# Applying the OAIS standard to CCLRC's British Atmospheric Data Centre and the Atlas Petabyte Storage Service

Corney, D.R., De Vere, M., Folkes, T., **Giaretta, D.,** Kleese van Dam, K., Lawrence, B. N., Pepler, S.J., Strong, B

**Central Laboratory for the Research Councils**

## Abstract

The OAIS reference model is being considered as a standard for adoption by the Digital Curation Centre. This work will form part of that assessment. The British Atmospheric Data Centre (BADC) at CCLRC is evaluating CCLRC's Atlas Petabyte Storage Service (APS) as its long-term data storage system. This paper critically examines the application of the OAIS standard to the proposed BADC/APS service, and draws conclusions about practical implications of implementing the OAIS standard. It identifies lessons that learned from running the archive service. Implications for improvements to the OAIS standard will also be drawn.

## 1. Background

The OAIS[1] reference model is being considered as a standard for adoption by the Digital Curation Centre[2]. This work being undertaken by CCLRC as part of the DCC consortium will form part of that foundation assessment.

The OAIS model and ISO standard has emerged as a useful tool in the challenges of digital preservation and curation. While the capacity to create and consume digital data has increased steadily, the capacity to manage the long term stewardship of data has been slower to develop. [Intro to OAIS, BF Lavoie, Jan 2004[3]].

The British Atmospheric Data Centre[4] (BADC) based at CCLRC is evaluating the use of CCLRC's Atlas Petabyte Storage Service[5] (APS) as its long term data storage system. These two independent components form separate parts of a repository that can be examined from the perspective of the OAIS model. The BADC fulfils the OAIS roles of Ingest and access and local storage, while the APS will fulfill the role of archival storage. Jointly, the BADC and APS will fulfil OAIS roles of Data Management and Administration.

This paper will critically examine the application of the OAIS standard to the proposed BADC/APS data service, and draw conclusions about the practical implications of implementing the OAIS standard, lessons that may be learned about the way that the archive service is run.

An OAIS repository has a set of mandatory responsibilities. It must be able to:

- Negotiate for and accept information from information producers
- Obtain sufficient control of the information in order to meet the long-term preservation objectives
- Determine the scope of the archives user community
- Ensure that the preserved information is independently understandable to the user community
- Follow documented policies and procedures to ensure information is preserved against all contingencies, and to enable dissemination of authenticated copies of the preserved information in its original form, or in a form traceable to the original.
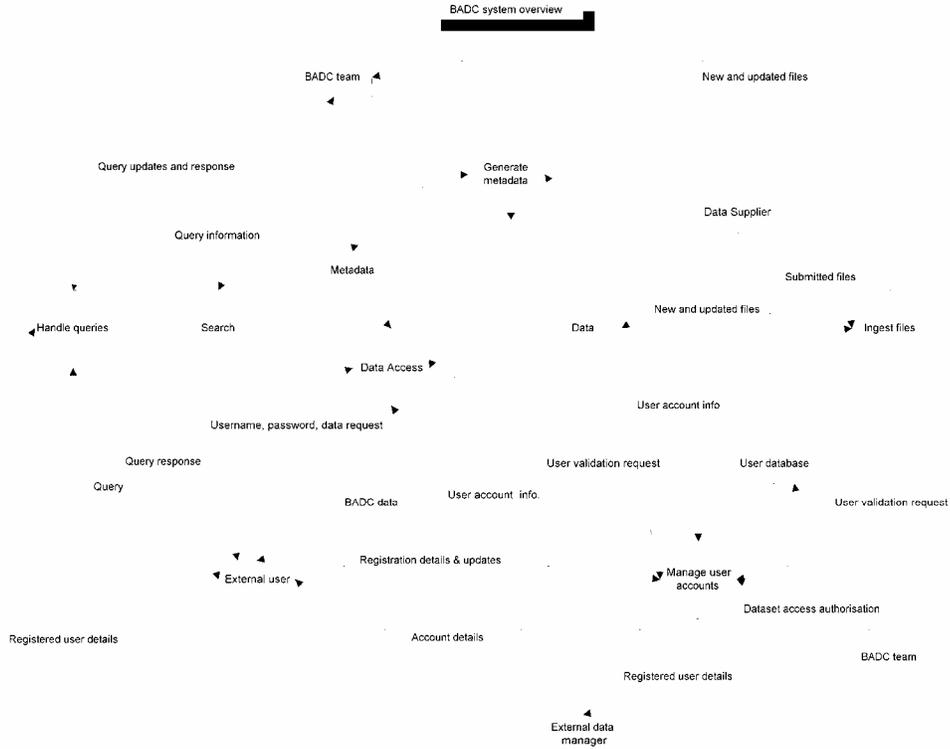- Make the preserved information available to the community.

