

Digital Preservation at Oxford and Cambridge

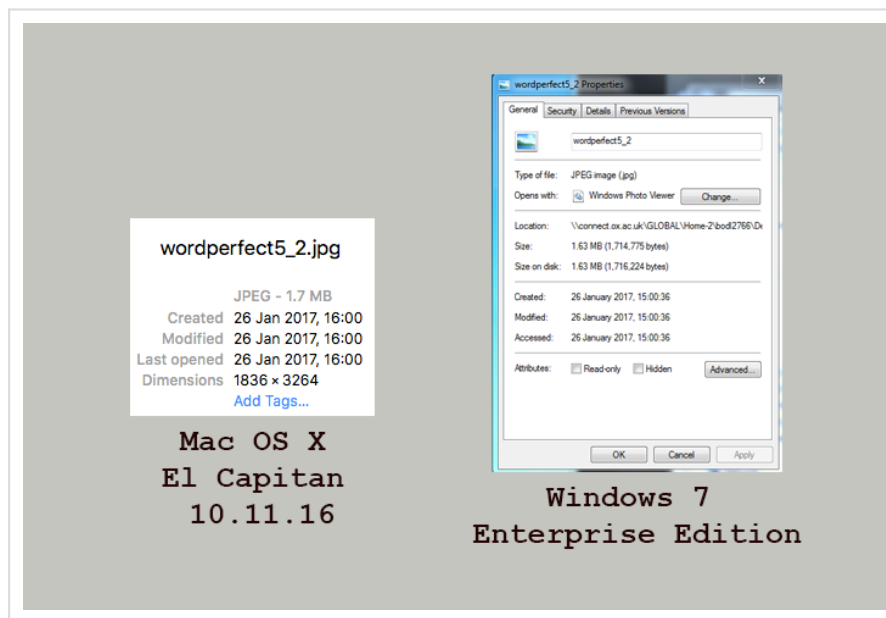
A collaborative research project to evaluate and provide sustainable recommendations for our digital preservation programmes

When was that?: Maintaining or changing ‘created’ and ‘last modified’ dates

Posted on [6 June, 2017](#) by [Sarah](#)

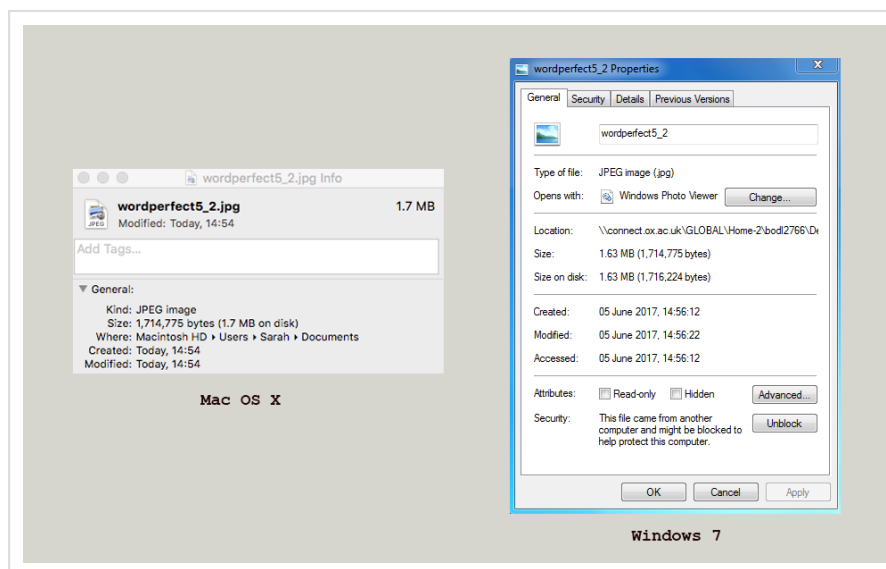
Sarah has recently been testing scenarios to investigate the question of changes in file ‘date created’ and ‘last modified’ metadata. When building training, it’s always best to test out what your advice before giving it and below is the result of Sarah’s research with helpful screenshots.

Before doing some training that involved teaching better recordkeeping habits to staff, I ran some tests to be sure that I was giving the right advice when it came to *created* and *last modified* dates. I am often told by people in the field that these dates are always subject to change—but are they really? I knew I would tell staff to put created dates in file names or in document headers in order to retain that valuable information, but could the file maintain the correct embedded date anyways? I set out to test a number of scenarios on both my Mac OS X laptop and Windows desktop.



