Digital Library Interoperability

Alex D. Wade
Director for Scholarly Communication
Microsoft External Research
Division within Microsoft Research focused on partnerships between academia, industry and government to advance research in fields that rely heavily upon advanced computing.

Supporting groundbreaking research to help advance human potential and the wellbeing of our planet.

Developing advanced technologies and services to support every stage of the research process.

Microsoft External Research is committed to interoperability and to providing open access, open tools, and open technology.

http://research.microsoft.com/collaboration/about/
Microsoft recognizes that we need to work more at engaging with the community in an open way on interoperability.

The principles:
- Open Connections to Microsoft Products
- Standards Support
- Data Portability
Open to Choice

Interoperability builds choice, so governments, developers, and citizens can decide what technologies work best for them. At Microsoft, we design our products to work with third-party software, as well as our own.

• Document Interoperability
  – When document format choices are limited, restrictions on technology choices generally follow.

http://www.microsoft.com/interoperability
Open to Innovation

Interoperability drives innovation within a thriving IT industry, creating technologies that improve citizen services and government efficiency. Governments and businesses alike are looking to new technologies, like cloud services, to enable innovative offerings.

• Liberating Data
  – Open Government Data Initiative
  – Open Data (OData) Protocol

http://www.microsoft.com/interoperability
When the IT community collaborates and develops new technologies that work together, new opportunities are created. Companies across the IT sector are working with partners, customers, and competitors to deliver the levels of interoperability customers want.

- IDC’s “Global Economic Impact Study” (2009)

http://www.microsoft.com/interoperability
• Built to be interoperable
• Web standards (HTTP, XML, SOAP, REST, etc.)
• Programming language support
  – .NET SDK
  – Ruby SDK
  – Java SDK
Open Data Protocol
A RESTful Interface for Data

Just HTTP
- Items as resources, HTTP methods (GET, PUT, ...) to act
- Leverage proxies, authentication, ETags, ...

Uniform URL convention
- Every piece of information is addressable
- Predictable and flexible URL syntax

Multiple representations
- Use regular HTTP content-type negotiation
- JSON and Atom (full AtomPub support)

http://www.odata.org
### Addressing lists and items

<table>
<thead>
<tr>
<th>Type</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of lists</td>
<td>.../_vti_bin/listdata.svc/</td>
</tr>
<tr>
<td>List</td>
<td>listdata.svc/Employees</td>
</tr>
<tr>
<td>Item</td>
<td>listdata.svc/Employees(123)</td>
</tr>
<tr>
<td>Single column</td>
<td>listdata.svc/Employees(123)/Fullname</td>
</tr>
<tr>
<td>Lookup traversal</td>
<td>listdata.svc/Employees(123)/Project</td>
</tr>
<tr>
<td>Raw value access</td>
<td>listdata.svc/Employees(123)/Project/Title/$value</td>
</tr>
</tbody>
</table>

### Presentation options

<table>
<thead>
<tr>
<th>Option</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorting</td>
<td>listdata.svc/Employees?$orderby=Fullname</td>
</tr>
<tr>
<td>Filtering</td>
<td>listdata.svc/Employees?$filter=JobTitle eq 'SDE'</td>
</tr>
<tr>
<td>Projection</td>
<td>listdata.svc/Employees?$select=Fullname,JobTitle</td>
</tr>
<tr>
<td>Paging</td>
<td>listdata.svc/Employees?$top=10&amp;$skip=30</td>
</tr>
</tbody>
</table>
OData Producers

- Applications
  - SharePoint 2010
  - IBM Websphere
  - Windows Azure Table Storage
  - SQL Azure
  - Zentity 2.0

- Live OData Services:
  - Facebook Insights
  - Netflix
  - Open Government Data Initiative
  - Open Science Data Initiative
  - DBPedia

OData Consumers

- Web Browsers
- Excel 2010
- LinQPad

- Client libraries for
  - Javascript
  - PHP
  - Java
  - iPhone (Objective C)
  - Windows 7 Phone
  - .NET

http://www.odata.org
Apache POI OpenXML Java API
Apache Stonehenge
AppFabric SDK for Java
AppFabric SDK for PHP
AppFabric SDK for Ruby
Bing Search Library for PHP
Eclipse Tools for Silverlight
Enhancements of Eclipse on Windows 7
Information Cards for C
Information Cards for Java
Information Cards for PHP
Information Cards for Ruby
Lab: Development Environment and Software Configuration Management (SCM) Interop
Lab: Directory Management Interoperability
Lab: Document & Records Management
Lab: Federated Identity
Lab: Interoperability with Open XML
Lab: Multi-platform System Management
Lab: Portal Aggregation and Integration
Lab: Remote Desktop Interoperability
Lab: Systems Management OData Client for Objective C (iPhone-Mac)
OData SDK for PHP
Office Binary to Open XML Translator
OpenXML Document Viewer/HTML Translator
OpenXML/ODF Translator
OpenXML/UOF Translator
Restlet Extension for OData
Samples for PHP with Silverlight
Samples for PHP with Webslices & Accelerators
SQL CRUD Application Wizard for PHP
SQL Server Reporting Services SDK for PHP
Toolkit for PHP with Virtual Earth (Bing Maps)
Windows Azure Command-line Tools for PHP
Windows Azure SDK for Java
Windows Azure SDK for PHP
Windows Azure Storage for WordPress
Windows Azure Tools for Eclipse
WSRP Producer for Microsoft SharePoint Server

http://www.interoperabilitybridges.com/
Zentity – a Research Output Repository Platform