BADC system overview

BADC team

New and updated files

Query updates and response

Generate
metadata

Data Supplier

Query information

Metadata

Submitted files

New and updated files

Handle queries      Search          Data          Ingest files

Data Access

User account info

Username, password, data request

Query response                 User validation request        User database

Query            BADC data      User account info.                         User validation request

Registration details & updates           Manage user
accounts

External user                                                Dataset access authorisation

Registered user details                   Account details                         BADC team

Registered user details

External data
manager

**Figure 1: Functional diagram of the BADC**

**Preservation Planning**

P
R
O
D
U
C
E
R

Descriptive
Info

**Data
Management**

Descriptive
Info

C
O
N
S
U
M
E
R

queries

result sets

SIP

**Ingest**

**Access**

orders

**Archival
Storage**

AIP

AIP
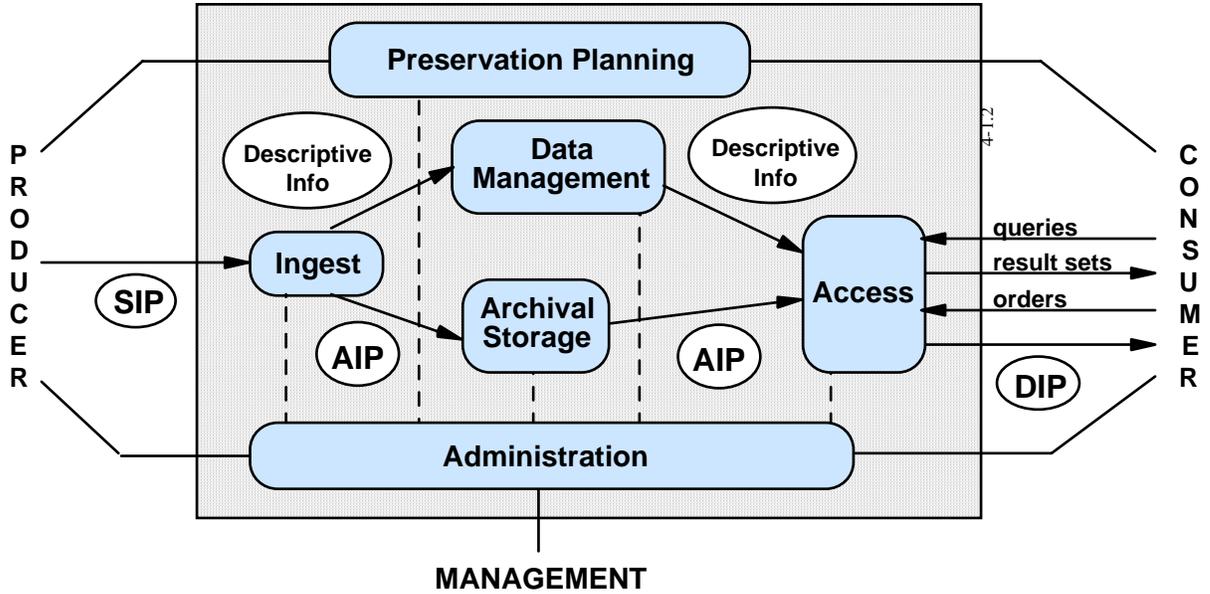
DIP

**Administration**

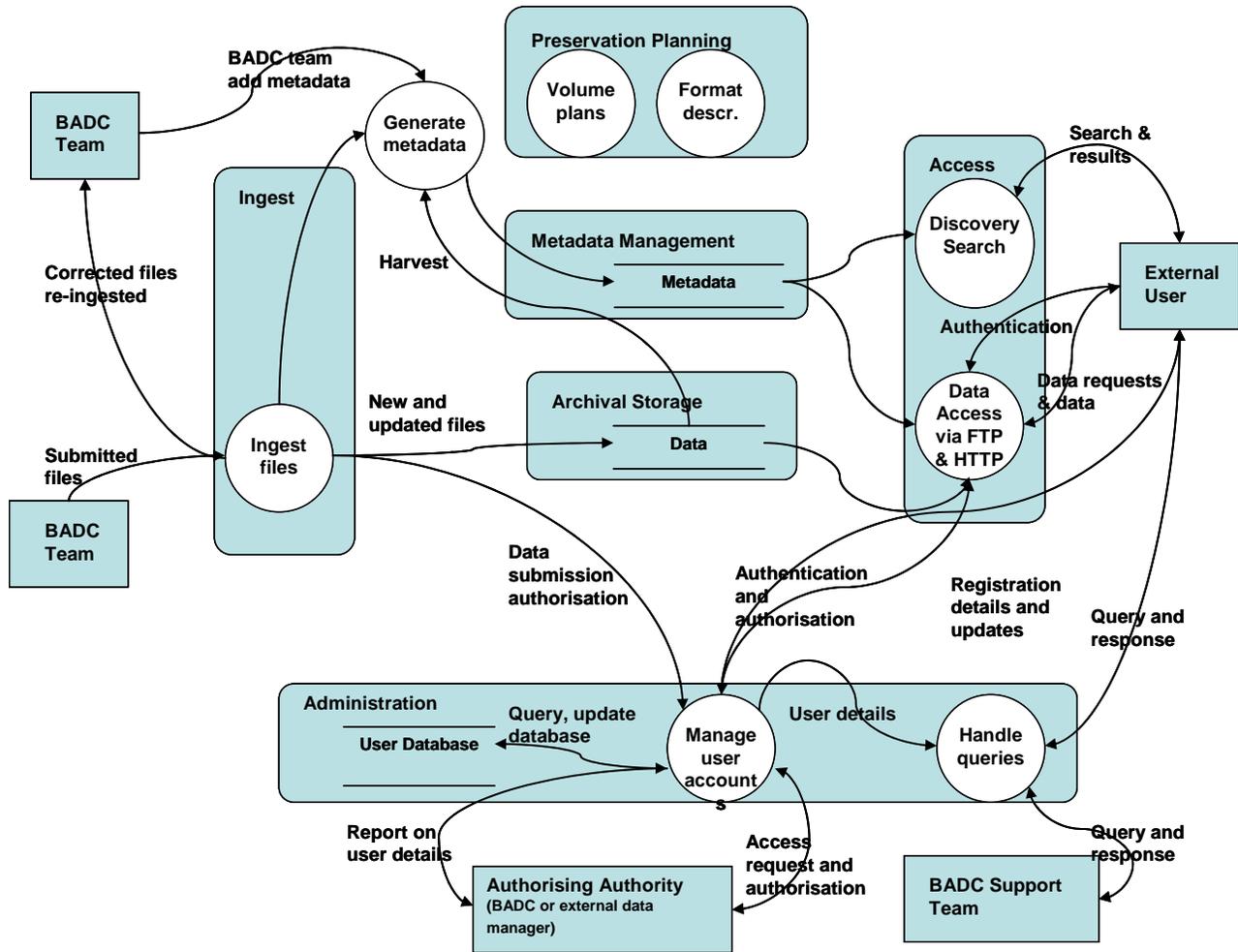**MANAGEMENT**

**Figure 2: OAIS Functional Model**

**Figure 3: BADC mapped to OAIS Functional Model**

A functional description of the BADC is provided in figure 1 and a high level functional description of the OAIS model is provided in figure 2. This will be the starting point for the proposed comparison and analysis.

## 2. OAIS Functional Model

The OAIS recognises the following functional entities.

