What is the Archives of Institutional Memory?

- The Archives of Institutional Memory is a digital repository for disseminating and preserving official Indiana University records with long-term, indefinite administrative, legal, fiscal or historical value.
What is an Institutional Repository?

- Clifford Lynch: “A university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.

- It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution.”
What is an Institutional Repository?

- A key part of the services that comprise an institutional repository is the management of technological changes and the migration of digital content from one set of technologies to the next as part of the organizational commitment to providing repository services.
Why create an Archives of Institutional Memory?

- AIM is designed to preserve and make available indefinitely authentic, digital records.
- This means that once records are put into AIM they are protected from unauthorized and undocumented alterations or deletions.
- It also means that over time managers of AIM will convert or migrate the records into new formats or software environments so that users can retrieve and use the documents.
- Finally, by placing records into AIM users can be assured that the records will be accessible and can be easily retrieved via an intuitive and functional user interface.
- In short, AIM is a repository designed to manage authentic, unchanging records that will be used for many purposes over long periods of time.
Why create an Archives of Institutional Memory?

- In part, AIM was created so that creators of records do not have to devote resources to and take responsibility for the long-term management and preservation of these important digital resources.
What is D-Space?

- DSpace is an open source software platform that enables organisations to:
- capture and describe digital material using a submission workflow module, or a variety of programmatic ingest options
- distribute an organisation's digital assets over the web through a search and retrieval system
- preserve digital assets over the long term
Who built DSpace?

- The MIT Libraries and Hewlett-Packard (HP) jointly developed DSpace.
- The system is now freely available to research institutions world-wide as an open source system that can be customized and extended.
Who manages DSpace?

- DSpace is freely available as open source software.
- The DSpace Community manages the code base and releases new versions of the software.
- An active community of developers, researchers and users worldwide contribute their expertise to the DSpace Community.
What kinds of DSpace services are other institutions building?

- Research institutions worldwide use DSpace to meet a variety of digital archiving needs:
- Institutional Repositories (IRs) for Research products
- Learning Object Repositories (LORs)
- eTheses
- Electronic Records Management (ERM)
- Digital Preservation
- Publishing
What type of content is included the Archives of Institutional Memory?

- AIM contains Indiana University records that have long-term, indefinite administrative, legal, fiscal or historical value. Records typically found in AIM include:
  - Mission critical documents such as IU Bulletins, Schedule of Classes, annual and strategic reports, and various types of planning documents
  - IU Policy documents
  - Publications of IU schools, departments and administrative offices
  - Publications created by official university groups - faculty, student or alumni - such as the Bloomington Faculty Council, the IU Student Association, and the Alumni Association
  - Records documenting IU events, such as commencement, founders' day, etc.
What other kinds of content might be included in AIM?

- Can also include at some point:
  - Images
  - Audio files
  - Video files
  - Web pages
How is the Archives of Institutional Memory different from the IUScholarWorks repository?

• IUScholarWorks is a digital repository for disseminating and preserving scholarly work created at Indiana University by its faculty, administrators and staff, and students. Examples of types of scholarly works that are appropriate for deposit in IUScholarWorks include publications, pre-publication scholarship, working papers, technical reports, supplementary research material not included in print journals, and conference papers.

• AIM, on the other hand, is designed to preserve and make available the official records with long-term value produced by the Indiana University administrative and academic community.

• In other words, AIM is designed to function as the repository for IU’s official, institutional records.
How is the Archives of Institutional Memory organized?

• At this point go to AIM Page and demonstrate

• AIM is organized into Communities (campuses, departments, research centers, or other groups) and into

• Collections (bulletins, policies, reports, newsletters or other types of official University records).
What metadata standards does DSpace support?

- Currently DSpace supports only the Dublin Core metadata element set with a few qualifications.
- But plans are underway to expand the metadata capabilities of the software.
How to does an office or department get their records into AIM?

- At this point, records can only be placed into AIM by the University Archivist, who is managing the repository.
- This will likely change in the future and a self-submitting strategy will be adopted.
How does one find an item in AIM?

- The organization of the AIM repository makes it easy to browse by community and by collection.
- Items can also be searched by keyword or phrase, or browsed by author, title, or date or other Dublin Core Metadata elements.
- Simple or advanced, boolean searches can be implemented.
- In addition, the metadata will be picked up by other similar search systems and by general search engines like Google.
Who can read the files in AIM?

- The default is open access to all deposited items for all users of the World Wide Web.
- However, if necessary and appropriate, access to records can be restricted to a defined community of users.
How are records preserved in AIM?

