Sustaining and Enhancing the Value of Digital Assets

DPC Briefing Day: Virtualization and Preservation

How cloud computing changes how we think about digital preservation'

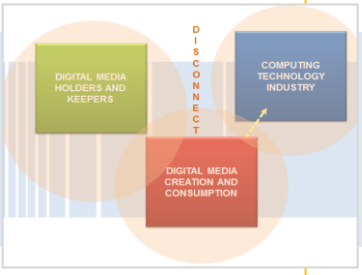
Cambridge, 22 July, 2014

Natasa Milic-Frayling
Principal Researcher
Microsoft Research Cambridge, UK

Overview

The Nature of Digital

Properties and paradoxes of digital



Sustainability of Digital

Methods for ensuring content accessibility
Gaps in the value chain

FILE → APPLICATION → DIGITAL CONTENT/ EXPERIENCE

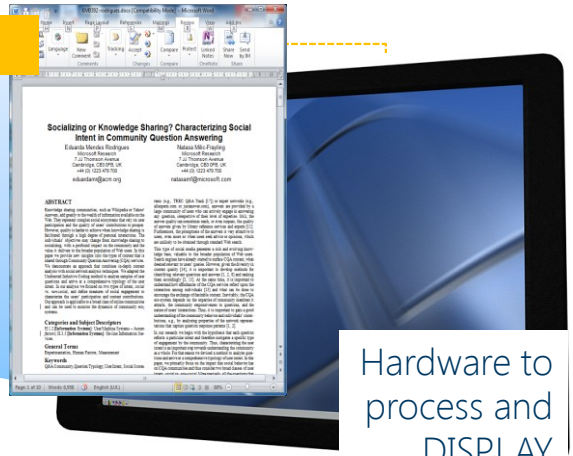
Persisted

Ephemeral

PRESERVATION = Persistence + Connection with the contemporary ecosystem.

DIGITAL ARTEFACT

SOFTWARE – decoder



FILE – digital object

Hardware to process and DISPLAY

Persisted part of the digital artefact

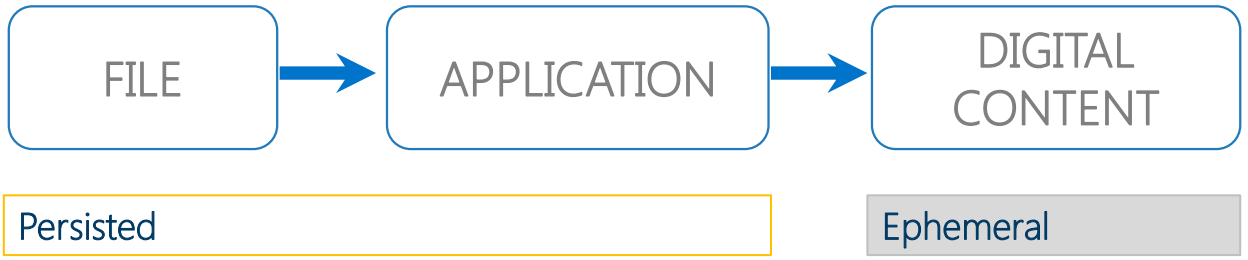
Paradox: we are concerned about storage, yet

Digital is inherently about processing bits, not about storing bits

Symbiosis of Files and Applications

Objective of preservation is to ensure that the persisted digital content and applications remain connected with the contemporary computing ecosystem.

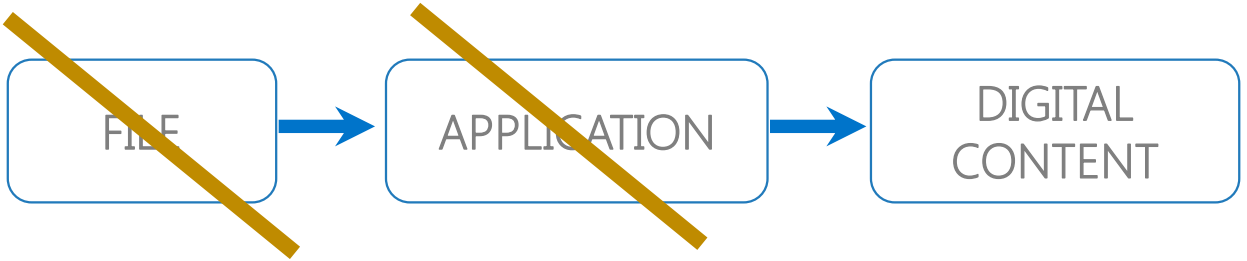
PRESERVATION = Persistence + Connection with the contemporary ecosystem.



What do you want to keep ‘unchanged’?

- If application is not running in the contemporary environment

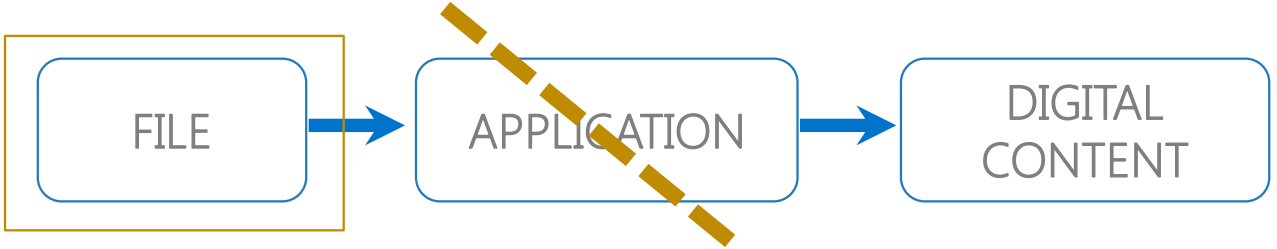
What do you want to keep ‘unchanged’?



- If application is not running in the contemporary environment
  - Migrate files and run with a contemporary software

(give up on both the original files and the application)

What do you want to keep ‘unchanged’?



- If application is not running in the contemporary environment
  - Retain the files and port the application to the new environment

(retain content files by give up on the application, at least partially)

What do you want to keep ‘unchanged’?

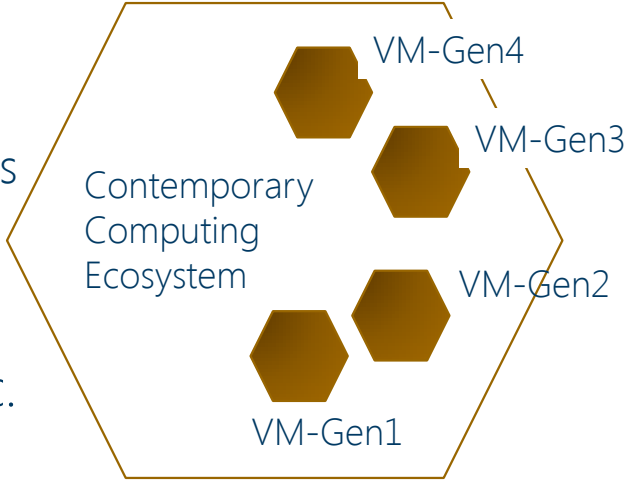
- If application is not running in the contemporary environment
  - Create a virtual machine with the old computing stack and run the original files and software.

