

Survey of Digital Assets at the University of Florida

January 7, 2004

Submitted by the Task Force on Institutional Repositories

Stephanie C. Haas, chairperson, Members: Suzy Covey, Vernon Kisling, Priscilla Williams, Cathy Mook, Winston Harris, Peter McKay, and Carl Van Ness.

BACKGROUND COMMENTS ON UF DIGITAL ASSETS

ARTICLES/RESEARCH

With Open Access and scholarly communication the rallying point of many proponents of Institutional Repositories (IR), the definition of what contents should be placed in repositories remains open to debate. While many of the early initiatives, e.g., eprints from the University of Southampton, eScholarship at University of California, Dspace at MIT, focused on the pre and post print versions of journal articles, the overriding difficulties remain faculty buy-in to the concept, the implications for peer-review and associated tenure issues, and the copyright issues. A survey by Ware of 45 institutional repositories found that the average number of documents per repository was only 1,245. (Ware, M. 2004. "Universities' Own Electronic Repositories Yet to Impact Open Access." *Nature.com's Web Focus: Access to the Literature*. <http://www.nature.com/nature/focus/accessdebate/4.html>] The recent initiative by NIH to mandate deposition of articles in PubMed will provide a tremendously powerful incentive to one of the nation's largest scientific communities to participate in repository building. Similar initiatives are occurring in Norway, Denmark, and the UK, where Wellcome Trust (an independent biomedical research funding charity which currently spends over £400 million each year) announced a new policy whereby "Wellcome Trust grantees will be required to deposit an electronic version of their peer reviewed research articles in PubMed Central (or the European PMC, once established) no later than six months after the date of publication." It is the belief of the Task Force that leverage tied to research funding will have a far greater impact on participation than other incentives. If all major funding agencies mandated deposition, it is likely that major repositories could be developed rapidly with critical mass of discipline-related documents. Susan Gibbons, University of Rochester, (IR conference) indicated that faculty felt much more connected to their disciplinary colleagues than their institution which would also support the idea of subject based repositories, e.g., PubMed, rather than institutional ones. Many of the speakers at the SPARC/IR conference mentioned that no individual IR will have enough material on any one subject to be of research value; only by synchronizing the building of subject content can a viable open access research collection be created. Because of the many issues associated with this genre of materials, the Task Force believes trends should be closely monitored, but these materials should not be the first to be addressed in IR planning.

Beyond articles, many institutions have used theses, dissertations, honor papers, and technical reports as the core for establishing repositories.

PUBLICATIONS

The University of Florida already has implemented a procedure for the submittal, storing and access to its theses and dissertations. Similarly, the Digital Library Center in collaboration with other campus units has begun to digitize and make available through the PALMM initiative some of the major

technical report series of the colleges, departments, and centers of the University. Although series are being digitized as resources permit, key series that are underway include:

1. Bulletin of the Florida Museum of Natural History (v.1-15 completed);
2. Annual Reports, Bulletins, Circulars, and Miscellaneous Publications of the Florida Extension Service;
3. Annual Reports, Bulletins, and Press Bulletins of the Florida Experiment Station;
4. Bulletins, Leaflets, Technical Progress Reports, and Florida Engineering News from the Florida Engineering Experiment Station; and
5. Technical Reports of the Howard T. Odum Center for Wetlands.

Additionally, discussions are underway with IFAS to determine the most appropriate way to archive and provide access to the historical versions of the EDIS publications as they are updated.

JOURNALS

Also, there are some examples of journals that have migrated to electronic format. Several of these have been developed by and in association with UF faculty and are served off of the servers at the Florida Center for Library Automation. These include: Florida Entomologist, Journal of Nematology, Nematologia Mediterranea, Nematropica, and Proceedings of the Florida State Horticultural Society.

SERIES

With legacy trails in the print universe, it is likely that major series that have crossed into digital realms can be readily identified and appropriate procedures established to prevent their disappearance and assure their accessibility. What remains unknown about UF's digital assets then are non-major series, individual items, class related items, and other ephemeral digital objects created by units, faculty, staff, and students.

METHODOLOGY TO DETERMINE UF'S DIGITAL ASSETS

STRATEGIES

Aware of the magnitude of the digital universe, the Task Force decided not to include Law and Health in this preliminary study, except where centers are run by multiple colleges. Review of digital assets concentrated on non-major reoccurring digital objects created by university units at the university administration, college, departmental, and institute/center level. No attempt was made to inventory the digital contributions of individuals: faculty, staff, or students unless they were codified at a higher level. While there is no doubt that the curricular related materials developed during the instructional process are valuable, time constraints tabled their inclusion until a future time.

