

# KEEP

KEEPING EMULATION ENVIRONMENTS PORTABLE

## Emulation: the debate, registries and PREMIS

DPC 22<sup>nd</sup> July 2014, Cambridge

**David Anderson and Janet Delve**

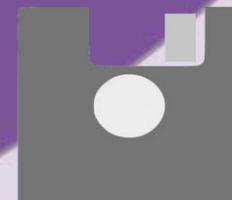
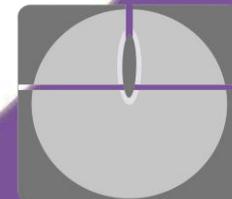
University of Portsmouth Future Proof Computing Group, UK



University of  
**Portsmouth**



FUTURE proof  
COMPUTING  
GROUP



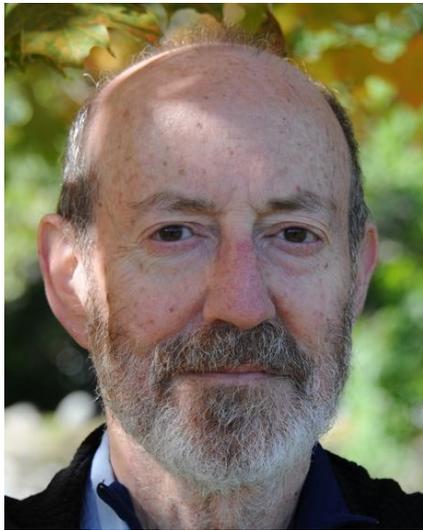
**TOTem**

Trustworthy Online Technical Environment Metadata

# Outline

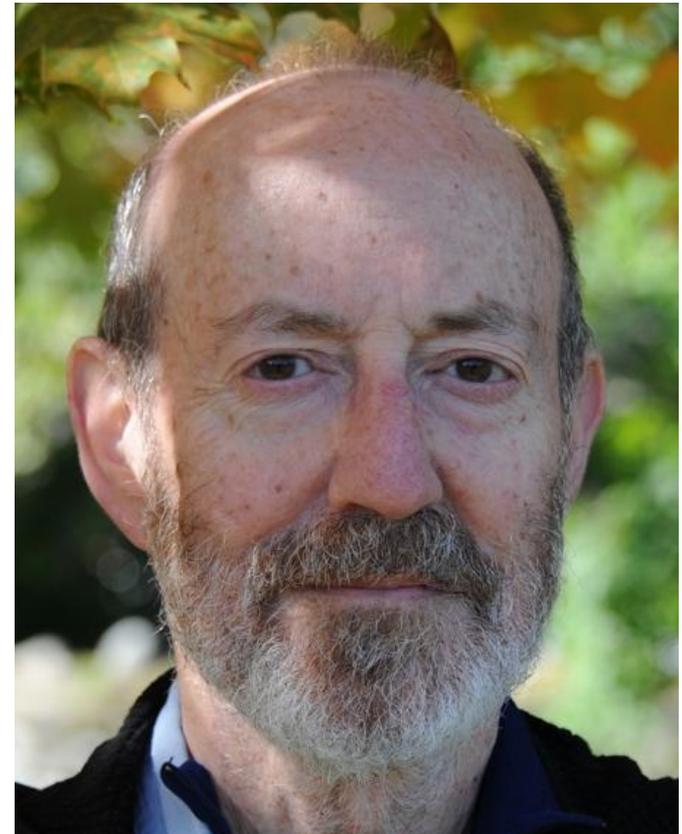
- Emulation: the debate (DA)
- The TOTEM technical registry (JD)
- PREMIS Environment Working Group (JD)
- Proposed technical registry (DA)

# Migration or Emulation?



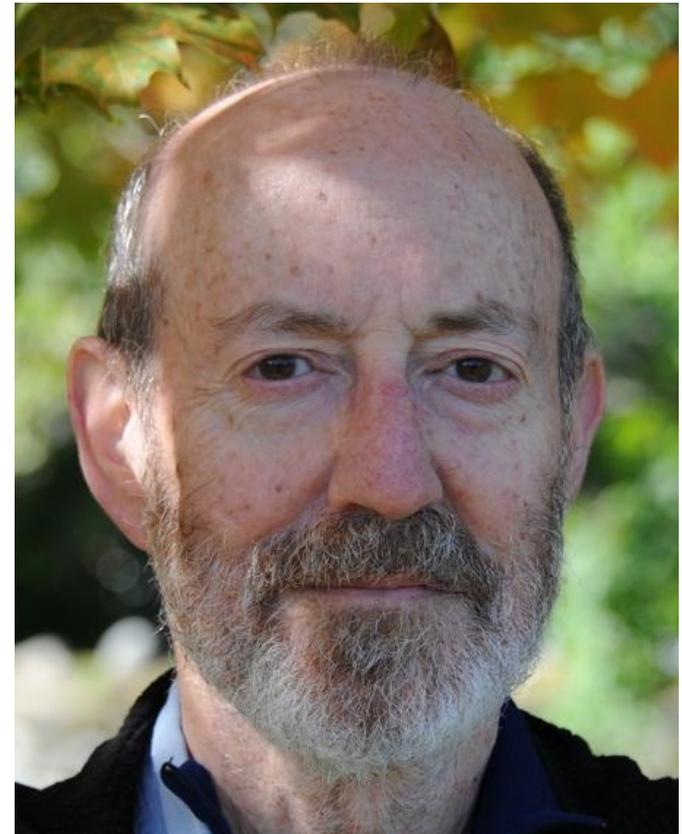
# Rothenberg's considered view on Migration

“While it may be better than nothing (better than having no strategy at all or denying that there is a problem), it has little to recommend it. .... however, to the extent that it provides merely the illusion of a solution, it may in some cases actually be worse than nothing. In the long run, migration promises to be expensive, unscalable, error-prone, at most partially successful, and ultimately infeasible.”  
(Rothenberg, 1999)