**Ingest**: This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers and prepare the contents for storage and management within the archive

**Archival Storage**: This entity provides the services and functions for the storage, maintenance and retrieval of Archival Information Packages

**Data Management**: This entity provides the services and functions for populating, maintaining, and accessing both descriptive information which identifies and documents archive holdings and internal archive administrative data.

**Administration**: This entity manages the overall operation of the archive system

**Preservation Planning**: This entity monitors the environment of the OAIS and provides recommendations to ensure that the information stored in the OAIS remain accessible to the Designated User Community over the long term even if the original computing environment becomes obsolete.

**Access**: This entity supports consumers in determining the existence, description, location and availability of information stored in the OAIS and allowing consumers to request and receive information products

Figure 3 identifies the corresponding entities within the BADC structure and maps them to the comparable OAIS entities. Each entity may be considered in turn, and all similarities and any differences between the BADC and the OAIS model considered to see what this tells us about the practical implementation of the OAIS model, and or the practical implementation of the BADC data service and archive.

All functions are carried out within the BADC apart from the Archival Storage entity, which is devolved to the Atlas Petabyte Storage facility, used to store the bulk data from the BADC. The APS also performs some Administration and Data Management.

It can be considered from two perspectives: a) its role as a back up for the BADC viewed from OAIS, and b) from a stand alone OAIS perspective. Since the BADC and APS are separate independent entities, the linkage between these is very low, although there are be lessons for the managerial and administration aspects, which need to be shared.

## 3. Comparison with OAIS of BADC/APS combination

Stepping through each of the OAIS mandatory responsibilities we can see that:
- the BADC does negotiate and accept information from producers
- it generally obtains control of the information
- the archive's user community is that supported by NERC, but non-UK users can also use the data

- the data is, currently understandable by the users. However users are often simply supplied with data and are expected to use whatever tools they have to convert this to usable information. Preservation Planning is probably not sufficiently emphasised.
- The BADC does, through the APS, have in place adequate policies to ensure the authenticity and bit-preservation of its data. Certainly within the APS there are checks at each internal data transfer. On the other hand transfers to the user largely rely of the reliability of whatever transfer mechanism is used e.g. FTP.
- BADC certainly makes the information available to the community.

The weakest area here is probably that of Preservation Planning (see Figure 6). As can be seen in Figure 3 there is little explicit in the area of monitoring technology or the Designated Community. However it is clear that this is done to some extent, at least informally.

On the other hand one must ask whether, to the extent it is done at all, this monitoring can be said to be done uniformly over all the datasets held by the BADC.

It may be that consistency with OAIS must be considered more finely grained – in fact on a dataset by dataset basis.