A semantic computing platform to store and expose relationships between digital assets

Flexible data model enables many scenarios and can be easily extended over time

Native support for RSS, OAI-PMH, OAI-ORE, AtomPub and SWORD

v.1 (v.2 available later this month!) :
http://research.microsoft.com/zentity/
Microsoft Academic Search
http://academic.research.microsoft.com
DL Opportunities/Challenges

- Clouds
- Data, data, data
- Preservation and File Formats
- Language
- Integration
Cloud Data Centers: Economies of Scale

- Data Centers range in size from “edge” facilities to megascale (100K to 1MK servers)
- Offer real economies of scale
  - Approximate costs for a small size center (1K servers) and a larger, 400K server center.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Cost in small-sized Data Center</th>
<th>Cost in Large Data Center</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
<td>$95 per Mbps/Month</td>
<td>$13 per Mbps/month</td>
<td>7.1</td>
</tr>
<tr>
<td>Storage</td>
<td>$2.20 per GB/Month</td>
<td>$0.40 per GB/month</td>
<td>5.7</td>
</tr>
<tr>
<td>Administration</td>
<td>~140 servers/Administrator</td>
<td>&gt;1000 Servers/Administrator</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Data Center estimates from James Hamilton
Azure Content Delivery Network

• Better performance and user experience by caching Azure blobs at strategically placed locations
  – 18 locations globally (United States, Europe, Asia, Australia and South America) and growing
This has happened before…

Electrical Grid Adoption

- 5% in 1900
- 40% in 1907
- 80% in 1930
- 90% in 1935
The Rationale for DLs into the Cloud

- We can expect digital library environments will follow similar trends to the commercial sector
  - Leverage computing and data storage in the cloud
  - Small organizations need access to large scale resources
  - Scientists already experimenting with Amazon S3 and EC2 services

- For many of the same reasons
  - Little/no resource-sharing across library infrastructures
  - High storage costs
  - Physical space limitations
  - Low resource utilization
  - Excess capacity
  - High costs of acquiring, operating and reliably maintaining machines is prohibitive
  - Little support for developers, system operators
Realizing Jim Gray’s Vision for Data-Intensive Scientific Discovery

- Jim Gray = eScience
- A Transformed Scientific Method

**eScience: What is it?**
eScience is where “IT meets scientists.” Researchers are using many different methods to collect or generate data—from sensors and CCDs to supercomputers and particle colliders. When the data finally shows up in your computer, what do you do with all this information that is now in your digital shoebox? People are continually seeking me out and saying, “Help! I’ve got all this data. What am I supposed to do with it? My Excel spreadsheets are getting out of hand!” So what comes next? What happens when you have 10,000 Excel spreadsheets, each with 50 workbooks in them? Okay, so I have been systematically naming them, but now what do I do?
The Fourth Paradigm
Data-Intensive Scientific Discovery

EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE
“The impact of Jim Gray’s thinking is continuing to get people to think in a new way about how data and software are redefining what it means to do science.”

— Bill Gates, Chairman, Microsoft Corporation

“One of the greatest challenges for 21st-century science is how we respond to this new era of data-intensive science. This is recognized as a new paradigm beyond experimental and theoretical research and computer simulations of natural phenomena—one that requires new tools, techniques, and ways of working.”

— Douglas Kell, University of Manchester

“The contributing authors in this volume have done an extraordinary job of helping to refine an understanding of this new paradigm from a variety of disciplinary perspectives.”

— Gordon Bell, Microsoft Research

http://research.microsoft.com/fourthparadigm/
Listed 7 key areas for action by Funding Agencies:

1. Fund both development and support of software tools
2. Invest at all levels of the finding ‘pyramid’
3. Fund development of ‘generic’ Laboratory Information Management Systems
4. Fund research into scientific data management, data analysis, data visualization, new algorithms and tools
Remaining three key areas for action relate to the future of Scholarly Communication and Libraries:

5. Establish Digital Libraries that support the other sciences like the NLM does for Medicine

6. Fund development of new authoring tools and publication models

7. Explore development of digital data libraries that contain scientific data (not just the metadata) and support integration with published literature
Project Trident — a Scientific Workflow Workbench

Author, Execute and Monitor Workflows

Share workflows via 

Compose and modify workflows via drag & drop canvas

View data products, performance metrics, and provenance data, and write them directly into repository

Version 1.2 (Open Source under Apache 2.0 License):
http://tridentworkflow.codeplex.com/
Data Curation Add-in for Microsoft Excel

• Microsoft Research, in partnership with California Digital Library’s Curation Center
  – Collaboration with Tricia Cruse & John Kunze
  – Part of the DataONE (an NSF DataNet Project)

• Proposed functionality **under consideration:**
  – **Versioning** - revision history and original raw data can be protected and recovered
  – **Time stamps** - easily determine when the data were created and last updated
  – “**Workbook builder**” - select from globally shared standardized layouts for capturing data
  – **Export metadata in a standard formats** (e.g., a DataCite citation or an EML document that describes the dataset(s) in a workbook) so that researchers can readily share their data,
  – **Globally shared vocabulary of terms for data descriptions** (e.g., column names), and as needed to add new terms to the globally shared vocabulary, to enable wide collaboration between researchers
  – **Import term descriptions from the shared vocabulary** and annotate them to refine local definitions
  – **Deposit data and metadata into a data archive** to preserve and publish research data
Using Azure cloud to host standard sequence data
Run BLAST in parallel “map reduce” style for a large collection input
Use Excel-based tool to search databases.