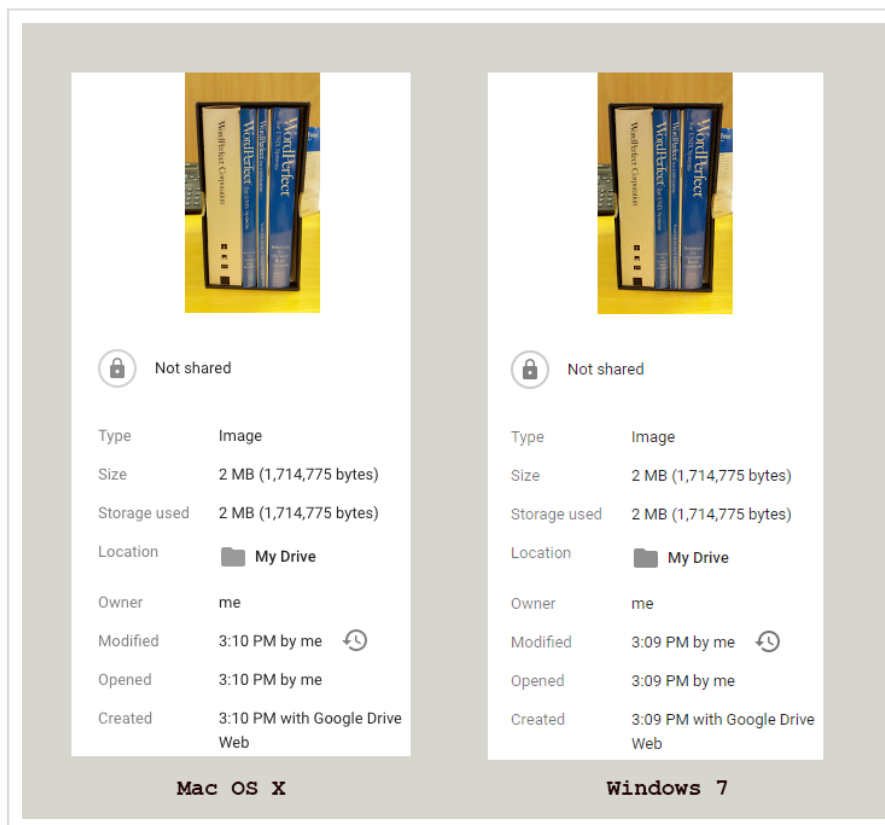
Scenario 1: Downloading from cloud storage (Google Drive)

This was an ALL DATES change for both Mac OS X and Windows.



Scenario 2: Uploading to cloud storage (Google Drive)



Once again this was an ALL DATES change for both systems.



Note: I trialled this a second time with the Google Drive for PC application and in OS X and found that *created* and last *modified* *dates* do not change when the file is uploaded or downloaded the Google Drive folder on the PC. However, when in Google Drive via the website, the *created* date is different (the date/time of upload), though the 'file info' will confirm the date has not changed. Just to complicate things.

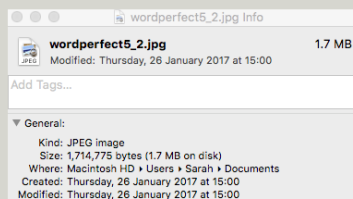


Not shared

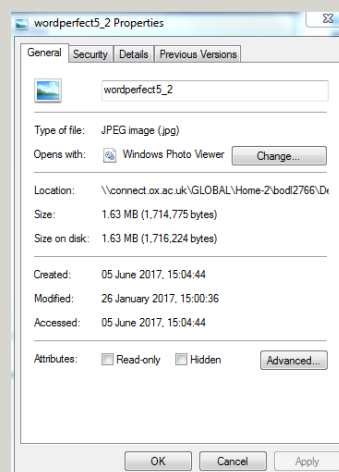
Type	Image
Size	2 MB (1,714,775 bytes)
Storage used	2 MB (1,714,775 bytes)
Location	 My Drive
Owner	me
Modified	Jan 26, 2017 by me 
Created	3:12 PM

Scenario 3: Transfer from a USB

Mac OS X had no change to the dates. Windows showed an altered *created* date, but maintained the original *last modified* date.