(retain original files and original application; maintain scaffolding)

Computational Cradles

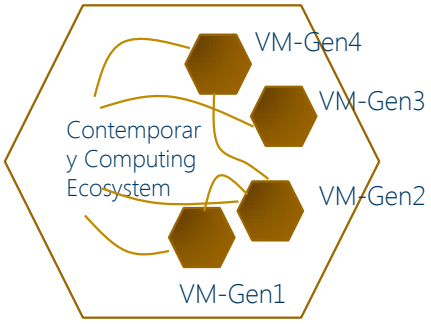
Sustain and increase the value of digtial through

- Virtualization of legacy software + Bridging Services
- Individual computational 'cells' for different generations of software stacks



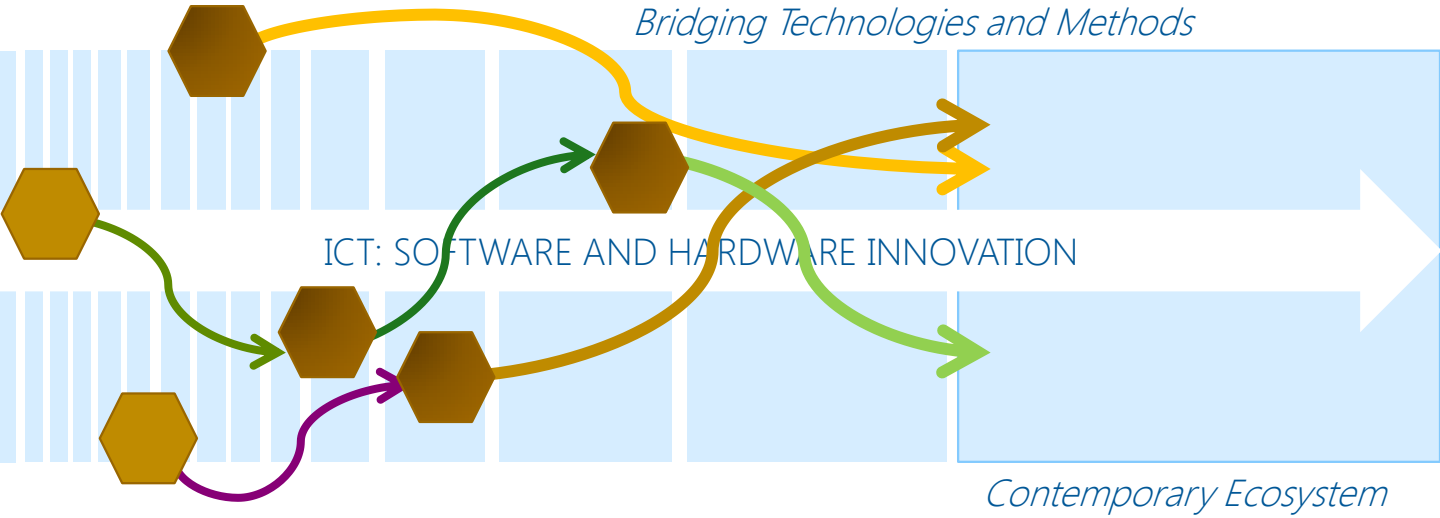
Bridging services: format translators, content extractors, etc.

Connecting Legacy with the Contemporary Ecosystem



Digital artifact always requires (some software) computation.

No need to give up on the original software!



Winning strategy

Virtualization of original software

Ensures access to the digital artefacts

Format transformation services

On demand transformation within a specific context.

Winning strategy

Virtualization of original software

Ensures access to the digital artefacts

Format transformation services

On demand transformation within a specific context.

extracting value from digital

FORMAT TRANFORMATION SERVICES

SCAPE–SCAlable Preservation Environments

- Develop scalabe services for planning and execution of preservation strategies
- Open source platform for semi-automated workflows for large-scale, heterogeneous collections of complex digital objects.

FP7 Project. Started February 2010. Sponsored for 3.5 years.

SCAPE Partners 2010-2014

AIT Austrian Institute of Technology GmbH

The British Library

Internet Memory Foundation

Ex Libris Ltd.

Fachinformationszentrum Karlsruhe, Gesellschaft für Wissenschaftlich-Technische Information GmbH

