

# UHT Digital Preservation

Dr David Tarrant

[davetaz@ecs.soton.ac.uk](mailto:davetaz@ecs.soton.ac.uk)

Electronics and Computer Science

# Overview

- Locating Tools
- Heavyweight Tools
- Global community involvement
- Self Reflection
- The new age of Distilled Digital Preservation

# Tools

- DROID
- PRONOM
- file
- fido
- FITS
- JHove
- FFmpeg
- Exiftool
- Office Analyser
- Open Office
- Metadata Extraction Tool
- ...

# Finding Tools - DROID

Google

**Search** About 25,200 results (0.32 seconds)

**Everything** [DROID \(Digital Record Object Identification\) – Freecode](#)  
[freecode.com/projects/droid](http://freecode.com/projects/droid)

**Images** **DROID (Digital Record Object Identification)** is a software tool to perform automated batch **identification of file** formats. It is designed to meet the fundamental ...

**Maps**

**Videos** [File Format Identification - Forensics Wiki](#)  
[www.forensicswiki.org/wiki/File\\_Format\\_Identification](http://www.forensicswiki.org/wiki/File_Format_Identification)

**News** 30 Nov 2010 – The UK National Archives developed the **Digital Record Object Identification (DROID)** tool, an "automatic **file format identification tool**." This tool ...

**Shopping**

**More** [SourceForge.net: DROID: News](#)  
[sourceforge.net/news/?group\\_id=160809](http://sourceforge.net/news/?group_id=160809)  
**DROID (Digital Record Object Identification)** is an automatic **file format identification tool**. It is the first in a planned series of tools developed by The National ...

**Southampton, UK**  
Change location

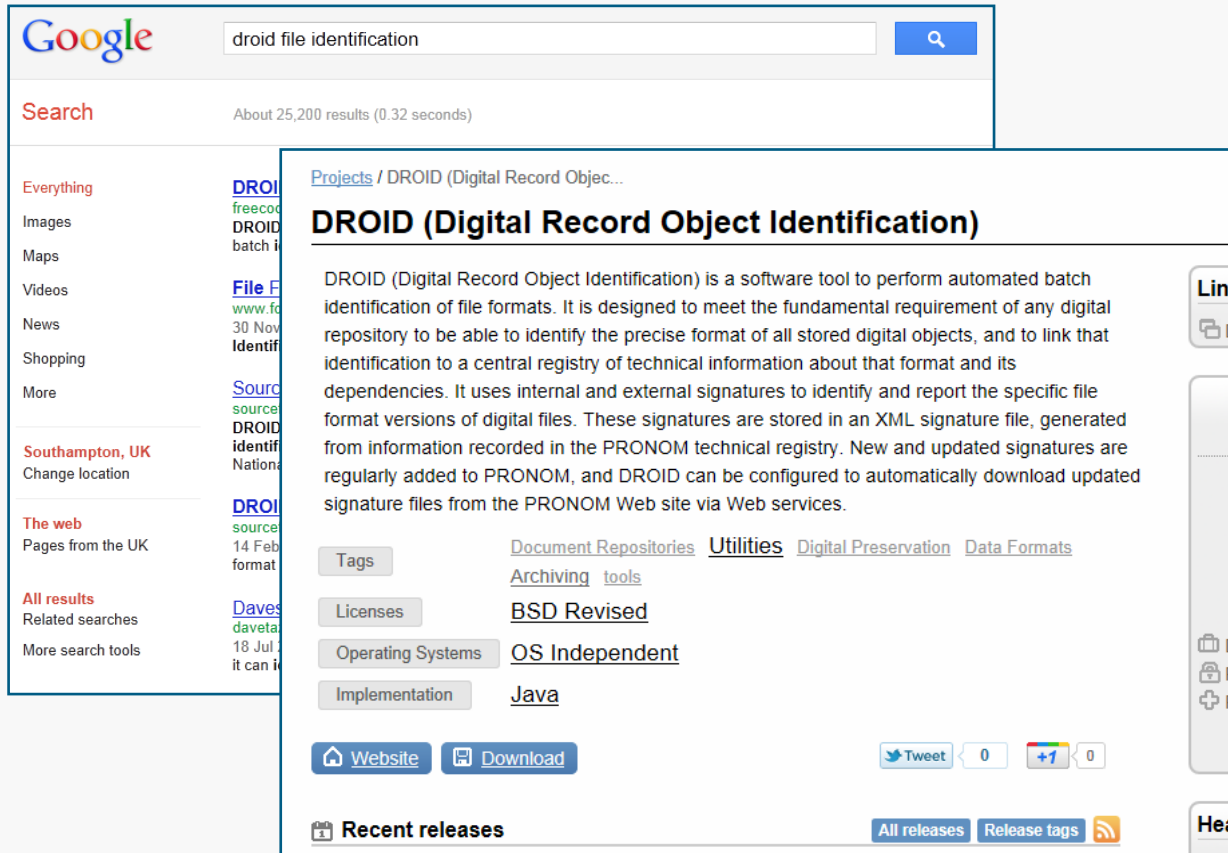
**The web** [DROID | Free Security & Utilities software downloads at ...](#)  
[sourceforge.net/projects/droid/](http://sourceforge.net/projects/droid/)  
14 Feb 2011 – **DROID (Digital Record Object Identification)** is an automatic **file format identification tool**. It is the first in a planned series of tools developed ...

