Structure Information, Semantic Information and Other Representation Information are both sub-types and components of Representation Information.

Representation Information is an Information Object that may have its own Data Object and its own Representation Information associated with understanding each Data Object, as shown in a compact form by the 'interpreted using' association. The resulting set of objects can be referred to as a **Representation Network**.

As an example, ISO 9660 (reference [D10]) describes text as conforming to the ASCII standard, but it does not actually describe how ASCII is to be implemented. It simply references the ASCII standard which is additional Representation Information that is needed for a full understanding. Therefore the ASCII standard is a part of the Representation Net associated with ISO 9660 and needs to be obtained by the OAIS in some form, or the OAIS needs to track the availability of this standard so that it may take appropriate steps in the future to ensure its ISO 9660 Representation Information is fully understandable.



Figure 4-11: Representation Information Object

4.2.1.3.2 Representation Networks

Representation Information, which is itself an Information Object, may be expressed in physical forms (e.g., a paper document) or in digital forms. When the Representation Information is in digital form, additional Representation Information is needed to understand the bits of the Representation Information as described in the previous subsection. In principle, this recursion continues until physical forms, which can be understood by the Designated Community, are encountered. For example, Representation Information expressed in ASCII needs the additional Representation Information for ASCII, which might

CCSDS 650.0

© ISO 2012 - All rights

be a physical document giving the ASCII standard. Each item of Representation Information can have multiple components, including multiple referenced Representation Information components; each with its own Representation Information.

To preserve the meaning of an Information Object, its Representation Information must also be preserved. This is most easily accomplished when the Representation Information objects are expressed in forms that are easily understandable, such as text descriptions that use widely supported standards such as ASCII characters for electronic versions. One problem with the use of only text descriptions is that such descriptions can be ambiguous. This is addressed by the use of standardized, formal description languages containing well-defined constructs with which to describe data structures. These languages may need to be augmented with text descriptions to convey fully the semantics of the Representation Information.

As the Knowledge Base of the Designated Community changes over time, the Representation Network may need to change accordingly. As noted in 2.2, an OAIS has a choice of whether to collect all the relevant Representation Information or to reference its existence in another trusted or partner OAIS Archive; this is an implementation and organization decision.

The Content Information must be defined and separated into Content Data Object and Representation Information. It is again an implementation and organization decision related to the way Data Objects are ingested and stored in the OAIS. For example, in the case of performing arts, the Content Data Object may be the score as a PDF document, and the Representation Information would include whatever information is needed to re-perform (as the way to use and understand) the piece, such as the way to display the PDF file, the audio processing software needed, placements of hardware such as loudspeakers, movement directions, and a description of how these relate to each other and to the Content Data Object, each of which may be quite complex, encoded in a separate way, and not easily described either simply as Structure or as Semantics. Alternatively, the Content Data Object may be multiple Data Objects including the score, the audio processing software needed, placements of hardware and movement directions. Each of these Data Objects will have its own Representation Information and there will need to be additional Representation Information that describes how the several Data Objects are related.

Two special types of Representation Information are **Representation Rendering Software** and **Access Software.** Representation Rendering Software is able to display the Representation Information in understandable forms. For example, the file and directory structure of many CD-ROMs conforms to ISO 9660. This standard is Representation Information describing how most CD-ROM file structures are to be implemented, and it may be obtained as a paper document. However, it may also be obtained as a digital object that needs to be understood as a PDF object. Rather than actually obtaining the documentation of PDF and writing software to understand the ISO 9660 object, an OAIS may use available PDF display software to render the ISO 9660 documentation humanly visible and readable. In this role the PDF display software is referred to as Representation Rendering Software because it is used to render the Representation Information. It also terminates the Representation Network. If the OAIS does not also obtain the associated description of PDF, it needs to record and track this fact because when PDF objects are no longer cost-effective for access and display, the ISO 9660 documentation expressed as a PDF object will need to be migrated to a new form.

Access Software presents some or all of the information content of an Information Object in forms understandable to humans or systems. It may also provide some types of access services, such as displaying, manipulating, processing, or sub-setting, to an Information Object. For some types of Digital Objects, such software may be widely available. It is not necessary for the OAIS to maintain or provide such software. The OAIS may want to maintain and provide this software for more specialized types of Digital Objects.

