Integrating Distributed Scholarly Content

Papyri.info as Case Study
Overview

• How a research community has embraced the online environment for research and teaching

• Supporting open-ended scholarly research tool within a library-based digital program

• The technical challenges of broadening the scope of what was originally a collection-based project

• Sustaining in such a project into the indefinite future
Presenters

• Rodney Ast (papyrologist), APIS Coordinator

• Stephen Davis (manager), Director, Libraries Digital Program

• Ben Armintor (programmer / analyst), Lead papyri.info Programmer
Rodney Ast
APIS Coordinator
Writing in the Ancient World
Ancient Texts

- Pharaonic Times – Middle Ages
- Many Different Languages
  - Egyptian (Hieroglyphic, Hieratic, Demotic, Coptic)
  - Greek
  - Latin
  - Arabic
- Many Different Types of Texts
  - Literature (Homer, Tragedy, Philosophy, History, Lyric Poetry)
  - Religious Texts
  - Imperial Decrees
  - Mundane Documents
Epistles of Paul

Gospel of Judas
Elegiac Poem by Cornelius Gallus (70-26 BC)

New Poetry of Sappho (7th/6th c. BC)
Love Charm to Win Heart of Another (4th c. AD)

Affidavit in a Case of Domestic Abuse (4th c. AD)
Order of the Roman Prefect Banning Oracles and Magic (198-199 AD)

“Encountering many who believed themselves to be deceived by the practices of divination I quickly considered it necessary . . . to enjoin all people to abstain from this hazardous inquisitiveness. Therefore, let no man through oracles, that is, by means of written documents supposedly granted under divine influence . . . pretend to know things beyond human knowlege and profess (to know) the obscurity of things to come . . . .”
Letter of a Woman to Her Mother
(early 2nd c. AD)

Thermouthas to Valerian, her mother, very many greetings and continued good health. I received from Valerius a basket in which were twenty pairs of cakes of wheat and ten pairs of loaves of bread. Send me the blankets of the price (agreed on) and fine wool, four shearings. Do you give them to Valerius. We have just been seven months pregnant. I salute Artemis and the little Nikarous and Valerius my lord - I long for him in my heart - and Dionysia and Demetrous, and the little Taesis often, and all those in the house. And send me news of how my father is, since he was ill when he left me. I salute grandmother, Rhodine salutes you. I have put her out at handwork: I need her again, but I am well. Phoebi 8:(Verso):Deliver to Philadelphia to Valerias, my mother.
Subjects of Study

- Ancient Economics
- Military, Political and Social History
- Church History and Religious Studies
- Literature
- Ancient Law
- History of Science and Medicine
- Paleography
- Linguistics
APIS and Related Projects

• **APIS**: Catalog with Metadata Related to Papyri & Clay Tables (incl. image links)

• **DDbDP**: Searchable Database of Ancient Texts

• **HGV**: Browsable Catalog of Bibliographic Information Pertaining to Ancient Documents (incl. links to APIS and other similar databases)
APIS
Advanced Papyrological Information System

Content as of 11/2007

• 31,000 Catalog Records
• 23,400 Unique Images
• 5,400 Translations
APIS Partners & Contributors 1

1995
- University of Calif., Berkeley
- Columbia University
- Duke University
- University of Michigan
- Princeton University
- Yale University

2003
- University of Pennsylvania
- University of Toronto

2004
- University of Chicago
- New York University

2005
- Badè Museum of Biblical Archaeology (PSR)
- California State Univ., Sacramento
- The Hermitage Museum
- University of Oslo
- Perkins School of Theology (SMU)
- Stanford University
- Washington State Univ., Pullman
- Univ. of Wisconsin at Madison
2006
- American Center of Oriental Research (ACOR), Petra Papyri
- Berenike Project

2007
- Oxford University, Center for Study of Ancient Documents
- Princeton Theol. Seminary

