

# De-mystifying OAIS compliance:

**Benefits and challenges of mapping the OAIS reference model to the GESIS Data Archive** by Natascha Schumann<sup>1</sup> and Astrid Recker<sup>2</sup>

## GESIS

### Abstract

Since its initial publication over a decade ago, the OAIS Reference Model, its concepts and terminology, have become essential to the digital preservation discourse. In this discourse, the topos – or myth – of “OAIS compliance” continues to play a central role as archives and repositories seek to demonstrate their fitness for the challenge of digital preservation. After briefly considering what OAIS is (and can be used for) and what it is not – namely, an abstract reference model, but not

an architecture that can be implemented directly –, we will use the GESIS Data Archive for the Social Sciences as

an example of mapping OAIS to an existing archive. We will then explore positive effects and benefits, as well as difficulties of completing this process. Thus, such a mapping can be taxing for an established archive: As most of the workflows have grown and proven their adequacy over a considerable period of time, taking a step back and viewing these processes from a new perspective is a challenge in itself<sup>3</sup>.

Keywords  
OAIS, standards, trusted digital repositories

### Introduction

The importance of the Reference Model for an Open Archival Information System (OAIS), which since the releases of its first draft versions in 1997 and 1999 has shaped and influenced digital preservation discourse like hardly any other model, is undisputed

(see, for example, Lee, 2012; Allinson, 2006; Oßwald, 2010). The OAIS standard has not only provided us with a common language – and thereby a common understanding of what it is that archives do when they preserve digital information objects; it has also given important impulses to move towards greater standardization in the field of digital preservation, including the development of criteria and procedures to analyze and assess archival preservation and dissemination practice (e.g. ISO 16363:2012 “Audit and

---

## Reference Model for an Open Archival Information System (OAIS) ...

---

certification of trustworthy digital repositories”). Despite – or possibly because – of the model’s influence, the ubiquity of its terminology and concepts, one frequently encounters misconceptions as to what OAIS is and what it is for. Often, these seem to be linked to a misunderstanding of what a reference model is. On a more concrete level, it is the notion of OAIS compliance and its – sometimes seemingly unreflected – use in archive self-portrayals or in the description of software packages which appears problematic.

### OAIS is a reference model

Often, one will hear or read about “OAIS being implemented” in some organization or another. What is usually meant by this is that a system is being built or adapted which conforms to the OAIS model in some way. Such statements are misleading, however, because as a reference model, OAIS can

by definition not be directly implemented: It is an abstract and highly generic conceptualization of a preservation and dissemination environment.

Thus, as defined by the Organization for the Advancement of Structured Information Standards, “[a] reference model is an abstract framework for understanding significant relationships among the entities of some environment...” (n.d.). As such, it “is not directly tied to any standards, technologies or other concrete implementation details, but it does seek to provide a common semantics that can be used unambiguously across and between different implementations” (ibid.). This is in accordance with the purpose of the OAIS model as given in the standard itself, which among other things states that the model:

- “provides a framework, including terminology and concepts, for describing and comparing architectures and operations of existing and future Archives”
- “provides a framework for describing and comparing different Long Term Preservation strategies and techniques” (CCSDS 2012, p. 1-1)

The OAIS reference model can be compared to a language operating on a meta-level, allowing us to speak about archives, their architectures and processes. Therefore, OAIS will not make a certain preservation strategy or technique a requirement. It will define the characteristics of a strategy it deems successful, but it will not prescribe a concrete, monolithic solution.

This means that OAIS cannot be used as a check list which can be ticked off as one builds an archival information system. Instead, to make this meta-language useful in building such a system, a translation process is required to create an architecture from the reference model which can then in turn be implemented. In this process, abstract OAIS concepts have to be translated into concrete system elements and processes tailored to work in a specific environment.

It is for this reason that to speak of an OAIS implementation is misleading. While this may seem quibbling over details, it is important to understand that the OAIS reference model will not translate into a real-world system seamlessly, and that this has an impact on the notion of OAIS compliance as put forward in the model, and as interpreted or translated by a given archive or preservation service provider.

