# Introduction

## purpose and scope

There is a well-recognized need to capture digital information associated with a great variety of activities in virtually all recognized disciplines. More specifically there are many well defined projects conceived and executed for which it is essential that associated information be retained for the long term. However it is widely recognized that many such projects are not able, for one reason or another, to leave a sufficient legacy of information that others can reuse and fully leverage the effort that has gone into the project. The purpose of this Recommended Practice is to provide a view of a generic project, broken into stages, and to identify the types of digital information that needs to be captured and/or generated, and retained, at each stage. By clarifying this focus it is expected that projects can significantly improve their information legacy to the benefit of the wider community.

This Recommended Practice identifies a generic project at a high level and thus it should not be hard for most projects to map their activities to the generic stages. At each stage there is a minimum set of data that should be considered for capture. There is Additional Information associated with this data that also needs to be captured. The types of Additional Information are informed by the OAIS reference model that provides a conceptual view of long term information preservation in an archive.

This Recommended Practice does not cover all aspects of a project. Aspects of the activities it does specify do not have to be carried out strictly sequentially, and indeed some may be revisited and improved at several of the stages. Project aspects such as costing, risk management, policies and workflow, and service architectures are covered by the specific project plans. These are therefore not addressed here except at a high-level. Each aspect could be addressed in a variety of ways. It is expected that full treatment of these issues will require additional, more focused, standards.

## Context

This Recommended Practice accomplishes the following:

* divides a generic project into 4 stages and identifies the data and Additional Information to be collected or improved at each stage;
* forms a basis for the specifications of Data Management Plans
* 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 project.

This Recommended Practice fits into the overall context defined by a number of other standards. Some relationships between the documents are illustrated in Figure 1-1.