**All results** [Daves thoughts on stuff: More on File Identification Tools](#)  
[davetaz-blog.blogspot.com/.../more-on-file-identification-tools.html](http://davetaz-blog.blogspot.com/.../more-on-file-identification-tools.html)  
18 Jul 2011 – **DROID** is an ever improving tool as the underlying data (number of **files** it can **identify**) expands. However certain decisions have led to it ...

**Related searches**

**More search tools**

# Finding Tools - DROID



The image shows a Google search interface. The search bar contains the text "droid file identification". Below the search bar, the search results are displayed. The main result is for "DROID (Digital Record Object Identification)". The description states: "DROID (Digital Record Object Identification) is a software tool to perform automated batch identification of file formats. It is designed to meet the fundamental requirement of any digital repository to be able to identify the precise format of all stored digital objects, and to link that identification to a central registry of technical information about that format and its dependencies. It uses internal and external signatures to identify and report the specific file format versions of digital files. These signatures are stored in an XML signature file, generated from information recorded in the PRONOM technical registry. New and updated signatures are regularly added to PRONOM, and DROID can be configured to automatically download updated signature files from the PRONOM Web site via Web services."

Below the description, there are several tags and categories:

- Tags: Document Repositories, Utilities, Digital Preservation, Data Formats, Archiving, tools
- Licenses: BSD Revised
- Operating Systems: OS Independent
- Implementation: Java

At the bottom, there are buttons for "Website" and "Download", and social media sharing options for "Tweet" (0) and "+1" (0). There is also a "Recent releases" section with buttons for "All releases" and "Release tags".

# Finding Tools - DROID

Google

Search About 25,200 results (0.32 seconds)

Everything  
Images  
Maps  
Videos  
News  
Shopping  
More

Southampton, UK  
Change location

The web  
Pages from the UK

All results  
Related searches  
More search tools

**DROID (Digital Record Object Identification)**

DROID (Digital Record Object Identification) is a software tool to perform automated batch identification of file formats. It is designed to meet the fundamental requirement of any digital repository to be able to identify the precise format of all stored digital objects, and to link that identification to a central registry of technical information about that format and its dependencies.

**DROID** by [adrian\\_brown](#), [davidunderdown](#), [fetherston](#), [mattpalmer1086](#), [robgtaylor](#), [rspencer1](#)

Summary Files Reviews Support Develop Hosted Apps Mailing Lists

Looking for the latest version? [Download droid-6.01.zip \(35.4 MB\)](#)

Home

Name	Modified	Size
<b>droid</b>	2011-02-04	
<b>droid-src</b>	2011-02-04	
<b>droid-technical</b>	2010-05-18	

Totals: 3 Items

# Finding Tools - DROID

The screenshot shows a search result for 'DROID' on SourceForge. The page header includes the SourceForge logo, a search bar with 'Find Open Source Software', and navigation links for Browse, Blog, Support, Jobs, Newsletters, and Resources. A 'Register' and 'Log In' link is also present. A sidebar on the left contains search filters like 'Everything', 'Images', 'Maps', 'Videos', 'News', 'Shopping', and 'More'. The main content area features a large green 'Download Now!' button with a downward arrow, powered by iLivid. Below this is a 'Featured Download' for Microsoft Office 365, with a 'Free Trial' button. To the right, there are two advertisements: one for QuickBooks Online with a '30 Day Free Trial' badge and a 'Try It Free' button, and another for BPMonline CRM with a starburst graphic and the text 'BPMonline CRM wins CRM Idol 2001!'. A footer at the bottom indicates 'Totals: 3 Items'.

