Opportunities and Future Directions
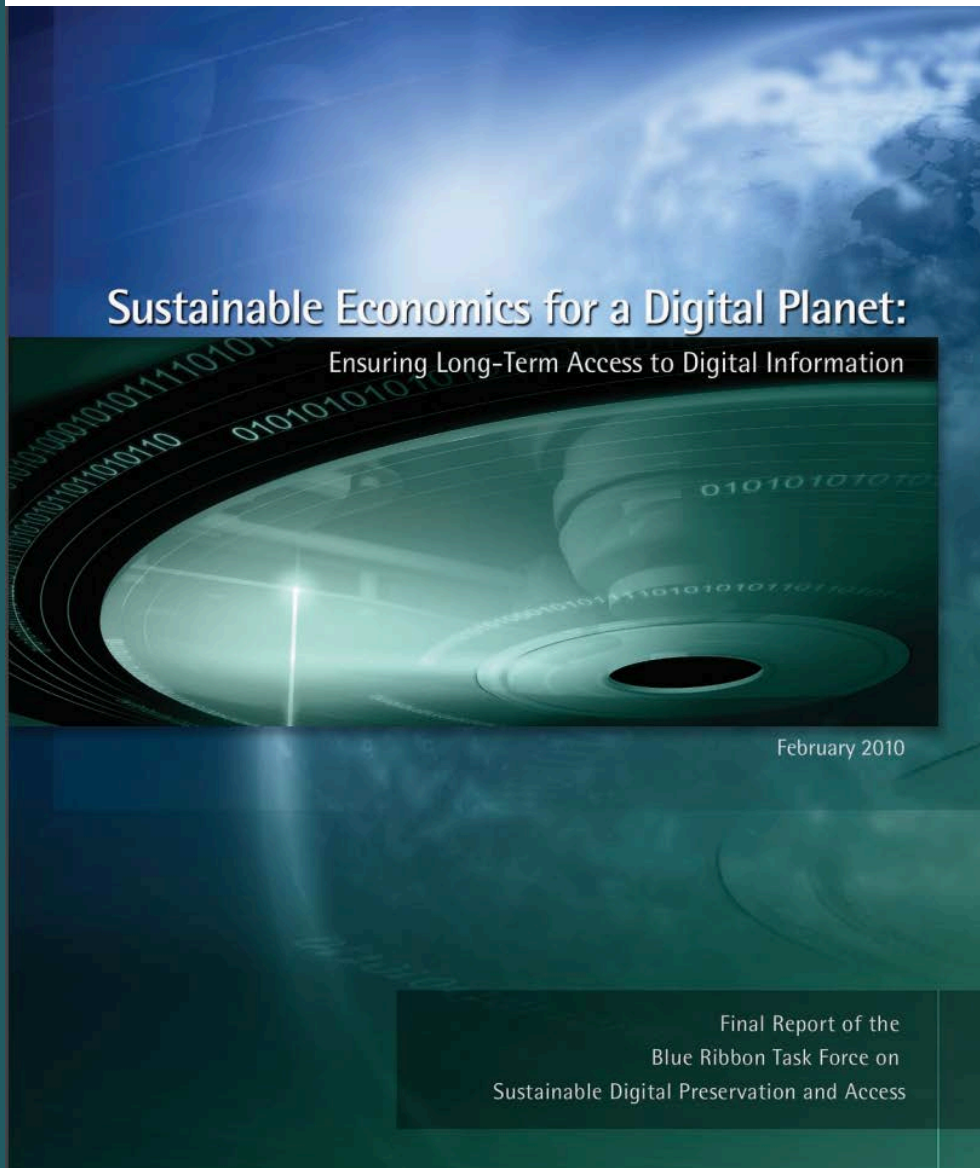
Digital Preservation

Jaye Weatherburn
Data Stewardship Coordinator
The University of Melbourne
@jayechats

# What does "digital preservation" mean to you?

Go to www.menti.com
and use the code 18 39 50

Sustainable Economics for a Digital Planet:
Ensuring Long-Term Access to Digital Information

February 2010

Final Report of the
Blue Ribbon Task Force on
Sustainable Digital Preservation and Access

"Sustainable digital preservation requires a compelling value proposition, incentives to act, and well-defined roles and responsibilities."