Whatever the level of granularity it may be that some benefit may be gained by using some of the data descriptions and tools being developed by the European Space Agency in the Advance Data Archive[8] (ADAR) project. There many types of Earth Observation data are described at the bit level and generic tools are used to subset and reformat the original data. In addition Data Dictionaries supplement the semantic information.
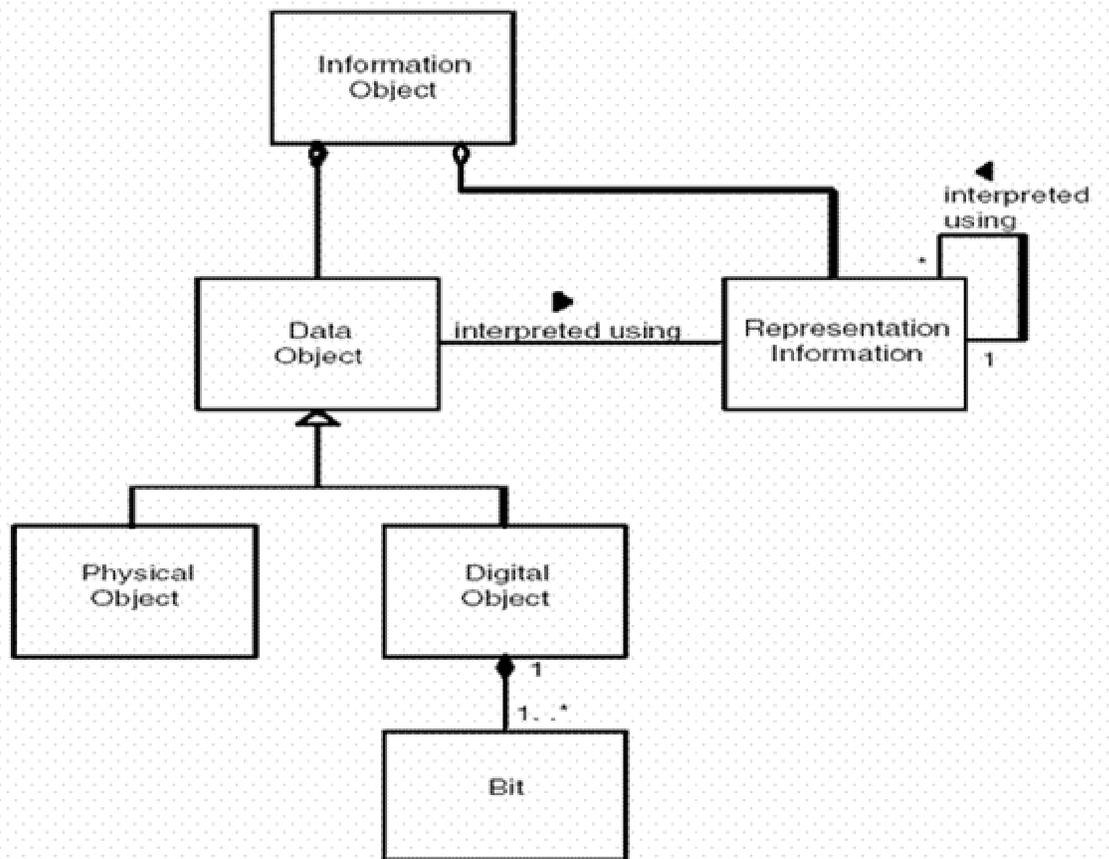
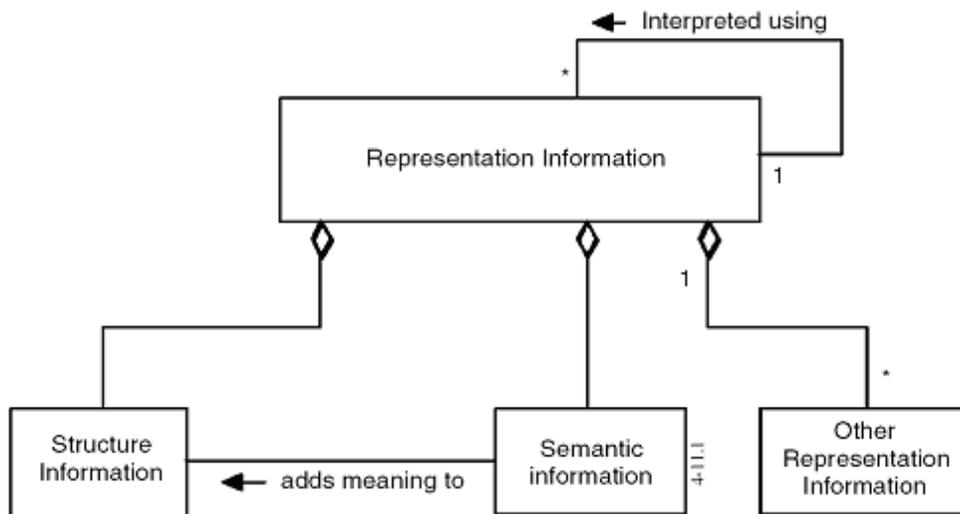**Figure 4: Information Object, showing Representation Network**
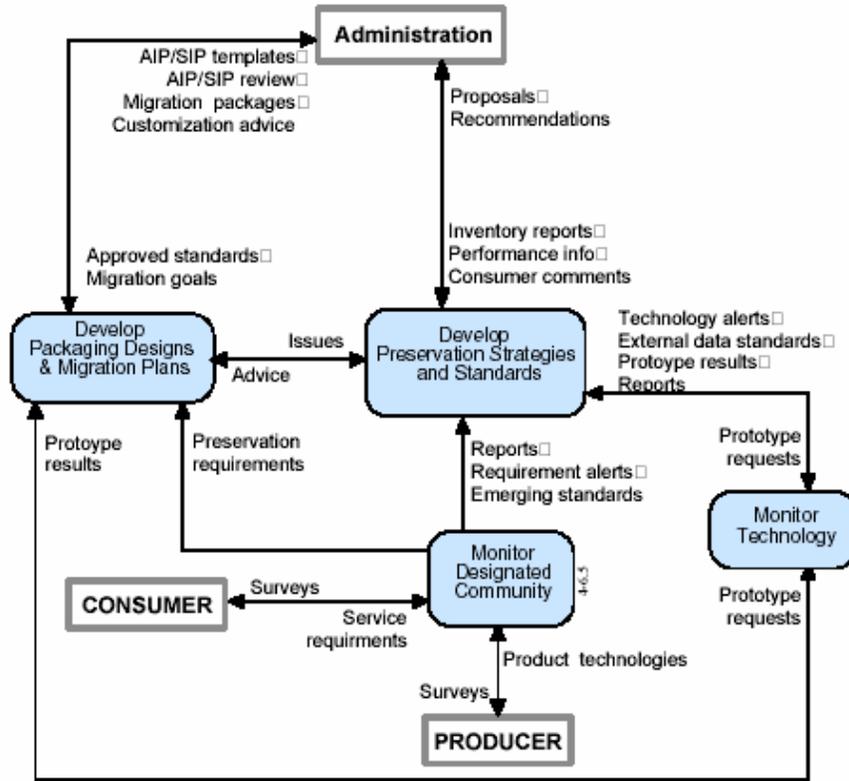


