OASIS ANNUAL REPORT, 2005–2006

Steering Committee: Kate Bowers, University Archives; Leslie Morris, Houghton Library (chair); Susan von Salis, Harvard University Art Museums; Michael Vandermillen, 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 was unchanged from last year, at 20. Changes of personnel required recertification at Law and Schlesinger, with Divinity in the process of recertification. The Ukrainian Research Institute, whose finding aids have been converted through Houghton’s LDI grant, have had to delay participation until a new Librarian is hired; the Fine Arts Library has expressed interest in participating, and has been assigned a buddy. Work with the Milman Parry Collection of Oral Literature archive (a LDI project) is in hiatus.

As of July 1, OASIS included 3,043 finding aids, up from 2,635 in June 2005, a 13% increase (up from 6% last year). With no major retrospective conversion projects underway this year, this may indicate the normal rate of growth to be expected. Contributions should increase significantly next year, as both Divinity and Baker undertake recon, and Schlesinger Library steps up its recon activity.

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. OASIS v. 2 provides some new types of statistics, which give a different angle of view:

<table>
<thead>
<tr>
<th></th>
<th>05–06</th>
<th>04–05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of components within the container listing section of finding aids (corresponding to a folder or item)</td>
<td>893,000+</td>
<td>715,000+</td>
</tr>
<tr>
<td>Number of name elements marked for indexing</td>
<td>919,000+</td>
<td>652,000+</td>
</tr>
<tr>
<td>Number of date elements marked for indexing</td>
<td>546,000+</td>
<td>516,000+</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>05–06</th>
<th>04–05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total searches (quick search and index search)</td>
<td>50,295</td>
<td>28,870 (since 26 January 2005) = 192 per day (number revised from last year)</td>
</tr>
<tr>
<td></td>
<td>138 per day</td>
<td></td>
</tr>
<tr>
<td></td>
<td>141 per day</td>
<td></td>
</tr>
<tr>
<td>Total browse requests</td>
<td>36,970</td>
<td>16,514</td>
</tr>
<tr>
<td></td>
<td>101 per day</td>
<td>118 per day</td>
</tr>
<tr>
<td>Total finding aids displayed</td>
<td>51,603</td>
<td>24,709</td>
</tr>
<tr>
<td></td>
<td>141 per day</td>
<td>176 per day</td>
</tr>
</tbody>
</table>

We anticipate that use of OASIS will increase now that it has been opened up to discovery by search engines such as Google.

**Developments in 2005–2006**

*System enhancements.* Not all of the work on OASIS planned for this year has been completed. The Steering Committee divided its enhancement list into two phases, in order to complete the more important and easier to implement changes first. Phase 1, completed with the release of OASIS v.2.2 on 26 June
2006, included some functional improvements, and a slightly different "look and feel" to the interface. Highlights include:

- OASIS is now open to discovery by search engines such as Google, providing much broader access to information about collections at Harvard.
- The persistent ID (URN) for each finding aid now appears in the header pane of the full finding aid display. This link will continue to work when the OASIS system itself changes again, and should be used if you want to link to a particular finding aid from another catalog, or from a web page.
- The "Questions or Comments" form has been modified to capture additional information useful to repositories in compiling user statistics, and has been integrated into the OIS problem-tracking system, Footprints. Users now also receive a confirming e-mail back when a comment is submitted; users now also receive a confirming e-mail back when a comment is submitted.
- Faster loading of "Easy print" (PDF) versions of finding aids.
- Migration of OASIS to EAD 2002. While invisible to our users, this important change brings the database into compliance with the most recent version of the EAD standard. This went relatively smoothly, with only 42 finding aids (ca. 1.5%) needing to be edited by hand to bring them into compliance.

Phase 2 of enhancements, just begun, includes implementing date-range searching/limiting, something much valued by our users, but focuses on performance tuning. OASIS is VERY SLOW, and this continues to be a major source of frustration to both outside users and Public Services staff. Large finding aids take a long time to load; the index searches are slower than Keyword; and “not” and “near” searches are so slow we feel obliged to warn users when they attempt them. Because the Steering Committee is unsure whether the planned “tweaking” planned for Phase 2 will significantly improve matters, we have included performance improvement on our priority list for next year.