Since Access software will incorporate some understanding of the Representation Information, some Archives may attempt to use Access Software as a substitute for full Representation Information. Access Software source code, which embodies at least a partial understanding of the associated Representation Information, may be used as documentation expressing such Representation Information. A problem with this approach is that the desired Representation Information may not be clearly identifiable as it may be mixed with various processing and display algorithms, and may be incomplete since the code assumes an underlying operating environment. It may be difficult to tell, from the software code, what Representation Information is missing. The use of Access Software executables, without the source code, such as may occur with proprietary formats, presents a much greater risk for loss of information because it is more difficult to maintain an operating environment for software than to migrate documentation over time. The practical use of emulation techniques to preserve working software is an area of active research. This is a significant issue for those desiring to preserve a look and feel to information access. Migration and software preservation are discussed more fully in section 5.

4.2.1.4 Taxonomy of Information Object Classes Used by OAIS

There are many types of information involved in the Long Term Preservation of information in an OAIS. Each of these types can be viewed as a complete Information Object in that it contains a Data Object and adequate Representation Information to understand the data. This subsection builds on the discussions in 2.2 about the types of supporting information needed to enable Long Term Preservation and the discussion in the previous subsection on the role of Representation Information. The information modeling in this subsection discusses several types of Information Objects that are used in the OAIS. The objects are categorized by their content and function in the operation of an OAIS including Content Information objects, Preservation Description Information objects, Packaging Information objects, and Descriptive Information Object. The following subsections discuss the contents of each of the types of Information Object. Figure 4-12 shows a taxonomy of those Information Objects used within the OAIS.

CCSDS RECOMMENDED PRACTICE FOR AN OAIS REFERENCE MODEL



Figure 4-12: Information Object Taxonomy

4.2.1.4.1 Content Information

The Content Information is the set of information that is the original target of preservation by the OAIS. Deciding what the Content Information is may not be obvious and may need to be negotiated with the Producer. The Content Information, which is an Information Object as shown in figure 4-12, is the Content Data Object together with its Representation Information. The Content Data Object in the Content Information may be either a Digital Object or a Physical Object (e.g., a physical sample, microfilm). Any Information Object may serve as Content Information.

The Representation Information for a digital Content Data Object (both semantic and syntactic) is needed to fully transform the bits into the Content Information. In principal, this even extends to the inclusion of definitions (e.g., dictionary and grammar) of any natural language (e.g., English) used in expressing the Content Information. Over long time periods the meaning of natural language expressions can evolve significantly in both general and in specific discipline usage.

As a practical matter, the OAIS needs to have enough Representation Information associated with the bits of the Content Data Object in the Content Information that it feels confident that the members of the Designated Community can enter the Representation Network with enough knowledge to begin accurately interpreting the Representation Information. This is a significant risk area for an OAIS, particularly for those with an expert Designated Community, because jargon and apparently widely understood terms may be short-lived. In such cases extra care needs to be exercised to ensure that the natural evolution of the Designated Community Knowledge Base does not effectively cause information loss from the Content Information.

As described above for an Information Object in general, the Representation Information can also be viewed as being augmented by Access Software that supports the presentation of the

CCSDS 650.0-M-

Content Information to the Consumer. Examples of this type of software include word processors supporting complex document format representations of Content Information and scientific visualization systems supporting representations of Content Information as a time series or a multidimensional array. Access Software may include rights enforcement tools that allow the access to protected content. The software uses its knowledge of the underlying Representation Information to provide these services.

Often required information will be embedded in the software packages used by the Designated Community to present and analyze the Content Information. A reason for preserving working Access Software arises from a convenience factor. Even with a complete set of Representation Information, practical access to all or part of a digital Content Data Object requires the use of Access Software. Thus a software module that provides useful access to a digital Content Data Object may be preserved in a working state as a matter of convenience.

This is not difficult to do as long as the environment, which supports the software module, is readily available. This environment consists of some underlying hardware and an operating system, various utilities that effectively augment the operating system and storage and display devices and their drivers. A change to any of these may cause the software module to no longer function, to function incorrectly, or to be unable to present results to the application or human user. The complexity of these interactions is what traditionally makes the preservation of working software such an arduous task.