Koninklijke Bibliotheek

KEEP SOLUTIONS LDA

Microsoft Research Limited

Österreichische Nationalbibliothek

Open Planets Foundation

Statsbiblioteket

Science and Technologies Facilities Council

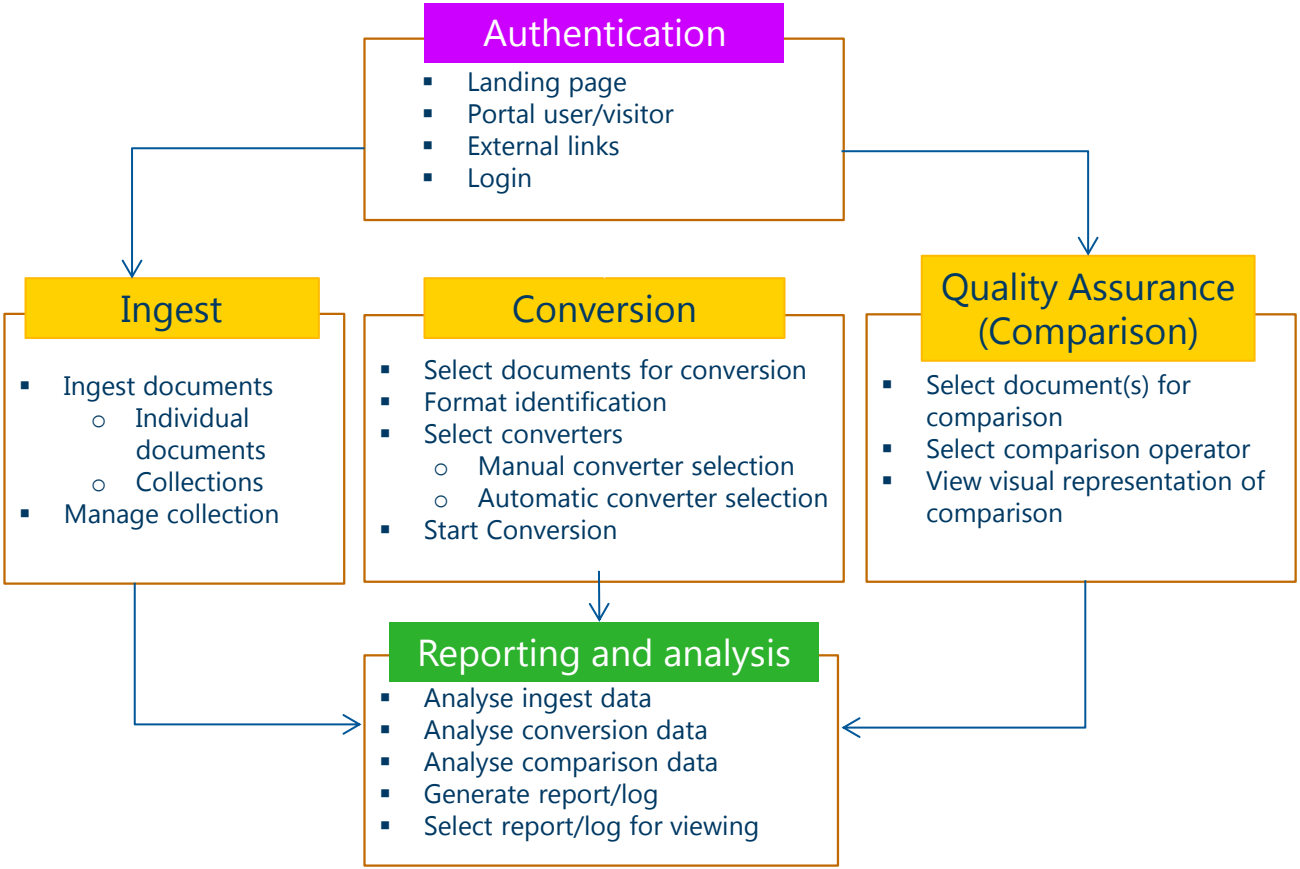
Technische Universität Berlin

Technische Universität Wien

The University of Manchester

Universite Pierre et Marie Curie – Paris 6

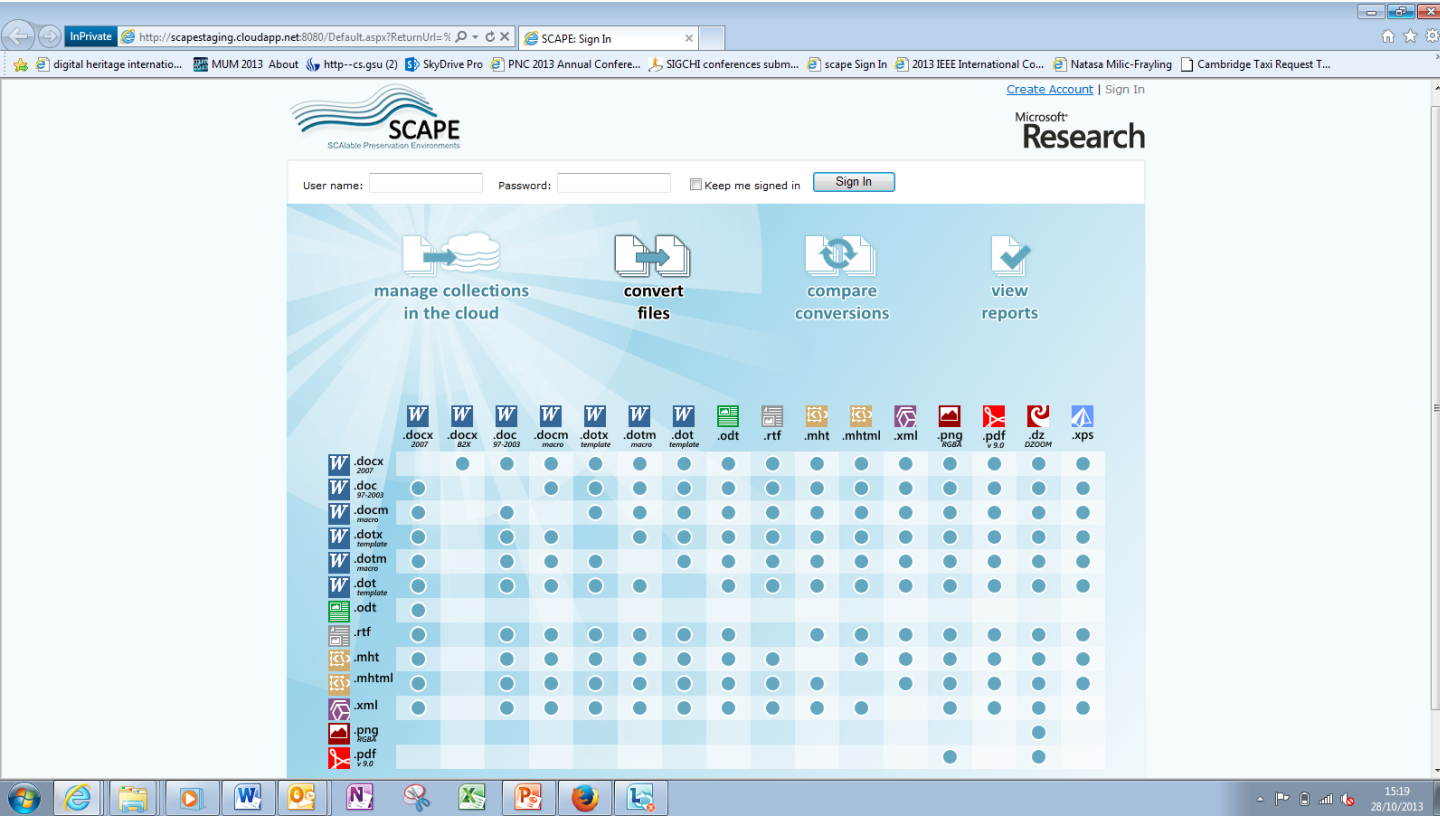
User Interaction with SAZ Services



Demo: Use of SCAPE Azure Services

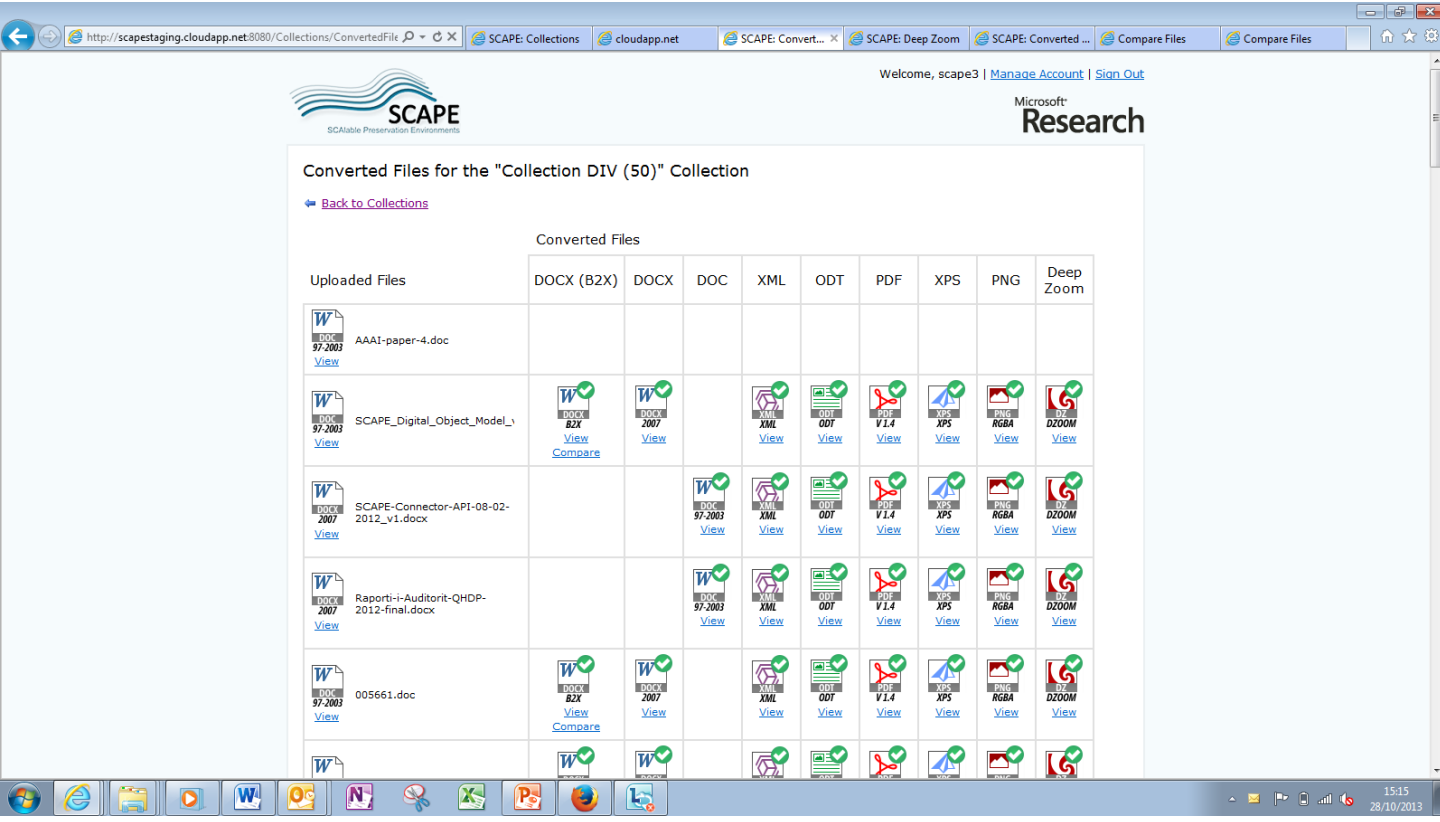
SCAPE: Format Transformation Services on Microsoft Cloud (Azure)