Google SOURCEFORGE Find Open Source Software Browse Blog Support Jobs Newsletters Resources Register Log In

Search

Everything  
Images  
Maps  
Videos  
News  
Shopping  
More

Southampton, Change location

The web  
Pages from the

All results  
Related searches  
More search tools

Your **DROID** download will start in 3 seconds...  
Problems with the download? Check your browser's security bar, or try a [direct link](#), or try another [mirror](#).

Mirror provided by: University of Kent UKMIRROR service

**Download Now!**  
Powered by iLivid

Featured Download

**Introducing Microsoft Office 365**  
Work together in the cloud – with Office, Exchange, SharePoint and Lync video conferencing, it all works together.

Free Trial

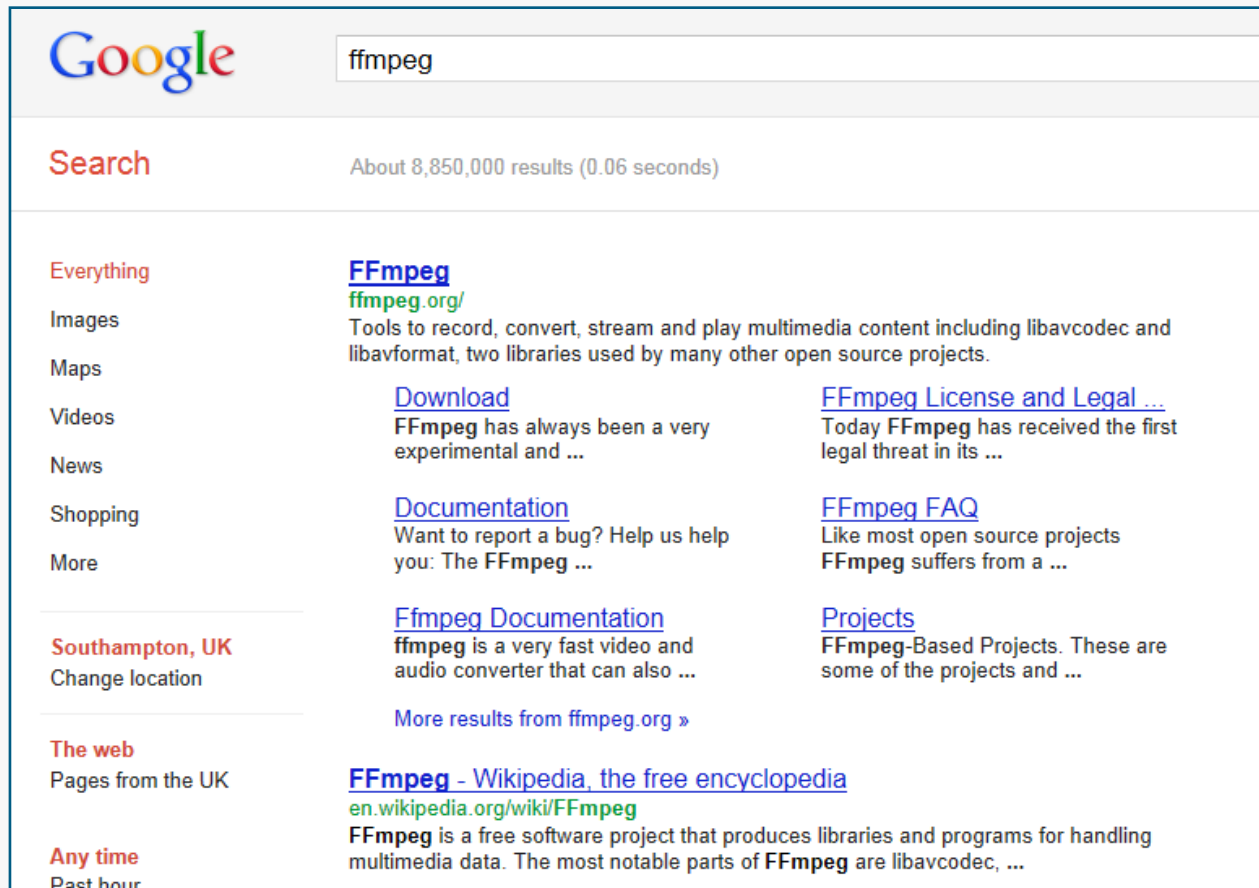
Time saving accounting software  
**QuickBooks Online**  
30 Day Free Trial  
Try It Free  
intuit

- ✓ Easy invoicing
- ✓ Track sales & expenses
- ✓ Anywhere, anytime access

**CRM Idol**  
BPMonline CRM wins CRM Idol 2001!

Totals: 3 Items

# Finding Tools - FFmpeg



The image shows a screenshot of a Google search for "ffmpeg". The search bar at the top contains the text "ffmpeg". Below the search bar, the word "Search" is displayed in red, followed by the text "About 8,850,000 results (0.06 seconds)".