2008 –
- British Museum
- Catholic University of America
- Amheida Excavations (Columbia U.)
- Fordham University
- The Kellis Project
- University of Lund
- Union Theological Seminary
<table>
<thead>
<tr>
<th>Item</th>
<th>APIS record: berenike.apis.102</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Receipt, August 6, 33 A.D.</td>
</tr>
<tr>
<td>Summary</td>
<td>Receipt for the delivery of a load of barley.</td>
</tr>
<tr>
<td>Full-Text (DBDP)</td>
<td>link to Perseus Project, O.Berenike:1:102</td>
</tr>
<tr>
<td>Inventory Id</td>
<td>Inv. BE97-13-A-150</td>
</tr>
<tr>
<td>Publications</td>
<td>O.Berenike I 102</td>
</tr>
<tr>
<td>Original Language</td>
<td>Greek</td>
</tr>
<tr>
<td>Physical Description</td>
<td>1 ostracon.</td>
</tr>
<tr>
<td>Notes</td>
<td>7 lines of text, Roman dump at Berenike.</td>
</tr>
<tr>
<td>Notes on Custodial History</td>
<td>Currently stored in a magazine of the Supreme Council of Antiquities in Quft.</td>
</tr>
<tr>
<td>Translation</td>
<td>&quot;-tios, lancer, to —, son of Didymous, greetings. I have received from you 6 artabas of barley, total 6, art., which Galus Iullus —, soldier, became surety for. Year 19 of Tiberius Caesar Augustus, Mesore 13.&quot;</td>
</tr>
</tbody>
</table>

Images
- large-sized image of front tile: 0.0
- medium-sized image of front tile: 0.0
- small-sized image of front tile: 0.0
- thumbnail-sized image of front tile: 0.0

Notice: Each library participating in APIS has its own policy concerning the use and reproduction of digital images included in APIS. Please contact the owning institution if you wish to use any image in APIS or to publish any material from APIS.
Barley Receipt (Aug. 6, 33 AD)
The Duke Databank of Documentary Papyri

*O. Berenike: Documents from Berenike*

Your current position in the text is marked in red. Click anywhere on the line to jump to another position.

Table of Contents

Go to 102

1.102.

Date: AD33; Location: Berenike

\[c 11\]. τις λογχέως

\[c 7\] Διδύμον χαίρειν:

\[έχω\] παρὰ σου {κ} ἄκε ἐνεργ.

\[ηπη\] Γάειος Ιόλυς

5 \[c 6\] στρατιώτης κοληῆς

[ἀρτάβαις] 6, (γίνονται) (ἀρτάβαι) 6, (ἔτους) 19 Τίβεριον


1 λογχευς Pap. 2 στρατιώτης Pap.
Klicken Sie auf den Band im "Heidelberger Gesamtverzeichnis", in dem sich der gesuchte Text befindet, sodann in der zweiten Spalte auf O.Ber. 1101.

Heidelberger Gesamtverzeichnis der griechischen Papyrusurkunden Ägyptens

- Publikation: O.Ber. 1102
- Datierung: 33, 6. Aug.
- Ort: Berenike
- Originaltitel: keiner
- Material: Ostrakon
- Abbildung: keine
  - Abbildung im Internet
- Text der DDBDP: Server in Somerville, Server in Berlin

Bemerkungen:

- Inhalt: Bestätigung, Empfang, Gerste, Soldaten

O.Bahri
O.Bahri Div.
O.BawitFAQ
O.Ber. I
O.Ber. II
O.Berl.
O.Bodl. I
O.Bodl. II
O.Bodl. III
O.Bru.
O.Bu Njem
O.Buch.
O.Cair. Cat.
O.Cair. GPW
O.Camb.
Stephen Paul Davis

Director, Libraries Digital Program
Putting It Together

APIIS + DDbDB + HGV + ?

= Papyri.info

(Papyrological Navigator)
Faculty & Scholars

Columbia University
Roger Bagnall, Professor Emer. of Greek & Latin Languages & Professor of History at Columbia & first Director, NYU Institute for the Study of the Ancient World

Duke University
Joshua D. Sosin, Assistant Professor, Classical Studies

Universität Heidelberg
Dr. James M. S. Cowey, Papyrologist, Heidelberg Documentary Papyri Project
Navigator Desiderata

- Integrated metadata search and retrieval
- Composite display of images, metadata, transcriptions & translations
- Visible data source branding
- Highly functional image display capability (e.g., zoomable)
- Full text searching of DDBDP Greek transcriptions
- Scalable content via additional data sources
- Personalization features
System Architecture Options

A. Real-time search, retrieval & display from remote data sources (Z39.50-like)

B. OAI harvesting to create locally hosted union catalogue
   or

C. Locally-maintained copies of source data
papyri.info Architecture A
Real-time meta-search

Navigator / Portal

Protocol

Internet

Protocol

Protocol

Protocol

Columbia APIS
Catalog records
Translations
Images & links

Duke Databank
Full text transcriptions
Editorial apparatus

Heidelberg Gv
Citations & references
Translations
Image links
papyri.info Architecture A
Real-time meta-search