Format Transformation Options



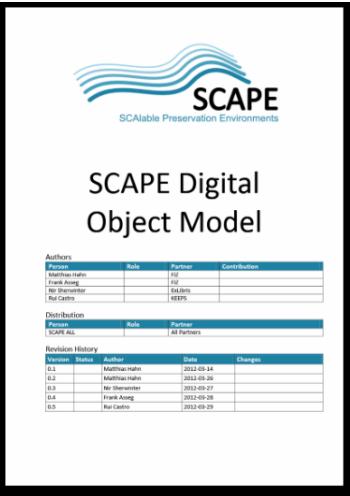
Prototype Features

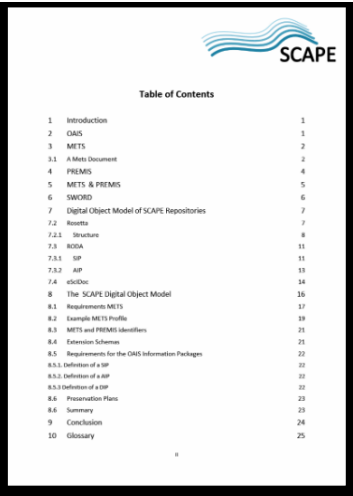
Results of the Conversion Process

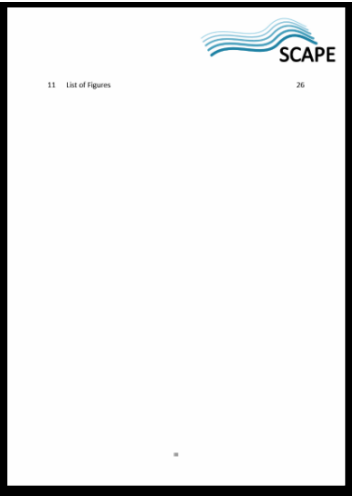


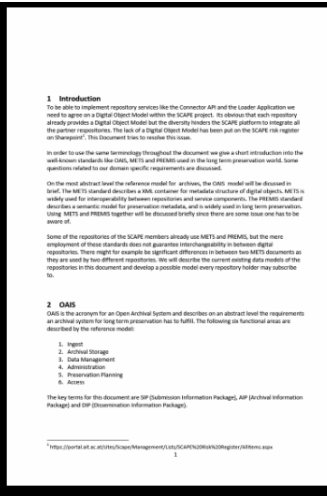
Prototype Features

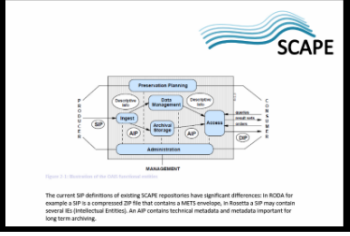
Word'97 Document Transformed into Deep Zoom Format