# Labour intensive and error-prone!

“migration is labor-intensive, time-consuming, expensive, error-prone, and fraught with the danger of losing or corrupting information. Migration requires a unique new solution for each new format or paradigm and each type of document that is to be converted into that new form. Since every paradigm shift entails a new set of problems, there is not necessarily much to be learned from previous migration efforts, making each migration cycle just as difficult, expensive, and problematic as the last. Automatic conversion is rarely possible, ... it is very likely to result in at least some loss or corruption, as documents are forced to fit into new forms.” (Rothenberg, 1999)

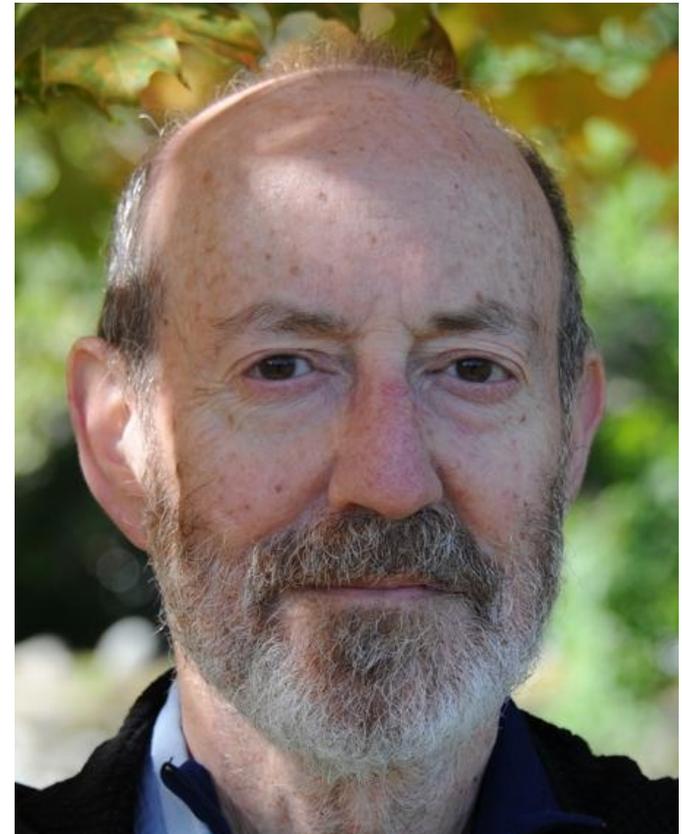


# Labour intensive and error-prone!

Emulation also requires considerable expenditure of time and effort in order to arrive at a successful outcome.

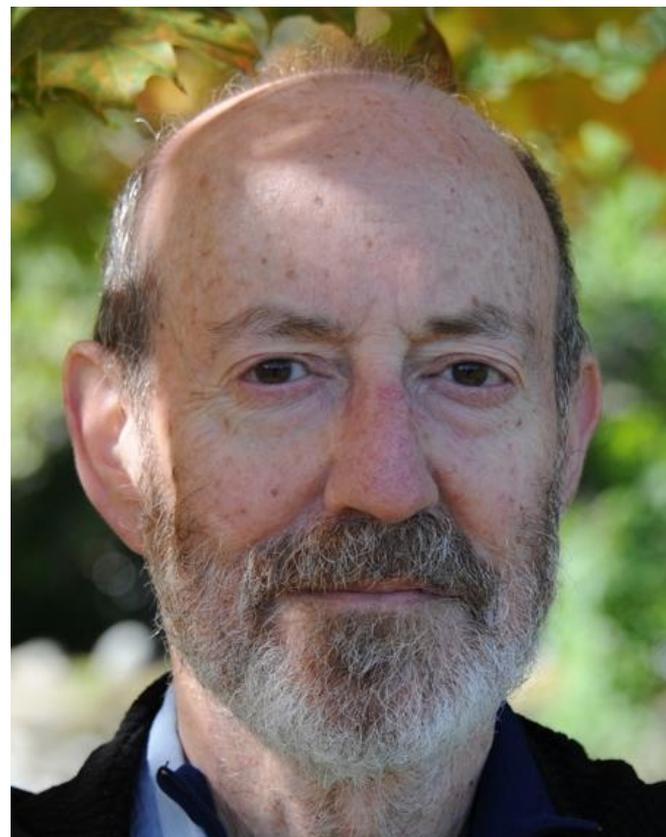
It is true that a great deal excellent work has been undertaken in the emulation community, which has provided benefits for the digital preservation community without giving rise thereby to any outlay of resources by museums, libraries or archives. But this software windfall should not be allowed to engender complacency.

Robust emulation software remains “labor-intensive, time-consuming, expensive” to develop.



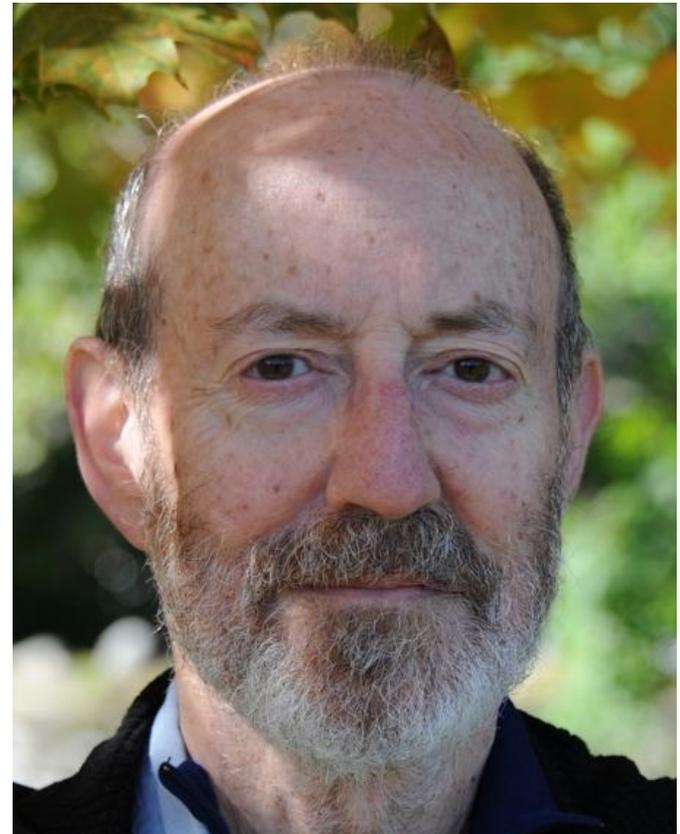
# Paradigms change!