**Figure 5: Representation Information**

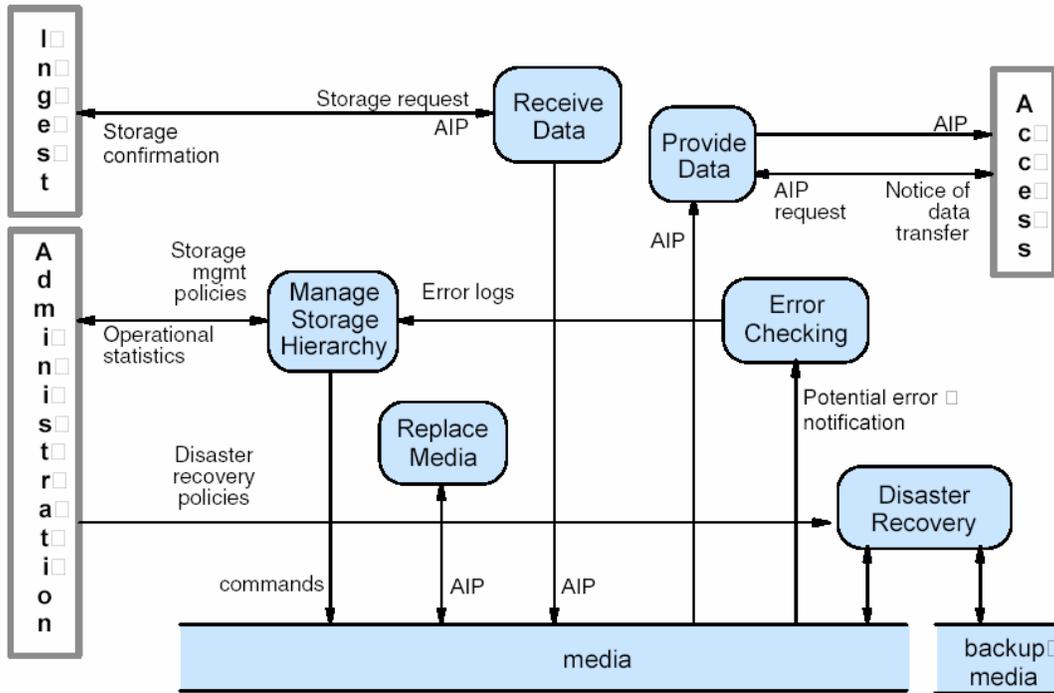**Figure 6: Functions of Preservation Planning**



**Figure 7 Functions of Archival Storage**

## 4. Comparison of OAIS with APS

It is an interesting exercise to see to what extent the APS by itself may be mapped to the OAIS model.

A key question concerns the extent to which we may call these bit sequences 'information'.

- Users such as the BADC send bit sequences and a limited amount of metadata to the APS for storage, and clearly this is a negotiated transfer.

The Designated Community must be, at least for the data they deposit, the BADC. They can, now, understand the bit sequences, given that the correct metadata, which must be the metadata which BADC originally sent when depositing the bits, is returned. To that extent we surely must agree that APS has adequate, however meagre, Representation Information to accompany the bits.

- The APS clearly has control of the information.
- The BADC is the Designated Community – at least for a limited amount of data – that which it originally deposited – and BADC can access the information.
- To ensure that the information is independently understandable by the BADC could be trivial.

If the BADC no longer is interested in, and/or no longer has the tools to use, a particular dataset (sequences of bits) held by the APS then it could be said that the Designated Community falls to zero.

- The APS does have in place policies to ensure that authentic copies are available.
- It certainly makes the information available to the BADC>

On the other hand clearly there is a question of funding and ensuring a planned succession of data holder. This consideration applies to BADC, APS or the combination.

## 5. Conclusions

The BADC/APS combination can clearly be mapped to the OAIS reference model. The latter provides, at the very least, a checklist to use to identify limitations in the system in terms of information preservation.

On the other hand, by not simply ticking each item on the checklist we can raise some important questions about granularity of compliance. By looking at what may be regarded as an extreme case, namely the APS by itself, we can raise other questions. These questions will inform the discussions of the RLG Certification Task Force[9] in its development of certification standards.

---

[1]

http://www.ccsds.org/CCSDS/documents/650x0b1.pdf

[2] Curating for the Future – the work of the Digital Curation Centre. Giaretta, D. Robinson, B., Lyon, L. AHM 2004 Poster

[3]

http://www.dpconline.org/docs/lavoie_OAIS.pdf

[4] http://badc.nerc.ac.uk/home/index.html

[5] http://www.e-science.clrc.ac.uk/web/services/datastore

[8]

http://earth.esa.int/rtd/Articles/ADAR_PVDST_Abstract.doc

[9]

http://www.rlg.org/en/page.php?Page_ID=367