In summary, the use of Access Software to replace Representation Networks is attractive from the point of view of minimizing the resources needed to ingest data and provide current users with access to data. However, the reliance on working software can provide major problems for Long Term Preservation when that software ceases to function. Indefinite Long Term information preservation requires a full and understandable description of the Representation Information. Subsection 5.2 (Preservation of Access and Use Services) discusses some techniques that can be used to preserve software over time and the risks associated with this approach.

An important function of the OAIS is deciding what parts of the Content Information are the Content Data Object and what parts are the Representation Information. This aspect is critical to a clear understanding of what is being preserved. The identification of digital Content Information with its Representation Information objects can be addressed by a series of steps, as follows:

- 1) Identify the bits comprising the Content Data Object of the Content Information.
- 2) Identify a Representation Information object that, in some way, addresses all the bits of the Content Data Object and converts them into more meaningful information.
- 3) For the Representation Information object identified, examine its content to identify if it requires additional Representation Information objects. If it does, obtain the required Representation Information objects. Repeat this step at least until no additional Representation Information objects are identified as required for the Designated Community.

CCSDS 650.0

- 4) Of the Representation Information objects addressed in step 3, for each that is held as a Digital Object, identify any required Representation Information object and repeat steps 3 and 4 until no new Representation Information objects are identified.
- 5) The Content Information consists of the Content Data Object and each of the Representation Information objects identified in steps 2 through 4.

As an example of this practice, consider an electronic file containing a sequence of values obtained from a sensor looking at the Earth's environment. There is a second file, encoded using ASCII, which provides information on how to understand the first file. It describes how to interpret the bits of the first file to obtain meaningful numbers. It explains what these numbers mean in terms of the physics of the observation being conducted. It provides the date and time period over which the observations were made, an average value for the observed values, and who made the observations. These two files are submitted to an OAIS for preservation.

Assume that the OAIS determines that the Content Information to be preserved is the observed bits together with their values as numbers and the physical meaning of these numbers. This information is conveyed by the bit sequence within the first file together with the Representation Information from the second file that is needed to transform the first file's bits into meaningful physical values. Neither the first file's underlying media nor the particular file system carrying the bits is part of the Content Information in this example. Only part of the second file's content is considered a part of the Content Information and this is the part that enables the transformation of the bits from the first file into meaningful physical values. In fact this second file does not carry all the Representation Information is needed:

- information that the second file is encoded in ASCII so that it can be read as meaningful characters;
- information on how the characters are used to express the transformations from bits to numbers to meaningful physics values.

This information, typically referred to as a combination of format information and data dictionary information, may also include instrument calibration values and information on how the calibrations are to be applied. All this information may be widely understandable once the ASCII characters are visible because it has all been expressed in English (or some other natural language), or some of it may be in more structured forms that will need additional Representation Information to be understood.

Therefore, the Representation Information of the second file needs additional Representation Information, and this information may need additional Representation Information, etc., forming a linked set of Representations of Representations. This is a good example of the complex Representation Net.

In the example above, there was a determination that the Content Information consisted of the observed sensor values and their meanings. This is by no means the only choice that could have been made. It could just as easily have been decided that the Content Data Object of the desired

Content Information was the bit sequences within the first file together with the all the bit sequences within the second file. The fact that some of these latter bit sequences are used to interpret the first file's bit sequences is just an example of a set of bits that is somewhat self-describing. It is irrelevant that some of the bits in the second file are the basis for information on the date and time period over which the observations were made, the average value for the observed values, and who made the observations. Once it has been determined that all these bits constitute the Content Data Object of the Content Information, then the Representation Information is that information needed to turn them into meaningful information. How extensive this meaning is to be carried and how far the Representation Network needs to be carried are local issues for the OAIS and its related Producer and Consumer communities.