**OASIS authoring Help site.** Glynn Edwards of Schlesinger Library took on the task of the first major update of the site in several years. The migration of OASIS to EAD 2002 will necessitate additional changes in the coming year, including the revision of the Harvard Encoding Guidelines.
**EAD workshops.** Training workshops, taught by Susan von Salis of HUAM and Kim Brookes of Simmons College, have been offered fifteen times since 1998. In August 2005, the first EAD in XML workshop was held, specifically for those switching from WordPerfect SGML to XML (12 trainees). A second workshop was held in February 2006 (11 trainees). This brings the total trained to 190.

**OASIS Users Group.** An interest group open to anyone interested in OASIS, whether as a participant or as a user, the Users Group reviewed documents for the updated OASIS authoring Help site, and provided valuable input on workflow and timing issues for the migration to EAD 2002.

**Steering Committee priorities for the coming year**

*Complete update of the OASIS authoring Help Site.* Migration to EAD 2002 requires changes in the Harvard Guidelines, as noted above.

*Investigate the functionality of the Archivist’s Toolkit.* As noted in past OASIS annual reports, the array of support issues surrounding the lack of a common EAD authoring tool used by Harvard repositories continues to absorb a disproportionate share of the Steering Committee’s time. While HOLLIS has Aleph, and VIA has OLIVIA, OASIS continues to rely on commercial software that is both expensive, and not designed for ease of use for EAD encoding. Additionally, as libraries and archives, and their user communities, gain sophistication in digital library applications, the need for more granular access to descriptive data in finding aids has become more apparent. The Archivists’s Toolkit, now in development with funding from the Mellon Foundation, offers a possible solution to both these problems, as well as potentially providing a common collections management tool for archival collections that might eliminate the need for locally maintained databases. For all these reasons (and because there is really nothing else out there that would solve even one of these problems), testing the Toolkit is a high priority.

*Submitted July 6, 2005*
OASIS enhancement requests, June 2006

ARC-01. OASIS management system. In order to provide more granular access to resources described in finding aids and to provide archivists with a simpler mechanism to create finding aids, this project will investigate the possibility of a database-driven finding aid management system, with a focus on an evaluation and participation in the ongoing effort in the archival community to create an open source Archivists's Toolkit, resulting in a set of functional and technical requirements and an implementation plan. This project includes assessment and evaluation, including installation and testing of beta releases of the Archivists Toolkit in multiple Harvard repositories to evaluate as possible OIS–supported metadata management system/creation software. An objective of this project is to develop a strategy and plan for deployment a new system if it appears warranted from testing.

Level of effort: Small/Medium

ARC-02. Add digital objects view to OASIS. Currently, when OCP needs to refer to specific subsets of archival material that have been digitized, they are forced to create an artificial page turned document that collectively points to the selected images. This is a problem since it displays the items completely out of the context of the complete finding aid. It is also a great deal of unnecessary work. This enhancement has been requested by OCP and approved by the OASIS steering committee. It will provide a new view within OASIS that can be linked to from the HOLLIS entry for a finding aid that will display the finding aid with only its digitized components visible, but also allow the full finding aid to be displayed as needed. This will eliminate the need for analytic records in HOLLIS for these digitized components and the creation of duplicate data in PDS.

Level of effort: Small

ARC-03. OASIS performance enhancement. OIS is currently investing effort within the Tamino XML database environment to improve the speed of certain searches in OASIS. There are three specific problem areas: (1) Not searches, which are important to reduce the size of a result set after an initial search turns up too many items, (2) Near searches, which are important for locating names when optionally containing middle names or initials, or in reverse order,
and (3) display time for large finding aids that have many highlighted search term hits within the document. Sometimes these operations can take 30 seconds to several minutes to complete. If new versions of Tamino to not resolve these problems, OIS proposes to experimentally migrate OASIS to use the XML database capabilities of Oracle. If that succeeds, we would replace the production OASIS and investigate the benefit of migrating the other OIS systems which rely on Tamino to Oracle as well VIA, TED, and Virtual Collections.

**Level of effort:** Small/Large (Large includes migrating all OIS systems away from Tamino)