

Reconsidering the OAIS Reference Model for Record Management and Archiving in a Cloud Computing Environment

Jan Askhoj, Mitsuharu Nagamori, Shigeo Sugimoto, University of Tsukuba, Japan

Email: janaskhoej@gmail.com, {nagamori, sugimoto}@slis.tsukuba.ac.jp

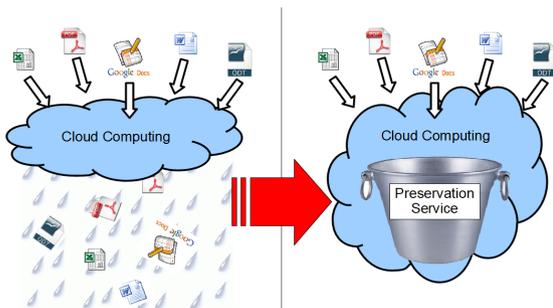


BACKGROUND

In recent years, cloud computing has become an increasingly popular technology. It is reported that the cloud computing market will grow to more than \$150 billion in 2013. But...

Cloud providers do not currently provide services for standards compliant archiving and preservation of records.

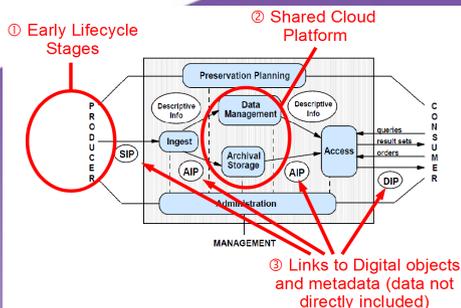
Moving to cloud computing with long-term preservation built in



If nothing is done, records without preservation management stored in the cloud could become inaccessible with the passage of time. As can be seen in the figure above, such poorly managed and obsolete records may fall out of the cloud and out of use. We aim to provide a model for cloud archiving systems to provide a suitable preservation service.

AN ESTABLISHED MODEL FOR ARCHIVES

The OAIS Model is the de-facto model for digital archiving systems. However, with cloud computing comes a number of changes not covered by OAIS.

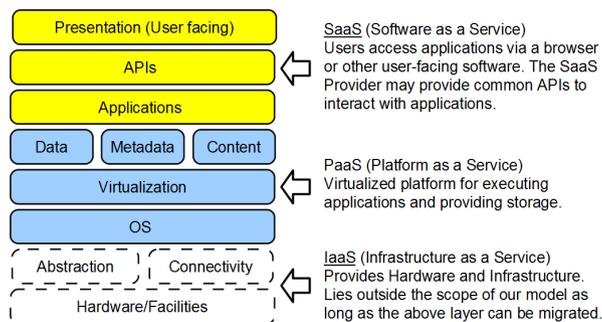


1. Early stages of the records lifecycle. With cloud computing and the shared platform it offers, digital objects can be immediately accessible to an archive, allowing early preservation planning.
2. Shared use of a trusted platform for storage, outside the control of an archive. Such functionality overlaps with several OAIS functions such as database administration and storage.
3. Digital objects and metadata not included in information packages. With a shared platform having everything in a package is unnecessary.

USING A LAYERED MODEL

Cloud services can be abstracted in a layered model, where each layer offers services to the layers above, allowing sharing between several systems. The below figure shows a general model of how cloud services are currently offered to customers who can then choose the service level they need.

A general layered model for cloud computing services

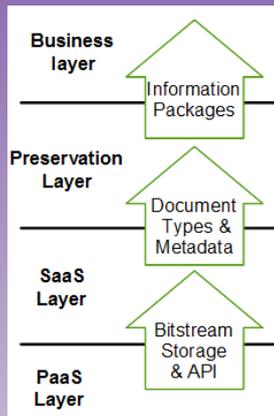


Our object is to offer preservation as defined in the OAIS model as a similar, easily accessible service. This will allow organizations to save records in the cloud, knowing preservation is being managed. It will also give archives access to securely stored records with standardized metadata.

SERVICE PROVISION

To provide preservation services, we have expanded on the general model above and added two new layers. Each layer is dependent on defined services from the layer below (see figure).

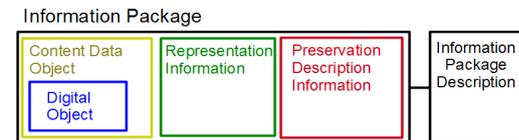
One of the main objects of our model is defining information types, formats and parameters for communication between layers. If these are clearly defined, the need for organizations to worry about the inner workings of below layers disappears.



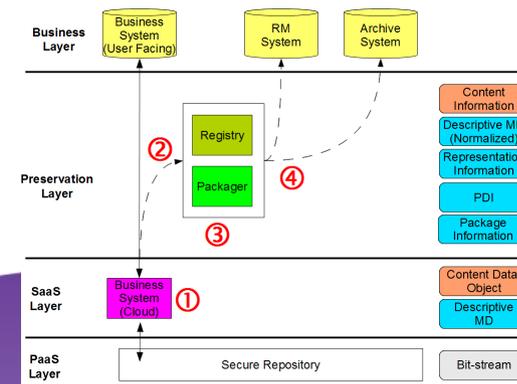
Business Layer: User facing Business Systems, Archives/Records Management Systems.
Preservation Layer: Ensures Information Objects can be accessed by Archives/Records Management Systems by offering: Registration, Harvesting, Conversion, Representation Information and Packaging.
SaaS Layer: Applications that represent bit-strings as documents used by systems and users.
PaaS Layer: Bit-level trusted repository and application platform.

INFORMATION FLOW

Many types of information are needed to create an Archive Information Package as defined in the OAIS model and as used in the Business Layer. An example of such a package can be seen below:



This information is provided by the entities in our layered model from the time of records declaration in the originating application to the creation of the archive package. The entire information flow is as follows:



1. Once a document is declared a record, the Packager is notified and record and metadata is harvested.
2. Record and metadata are normalized (converted into a standard XML format and crosswalked) based on information in the Registry.
3. Additional Metadata for preservation is added.
4. An Information Package is created based on information from the originating system, the package creation process and registries of agents, schemas and digital objects.

CONCLUDING REMARKS

We believe our proposed model has a number of advantages when developing a cloud archive system framework.

- 1) It builds on OAIS model concepts and information types.
- 2) It helps expand archive services such as trusted storage and preservation to earlier stages in the document lifecycle.
- 3) It makes archive system design simpler by allowing organizations to differentiate between different layers of service and disregard lower level functionality.