As another example, consider an electronic file containing a word processing document. This binary Data Object will have a complex format that can be seen as a document only after it has been viewed through use of associated Representation Information. In common practice, this viewing will be provided by Access Software that can use internal, or external, Representation Information. The Content Data Object is most likely to be defined as the bit sequence content of the electronic file. The Representation Information is a description of the word processing format, at a minimum, and may include information deemed needed to adequately understand the meaning of the document as viewed. If the word processing format is proprietary, and if adequate Representation Information cannot be acquired which will at the least allow simply viewing, to ensure its Long Term Preservation it may be necessary to migrate the document to another (possibly non-proprietary) format for which Representation Information is more openly available.

As a variation on the above example, it may be decided that the Content Information to be preserved is not the full word processing view of the document, but simply a sequence of text paragraphs that can be adequately represented by ASCII characters. In this case, the OAIS may decide to extract the relevant text characters and save them as a text file. The Content Data Object would be defined, most likely, as the bit stream made up of these characters. The Representation Information would be a description of how to interpret this bit stream as characters, together with any additional information deemed needed to adequately understand the meaning of the text.

4.2.1.4.2 Preservation Description Information

In addition to Content Information, the Archival Information Package must include information that will support the trust in, the access to and context of the Content Information over an indefinite period of time. The specific set of Information Objects, which are required for this function, is collectively called **Preservation Description Information** (PDI). The PDI must include information that is necessary to adequately preserve the particular Content Information with which it is associated. It is specifically focused on describing the past and present states of the Content Information, ensuring it is uniquely identifiable, and ensuring it has not been unknowingly altered.

This information is typical for all types of Archives and has been classified in the context of traditional Archives. However, the class definitions must be extended for digital Archives.

CCSDS 650.0

The following definitions are largely based on the categories discussed in the paper 'Preserving Digital Information' (reference [D2]). The relationship between the concepts in OAIS Reference Model and the Preserving Digital Information paper are discussed in annex B of this document. Table 4-1 provides illustrative examples of this information for various popular Content Information types.

- Reference Information identifies, and if necessary describes, one or more mechanisms used to provide assigned identifiers for the Content Information. It also provides those identifiers that allow outside systems to refer, unambiguously, to this particular Content Information. Examples of these systems include taxonomic systems, reference systems and registration systems. In the OAIS Reference Model most if not all of this information is replicated in Package Descriptions, which enable Consumers to access Content Information of interest.
- Context Information documents the relationships of the Content Information to its environment. This includes why the Content Information was created and how it relates to other Content Information objects existing elsewhere.
- Provenance Information documents the history of the Content Information. This tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated, providing an audit trail for the Content Information. This gives future users some assurance as to the likely reliability of the Content Information as it contributes to evidence supporting Authenticity. Provenance can be viewed as a special type of context information.
- Fixity Information provides the Data integrity checks or validation/verification keys used to ensure that the particular Content Information object has not been altered in an undocumented manner. Fixity Information includes special encoding and error detection schemes that are specific to instances of Content Objects. Fixity Information does not include the integrity preserving mechanisms provided by the OAIS underlying services, error protection supplied by the media and device drivers used by Archival Storage. The Fixity Information may specify minimum quality of service requirements for these mechanisms.
- Access Rights Information identifies the access restrictions pertaining to the Content Information, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the Submission Agreement, related to both preservation (by the OAIS) and final usage (by the Consumer). It also includes the specifications for the application of rights enforcement measures.

These classifications provide a minimum set of PDI; they do not specify a data structure.

Content Information	Poforonoo	Provenence	Contoxt	Eivity	Access Dights
Space Science Data	Object identifier	Instrument description	 Calibration history 	CRC Checksum	Identification of the properly
	 Journal reference Mission, instrument, title, attribute set 	 Principal Investigator Processing history Storage and handling history Sensor description Instrument Instrument mode Decommutati on map Software interface specification Information Property Description 	 Related data sets Mission Funding history 	Reed- Solomon coding	authorized Designated Community (Access Control) Permission grants for preservation and for distribution Pointers to Fixity and Provenance Information (e.g., digital signatures, and rights holders)
Digital Library Collections	 Bibliographic description Persistent identifier 	 For scanned collections: metadata about the digitization process pointer to master version For born-digital publications: pointer to the digital original Metadata about the preservation process: pointers to earlier versions of the collection item change history Information Property Description 	Pointers to related documents in original environme nt at the time of publication	 Digital signature Checksum Authenticity indicator 	 Legal framework(s) Licensing offers Specifications for rights enforcement measures applied at dissemination time Permission grants for preservation and for distribution Information about watermarking applied at submission and preservation time Pointers to Fixity and Provenance Information (e.g., digital signatures, and rights holders)

