MAYO CLINIC

INFORMATICS PROFESSIONALS. LEADING THE WAY.

MedTator: A Lightweight Interactive Multi-Document Annotation Tool

S36: Systems Demonstrations - Improving Workflow:
Informatics Tools No Researcher Should Be Without

Huan He, PhD*

Sunyang Fu, PhD, Liwei Wang, MD, PhD, Andrew Wen, MS, Sijia Liu, PhD, and Hongfang Liu, PhD †

Department of AI and Informatics, Mayo Clinic, Rochester, MN, USA

* He.Huan@mayo.edu
† Liu.Hongfang@mayo.edu

Disclosure

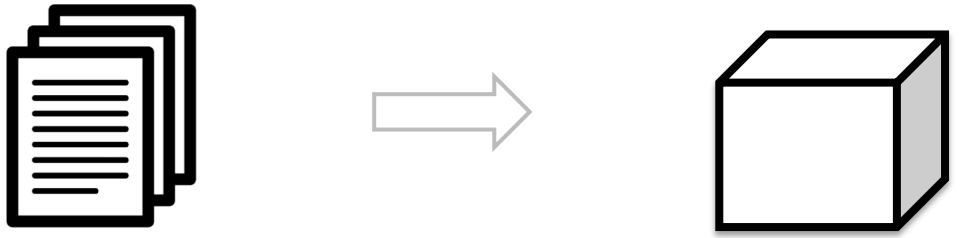
- I have no relevant relationships with commercial interests to disclose.

Learning Objectives

- Learn the serverless architecture for tool development
- Learn how to use MedTator for corpus annotation

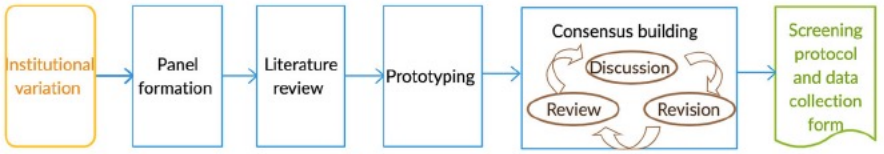
Background

“Hi, I have wonderful data, could you _____ with AI?”

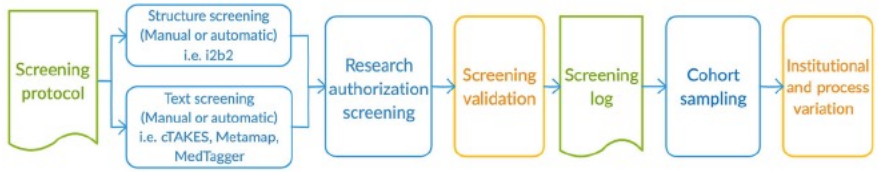


Adventure starts here

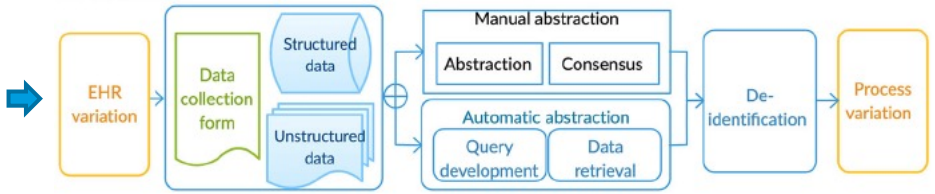
"Data"