“As has been proven repeatedly during the short history of computer science, formats, encodings, and software paradigms change often and in surprising ways. Of the many dynamic aspects of information science, document paradigms, computing paradigms, and software paradigms are among the most volatile, and their evolution routinely eludes prediction.” (Rothenberg, 1999)



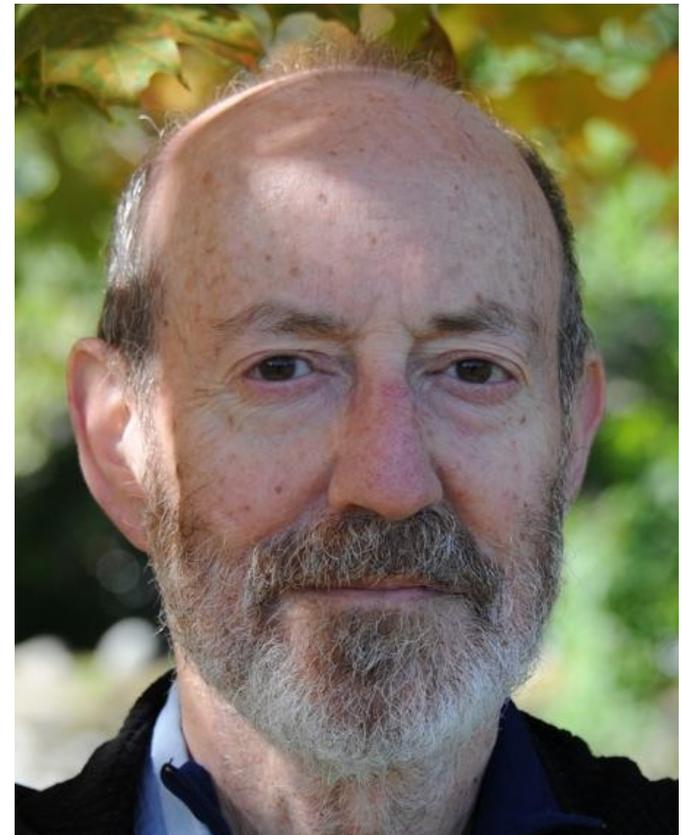
# Paradigms change!

There is nothing about emulation in and of itself, which makes it immune to the disruption caused by the introduction of new approaches to computing and the inevitable obsolescence of the old.



# Migration (unlike emulation) involves urgency.

“... there is a degree of urgency involved in migration. If a given document is not converted when a new paradigm first appears, even if the document is saved in its original form (and refreshed by being copied onto new media), the software required to access its now-obsolete form may be lost or become unusable due to the obsolescence of the required hardware, making future conversion difficult or impossible.”

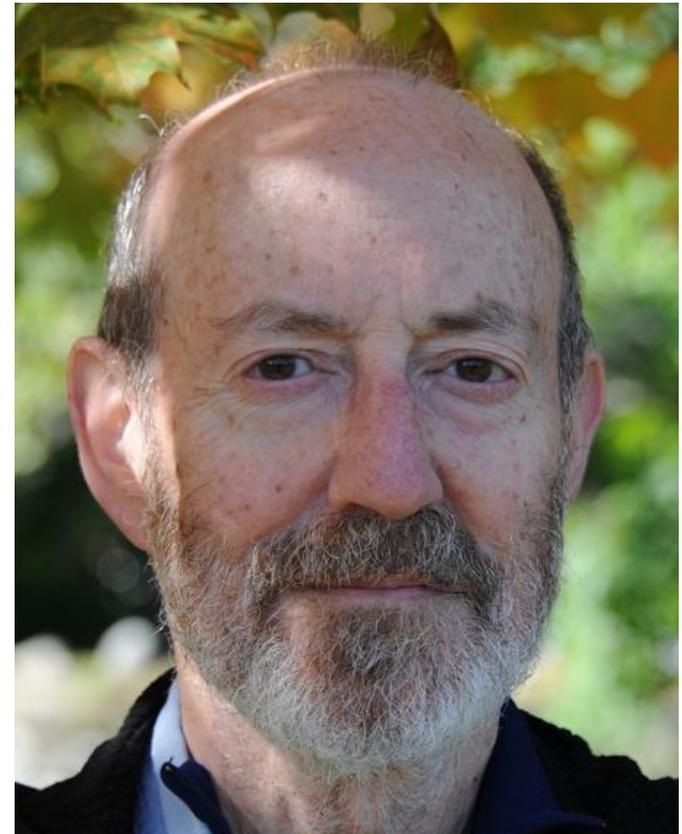


# Migration (unlike emulation) involves urgency.

The introduction of a new hardware paradigm would mean that every emulator written to run on the previous paradigm would no longer function on the new device. This would leave us with two options:

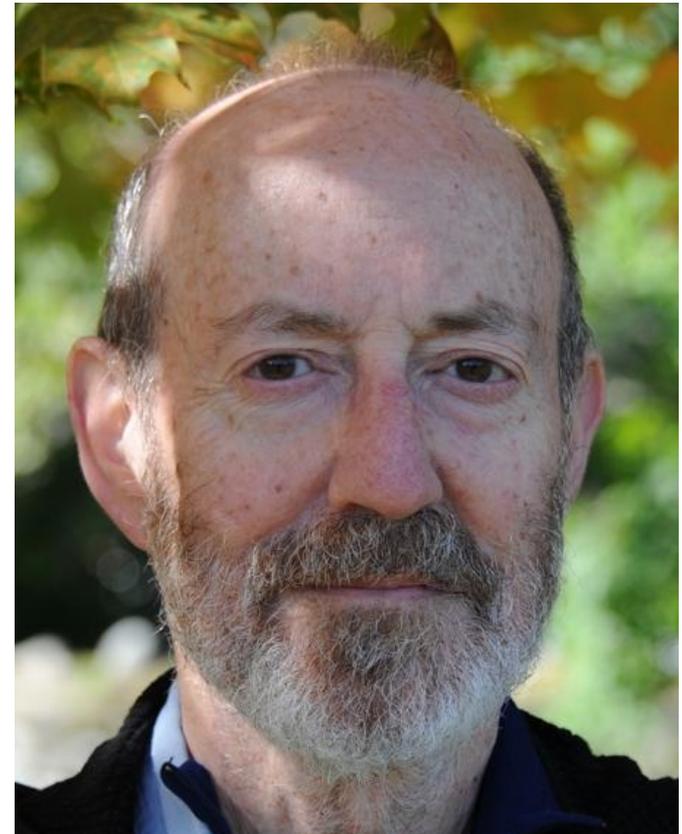
Write new emulators

Migrate the old emulators to run on the new platform



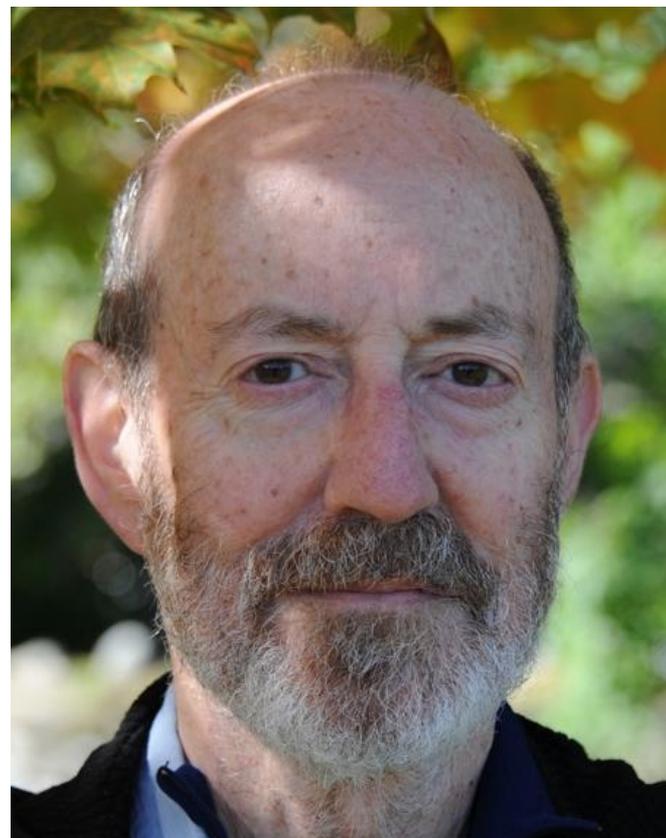
# Migration (unlike emulation) is an ongoing activity.

“Worse yet, this problem does not occur just once for a given document (when its original form becomes obsolete) but recurs throughout the future, as each form into which the document has migrated becomes obsolete in turn. Furthermore, because the cycles of migration that must be performed are determined by the emergence of new formats or paradigms, which cannot be controlled or predicted, it is essentially impossible to estimate when migration will have to be performed for a given type of document—the only reliable prediction being that any given type of document is very likely to require conversion into some unforeseeable new form within some random (but probably small) number of years.” (Rothenberg, 1999)



Migration (unlike emulation) is an ongoing activity.

It is to be expected that the mean time between preservation interventions is shorter with some preservation strategies than others. It is difficult to see how this could be proven in advance or determined with any great accuracy but if this information were available it would be valuable (but not decisive) in helping to determine the strategy adopted by individual institutions.



