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
Environment

- Software
- Hardware
- A format
- A document
  - A policy document
  - A manual
  - Documentation
- A cheat sheet
- A user behaviour study
- A process
- “Other representation information”
PREMIS Environment Working Group Goal

- High-level data model for Environments
- Capture the required relationships to other DP entities
- Capture desirable characteristics
- Standardized way of treating Environments
- Information sharing / exchange
- Repositories and registries

Not:
modelling the internals of a given Environment category – as e.g. TOTEM
Requirements

- Environments may be digital or non-digital
- Environments may be generic or instances (abstract description or concrete digital object)
- Environments may be tools or services
- Environments have no simple software / hardware distinction (Virtual machines blur the distinction)
PREMIS – Environment Metadata

1.8 **environment**
   1.8.1 environmentCharacteristic
   1.8.2 environmentPurpose
   1.8.3 environmentNote

1.8.4 **dependency**
   1.8.4.1 dependencyName
   1.8.4.2 dependencyIdentifier
      1.8.4.2.1 dependencyIdentifierType
      1.8.4.2.2 dependencyIdentifierValue

1.8.5 **software**
   1.8.5.1 swName
   1.8.5.2 swVersion
   1.8.5.3 swType
   1.8.5.4 swOtherInformation
   1.8.5.5 swDependency

1.8.6 **hardware**
   1.8.6.1 hwName
   1.8.6.2 hwType
   1.8.6.3 hwOtherInformation

1.8.7 environmentExtension

13 Nov 2012
The PREMIS Data Model

Slide adapted from S. Peyrard
Requirements

- Environment Type and SubType
- Environment Identifier
- Environment Designation (Name and Version)
- Environment Registry
- Environment Storage
- Relationship to other Environment: structural, replacement, dependency, generalization, reference, ...
- Link to Object, Agent and Identifier
Why PREMIS 3

- The environment description can be **modularized** and **shared** across different objects
- Express relationships between environments:
  - **whole/part**: environments can bundle together different pieces of environments
  - **replacement**: environments can be superseded by more recent ones
  - **dependency**: environments can be related to other environment that support their use
- Possibility to associate a **registryKey** with an environment
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)
relationship type: conceptual
subType: is specialized in
object file .img
urn:x-nbn:de:o1

relationship type: structural
subType: includes
environment DNB catalogue
urn:x-nbn:de:y1

role: render
MS-DOS operating system
urn:x-nbn:de:y2

relationship type: conceptual
subType: is specialized in
MS-DOS 7.1 operating system
urn:x-nbn:de:y4

relationship type: emulation
subType: is emulated by
DOSBox software emulator
urn:x-nbn:de:y6

QEMU 1.2 hardware emulator
urn:x-nbn:de:y7

relationship type: dependency
subType: requires
Emulation Framework 2.0.0
urn:x-nbn:de:y8

relationship type: dependency
subType: requires
DNB reading room environment
urn:x-nbn:de:y9

13 Nov 2012
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/](http://www.keep-totem.co.uk/).
- Several iterations?
- Hardware emulators and if necessary, software emulators can then be specified.
- We need an emulation platform, e.g. KEEP Emulation framework (EF) [http://emuframework.sourceforge.net/](http://emuframework.sourceforge.net/) to run these emulators. (Also bwFLA)
- Finally need reading room environment to run this EF.
## Trustworthy Online Technical Environment Metadata (TOTEM) compatibility search

**Search Results for: Adobe Acrobat Reader and Acrobat 6.0**

<table>
<thead>
<tr>
<th>OS Name</th>
<th>Windows NT</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS Version Name</td>
<td>Windows 2000</td>
</tr>
<tr>
<td>Release Date</td>
<td>2000-02-17</td>
</tr>
<tr>
<td>Description</td>
<td>Four editions of Windows 2000 were released, listed here in increasing ranking: Professional, Server, Advanced Server, and Datacenter Server.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OS Name</th>
<th>Windows NT</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS Version Name</td>
<td>Windows 2000</td>
</tr>
<tr>
<td>Release Date</td>
<td>2000-02-17</td>
</tr>
<tr>
<td>Description</td>
<td>Four editions of Windows 2000 were released, listed here in increasing ranking: Professional, Server, Advanced Server, and Datacenter Server.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OS Name</th>
<th>Windows NT</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS Version Name</td>
<td>Windows XP 64-bit Edition</td>
</tr>
<tr>
<td>Release Date</td>
<td>2001-10-25</td>
</tr>
<tr>
<td>Description</td>
<td>The first Microsoft operating system for 64-bit processors designed for working with large amounts of memory and projects such as movie special effects, 3D animations, engineering, and scientific programs.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OS Name</th>
<th>Windows NT</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS Version Name</td>
<td>Windows XP 64-bit Edition</td>
</tr>
<tr>
<td>Release Date</td>
<td>2001-10-25</td>
</tr>
<tr>
<td>Description</td>
<td>The first Microsoft operating system for 64-bit processors designed for working with large amounts of memory and projects such as movie special effects, 3D animations, engineering, and scientific programs.</td>
</tr>
</tbody>
</table>
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 who can expect this award! See the DPC

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

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.
Conclusion

- Design Environment stand-alone entity

- Validate on community-provided use cases: TOTEM technical registry, IIPC, DAITSS, TIMBUS project, New York University

- Propose to PREMIS Editorial Committee