### A mythical creature: OAIS compliance

Because of the OAIS model’s abstract character, the notion of OAIS compliance is, as has been pointed out repeatedly, “necessarily vague” (Lavoie, 2004, p. 17). To comply with the OAIS model means complying with a set of very abstract requirements which themselves need interpretation, translation, and concretization if they are to be useful.

Thus, the standard itself makes only two requirements for compliance:

- 1 “Support” (itself a rather vague notion) of the OAIS information model described in chapter 2.2, including among other things the concept of information packages composed of content information and accompanying metadata.

- 2 Fulfill the set of mandatory responsibilities described in chapter 3.1 of the standard (see CCSDS, 2012, p. 1-3). The latter (see box 1) are high-level requirements that, as Beedham et al. note, “it would be difficult for any functioning archive not to comply with” (2005, p. 10).

#### Box 1: OAIS Mandatory Responsibilities

The OAIS shall:

- Negotiate for and accept appropriate information from information Producers.
- Obtain sufficient control of the information provided to the level needed to ensure Long Term Preservation.
- Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided, thereby defining its Knowledge Base.
- Ensure that the information to be preserved is Independently Understandable to the Designated Community.
- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive, ensuring that it is never deleted unless allowed as part of an approved strategy. There should be no ad-hoc deletions.
- Make the preserved information available to the Designated Community and enable the information to be disseminated as copies of, or as traceable to, the original submitted Data Objects with evidence supporting its Authenticity. (CCSDS, 2012, p. 3-1)

Regardless of this vagueness, “OAIS compliance” has almost become a topos in digital preservation discourse, a label that is applied to repositories and their hosting institutions “to underscore [their] trustworthiness” (CCSDS, 2011, pp. 1-1). Yet this label remains largely meaningless without context and specification. Thus, for any organization or repository labeling itself as OAIS-compliant, it has to be clear what this is taken to mean – that is, how the vagueness of OAIS compliance has been translated into a concrete set of criteria in a given case. These criteria might be something so complex as those laid down in the above-mentioned ISO standard. But, as Lavoie explains, to be OAIS-compliant could also quite simply involve using “OAIS concepts, terminology, and the functional and information models” when building a digital archive or preservation system; or OAIS-compliance can be the result of a mapping process in which “the various components in the archival system [are matched with] the corresponding features of the reference model” (2004, p. 17). But OAIS compliance could also mean “explicit application of OAIS concepts, terminology, and the functional and information models” or “that the OAIS concepts and models are ‘recoverable’ from the implementation – in other words, it is possible to map, at least from a high-level perspective, the various components in the archival system to the corresponding features of the reference model” (Lavoie, 2004, p. 17).

We would argue that there are good reasons to include the OAIS functional model in compliance testing as suggested by Lavoie. Thus, in particular, it can be assumed that in order to fulfill the OAIS mandatory responsibilities, an archival information system also has to perform the functions described in the standard. Accordingly, it is the second approach described by Lavoie that the GESIS Data Archive adopted in testing OAIS compliance; it thus followed in

the steps of the UK Data Archive and the ICPSR (see Beedham et al., 2005; Vardigan & Whiteman, 2007).

**Mapping the GESIS Data Archive to OAIS**

The GESIS Data Archive was originally founded in 1960 at the University of Cologne as the Central Archive for Empirical Social Research (Zentralarchiv für empirische Sozialforschung), Europe's first data archive in the social sciences. In 1986, it became a member of the newly founded Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS). Since 2007, the Data Archive is one of five scientific departments of GESIS – Leibniz-Institute for the Social Sciences, Germany's biggest research-based social sciences infrastructure institution. It is also a member of CESSDA, the Council of European Social Science Data Archives, dedicated to improving standardized access to social science research data in Europe (see <http://www.CESSDA.org>).

Since its foundation, the GESIS Data Archive has undertaken continual responsibility for preserving social science research. Collecting primarily digital data from empirical social research, the Data Archive currently holds more than 5,100 studies equaling over 600,000 files. To consolidate and demonstrate its status as a trustworthy provider of preservation services, the Data Archive has embarked on a series of self-audit and certification activities. The first step, now almost completed, is the application for the Data Seal of Approval (<http://datasealofapproval.org/>). From these activities resulted a decision to test OAIS compliance by carrying out a mapping of the GESIS Data Archive to the OAIS reference model. The objectives of this mapping are the following:

- gain a more structured overview of workflows and preservation/ dissemination processes;
- identify and close possible gaps in these workflows and processes;
- introduce OAIS terminology and concepts to support communication within the Archive and with other organizations.