# TOTEM

# KEEP Overview.

# TOTEM

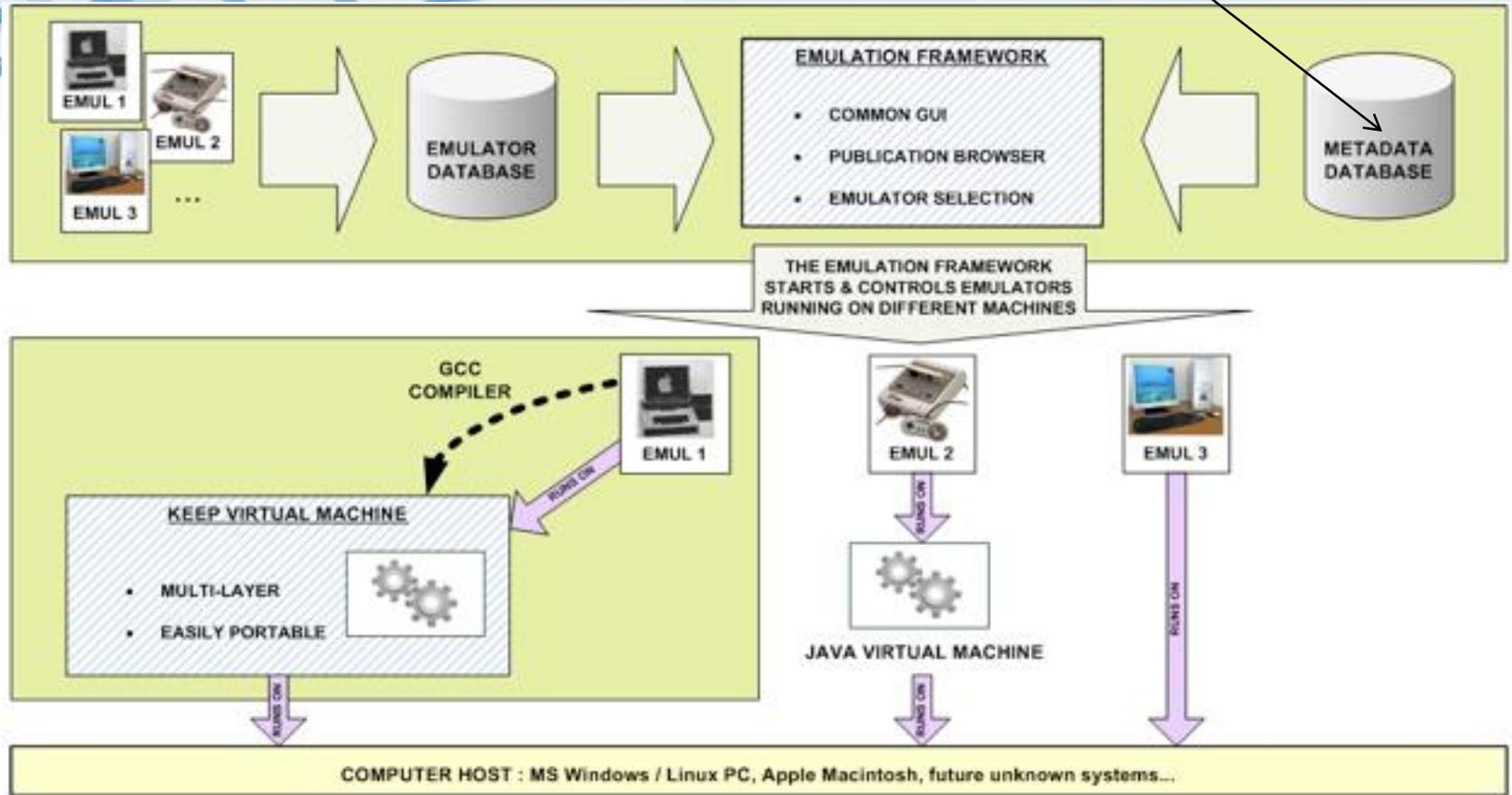


Figure 1: New integration strategy schema

# The TOTEM Registry

- Trustworthy Online Technical Environment Metadata Database / Registry
- TOTEM contains technical environment metadata: S/W, O/S, H/W compatible *versions*
- Suite of data models and Use cases
- Online MySQL database
- Metadata schema, XML
- Book – series editor Professor Manfred Thaller
- Contribution to PREMIS Environment Working

## Verlagsprogramm

[Übersicht](#)

[Fachgebiete](#)

[Kommunikations- &  
Medienwissenschaften](#)

[Informatik](#)

[Kölner Beiträge zu einer  
geisteswissenschaftlichen  
Fachinformatik](#)



**Janet Delve & David Anderson**

**The Trustworthy Online Technical  
Environment Metadata Database –  
TOTEM**

**- in englischer Sprache -**

**Kölner Beiträge zu einer  
geisteswissenschaftlichen Fachinformatik,  
Band 4**

**Hamburg 2012, 210 Seiten,  
ISBN 978-3-8300-6418-3**

# TOTEM Home page



FUTURE proof  
COMPUTING  
GROUP

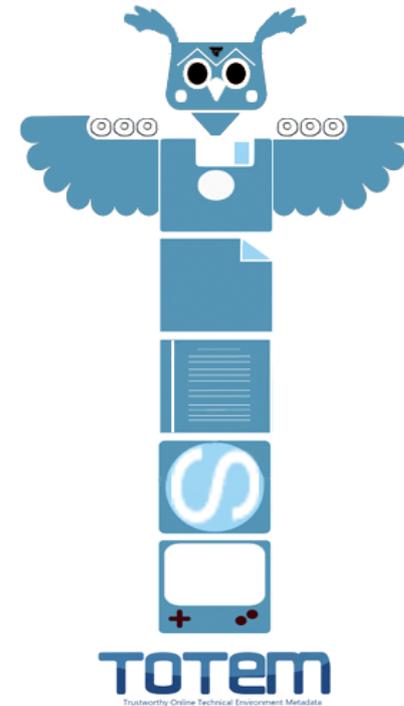
[Home](#) [Sign in](#) [Register](#) [Contact](#)

## Welcome to TOTEM - the Trustworthy Online Technical Environment Metadata Registry

Totem poles are monuments created by First Nations of the Pacific Northwest to represent and commemorate ancestry, histories, people, or events, and would be erected to be visible within a community.

The TOTEM registry is used, analogously, to record the complex hardware and software relationships which apply to digital objects, and to make them visible to the digital preservation community.

TOTEM was created by the University of Portsmouth, Future Proof Computing Group



# TOTEM Simple Search



FUTURE proof  
COMPUTING  
GROUP

[Home](#) | [PC Architecture](#) | [C64 Architecture](#) | [Console Architecture](#) | [Sign out](#)

TOTEM > Console Architecture > Hardware Search > Summary for "Nintendo Entertainment System (NES)"

## Summary for: [Nintendo Entertainment System \(NES\)](#)

Look at: [Summary](#) || [Console Versions](#)

**Name:** Nintendo Entertainment System (NES)

**Manufacturer:** Nintendo

**Type:** Video game console (3rd generation)

University of Portsmouth, Future Proof Computing Group (c) 2008 - 2012



KEEPING EMULATION ENVIRONMENTS PORTABLE

# TOTEM Super Mario Bros. Versions



FUTURE proof  
COMPUTING  
GROUP

[Home](#) | [PC Architecture](#) | [C64 Architecture](#) | [Console Architecture](#) | [Sign out](#)

TOTEM > Console Games > Software Search > Software Versions for "Super Mario Bros. 3"

## Software Versions for: Super Mario Bros. 3

Look at: [Summary](#) || [Software Versions](#)

### Version Name: SMB3 European Version

**Description:** Super Mario Bros. 3, also referred to as Super Mario 3 and SMB3, is a platform video game developed and published by Nintendo for the Nintendo Entertainment System (NES), and is the third game in the Super Mario series.

