University of Kentucky
UKnowledge

Library Presentations | University of Kentucky Libraries

4-25-2024

Preserving the Uncrawlable: Serving the Server

Andrew McDonnell
University of Kentucky, mcdonnell@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/libraries_present

Part of the Archival Science Commons, Byzantine and Modern Greek Commons, and the Digital Humanities Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation

McDonnell, Andrew, "Preserving the Uncrawlable: Serving the Server" (2024). *Library Presentations*. 273.
https://uknowledge.uky.edu/libraries_present/273

This Presentation is brought to you for free and open access by the University of Kentucky Libraries at UKnowledge. It has been accepted for inclusion in Library Presentations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu, rs_kbnotifs-acl@uky.edu.

Preserving the Uncrawlable: Serving the Server

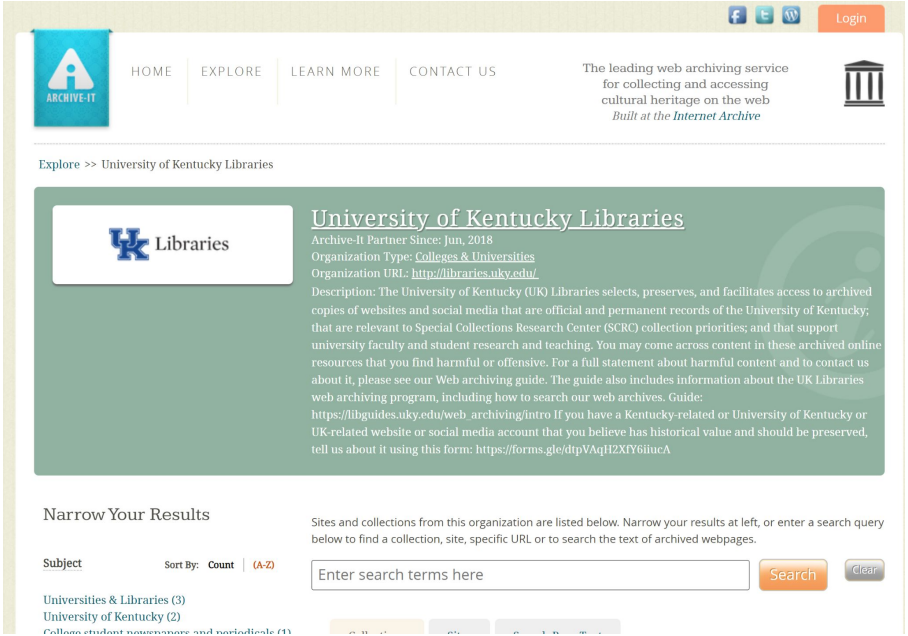
tiny.cc/ukweb2024

•••

Andrew McDonnell, Digital Archivist
University of Kentucky Libraries
Special Collections Research Center
April 25, 2024

University of Kentucky Libraries Web Archiving

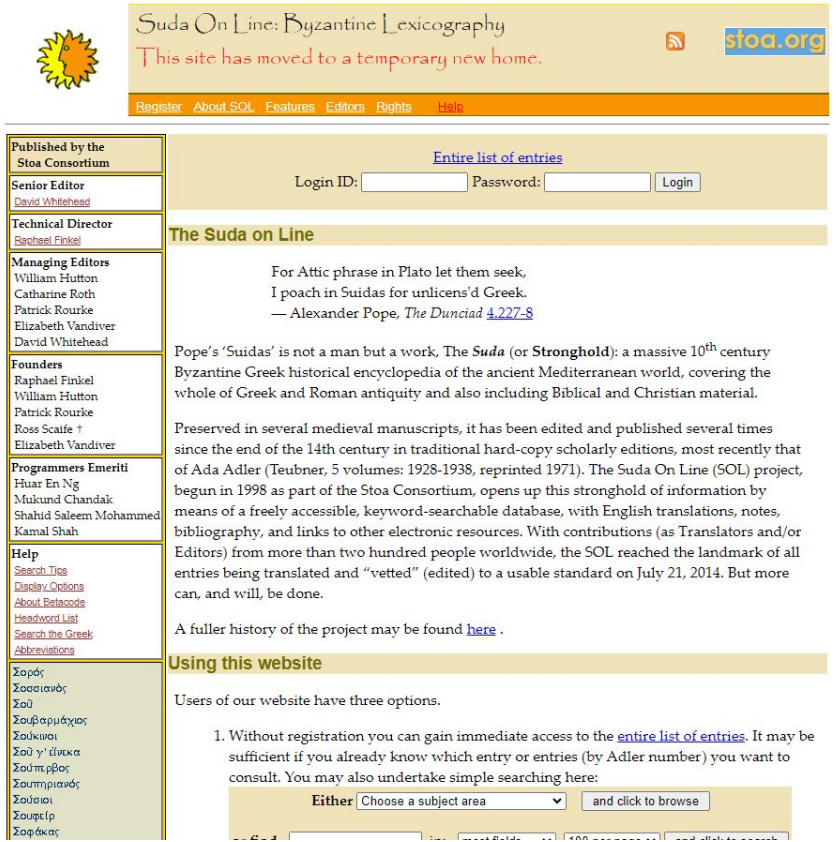
Established in 2018
https://archive-it.org/organizations/915