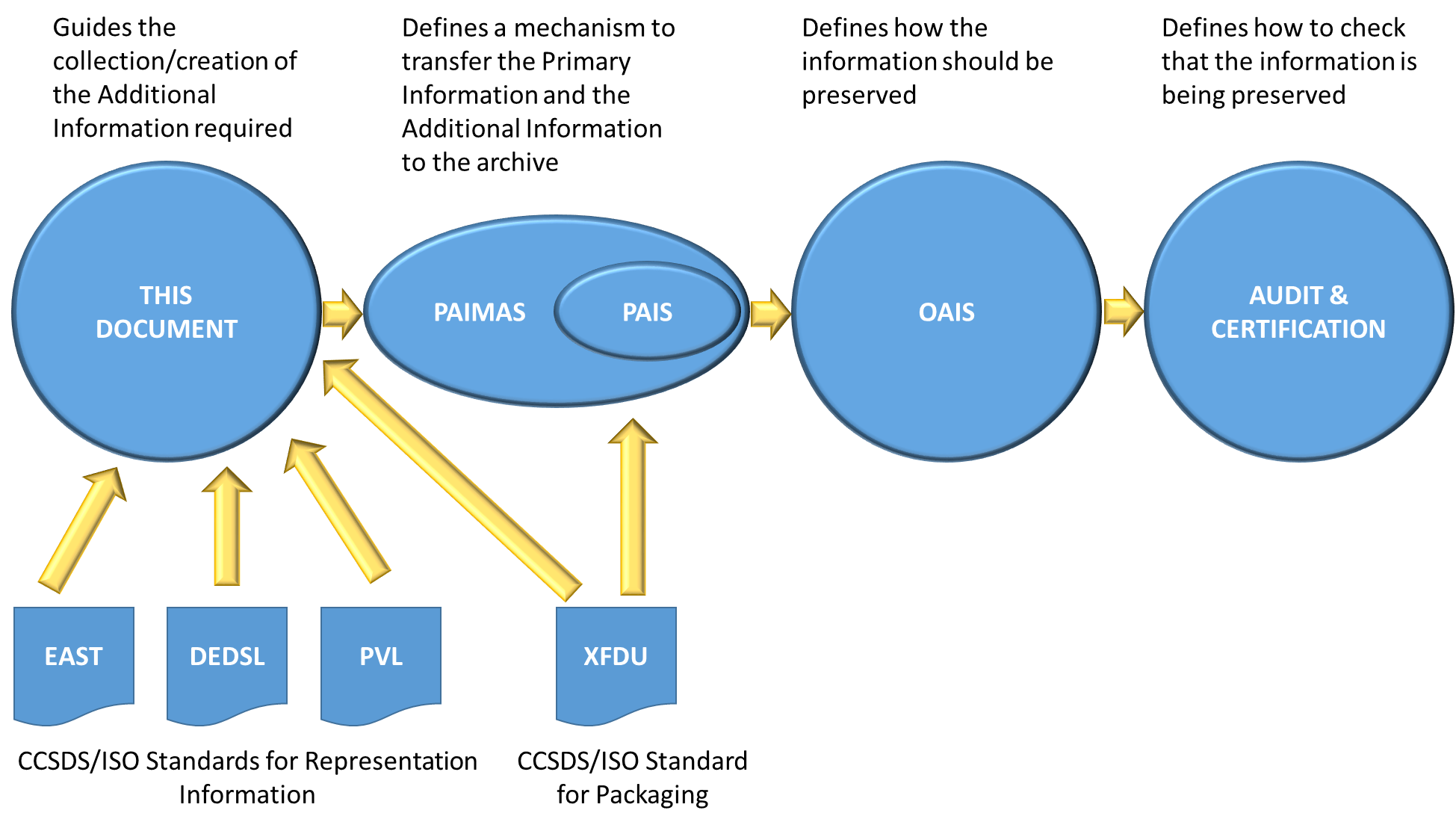


Figure ‑ Relationship between standards

OAIS **Error! Reference source not found.** is one of the most widely recognized and applied archival standards available today. An OAIS is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community.

PAIMAS **Error! Reference source not found.** defines a methodology for transferring data from an Information Producer to an Archives based on the four following phases: Preliminary, Formal Definition, Transfer, Validation. Required activities during each phase are identified.

PAIS **Error! Reference source not found.** provides the abstract syntax and an XML implementation of descriptions of data to be sent to an archive. These descriptions are negotiated agreements between the data Producer and the Archive and facilitate production of agreed data by the Producer and validation of received data by the Archive. The Recommended Standard includes an abstract syntax and one possible concrete implementation for the packages.

The Audit and Certification of Trustworthy Digital Repositories Recommended Practice **Error! Reference source not found.** provides metrics for use in assessing the trustworthiness of digital repositories or archives.

In addition there are other CCSDS/ISO standards may be used to create Representation Information (**Error! Reference source not found.**,**Error! Reference source not found.**and **Error! Reference source not found.**) and also to package information **Error! Reference source not found.**. There are many other techniques for creating Additional Information but these are outside the scope of this document.

In addition, the archival community has an existing, well established, set of concepts and terminology. The relationship with these and the OAIS concepts that underpin this document is described in Annex C

## Applicability

While this recommendation originates in the space community, it is being designed in a generic way and should be applicable to any science domain and to the wider records management and archival communities, to information created in an individual project or, perhaps, by an organisation as a whole. It should also be applicable to projects where the data already exists as well as where data is to be created in the future. It is also applicable to projects where the associated data is not the main focus of the project.

This document should be of use to funders, researchers, information creators, archive managers and end-users by helping to increase the effectiveness of preservation activities and the exploitation of information.

This guidance can form the basis on which plans, including Data Management Plans, can be constructed, updated and monitored, to achieve the objectives noted above.

## Rationale

Data that is collected or created needs to have Additional Information associated with it if it is to be independently understandable, usable and trusted as being authentic. That Additional Information changes over time, as hardware, software, the general environment and users’ tacit knowledge changes. OAIS uses the terms Representation Information and Preservation Description Information for this associated information. It must be accumulated over the life of the project. For example Provenance Information will accumulate over time, recording the things that have happened to the data.

In the case of information created by individual projects, funders are increasingly asking that Data Management Plans accompany any request for project funding, however these tend not to evolve with the project and are difficult to monitor.

Many project models have been proposed. However they do not focus on the activities needed at each stage that will help to ensure that the data can be optimally exploited over the long term.

There are a small number of generally applicable stages in the project where, typically, the responsibility is handed on from one individual or team to another. Each of those individuals or teams has specific knowledge about the information which subsequent individuals or teams may not possess. Therefore there is a need to specify the information to be captured at each of those stages. Improvements may be needed to, for example, the Representation Information, which was recorded in an earlier stage; this may arise if the information is better understood or reformatted or re-processed in later stages.

Therefore there is a need for guidance as to what additional information should be captured or improved through the various stages of the project.

This document should enable:

* the Producer (including for example scientists who create the data) to capture and record the relevant information in a timely manner;
* the Archive to be assured that it will receive adequate information to enable it to perform preservation activities and support exploitation (e.g. re-use or secondary use) of the information
* the user to re-use information more easily
* the funder/sponsor to be assured that the resources that they contribute to the creation of the information will have suitable pay-back