



# Media Preservation at Indiana University and the Scholarly Data Archive

Jon Dunn, Kurt Seiffert  
PASIG – January 12, 2012



INDIANA UNIVERSITY



# Outline

- IU Bloomington Media Preservation Initiative
- Requirements and strategy for preservation storage
- Scholarly Data Archive and cost model for bit storage



# Motivations for Media Preservation at IU

- Rich set of media resources
  - Jacobs School of Music
  - Ethnographic audio and video collections
  - Film archives
- History of work in digital media
  - Variations digital music library
  - EVIA Digital Archive
  - Sound Directions audio preservation best practices



## Media Preservation Survey (2009)

- Commissioned by Vice Provost for Research
  - IU Bloomington campus
- Results:
  - 560,000+ audio, video, and film objects
  - 50+ physical formats in over 80 organizational units
  - 180,000 objects at high risk of loss in <10 years
  - 44% unique or rare



# Solutions

- Digitization is only solution due to obsolescence
- Some digitization already going on
- At current pace, will be done in 120 years
  
- Remember:  
180,000 objects at high risk of loss in <10 years



# IUB Media Preservation Initiative

- Year-long planning process (2010-2011)
  - VP Research, Libraries, UITs, Arts and Sciences, others
  - AVPS as consultant
- Focused on planning for:
  - Preservation (technical approach, workflow)
  - Facility and staffing needs
  - Access
  - Technology infrastructure
  - Campus engagement



# Media Preservation Plan

- Preserve rare and unique items within 15 years
- Combination of insource and outsource
- Minimal descriptive metadata to support workflow; extensive technical metadata
- Develop prioritization plan
- Digitize film for access; audio/video for preservation
- Use and extend existing data storage, repository, and media delivery infrastructure
- Effective access important: *Variations on Video*
- Explore partnerships



INDIANA UNIVERSITY

# Media Preservation Plan

[www.indiana.edu/~medpres/](http://www.indiana.edu/~medpres/)







# Technical Requirements

- 39 petabytes over 15 years
  - Uncompressed SD video; 96/24 stereo audio
- 200-400TB interim storage to support workflow
- Scholarly Data Archive
  - Primarily tape-based HSM storage mirrored between Bloomington and Indianapolis
- Extend Fedora repository to support preservation
- Need to explore out-of-region storage options

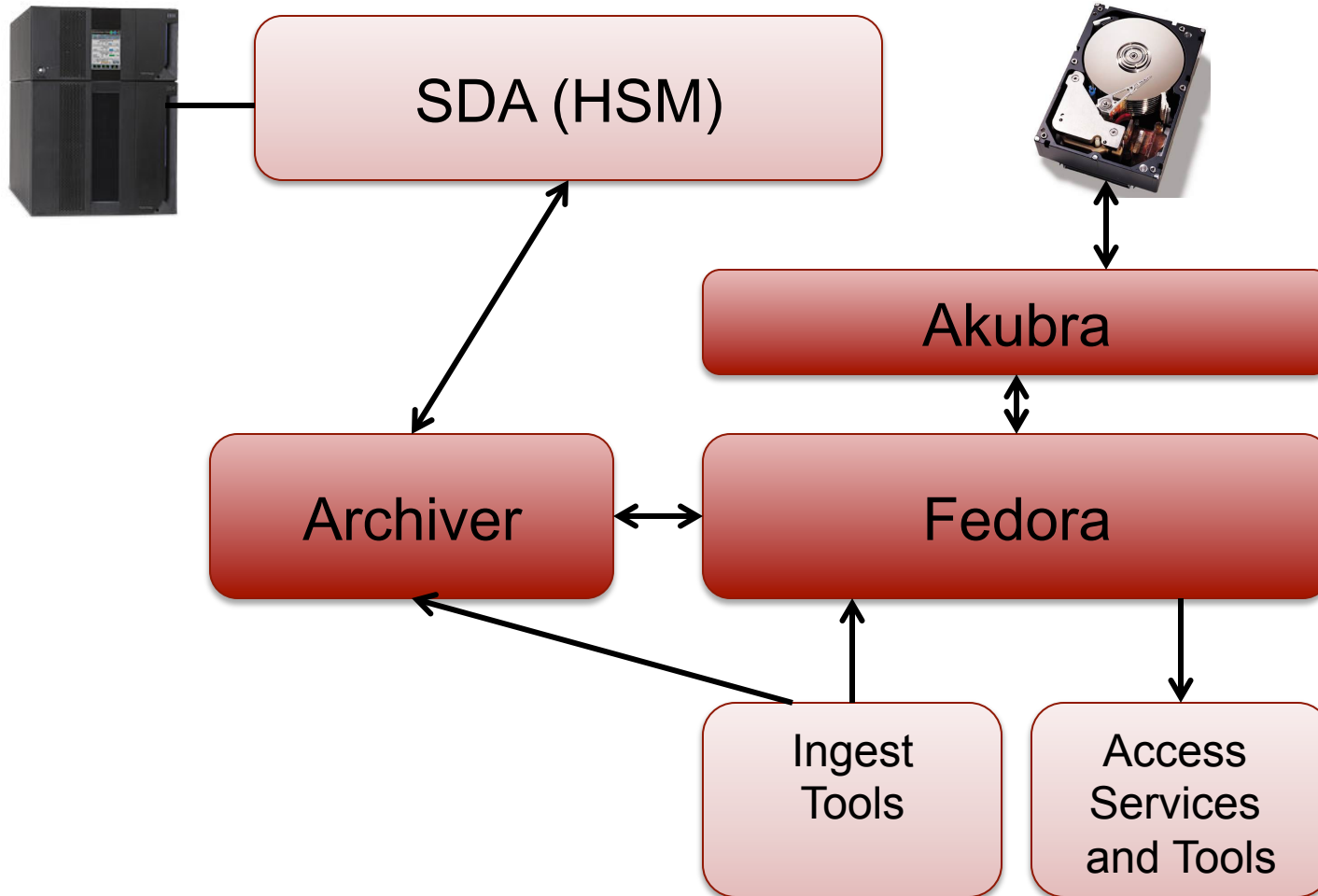


# IU Fedora Repository

- IU has been working with Fedora since 2002
  - Home for new digital collections
  - Migration of legacy collections
- Dual focus on access and preservation
- Hosted in “internal cloud” – IU Intelligent Infrastructure
- >700,000 objects; 700GB on disk; 15TB master files on tape
- How do we scale from 15TB to 39PB?



# Integration of SDA with Fedora





## Scholarly Data Archive

- Central storage resource for IU since 1999
- Data replicated between Indianapolis and Bloomington
- Implemented on HPSS
- 40PB IBM TS1140 based capacity, 15PB tape
- Available from desktop, web, high speed transfer protocols
- Currently holds ~6PB of data in 36 million files



INDIANA UNIVERSITY

# Economic Model for Archive Data





## SDA Funding Model

- Leverages expectations in data density growth of tape
- Working towards one time charge cost rather than annual costs
- Pricing includes replacing servers, tape drives, media, staff, licensing, and support
- Actively managed by migrating bits over time.



## More Information

- Media Preservation Initiative
  - <http://www.indiana.edu/~medpres/>
- Variations on Video
  - <http://variationsonvideo.org/>
- Scholarly Data Archive
  - <https://pti.iu.edu/storage/sda>