- Performance: speedup
- 45x with 50 roles; 94x with 100 roles;
- with 300 roles computation time is reduced to 4 hours!
<table>
<thead>
<tr>
<th>Date</th>
<th>LAT</th>
<th>LON</th>
<th>DEPTH</th>
<th>MAG</th>
<th>TYPE</th>
<th>CATALOG</th>
<th>EVENT_CC</th>
<th>MAG</th>
<th>COMP_CODE</th>
<th>REGION</th>
<th>REGION_TEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/3/2010 22:24</td>
<td>3.78</td>
<td>127.11</td>
<td>63.5</td>
<td>4.7M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>263</td>
<td>TALAUD IS INDONESIA</td>
<td>TALAUD IS INDONESIA</td>
</tr>
<tr>
<td>3/3/2010 22:17</td>
<td>3.61</td>
<td>127.18</td>
<td>60</td>
<td>4.6M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>263</td>
<td>TALAUD IS INDONESIA</td>
<td>TALAUD IS INDONESIA</td>
</tr>
<tr>
<td>3/3/2010 21:37</td>
<td>3.68</td>
<td>127.25</td>
<td>35.3</td>
<td>4.8M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>263</td>
<td>TALAUD IS INDONESIA</td>
<td>TALAUD IS INDONESIA</td>
</tr>
<tr>
<td>3/3/2010 21:24</td>
<td>-38.46</td>
<td>-73.7</td>
<td>35</td>
<td>5.1M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 17:44</td>
<td>-36.4975</td>
<td>-73.1839</td>
<td>35</td>
<td>5.9 MW</td>
<td>MW</td>
<td>NEICALRT</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 17:44</td>
<td>-36.5</td>
<td>-73.18</td>
<td>35</td>
<td>5.9M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 17:44</td>
<td>-36.45</td>
<td>-73.07</td>
<td>19</td>
<td>6M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 15:25</td>
<td>-34.65</td>
<td>-72.43</td>
<td>33.5</td>
<td>4.9M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 12:50</td>
<td>-34.3</td>
<td>-72.25</td>
<td>35</td>
<td>4.9M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 11:13</td>
<td>3.16</td>
<td>127.08</td>
<td>64.5</td>
<td>5M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>263</td>
<td>TALAUD IS INDONESIA</td>
<td>TALAUD IS INDONESIA</td>
</tr>
<tr>
<td>3/3/2010 08:49</td>
<td>-17.18</td>
<td>-177.57</td>
<td>353.6</td>
<td>4.6M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>181</td>
<td>FIJI ISLANDS REGION</td>
<td>FIJI ISLANDS REGION</td>
</tr>
<tr>
<td>3/3/2010 07:51</td>
<td>5.82</td>
<td>-73.06</td>
<td>157.3</td>
<td>4.7M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>99</td>
<td>NORTHERN COLOMBIA</td>
<td>NORTHERN COLOMBIA</td>
</tr>
<tr>
<td>3/3/2010 07:09</td>
<td>-35.97</td>
<td>-72.61</td>
<td>35.1</td>
<td>4.9M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 06:16</td>
<td>-33.62</td>
<td>-71.96</td>
<td>35</td>
<td>5.1M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 05:49</td>
<td>-32.5</td>
<td>-71.71</td>
<td>35</td>
<td>4.7M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 04:35</td>
<td>-37.53</td>
<td>-73.77</td>
<td>30.2</td>
<td>5.1M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>135</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
<td>NEAR COAST OF CENTRAL CHILE</td>
</tr>
<tr>
<td>3/3/2010 04:35</td>
<td>35.55</td>
<td>-97.28</td>
<td>5</td>
<td>2.5M</td>
<td>M</td>
<td>FINGER</td>
<td>NEIC</td>
<td>NEIC</td>
<td>499</td>
<td>OKLAHOMA</td>
<td>OKLAHOMA</td>
</tr>
</tbody>
</table>

SeisimQuery: Event Query

Event selection criteria: magnitude between 1 and 9.9
Sustaining the Digital Investment

Issues and Challenges of Economically Sustainable Digital Preservation

February 2010


17 member group of experts from economics, computer science, libraries, archives, museums, and related fields. Supported by NSF, Andrew Mellon Foundation, JISC, US Library of Congress

What are technical solutions?