**Release Date:** 1991-08-21

**Source:** [http://en.wikipedia.org/wiki/Super\\_Mario\\_Bros.\\_3](http://en.wikipedia.org/wiki/Super_Mario_Bros._3)

### Version Name: SMB3 Japanese Version

**Description:** The original Japanese version of this game featured several differences from the version of the game that would later be released in North America and Europe. In the Japanese game, Mario always reverts back to small Mario when he is hit, but in the American version he'll change to Super Mario if he is hit while he has a power-up. Peach's joke at the end of the American/PAL version was added in (the original Japanese text simply said, "Thank you! Finally, peace returns to the Mushroom World. The end!") The Japanese version of the game featured different levels names than (most) versions of the American/PAL game (Desert Hill became Desert Land, Ocean Side became Water Land, Big Island became Giant Land, The Sky became Sky Land, Iced Land because Ice Land, Pipe Maze became Pipe Land, and Castle of Kuppia became Dark Land). There were other small changes made as well to specific levels (the end of the Grass Land fortress was slightly altered, as was a level in World 5)

**Release Date:** 1988-10-23

**Source:** [http://en.wikipedia.org/wiki/Super\\_Mario\\_Bros.\\_3](http://en.wikipedia.org/wiki/Super_Mario_Bros._3)

### Version Name: SMB3 North American Version

**Description:** The original Japanese version of this game featured several differences from the version of the game that would later be released in North America and Europe. In the Japanese game, Mario always reverts back to small Mario when he is hit, but in the American version he'll change to Super Mario if he is hit while he has a power-up. Peach's joke at the end of the American/PAL version was added in (the original Japanese text simply said, "Thank you! Finally, peace returns to the Mushroom World. The end!") The Japanese version of the game featured different levels names than (most) versions of the American/PAL game (Desert Hill became Desert Land, Ocean Side became Water Land, Big Island became Giant Land, The Sky became Sky Land, Iced Land because Ice Land, Pipe Maze became Pipe Land, and Castle of Kuppia became Dark Land). There were other small changes made as well to specific levels (the end of the Grass Land fortress was slightly altered, as was a level in World 5)

**Release Date:** 1990-12-02

**Source:** [http://en.wikipedia.org/wiki/Super\\_Mario\\_Bros.\\_3](http://en.wikipedia.org/wiki/Super_Mario_Bros._3)

University of Portsmouth, Future Proof Computing Group (c) 2008 - 2012



KEEPING EMULATION ENVIRONMENTS PORTABLE

# TOTEM Compatibility Search (C64, S/W – file)



[Home](#)

[PC Architecture](#) ▾

[C64 Architecture](#) ▾

[Console Games](#) ▾

[Sign out](#)

*CiuffreA*

TOTEM > C64 Architecture > Compatibility Search (Software) > Search Results for "Commodore Executable Program in BASIC (PRG)" and "PRG"

Search Results for: **Commodore Executable Program in BASIC (PRG) and PRG**

**Software Name:** SpeedScript

**Software Version:** SpeedScript 3.0

**Release Date:** 1985-10-01



KEEPING EMULATION ENVIRONMENTS PORTABLE

# TOTEM Drop down Menus

The screenshot shows a web browser window with a dropdown menu open. The menu is titled 'TOTEM > C' and lists the following items:

- 1292 Advanced Programmable Video System
- 1292 Advanced Programmable Video System
- 1392 Advanced Programmable Video System
- 1392 Advanced Programmable Video System
- 3DO Interactive Multiplayer
- Advanced Programmable Video System
- Amiga CD32
- Arcadia 2001
- Astrocade
- Atari 5200
- Atari 7800
- Atari Jaguar
- ColecoVision** (highlighted)
- Database (Videomaster)
- Database (Waddington/Voltmace)
- Dreamcast
- Family Computer (Famicom)
- FM Towns Marty
- G7000
- Game Boy

The background of the page includes the University of Portsmouth logo and the text 'FUTURE proof COMPUTING GROUP'. Other visible elements include 'Home', 'Console Architecture', 'Sign out', 'UPHEC', and a 'Submit Query' button.

# TOTEM RDF mapping – University of Cologne

- TOTEM team and the University of Cologne (Johanna Puhl) have shown in a proof-of-concept that emulation metadata from the TOTEM database can be mapped into an OWL/RDF-presentation of preservation metadata resulting from the Planets-project: the Planets ontology, which can be queried via SPARQL.