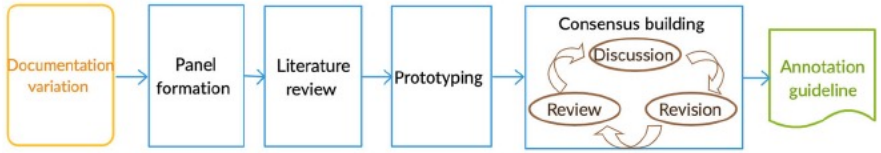
1. Protocol development



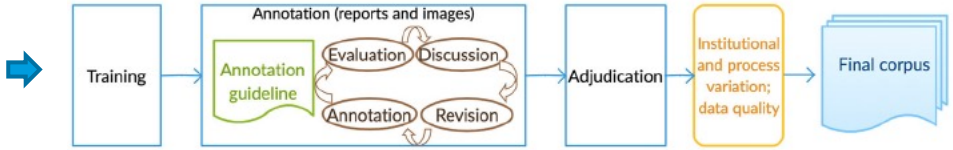
2. Data collection



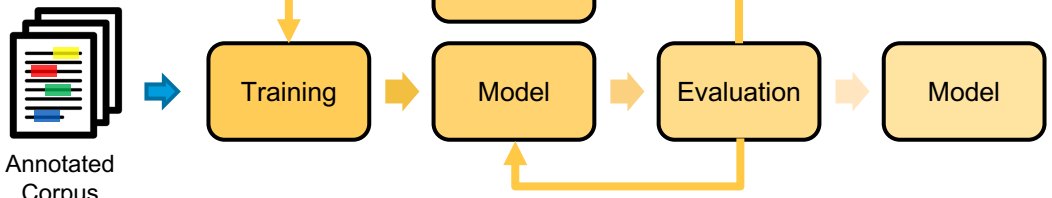
3. Cohort screening



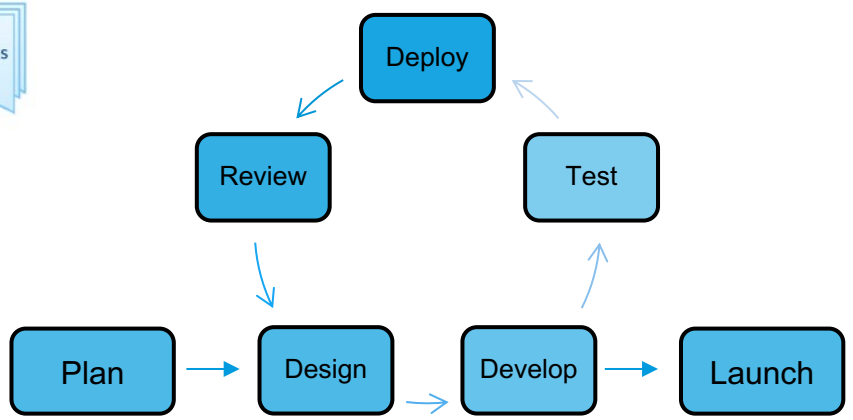
4. Guideline development



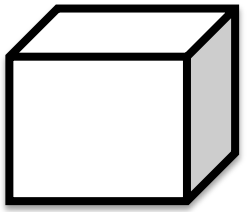
5. Corpus annotation



6. Model development



7. Software development



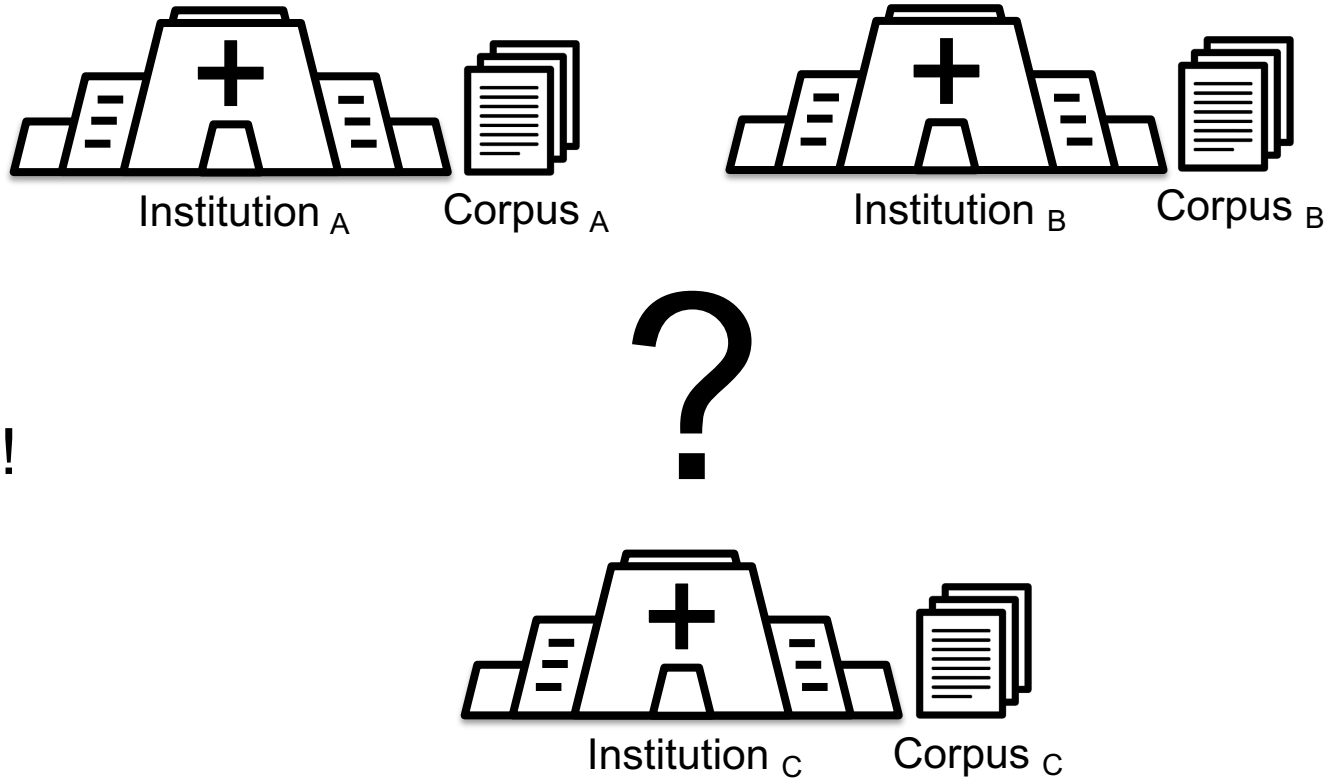
A fancy system

Fu, S., Leung, L.Y., Raulli, AO. *et al.