To achieve these goals, a mapping between the Archive and the OAIS functional model, as well as an application of the concepts

from the OAIS information model, are currently being carried out. In the following, we report briefly on the procedure and first results of our functional model mapping.

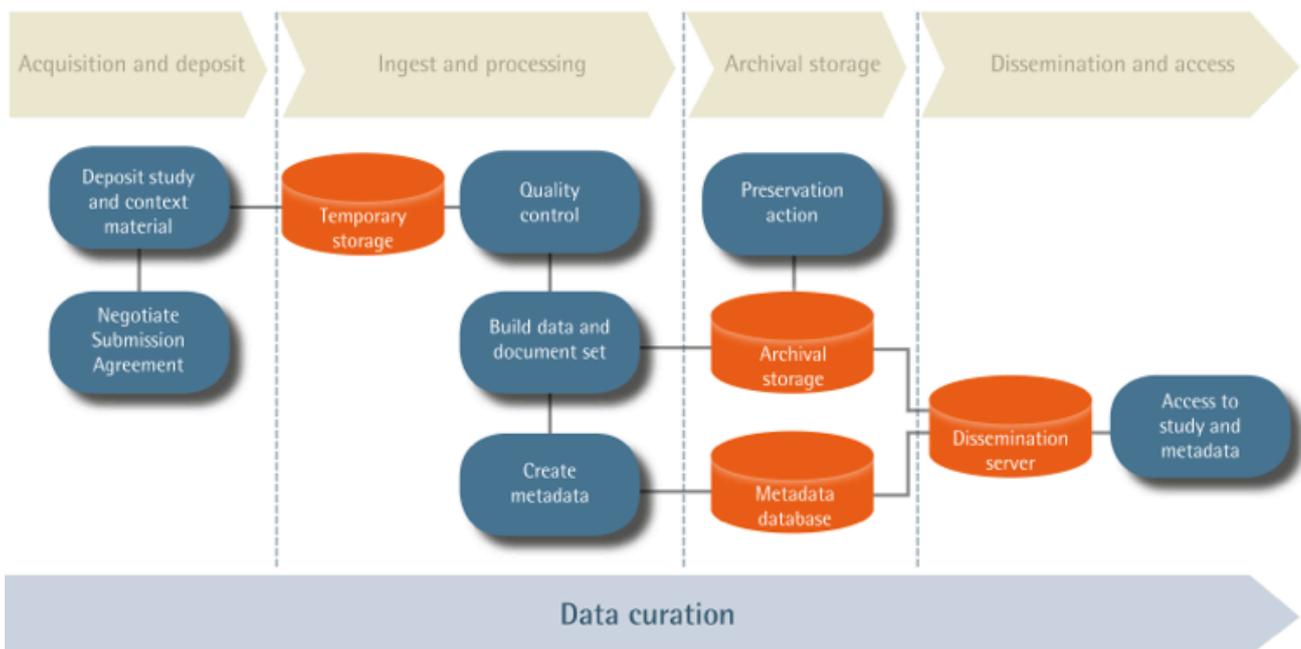
**Functional Model Mapping**

The main tool to carry out mapping was a simple spreadsheet listing OAIS functions and the different processes/responsibilities that these comprise. For each of these sub-processes we then determined the following:

- Who is responsible within GESIS (organizational unit down to team level)?
- Is the process carried out by a human staff member and/or is it supported by a technical system?
- How is the process incorporated into Archive workflows? (e.g. is it a routine activity carried out on a regular basis?)
- Are our activities sufficient?
- Any open questions or comments

At the same time we created a simplified diagram of the current archive workflow containing the main top-level functions performed as data are acquired, deposited, archived, and disseminated (see figure 1). This helped in creating a general overview of where and when processes were taking place, and to match these with the functional entities of the OAIS model. We then started increasing the granularity of the different sections of the overview diagram by spelling out the steps carried out in a given phase and by matching them to OAIS functions.