# PREMIS

# Describing Digital Object Environments in PREMIS (June 2011-)

**Angela Dappert**

Digital Preservation Coalition

**Sébastien Peyrard**

National Library of France

**Janet Delve**

The University of Portsmouth

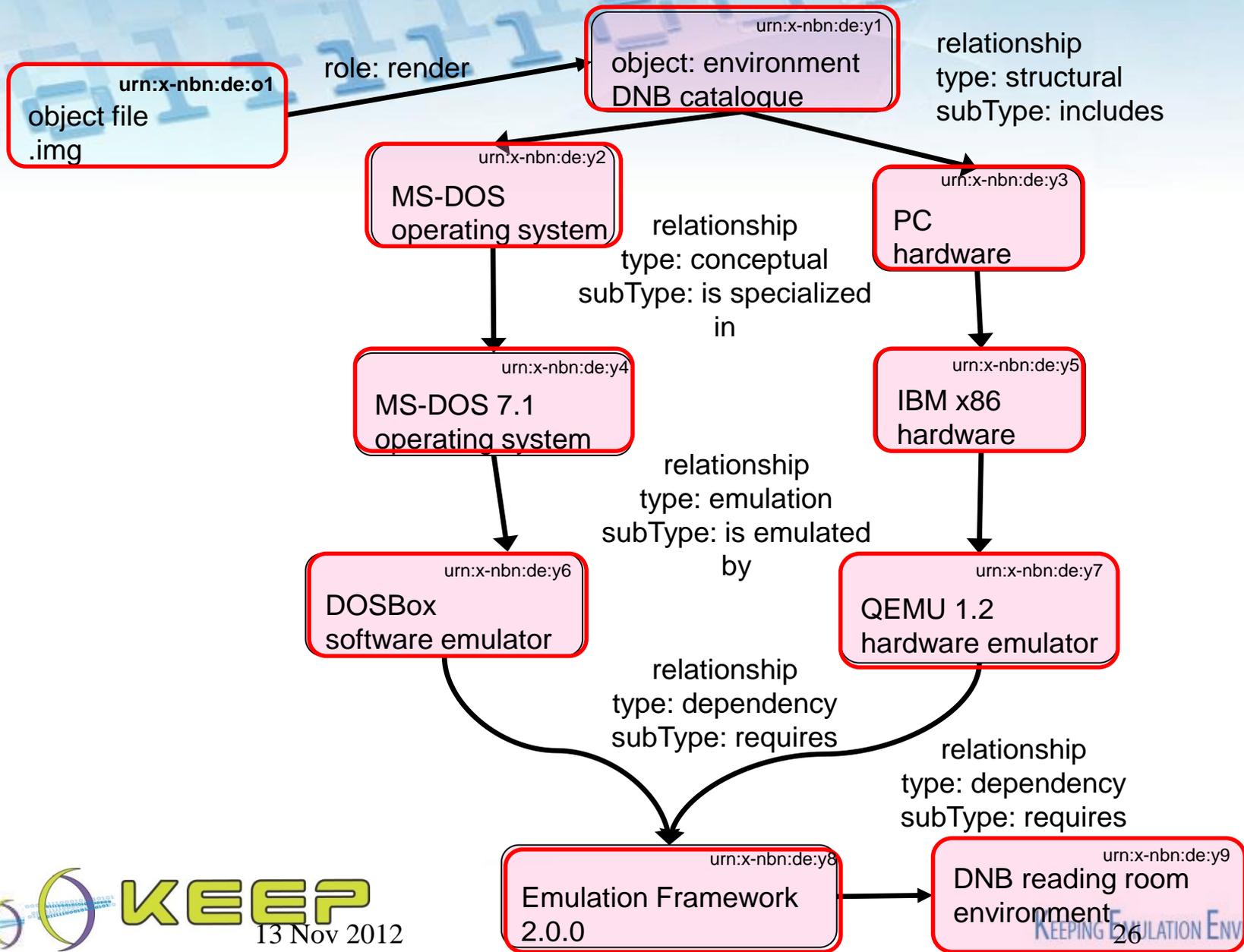
**Carol C.H. Chou**

Florida Digital Archive

## Use Case 3: Environment Used for emulation preservation action



- Example from DNB for EC KEEP project. Digital Object is radar simulation for racing boat training package (1999).
- Vague systems requirements in catalogue metadata:  
**PC (hardware) and MSDOS (operating system)**



# Use Case 3: Environment Used for emulation preservation action

- For emulation there are two issues: we need **versions**, and they need to be **compatible**. These can be found in **TOTEM**  
<http://www.keep-totem.co.uk/> .
- Several iterations?
- **Hardware emulators** and if necessary, **software emulators** can then be specified.

# Use Case 3: Environment Used for emulation preservation action

- We need an **emulation platform**, e.g. KEEP Emulation framework (EF) <http://emuframework.sourceforge.net/> to run these emulators. (Also bwFLA)
- Finally need **reading room environment to run this EF.**

## Welcome to the Emulation Framework project

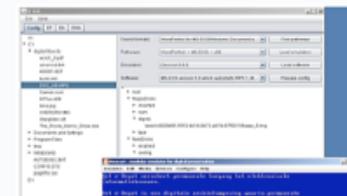
The Emulation Framework is software developed by the international KEEP project, co-funded by the European Union's 7th Framework Programme. It is available as open source software under the Apache 2.0 license.

## EF nominated for DPC Award

**8 Oct 2012** - The Emulation Framework is nominated for the R&D award of the Digital Preservation Coalition (DPC). The project is one of the four on the shortlist for the award. The DPC award is meant for groundbreaking new developments in the field of digital preservation leading to new innovations and helping organisations to deal with the challenges that our digital era poses to us. On 3rd of Dec 2012 the winner for the award will be announced. Before then the judges will take a closer look to the project and its outcomes and there will be an online voting. We will let you know how you can vote for this project. On the 30th

## In this release (2.1.0)

- × 6 platforms supported: x86, C64, Amiga, BBC Micro, Amstrad, Thomson TO7
- × 7 emulators included: Dioscuri, Qemu, VICE, UAE, BeebEm, JavaCPC, Thomson
- × 30+ file formats supported: PDF, TXT, XML, JPG, TIFF, PNG, BMP, Quark, ARJ, EXE, disk/tape images and more
- × Integration with format identification tool FITS
- × Web services for software and emulator archives



# Link to Technical Registry TOTEM

*{version of original hardware platform as located in environment registry that is compatible with the software version chosen above}*

environmentIdentifier (M, R)

environmentIdentifierType (M, NR): **URN**

environmentIdentifierValue (M, NR): **urn:x-nbn:de:y3 (not real identifier)**

environmentDescription (O, R)

environmentName (O, NR): **PC**

environmentVersion (O, NR): **IBM x86** *{chosen as it was current in 1996 and is compatible with MSDOS 7.1}*

environmentRegistry (O, R)

environmentRegistryName (M, NR): **TOTEM** *{hardware is not in PRONOM}*

environmentRegistryKey (M, NR): **TUID-xxxx**

environmentRegistryRole (O, NR): **external**

environmentNote (O, R): **Developed 1991.**

# Emulation Complexity

- **Version details** are vital for SW, OS, HW etc.
- There are **complex interdependencies** between SW/OS/HW
- We need **iterative technical registry calls** to determine these.
- We have **stacked environments**.
- Finding emulation information is not straightforward, so it is imperative we **keep emulator details**.
- Running Emulation Frameworks is new for memory institutions, and is complicated, so important to record **reading room environment details** where this happens.