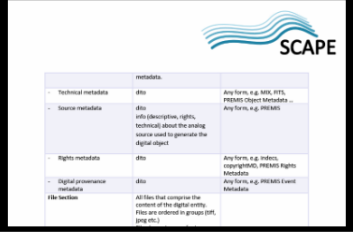


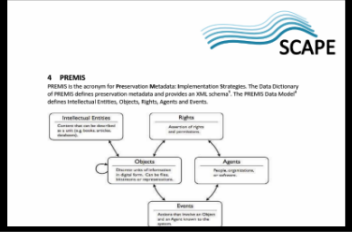


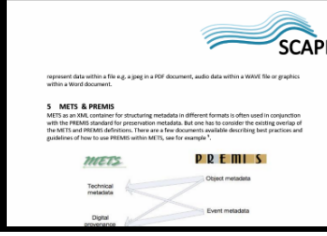








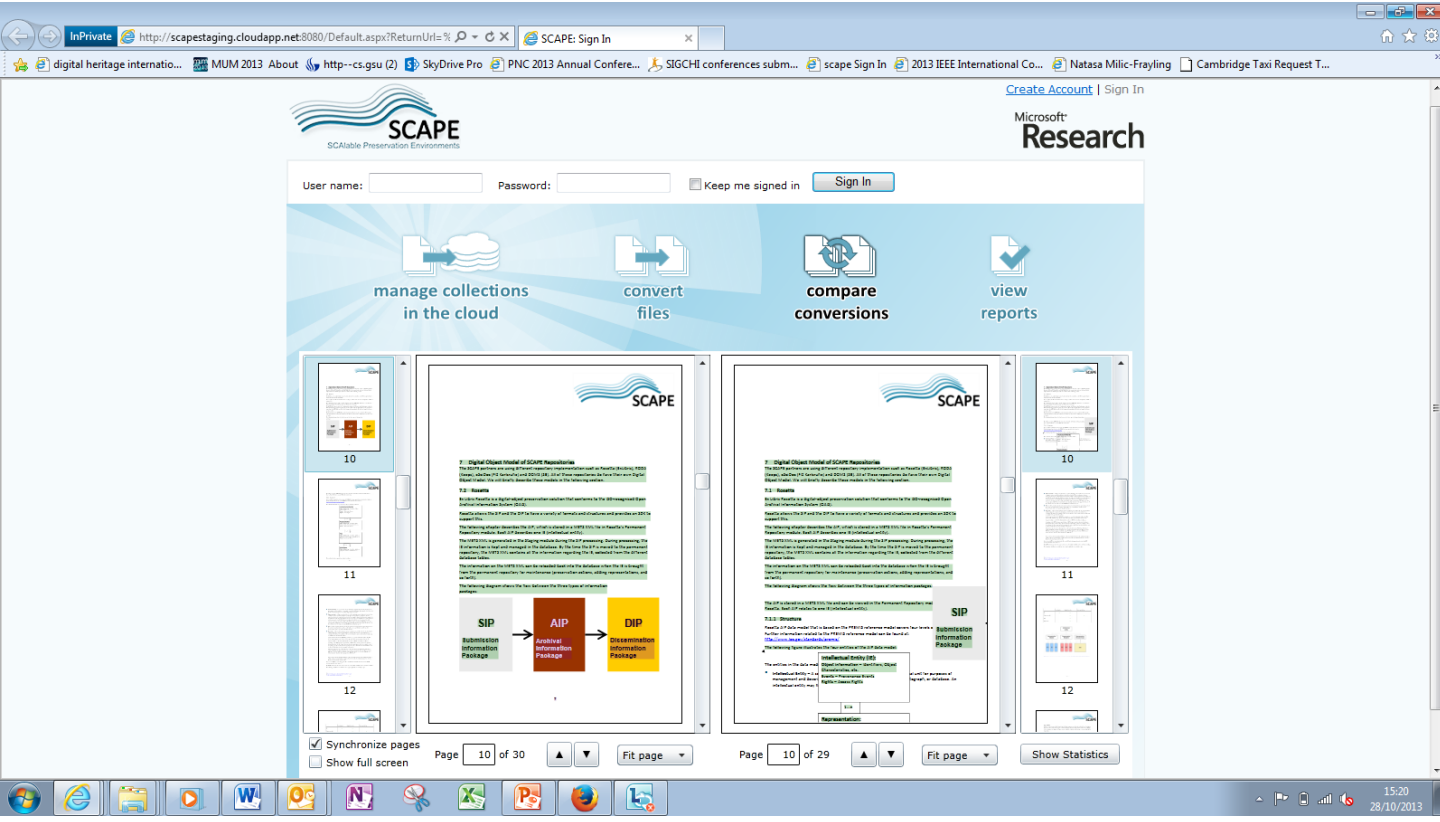




Show full screen | Zoom In | Zoom Out | Reset

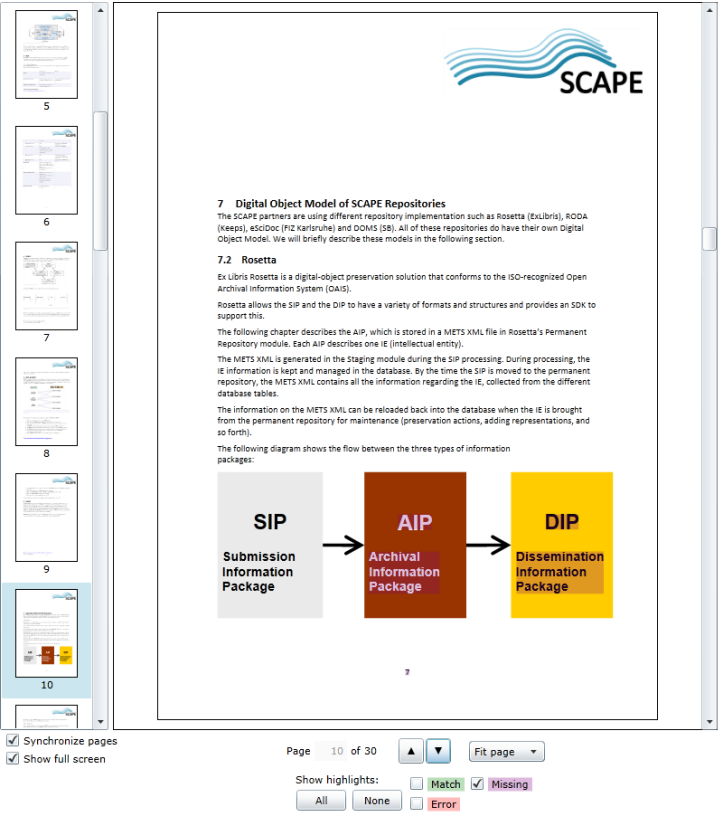
SCAPE: Format Transformation Services on Microsoft Cloud (Azure)

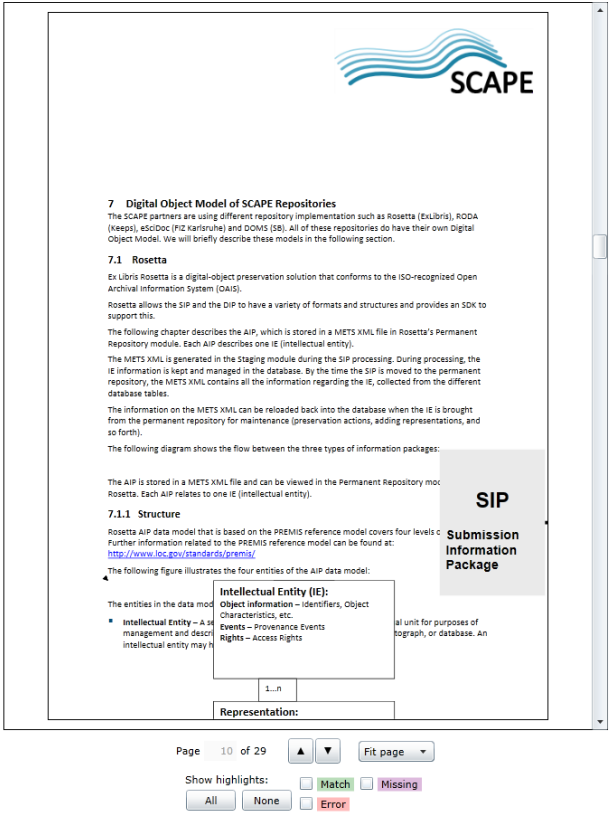
Quality Assessment of the Conversion Process



Prototype Features

Differences are Assessed and Classified based on 'Importance'





