

CASPAR Validation

- slide 1** Although it is quite simple to claim to be doing valid work for digital preservation, it is not so easy to collect evidence of it, the CASPAR workgroup even decided to develop a way to validate the approach taken to preserve digital objects. This presentation describes this assessment approach, the metrics which have been set and several types of evidence that have been collected by using these metrics to validate what CASPAR is doing for digital preservation. Any project involved in this issue should be able to use the same metrics to provide similar kinds of evidence.
- slide 2** The metrics being proposed cover a very wide range of materials. A first group of metrics is appointed to demonstrate a sound theoretical basis for the approach taken; it should be based on the mainstream thinking of digital preservation so probably, it would be a reasonable metric for any project to adopt. The second broad group of metrics is useful to provide practical demonstrations by means of 'accelerated lifetime' tests. A similar kind of test is done in the commercial world for all types of pieces of equipment, e.g. hitting the objects a lot of times and so showing that it survives. What we have to do in the preservation environment is to pretend that time passes to simulate changes in software, hardware, environment and also in the Knowledge Base of the Designated Community; all of which can be discussed in terms of testbeds. A third metric has recently become possible and is to show that a repository gains in trustworthiness by using the techniques that are proposed. In other words, a draft of a standard has been prepared and is going through the review process; on the metrics proposed in this standard it is possible to build a certification process checking that all the expected requirements have been met.
- slide 3** A fourth group of metrics is perhaps only specific to CASPAR; it concerns the methodology and includes metrics that regard framework evaluation and testbed evaluation. Finally, there are some metrics concerning virtualization which is fairly generally agreed to be a good thing for preservation since it makes independent from the layers below, in information terms. These techniques should be tested across a number of disciplines in order to demonstrate the ability to deal with digital rights management, authenticity and the many different sorts of Preservation Description Information (PDI) components that are needed according to OAIS.

So, again, the basic compatibility with OAIS is what these metrics are trying to demonstrate.

- slide 4 In addition, as any other project, CASPAR is interested in showing the impact that it has on digital preservation.

Three metrics have been developed for this purpose: one to show its contribution to standards; a second to show that the techniques are being adopted and that there is a better awareness of the problems and the techniques; and finally, a third one, to show the effectiveness of the dissemination process.
- slide 5 The CASPAR approach, for the main part, revolves around capturing Representation Information (RepInfo) and RepInfo Networks.

(Separate presentations deal with OAIS, with RepInfo and with the tools that can be used to capture it.)
- slide 6 Much should be said about RepInfo and Archival Information Packages (AIP),
- slide 7 about where to get the materials,
- slide 8 and about preservation analyses and preservation workflows.
- slide 9 All of these are illustrated in the detailed testbeds presentations in the training session.
- slide 10 This process of validation has demonstrated that a certain quantity of evidence is contained in the CASPAR approach.

Firstly, the OAIS compatibility, which means the application of the information model which consists of the RepInfo, packaging description and so on.

Another side of OAIS conformance has to do with the OAIS responsibilities which are listed in the standard. Looking at these, most of them were outside of the scope of the project because they concerned other activities of repositories, but CASPAR certainly does help with issues like the definition of Designated Community and with authenticity that are two key parts of the responsibilities of repositories.
- slide 11 CASPAR theoretical underpinning consists of many papers submitted to peer-review journals, where, in particular, a lot of work has been done on the role of knowledge management in preservation.

Although it must be admitted that CAPSAR does not give a full theoretical treatment of preservation, there are many of the needed components and the general idea of how to take it forward.
- slide 12 Concerning trustability, existing archives, either within the CASPAR consortium or very closely related, have been looking at this in detail and it is clear that

almost all the archives have neither adequate RepInfo, nor the logical AIP which contains all of the information that is needed for the preservation of an object, both of which will certainly be improved by the use of CASPAR. So an archive will almost automatically achieve increased trustworthiness from the use CASPAR.

- slide 13** In terms of virtualization, there are separate presentations in the training part concerning digital rights, its virtualization and preservation, in addition to those regarding authenticity and provenance.
However, it must be said that the aims of the CASPAR project in terms of persistent identifiers have not actually been achieved.
- slide 14** In terms of impacts, the CASPAR project has had significant input in at least three areas: the treatment of authenticity, which has been very much improved; significant properties, which had not been dealt with in OAIS where, for various reasons, they are called Transformational Information Properties, and are dealt with in a very coherent way here; digital rights (Separate presentations concerning all of these aspects are contained in the training part).
CASPAR also had significant input to the preparation of the audit and certification standard for digital repositories.
- slide 15** Other impacts regard the standards in the commercial world for the Storage Networking Industry Association (SNIA).
Some implementations come from IBM, and also an industry standard, the Self-contained Information Retention Format (SIRF), which is essentially an implementation of AIPs, but at an industry storage level.
- slide 16** The evidence of adoption and dissemination are available in the CASPAR deliverables; moreover a number of details and dissemination materials are available on the CASPAR website and related links to it.
- slide 17** The bulk of the work has been done in terms of 'accelerated lifetime' tests.
During the period of the project, significant data sets and threats to them have been identified. These data sets are complex digital objects which are made of many components that need to be preserved.
Some real life scenarios, of course in accelerated time frame, have also been identified where there are real threats to changes in hardware, software, environment and Knowledge Base of the Designated Community. The researchers have run through these scenarios showing where CASPAR can contribute to their preservation despite all of these changes.
- slide 18** Evidence that CASPAR has allowed digital objects to be preserved has just been shown. The down side of this is that really complex examples have been chosen, so it is quite difficult to tell their story; for each of them in their separate presentations and testbeds, there may be a great deal of background

information to communicate, terminology that needs to be understood just to comprehend the issues, the description of the scenario and of which objectives can be reached by using the CASPAR approach. Then, for each of the testbeds there will be conclusions where, in many cases, the people responsible for maintaining that digital object will say why they think what has been done or what can be done using the CASPAR approach really contributes to their confidence about being able to preserve these complex digital objects.