

# 2015-10-23 HydraDAM2 at IU Meeting notes

## Date

23 Oct 2015

## Attendees

- [Julie Hardesty](#)
- [Heidi Dowding](#)
- [Jon Dunn](#)
- [Will Cowan](#)
- [Brian Wheeler](#)
- [Randall Floyd](#)
- [Nianli Ma](#)

## Agenda

- Review previous meeting notes
  - corrections and links added - thanks!
- HPSS meeting (with Kristy Kallback-Rose and Danko Antolovic) - Heidi and Brian
  - next version will solve some of our issues but no date for migration to next version
  - next version has checksums going through entire workflow vs now (no checksum between things going from disk cache to tape)
  - drives HPSS purchased do check as things are being written to tape, but errors aren't passed up the stack to know it was part of a transfer
  - 3rd party products can do this work but have never been involved in IU's HPSS environment, other places use them
    - listens for SCSI commands going by
  - reader and writer will both be checked in new workflow (both sides of where HPSS gets a file and where it puts that file)
  - regularly scheduled fixity validation (suggested every 2-5 years if it can't be yearly)
  - derivatives - check fixity on these? could bargain with these to have all masters checked if derivatives are not
    - Brian running tests on MDPI data to do fixity checking and see how much bandwidth is required (estimate)
  - HPSS folks will work with us whatever we need
  - what do you do when you find an error? go to original source to get another copy
    - if there are 2 copies in HPSS (IUB and IUPUI), go to other copy, but requires manual intervention so file on IUB tape is marked unavailable
  - also assuming another copy is stored in non-HPSS system (DPN) - need for fixity checking goes down at that point, assuming fixity checking occurs regularly in non-HPSS system
  - migrating into new formats "relatively often"
  - network is real problem - storage silos everywhere but getting it from one place to another is a problem
- Additional discussion
  - HydraConnect meeting on digital preservation - Jon brought up idea of using Fedora/Hydra functions to manage checksum fixity checking; folks said that wasn't doing preservation because that is trusting a different system than where preserved items are located (same can be said for HPSS environment to certain degree)
    - can't do all fixity checking in place - takes too much bandwidth, have to farm that out
  - Survey of fixity checking practices - [NDSA storage survey 2012](#)
  - Fixity checking on regular schedule to start but hopefully don't need as much regular checking and can do random sampling or as files are pulled for use in future
  - At this point, have to be good HPSS citizens and not lock everyone out; also need to make sure SDA meets our needs and advocate for more funding if necessary
- Next steps / plan moving forward
  - January - Dark Avalon can take in HPSS data; will need a download master feature/tool of some kind
  - long term - HydraDAM2 comes in between HPSS and Dark Avalon and will also have that master download feature/tool
  - HydraDAM2 - manual fixity check calls, might need to be requests with reports you can get at later time
    - ingest to HydraDAM2 probably needs to take static XML and convert to RDF properties - need data model
  - Brian can get bag examples to Julie for documenting bag examples on wiki
  - what metadata are we mapping into RDF for HydraDAM2?
  - for items with IUCAT records - use catalog key to put into Avalon; otherwise POD info in MODS in bag has to be called up
  - **USE CASE [not in Func Req - HDM-234]:** HydraDAM2 - might need descriptive metadata versioning, ability to compare current IUCAT record with preserved MODS, but low priority
  - on-demand fixity checking based on events - are there standards for that? batch our fixity checking requests (looking on a single tape for certain criteria - have to do things somehow based on physical underlying system); need to gather up requests as they occur; if there is checksum question on master file, it needs to be taken out of circulation
    - [AHEYM 5th video file example](#) - truncated on upload to HPSS, that changed the file and checksum didn't match in IUB; no validation after upload; different than process we do now (earlier use of HPSS for Sound Directions)
  - when user goes through HydraDAM2 to retrieve master file, there should be fixity checking happening at that point
    - would need fixity check message stored somewhere
    - call to download master file will grab from HPSS and place in dropbox, fixity check happens then automatically through HPSS and transfer fails if fixity fails
    - would be up to the end user when downloading from dropbox to check the checksum at that point
  - see if there are additional use cases for HydraDAM2; create stories for them in backlog
  - see if any of those stories need to be reflected in Dark Avalon (separate group working on Dark Avalon now)
  - review HydraDAM2 functional requirements page and see if these use cases are included there already
  - might need meeting again this year depending on how HydraDAM2 work and Dark Avalon work goes

- Dark Avalon work being worked into current Avalon development in terms of more automated deposit into Avalon; will need a group for actual process of getting MDPI items into Dark Avalon

## Action items

- MDPI bag examples for audio formats to create wiki documentation - **Brian and Julie**
- Consider metadata to map into HydraDAM2 RDF from example MDPI bags - **Julie**
- Review [HydraDAM2 functional requirements](#) to see if use cases brought up in these meetings already included there - **Julie [DONE]**
- Add use cases not already covered by functional requirements to HydraDAM2 backlog - **Julie [DONE]**