Synchronize pages
Show full screen

Page 10 of 30 ▲ ▼ Fit page

Show highlights: Match Missing
All None Error

Page 10 of 29 ▲ ▼ Fit page

Show highlights: Match Missing
All None Error

Show Statistics

Prototype Features

Pivot View of the Collection Metadata

Winning strategy

Virtualization of original software

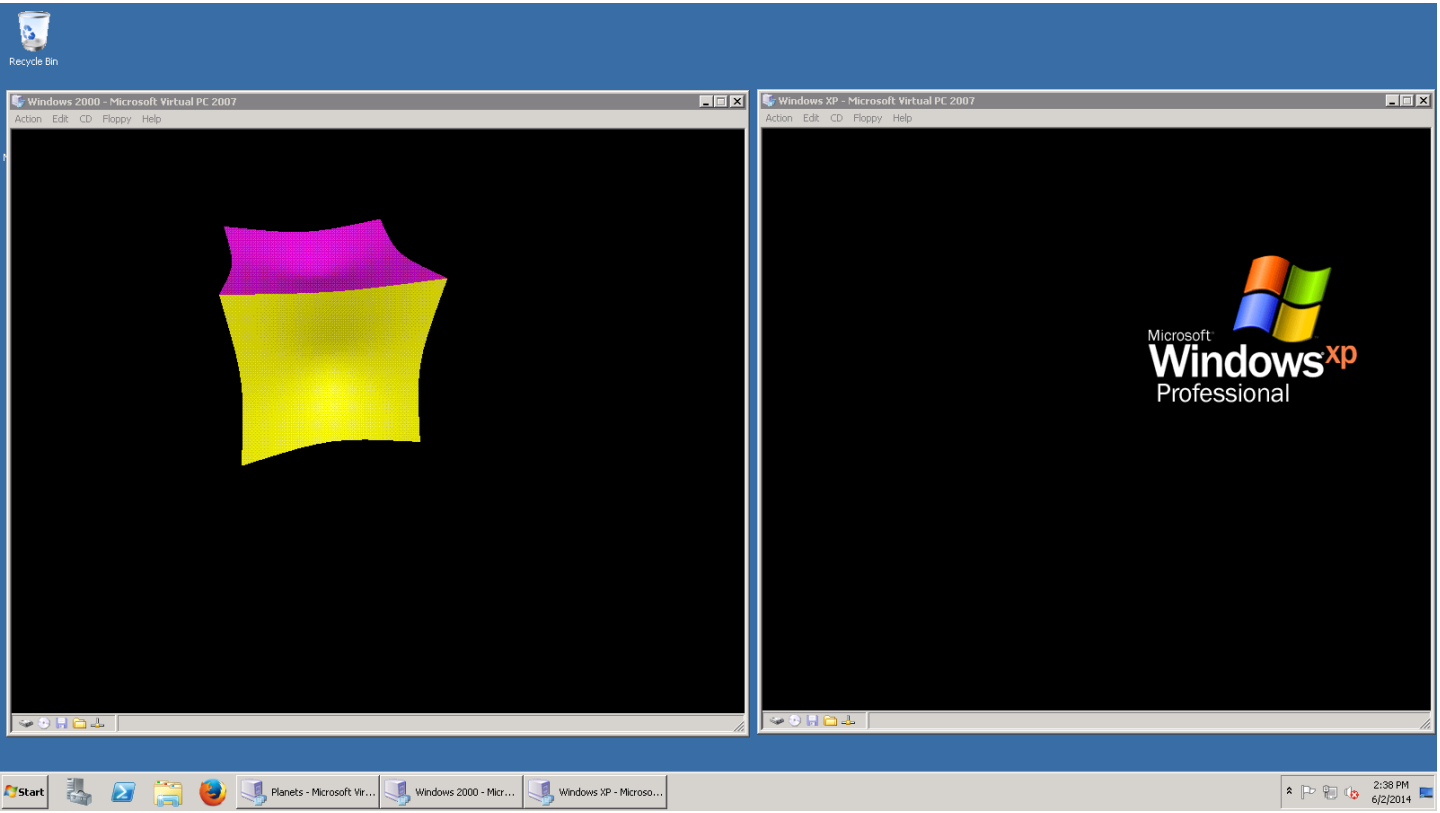
Ensures access to the digital artefacts

Format transformation services

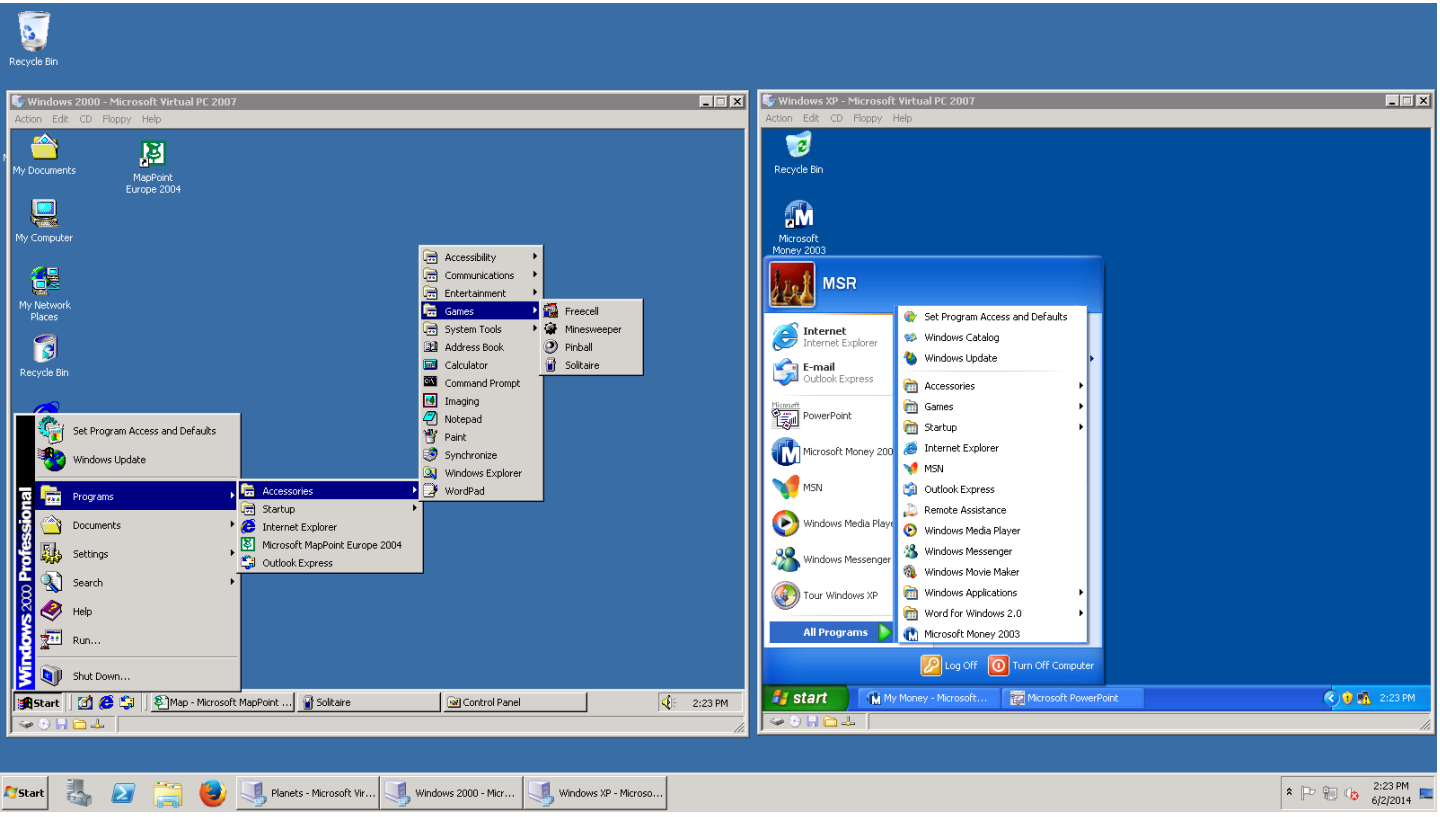
On demand transformation within a specific context.

preserving computation

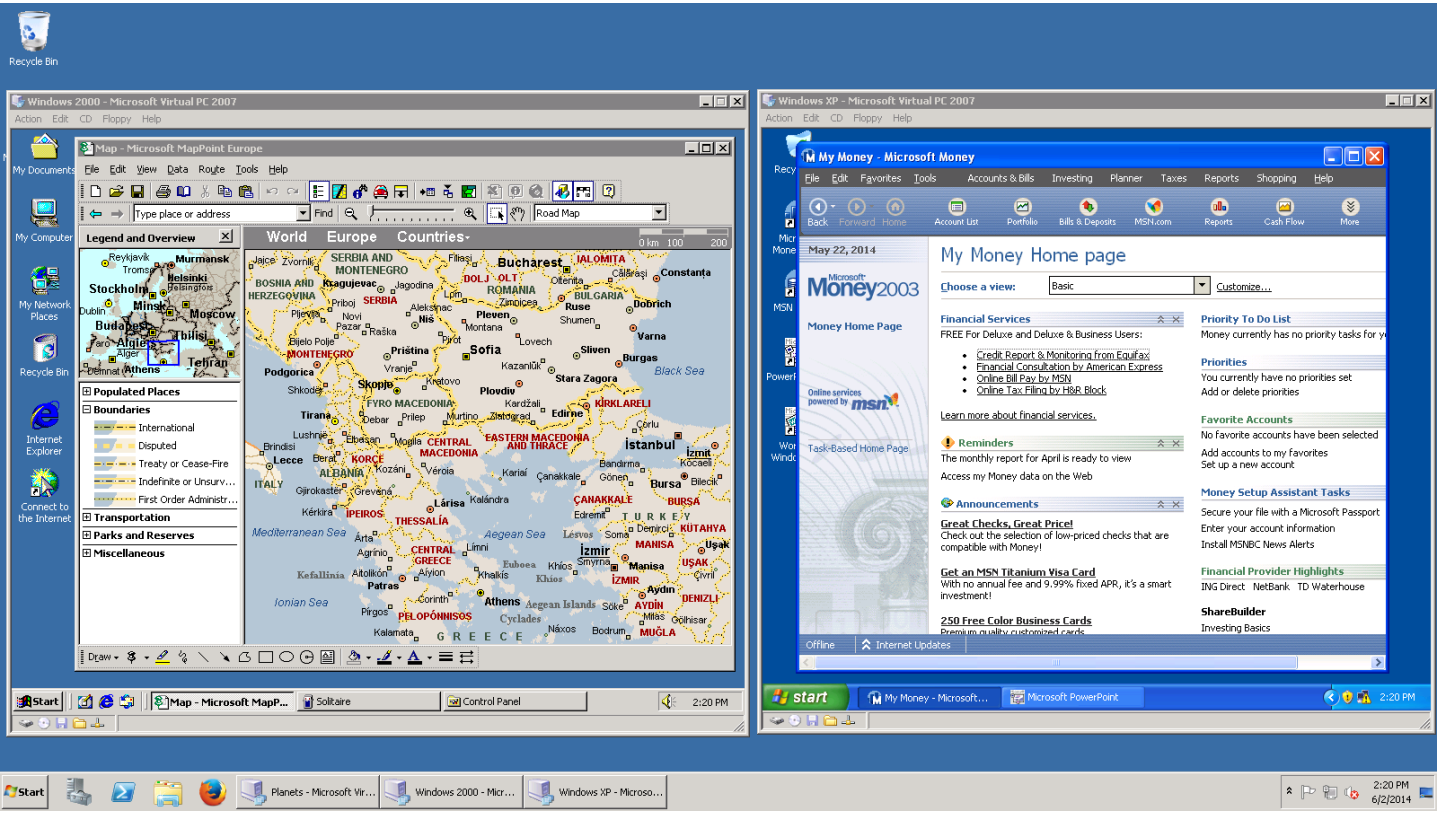
VIRTUALIZATION OF LEGACY SOFTWARE



Virtual Machine with Windows 2000 (left) and Windows XP (right), running on Microsoft Cloud (Azure)



