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

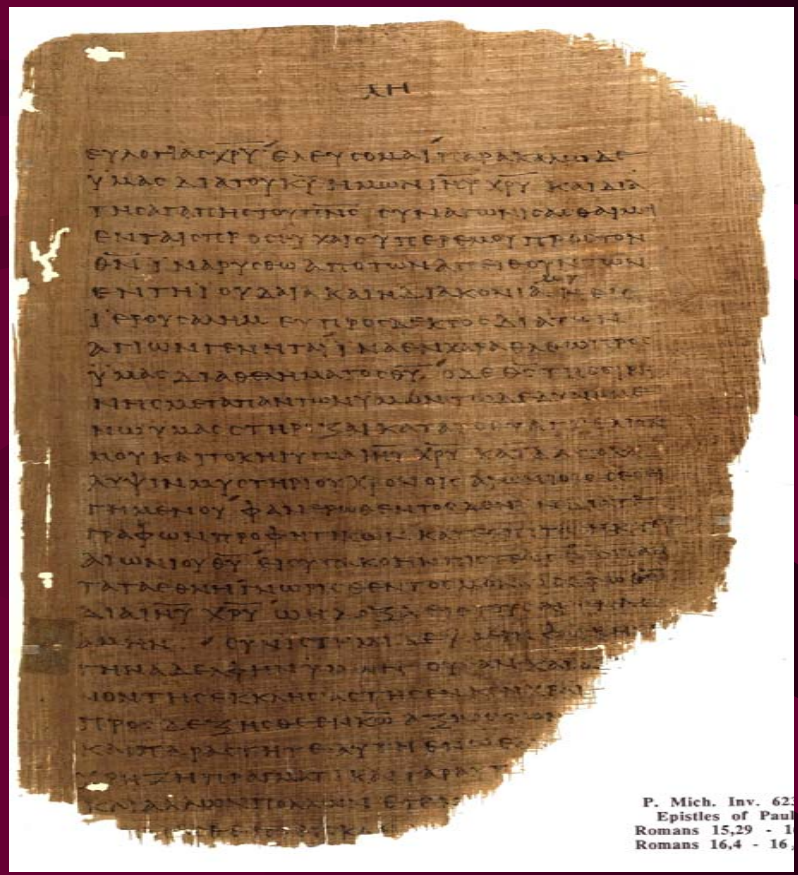
Writing in the Ancient World

Ancient Texts

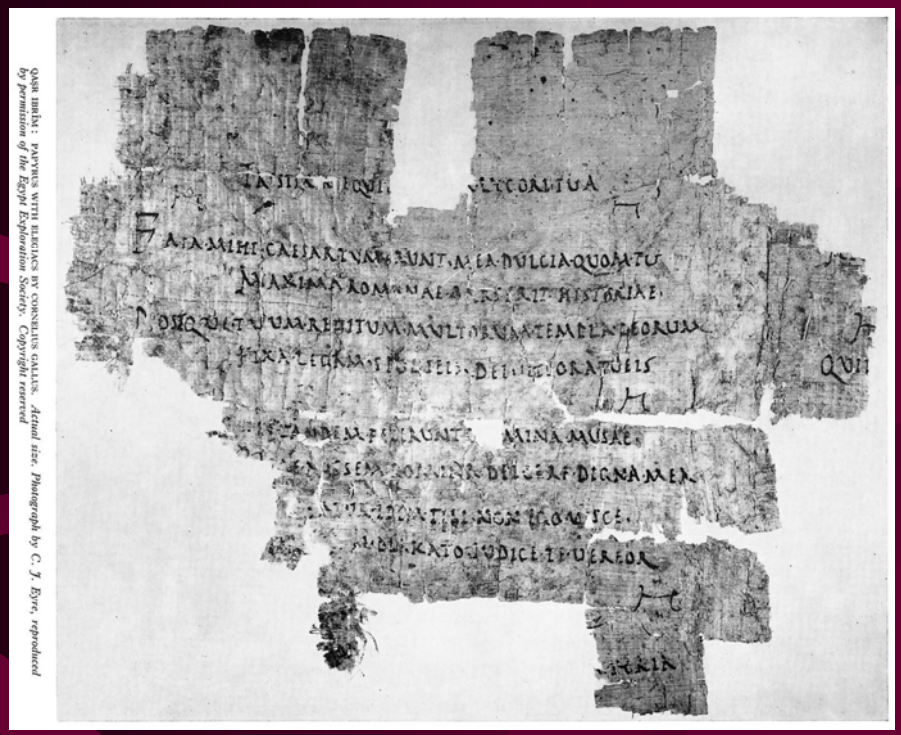
- Pharaonic Times – Middle Ages
- Many Different Languages
  - Egyptian (Hieroglyphic, Hieratic, Demotic, Coptic)
  - Greek
  - Latin
  - Arabic
- Many Different Types of Texts
  - Literature (Homer, Tragedy, Philosophy, History, Lyric Poetry)
  - Religious Texts
  - Imperial Decrees
  - Mundane Documents

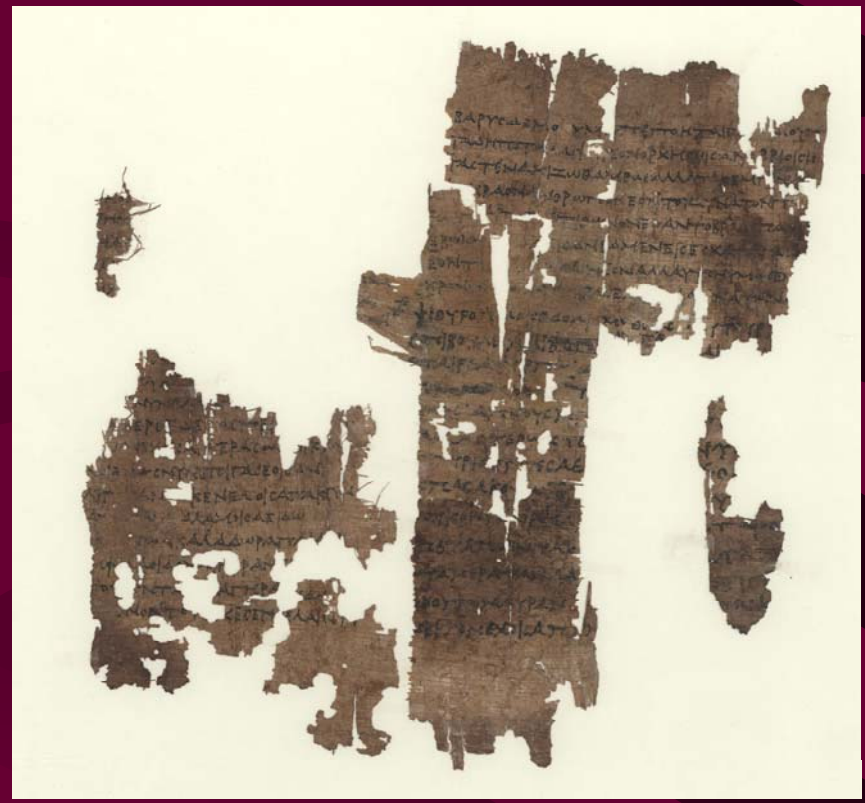
Gospel of Judas

Epistles of Paul



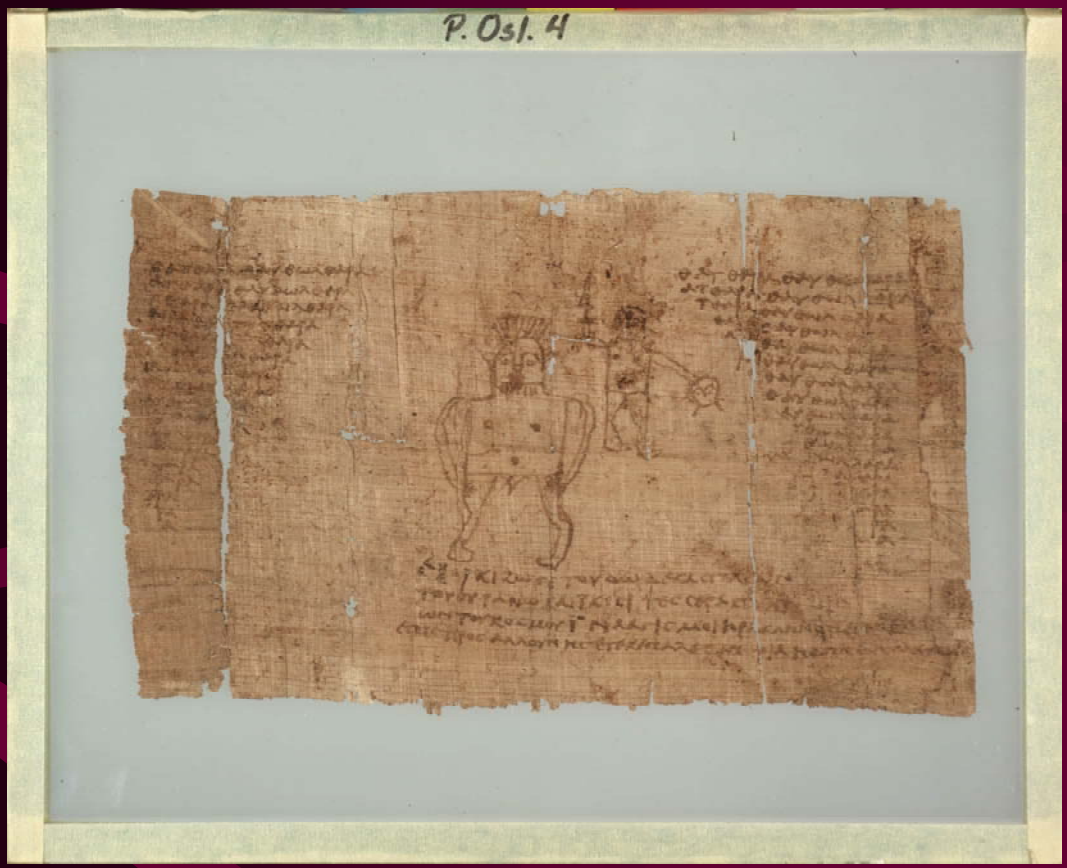
Elegiac Poem by Cornelius Gallus (70-26 BC)

