

NDSA Fixity Working Lunch

November 10, 2016

This meeting was held during the NDSA Digital Preservation 2016 conference for interested parties to talk about fixity and the plans the Standards and Practices interest group and the Infrastructure interest group have in forming a Fixity Working Group. The following is a summary of that discussion.

A few slides were shown by Aaron Collie to get people up to speed on what the two Interest groups had discussed. [People had questions about normal activities surrounding the use of fixity checking and the groups discussed using the Fixity document to come up with a survey to send out. Also to document real use cases of when and why fixity was checked/used.]

Related Documents:

- Storage Report Findings: <http://www.dlib.org/dlib/may13/altman/05altman.html>
- National Agendas: Both had a call to action related to fixity.
<http://www.digitalpreservation.gov/documents/2014NationalAgenda.pdf> and
<http://www.digitalpreservation.gov/documents/2015NationalAgenda.pdf>
- Fixity: Checking your Digital content...
<http://ndsa.org/documents/NDSA-Fixity-Guidance-Report-final100214.pdf>

The Storage survey has some fixity related questions. If this is being sent out again, might want to see if we can use this data - and add to it if necessary. Comment about [Survey Document]: surveys are good, but the use cases will add value. Vote for adding use cases to final report.

Interest in:

- When is fixity generated and when is it verified? [thinking about as content moves around, within/between systems... where and how is fixity generated, checked and monitored.]
- BagIt was mentioned as a useful tool
- How can you use side car files...
- How can you communicate to vendors that fixity is something we want initially generated from them, and in what forms do we expect it in

Communicating to Vendors:

- Explaining what fixity is, why it is important, why we want it generated by you, and what forms we expect it to be in

- As a team, we need to share the specifications that we put into Statements of Work... so we can start to ask for the same things. Then they might become more mainstream and expected from us as a whole.

Lifecycle

- Fixity is not just a one time thing... it is not just something done when it gets to the final storage / preservation location. Fixity should be done along the entire lifecycle as we are working on the files. But what is enough for you?

Informal Poll of the room

- Most are doing fixity..... in the room... but not all...
 - ex: isilon device to store files and that monitors checksums and gives report once a month, and then get once a year and add or delete files if we need to...
[\[https://open-systems.ufl.edu/content/what-isilon\]](https://open-systems.ufl.edu/content/what-isilon)

Challenges

- about the resource expenditure... (computational, or people resources)
- Prioritizing things is a challenge, workflows are still very manual, and archivists will not give up on idea that snowflake and needs to be managed so carefully... do the digitized files all need to be monitored or are born digital more 'valuable'
 - Minnesota Historical Society... implementing storage tiers to appraise materials... Assigned level dictates the quality of storage provided... some get tape with checksums, but others get Exact File per collection on the higher... so in general the most fragile stuff gets the most attention... The appraisal is not based on the content, it is based on born digital/digitized, who has the copies, what is the legal obligation to keep things are...
 - Michigan State... also uses different levels of preservation. Things with a higher level of preservation, uses a higher amount of resources including fixity resources

Tools / Methods Mentioned

- Use tools that use RSYNC... built into the process of moving things around... incremental backups... fixity events that take place... get alerts when things go wrong...
- Bag It being used in multiple places
- Use of Block Chains and Tokens
- Splunk... for the monitoring envt? [Question asked if anyone uses it] No response... LOC and NARA have it but don't use it for that... it had others purposes...
- code.gov new repository [20% code due this week]

- audit control envt (ACE), sha256 and tokens for the checksums with time

Possible Outcomes from Use Cases

- decision tree... what cluster am i in... digitization files from vendor... state archives, special collections... then where do I go from that
- series of blog posts as these use cases
- stories about people about using the info and what they did do with the info... horror stories

Why are You Here Today: What is keeping you up at night?

- build tools, sticking points, and like to develop tools.... using the concepts of bag it and checksums not in bag it... side cars... see that a lot... option developing a script would support that... match and put together and match things... OR... approach with policy... can you come up with set of use case from NDSA perspective with vendors... worked well with file formats... come up with a few scenarios that we can push... ndsa recommended practices...
- interested in hearing what others are doing... have checksums for the files but hundreds of billions... not fixity checking them all monthly... capture, move things around, all the system guys check after things moved, do check fixity day after deposit.... disk went down, but the software didn't know the disk went down... 24 hours after deposit... and build out infrastructure and want to do scheduling but the more you touch it,
- what we identify as fixity... system admin... vs time and date changes...
- md goes in a database.... it does not go into the file name...
- they understood that checksum and fixity that a good thing, but don't understand the utility of it (fedora). When is that info created... interested in hearing what others were doing and the utility of what people we doing...
- triple constraint issue... can check all the time... can't get to all, might damage.... so dial it back a bit... once every two years, is that enough... figuring out in community discussion, given what you situation... what is the place to put the cursor on the dial....
- use migration... don 't check at rest but since we migrate at a regular basis...
- Use the NDSA levels.... take them to the next step...
- points at which great risk.... when to check, but when else should you do things... how often... or if there are access vs dark archives...

- reference levels of compliance... and Leslie Johnson's document...