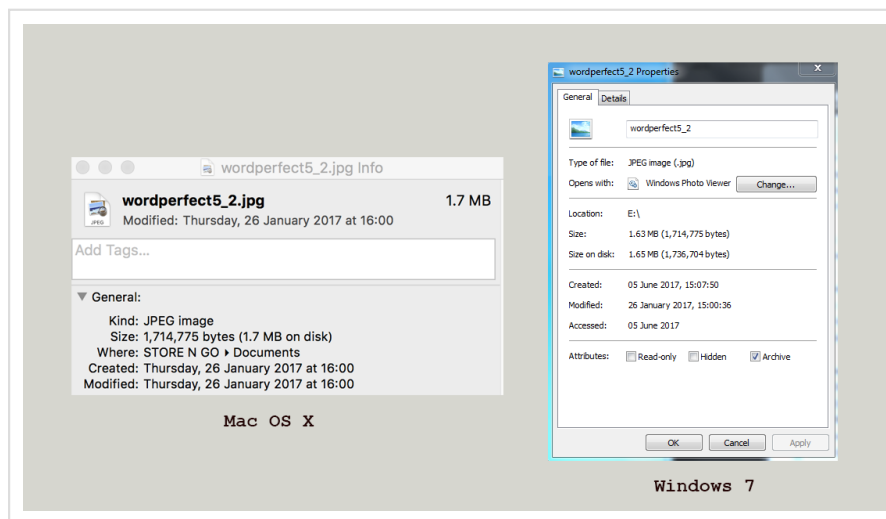
Mac OS X



Windows 7

Scenario 4: Transfer to a USB

Once again there was no change of a dates in the Mac OS X. Windows showed an altered *created* date, but maintained the original *last modified* date.



Note: I looked into scenarios 3 and 4 for Windows a bit further and saw that [Robocopy](#) is an option as a command prompt that will allow directories to be copied across and maintains those date attributes. I copied a 'TEST' folder containing the file from the Windows computer to the USB, and back again. It did what was promised and there were no changes to either dates in the file. It is a bit annoying that an extra step is required (that many people would find technically challenging and therefore avoid).

```

C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\bod12766>robocopy \\connect.ox.ac.uk\GLOBAL\Home-2\bod12766\Documents\TEST E:\

ROBOCOPY :: Robust File Copy for Windows

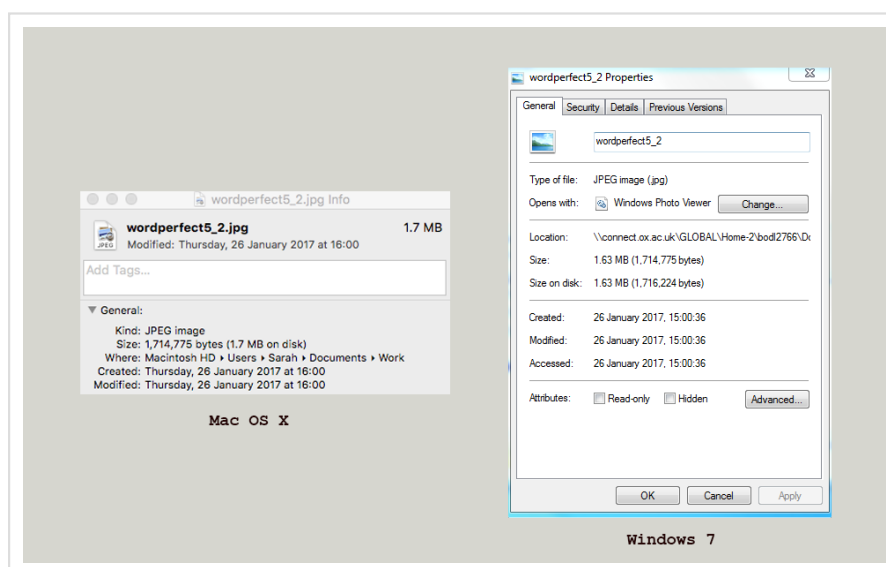
-----
Started : Mon Jun 05 15:26:16 2017
Source : \\connect.ox.ac.uk\GLOBAL\Home-2\bod12766\Documents\TEST\
Dest = E:\
Files : *.*
Options : *.* /COPY:DAT /R:1000000 /W:30

-----
nts\TEST\          1      \\connect.ox.ac.uk\GLOBAL\Home-2\bod12766\Docume
*EXTRA Dir         -1      E:\.fsevents\
*EXTRA Dir         -1      E:\.Spotlight-V100\
*EXTRA Dir         -1      E:\.TemporaryItems\
*EXTRA Dir         -1      E:\.Trashes\
*EXTRA Dir         -1      E:\Documents\
*EXTRA Dir         -1      E:\F DRIVE\
*EXTRA Dir         -1      E:\MSLA templates\
*EXTRA Dir         -1      E:\MSLA_USB\
*EXTRA Dir         -1      E:\SLU final presentations\
*EXTRA File        4096      ._.TemporaryItems
*EXTRA File        4096      ._.Trashes
*EXTRA File        4096      ._20170531_PersonalDigitalArchiv
ing.pptx           4096      ._.wordperfect5_2.jpg
g.pptx            139.0 m     20170531_PersonalDigitalArchivin
100%              *EXTRA File        1.6 m     wordperfect5_2.jpg.zip
                  New File         1.6 m     wordperfect5_2.jpg

-----
Dirs :      Total      Copied   Skipped  Mismatch  FAILED  Extras
Files :      1          0        1         0         0        9
Bytes :    1.63 m     1.63 m        0         0         0    140.64 m
Times :    0:00:01    0:00:01        0         0         0    0:00:00
  