bit.ly/2qg1Hge
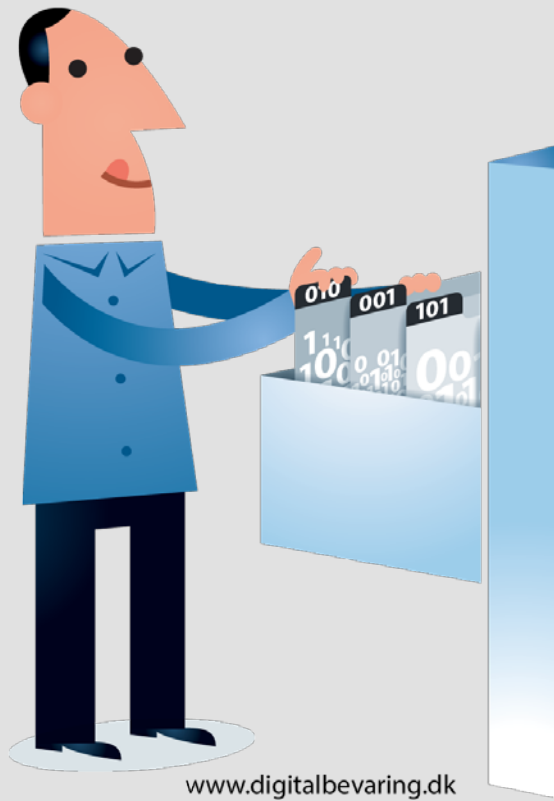
Why do we need digital preservation?

Jaye Weatherburn
@jayechats

Hey #digipres community, what is your favourite go-to answer for the question "Why should I care about digital preservation?" I want to share these at a talk I'm doing on #digitalpreservation (audience focus/knowledge is academic sector: institutional repositories) #AusPreserves

Stephanie Morton
@BooktopusSteph

Replying to @jayechats

Is it worth losing all the hours of effort you spent gathering & analysing data and labouring over the publications that came from it (+the data/pubs) simply because the only drive you stored it on died or proprietor of one of the file formats decides to stop supporting it?

I like this example from Data Management for Researchers by @KristinBriney
pelagicpublishing.com/products/data- …

THE DATA PROBLEM

On July 20, 1969, Neil Armstrong climbed out of his spacecraft and placed his feet on the moon. The landing was broadcast live all over the world and was a significant event in both scientific and human history. Today, we can still watch the grainy video of the moon landing but what we cannot do is watch the original, higher quality footage or examine some of the data from this mission. This is because much of the data from early space exploration is lost forever.

Among the lost data are the original Apollo 11 tapes containing high-quality video footage of the moon landing. Their loss first came to light in 2006 (Macey 2006) and NASA personnel spent the next three years searching for the tapes across multiple continents before concluding that they were likely wiped and reused for data storage sometime in the 1970s (NASA 2009; O'Neal 2009; Pearlman 2009). Other data from this

Katrina Trewin
@trewkat

Replying to @jayechats

Ok: "Study the past if you would define the future" ~Confucius. Can't study something that's no longer accessible..

somaya langley
@criticalsenses

Replying to @jayechats

how much would it cost you/how much effort would it take to redo your research if you lost access to it (or could it even be redone)?

Jaye Weatherburn
@jayechats

Replying to @criticalsenses

And I guess, for repository managers responsible for looking after said research, how much would it cost/how much effort would it take to recover, if access was lost? Or could it be recovered at all?

Stephanie Morton
@BooktopusSteph

Replying to @jayechats @criticalsenses

Exactly! Also, from a repository standpoint, we can't preserve & curate what you can't provide! I've had so many researchers deposit metadata records for pubs then can't actually provide me with the pub or other verification material needed. If YOU don't have it...

somaya langley
@criticalsenses

Replying to @jayechats

yeah - for "respository managers" (who i really feel it isn't suitable to put all of that responsibility on one role in a whole digital content lifecycle), u can really emphasise that due to the non-tenured nature of many of academics, a whole lotta researchers will have moved on

Leslie Johnston
@lljohnston

Following

I've occasionally been asked when we will be "done," as in all formats identified, all decisions documented, and all formats migrated to something sustainable. I hate seeing the disappointment when I explain it's an ongoing, cyclical process of monitoring/maintenance.

Jason Scott
@textfiles

Follow

I'd just say "When is the yardwork done?"

But putting stuff into a repository IS

preserving it?

