

OASIS ANNUAL REPORT, 2004-2005

Steering Committee: Kate Bowers, University Archives; Leslie Morris, Houghton Library (chair); Susan von Salis, Harvard University Art Museums; Julie Wetherill, Lee Mandell (through May), and Michael Vandermillen (from June), OIS; Susan Lee, ULC Liaison.

Overview: A publicly available Harvard catalog since July 1998, OASIS provides centralized access to a growing percentage of finding aids for collections at Harvard. These finding aids are detailed descriptions of collections that contain a wide variety of materials, including letters, diaries, photographs, drawings, printed material, and objects. The search engine behind OASIS is TED/Tamino, an XML database. The OASIS Steering Committee meets every other month, alternating with the OASIS Users Group, which also meets every other month.

OASIS began with seven participating repositories; this year membership reached 20 (up from 18 last year), with the addition of the Harvard University Art Museums and Harvard Yenching Library. Baker Library resumed submitting finding aids after a long hiatus. Still in the very early stages of OASIS preparation are the Ukrainian Research Institute and the Milman Parry Collection of Oral Literature archive (a LDI project).

As of July 1, OASIS included 2,635 finding aids, up from 2,255 in June 2004, a 6% increase. This more modest growth rate can be attributed to the end of Schlesinger Library's recon (phase 1), and the near-completion of the Houghton Library's recon project. We expect contributions to increase significantly next year, as Baker Library begins recon, and Schlesinger Library undertakes phase 2.

Since finding aids vary greatly in length, and level of description, the raw number of finding aids in the system has never satisfactorily conveyed the scope and size of the database. The new OASIS provides some new types of statistics, which give a different angle of view:

Number of components within the container listing section of finding aids (corresponding to a folder or item)	715,000+
Number of name elements marked for indexing	652,000+
Number of date elements marked for indexing	516,000+

How is the database used? OASIS v.2 only provides a very broad picture, as in the past, these figures must be used with caution. There is still a pressing need for reporting that will help us see more clearly our users' search strategies; which finding aids are used most often; the proportion of Harvard and non-Harvard users; etc.

Total searches (quick search and index search)	41,838 (since 26 January 2005) =299 per day
Total browse requests	16,514 = 118 per day
Total finding aids displayed	24,709 =176 per day

Developments in 2004-2005

System reimplementation. The long-anticipated reimplementation of OASIS began in June 2004, when the OASIS User Interface Working Group began meeting with designer Roz Ault on the general “look and feel” of the new system. The overall design principle was to make it similar to the interface of VIA, since much functionality is shared. We also hoped this similarity would make it easier for users moving between the two systems. The camels, however, are unique to OASIS.

New features in the OASIS interface include the ability to:

- refine a search by searching within a results set
- browse long finding aid lists more quickly using an alphabet bar
- browse all OASIS finding aids in one list, regardless of repository
- see, re-execute, or combine previous searches in Search History
- download and print finding aids in PDF format
- create portfolios of multiple finding aids

The new OASIS catalog is XML (not SGML) based. The public interface is based on TED (the generic catalog for XML-based information also used for BIL and VIA). Search results created in TED are then delivered through a java application that includes some extensions that allow highlighted hits-in-context, portfolios, and other customized styling. The target date for the release of the public interface was mid-December 2004; the actual release was a month late, on 24 January 2005.

OASIS phase 2, the maintenance interface, began immediately after the release of the public interface. Phase 2 included redaction (the ability to remove from public display sections of a finding aid, for example, the names of students involved in disciplinary cases while those files are closed), real-time preview (so that processors can see what the finding aid looks like before it “goes live”), automatic naming (assignment of URN) and archiving of the finding aid (no more need to use NRS and DRS), and other enhancements to make the process of submitting finding aids easier and smoother. Phase 2 was completed in April 2005.

Reception of the new OASIS has been positive. Its major drawbacks are reflected in the list of remaining bugs appended to this report: some searches are VERY SLOW. Limiting by date is so slow as to be unusable, and we had to remove it as an option. The root of the problem seems to be that Tamino does not use the indexes when performing "near" and "not" searches if there are intervening words; nor does it use the index in "name" searches, because "name" can occur at any level. Instead, in these cases it searches the entire database. We are promised that the speed problem will be significantly reduced in the next Tamino release.

Ad hoc committees. During pre-release testing the Steering Committee found that few individuals were able to find the time to work intensively with the system, and so anticipated that once it was publicly released (and therefore the only option available) there might need to be some revisions. This proved to be the case, and immediately following the release of the public interface in January, two committees were formed: an Ad Hoc Committee on the Questions/Comments Form, charged with examining its content and design to make it more useful for repository statistics-gathering; and the Ad hoc Committee on OASIS System Help Page, charged to look closely at the Help pages (which were compiled somewhat hastily). Both committees reported promptly, and their recommendations are incorporated into the enhancements list.