For the pre-ingest and ingest phase this resulted in the realization that the ingest process at the GESIS Data Archive (which as a social sciences archive puts a strong emphasis on extensive quality control, data processing and enhancement) cannot be adequately captured by OAIS in this form and detail (see also Vardigan & Whiteman, 2007). Thus, quality controls carried out during ingest, include: disclosure control; technical control of the files (format, readability, presence of malware, etc.); control of completeness; plausibility, consistency and weightings; as well



**Figure 1** GESIS Data Archive digital preservation workflow

as format conversions. If any problems are discovered, further communication with the data depositor may be necessary in order to clarify the discovered issues and to correct mistakes. Although this does not pose a problem for OAIS compliance, as the standard itself acknowledges that “[t]he complexity of this ingest process can vary greatly from OAIS to OAIS, or from Producer to Producer within an OAIS” (CCSDS, 2012, p. 4-52), it does complicate the mapping process.

It further became clear that some of the activities performed during ingest at the Data Archive are placed in different functional entities in the OAIS model. This caused us to “re-allocate” some of the functions to accommodate the actual Data Archive workflow<sup>4</sup>. As a consequence, our ingest comprises of functions from the OAIS functional entities ingest and administration among others, which are performed by several members of archive staff. It should be noted that this, too, is accounted for by the standard, which clearly states with regard to the functional model: “However, this is not to be taken as a recommended design or implementation, and actual implementations are not expected to have a one-to-one mapping to the functions shown, and may for example choose to combine functions or break out functionality differently” (CCSDS, 2012, p. 4-3). Yet, this makes the mapping less straightforward and hence more time-consuming.

**Benefits and Challenges**

We primarily benefited from the mapping in three areas: communication, self-reflection, and process evaluation. As noted in Beedham et al. (2005, p. 82), OAIS – with its clearly defined vocabulary – can support communication within, or between organizations, by offering a common language. Thus, one stated purpose of the OAIS standard is to provide the digital preservation community with a vocabulary composed of terms “that are not already overloaded with meaning so as to reduce conveying unintended meanings. Therefore it is expected that all disciplines and organizations will find that they need to map some of their more familiar terms to those of the OAIS Reference Model” (CCSDS, 2012, p. 1-5). This mapping process has begun at the GESIS Data Archive, and while the introduction of this new vocabulary and its establishment in everyday communication is a gradual process taking its time, OAIS terminology’s potential to help ensure that staff are speaking about the same things is already apparent<sup>5</sup>.

At the same time, mapping the elements and processes of the GESIS Data Archive to the OAIS functional model and vocabulary has fostered self-reflection. As mentioned above, the Data Archive has grown over decades, and while we are certain that our digital collections are in expert hands at the Archive, mapping to OAIS gives us the opportunity to – figuratively speaking – take a step back to analyze our daily routines and procedures. Looking at these routines through “OAIS glasses” and applying OAIS terminology to them, has helped us gain a more systematic understanding of the workflows that take place as we go about our daily work, the communication processes supporting them, and the roles and responsibilities of the actors involved.

Finally, undertaking the mapping helped us not only to identify and name processes, it also allowed us to evaluate them. Thus, we were able to spot gaps in our routines and to plan and implement strategies to close them.

As was to be expected, the mapping process was not without challenges. Leaving aside the difficulty that the scrutinizing and questioning of accustomed daily routines can pose for any established organization, some features of OAIS functional model itself contribute to making the mapping more difficult. While the following account is certainly not exhaustive (it does, for example, leave aside the problem of vagueness mentioned earlier), some of the problems we identified can be summarized under the headings ‘simplicity vs. complexity’ and ‘formalism vs. pragmatism’.

**Simplicity vs. Complexity**

As already discussed for the ingest functional entity, mapping the Data Archive to the OAIS functional model was complicated by the fact that often OAIS did not seem complex enough to model functions and processes performed by the Data Archive and its staff.