• We have two main strategies:

  content migration
  – Migration to standards that are likely to be supported in the future.

  emulation and simulation
  – Create emulators of hardware and simulators of software systems to enable old programmes to run and old data to be used.
Why should Microsoft Engage

• We and computing industry are the closest to the problems
  – We create the technology. We can make it live longer and continue to bring benefits to all of us.
  – All our research areas are applicable, from technologies for media processing to retrieval, security, program checking and verification, and system emulation.
• 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
PLANETS Partners

The British Library
National Library, Netherlands
Austrian National Library
State and University Library, Denmark
Royal Library, Denmark

National Archives, UK
Swiss Federal Archives
National Archives, Netherlands

Hatii at University of Glasgow
University of Freiburg
Technical University of Vienna
University at Cologne

Tessella Plc
IBM Netherlands
Microsoft Research, Cambridge
ARC Seibersdorf research
Towards Effective Digital Preservation

**TECHNOLOGY**
- Tools, methods, training
- Support for data processing and curation
- Standards for content handling and representation

**SOCIAL**
- Definition of new roles and skill sets for librarians and archivists

**ECONOMIC**
- Cost of preservation is enormous. Who will pay for it? Who will ensure that it remains accessible.
- Requires development of an ecosystem to support economically viable digitization efforts

**LEGAL**
- Legal framework to enforce and protect the rights of authors and publishers, privacy and security.
Microsoft & PLANETS: Preserving Office Documents

- Microsoft Research role within PLANETS:
  - Conversion of binary Microsoft Office Documents into Office Open XML File Format (OpenXML)

- We extended the effort to include other formats
  - More legacy formats, e.g. WordPerfect
  - Other open standards, e.g. Open Document Format.
Document Conversion Tools – Our Approach

• Three-step approach, resulting in a modular and extendible infrastructure
  – **Identify** existing conversion tools and libraries
  – **Wrap** these tools and libraries into re-usable components
  – **Integrate** these components into PLANETS and other systems.

• If possible, do not use the office applications (e.g., Microsoft Office or OpenOffice.org)
  – They are designed as interactive applications
  – Message boxes might pop up ("Do you want ...")
  – Unclear license question when running on a server.
Supported Formats

- **Source formats**
  - WordPerfect 5
  - WordPerfect 6
  - DOS Word
  - Word 2, 6, 95
  - Word 97-2003
  - RTF
  - ODF
  - OpenXML

- **Target formats**
  - OpenXML
  - ODF
  - UOF
  - HTML
  - XCDL (format defined in PLANETS/PC)
Conversion service

Convert your documents by e-mail.

- convert up to three documents for free
- have converted documents sent to you by e-mail
- no registration required
- available for multiple formats

1. Select the format of your current document.
   Convert From

2. Select the format you want your document to convert to.
   Convert To

3. Send an email with your files attached to the following address:

Please remember that only 3 documents can be converted.

Want more?

Register with our service now. It's free!
Conversion service

Convert your documents by e-mail.
- convert up to three documents for free
- have converted documents sent to you by e-mail
- no registration required
- available for multiple formats

1. Select the format of your current document.
   WordPerfect 6 (*.wp,.wpd)

2. Select the format you want your document to convert to.
   Microsoft Word 2007 (*.docx)

3. Send an email with your files attached to the following address.
   convert2docx@mcp.dialogika.de

Please remember that only 3 documents can be converted.
The conversion of your document was successful

Thank you for using DOC Conversion Services!
Please find attached your converted document.

Do you want MORE?
Register for our service now. It's free!

- view, convert and compare documents online
- convert up to 200 MB at a time
- store documents for up to 30 days
- share documents with friends and colleagues

[Register online for free]
## Conversion service

### File Manager

<table>
<thead>
<tr>
<th>Uploaded Files</th>
<th>Converted Files</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>File Name</strong></td>
<td><strong>File Format</strong></td>
</tr>
<tr>
<td>GREEK.odt</td>
<td>ODT</td>
</tr>
<tr>
<td>GREEK.docx</td>
<td>DOCX</td>
</tr>
<tr>
<td>WP5_GREEK.docx</td>
<td>DOCX</td>
</tr>
<tr>
<td>chi2010pubsformat.docx</td>
<td>DOCX</td>
</tr>
<tr>
<td>chi2010pubsformat.odt</td>
<td>ODT</td>
</tr>
<tr>
<td>SystemQuality-CKM2009-draft8.5.odt</td>
<td>ODT</td>
</tr>
<tr>
<td>SearchMobilPoster-WWW12.docx</td>
<td>DOCX</td>
</tr>
<tr>
<td>GREEK.odt</td>
<td>ODT</td>
</tr>
</tbody>
</table>

Select:  All None

Download Remove
Επίτροπη
ΚΑΝΟΝΙΣΜΟΣ (ΕΚ) αριθ. ΤΟΥ ΣΥΜΒΟΥΛΙΟΥ
της
για την καθοδηγήσεις για την περίοδο εμπορίας 1996-97, των τιμών βάσεως και
tιμών εγγράφων των εφαρμοσμένων στον τομέα των οπωροκηπευτικών

ΤΟ ΣΥΜΒΟΥΛΙΟ ΤΗΣ ΕΥΡΩΠΑΪΚΗΣ ΕΝΟΣΗΣ.

