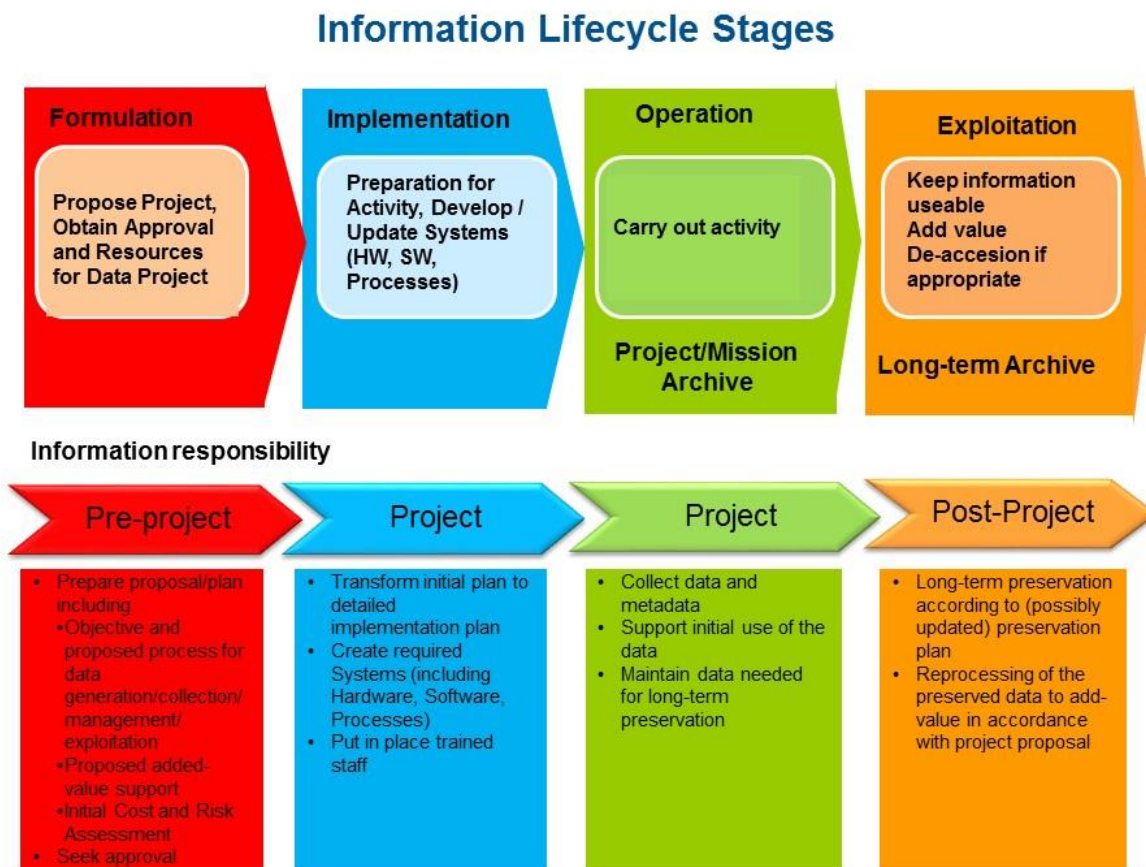
The figure identifies the stages of the information lifecycle.

1) The Formulation stage is responsible for proposing the project, which acquires or generates, manages or supports exploitation of data or information and then gaining approval and resources to carry out that activity.

2) The Implementation stage is responsible for preparing to carry out the activity. The project develops (or updates) systems (hardware, software and processes to meet the needs of the proposed activity.

3) The Operation stage is responsible for carrying out the activity to acquire or generate, manage, and support exploitation of data or information from this activity

4) The Exploitation stage is responsible for keeping the data or information useable and adding value if possible and proposed as part of activity. It also handles de-accessioning of the data if appropriate.

Within each of these stages, the high-level activities and typical deliverables will be identified to delimit the stages. Once this framework has been defined, we will be able to provide guidelines for activities that need to be performed during these stages to address a number of different data and information stewardship concerns. Within this first Recommendation we will focus on preservation. We intend to identify and provide guidelines for preservation concerns and activities that would ideally take place at each stage. Subsequent Recommendation will address additional aspects that should be addressed at each stage of the Information Lifecycle Framework. For example, future Recommendations or issues of this first Recommendation could address Data Management Plans, Risk Management issues, etc.

•

7 CONTEXT

There are a few similar and complementary efforts underway within the archiving community. However those other efforts are either restricted to limited communities or they do not lead to de-jure standards. It is our intent to partner with a couple of these efforts to move our project forward swiftly.

The first partnership is with the European Union Earth Observation community's Long-Term Data Preservation (LTDP) project. In fact their work is one of the primary inputs for our standardization project. They have reached out to us as well as to other organizations such as the Committee on Earth Observation Satellites (CEOS). We anticipate that some of their authors, members and reviewers will work with us on this effort. The differentiation of our effort from theirs will be that this effort will be aimed at a wider community than just the Earth Observation community. We anticipate that our lifecycle framework will begin earlier (early formulation stage) than is currently addressed in the LTDP efforts, and our effort will result in both CCSDS and ISO standards.

Another partnership that we intend to pursue is with the Research Data Alliance (RDA) Active Data Management Plans (ADMP) working group. Participants in our working group also participate or are involved in the management of the RDA working group. For the first phase of our project, we will focus on activities that capture and record important preservation metadata throughout the information lifecycle. ADMPs will likely be generating or capturing some this same metadata. It is anticipated that RDA ADMP project will also generate material that will be used in later phases of our efforts to extend the CCSDS and ISO standard to cover standardization of data management plans.

There are also a number of efforts that are underway at national space agencies. There may significant overlap in the interests and concerns of those projects with our efforts and we hope that those projects will allow participation from individuals to serve as gateways for information and idea exchange between these projects and our standardization effors.