| Table 1. OAIS function “Establish Standards and Policies” – involvement of GESIS organizational units   |  |
|---|--|
| <p>Tasks and responsibilities (summarized from CCSDS, 2012, pp. 4-12-4-13):</p> <ul style="list-style-type: none"> <li>• establish and maintain the Archive (system) standards and policies.</li> <li>• receive budget information and policies from Management.</li> <li>• provide Management with periodic reports.</li> <li>• receive recommendations for Archive system enhancement, proposals for new Archive data standards, and periodic risk analysis reports from Preservation Planning.</li> <li>• face risks from unforeseen events; make appropriate decisions to minimize the risk of not fulfilling the Archive’s commitments.</li> <li>• receive performance information and Archive holding inventories from Manage System Configuration.</li> <li>• provide approved standards and migration goals to Preservation Planning.</li> <li>• develop storage management policies, including migration policies and database administration policies.</li> <li>• develop disaster recovery policies.</li> <li>• determine security policies for the contents of the Archive, including those affecting Physical Access Control and the application of error control techniques.</li> </ul> | <p>Fulfillment of these tasks requires activities by and communication between at least the following departments and teams</p> <ul style="list-style-type: none"> <li>• GESIS Data Archive: <ul style="list-style-type: none"> <li>○ Director and team leaders</li> <li>○ Team Acquisition, Preservation, Access</li> <li>○ Team Archival Tools and Processes</li> </ul> </li> <li>• GESIS president</li> <li>• GESIS administration</li> <li>• IT department</li> <li>• Department Knowledge Technologies for the Social Sciences</li> </ul> |

On the other end of that scale, we find OAIS functional entities and functions which seem relatively simple and straightforward, but turn out to be highly complex in mapping. Such unexpected complexity can occur particularly when single functions are performed jointly by several teams and/or departments. In this case, mapping the function entails documenting all the communication processes taking place between the actors and the systems involved in this process. As Beedham et al. observe for the data management function, this leads to “an ‘explosion’ of mappings to all the different systems and processes that an archive performs” (2005, p. 47). Another example of this is “Establish standards and policies,” which is part of the administration function, and which at the GESIS Data Archive cuts across different teams and departments (see table 1)

**Formalism vs. pragmatism**

As Beedham et al. observe with regard to the UK Data Archive and The National Archive,

“the OAIS standard can sometimes be overly bureaucratic and over-concerned with processes. Realistically organisations like UKDA have to be more pragmatic in their approach to decision making. . . The OAIS reference model only provides a formalised view of the functions of digital archiving; it does not prescribe implementation strategy or management style. Nevertheless, a real archival organisation never operates quite as ‘cleanly’ as the OAIS model envisages” (2005, p. 53).

The same is certainly true for the GESIS Data Archive, as problem-solving, planning, and decision-making processes can be less formalized than those described in the OAI functional model. Thus, we will often find that the Data Archive performs all the processes of which an OAI function is composed. However, many of these processes take place as part of routine, team (or department) internal communication, which may take many different forms (as well as degrees of formality) and which will not always be explicitly labeled as pertaining to a given OAI function. Thus, the fulfillment of a certain OAI function becomes a side-effect of certain communication processes.

Without question, the mapping process helped us identify possibly critical processes where we need to introduce more formality; for example, in the case of the Monitor Technology function. Greater formalization may also be needed in cases where functions, which according to OAI communicate with each other, are fulfilled by one and same person. As in this case no real communication (e.g. in form of requests and responses to these requests) takes place, the use of additional documentation tools (e.g. check lists) will have to be considered to create more transparency and to ensure that all necessary steps are taken.

However, as Beedham et al. point out, the OAI reference model only lists and describes functions without specifying how exactly they should be implemented (see 2005, p. 53). This means that it is really up to an archive to decide how frequently, and in which form, a process should take place. Does fulfilling the "Establish Standards and Policies" function require regular meetings (on team, department or institutional level)?; or, can certain processes be adequately addressed in ad-hoc communication between the staff members involved?

As already mentioned, what helped in assessing our current practice in the course of the mapping process, was to identify the level of "formality" with which a certain OAI function is fulfilled by the Data Archive. This was achieved, for example, by indicating

- whether a function is performed routinely (that is on a regular basis, or for every dataset submitted to the archive), or only on request/as required (and hence reactively rather than proactively; on this aspect see also Beedham et al., 2005, p. 53);
- whether checklists, minutes, or other records exist to document the process;
- and whether in our opinion the level of formality was sufficient or not.

However, in some instances the mapping may also result in a conscious decision to deviate from OAI entirely for reasons dictated by the specific environment in which the archive operates.