Έχει εκδηλώσει:

η συνθήκη για την ίδρυση της Ευρωπαϊκής Κοινότητας,
tον κανονισμό (ΕΟΚ) αριθ. 1035/72 του Συμβουλίου, της 18ης Νοεμβρίου 1972, περί κοινώς
οργανώσεως της αγοράς στον τομέα των οπωροκηπευτικών, ένας τροποποιημένος
tην πρόταση της Επιτροπής,
tη γνώμη του Ευρωπαϊκού Κοινοβουλίου,
tη γνώμη της Οικονομικής και Κοινωνικής Επιτροπής,

Εκτιμάται:

(PPRIX:EL/0011.wpd)
E1-1

Πρόταση
ΚΑΝΟΝΙΣΜΟΣ (ΕΚ) αριθ. ΤΟΥ ΣΥΜΒΟΥΛΙΟΥ
της
για την καθοδηγήσεις για την περίοδο εμπορίας 1996-97, των τιμών βάσεως και
tιμών εγγράφων των εφαρμοσμένων στον τομέα των οπωροκηπευτικών

ΤΟ ΣΥΜΒΟΥΛΙΟ ΤΗΣ ΕΥΡΩΠΑΪΚΗΣ ΕΝΟΣΗΣ.

Έχει εκδηλώσει:

η συνθήκη για την ίδρυση της Ευρωπαϊκής Κοινότητας,
tον κανονισμό (ΕΟΚ) αριθ. 1035/72 του Συμβουλίου, της 18ης Νοεμβρίου 1972, περί κοινώς
οργανώσεως της αγοράς στον τομέα των οπωροκηπευτικών, ένας τροποποιημένος
tην πρόταση της Επιτροπής,
tη γνώμη του Ευρωπαϊκού Κοινοβουλίου,
tη γνώμη της Οικονομικής και Κοινωνικής Επιτροπής,
Open XML to HTML Translator Overview

- Resulted from direct feedback from the Document Interoperability Initiative

- The main goal of the OpenXML Document Viewer project is to create software tools, plus guidance, showing how documents created using Open XML Format can be translated to HTML

- The OpenXML Viewer is available under the open source Microsoft Public License (Ms-PL)
Open XML to HTML Translation Steps

1. Unzip the document
2. Parse docx into single XML
3. Pre-process parsed XML
4. Convert VML to SVG
5. Apply XSL Transforms
Open XML to HTML Translator Architecture

- Application Interface allows transformation engine to be leveraged across multiple browsers and operating systems
- Transformation Service provides a reusable core engine
- Utility classes provide support functions including XSL transformation, which does the bulk of the heavy lifting
More information

• Microsoft Interoperability Blog

• Open XML Projects
  • HTML Translator – [http://openxmlviewer.codeplex.com](http://openxmlviewer.codeplex.com)
  • Binary Translator – [http://b2xtranslator.sourceforge.net](http://b2xtranslator.sourceforge.net)

http://www.documentinteropinitiative.org/
Evolve from DLs being a ‘place’ where one goes to store/discover digital objects to the DL as a set of services woven into the tools & workflows of the researcher.
SWORD Protocol

Facilitating Deposit
Article Authoring Add-in for Word

Read, convert, and author NLM XML documents

Repository deposit via SWORD

ORE Resource Map creation

v.2 beta 3:
http://research.microsoft.com/authoring/
From a SharePoint site researchers can...

Select any file stored in SharePoint:
- Document
- Presentation
- Image
- Data files

and publish it to any repository (via SWORD)

- SWORD endpoints are managed as a custom list, so new locations are easily added
Send this file to Repository X

Send To:
- arXiv.org
- PubMed
- eprints
- Zentity
- DSpace
Conference Management Tool

http://cmt.research.microsoft.com

CMT to arXiv via SWORD
The Setup (Conference Chair)
Enter Conference Account information from arXiv
List of available collections from arXiv

Select data to deposit

Select data to deposit

List of available collections from arXiv

Pre-configured question(s) supplied by arXiv.

For multi-track conferences, select which ones are published to arXiv.
Publishing Workflow (Conference Chair)
Conference Chair selects papers to publish to arXiv, and an email confirmation request is sent to the authors.
### Publishing Workflow (Author)

<table>
<thead>
<tr>
<th>Last Modified Time</th>
<th>File Uploaded On</th>
<th>Paper Decision</th>
<th>Camera-Ready</th>
<th>Presentations</th>
<th>Publishing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>arXiv</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Request Status: Not Responded</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><a href="#">Accept</a>  <a href="#">Decline</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **File Not Uploaded**
- **File Not Uploaded**
- **File Not Uploaded**
- **File Not Uploaded**
- **File Not Uploaded**
Publishing Workflow (Author)

If author accepts, then Collection & primary category must be specified

Author must agree to arXiv licensing terms
Once accepted by the Author, the Chair can publish directly to arXiv.
#depositmo

- Interactive Multi-Submission Deposit Workflows for Desktop Applications
  - “Changing the culture, embedding deposit into the natural everyday workflow of researchers and lecturers”

http://blogs.ecs.soton.ac.uk/depositmo
Article Authoring Add-in for Word

Read, convert, and author NLM XML documents

Repository deposit via SWORD

ORE Resource Map creation

v.2 beta 3: http://research.microsoft.com/authoring/
Abstract Dynamics of Plasmodium Sporogony

Abstract

The population dynamics of Plasmodium sporogony within mosquitoes consists of an early phase where parasite abundance decreases during the transition from gametocyte to oocyst, an intermediate phase where parasite abundance remains static as oocysts, and a later phase where parasite abundance increases during the release of progeny sporozoites from oocysts. Sporogonic development is complete when sporozoites invade the mosquito salivary glands. The dynamics and efficiency of this developmental sequence were determined in laboratory strains of Anopheles dirus, Anopheles minimus and Anopheles sawadwongporni mosquitoes for Plasmodium vivax parasites circulating naturally in western Thailand.