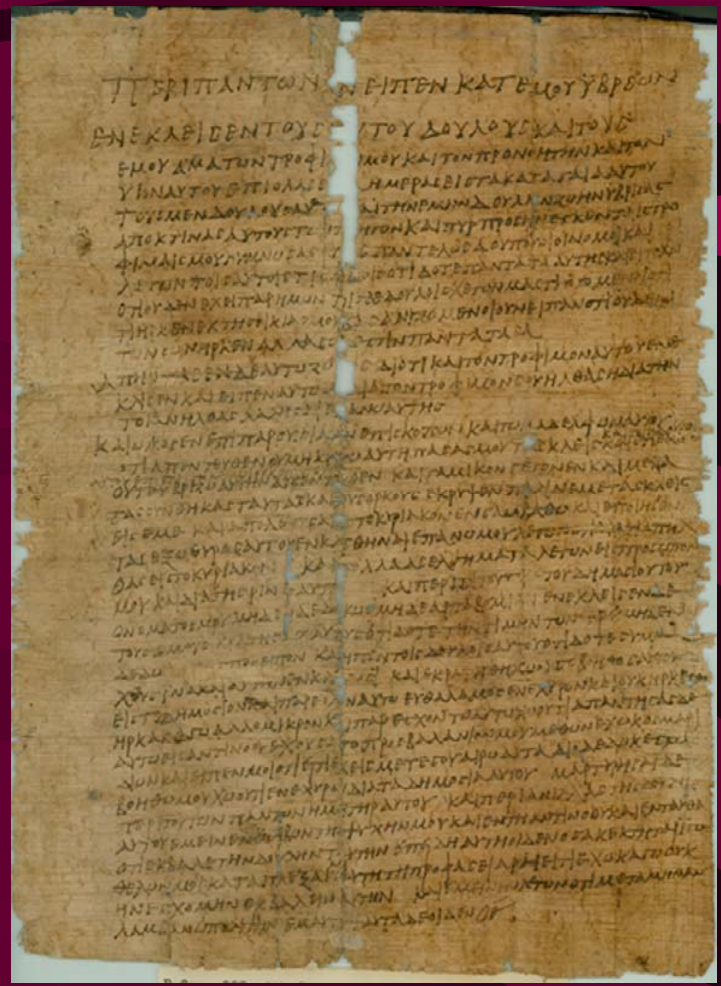




New Poetry of Sappho ($7^{th}/6^{th}$ c. BC)

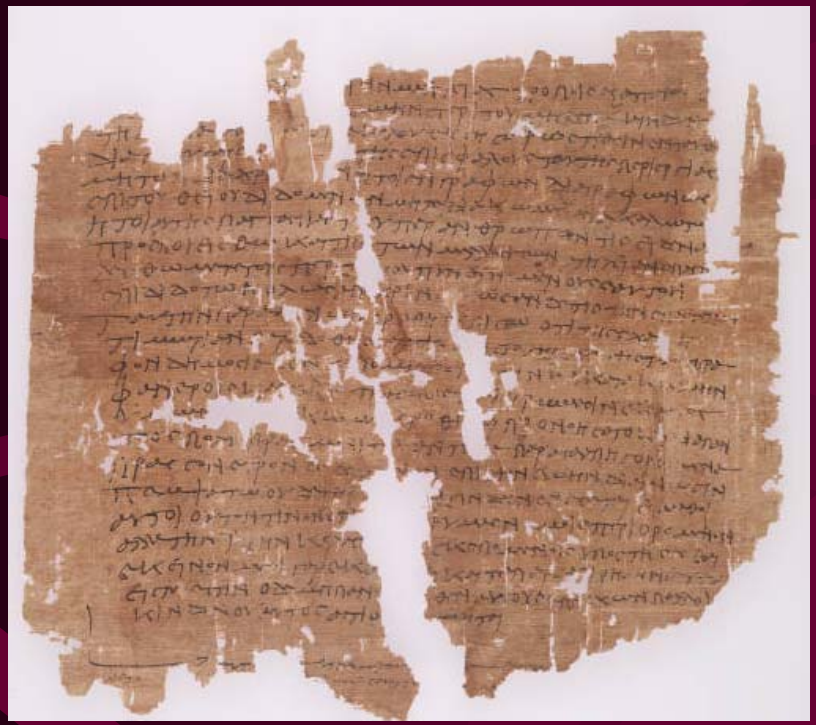
Love Charm to Win Heart of Another (4th c. AD)





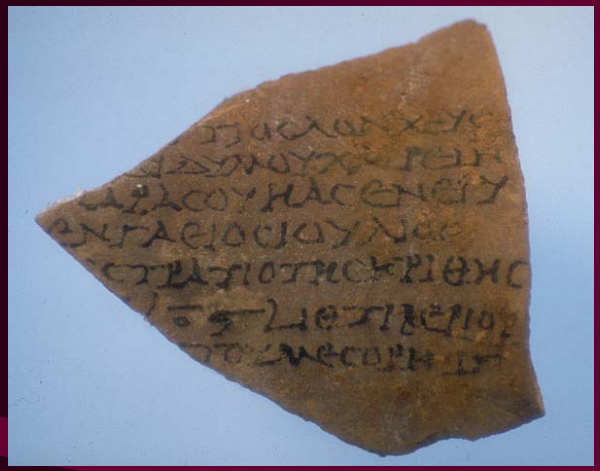
Affidavit in a Case of Domestic Abuse (4th c. AD)

Order of the Roman Prefect Banning Oracles and Magic (198-199 AD)

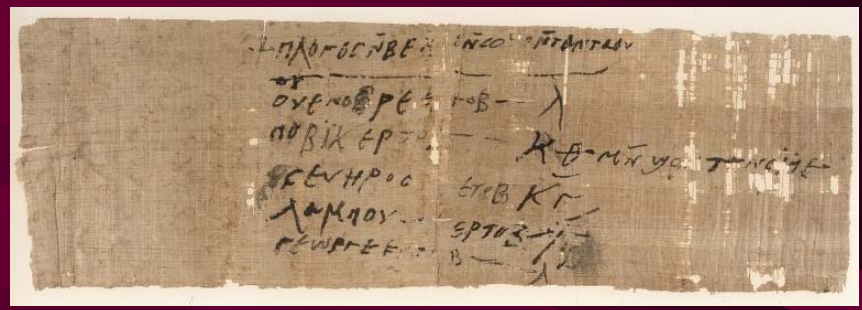


"Encountering many who believed themselves to be deceived by the practices of divination I quickly considered it necessary . . . to enjoin all people to abstain from this hazardous inquisitiveness. Therefore, let no man through oracles, that is, by means of written documents supposedly granted under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

Barley Receipt (Aug. 6, 33 AD)

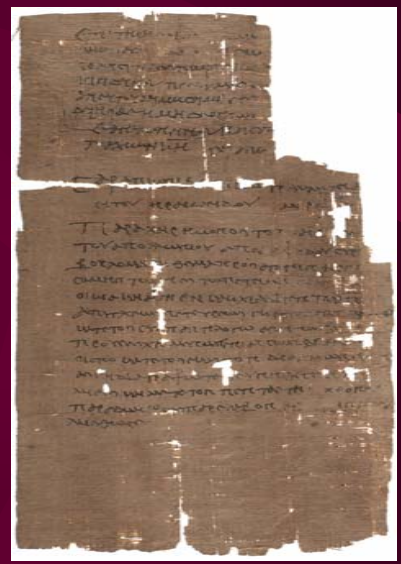


Account of Salaries in Coptic
(4th/5th c.)