- For text:
  - PDF file
  - and a PDF/A file
PDF/A: Time Line for Part I

- October 2002 Initial meeting of AIIM/NPES PDF/A committee
- April 2003 Initial Working Draft (WD)
- August 2003 New Work Item (NWI) approved and
- Joint Working Group (JWG) formed
- December 2003 First Committee Draft (CD) approved
- September 2004 Second CD approved
- June 2005 Draft International Standard (DIS) unanimously approved
September 28, 2005 the International Standards Organization (ISO) approved a new Standard governing electronic document archiving:

The PDF/A standard

• PDF/A-2 is targeted for 2009 approval and release
• PDF/A-2 is based on PDF 1.7 / ISO 32000
ISO 19005-1 defines “a file format based on PDF, known as PDF/A, which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files.” (from ISO 19005-1).

The Standard does not define an archiving strategy or the goals of an archiving system. It identifies a “profile” for electronic documents that ensures the documents can be reproduced in years to come.
What is PDF/A?

- PDF/A *alone* does not guarantee preservation
- PDF/A *alone* does not guarantee exact replication of source material
- The intent of PDF/A is *not* to claim that PDF-based solutions are the best way to preserve electronic documents
- But once you have decided to use a PDF-based approach, PDF/A defines an archival profile of PDF that is *more* amenable to long-term preservation
Why PDF/A?

- The feature-rich nature of PDF can create difficulties in preserving information over the long-term, and some useful features of the PDF file format are incompatible with the demands of long-term preservation.
- For example, PDF documents are not necessarily self-contained, drawing on system fonts and other content stored external to the original file.
Why PDF/A?

- Additionally, because of the lack of standardization among the many PDF development tools on the market, there is inconsistency in the implementation of the file format.
- Tremendous quantities of valuable information are currently be created and saved all over the world as PDF, and a specification solution is needed to ensure that digital PDF documents remain readable, renderable and accessible for the long-term.
What long-term preservation needs does PDF/A-1 address?

- Characteristics identified as objectives for PDF/A were:
  1. Device Independent - Can be reliably and consistently rendered without regard to the hardware or software platform.
  2. Self-contained - Everything that is necessary to render or print a PDF/A-1 file must be contained within the file. All of the information necessary for displaying the document in the same manner every time is embedded in the file. This includes all visible content like text, raster images, vector graphics, fonts, color information and more.
What long-term preservation needs does PDF/A-1 address?

3. Self-documenting - Contains its own description - PDF/A-1 requires Adobe Extensible Metadata Platform (XMP) be used for embedding metadata in PDF files.

4. Unfettered - Absence of technical file protection mechanisms – For example, PDF/A-1 prohibits encryption and compression. This prohibition means that User IDs and/or Passwords are not needed to do anything with a PDF/A-1 file. PDF/A-1 files are open and available to anyone or any software that processes the file. Implementers that require access controls can provide these access controls outside of the file format.
What long-term preservation needs does PDF/A-1 address?

- 5. Available - PDF/A-1 is based on an authoritative specification that is publicly available.
- Anyone can use the PDF Reference and XMP Specification in conjunction with PDF/A-1 to create applications that read, write, or process PDF/A-1 files. Adobe has granted a general royalty free license to use certain of its patents to create applications that process PDF/A-1 files.
What long-term preservation needs does PDF/A-1 address?

6. Adoption - Widespread use may be the best deterrent against preservation risk - PDF/A-1 was designed for flexibility of implementation to promote its wide adoption. Market support of PDF/A will help ensure the viability of PDF/A and extend the length of time that PDF documents can be maintained as PDF/A.
PDF/A

- PDF/A-1 is further subdivided into two levels of compliance: PDF/A-1a and PDF/A-1b.
- PDF/A-1a (Level A Conformance) denotes full compliance with the currently approved PDF/A Standard ISO 19005-1: Part, including those related to structural and semantic properties of documents.
PDF/A-1a

- The standard recommends that creators "should attempt to capture a document's logical structure hierarchy to the finest granularity possible." Nevertheless, the standard also indicates that "PDF/A-1 writers should not add structural or semantic information that is not explicitly or implicitly present in the source material solely for the purpose of achieving conformance." Hence, the logical structure of a document will only be represented to the degree the creator or process during creation takes steps to incorporate relevant structural tagging.
PDF/A-1b

- There is also a “minimal compliance” level for PDF/A: PDF/A-1b (Level B Conformance). PDF/A-1b requirements are meant to ensure that the rendered visual appearance of the file is reproducible over the long-term but not necessarily exactly as the original.

- PDF/A-1b does not require representation of the logical structure of the document as specified in section 6.8 of ISO 19005-1 for PDF documents.