```

Scenario 5: Moving between folders

No change across either systems. This was a relief for me considering how often I move files around my directories.



Conclusions

When in doubt (and you should always be in doubt), test the scenario. Even when I tested these scenarios three of four times, it did not always come out with the same result. That alone should make one cautious. I still stick to putting *created* date in the file

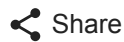
name and in the document itself (where possible), but it doesn't mean I always receive documents that way.

Creating a zip of files/folders before transfer is one method of preserving dates, but I had some weird issues trying to unzip the file in cloud storage that took a few tries before the dates remained preserved. It is also possible to use [Quickhash](#) for transferring files unchanged (and it generates a checksum).

I ignored the last accessed date during testing, because it was too easy to accidentally double-click a file and change it (as you can see happened to my Windows 7 test version).

Has anyone tested any other scenarios to assess when file dates are altered? Does anyone have methods for transferring files without causing any change to dates?

SHARE THIS:



This entry was posted in [case study](#), [digital preservation](#), [personal digital archiving](#), [technology](#) by [Sarah](#). Bookmark the [permalink \[http://www.dpoc.ac.uk/2017/06/06/when-was-that-maintaining-or-changing-created-and-last-modified-dates/\]](http://www.dpoc.ac.uk/2017/06/06/when-was-that-maintaining-or-changing-created-and-last-modified-dates/).

About Sarah

Digital Preservation Specialist - Outreach and Training:
Bodleian Libraries, Oxford University

[View all posts by Sarah](#) →

12 THOUGHTS ON "WHEN WAS THAT?: MAINTAINING OR CHANGING 'CREATED' AND 'LAST MODIFIED' DATES"

Sarah

on [10 May, 2018 at 09:49](#) said:

Excellent points. I think a lot does come down to the subtleties in using different file systems, but each OS does seem to preference a different file system (also there are compatibility issues to consider) that makes thinking about different ones a good starting point. My Windows using NTFS and I am fairly certain the MAC (Unix) uses APFS.

Both of those would definitely have different issues and impact index information like date created and date modified in different ways.

Deciding which value is write is probably trickier – it depends on your use case. The dates might not actually be a part of the file, but they are important metadata nonetheless. I think what it is important to remember (and what the tests show) is that it's not a trustworthy piece of metadata and should be used with a grain of salt. It doesn't mean we shouldn't record them or consider them valuable, but be aware of the limitations of these dates and the complexity around file systems and OS (and Google, because I truly have no idea how Google stores documents on the cloud, but perhaps someone does, in which case I would be grateful to learn more!).

Ed
on **9 May, 2018 at 03:11** said:

A thoughtful piece of work.

From my own reading around the topic, I think you may find there is a subtlety in that it can be affected by the File system (FAT, etc, etc) which is used to host the file (as opposed to the OS or Internet host), in that some file systems or may not actually support a field to hold a value, as well as the specific copy-executable that being used to copy – and whether it copies or updates the data correctly in those fields.

In deciding which value it right, there is perhaps also a question of definitions – whether copying a file modifying it?, or even creating it? – personally I would say it is neither and is simply an access as it does not actually change any data in the file (- that is certainly more meaningful for synching software, but I am not the authority on such definitions).