Maintain hardware & software

Emulation > https://bit.ly/2D81R1S

Combination of strategies

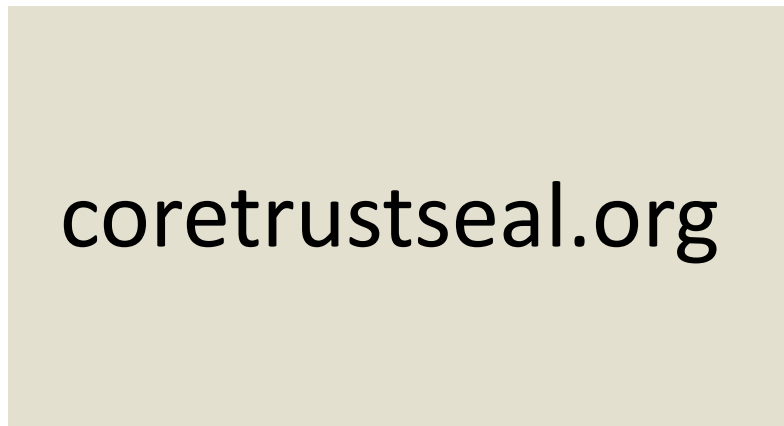|	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs
Metadata	- Inventory of content and its storage location - Ensure backup and non-collocation of inventory	- Store administrative metadata - Store transformative metadata and log events	- Store standard technical and descriptive metadata	- Store standard preservation metadata
File Formats	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

Trust
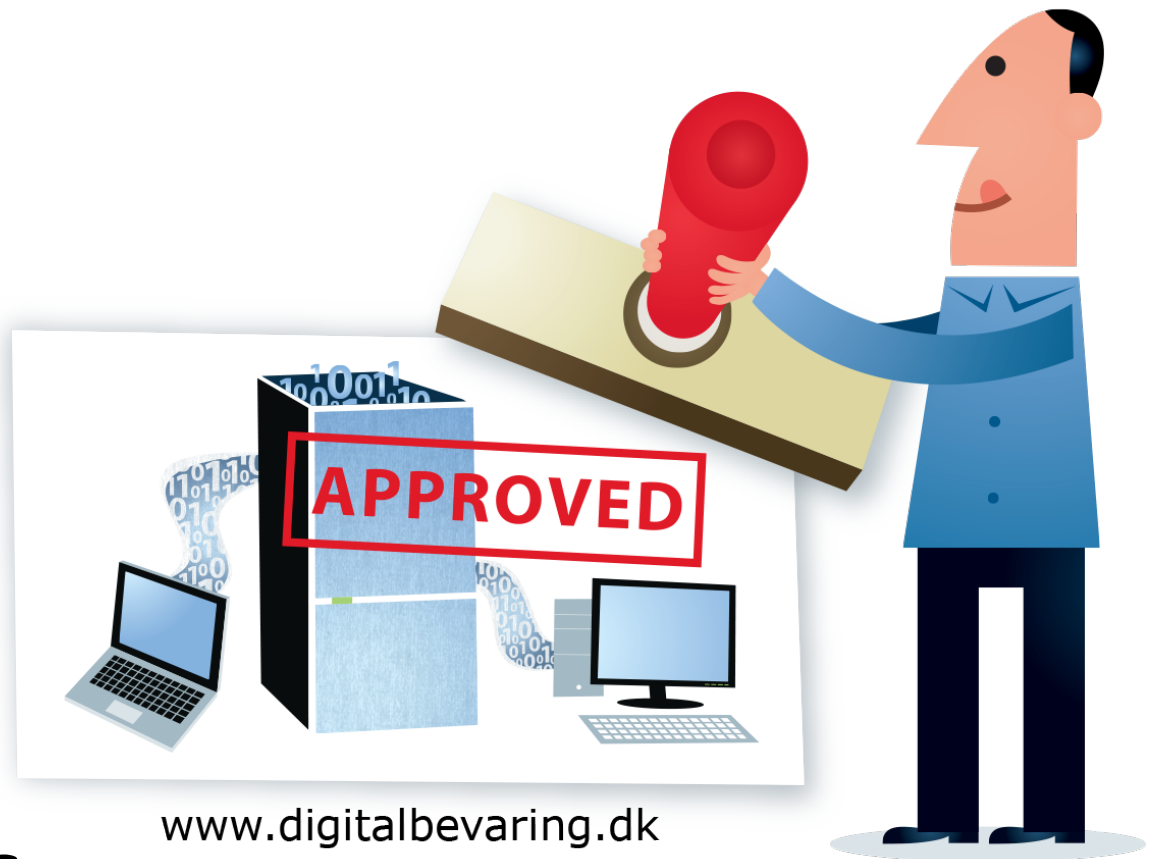Self-assessment
Audit
Certification