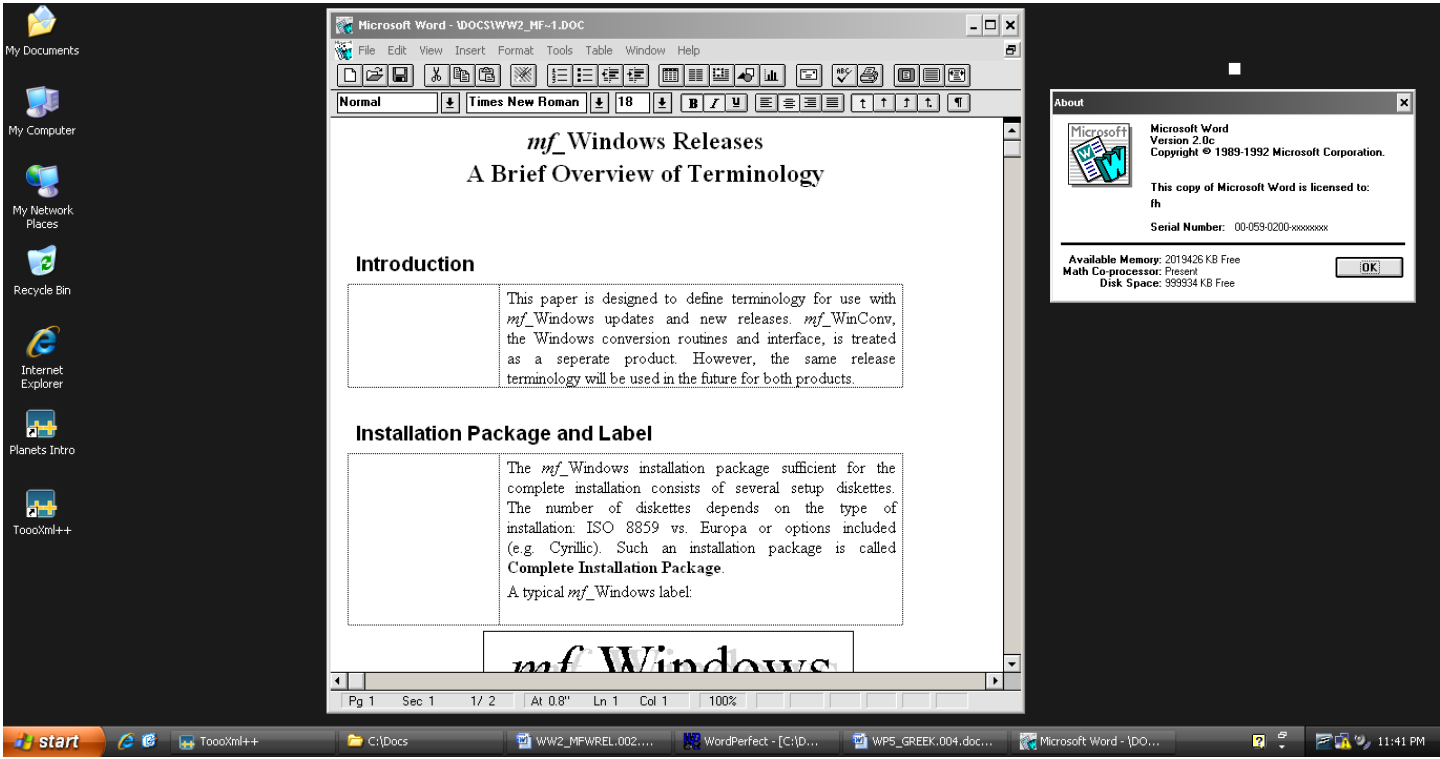
Start menus for Windows 2000 (left) and Windows XP (right),



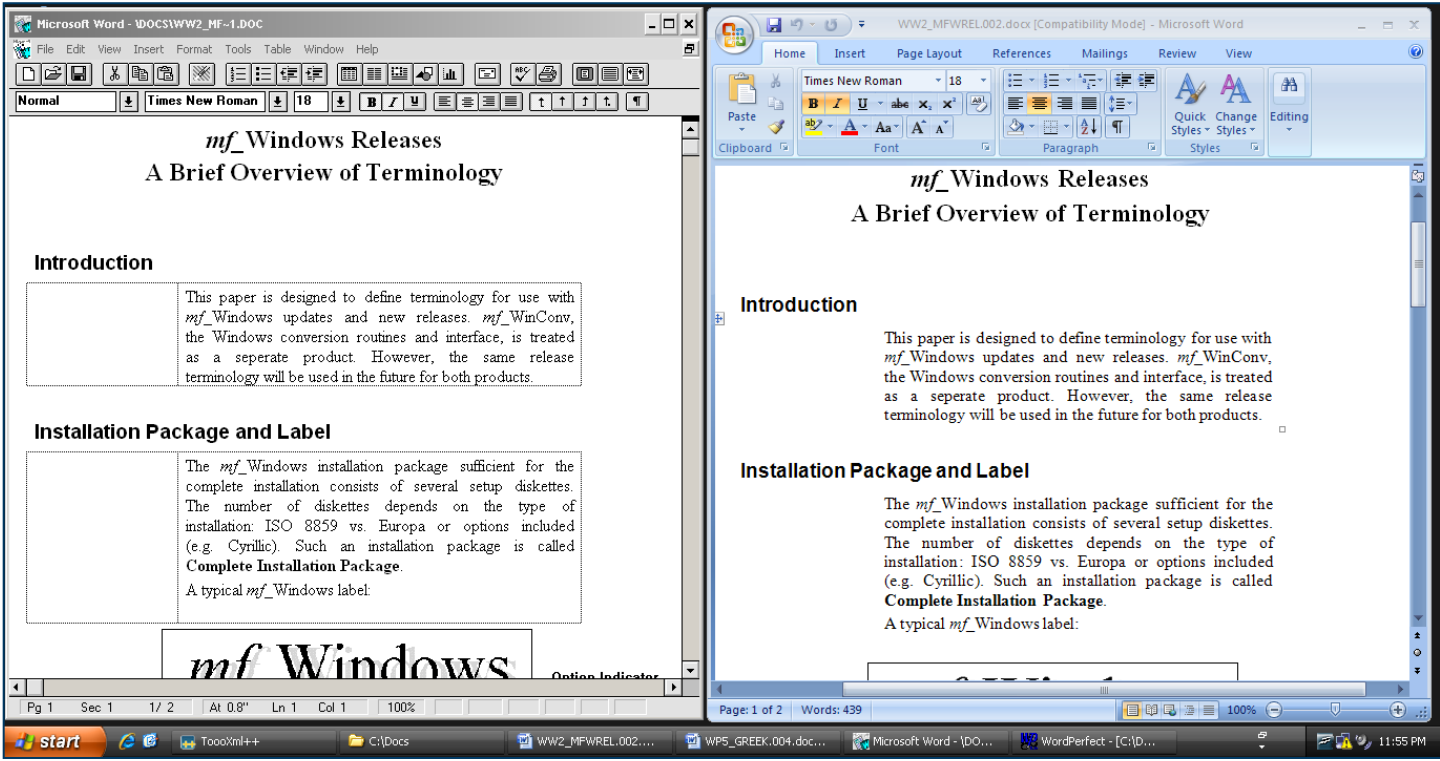
MS Map Point application running on Windows 2000 (left) and MS Money 2003 running on Windows XP (right),