On the left side, there is a vertical menu with the following items: "Everything", "Images", "Maps", "Videos", "News", "Shopping", "More", "Southampton, UK", "Change location", "The web", "Pages from the UK", "Any time", and "Past hour".

The main search results are as follows:

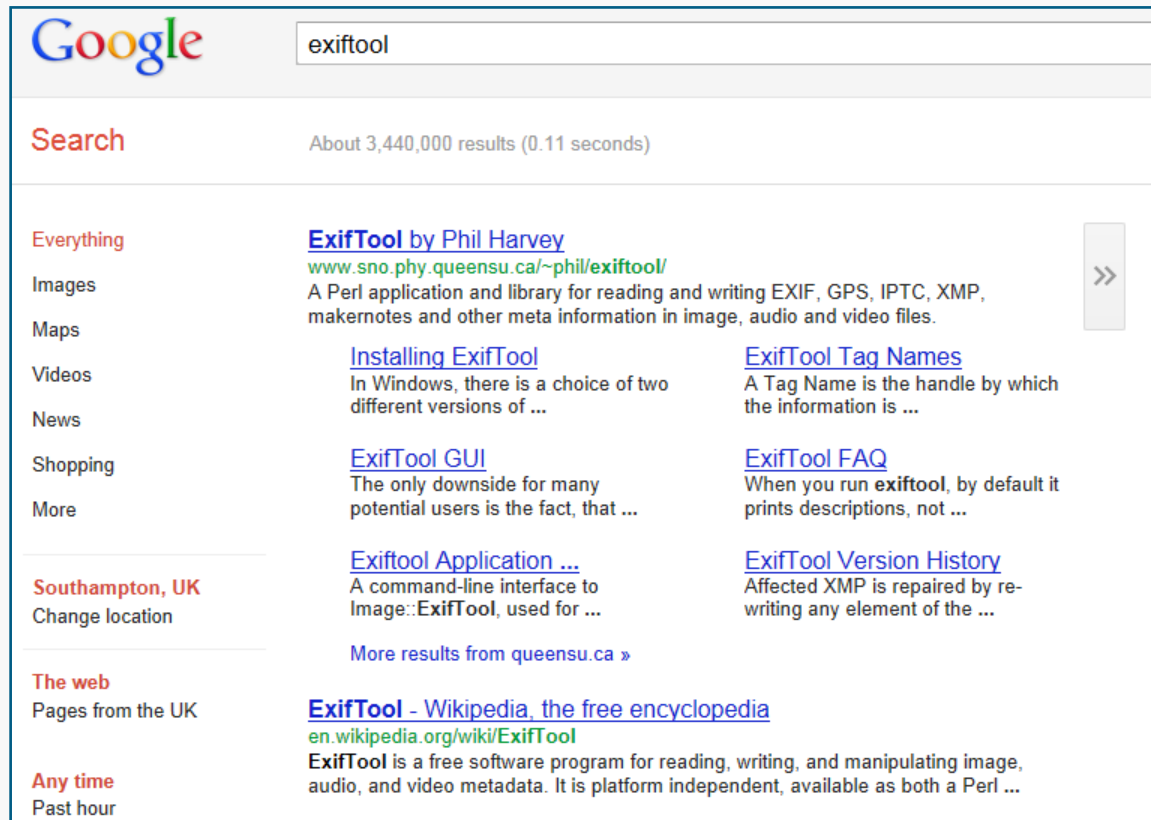
- FFmpeg**  
[ffmpeg.org/](http://ffmpeg.org/)  
Tools to record, convert, stream and play multimedia content including libavcodec and libavformat, two libraries used by many other open source projects.
- [Download](#)  
FFmpeg has always been a very experimental and ...
- [FFmpeg License and Legal ...](#)  
Today FFmpeg has received the first legal threat in its ...
- [Documentation](#)  
Want to report a bug? Help us help you: The FFmpeg ...
- [FFmpeg FAQ](#)  
Like most open source projects FFmpeg suffers from a ...
- [Fmpeg Documentation](#)  
ffmpeg is a very fast video and audio converter that can also ...
- [Projects](#)  
FFmpeg-Based Projects. These are some of the projects and ...