Suda On Line: Byzantine Lexicography

Founded in 1998

Still live:
www.cs.uky.edu/~raphael/sol/sol-html

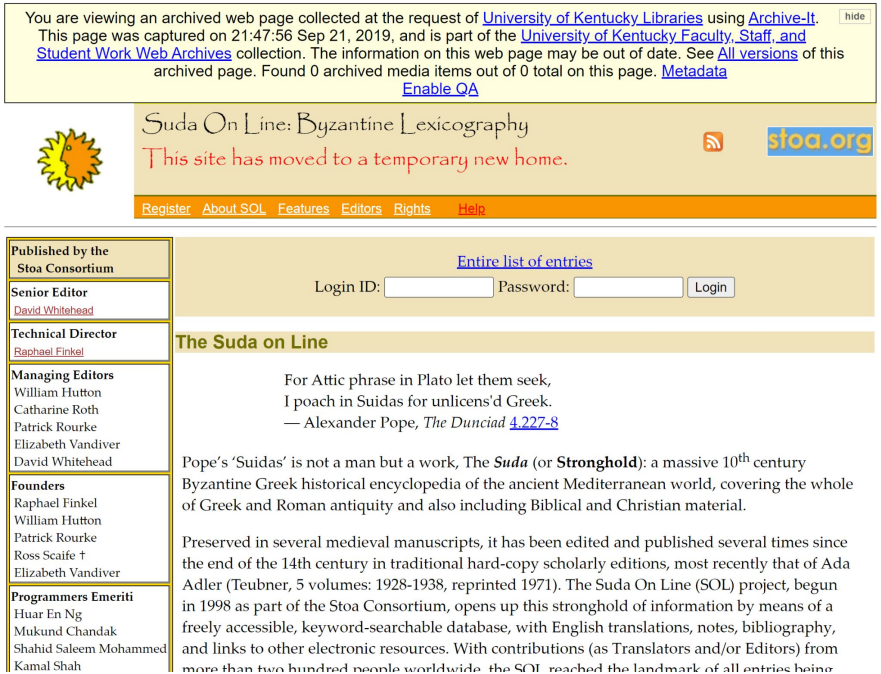


Suda On Line: Byzantine Lexicography

Founded in 1998

https://wayback.archive-it.org/11384/*/https://www.cs.uky.edu/~raphael/sol/sol-html/

Looks Great!



House of lies.

Suda On Line: Byzantine Lexicography

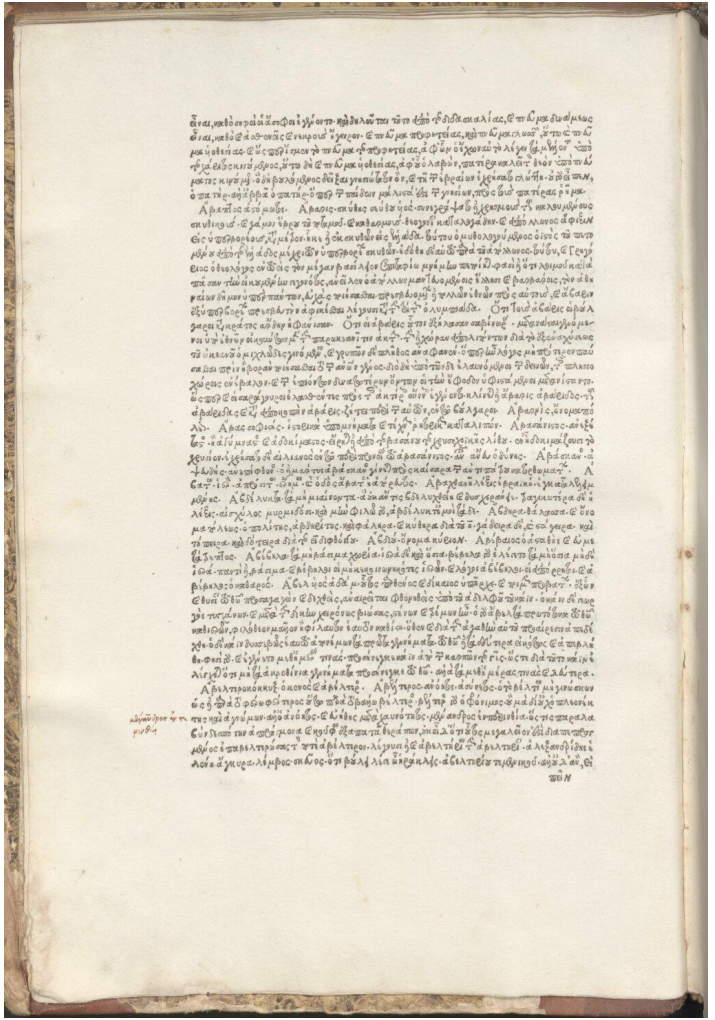
Founded in 1998

- What and Why?(Appraisal)
- How? (Capture/Ingest)
- Larger philosophical questions (Context/Respect des fonds)