Birth Certificate in Latin (April 29, 145 AD)

Lease of Confiscated
Property (May 23, 142 AD)



Letter of a Woman to Her Mother (early 2nd c. AD)

Θερμουθᾶς Οὐαλεριᾶτι τῇ μητρὶ πλεῖστα χαίρειν καὶ διὰ παντὸς ὑγιαίνειν. ἔλαβον παρὰ Οὐαλερίου σπυρίδιν ὅπου λαγάνια ζεύγη εἴκοσι καὶ κολλούρια
5 δέκα ζεύγη· πέμψον μοι τὰ λωδίκια τῆς τιμῆς καὶ ἐρίδια καλὰ πόκους τέσσαρας. δὸς σῦ αὐτὰ Οὐαλερίῳ. καὶ ἑπτάμηνον ἦμεν ὄρτι. καὶ ἀσπάζομαι Ἄρτεμιν καὶ
10 Νικαροῦν τὴν μικρὰν καὶ Οὐαλέριν τὸν κύριον μου ἐπιθυμῶ αὐτὸν ἐν τῷ νῷ καὶ Διονυσίαν καὶ Δημητροῦν πολλὰ καὶ Ταῆσιν τὴν μικρὰν πολλὰ καὶ τοὺς ἐν οἴκῳ πάντας.
15 καὶ πῶς ἔχει μου ὁ πατήρ, πέμψεις μοι τὴν φάσιν, ὅτι ἀσθενῶν ἀπῆλθε ἀπ' ἐμοῦ. ἀσπάζομαι Μάμαν. ἀσπάζεται ὑμᾶς Ῥοδίνη. βέβληκα αὐτὴν εἰς τὴν τέχν[η]-
20 ν· σῦ χρῄζω αὐτῆς, ἀλλὰ ἤδέως ἔχω. Φαῶφι 8.

ν

ἀπόδος εἰς Φιλαδέλ(φιαν) Οὐαλεριᾶτι τῇ μητρί.

Thermouthas to Valerias, 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 Valeris 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. Phaophi 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

APIS Partners & Contributors 2

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

APIS Papyrus Entry: berenike.apis.102 - Windows Internet Explorer
http://wwwapp.cc.columbia.edu/ldpd/app/apis/item?mode=item&key=berenike.apis.102
File Edit View Favorites Tools Help
Google Go Bookmarks 171 blocked Check AutoLink AutoFill Send to
APIS Papyrus Entry: berenike.apis.102

APIS Advanced Papyrological Information System

Item	APIS record: berenike.apis.102
Title	Receipt, August 6, 33 A.D.
Summary	Receipt for the delivery of a load of barley.
Full-Text (DDBDP)	link to Perseus Project, O.Berenike:1:102
Inventory Id	Inv. BE97-13-A-150
Publications	O.Berenike I 102
Original Language	Greek
Physical Description	1 ostracon
Notes	7 lines of text. Roman dump at Berenike.
Notes on Custodial History	Currently stored in a magazine of the Supreme Council of Antiquities in Quft.
Translation	"--tios, lancer, to --, son of Didymous, greetings. I have received from you 6 artabas of barley, total 6 art., which Gaius Iulius --, soldier, became surety for. Year 19 of Tiberius Caesar Augustus, Mesore 13."

View Full Metadata

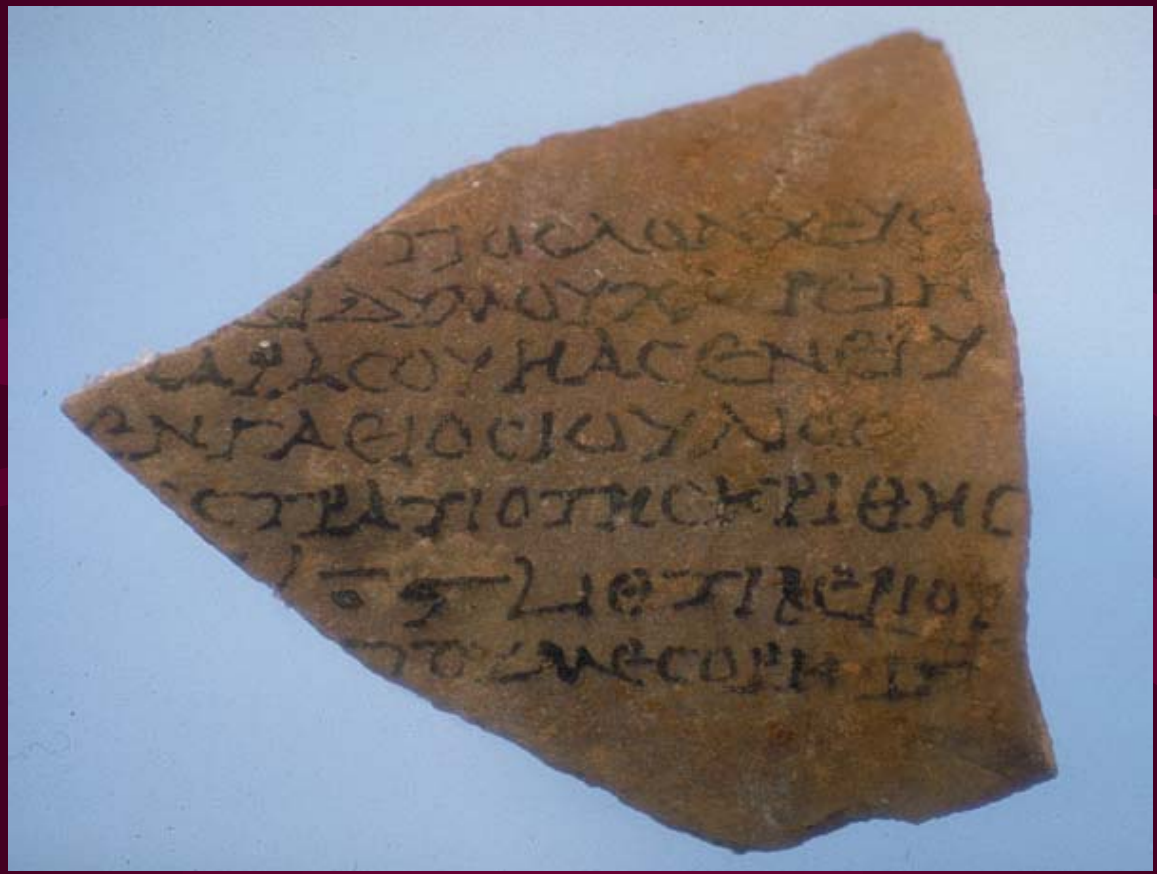
Images

- 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.

start DLF_Nov07 APIS Papyrus Entry: ...

Barley Receipt (Aug. 6, 33 AD)



Duke Data Bank of Documentary Papyri:
Papyri collection contents
About the Papyri collection

Plot:
- dates in this document

Contents:
- Volume 1

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.

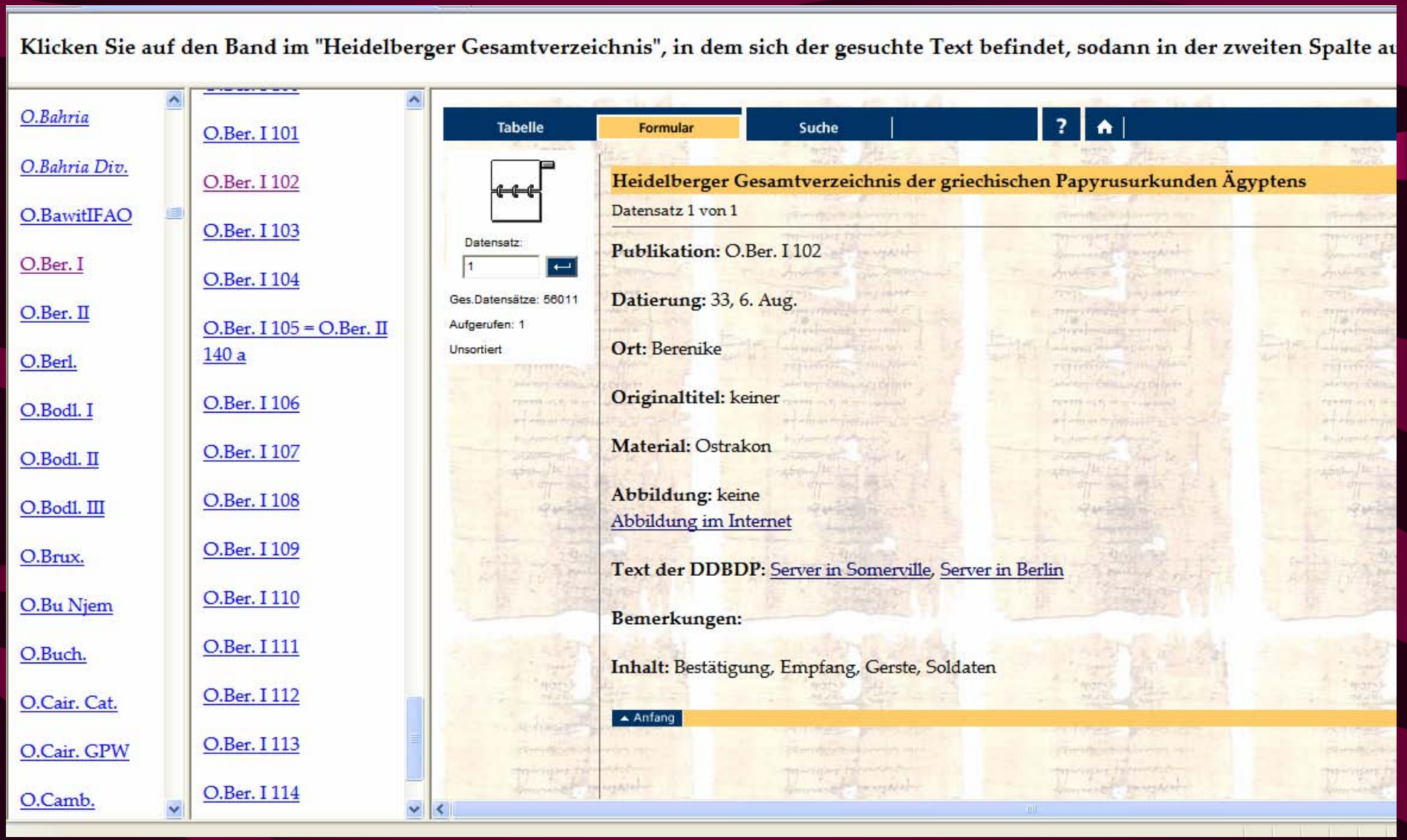
Table of Contents Go to 102

1.102.