Table 4-1: Examples of PDI

CCSDS RECOMMENDED PRACTICE FOR AN OAIS REFERENCE MODEL

Content Information Type	Reference	Provenance	Context	Fixity	Access Rights
Software Package	 Name Author/ Originator Version number Serial number 	 Revision history Registration Copyright Information Property Description 	 Help file User guide Related software Language 	 Certificate Checksum Encryption CRC 	 Designated Community Legal framework(s) Licensing offers Specifications for rights enforcement measures applied at dissemination time Pointers to Fixity and Provenance Information (e.g., digital signatures, and rights holders)

The OAIS needs to explicitly decide what the exact definition of Content Information is in order to be able to ensure that it also has the PDI needed to preserve the Content Information. Once the Content Information has been determined, it is possible to assess the Preservation Description Information.

4.2.1.4.3 Packaging Information

The **Packaging Information** is that information which, either actually or logically, binds or relates the components of the package into an identifiable entity on specific media. For example, if the Content Information and PDI are identified as being the content of specific files in a TAR file, then the Packaging Information may include the name of the TAR file and the fact that it is a TAR file including details of any specific encoding. On the other hand if the Content Information and PDI are files on a CD-ROM, then the Packaging Information may include the ISO 9660 volume/file structure on the CD-ROM. These choices are the subject of local Archive definitions or conventions. The Packaging Information does not necessarily need to be preserved by an OAIS since it does not contribute to the Content Information or the PDI. However, there are cases where the OAIS may be required to reproduce the original submission exactly. In this case the Content Information is defined to include all the bits submitted.

The OAIS should also avoid holding PDI or Content Information only in the naming conventions of directory or file name structures. These structures are most likely to be used as Packaging Information. Packaging Information is not preserved by all Digital Migrations. Any information saved in file names or directory structures may be lost when the Packaging Information is altered. The subject of Packaging Information is an important consideration to the Migration of Information within an OAIS to newer media. This subject is addressed in detail in section 5 of this document.

CCSDS 650.0-M-

4.2.1.4.4 Descriptive Information

The Information Objects described previously in this section provide the information necessary to enable the Long Term Preservation function of the Archive. In addition to preserving information, the OAIS must provide adequate features to allow Consumers to locate information of potential interest, analyze that information, and order desired information. This is accomplished through a specialization of the Information Object called Descriptive Information, which contains the data that serves as the input to documents or applications called **Access Aids**. The Descriptive Information is generally derived from the Content Information and PDI. The Descriptive Information can be viewed as an index to enable efficient access to the associated Information Package via associated Access Aids. **Access Aids** are documents or applications that can be used to locate, analyze, retrieve, or order information from the OAIS.

4.2.2 LOGICAL MODEL OF INFORMATION IN AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)

The previous subsection defines the types of Information Objects that are needed by an OAIS to enable the Long Term Preservation of information and effective access to the preserved information by the Designated Community. This subsection uses those Information Object descriptions to model the conceptual information structures required to accomplish these functions. The models presented in this subsection are not intended to imply an implementation, but rather to highlight the relationship among the types of information needed in the archival process.

4.2.2.1 Information Package

The conceptual structure for supporting Long Term Preservation of information is the Information Package. An Information Package is a container that contains two types of Information Objects, the Content Information and the Preservation Description Information (PDI); the Information Package can be associated with two other types of Information Objects, Packaging Information and Package Descriptions. There are several types of Information Packages that are used within the archival process. These Information Packages may be used to structure and store the OAIS holdings; to transport the required information from the Producer to the OAIS, or to transport requested information between the OAIS and Consumers. There are differing information requirements for each of these functions. The UML diagram in figure 4-13 illustrates the conceptual view of an Information Package. This UML diagram shows that an Information Package contains zero or one Content Information objects, zero or more PDI objects, and is associated with exactly one piece of Packaging Information, which identifies and delimits the Information Package. The Information Package is also associated with one or more Package Descriptions that describe the Content Object to enable efficient access.