At the bottom of the results, there is a link: [More results from ffmpeg.org »](#)

Below the results, there is a section for "The web" with the text "Pages from the UK" and a link: [FFmpeg - Wikipedia, the free encyclopedia](#)  
[en.wikipedia.org/wiki/FFmpeg](http://en.wikipedia.org/wiki/FFmpeg)  
FFmpeg is a free software project that produces libraries and programs for handling multimedia data. The most notable parts of FFmpeg are libavcodec, ...



# Finding Tools – Exiftool



The image shows a Google search interface for the term "exiftool". The search bar at the top contains the text "exiftool". Below the search bar, the word "Search" is displayed in red, followed by the text "About 3,440,000 results (0.11 seconds)".

On the left side, there is a vertical navigation menu with the following items: "Everything", "Images", "Maps", "Videos", "News", "Shopping", and "More". Below this menu, there are three sections: "Southampton, UK" with a "Change location" link, "The web" with a "Pages from the UK" link, and "Any time" with a "Past hour" link.

The main search results area displays the following:

- ExifTool by Phil Harvey**  
[www.sno.phy.queensu.ca/~phil/exiftool/](http://www.sno.phy.queensu.ca/~phil/exiftool/)  
A Perl application and library for reading and writing EXIF, GPS, IPTC, XMP, makernotes and other meta information in image, audio and video files.
- Installing ExifTool**  
In Windows, there is a choice of two different versions of ...
- ExifTool Tag Names**  
A Tag Name is the handle by which the information is ...
- ExifTool GUI**  
The only downside for many potential users is the fact, that ...
- ExifTool FAQ**  
When you run **exiftool**, by default it prints descriptions, not ...
- Exiftool Application ...**  
A command-line interface to Image::ExifTool, used for ...
- ExifTool Version History**  
Affected XMP is repaired by re-writing any element of the ...

At the bottom of the results, there is a link: "More results from queensu.ca »".

# Finding Tools – Exiftool

The image shows a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text 'exiftool'. Below the search bar, the word 'Search' is displayed in red, followed by the text 'About 3,440,000 results (0.11 seconds)'. On the left side, there is a vertical menu with the following items: 'Everything' (selected), 'Images', 'Maps', 'Videos', 'News', 'Shopping', and 'More'. Below this menu, there are three sections: 'Southampton, UK' with a 'Change location' link, 'The web' with a 'Pages from the UK' link, and 'Any time' with a 'Past hour' link. The main content area on the right is partially obscured by a blue-bordered box containing text about installing ExifTool.

## Installing ExifTool

See the appropriate section below with instructions for installing or uninstalling ExifTool on your specific platform:

- [Windows](#)
- [Macintosh OS X](#)
- [Unix Platforms](#)

Also see [these instructions](#) for help running ExifTool.

## Windows

In Windows, there is a choice of two different versions of ExifTool to install. The Perl distribution requires Perl to be installed on your system. (A good, free Perl interpreter can be downloaded from [activeperl.com](#).)

If you don't already have Perl, it is easier to install the stand-alone ExifTool executable, but note that the stand-alone version doesn't include the HTML documentation or some other files of the full distribution.

### Stand-Alone Executable

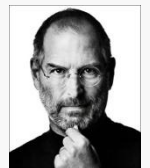
1. **Download** the **Windows Executable** from the [ExifTool home page](#).  
(The file you download will have a name like "exiftool-#.##.zip".)
2. **Extract** "**exiftool(-k).exe**" from the ".zip" file, and place it **on your Desktop**.

Tools **NEED** wide support, both in terms of adoption and development from the community.

Tools NEED wide support, both in terms of adoption and development from the community.