Navigator / Portal

Protocol

Internet

Protocol

Columbia APIS
Catalog records
Translations
Images & links

Protocol

Duke Databank
@ Perseus
Full text transcriptions
Editorial apparatus

Protocol

Heidelberg Gv
Citations & references
Translations
Image links

Not implemented because:
- No extent protocol
- Perseus problems
- Heidelberg technology
papyri.info Architecture B
Offline harvesting

Navigator / Portal

Composite database

Protocol

Internet

Protocol

Columbia APIS
Catalog records
Translations
Images & links

Protocol

Duke Databank
Full text transcriptions
Editorial apparatus

Protocol

Heidelberg Gv
Citations & references
Translations
Image links
Papyri.info Architecture B

Offline harvesting

Navigator / Portal

Composite database

Protocol

Internet

Not implemented because:
- OAI not workable
- Limited tech support
- Perseus problems

Protocol

Columbia APIS
Catalog records
Translations
Images & links

Protocol

Duke Databank
Full text transcriptions
Editorial apparatus

Protocol

Heidelberg Gv
Citations & references
Translations
Image links
papyri.info Architecture C
Local Copies of Source DBs

Navigator / Portal

Local copies of databases

- Columbia APIS
- Duke Databank
- Heidelberg Gv
**papyri.info Architecture C**

**Local Copies of Source DBs**

- **Navigator / Portal**
- **Implemented because:**
  - Stable environment for prototyping
  - No protocol development
  - DDBDP public access transferred to Columbia

- **Local copies of databases**
  - Columbia APIS COPY
  - Duke Databank COPY
  - Heidelberg Gv COPY

Columbia University Libraries
Digital Program
Data sources

Columbia APIS
Catalog records
Translations
Images & links

Duke Databank
Full text transcriptions
Editorial apparatus

Heidelberg Gv
Citations & references
Translations
Image links
**papyri.info Overview 2**

### Data sources

- **Columbia APIS**
  - Catalog records
  - Translations
  - Images & links

- **Duke Databank**
  - Full text transcriptions
  - Editorial apparatus

- **Heidelberg Gv**
  - Citations & references
  - Translations
  - Image links

### Portal tasks

- Item matching
- Metadata merger
- Metadata indexing
- Full text indexing
- Image mgmt & display
papyri.info Overview 3

Data sources

**Columbia APIS**
- Catalog records
- Translations
- Images & links

**Duke Databank**
- Full text transcriptions
- Editorial apparatus

**Heidelberg Gv**
- Citations & references
- Translations
- Image links

Portal tasks

- Item matching
- Metadata merger
- Metadata indexing
- Full text indexing
- Image mgmt & display

Portal features

*In production (2007):*
- Integrated
  * Displays
  * Metadata searching
- Flexible presentation
- Source ‘branding’
- Scalable content

*To come by 2008:*
- Greek text searching
  * String
  * Lemmatized
- Personalization
Putting It Together (continued)
Rodney Ast
Papyri.info & the Papyrological Navigator

- Conversion of the textual database, the DDbDP, to a TEI standard called EpiDoc
- Integration of the three major digital projects in the humanities:
  1) APIS 2) DDbDP and 3) HGV
- Innovative image presentation with ERez server and FSI viewer
P.Oxy. IV 744

Metadata for toronto.apis.17

Title: Letter, 17 June 1 B.C.
Summary: Hilari\non writes from Alexandria to his wife Alla, t\nelling her that he must remain there and offering re\nassurance. The instruction given at lines 8-10 is am\nbiguous and the subject of much discussion.
Language: Greek
Inv. Id: F0827
Phvs.: 1 papyrus; 25 x 14.3 cm.
Desc.: 

APIS Images

Metadata for P.Oxy. IV 744

Title: Letter of Ilarian
Publication: P.Oxy. IV 744
Number: Sel. Pap. I 105
Post: VIII, S. 237; IX, S. 181; XI, S. 145
Concordance BL
Entries
Material: Papyrus
Date: 001 B.C Jun 17
Image Notes: Finegan, Light from the Ancient Past, Plate 135 (nach S. 319); Detersmann, Licht vom Osten, Abb. 25, S. 135; Laudien, Griechische Papyri aus Oxyrhynchus, S. 1; Milligan, Selections from the Greek papyri, Romische, 7; Wirtschaft der griechisch-römischen Welt S. 1

