Microsoft Azure

Microsoft Azure and the Azure4Research Program

Alex Wade
Director – Scholarly Communication
Microsoft Research

@alexwade
Bringing cloud computing to researchers

Learn about the new Azure for Research program

Azure Research Awards (>320 to date)
Training and Webinar series
Technical papers & curriculum
Research community engagements

www.azure4research.com
• Ensure long-term access to Europe’s cultural and scientific heritage
  – Improve decision-making about long term preservation
  – Ensure long-term access to valued digital content
  – Control the costs through automation, scalable infrastructure
  – Ensure wide adoption across the user community
  – Establish market place for preservation services and tools
• Build practical solutions
  – Integrate existing expertise, designs and tools
  – Share and build
SCAPE—SCAlable Preservation Environments

- Develop scalable 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

Österreichische Nationalbibliothek

Open Planets Foundation

Statsbiblioteket

Science and Technologies Facilities Council

Technische Universität Berlin

Technische Universität Wien

The University of Manchester

Université Pierre et Marie Curie – Paris 6
# Format migration of Office documents

<table>
<thead>
<tr>
<th>Source formats</th>
<th>Target formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>WordPerfect 5</td>
<td>OpenXML</td>
</tr>
<tr>
<td>WordPerfect 6</td>
<td>ODF</td>
</tr>
<tr>
<td>DOS Word</td>
<td>UOF</td>
</tr>
<tr>
<td>Word 2, 6, 95</td>
<td>HTML</td>
</tr>
<tr>
<td>Word 97-2003</td>
<td>RTF</td>
</tr>
<tr>
<td>RTF</td>
<td>ODF</td>
</tr>
<tr>
<td>ODF</td>
<td>HTML</td>
</tr>
<tr>
<td>OpenXML</td>
<td>XCDL (format defined in PLANETS)</td>
</tr>
</tbody>
</table>
Workflow supported by the SCAPE Azure Services

**Authentication**
- Landing page
- Portal user/visitor
- External links
- Login

**Ingest**
- Ingest documents
  - Individual documents
  - Collections
- Manage collection

**Conversion**
- Select documents for conversion
- Format identification
- Select converters
  - Manual converter selection
  - Automatic converter selection
- Start Conversion

**Quality Assurance (Comparison)**
- Select document(s) for comparison
- Select comparison operator
- View visual representation of comparison

**Reporting and analysis**
- Analyse ingest data
- Analyse conversion data
- Analyse comparison data
- Generate report/log
- Select report/log for viewing
Available conversions

- Document conversions supported by MSR Azure Services
- General principle: repurpose existing tools.
SCAPE Azure Service demo
Figure Error: No text or specified style to document a text element in the document.

An object in PREMIS has three subtypes: file, representation and stream.
What does Cloud offer us?

Extendible functionality
Extendible data store
Scalable computation
Virtualization
Common platform for creating services
Support for client applications on diverse computing platforms
Thank you!

@alexwade