Increasing value of legacy content

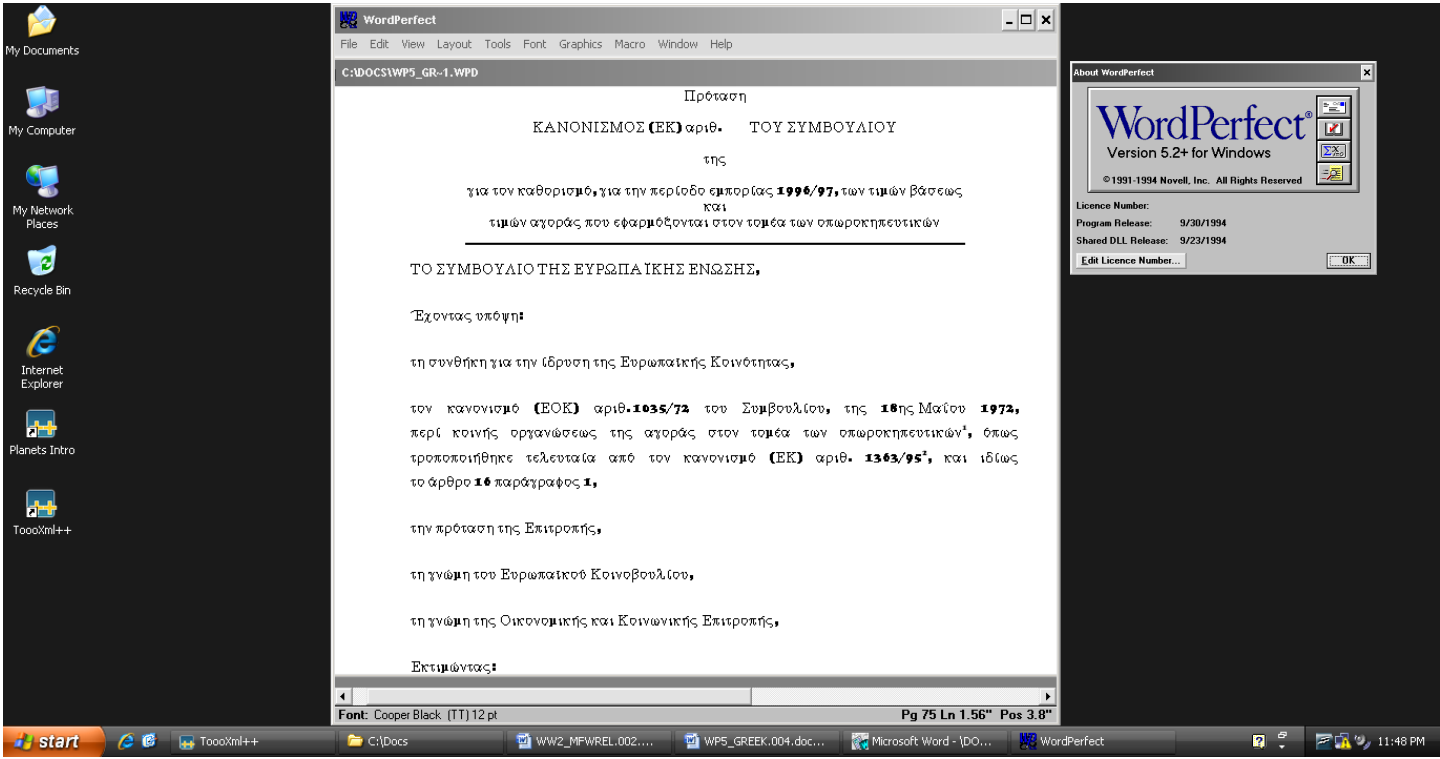
FORMAT TRANSFORMATION



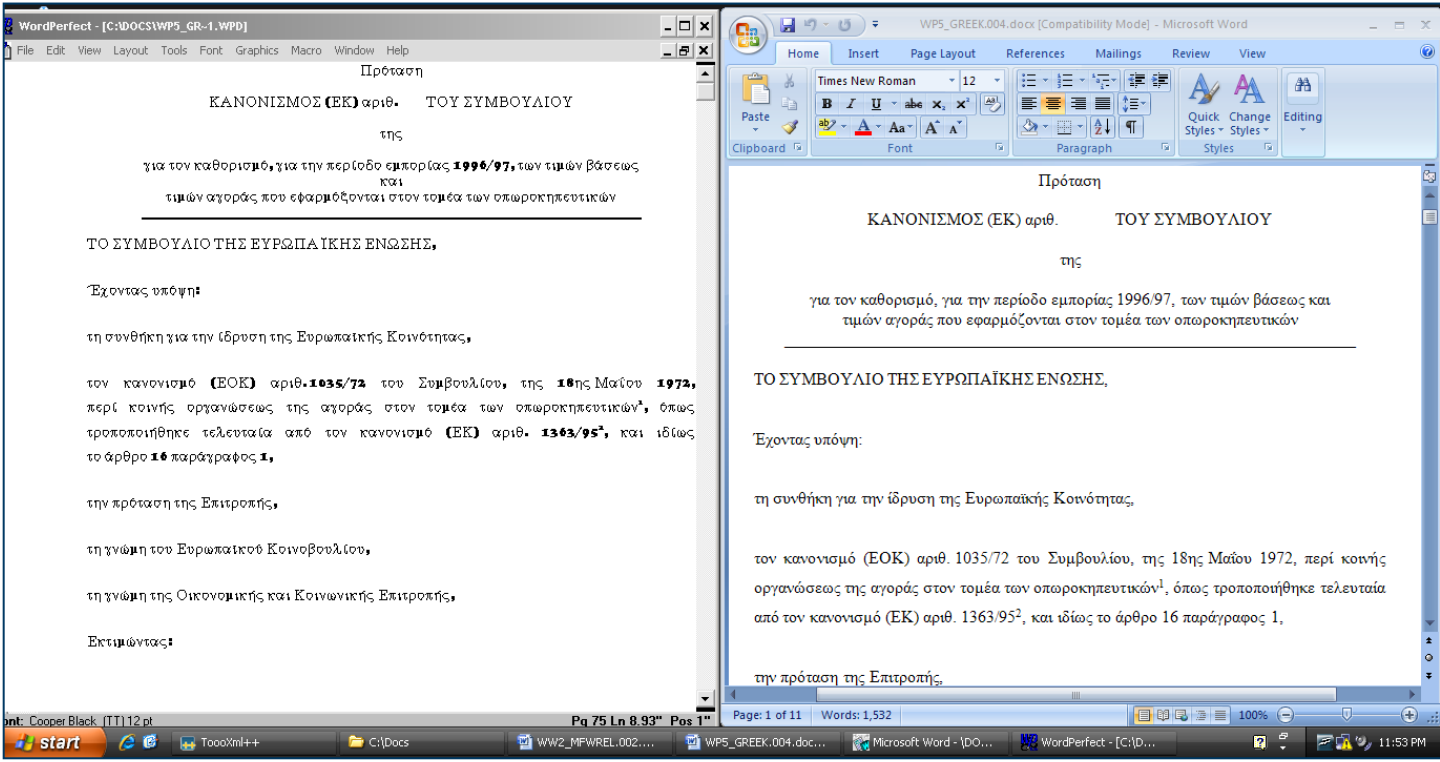
Word document shown in Microsoft Word 2.0 (from 1992)
Running in the Virtual Machine with Windows XP



Word document in MS Word 2.0 (from 1992) and converted to Open XML format, shown in Office 2007 (right)



Word Perfect document, shown in WordPerfect 5.2 (from 1994)
Running in the Virtual Machine with Windows XP



Word Perfect document in WordPerfect 5.2 (from 1994) and converted to Open XML format, shown in Office 2007 (right)

How to cover the cost of long term access?

Manage the services cost to achieve a positive balance between

- value created through immediate explorations of content
- investment needed to sustain perceived value of content in the future.

Assessing the potential value of digital content in the future is difficult.

We can make projections into the near future based on the current needs and opportunities.

Sustainability is possible if the current use scenarios extract sufficient value from the legacy digital assets

What does the Cloud paradigm offer?

Distributed IT cost and opportunities for extracting value from legacy content:

- Extendible functionality
- Extendible data store
- Scalable computation
- Virtualization
- Common platform for creating services
- Support for client applications on diverse computing platforms.

Cloud may enable digital future

under the assumptions that:

- Access to digital media becomes one of the primary drivers for innovation and evolution of the ICT ecosystem
  - Customers/digital media producers should demand and pay for long term access provisions at the time of technology acquisition.
- Digital media curation and education become an essential component of digital media services
  - Content creators and content holders need to demonstrate that there is value in combining contemporary and past information to provide compelling and competitive services.

Thank you

Natasa Milic-Frayling
natasamf@microsoft.com
Integrated Systems
Microsoft Research Cambridge UK