DDBDP Full Text

Date: 186
Location: Oxyrhynchus

Recto

1. ὕπατος Ἀλία τῇ ἀδηλίᾳ πλέον χαί\n2. ρώμῃ καὶ βασιλείᾳ τῆς κυρίας μου καὶ Ἀπολλω\n3. νωτίς, γάρ όσα ἔχεις καὶ γιὰ νὰ πάεις ἀλλάζου\n4. ἵππος σκεπάζει τὴν εἶναι τὸν ἐν τῷ ἔρι\n5. νείριτα τῷ ἐν τῇ ἀρετῇ, τὸν ἐν τῇ ἀρετῇ, τὸν μὲν ἐ\n6. πολέμιον σε ἐν τῇ ἀρετῇ σε ὑπάρχειν
7. ἐν τῷ ἔριτρῳ καὶ ἐν τῷ ἐν τῷ ἐφεύρειν
8. ἐν τῷ ἐν τῇ ἀρετῇ ἀποτελεῖ σοι δικαία, ἄν

9.
Implementing the Navigator

Considerations for Inter-Database Applications
Portlet Architecture

• Pros
  – “Widget-like” approach to varied data sources
  – Preferences architecture out-of-box
  – Some display control out-of-box

• Cons
  – Developmental complexity: Java, Velocity, etc.
    • Coding against the organizational metaphors of the framework
  – Performance: Many servlets per page
eRez / FSI

- Imaging server and Flash plugin
- Live example: [link]

**Pros**
- Great images
- AJAX-like XML configuration
- Scriptable Flash bindings
- Single-source on server: TIFF (no permanent derivatives)

**Cons**
- Uniform user experience requires local images, unlike vanilla HTML
Searching APIS Data

- MARC-based input format
- Unique ID
  - Assigned by contributing institution
  - {institution code}.apis.{digit string}
- Lucene-indexed back-end
  - MARC-like data transformed to XML

Columbia University Libraries
Digital Program
Searching HGV Data

- Published online from FM7
- Principal record identifier is the manually maintained most-recent publication number
- Simple, non-normalized data structure means principal id is not unique
- Limited technical support makes ‘live’ incorporation of search unrealistic
DDBDP at Perseus

- Accessible as XML, text encoded in Beta Code
- Unique document ID’s abstracted from publication numbers
- E.g. Perseus:text:1999.05.0094: document=188
- Unique identifiers and XML output make it a viable candidate for ‘live’ data inclusion
### Composite Searching

**The Navigator Interface**

#### Metadata Search

<table>
<thead>
<tr>
<th>Search by Keyword</th>
<th>Publication Number</th>
<th>Inventory Number (apis only)</th>
<th>APIS Number</th>
<th>Search by Number</th>
<th>Provenance</th>
<th>Language (apis only)</th>
<th>APIS Collection</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Series: P.Col.</td>
<td>e.g. '185a'</td>
<td>Institution:</td>
<td>Publication Number:</td>
<td>philadelphia</td>
<td>Arabic</td>
<td></td>
<td>On or after</td>
</tr>
<tr>
<td></td>
<td>Volume:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Atomic</td>
<td>Select</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Document:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CTRL-click (Mac users command-click) to select multiple languages</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Options

- Show records with:
  - images first
  - publications first
  - translations first

- Show results:
  - in brief
  - in detail

- Display 100 hits per page

- Sort by APIS Control Name
Composite Searching
Step 1: Mapping the Data Models

<table>
<thead>
<tr>
<th>APIS</th>
<th>MARC</th>
<th>HGV</th>
<th>Indexed XML</th>
</tr>
</thead>
<tbody>
<tr>
<td>dd001</td>
<td>001</td>
<td>Publikation-L</td>
<td>controlName</td>
</tr>
<tr>
<td>dd510</td>
<td>510 $a</td>
<td>Publikation-L, Andere Publikation</td>
<td></td>
</tr>
<tr>
<td>dd041</td>
<td>041 $a</td>
<td>language</td>
<td></td>
</tr>
<tr>
<td>dd046</td>
<td>008/06-14, 046</td>
<td>J/M/T [2]</td>
<td>date[1 2]</td>
</tr>
<tr>
<td>dd561 *</td>
<td>561 $a</td>
<td>Ort</td>
<td>provenance</td>
</tr>
<tr>
<td>dd130</td>
<td>130</td>
<td>Origineltitel</td>
<td>title</td>
</tr>
<tr>
<td>dd655_phy</td>
<td>655 $a</td>
<td>Material</td>
<td>Material?</td>
</tr>
<tr>
<td>dd650</td>
<td>650</td>
<td>Inhalt</td>
<td>keyword</td>
</tr>
<tr>
<td>dd090</td>
<td>090 $a</td>
<td>inventory</td>
<td></td>
</tr>
</tbody>
</table>
Composite Searching
Step 2: Normalizing the Data Values