### Degrees of compliance

As the GESIS Data Archive's experience with the ongoing mapping process illustrate, OAI compliance – in contrast to compliance with a certification standard or criteria catalog such as ISO 16363:2012 – is not 1 or 0, Yes or No. Rather, we would argue, compliance comes in degrees.

Thus, what OAI compliance means is really a matter of interpretation, an act of filling in the gaps that the reference model necessarily leaves with context information from of our own organizations. Sometimes, this context may even lead to the decision to not comply with a certain aspect of OAI. We would

argue that such decisions, as long as they are well-founded, do not necessarily compromise OAI compliance. However, this illustrates once more that without providing enough of the context information specific to the archival system (which as a reference model OAI must necessarily ignore), and without spelling out and making transparent the acts of interpretation performed in translating the reference model into something like a checklist, the statement "We are OAI compliant" remains utterly meaningless.

### References

- Allinson, J., 2006. OAI as a reference model for repositories: an evaluation. Available at: <<http://www.ukoln.ac.uk/repositories/publications/oais-evaluation-200607/Drs-OAIS-evaluation-0.5.pdf>> [Accessed 20 August 2013].
- Beedham, H., Missen, J., Palmer, M. and Ruusalepp, R., 2005. Assessment of UKDA and TNA compliance with OAI and METS Standards. Colchester, Essex: UK Data Archive, University of Essex. Available at: <<http://www.jisc.ac.uk/media/documents/programmes/preservation/oaismets.pdf>> [Accessed 28 June 2013].
- CCSDS, 2011. Audit and Certification of Trustworthy Digital Repositories. Recommended Practice. Available at: <<http://public.ccsds.org/publications/archive/652x0m1.pdf>> [Accessed 28 June 2013].
- CCSDS, 2012. Reference Model for an Open Archival Information System (OAI). Recommended Practice. Available at: <<http://public.ccsds.org/publications/archive/650x0m2.pdf>> [Accessed 28 June 2013].
- Lavoie, B.F., 2004. The Open Archival Information System Reference Model: Introductory Guide. DPC Technology Watch Series Report 04-01. Available at: <[http://www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://www.dpconline.org/docs/lavoie_OAIS.pdf)> [Accessed 28 June 2013].
- Lee, C.A., 2010. Open Archival Information System (OAI) Reference Model. In M. J. Bates & M. N. Maack, eds. Encyclopedia of Library and Information Sciences. Taylor & Francis, pp. 4020–4030.
- Oßwald, A., 2010. Kapitel 4: Das Referenzmodell OAI - Open Archival Information System. In H. Neuroth et al., eds. nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3. Available at: <<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2010062438>>. [Accessed 20 August 2013]
- Vardigan, M. and Whiteman, C., 2007. ICPSR meets OAI: applying the OAI reference model to the social science archive context. *Archival Science*, 7(1), pp.73–87.

### NOTES

1. Natascha Schumann is affiliated at the Data Archive for the Social Sciences at the GESIS Leibniz Institute for the Social Sciences in Cologne. The main focus of her work is on digital curation of social science research data and audit and certification in this area. Her contact email is [natascha.schumann@gesis.org](mailto:natascha.schumann@gesis.org).
2. Dr. Astrid Recker works at the GESIS Data Archive where she is responsible for the design and delivery of digital preservation workshops for the "Archive and Data Management Training Center." She is also involved in the EU-funded project Data Service Infrastructure for the Social Sciences and Humanities (DASISH). Her contact email is [astrid.recker@gesis.org](mailto:astrid.recker@gesis.org).
3. This paper is an updated version of a presentation given at the IASSIST 2013 conference in the session "Beyond Bits and Bytes: The Organizational Dimension of Digital Preservation."
4. A similar observation is made in Beedham et al. for the administration function (2005, p. 48).
5. It is similarly clear, however, that OAI terminology will not become the only vocabulary with which the Data Archive operates. For example, the communication with stakeholders – data producers

and users in particular – requires the use of a different, less “technical” vocabulary. Similarly, internal communication, too, will continue to use non-OAIS terms and concepts where we regard these as (more) appropriate. Yet, it is important to be able to use OAIS terminology as a point of reference in cases of ambiguity or unclarity.