

## Characterisation: Making sense of your collections



Know your enemy!

Pronom + Droid

- Duplication
- Fixity
- Characterisation

Some others

## File formats are a problem for preservation ... sort of



Changing file formats

vs

'Robust' formats

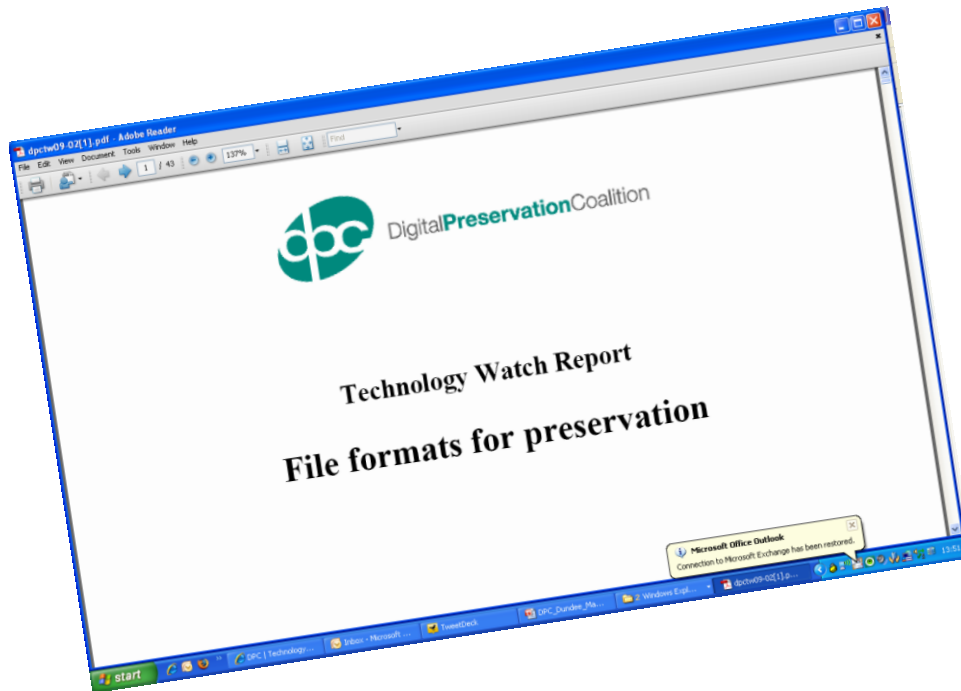
vs

Proliferating formats

vs

Conformant data containers

## How to pick a winner ...



Adoption  
Dependency  
Disclosure  
Transparency  
Metadata support  
Interoperability  
Complexity  
Stability  
Rights management

Todd, M 2009 'File formats for preservation', DPC Technology Watch Report 02/09, online at <http://www.dpconline.org/advice/technology-watch-reports.html>



## How to pick a winner ...

beyond and **potentially over-writing the criteria** ...  
repository managers should **align** the recognition and  
weighting of criteria with a **clear preservation strategy**  
that articulates the **purpose** of the repository and the  
**needs of its designated community**;

Todd, M 2009 'File formats for preservation', DPC Technology Watch Report 02/09, online at <http://www.dpconline.org/advice/technology-watch-reports.html>

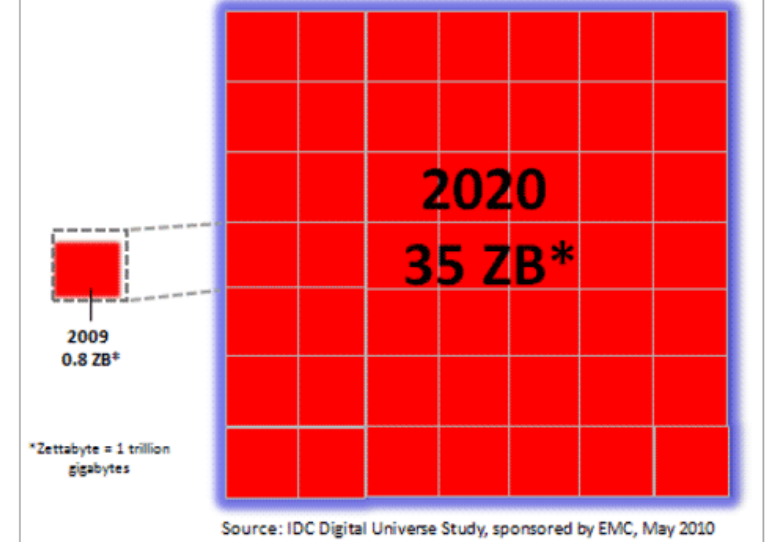
## How to pick a winner ...