Methods

Mosquitoes were fed blood from 20 symptomatic Thai adults via membrane feeders. Absolute densities were estimated for macrogametocytes, round stages (= female gametes/zygotes), ookinetes, oocysts, haemolymph sporozoites and salivary gland sporozoites. From these census data, five aspects of population dynamics were analysed: 1) changes in life-stage prevalence during early sporogony, 2) kinetics of life-stage formation, 3) efficiency of life-stage transitions, 4) density relationships between successive life-stages, and 5) parasite aggregation patterns. (Fink, et al., 2010)

Results

There was no difference among the three mosquito species tested in total losses incurred by P. vivax populations during early sporogony. Averaged across all infections, parasite populations incurred a 68-fold loss in abundance, with losses of ca. 16-fold, 2-fold and 2-fold at the first (=
Abstract

The population dynamics of Plasmodium sporogony within mosquitoes consists of an early phase where parasite abundance decreases during the transition from gametocyte to oocyst, an intermediate phase where parasite abundance remains static as oocysts, and a later phase where parasite abundance increases during the release of progeny sporozoites from oocysts. Sporogonic development is complete when sporozoites invade the mosquito salivary glands. The dynamics and efficiency of this developmental sequence were determined in laboratory strains of Anopheles dirus, Anopheles minimus and Anopheles sawadwongporni mosquitoes for Plasmodium vivax parasites circulating naturally in western Thailand.

Methods

Mosquitoes were fed blood from 20 symptomatic Thai adults via membrane feeders. Absolute densities were estimated for macrogametocytes, round stages (= female gametes/zygotes), ookinetes, oocysts, haemolymph sporozoites and salivary gland sporozoites. From these census data, five aspects of population dynamics were analysed: 1) changes in life-stage prevalence during early sporogony, 2) kinetics of life-stage formation, 3) efficiency of life-stage transitions, 4) density relationships between successive life-stages, and 5) parasite aggregation patterns. (Fink, et al., 2010)

Results

There was no difference among the three mosquito species tested in total losses incurred by P. vivax populations during early sporogony. Averaged across all infections, parasite populations incurred a 68-fold loss in abundance, with losses of ca. 19-fold, 2-fold and 2-fold at the first /=
Abstract

The population dynamics of Plasmodium sporogony within mosquitoes consists of an early phase where parasite abundance decreases during the transition from gametocyte to oocyst, an intermediate phase where parasite abundance remains static as oocysts, and a later phase where parasite abundance increases during the release of progeny sporozoites from oocysts. Sporogonic development is complete when sporozoites invade the mosquito salivary glands. The dynamics and efficiency of this developmental sequence were determined in laboratory strains of Anopheles dirus, Anopheles minimus and Anopheles sawadwongporni mosquitoes for Plasmodium vivax parasites circulating naturally in western Thailand. (Bockarie & Dagoro, 2006)

Methods

Mosquitoes were fed blood from 20 symptomatic Thai adults via membrane feeders. Absolute densities were estimated for macrogametocytes, round stages (= female gametes/zygotes), ookinetes, oocysts, haemolymph sporozoites and salivary gland sporozoites. From these census data, five aspects of population dynamics were analysed: 1) changes in life-stage prevalence during early sporogony, 2) kinetics of life-stage formation, 3) efficiency of life-stage transitions, 4) density relationships between successive life-stages, and 5) parasite aggregation patterns. (Fink, et al., 2010)

Results

There was no difference among the three mosquito species tested in total losses incurred by P. vivax populations during early sporogony. Averaged across all infections, parasite populations incurred a 68-fold loss in abundance, with losses of ca. 18-fold, 2-fold and 2-fold at the first (=
Chemistry Add-in for Word

Author/edit 1D and 2D chemistry. Change chemical layout styles.

Intent: Recognizes chemical dictionary and ontology terms

Data: Semantics stored in Chemistry Markup Language

Intelligence: Verifies validity of authored chemistry

Relationships: Navigate and link referenced chemistry

• Peter Murray-Rust
• Joe Townsend
• Jim Downing

Open Source Project (Apache 2.0 License)
http://research.microsoft.com/chem4word/
1.1 Specific experimental procedures

1.2 Experimental procedures for chapter 3: decaestrechinate D

1.2.1 Preparation of (2E) – 3 – (3 – methyloxiran – 2 – yl)prop – 2 – en – 1 – ol (173)

Diisobutylaluminium hydrid (17.70 ml of a 1 M solution in THF, 17.70 mmol) was slowly added to the epoxide ester 172 (1.20 g, 7.68 mmol) in THF (10 ml) at -78°C. After stirring at this temperature for 1h, methanol (10 ml) was added slowly and the resultant solution was warmed to rt. TEA (8 ml) was subsequently added and the mixture was stirred at rt overnight. Filtration through a pad of Celite® followed by washing with Et2O (150 ml)
1.1 Specific experimental procedure