Two strategies were used to determine digital assets. The first was an online survey that was sent to the Dean's list. Seventy-five responses were received but that produced only 19 URLs to digital objects. It appears that list reaches a different constituency than the one creating Web sites within University units. There were many responses from health related units but none of these except the Student Health Center itemized any sites.

The second strategy involved each of the Task Force members reviewing a portion of the University's

digital presence. Again, the parameter for URL inclusions was some indication of reoccurring content, or pseudo-seriality, such as collections of monographic works, e.g., technical report series.

Task Force members were assigned web exploration tasks in relevant subject areas:

- Carl- university administrative units
- Peter-College of Business Administration & School of Accounting and departmental sites for Social Sciences: Anthropology, Political Science, Psychology, Sociology, Journalism
- Vernon-sciences, excluding agriculture
- Stephanie -College of Agriculture and its departments, a portion of the Centers, and the College of Health and Human Performance
- Priscilla-African & Asian Languages and Literatures Dept., Communication Sciences & Disorders Dept., English Dept., Philosophy Dept., and Romance Languages and Literatures Dept
- Cathy-Classics, Criminology, German and Slavic studies, History and Religion. Religion, Education, and she sampled the Department web sites from the College of Fine Arts.
- Stephanie, Suzy, and Winston- reviewed the Web sites maintained by the various Centers of the University

For each site, the following information was collected: unit name, title, URL, format, frequency, audience, notes.

Sites were reviewed from November 30 to December 14 and the results were integrated into an Excel spreadsheet, including the unique titles from the Dean’s list survey. A total of 313 sites were visited. Of these, 244 had identifiable titles that were of potential interest, either in terms of research or historical value, including documenting activities of the university community. The 69 other sites either had no content that could be identified or the unit itself could not be located. This happened with several of the centers.

SURVEY RESULTS

Results of the survey indicate that the Web is being rapidly assimilated into the administrative, research, and outreach efforts of the university community. To provide a general context for reviewing UF’s digital assets, some broad categorization of results is useful.

Table 1 indicates the units identified and the number of sites identified that seem to have research interest.

University unit/college	URL Count
University Administration	24
Agricultural and Life Science	100
Business Administration	23
Design, Construction, and Planning	9
Education	5
Engineering	26
Florida Museum of Natural History	8
Journalism and Communication	11
Liberal Arts & Sciences	40

Other categorizations that establish an overview of the digital assets are genre (Table 2) and format (Table 3). The genre field shows the types of materials being created and served on the Web. Because the Task Force was looking for pseudo-serial types of digital assets, a cursory review of results differentiated the following types: annual reports; newsletters; technical series that could include report series, identification keys, lectures, journal articles, etc.; administrative documents; journals; image collections; and publicity series. While pdf and html formats dominated, Table 3 indicates counts for all formats identified. It should be noted that for some objects, multiple formats are available, e.g., pdf, html or pdf, Flash

Table 2. Number of genre types identified.

Genre	Count
Administrative documents	23
Annual reports	9
Databases	3
Image collections	2
Journals	8
Newsletters	76
Publicity series	5
Technical series	105

Table 3. Formats used

Format	Instances
html	76
asp	3
pdf	140
wmv	2
ppt	3
mov	1
flash	1
jpg	1

Predominate Genre: Technical series, Newsletters, Administrative documents, and Journals

As noted above, technical series, newsletters and administrative documents make up the greatest proportion of documents on the web. Table 4 provides analysis of these genre.

Table 4.		
Genre (Number of sites)	College	Notes
Technical series (105)	College of Agricultural and Life Sciences (19)	Format: html and pdf Lectures in audio formats: .wmv and .mov ALEPH cataloging: Largely missing for series title and/or individual titles, e.g., neither <i>Publications on Economic Implications for Florida of the Terrorist Attacks in New York and Washington DC, Sept. 11, 2001</i> nor the eight papers of this series were in the catalog.
	Liberal Arts & Sciences (19)	
	Business Administration (11)	
	Engineering (11)	
	Education (5)	
	Design, Construction and Planning (4)	
	All other units (36)	