Date: AD33 ; Location: Berenike

[.c 11].τιος **λογχεὺς**[1]
[.c 7] Διδύμου χαίρειν:
[ἔχω] παρὰ σοῦ {κ} ἃς ἐνεγύ-
[ησ]εν Γάειος Ἰούλιος
5 [.c 6]ς **στρατιώτης**[2] κριθῆς
[ἀρτάβας] 6, (γίνονται) (ἀρτάβαι) 6. (ἔτους) 19 Τιβερίου
[Καίσαρος Σεβα]στοῦ Μεσορὴ 13.

[1] λονχευς Pap. [2] στρατιοτης Pap.


Klicken Sie auf den Band im "Heidelberger Gesamtverzeichnis", in dem sich der gesuchte Text befindet, sodann in der zweiten Spalte au
O.Bahria
O.Bahria Div.
O.BawitIFAO
O.Ber. I
O.Ber. II
O.Berl.
O.Bodl. I
O.Bodl. II
O.Bodl. III
O.Brux.
O.Bu Njem
O.Buch.
O.Cair. Cat.
O.Cair. GPW
O.Camb.
O.Ber. I 101
O.Ber. I 102
O.Ber. I 103
O.Ber. I 104
O.Ber. I 105 = O.Ber. II 140 a
O.Ber. I 106
O.Ber. I 107
O.Ber. I 108
O.Ber. I 109
O.Ber. I 110
O.Ber. I 111
O.Ber. I 112
O.Ber. I 113
O.Ber. I 114
Tabelle
Formular
Suche
Datensatz:
1
Ges.Datensätze: 56011
Aufgerufen: 1
Unsortiert
Heidelberger Gesamtverzeichnis der griechischen Papyrusurkunden Ägyptens
Datensatz 1 von 1
Publikation: O.Ber. I 102
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
Anfang

Stephen Paul Davis
Director, Libraries Digital Program

Putting It Together

APIS + 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

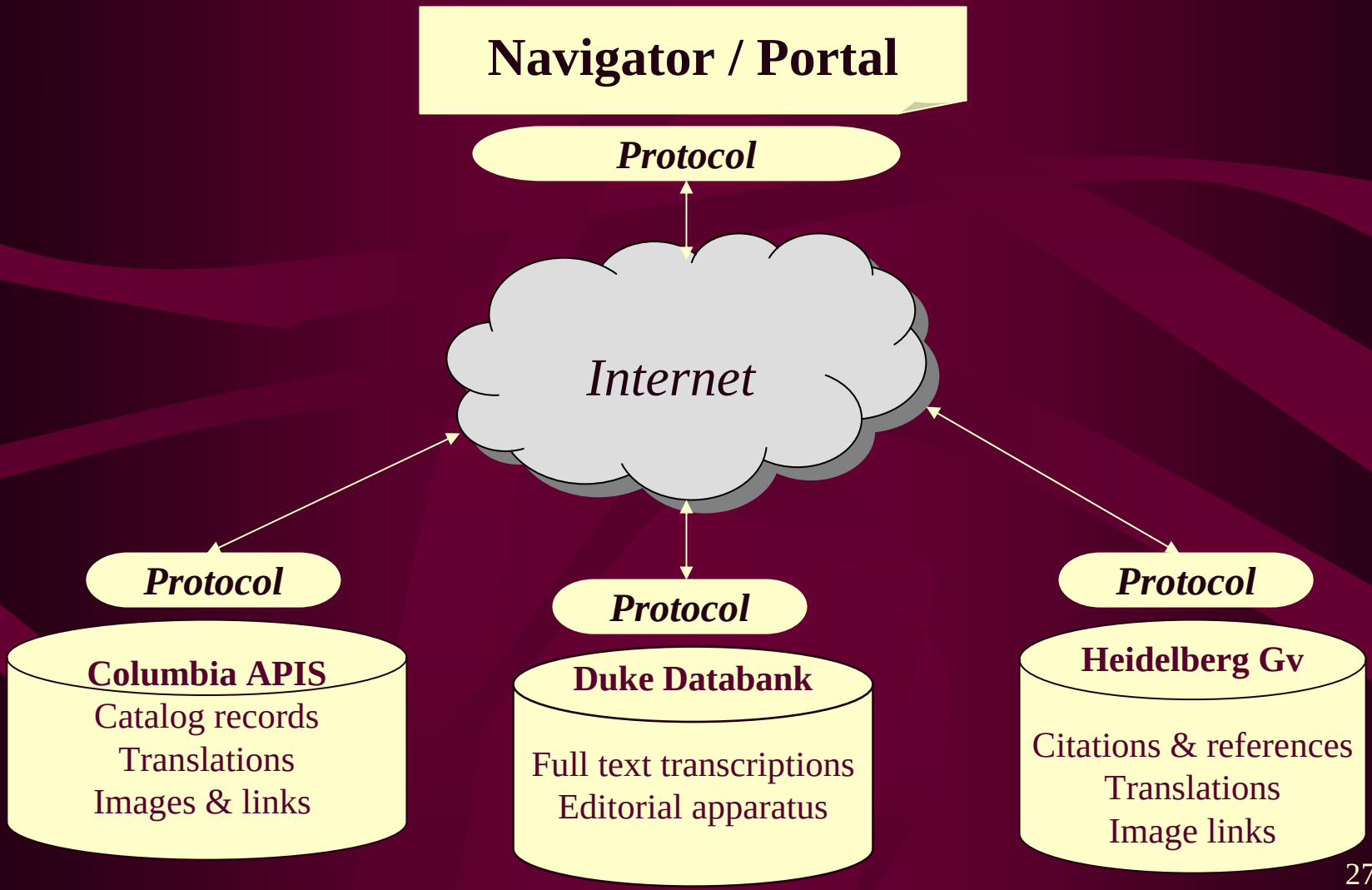
A. Real-time search, retrieval & display from remote data sources (Z39.50-like)