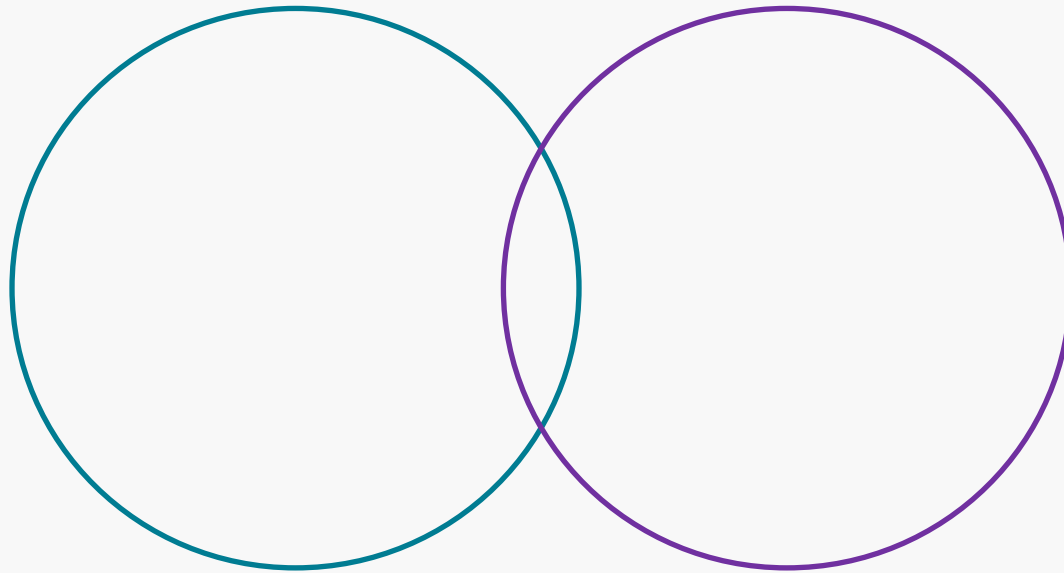
(development part is hard, it helps if the project has a ~~dictator~~ clear minded individual as a leader)

“Something is only worth doing if you are going to be the best at it!”



# Tools and Projects

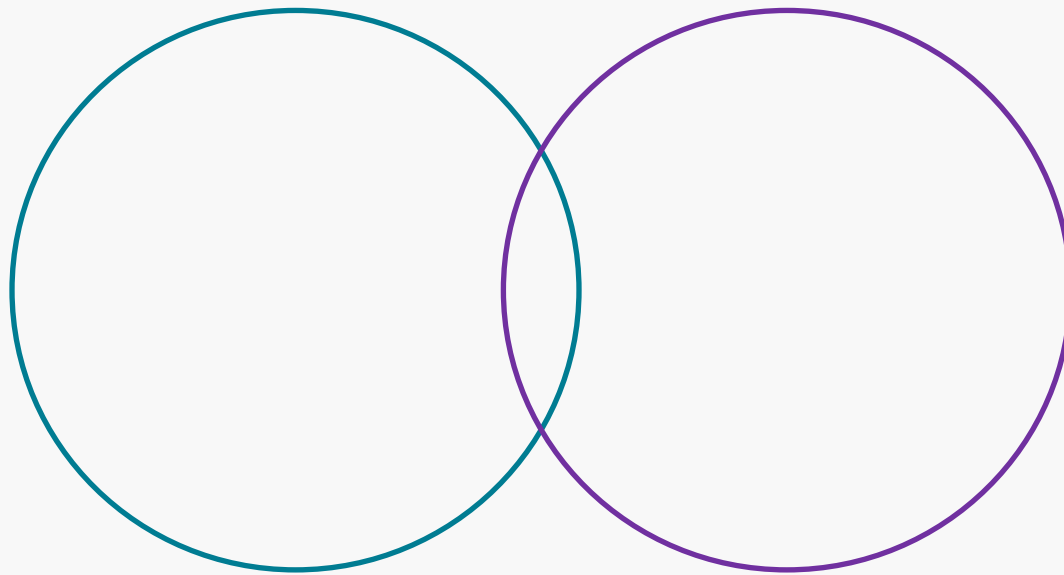
...