Suda Off Line: Byzantine Lexicography

1499 edition housed by the Bavarian State Library and shared online by the Munich DigitiZation Center (MDC)

https://daten.digitale-sammlungen.de/bsb00050756/image_16



Suda On Line: Byzantine Lexicography

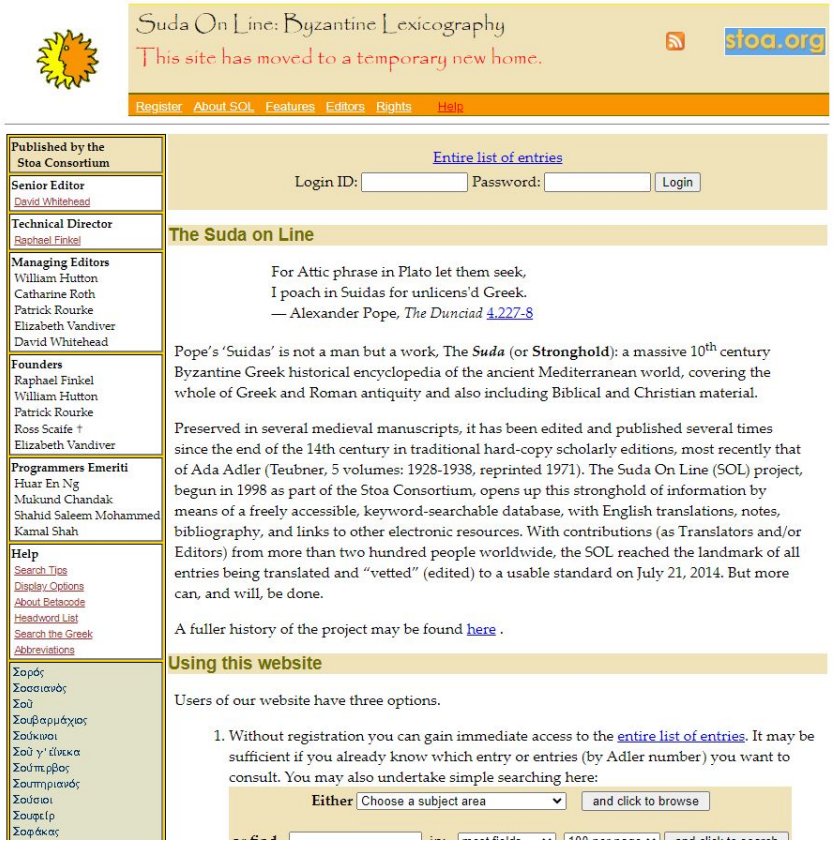
Founded in 1998

- Translation of 10th century encyclopedia written in Greek
- Crowd-sourced project, pre-dating Wikipedia
- Product of the Stoa Consortium for Electronic Publication in the Humanities
- Only English translation of the Suda, in any format

Suda On Line: Byzantine Lexicography

Founded in 1998

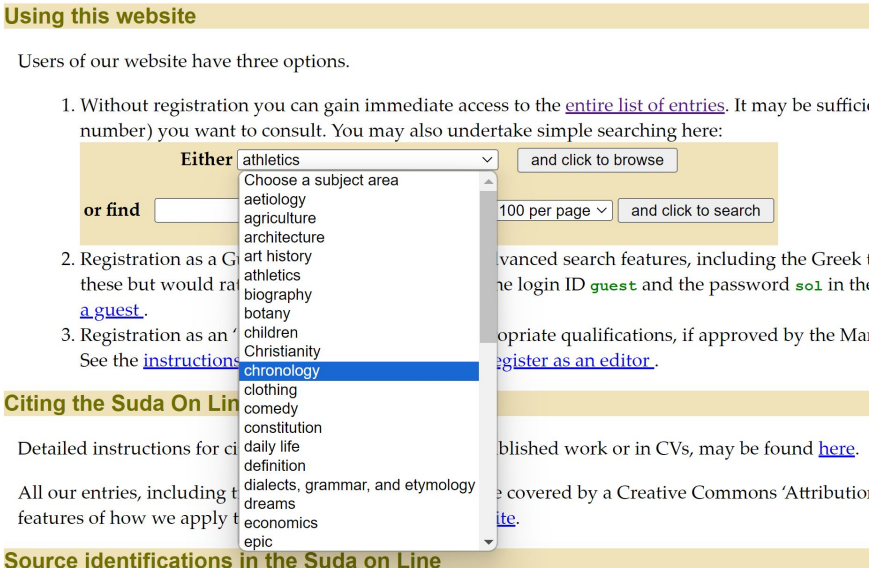
Still live:
www.cs.uky.edu/~raphael/sol/sol-html