# Proposed technical registry

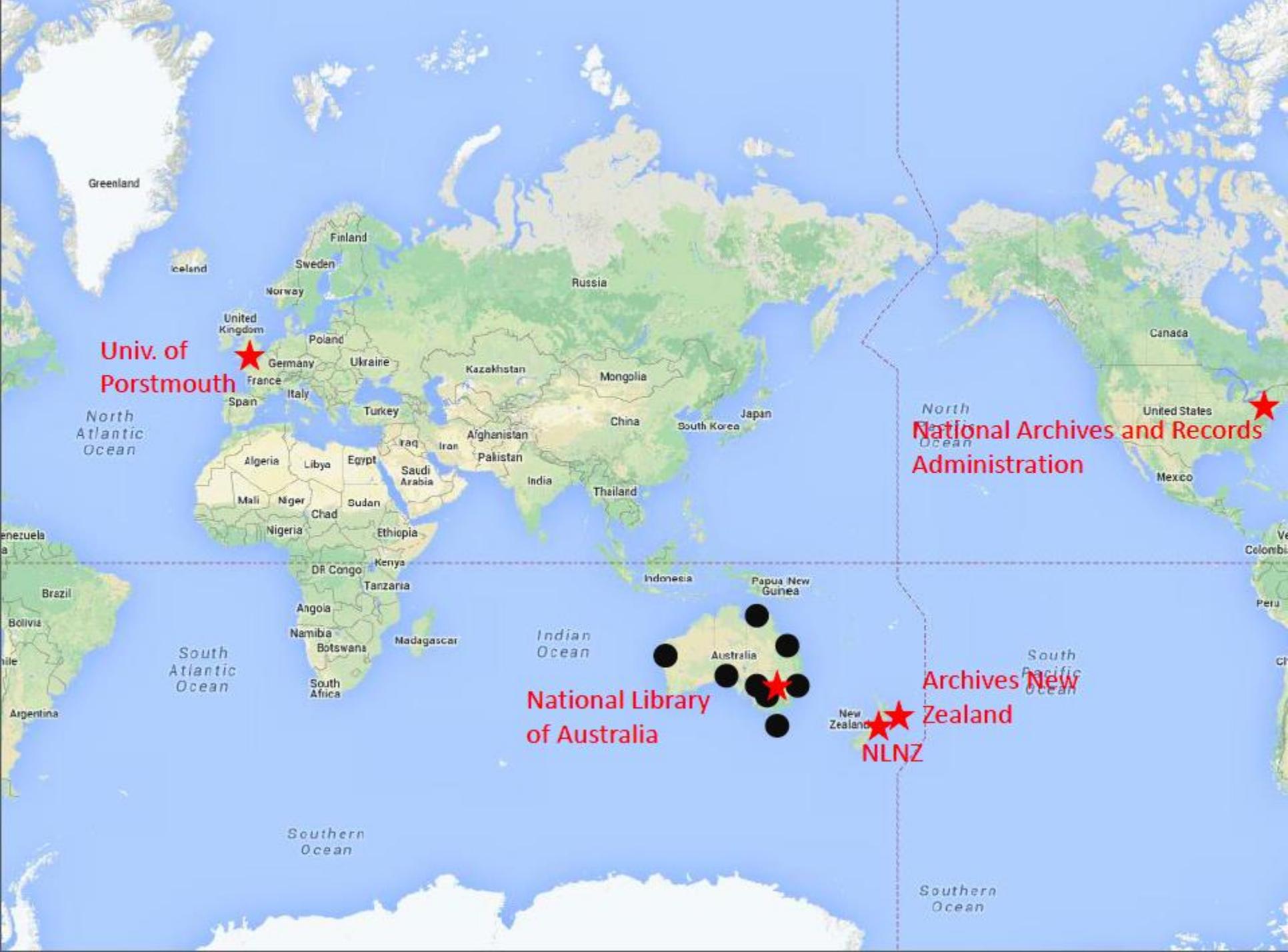
# Why have a technical registry?

A Technical Registry needs to:

- support preservation risk management, planning and action as central to an ongoing active digital preservation programme
- be at the centre of a mature digital preservation market supporting a range of digital preservation products, services and programmes
- be able to be deployed to any preservation repository (e.g. Rosetta, SDB, FEDORA, DuraSpace)
- be the key knowledge-base from which other products and services will draw to support individual organisation's digital preservation programmes (e.g. DROID, JHOVE, NLNZ MET)
- be flexible as described above in order to be successful.

# What's already out there?

- Current offerings:
- are dependent on the PRONOM development roadmap to the detriment of the wider digital preservation community needs
- reflect no other relevant sources that are available across the wider digital community
- lack depth of information for formats and software
- lack breadth of information for formats and software
- lack complex linking between information, which is at the very heart of preservation activities
- contain little or no information on hardware and peripherals
- do not allow for the development of community knowledge and best-practice.



**Univ. of  
Portsmouth**



**National Archives and Records  
Administration**



**National Library  
of Australia**



**Archives New  
Zealand**



**NLNZ**

# What are we planning to deliver?

- a Digital Preservation Technical Registry that captures and relates all aspects of technical information relating to digital preservation, including, but not necessarily limited to: formats, environments, applications, risks, plans, etc.

# What are we planning to deliver?

a repository of key technical information and relationships that will:

- primarily support the digital preservation community in understanding, characterising, validating, risk identification, and preservation of digital objects
- secondly provide a resource for organisations and individuals becoming involved in, or learning about, digital preservation.
- provide a comprehensive, consolidated, accurate, and open information source that can be used in conjunction with any digital preservation repository.

# Key capabilities of a Technical Registry

- The key capabilities of the solution are the ability to:
- import information from current and potential future source registries.
- store past versions of the external source registry records.
- support internal registries and online maintenance of the internal registries.
- link flexibly records within and across external source and internal registries.
- define the valid link types that can exist between records.
- configure what a user, role, or institution can view by allowing information to be filtered based on these attributes.
- support creating and running reports across external source and internal registries.

# Simple model