Tools produced by Academic Digital  
Preservation Projects (the experts)

# Tools and Projects

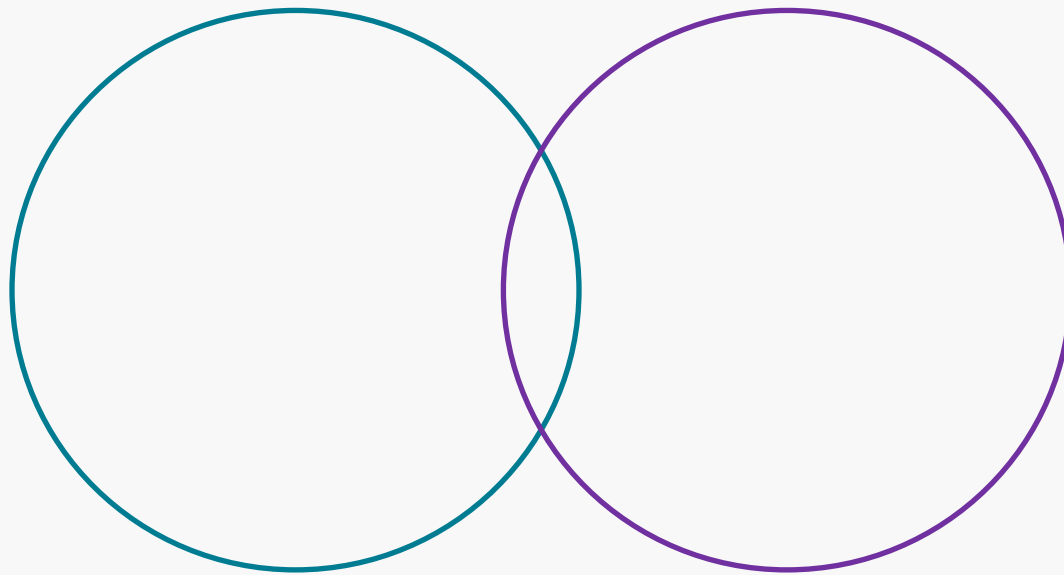
Tools **supported** by the global open source community



Tools produced by Academic Digital Preservation Projects (the experts)

# Tools and Projects

Tools **known** by the global open source community?



Tools produced by Academic Digital Preservation Projects (the experts)

# It is hard to preserve:

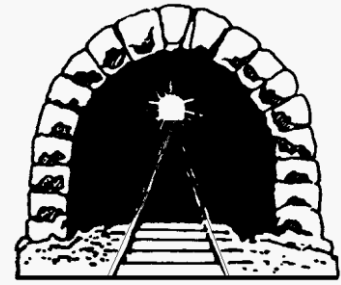
- Complex stuff
- Stuff with hard to preserve dependencies
- Stuff with too many dependencies
- Stuff which is tied to specific dependencies and hard to re-create scenarios
- Stuff of which the use and function is unclear





How good is the Digital Preservation community at listening to itself?

# Light at the end of the tunnel



- Separating R from D
- Digital Preservation Research is Essential.
- Development on well defined, crafted tools is essential.
- These two do not marry together.



# UHT Digital Preservation

- Free form tool development will produce a full fat tool.
- Full fat is bad for you and has a very short shelf life.
- Removing the fat (via UHT) makes it taste less good but lasts longer.
- Nothing to stop there being many flavours and brands of UHT! (as long as they are not created by taking an existing brand and adding to it, you end up with full fat again)

# Distilled Digital Preservation

- PRONOM – The best registry for data in the world, open data, good APIs, simple to understand, **limited scope, widely used.**
- DROID – **Widely used** but too complex.
- Other tools – Research or Development?

# Call to Action

- **Simplify** – define in a sentence
- **Inform** – google, wikipedia, short documentation
- **Modularise** - separate apps
- **Package** – windows, os x, linux
- **Lighten** – reduce the number of use cases
- **Enrich** – don't be the main application
- **Support** – make upgrades easy



# The Good News

- The Digital Preservation Community is Growing
- The number of people actively using the tools is growing
- Awareness training has had a great affect.

# Support



- Things are still changing, old software, libraries and signature files are a preservation risk!
- More file formats, more data, more guidance.
- Risk analysis needs work. (Not much research in this area).

# OpenPlanets Aims

- Identify and support lightweight tools for digital preservation
- Support those who are developing such tools
- Improve user experiences by educating and communicating with users and developers



# Connecting Projects to Data

- OpenPlanets REF
- Extension of the P2-Registry, collected data is input, collated and processed to help improve both tools and user experience.
- Intending to collect data and preserve data from digital preservation projects.

# Connecting Tools to Users

- Support the packaging and support mechanisms for tools
- Develop a new connection tool which can utilise existing tools and provide more focussed preservation information (MUPPET and OPFFS are the realisation of this)

# Conclusion

- Improve Presence
- Lighten Tools
- Support the global community
- Digital Preservation tools should be digitally preservable
- Support users not projects