Free-Form APIS to Structured HGV
Free-form text note: “Thmoisepho toparchy of the Oxyrhnchite nome”
Place name: “Oxyrhnchite”
Matched variant: “Oxyrhynchite”
Indexed Canonical form: “oxyrynchite”

HGV Fields to Regularized APIS
Date normalization: Centuries to commonly formatted, index-able dates
Publication numbers: Tracking variant abbreviations of serial publication names
Composite Searching
Step 3: Mapping the Entity Relationships

Building Composite Documents

Diagram:

- APIS
  - (1,1)

- DDbDP
  - (1,1)

- HGV
  - (1,1)
Composite Searching

Step 3: Mapping the Entity Relationships

Finding Related Items
Reconciling Relationships

• Negotiating the relationships through externally defined identifiers
  – Checklists/Authorities
  – Accommodating Variants

• Presenting 0,n relationships
  – Clusters of related documents vs. composite documents
  – The perspective of interest
    • 1 APIS – n HGV [columbia.apis.p74]
    • n APIS – 1 HGV [P.Dura 44]
Lessons Learnt in Inter-Db Searching

• **Data**
  – From people, transformed for processes, displayed to people

• **Unique ID’s**
  – Externally Negotiable: Are the data sources query-able?
  – Ontological considerations: Are the items in the data sources of the same type? Is there a type?

• **Relating data**
  – Structure: Are the data models sufficiently analogous to be searched?
  – Format: Can differences in convention be overcome to make searching sensible?
  – Entity Relationships: Is a search result a composite document, or a cluster of related items?
  – Search Precedence: Are any of the data sources consistently preferred for a particular category of searchable data?
Lessons Learnt in Inter-Db Searching

- Local/Cached & Remote data
  - Is the external data source a good candidate for remote access?
    - Directly query-able for data in usable form
    - Technical support to meet the needs of the project
  - Can a workflow be developed to keep a local copy fresh?

- Search Strategies
  - Simple filters vs. Faceting: Utility, performance consideration in implementation
    - Exploration of relationships
    - Paged view, set-wide facets (SOLR)
    - Client-side faceting (Exhibit)
  - Lucene: Flexibility in indexing, Limitations in data structure
Thoughts on Sustainability

Stephen Paul Davis

Director, Libraries Digital Program
Thoughts on Sustainability 1

Engage Scholars

• Develop new projects in partnership with scholars and researchers in addition to librarians and technology staff

• Be prepared to meet scholars’ specific research needs & preferences, and to evolve them within the context of evolving digital library standards

• Establish meaningful, ongoing advisory and/or governance groups
Thoughts on Sustainability 2

Engage Institutions

• Work to insure that faculty, staff and students at all levels of host institution recognize the value of the project

• Work to establish some level of internal operational funding to continue projects at base level

• Work towards model of institutional & multi-institutional responsibility

• Involve funding agencies in ongoing discussions of strategies & priorities
Thoughts on Sustainability 3
Think Beyond Collection Building

• Work with scholars and researchers to develop alternative scenarios for the future

• Consider partnerships and ‘interoperation’ with complementary digital projects worldwide

• Develop tools and technologies adaptable to a highly distributed digital research environment

• Open up content and tools to other projects and types of users (e.g., with published APIs, web services, XML access)
Thoughts on Sustainability 4

Develop Diverse Funding Strategies

• Build core institutional program funding

• Pursue flexible grant strategies

• Develop multi-institutional governance and funding

• Explore endowment building for multiple interrelated projects rather than single projects

• Do not expect end users to pay for your content
Thoughts on Sustainability 5

Build Change into Planning

- Assume key players will someday move on
- Assume projects may move from one institution to another
- Assume that all technology components will need life-cycle management and funding
- Recognize that an indispensable research tool cannot easily be discontinued
- Have a realistic exit strategy
Thank you

Questions?

www.papyri.info