* Assessment of the impact of EHR heterogeneity for clinical research through a case study of silent brain infarction. *BMC Med Inform Decis Mak* **20,** 60 (2020). https://doi.org/10.1186/s12911-020-1072-9

To conduct multi-site annotation

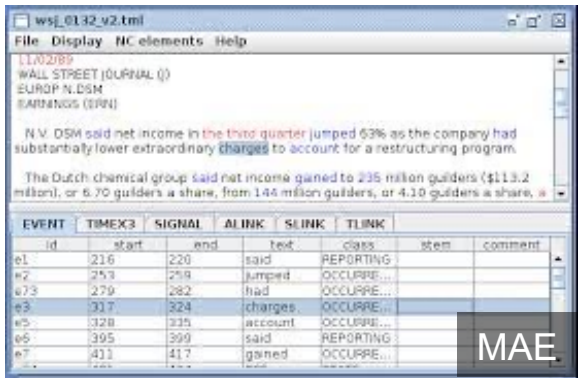
- Data security
- Corpus management
- Multi-site collaboration

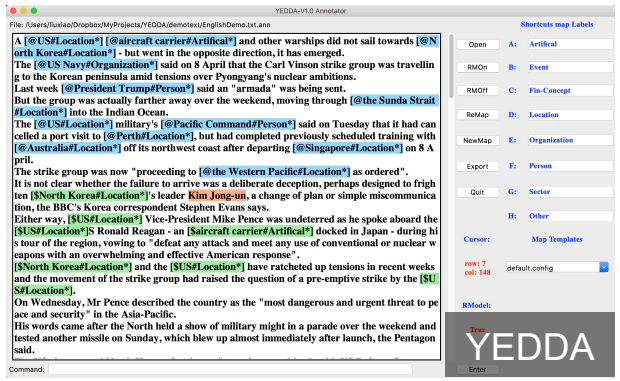
A text annotation tool is needed!

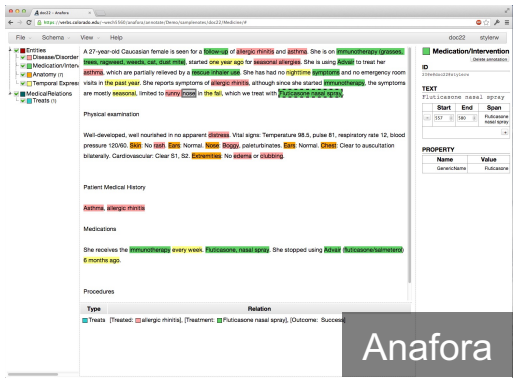


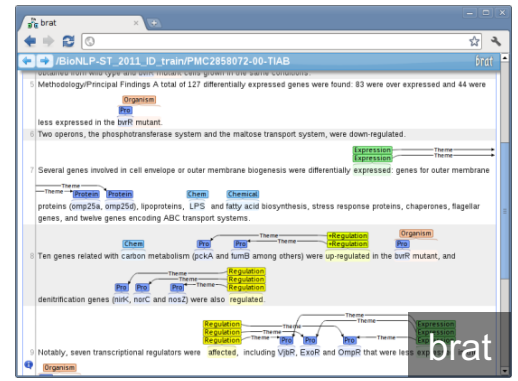
Text Annotation Tools

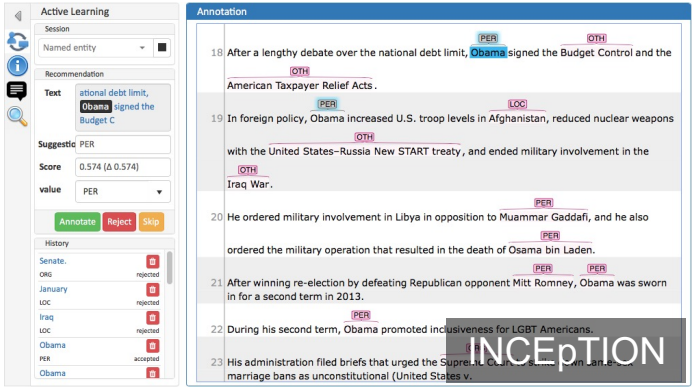
https://github.com/mariananeves/annotation-tools