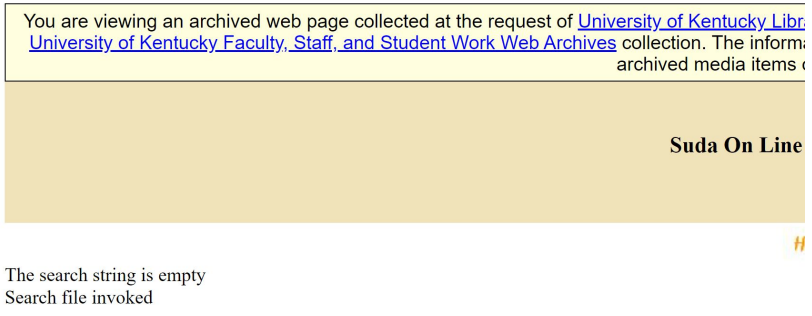
Why Won’t you Just Crawl?

…

Server-Side Operations



Preserving function, not just form.



Idea 1:
Preserve the Whole Server
...

The Importance of the Donor

Passive vs. Active Web Archiving

A bird in hand.

OVF/OVA file

-Open Virtualization Format/Appliance

-Preserve the full disk image as a virtual machine

-Document content and instructions

-Easy!

A bird in hand.

OVF/OVA file

-Open Virtualization Format

-Preserve the full disk image as a virtual machine, import in VirtualBox

-Document content and instructions

~~-Easy!~~

Possible?

———

A bird in hand.

OVF/OVA file

-Open Virtualization Format

-Preserve the full disk image as a virtual machine

-Document content and instructions

~~-Easy!~~

~~Possible?~~

Possible!

The .OVA
It's something.

Pros

-The full server is preserved

-It works

Cons

-Offline storage limits access

-Long-term sustainability?

-Requires full access to server

-Requires all website components share one server (no external resources/API integration)

Idea 2:
ReproZip-Web

•••

ReproZip-Web

They’re presenting here tomorrow!

ReproZip-Web

Capture server-side actions

Also capture external website elements

Package & reproduce sites far more complex than Suda On Line

ReproZip-Web

Pros:

-Potentially ideal for efficiently capturing DH projects

-Smaller output than server image

-Lower barrier to serving access than VM

ReproZip-Web

Challenges:

-Technical complexity

-Requires access/permissions to servers

-Widespread acceptance of .rpz standard/sustainability

ReproZip-Web

That being said:

Oui.

.RPZ

—

A Larger Question

...

What is a website when it’s no longer online?

The Offline Suda On Line

To what extent is SOL a tool?

To what extent a text?

—

The Offline Suda On Line

Aboutness & Context.

The Future of the Suda On Line

The conversation continues...

References

Boss, Katherine, Ilya Kreymer, Vicky Rampin, and Rémi Rampin. “Server-Side Web Archiving Using Reprozip-Web”. *iPRES 2023 Proposal Submission*. https://drive.google.com/file/d/1D9h_qc7jd4KP58DxlsFhzNxr7ckiPtPy/view . Accessed 2024 April 14.

Mahoney, Anne. “Tachypaedia Byzantina: The Suda On Line as Collaborative Encyclopedia”. *digital humanities quarterly*. 3:1, 2009, https://www.digitalhumanities.org/dhq/vol/003/1/000025/000025.html . Accessed 2024 April 14.

Suda On Line: Byzantine Lexicography. https://www.cs.uky.edu/~raphael/sol/sol-html/index.html . Accessed 2024 April 14.

Suda, 1499, National Library of Munich, https://daten.digitale-sammlungen.de/0005/bsb00050756/images/index.html?fip=193.174.98.30&id=00050756&seite=1 . Accessed 2024 April 14.

Theimer, Kate. “Archives in Context and as Context.” Journal of Digital Humanities. June 2012, https://journalofdigitalhumanities.org/1-2/archives-in-context-and-as-context-by-kate-theimer/ . Accessed 2024 April 14.

University of Kentucky Web Archives, https://archive-it.org/organizations/915 . Accessed 2024 April 14.

Thank you!

Questions?

Andrew McDonnell
Digital Archivist
University of Kentucky Libraries
Special Collections Research Center
mcdonnell@uky.edu

Presentation Link:
tiny.cc/ukweb2024