1.2 Experimental procedure

1.2.1 Preparation of (2S,3S)-5-[(3,3-dimethylallyl)prop-2-en-1-yl]prop-2-en-1-ol (173)

Diisobutylaluminium hydride (1.06 mmol, 3.57 g, 20% solution in THF, 17.70 mmol) was slowly added to the epoxide ester 172 (1.20 g, 7.68 mmol) in THF (10 ml) at -78°C. After stirring at this temperature for 1h, methanol (10 ml) was added slowly and the resultant solution was warmed to rt. TEA (8 ml) was subsequently added and the mixture was stirred at rt overnight. Filtration through a pad of Celite® followed by washing with Et₂O (150 ml)
1.1 **Specific experimental procedures**

1.2 **Experimental procedures for chapter 3: decaestrictine D**

1.2.1 Preparation of (2E) - 3 - (3 - methyloxiran - 2 - yl)prop - 2 - en - 1 - ol (173)

---

Diisobutylaluminium hydrid (17.70 ml of a 1 M solution in THF, 17.70 mmol)
Worldwide Telescope

http://www.worldwidetelescope.org/

TIME magazine
“50 Best sites on the Internet 2009”
### Query Results from the Astronomy Database

Retrieved **200** abstracts, starting with number **1**. Total number selected: **393**.

<table>
<thead>
<tr>
<th>#</th>
<th>Bibcode</th>
<th>Authors</th>
<th>Score</th>
<th>Date</th>
<th>List of Links</th>
<th>Access Control</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2009ApJ...700.1609M</td>
<td>Myers, Philip C.</td>
<td>1.000</td>
<td>08/2009</td>
<td>A Z E F L X</td>
<td>R C S</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>Filamentary Structure of Star-forming Complexes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2009ApJ...700.1190D</td>
<td>Desai, Vandana; Soifer, B. T.; Dey, Arjun; LeFloc'h, Emeric; Armus, Lee; Brand, Kate; Brown, Michael J. I.; Brodwin, Mark; Jannuzi, Buell T.; Houck, James R.; and 8 coauthors</td>
<td>1.000</td>
<td>08/2009</td>
<td>A Z E F L X</td>
<td>R C S</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>Strong Polycyclic Aromatic Hydrocarbon Emission from z = 2 ULIRGs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The initial conditions of isolated star formation - IX. Akari mapping of an externally heated pre-stellar core</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A spatial study of the mid-IR emission features in four Herbig Ae/Be stars</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mid-infrared spectroscopy of infrared-luminous galaxies at z ~ 0.5-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“alpha” Faceted Topic Search in ADS (courtesy of Michael Kurtz & Alberto Accomazzi)
list of objects with links to WWT browser
And now we got to NGC 7023 by using the literature as a filter.
Multilingual

WORLDWIDE SCIENCE.ORG Beta

View Demo
Consider this . . .
While English is the *lingua franca* for science, these are the world’s most widely spoken languages:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Language</th>
<th>Estimated Number of Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mandarin Chinese</td>
<td>1,051,000,000</td>
</tr>
<tr>
<td>2</td>
<td>English</td>
<td>510,000,000</td>
</tr>
<tr>
<td>3</td>
<td>Hindi/Urdu</td>
<td>490,000,000</td>
</tr>
<tr>
<td>4</td>
<td>Spanish</td>
<td>420,000,000</td>
</tr>
<tr>
<td>5</td>
<td>Arabic</td>
<td>280,000,000</td>
</tr>
<tr>
<td>6</td>
<td>Russian</td>
<td>255,000,000</td>
</tr>
<tr>
<td>7</td>
<td>Portuguese</td>
<td>230,000,000</td>
</tr>
<tr>
<td>8</td>
<td>German</td>
<td>229,000,000</td>
</tr>
<tr>
<td>9</td>
<td>Bengali</td>
<td>215,000,000</td>
</tr>
<tr>
<td>10</td>
<td>French</td>
<td>130,000,000</td>
</tr>
<tr>
<td>11</td>
<td>Japanese</td>
<td>127,000,000</td>
</tr>
</tbody>
</table>

(Source: Wikipedia)
Of the world’s “top 400” institutional repositories, 250, or 63%, have some or all non-English content.

Examples:
- HAL CNRS -- French
- Kyoto University Research Repository – Japanese
- Leiden University Digital Repository -- Dutch
- CSIC (Spanish National Research Council)

(Source: Cybermetrics Lab, Spain)
Major Non-English Science “Producers”
• Japan
• France
• Germany
• Brazil

. . . and many other countries.
To further accelerate access to science, multilingual translations are needed in both directions:

- Translation of English content for non-English speakers . . . and . . .
- Translation of non-English content for English speakers
Now, we have the essential ingredients for real-time translation of science

- National science databases in multiple languages
- Federated search
- Multilingual translation on both front and back end of the user experience

A public-private partnership, introduced as *Multilingual WorldWideScience.org*\(^\text{Beta}\)
Here’s how it works . . .

2. MWWS.org uses Microsoft to translate the Chinese query into individual languages of source databases (English, French, Portuguese, Russian, etc.)
3. MWWS.org sends the translated queries to corresponding databases, which search their contents and return results in native languages to MWWS.org.
4. MWWS.org uses Microsoft to translate native language results into Chinese and presents results to the user in relevance-ranked order.

Conversely, an English-speaking user could have a query translated into languages of non-English databases and then get results back in English.
With the launch of Multilingual WorldWideScience.org, we are . . .