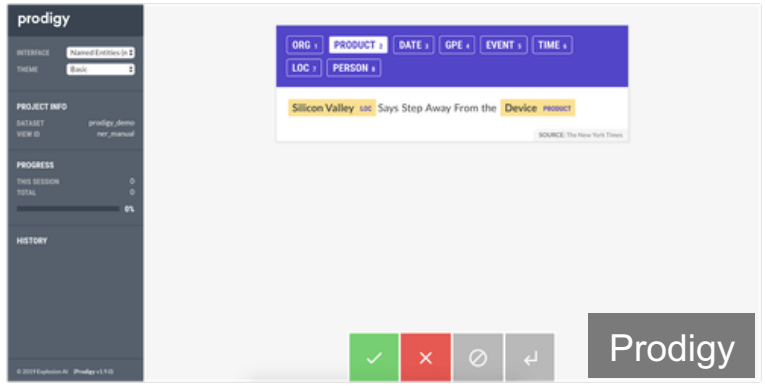


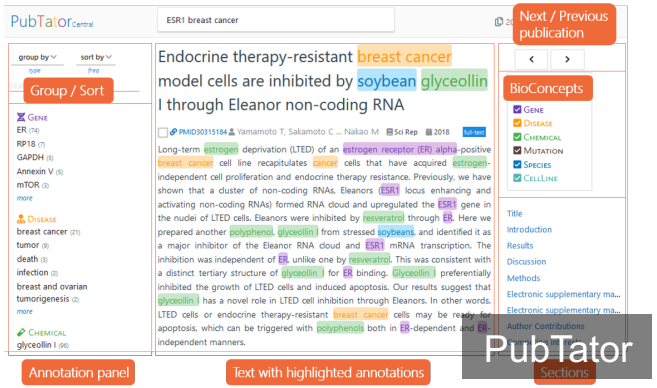


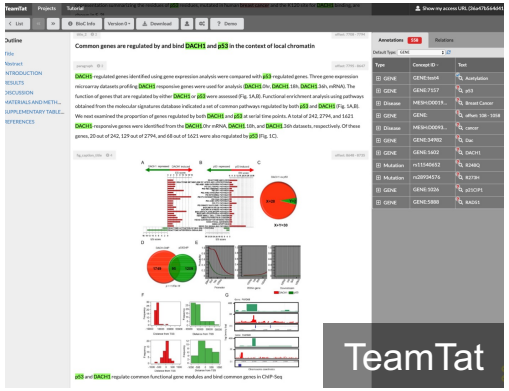


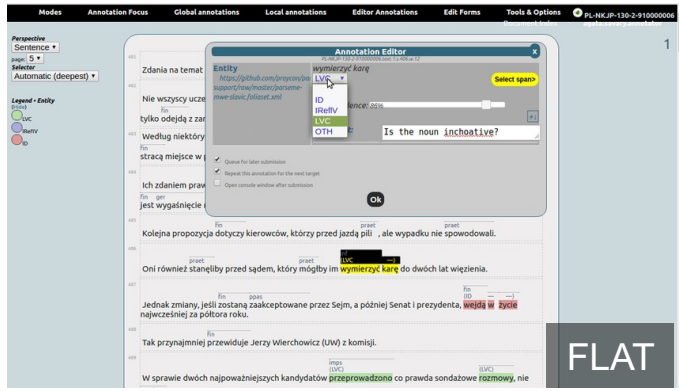


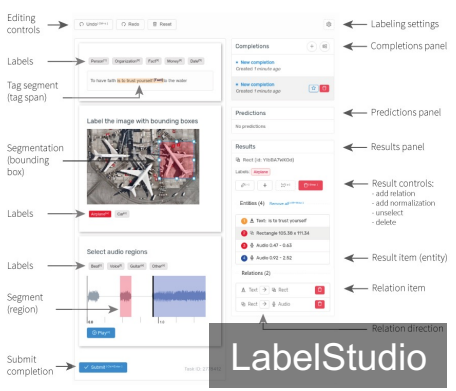


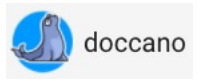


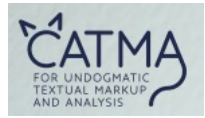












and many!