Em Johnson
@MissEmilieLib

I'd like to survey my researchers and ask them what their awareness/trust levels are for my IR - for institutional repositories generally, but SPECIFICALLY their own designated institutional repository #APOForum2018 #OAWeek18

1:45 PM - 25 Oct 2018

CoreTrustSeal certification

- Trustedness of container, not the data
- Emerged out of WDS, DSA certification
- Assess maturity through documented procedures on ingesting and making data available, underlying technical and organisational infrastructure
- Australia was lagging behind

coretrustseal.org

The Trusted Data Repository Program

Ran 2016 till 2018
NIF: Multi Node National facility
ADA: Disciplinary National Facility
CSIRO: Institutional repository

www.digitalbevaring.dk

Self-assessment
Audit frameworks

ICPSR case study, p. 94

Key reasons for needing a
trustworthy repository system

1. Accurate replication and validation

2. User community looks to the
repository as a trustworthy source
of info

3. Increasing funder requirements

4. Standards of trustworthiness
help repositories self-assess
capabilities

BUILDING TRUSTWORTHY
DIGITAL REPOSITORIES

THEORY AND IMPLEMENTATION

EDITED BY PHILIP C. BANTIN

Self assessment or audit?
What framework to use?

What do you want to achieve?
What resources are available?
What type of content?

iPRES 2018
the 15th International
Conference on Digital
Preservation

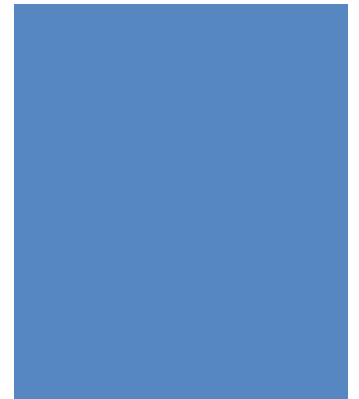Conference proceedings: osf.io/u5w3q

To join: groups.google.com/forum/#!forum/australasia-preserves/join

Australasia Preserves
DP community of practice
bit.ly/2EOmRwe

SLNSW videos

bit.ly/2vDn8dw

DP Carpentry: bit.ly/2JkFkiu

Dia mundial da
preservação digital
29 novembro 2018

Παγκόσμια Ημέρα
Ψηφιακής Διατήρησης
29 Νοεμβρίου 2018

Welttag Digitale
Erhaltung
29. November 2018

Giornata mondiale della
preservazione digitale
29 novembre 2018

Journée mondiale de la
préservation numérique
29 novembre 2018

World Digital
Preservation Day
29 November 2018

Dia mundial de la
preservación digital
29 noviembre 2018

اليوم العالمي للحفظ الرقمي
29 نوفمبر 2018

Werelddag van de
Digitale Duurzaamheid
29 november 2018

bit.ly/2F04VyX

#WDPD2018

World Digital
Preservation
Day
29 November 2018
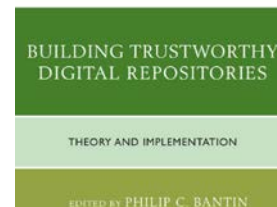
APPROVED
www.digitalbevaring.dk

Thank you

F
A
I
R

"Digital preservation is a challenge for all of society because we all benefit from reliable, authentic information now and into the future. Done well, all of society will reap the benefits of digital stewardship."

Jaye Weatherburn
Data Stewardship Coordinator
The University of Melbourne
@jayechats

TRUST
MIN   MAX

Sustainable Economics for a Digital Planet:
Ensuring Long-Term Access to Digital Information

BUILDING TRUSTWORTHY
DIGITAL REPOSITORIES

THEORY AND IMPLEMENTATION

EDITED BY PHILIP C. BANTIN

Final Report of the
Blue Ribbon Task Force on
Sustainable Digital Preservation and Access

DIGITISATION
IS NOT
DIGITAL PRESERVATION

DIGITISATION *NEEDS* DIGITAL PRESERVATION