Sarah
on [12 June, 2017 at 20:58](#) said:

> Such information may not always be metadata or technical metadata, and can come from many sources. If possible I would like to find a way to store and curate it, independent of these fiddly little file properties. Perhaps I'm saying I'd like to move towards a form of triangulation, where we can depend on more than one source to give us a complete picture of the truth.

Excellent points! I especially like the idea of triangulation. Discussions like these leave me wondering what conclusions and discoveries researchers will make and what stories they will tell.

Sarah
on [12 June, 2017 at 20:48](#) said:

It has quite a few options and controls if you want to do some complex transfers. I will certainly get it onto a Windows computer and run a few tests. I am also running some tests with Exactly after some suggestions, but I haven't looked into embedded metadata yet.

It's tricky, because I always try to find solutions that do not require downloading or using too many programs for content creators. Or at least I try to provide levels of options. I am looking at Fixity and I know that it provides some interesting options for checking, but often when using Mac OS X I just run the `shasum -a 256` command on my files in and output to a text file (for my personal collections). You can run a check on the text file (`shasum -c`) and it will verify all the stored checksums. It is quick and easy and requires no additional software. But there are extra features that a GUI program like Fixity would provide.

Kirsty Lee

on [9 June, 2017 at 14:38](#) said:

I would recommend RichCopy. It has evolved out of Robocopy and is considered a more advance tool than its predecessor. I've found that this maintains both the 'date created' and the 'date modified' fields intact. Its a really easy tool to use, but it doesn't come with checksum creation or validation. I'm currently testing the Exactly tool, but not sure yet how it treats embedded metadata.

Ed Pinsent

on [8 June, 2017 at 16:19](#) said:

> It definitely made me think about my own practices and those of our researchers and other creators. We do so much thinking and work to save them when we transfer and ingest them into our repositories, but what are they really? Are they really that significant of a property when they can change so easily? Or that a checksum doesn't register them when they change?

I would look at it from a slightly different angle.

Individual date properties may or may not be "significant", but I think the cumulative effect of dates changing and original dates of creation getting overwritten would be to cast doubt on the overall authenticity of our collections. From an archivist point of view, date of creation is a ground truth, and one that future historians are going to obsess over. It's massively important to know what happened when, and in what order.

In terms of our practices and working with content creators, I'm in favour of assembling as much contextual information as we can get; it could be external to the deposited collection, could be compiled by the creator, or by the archivist. I still cling to old-fashioned documents like transfer lists, and in my perfect world there would be a digital equivalent to a transfer list in every SIP we get. I also like manifests, system audits, and directory audits...anything that can give us a view of when (and how) files were arranged in a directory.

Such information may not always be metadata or technical metadata, and can come from many sources. If possible I would like to find a way to store and curate it, independent of these fiddly little file properties. Perhaps I'm saying I'd like to move towards a form of triangulation, where we can depend on more than one source to give us a complete picture of the truth.

Sarah
on **8 June, 2017 at 11:04** said:

Ed got me on that post via Twitter! I had a really good read of it, as I have some Google Docs that I will have to migrate at some point when they are finished and I am now keenly aware they won't retain important dates (wish I had noticed that sooner). I hadn't thought about the effect on comments and revision information either. There's so much extra functionality in a Google Doc than you realise. Your methodology was great as well—I am keen to see what you find out with Google Sheets.

Robocopy is already installed on the computer and run with a simple command prompt, so it made it fairly quick to use (you can Google how to use it and you'll get plenty of advice). My only issue was that it copied at the directory level and not individual files—so not bad if you plan to work at scale.

I wasn't sure about the checksum changing when the date metadata alters. I've read that it could for an MD5 since it calculates created dates in it. So, I tried an MD5 on the same file with two different dates and it remained unchanged. I also tested it with a SHA-256 and found no change then either. So, it seems to have no impact on the checksum.

It definitely made me think about my own practices and those of our researchers and other creators. We do so much thinking and work to save them when we transfer and ingest them into our repositories, but what are they really? Are they really that significant of a property when they can change so

easily? Or that a checksum doesn't register them when they change?

Sarah

on **8 June, 2017 at 09:40** said:

Thanks, Somaya. Some good tools and information in your comment that I think would be useful to a lot of people. I would say that if people are taking the trouble to download DROID or Exactly then also installing Java is generally not a difficult step. I didn't know when I got my Mac that it wasn't installed with Java and when I went to get a Java-dependent program it prompted me to get Java. Overall, it was probably the most painless step in the process of installing and getting the program running. I know your post pointed to times where it could be a potential problem.