Newsletters (76)	Agriculture and Life Sciences (37)	Format: html and pdf; 1-jpg and 1-asp; some are in multiple formats. ALEPH cataloging: 65 have not been cataloged in ALEPH, 7 have records for the print version only, and 4 have cataloging for the electronic version.
	Liberal Arts & Sciences (14)	
	Engineering (8)	
	Florida Museum of Nat. Hist.(6)	
	Journalism (4)	
	All others (7)	
Journals (8)	Liberal Arts & Sciences (6)	Format: html and pdf; 1-flash ALEPH catalog: 6 of 8 titles
	All others (2)	

Technical series include a diverse set of materials. They may be technical reports, bulletins, briefs, circulars, lectures, data sets, monographic series, position papers, proceedings, general collections, presentations, fact sheets, project papers, etc. The content of these series indicates that they have been created for more serious business and research/educational endeavors. Audiences named include policy makers, industry groups, federal agencies, researchers, regulators, academics, and informed public. Many of these publications are issued on an irregular basis.

Newsletter frequency ranges from once a year to bi-weekly, and issues available online range from one to archives containing several years. The intended audience of newsletters was frequently the staff, students, and alumni of the given department; others were focused on providing timely information to the practitioners in the specified field; and still others appeared to be intended for general public education. This was particularly true of newsletters produced by IFAS extension, e.g., Family Nutrition Newsletter. Although archiving policies could not be determined, one surmises that archived issues indicate a value determination.

University administrative documents include budget documents, handbooks for faculty and students, program reviews, schedules of courses, etc. They document the official business of both the university and the departments. Of the titles, fourteen were in ALEPH: seven had entries for the print version only, and one has a record for the online version only of the University of Florida Student Guide.

Other series

The salient points concerning the other types of series are:

- All of the annual reports are in pdf format. Some sites have archives from previous years.
- Databases include the Archie Carr Sea Turtle database that is created in ProCite, massaged into a pseudo-MARC record, and served by FCLA. The Aquatic and Invasive Plants database is served off a dedicated server by the Center for Aquatic and Invasive Plants. DOCWEB is a database to the documents of Computing & Network Services.
- The two image collections identified are served in jpg formats. Copyright, use restrictions, and fees are known to apply to the images at the Center for Aquatic and Invasive Plants. Images at the North Florida Research and Education Center site appear freely available.

- Publicity objects are in html, pdf, wmv, and asp. Most appear to be news releases.

ANALYSIS BY COLLEGE

Four units account for 77% of the sites identified in this survey. They are the colleges of Agriculture and Life Science (100), Liberal Arts & Sciences (40), and Engineering(26). The fourth is a cluster of university administration units (24). A brief discussion of each of these units is given below. Data for all colleges/units is available on the Task Force web site at <http://www.uflib.ufl.edu/digital/Temporary/IR/Relateddocs.htm>. Each of the colleges has some unique sites. These are discussed briefly below.

Agricultural and Life Sciences

Because of the breadth of functions of this college, it represents a microcosm of the types of digital sites found throughout the university community with a predominance of technical and newsletter series. Some of the sites are extremely complex, e.g., keys to insect identification created by the Entomology and Nematology Department, and the equally complex Singing Insects of North America [<http://buzz.ifas.ufl.edu/>] that includes species information including sounds in wav file format. An additional aspect of the latter site is that images and sounds are collected from all over the world and the contributor retains the copyright for contributions.

Liberal Arts and Sciences

Of interest in the College of Liberal Arts and Sciences are their online journals and the audio clips associated with the children's' culture program entitled *Recess*. Technical series, newsletters and journals are the top three genre. Technical series include DNA sequences, sea turtle tag data, astronomical data (faculty copyrighted), and meeting reports.

College of Engineering

Video project descriptions are available through Engineering in wmv format. Again, technical series and newsletters are the major genre.

University Administrative Units

Administrative series are documents related to the operations of the institution including catalogs, student guides, miscellaneous athletic guides, publicity, journals, annual reports, etc. They are all in html/pdf formats. Many have archive files associated with them.

One set of documents that was previously mentioned was the honors papers. Although not a focus of this survey, honors papers are published in the Journal of Undergraduate Research which is available online at <http://web.clas.ufl.edu/CLAS/jur/>. All issues are available online and searchable with the UF Google interface.

DISCUSSION

This section relates the survey results to the Task Force objectives. It should be noted that certain objectives were purposely left undone because the labor needed was not available and because overriding policy decisions concerning the content of a UF institutional repository need to be established before further effort is warranted. Objectives of the Task Force are:

1. Review existing lists of intellectual products produced by all units of the University of Florida.

313 sites were reviewed yielding 244 identifiable titles of potential interest. The review purposely excluded the Health Center and Law units.