Authoring software. It was planned that OASIS v.2 would accept only XML EAD finding aids. Many OASIS contributors have been using WordPerfect 8, which supports SGML but not XML. XMetal, also used by some OASIS contributors, does support XML. It is good for creating finding aids from scratch, but awkward to use for training and recon, and cannot easily be used for print products such as folder labels. So Houghton, the major WordPerfect user, investigated whether WordPerfect 12 (which does support XML) could replace WordPerfect 8. It has been a frustrating and time-consuming process, a minefield of undocumented quirks and features that have absorbed many, many hours of time of the Steering Committee, Houghton, and HCL ITS. **HOLLIS has Aleph, and VIA has OLIVIA, for the creating of bibliographic data, both supported by OIS; OASIS relies on commercial software that is not designed for authoring finding aids.** Fortunately, the Mellon Foundation has funded a project called the Archivist's Toolkit, a suite of tools that will include EAD authoring software. The OASIS Steering Committee will be following developments closely—helped by the fact that Lee Mandell, formerly the OIS OASIS Liaison, is now the Lead Developer for the project. It is our heartfelt hope that the Toolkit will be ready soon, and that it will be possible for OIS to support it for OASIS contributors.

EAD workshops. Training workshops, taught by Susan von Salis of HUAM and Kim Brookes of Simmons College, have been offered thirteen times

since 1998. Usually offered twice-yearly, only one workshop was offered this year because of the above-described problems with finding software to teach the workshop in XML. In the fall of 2004, 12 individuals were trained, bringing the total to 167.

OASIS Users Group. An interest group open to anyone interested in OASIS, whether as a participant or as a user, the Users Group was quite busy this year, being involved in reviewing the design and functionality of OASIS v.2 at various stages of reimplementation, as well as commenting on the work of the two Ad Hoc committees, and seeing demos of both WordPerfect 12 and XMeta1 4 as EAD authoring tools.

Steering Committee priorities for the coming year

Updating the OASIS Support Site. The reimplementation of OASIS makes this a high priority, as the new finding aid submission tools change each repositories workflow, and new authoring tools will need new “tips and tricks” sections. The number 1 enhancement priority for FY06, a migration to EAD 2002, will also require changes in the Harvard Guidelines, as some elements are deprecated, and new elements added.

Submitted July 6, 2005

OASIS enhancement request, June 2005

We have divided the list into three categories: development priorities, enhancement priorities, and bug fix priorities. While we believe that all the “bugs” require work by Tamino staff rather than OIS staff, resolution will require that OIS expend time and effort to make sure the bugs are indeed fixed. Many of the bugs represent extremely important functionality issues, and we do not want them to be overlooked.

Tamino development priority

Ability to search for 2 separate strings within one tag, e.g. “Henry” and “James” within the same <persname> tag, or “organization” and “women” within the same <corpname> tag. With more than 650,000 names now in OASIS, this would vastly improve search results.

Enhancement priorities

1. Upgrade OASIS to use EAD 2002.
2. When a user of OASIS submits a query via “Questions and Comments,” the system should send a confirming e-mail to the user that incorporates the text of their query.
3. User interface enhancements, as follows:
 - a. Incorporate the URN for the finding aid into the full finding aid display, so that it is apparent what link needs to be used if the f.a. is linked to from other sites.
 - b. Move the “Questions and comments”, copyright, and search links from the bottom of the title frame to the top right corner in the single f.a. view, so that they do not scroll off.
 - c. Make the OASIS shield a clickable link to OASIS home.
 - d. Change to compliant labels for banner images, so that they display in all browsers.
4. Add to Search the ability to search “not repository”, e.g. Unitarian NOT Divinity School Library.
5. Add ability to limit searches to finding aids with linked images and/or sound. We believe this could be done by a parameter limiting to <extref> within <unitid> or <unittitle> and <dao> <daoloc> and <daogrp> anywhere.
6. Further improvements to the PDF formatting.
7. Add ability to perform complex searches on result sets after having done a number of searches, e.g.:
 - Conduct a number of searches
 - Go to “search history”
 - Select two previously done searches and perform search operations on just the specific **result sets** returned (for example, I want to see everything that was a hit in search #1 and also a hit in search #3 but not if they contain the term “frigate”.)

Bug fix priorities

1. Tamino must support “round-tripping” UTF-8 encoded documents (i.e. UTF-8 in, get UTF-8 out).
2. Speed must be improved in:
 - a. faster display of large finding aids when displaying hits-in-context
 - b. faster searching

- c. particularly, faster proximity searching
- 3. Implement date range searching.
- 4. From "browse", ability to add a finding aid to a portfolio from full finding aid display.