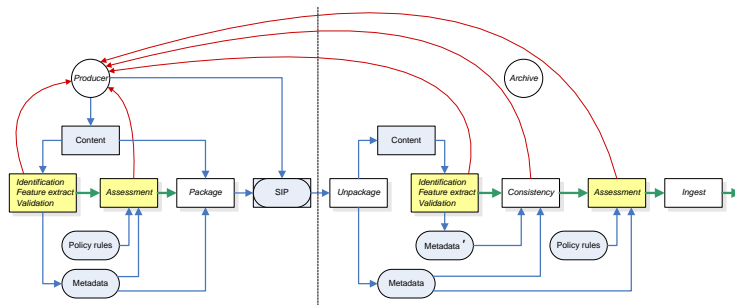
## 'Digital Universe' Nears A Zettabyte

May 4th, 2010 : Rich Miller

**Figure 1: The Digital Universe 2009 – 2020**  
*Growing by a Factor of 44*



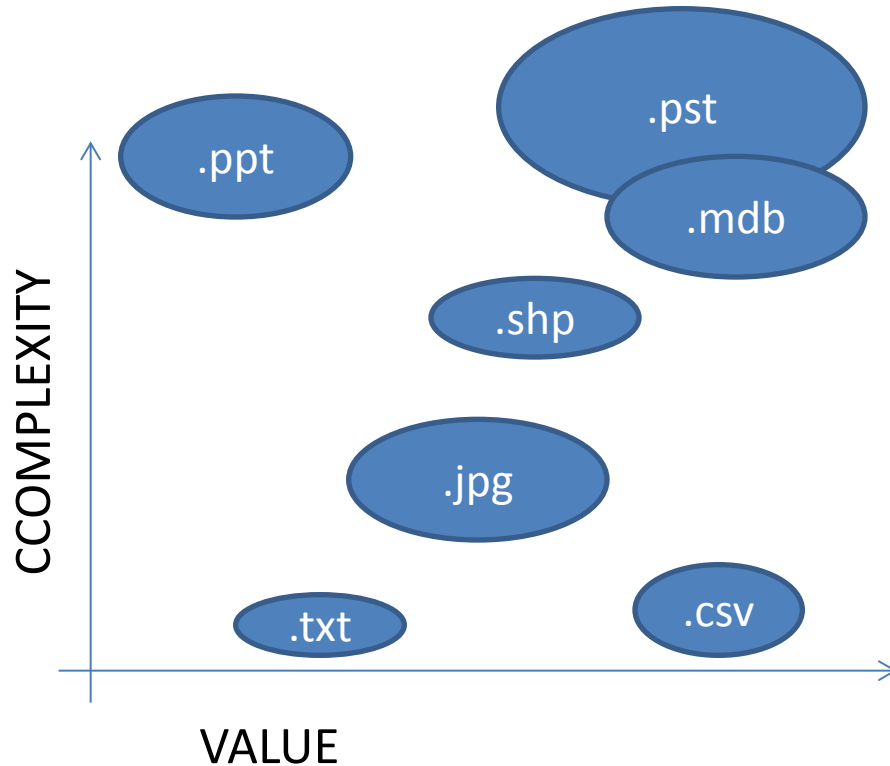
The Great Recession hasn't slowed the breakneck growth of the Digital Universe. In 2010 the volume of digital information created and duplicated in a year will reach 1.2 zettabytes, according to new data from IDC.



... it's not going to be about  
obsolescence so much as  
workflow and capacity



## ‘Characterisation’: the process of knowing what you have



Technology not content  
Is it what it says it is?  
Is it a problem?  
What are our priorities?  
Are these the same priorities  
as other people?

## **Pronom + Droid: eh?**

Pronom: a register of file formats and their behaviours  
(probably the world's most boring database)

DROID: a tool that analyses the files on a system (using the  
most boring database in the world)

<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>



Digital Preservation Coalition

# file formats and their characteristics

**The National Archives**

[myPage \(not signed in\)](#)

Search the archives

Search

[Advanced search](#)

[About us](#)

[Education](#)

[Records](#)

[Information management](#)

[Shop online](#)

You are here: [Home](#) > [Services for professionals](#) > [Preservation](#) > PRONOM



The **technical registry**  
**PRONOM**

Welcome : [About](#) [Add an entry](#)  
[Search](#) [Help](#) [Information resources](#)

Welcome to PRONOM

[PRONOM changes and DROID signature file release notes.](#)

Find out more about our plans to make PRONOM's data available

The online registry of technical information. PRONOM is a repository of software products and other technical components required to preserve digital information of historical or business value. Find out about the future of PRONOM

[Search PRONOM](#)

[Simple search](#) [File format](#) [PRONOM Unique Identifier](#) [Software](#) [Vendor](#) [Lifecycles](#) [Migration Pathways](#)

Details for: Tagged Image File Format 3

[Save as...](#) [XML](#) [CSV](#) [Print](#)

Go to: [Summary](#) | [Documentation](#) > | [Signatures](#) > | [Compression](#) > | [Character encoding](#) > | [Rights](#) > | [Reference files](#)  
> [Properties](#) >

## Summary

Name	Tagged Image File Format
Version	3
Other names	TIFF (3)
Identifiers	PUID: fmt/7 MIME: image/tiff Apple Uniform Type Identifier: public.tiff
Family	
Classification	Image (Raster)
Disclosure	Full
Description	The Tagged Image File Format (TIFF) is a raster image format originally developed by the Aldus Corporation, primarily for use in scanning and desk-top publishing. When Adobe Systems Incorporated purchased Aldus in 1994, they acquired the rights to the TIFF format and have maintained it since then. TIFF files comprise three sections: an Image File Header (IFH), an Image File Directory (IFD), and the image data. TIFF files can contain multiple images (multi-page TIFF), and each image has a separate IFD. The IFH always appears at the beginning of the file, and is immediately followed by a pointer to the first IFD. The IFD contains metadata which describes the associated image, stored as a series of tags. The IFD also contains a pointer to the actual image data. TIFF 3.0 supports colour depths from 1 bit to 24 bit (e.g. monochrome to true colour), and a range of compression types (LZW, and CCITT Group 3 and Group 4).





Digital**Preservation**Coalition



*DROID: Search and report on files from an entire network*

*Identify files by 'extension'*

*Identify files by 'signature'*

*Identify subtypes and versions*

*Report errors and concerns*

*Scalability ...*

*Built for TNA ...*

*Only as good as data ...*



*Also in this space:*

- *C3PO*
- *JHOVE*
- *TIKKA*
- *FITS*
  
- *UDFR*
- *TOTEM*

*Only as good as their data:  
SO JOIN IN!*

- *CRISP*
- *JUST FIX IT*
- *PRONOM*
- *DROID*
- *...*