2. Determine what products are currently being published and in what format, i.e., print or digitally, and the digital format, e.g., jpg, tiff, pdf, html. Determine the physical extent of these products. Where necessary, units will be contacted to verify production. Identify a source for obtaining copies of the products identified. The information collected in this objective will be used by the Digital Library Center to determine costs for digitization and archiving.

A database was compiled listing academic unit, title, URL, format, frequency, intended audience, and notes. For the majority of URLs identified, digital formats could be identified. Although year ranges were recorded for many archived series, no attempt was made to verify the physical extent, i.e., file sizes of any of the archived titles. Unencrypted PDF files will offer few challenges to capture and archiving, but the HTML files with their many external links will pose intellectual and probably legal issues. Some preliminary work on file sizes of IFAS documents was done and will be discussed under Costs in the Conclusion section below.

3. Analyze intellectual production in terms of ingesting and archiving feasibility; in other words, what formats can be processed without major reengineering. Reference [UF ETD acceptable formats](#) and [FCLA Digital Archive acceptable formats](#). Where products are already being issued in electronic format, determine what programming and software will be needed to capture the objects for repository inclusion.

The list of formats and acceptability levels for the FCLA Digital Archives is given below. Although HTML files can be handled, the complexity of external linkages used to populate and provide functionality to these pages must be considered carefully prior to creating archival packages. According to Priscilla Caplan, the entire package will need to be harvested before submitting to FCLA for archiving.

Format	Instances	FCLA Digital Archive Acceptability Level
html	76	Acceptable Bit level preservation only
asp	3	

pdf	140 Embedded Fonts; no encryption	Encryption
wmv	2	Bit level preservation only
ppt	3	Bit level preservation only
mov	1	Bit level preservation only
flash	1	
jpg	1 JPEG/JFIF (*.jpg)	

3. Define potential audience(s) for these materials. The Task Force may consult with any library staff member or any faculty member or faculty group as necessary to determine potential uses and audiences of any products identified.

Audiences ranged from faculty, student, staff, and alumni to professionals in various disciplines.

4. Determine what if any copyright restrictions apply.

Copyright designations on UF web sites appear to vary greatly. Some sites bear the copyright symbol preceded by the departmental name, others use the copyright symbol followed by a year and the University of Florida. Some of the IFAS sites have no designation of any type; others have the following complex statement:

This document is copyrighted by the University of Florida, Institute of Food and Agricultural Sciences (UF/IFAS) for the people of the State of Florida. UF/IFAS retains all rights under all conventions, but permits free reproduction by all agents and offices of the Cooperative Extension Service and the people of the State of Florida. Permission is granted to others to use these materials in part or in full for educational purposes, provided that full credit is given to the UF/IFAS, citing the publication, its source, and date of publication.

It is apparent that if a title is selected for inclusion in the institutional repository that verification of copyright status will need to be confirmed. Since the intent of the repository will be to provide long-term access, it is believed that most departments will cooperate willingly with submittal/ingest of titles.

5. Based on its findings, the Task Force will recommend collection management policies that facilitate the building of the repository.

Task Force members suggested that collection management policies of UF digital assets cannot be developed without the input and cooperation of numerous units across campus. Selector input on digital content will be addressed in the concluding section.

CONCLUSIONS

As numerous studies and speakers have indicated, successful institutional repositories require sufficient human and technical support from all levels of an institution. While the Task Force has attempted to

canvas UF's digital assets, no member believes that this represents a comprehensive review. Nonetheless, it does permit us to set forth a fairly accurate view of the current balance sheet.

Summarily, it appears that the UF community is actively involved in presenting its research, news, outreach, and educational activities through Web publishing. The content developed by departmental efforts flavor each college's presence. Some departments have extremely well organized sites with archival runs of publications, e.g., Business Administration, Journalism, some IFAS departments, while others display little consistency.

In terms of catalog access to titles, when earlier print versions of series existed, records exist in ALEPH, but many electronic versions do not have records.

Copyright designations vary across the sites. When numerous, non-UF individuals/agencies have contributed to site development, the copyright issues become extremely complex. In some cases, copyright is attributed to an individual faculty member. Another factor in this arena is the economic models used to develop certain serial titles. Some departments sell their publications, both in print and downloadable files, as a means of generating income.