I was trying to point out recordkeeping and file management from an active life. We go to so much trouble to preserve these dates, but do they really reflect what they think they reflect? I moved files from a 3.5 floppy about 12 years ago and then transferred them years later to Google Drive as my laptop died. The created date reflects 2014, but I know for a fact I wrote that essay in 2000.

I was also thinking how to lower the bar for staff to have better record-keeping habits and to pass them on to researchers. In an effort to improve the quality of what might end up in archives later, but also how we manage deposits (though dates for AAMs are probably less crucial considering the other date metadata we collect like accepted & published dates). Programs are great, but lots of people just won't bother. Even I forget when I am in a hurry. When in doubt, put the date elsewhere.

Somaya Langley

on **7 June, 2017 at 17:46** said:

There's also a good whitepaper by the SANS Institute that looks into timestamps and how they're handled in NTFS, FAT32 and exFAT in combination with selected Windows Operating Systems. See: <https://www.sans.org/reading-room/whitepapers/forensics/filesystem-timestamps-tick-36842>

In terms of tools for doing file transfer and retaining disk file system metadata, I've also used Robocopy, but with a bunch of different flags e.g. COPYALL (though I can't remember exactly which ones off the top of my head). For command line copying in the Mac OSX environment, in the past I've used: `rsync -av`

I've also used TeraCopy (especially in scenarios where the network architecture doesn't allow you to address the drive by its letter – yes, it has happened) – <http://www.codesector.com/teracopy>

There's also a couple of other tools, such as AV Preserve's Exactly – <https://www.avpreserve.com/tools/exactly/>

However as you'll notice with DROID and Exactly, they're Java-based. There aren't GUI tools that don't depend on Java for the Mac environment (and current Mac OSes don't come with Java installed).

I wrote about this issue in one of my blog posts last year (<http://www.dpoc.ac.uk/2016/11/18/the-gap-in-digital-preservation/>) and referenced a call out I made to the Digital Curation Google Group about needing a non-Java dependent GUI tool for the Mac to do file transfers. See: <https://groups.google.com/forum/#!topic/digital-curation/iARnIMwZXn8>

But as you say, better record-keeping and file management is important.

Jenny Mitcham

on **7 June, 2017 at 16:41** said:

This is the sort of blog post that I really like to read! Makes me question my own practices. I need to think about this stuff more!

Not tried Robocopy but perhaps I should – interested to hear what other people do.

Am I right in thinking that a change of creation or last modified date won't impact the checksum?

You may have seen my blog post on preserving Google documents last month (<https://digital-archiving.blogspot.co.uk/2017/04/how-can-we-preserve-google-documents.html>)? I was interested to discover that if you use the Google Takeout service to export a batch of files out of Google Document format and into MS Word for example, it somehow *did* manage to retain the system dates. However, if you individually export a single file it doesn't. Just thought I'd add that into the mix!

Sarah
on **7 June, 2017 at 12:05** said:

It's a really excellent article by Moran and Gattuso, from good view points as well. In terms of policy and workflows, how we manage the "pre-conditioning" step and how we'll define it is something we're debating about. It's also interesting to consider what metadata matters in what case — sometimes those dates are less important than others because of other dates stored in the descriptive metadata (like publication or acceptance date for AAMs in our research archive). Might have to pull that article out of the archive again though!

But when we think about getting staff to have better recordkeeping habits, sometimes the easiest thing you can do is get them to record the date in the file name. Testing out these scenarios and seeing what changes and when (I also did some random changes using the touch command in bash). It's not always when you think it would and it's really easy to do over the active life of a digital file (and beyond).

Bertrand Caron
on **7 June, 2017 at 11:29** said:

If I remember well, there was a paper from J. Moran and J. Gattuso (NLNZ) at iPRES 2015: “Beyond the Binary: Pre-Ingest Preservation of Metadata” (Proceedings, p. 137, <https://phaidra.univie.ac.at/view/o:429524>), which could be of interest for this topic. Basically, the authors suggest presenting a “pre-conditioning policy” which would describe the operations that a repository performs on digital objects before ingest in order to ensure long-term preservation, and recording original filenames and file dates as provenance metadata.

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)