• Opening vast reservoirs of heretofore under-utilized scientific knowledge
• Providing equal access to science for anyone on the Internet
• Promoting scientific collaboration, participation, and transparency

. . . and accelerating scientific discovery!
Strategic groundwater resources in the Tagliamento River basin (northern Italy): hydrogeological investigation integrated with geophysical exploration

1. Rapti-Caputo, Bratus; Santarato
2. Hydrogeology journal 2009-01-01
3. Refdoc (French)
Los recursos estratégicos de las aguas subterráneas de la cuenca del río Tagliamento (norte de Italia): investigación hidrogeológica integrado exploration
Ressources subterraines géophysiques aqua estratégico

cuenca de el río Tagliamento (norte de Italia) con: investigaciones hidrogeológicas junto con la exploración de aguas subterráneas en la cuenca del Río Tagliamento (Norte de Italia) estratégicos
géophysique
Recursos: una investigación hidrogeológica integrada con la exploración
de agua subterránea na bacia Rio Tagliamento (Norte de Italia) hidrográfica: investigación hidrogeológica contemplando prospección geofisica

Original Title: Strategic groundwater resources in the Tagliamento River basin (northern Italy): hydrogeological investigation integrated with geophysical exploration

Ressources souterraines stratégiques en eau du bassin de la rivière Tagliamento (Italie du Nord): investigations hydrogéologiques couplées avec une exploration géophysique

Recursos estratégicos de aguas subterráneas en la cuenca del Rio Tagliamento (Norte de Italia): una investigación hidrogeológica integrada con la exploración de agua subterránea na bacia Rio Tagliamento (Norte de Italia); investigación hidrogeológica contemplando prospección geofisica

Rapti-Caputo; Bratus; Santarato

Hydrogeology journal 2009-01-01
Titre du Document
Strategic groundwater resources in the Tagliamento River basin (northern Italy)
exploration

Ressources souterraines stratégiques en eau du bassin de la rivière Tagliamento
une exploration géophysique

Recursos estratégicos de aguas subterráneas en la cuenca del Río Tagliamento
la exploración geofísica

Recursos estratégicos de agua subterránea na bacia hidrográfica do Rio Tagliamento
prospeção geofísica

Auteur(s)
Rapti-Caputo Dimitra ; Bratus Antonio ; Santarato Giovanni

Résumé
Le secteur Ouest du bassin de la rivière Tagliamento (Région Frioul-Venezia-Jul

Titulo del documento
Los recursos estratégicos de las aguas subterráneas de la cuenca del río Tagliamento
con exploración geofísica

Cuenca de agua estratégica de los recursos de aguas subterráneas del río Tagliamento
con una exploración geofísica

Recursos estratégicos águas subterráneas en la cuenca del Río Tagliamento
la exploración geofísica con integrada

Recursos água subterrânea estratégicos na bacia Rio Tagliamento (Norte do
prospeção geofísica

Autor(s)
Dimitra Rapti-Caputo, Antonio; Giovanni Santarato Bratus

Resumen
La cuenca occidental del Tagliamento de río (región de Friul-Venecia-Julía, norte

The Microsoft Translator web page widget allows you to bring real-time, in-place translation to your web site. Users can see your pages in their own language, without having to go to a separate translation web site, and share your page with friends in multiple languages. You can learn more about how to use the widget, get help or interact with other website owners on the Microsoft Translator forums.

Site info

* Site address: http://www.example.com/

Input the site address where you would be displaying the widget. Learn more here.

* Site language: English

Select the content language of your website.

Options

Translation:

If you choose to display the widget then any visitor to your site can invoke translation by clicking it (manual). In addition you may choose to have the translation toolbar appear to visitors whose browser language is different than your site language and offer to translate the page (notify) or to do so automatically (auto), in which case displaying the widget is optional.

- Manual
- Notify
- Auto

Display widget control

Color:

- Blue
- Black
- Green
- Orange
- Red
- Yellow
- Grey

Widget width:

200

Preview:

Translate this page

Microsoft® Translator
Translation is a tool, not a destination. Following up on this philosophy, Microsoft Translator developer offerings provide developers a wide spectrum of translation and language APIs for integration into their applications and services. For commercial licensing information, you can contact mtlic@microsoft.com.

**AJAX Interface**
The AJAX interface provides your web applications with JavaScript based access to translation services.

**SOAP Interface**
The SOAP interface provides a strongly typed, web service standards based programming model supporting rich development environments such as the .NET framework.

**HTTP Interface**
The HTTP interface provides a high degree of interoperability for any network enabled application and supports a broad range of development environments.

**MSDN reference for Microsoft Translator APIs**
Comprehensive reference for Microsoft Translator APIs on MSDN.

**Interactive SDK**
Start developing right now with online demos and source code with the Microsoft Translator interactive SDK. Try out the various APIs and look at the code behind.

**Getting started guide (ASP.NET)**
Use this getting started guide to learn how to quickly enable your ASP.NET application to consume translations.

**News, Feedback & Support**
- Microsoft Translator Blog
- Microsoft Translator Developer Forum
Questions?

Alex Wade
Director — Scholarly Communication
Microsoft Research
awade@microsoft.com
http://research.microsoft.com/people/awade

URL – http://www.microsoft.com/scholarlycomm/
Facebook: Scholarly Communication at Microsoft