In terms of IR functions, the Task Force believes that both access and archival preservation are important. Access can be provided by UF through its own servers; whereas, archiving should be outsourced to FCLA. Different formats will be used for access vs. archiving.

The question of access/archiving actually requires a two node infrastructure. Access can be accomplished in a number of ways but ingesting digital documents and creating searchable metadata appear most foolproof against the lost link syndrome. True digital archiving with implied forward migrations, refreshing, etc. is available only by submitting documents to FCLA's digital archive. This is not a searchable archive, has a fee structure, and should be used for titles that are expected to have lasting value.

Currently, Greenstone open source software is being acquired to facilitate the future development of UF textual and graphic digital collections. This software can be used for institutional repository development as well. As soon as the new version is available, its IR functionality should be tested. An example of the ingesting mechanism of Greenstone can be seen at the Indian Institute of Science Publications Database <http://vidya-mapak.ncsi.iisc.ernet.in/cgi-bin/library> under "Add Publications." The Task Force believes that until the system infrastructure is in place and tested, content building and IR promotion should be limited.

It appears that many institutions have taken the approach of building a repository using Dspace, eprints, or some other software and then assuming that faculty will be interested in archiving digital titles. Since this does not appear to be a valid assumption, the Task Force recommends a more circumspect approach to the development of an institutional repository. As mentioned at the outset, UF already has IR content that is being created. (Please see *Background Comments*.) In the cases of both the electronic theses and dissertations and the series digitized by the Digital Library Center, the digital titles have

access and archiving strategies in place. The theses and dissertations are being served by FCLA and they will be archived in FCLA's digital archive, as will the UF titles already digitized by the DLC.

RECOMMENDATIONS FOR INITIATING UF's DIGITAL COLLECTION

Specific recommendations by the Task Force concerning the next steps in developing a UF repository are outlined below. Unless additional personnel is hired, selectors will bear the brunt of labor involved in building the content of the IR. This means that they must fully buy into the importance of developing UF's IR and must also understand the architecture and functioning of Greenstone so that they can explain it fully to departments and others.

Institutional Repository Architecture

- 1) Existing UF collections, e.g., electronic theses and dissertations, Florida Agriculture and Rural Life, and Florida Environments Online already constitute a nascent institutional repository, they simply have never been clustered formally as a UF digital collection/repository.
- 2) The loading, testing, and maintenance of Greenstone's institutional repository functions are paramount. Here the systems department and DLC director and metadata coordinator will play critical roles..
- 3) Concurrently, the metadata format for repository items should be developed jointly by cataloging and DLC.
- 4) Once testing is complete, procedural guidelines for self-archiving various types of digital files must be compiled, and all selectors trained.
- 5) If UF is desirous of having all of its digital assets searched through a central point, metadata for these existing materials could be harvested into the IR with links to the distributed full text files.

Content Development

While the Task Force recognizes the potential need for a centralized UF digital collection, the effort involved in building it requires campus wide support and input. While the library system may take a leadership role in its development, its success and sustainability depend on a critical perception by the university community as to its value, and at a more basic level, a belief in the long-term value of the Web publications they are creating.

As builders and promoters of this digital collection, CM selector participation becomes essential. The activities defined below are dependent on selector involvement:

- 1) Selectors need to take the sites already identified, complete the departmental web investigation, and enumerate titles of research potential. It should be noted that some of the Task Force members felt that the sites they reviewed were of

minimal research value and questioned the need for archiving in any sense of the word.

2) Using their compiled lists, selectors should directly contact the department to determine how it views the research value of the site. If concordance is high, a discussion of self archiving the site to UF's institutional repository and beyond to FCLA's digital archive should follow. The Task Force believes that departments must be involved in evaluating the research worth of their own publications.

3) As the selectors complete #1 above, the Task Force members suggest that three collecting priorities have already been established based on print materials: the administrative documents of the university, IFAS document series, and theses and dissertations. Similarly, the Digital Library Center should continue to pursue its current negotiation with IFAS to capture documents in the UF IR and archive them in the FCLA digital archives. Appropriate procedures for incorporating theses/dissertations into the UF digital collection/repository should also be explored.