B. OAI harvesting to create locally hosted union catalogue

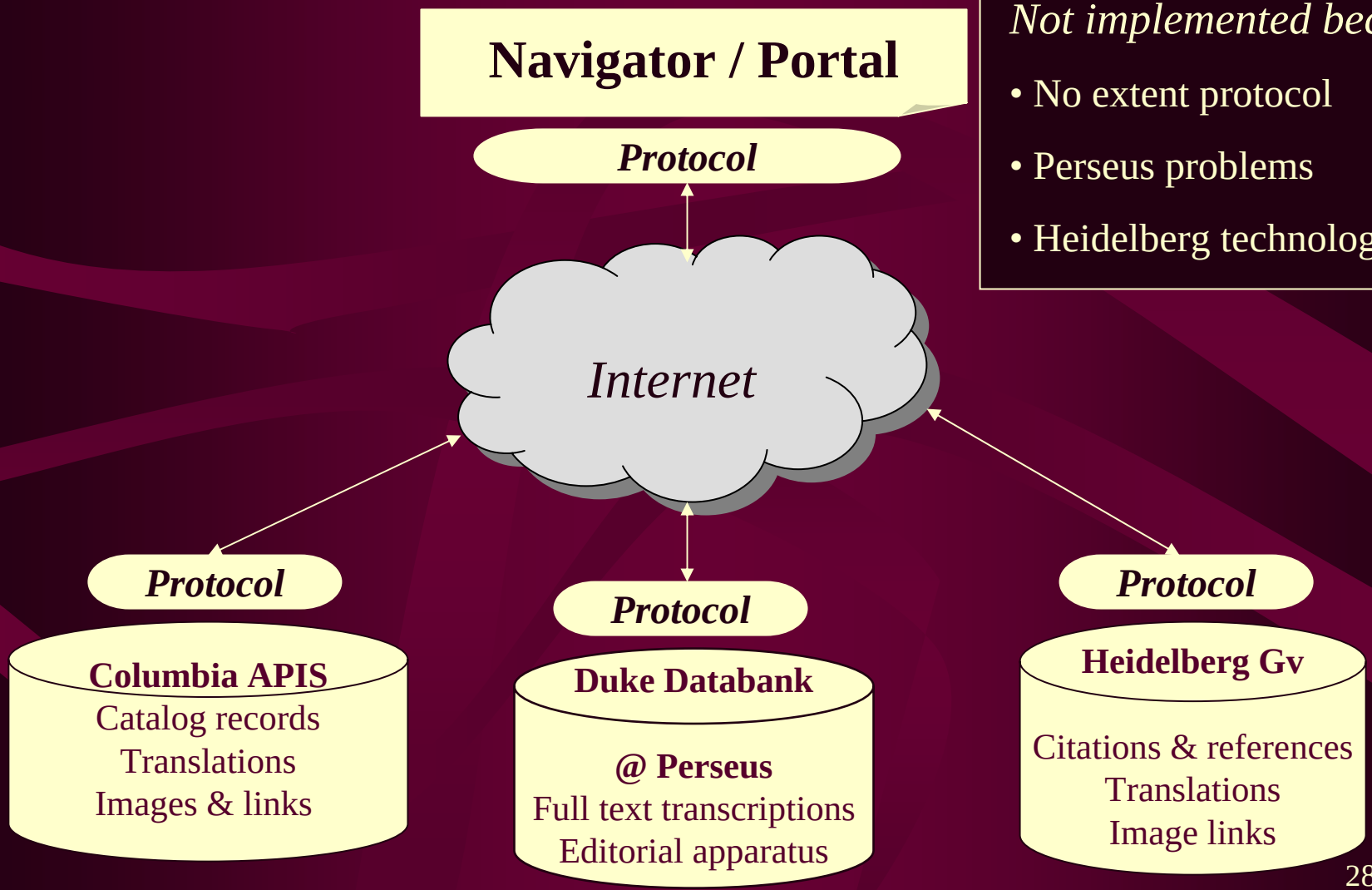
or

C. Locally-maintained copies of source data

papyri.info Architecture A
Real-time meta-search



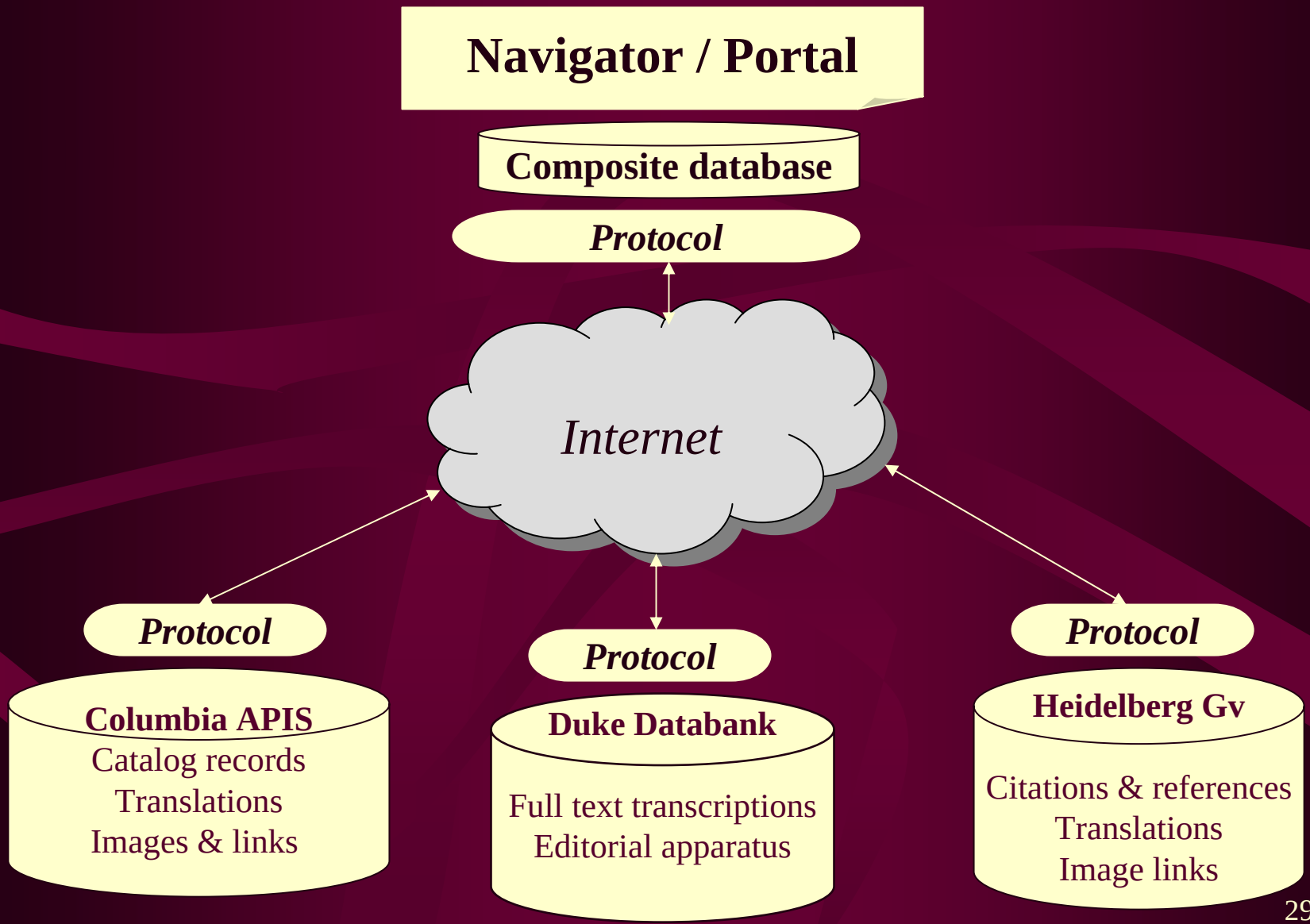
papyri.info Architecture A
Real-time meta-search



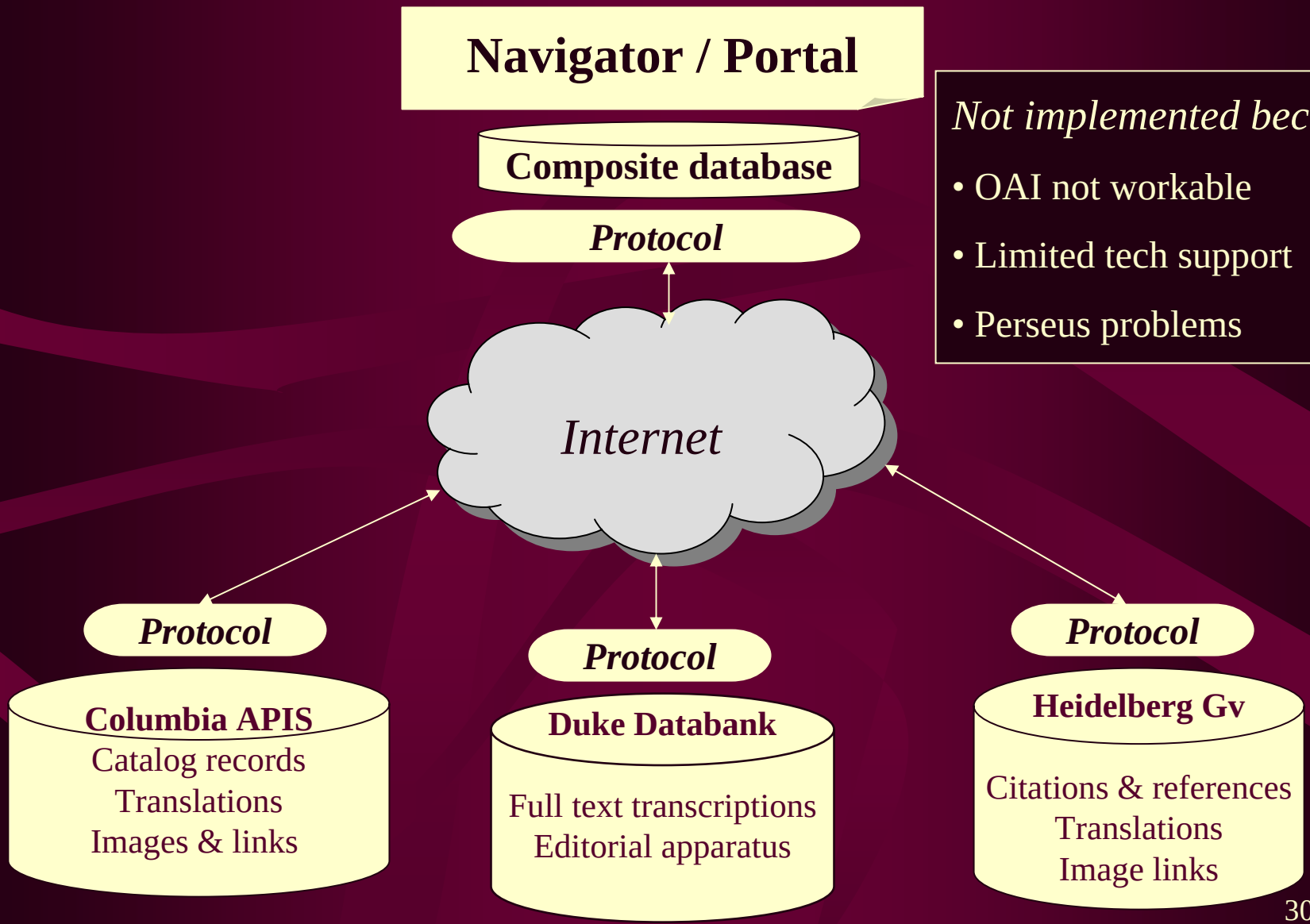
Not implemented because:

- No extent protocol
- Perseus problems
- Heidelberg technology

papyri.info Architecture B
Offline harvesting



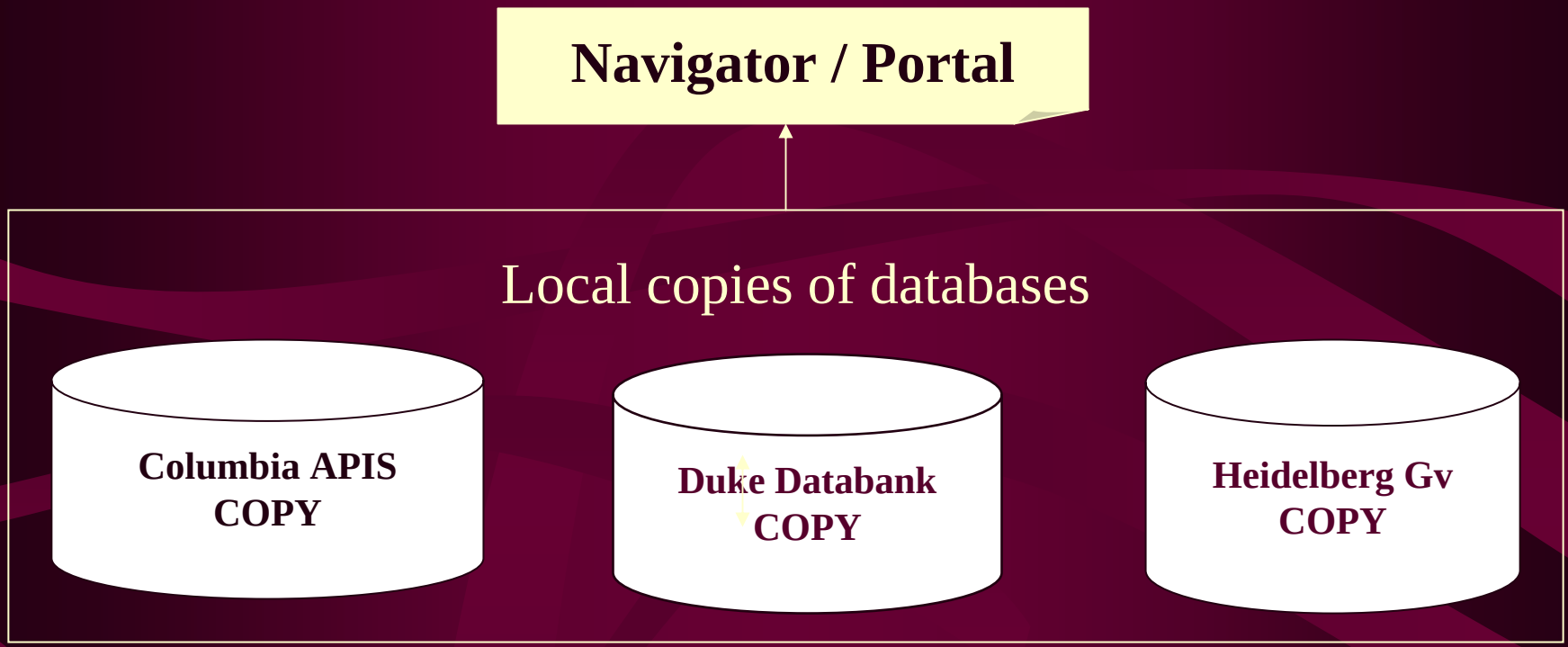
papyri.info Architecture B
Offline harvesting



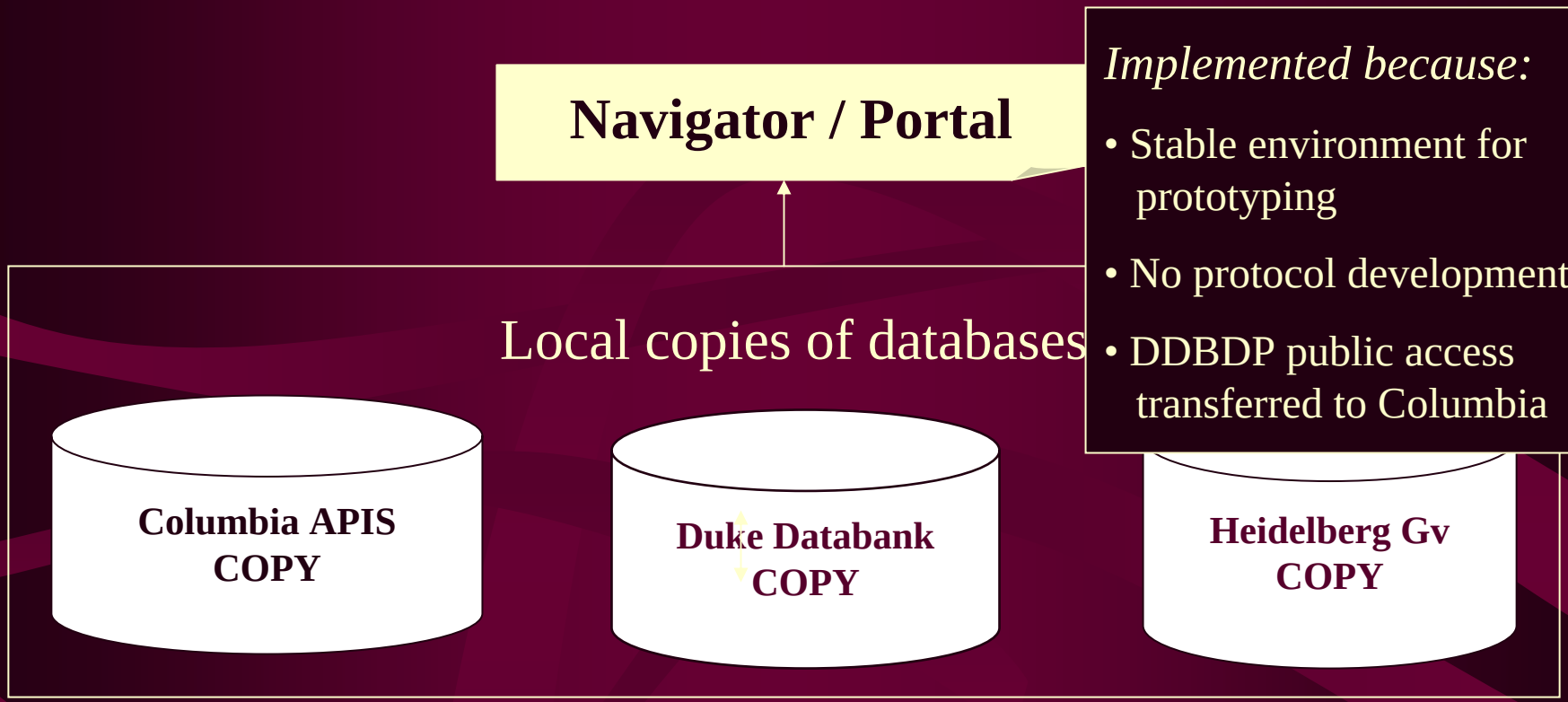
Not implemented because:

- OAI not workable
- Limited tech support
- Perseus problems

papyri.info Architecture C
Local Copies of Source DBs



papyri.info Architecture C
Local Copies of Source DBs



Implemented because:

- Stable environment for prototyping
- No protocol development
- DDBDP public access transferred to Columbia

papyri.info Overview 1

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

- http://apptest.cul.columbia.edu:8082/navigator/portal/apisfull.psml?controlName=toronto.apis.17 *[link]*

File Edit View Favorites Tools Help
Google Go Bookmarks 171 blocked Check AutoLink AutoFill Send to Settin
Metadata Overview Metadata Overview Metadata Overview Page Tools

APIS Metadata toronto.apis.17

Metadata for toronto.apis.17

Title Letter, 17 June 1 B.C.

Summary Hilarion writes from Alexandria to his wife Alis, telling her that he must remain there and offering reassurance. The instruction given at lines 8-10 is ambiguous and the subject of much discussion.

Language Greek

Inv. Id F0827

Phys. Desc. 1 papyrus; 25 x 14.3 cm.

APIS Images

HGV Metadata P.Oxy. IV 744

Metadata for P.Oxy. IV 744

Title Letter of Ilarion

Publication Number P.Oxy. IV 744
Sel. Pap. I 105

Post-Concordance BL Entries VIII, S. 237; IX, S. 181; XI, S. 145

Material Papyrus

Date 001 BCE Jun 17

Image Notes Finegan, Light from the Ancient Past, Plate 135 (nach S. 318); Deissmann, Licht vom Osten, Abb. 25, S. 135; Laudien, Griechische Papyri aus Oxyrhynchos, S. 1; Milligan, Selections from the Greek Papyri, Frontispiece; Kloft, Wirtschaft der griechisch-römischen Welt, S. 74

DDBDPText 0181;4;744

DDBDP Full Text

Date: 1BC
Location: Oxyrhynchus

Recto

1 Ἱλαρίων Ἀλιτι τῆι ἀδελφῆ πλεῖστα χαί-
2 ρειν καὶ Βερούτι τῆ κυρίᾳ μου καὶ Ἀπολλω-
3 ναρίῳ . γίνωσκε ὡς ἔτι καὶ νῦν ἐν Ἀλεξαν-
4 δρε<ί>α <ἐ>σμεν : μὴ ἀγωνιᾶς ἐὰν ὅλως εἰσ-
5 πορεύωνται , ἐγὼ ἐν Ἀλεξανδρε<ί>α μενῶ .
6 ἐρωτῶ σε καὶ παρακαλῶ σε ἐπιμελή-
7 θ<ητ>ι τῷ παιδίῳ καὶ ἐὰν εὐθὺς ὀψώνι-
8 ον λάβωμεν ἀποστελῶ σοι ἄνω. ἐὰν

Done Internet 100%
start DLF_Nov07 PowerPoint Slide Sho... Metadata Overview -... Inbox - Outlook Expr... EN 21:44

Implementing the Navigator

Considerations for Inter-Database Applications

Portlet Architecture

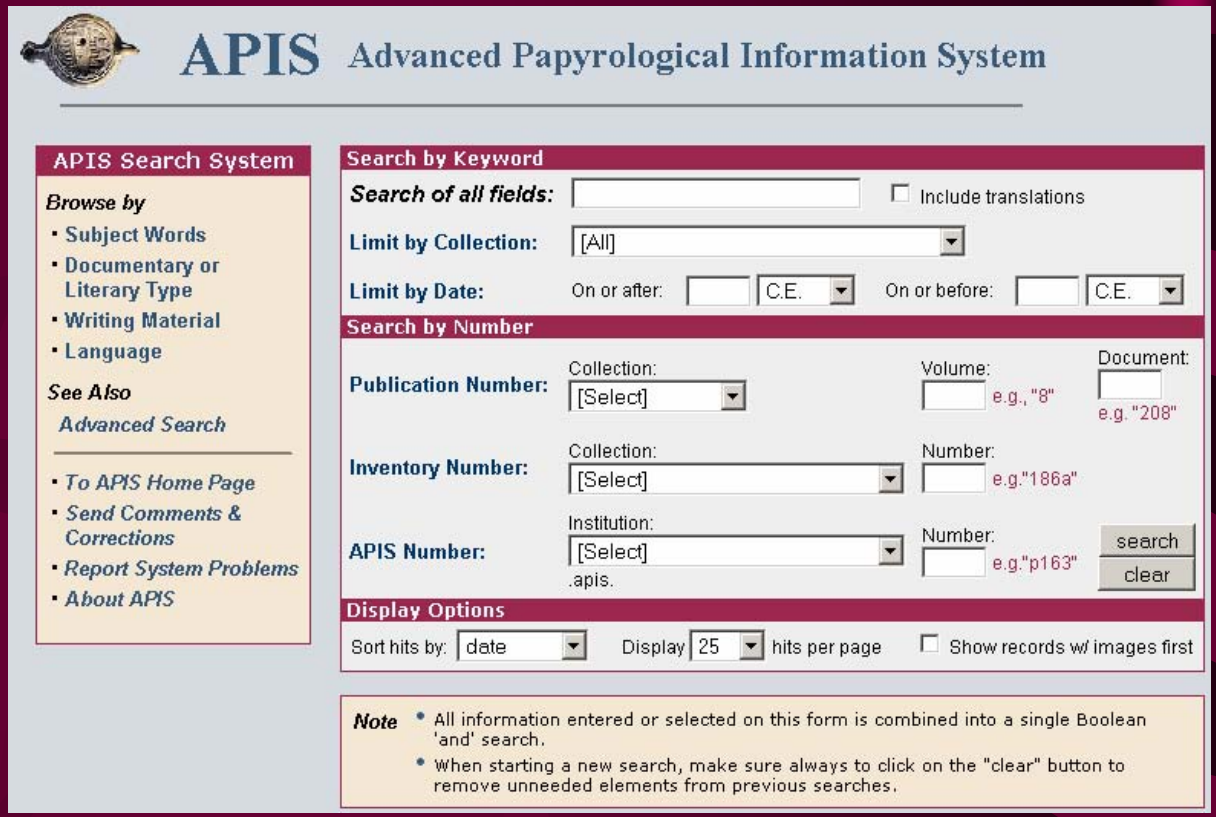
- Pros
  - “Widget-like” approach to varied data sources
  - Preferences architecture out-of-box
  - Some display control out-of-box
- Cons
  - Developmental complexity: Java, Velocity, etc.
    - Coding against the organizational metaphors of the framework
  - Performance: Many servlets per page