Constituency Building and the Future

While the Task Force members understand the "glamour" associated with open access to faculty publications, the initial UF digital collection/repository and archiving activities should be directed toward official university documents and publications that have research/historical value but are not in the national spotlight. By focusing on access and archiving of administrative and departmental/unit digital assets, the value of the institutional collection/repository should be easy to establish and explain. As departmental contributions become established and monitored, efforts can be expanded to alert individual faculty members to the campus-wide IR initiative. Solicitation of curricular related materials might be considered as a next phase.

Two current leverages provide incentives for participating in repositories: administrative and legal mandates and monetary considerations, i.e., funding sources require copy deposition. At UF, state law and university mandates require the University to retain core documents that have permanent historical and administrative value. Traditionally, these have been deposited in the University Archives. Because the Task Force believes funding agency leverage will play a critical role in developing open access to journal articles, they advise a spectator role in this arena until the success of the NIH/PubMed initiative can be evaluated. Nonetheless, it was noted that some faculty are already making copies of their publications available on departmental and individual web sites. Technically, it is possible to ingest these documents, but systematic collection of journal articles raises compelling legal issues. One interesting angle is the *Minds of Carolina* project at the University of North Carolina that focuses on retiring faculty. "Faced with the question of what will happen to their scholarly contributions on retirement, retiring faculty members are receptive to the stewardship and preservation of an IR." –Library Technology Reports, July-August 2004, 40(4): 58. An additional strategy is to determine faculty with high research profiles (using citation counts) and recruit them as contributors.

Cost Factors

The annual operating cost (salaries, benefits, operating expenses, and business escrow) for MIT's Dspace repository is \$285,000, and the University of Rochester's repository is \$200,000. These costs do not include what may prove to be the most costly component: the future cost of preservation.

Although the Task Force realizes the danger of making estimates without sufficient data, the following section provides a draft of the staffing, equipment, and file storage requirements that might be needed, based on the recommendations above.

Staffing

Time	Title	Description
1 FTE	Program Developer	Coordinate all activities of the institutional repository, develop guides and manuals related to the institutional repository, train selectors, and as needed train departmental staff, set timelines, monitor submittals and use, and document development of repository
.25FTE	Greenstone Programmer	Develops functionality for institutional repository including ingest, normalization, metadata template and display
.5FTE	Normalization Tech	Responsible for taking ingested documents and converting them to acceptable formats for digital archiving, e.g., encrypted pdfs should be unencrypted before submittal to FCLA.
.5FTE	Metadata Cataloger	Reviews ingest template and records created from its use. Provide authority control for metadata.
.1 FTE (4hrs/wk) until review is finished; repeat review every two years	Selectors	Review departmental sites, reach consensus with departmental representatives on research value, alert program developer to titles to be ingested. Confirm ingestion.
.1FTE (4hrs/wk) until review is finished; repeat review as needed	University Archivist	Assigned to work with Dennis Kovac, UF's Record Manager to determine university titles that need permanent retention in the digital archive.

Equipment

Web servers should be purchased and maintained in the systems department. One for digital content storage and one developmental server to test system modifications without impacting access to content. Specifications have been developed by Bill Covey and Erich Kesse.

System backup either magnetic tapes, off-site storage, or redundant servers. Specifications need to be developed by Bill Covey.

Greenstone software is free ware, but may require additional software purchases to provide functionality desired. Confirm needed software with Erich Kesse and Bill Covey.

Storage file size

The first year will include the ingesting of 50 digital administrative unit titles (est. by C. Van Ness), the current 3,600 digital IFAS documents, and the first 150 digital titles identified by selectors.

Estimating file size is inexact and should be revalidated before actual budgeting occurs. For html files, all of the peripheral files that define a page content must be captured and this creates the potential for ingesting tremendously expanded files. The development team should be particularly aware of this and make appropriate policy decisions concerning the depth of capture. The chart below indicates approximate file sizes for various genre and formats to be ingested in the first year. The sizes for the IFAS documents are based on a sample of 30 titles. A sample of file sizes for administrative documents is also given. Estimates of storage cannot be given until all titles and formats have been determined.

Unit	Calculation Method/Title	Format	# proposed for ingest	Storage needed
IFAS	Based on 30 EDIS title sample	html=19 KB /title	3,600	67MB -does not include images and other information stored on peripheral pages
		pdf=719 KB /title	3,600	2,728MB

Digital Archiving

Because FCLA will be the first functional digital archives in the U.S., there is no way to estimate storage costs for archiving. It is generally understood that FCLA will provide storage for the first year for free until enough data is available to make reasonable cost projections. Once data is available a cost structure will be developed and then the UF Library can determine its policies of retention.