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

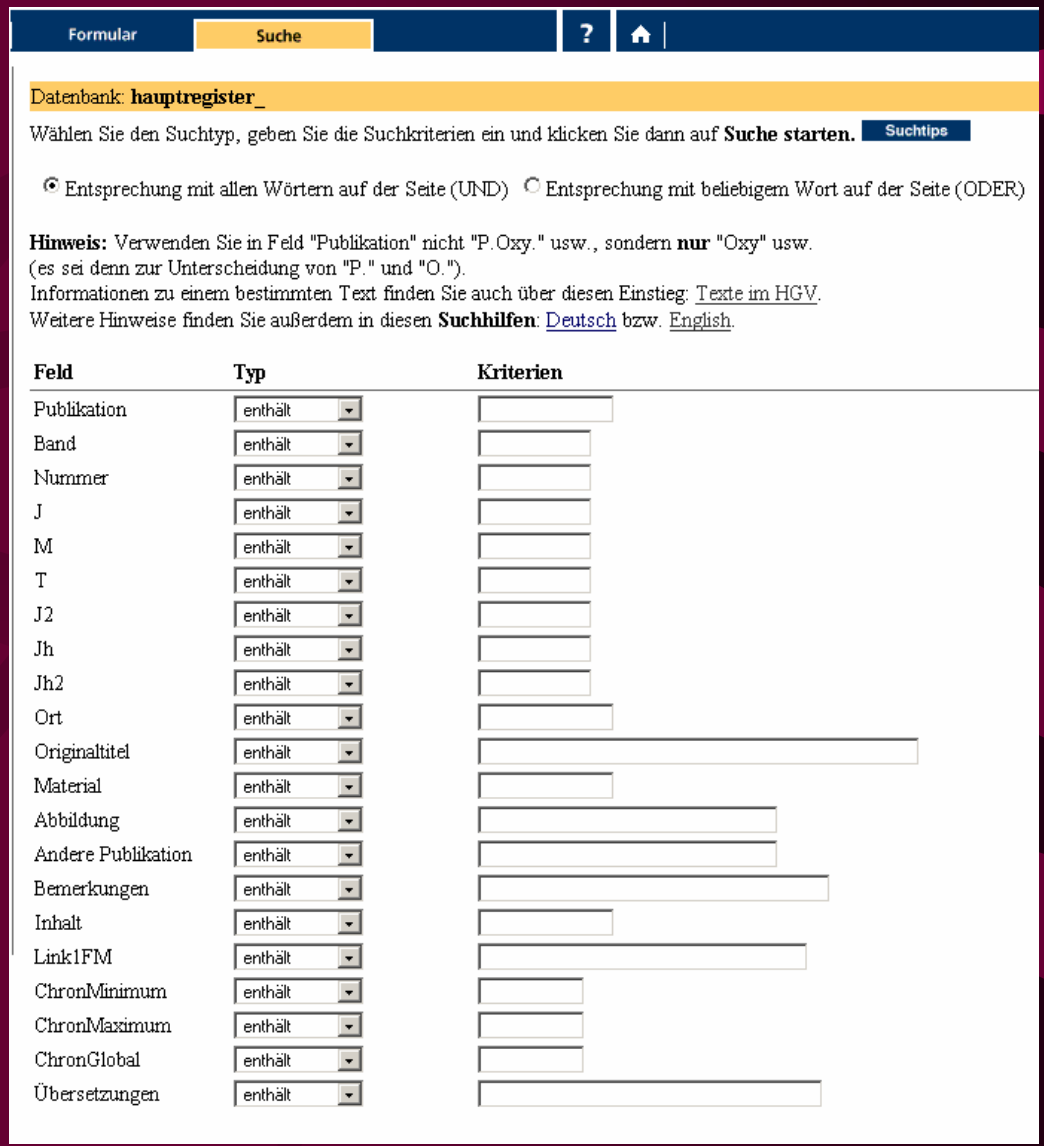
Searching APIS Data

- 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



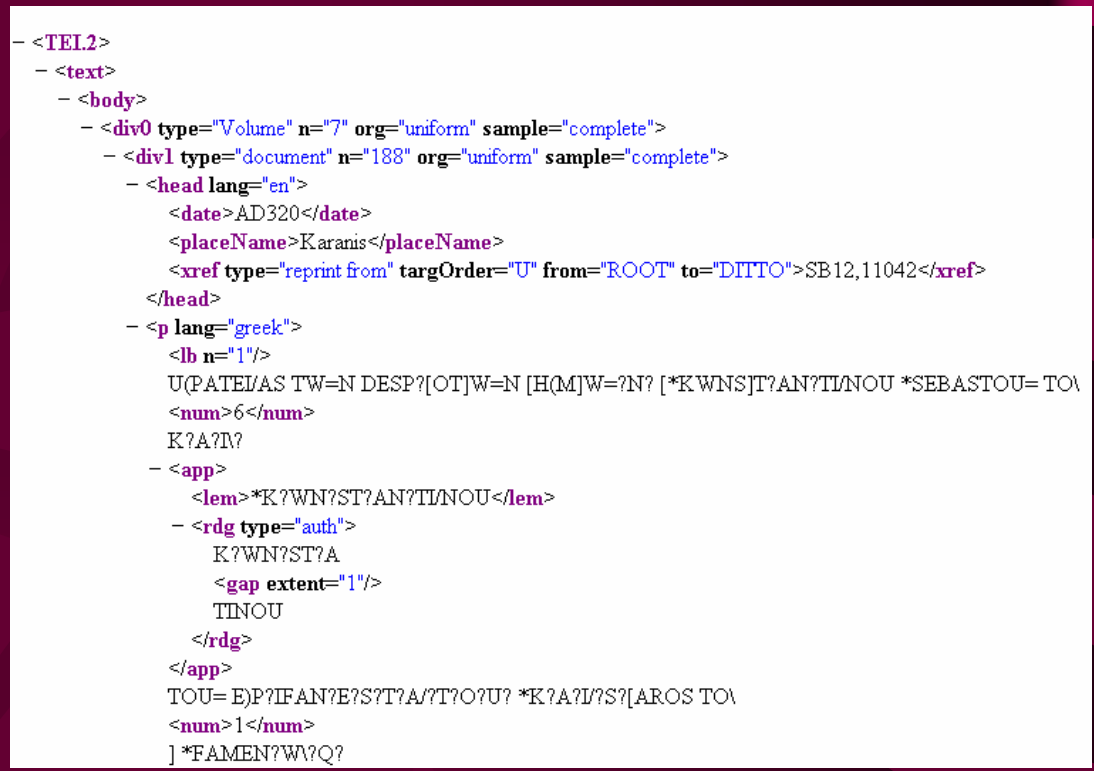
Searching HGV Data

- 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



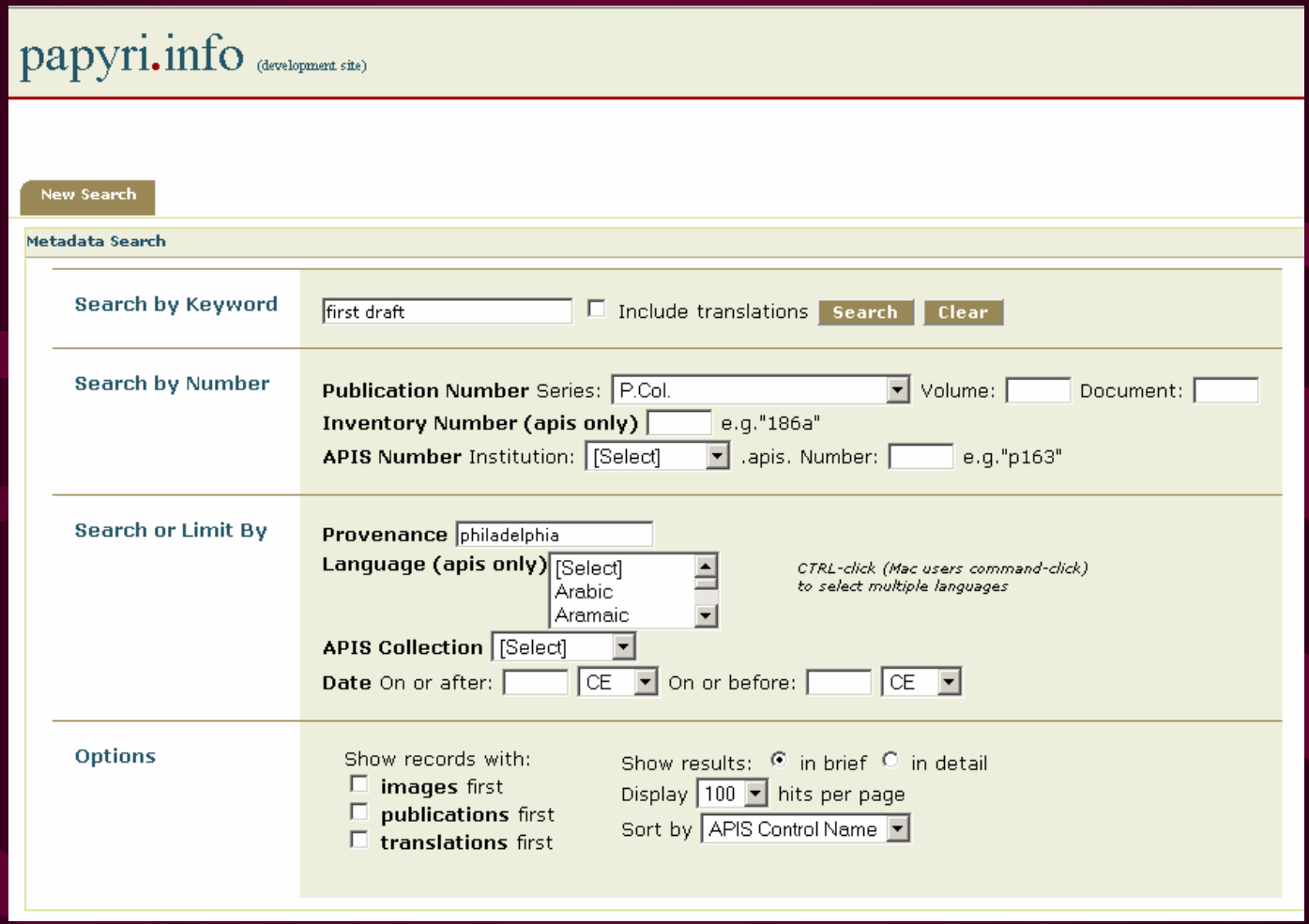
DDbDP at Perseus

- 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



A search for key words "first," "draft" in Philadelphia [link]

Composite Searching

Step 1: Mapping the Data Models

APIS	MARC	HGV	Indexed XML
dd001	001	Publikation-L	controlName
dd510	510 $a	Publikation-L, Andere Publikation	
dd041	041 $a		language
dd046	008/06-14, 046	J/M/T [2]	date[1 2]
dd561 *	561 $a	Ort	provenance
dd130	130	Origineltitel	title
dd655_phy	655 $a	Material	Material?
dd650	650	Inhalt	keyword
dd090	090 $a		inventory

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

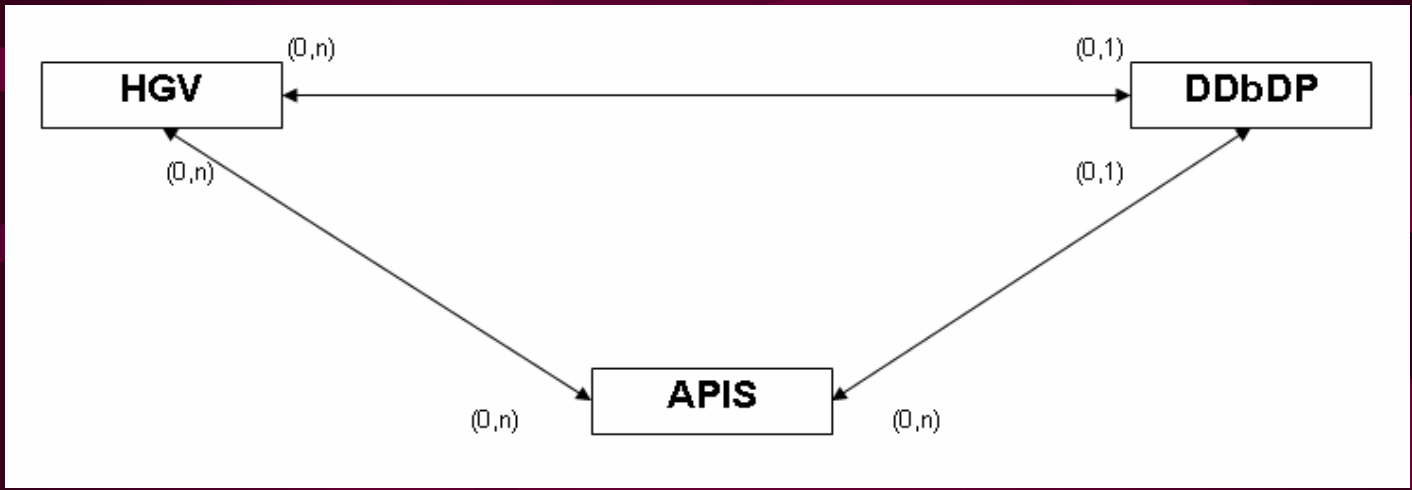
Step 3: Mapping the Entity Relationships

Building Composite Documents

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