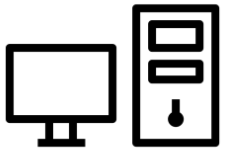
The Installation of Existing Tools

1. Prepare a server and get permissions
2. Download the package of an annotation tool
3. Read through README, instruction of installation
4. Install the prerequisites
   - Oh, something wrong, version conflict, not found, exception, error …
5. Install the tool and configure it
   - Oh, something wrong but different error
6. Finally, tool is installed but how to make it work?
7. Read through manual and try again
   - I did what they said in the manual, but it still didn't work, why?

The Architecture of Existing Tools

- Standalone tools
  - MAE
  - AGTK
  - Analec
  - @Note
  - BioAnnotate
  - Callisto
  - Ellogon
  - Glozz
  - MMAX2
  - NOMAD
  - RAD
  - ...

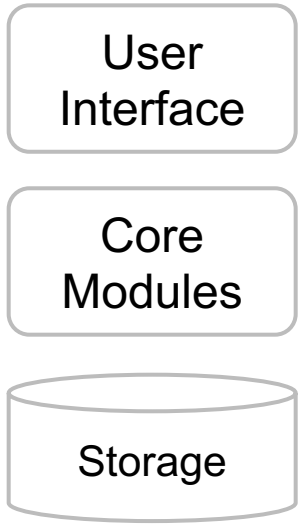
- Web-based tools
  - Anafora
  - eHost
  - YEDDA
  - BioQRator
  - brat
  - Catma
  - FLAT
  - WebAnno
  - INCEpTION
  - MAT
  - TextAE
  - …



https://github.com/mariananeves/annotation-tools

System Architecture Perspective

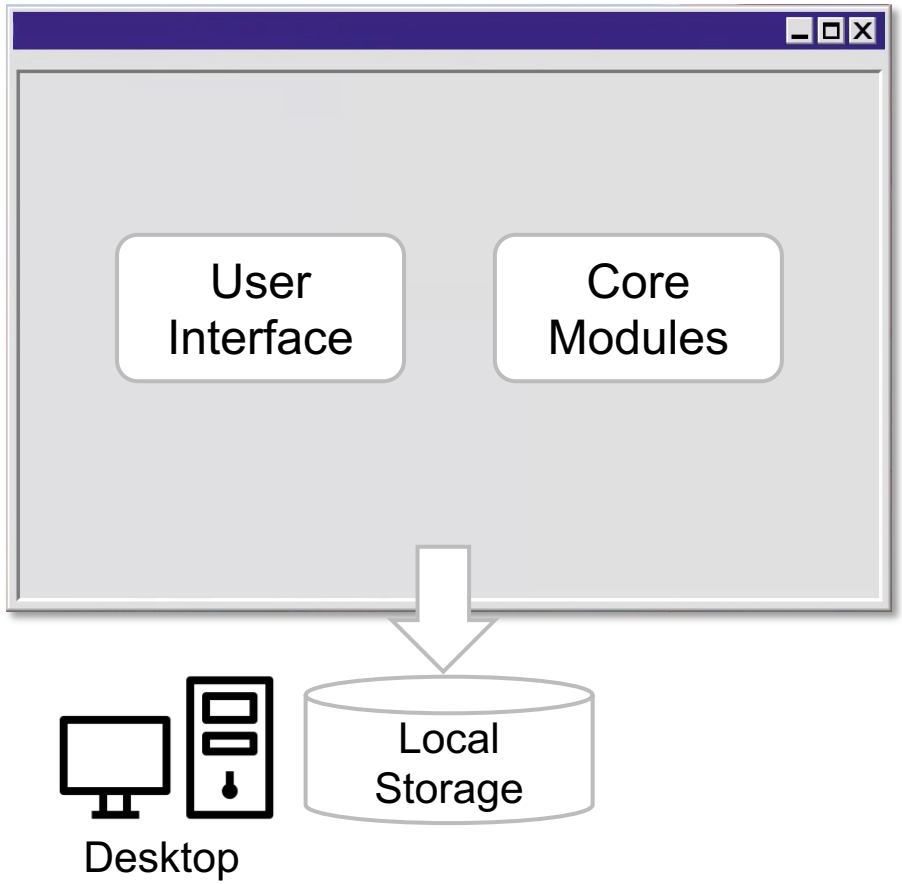
- Input
  - How the data enter system?
- Process
  - What to do with the entered data?
- Output
  - Where to save the processed data?



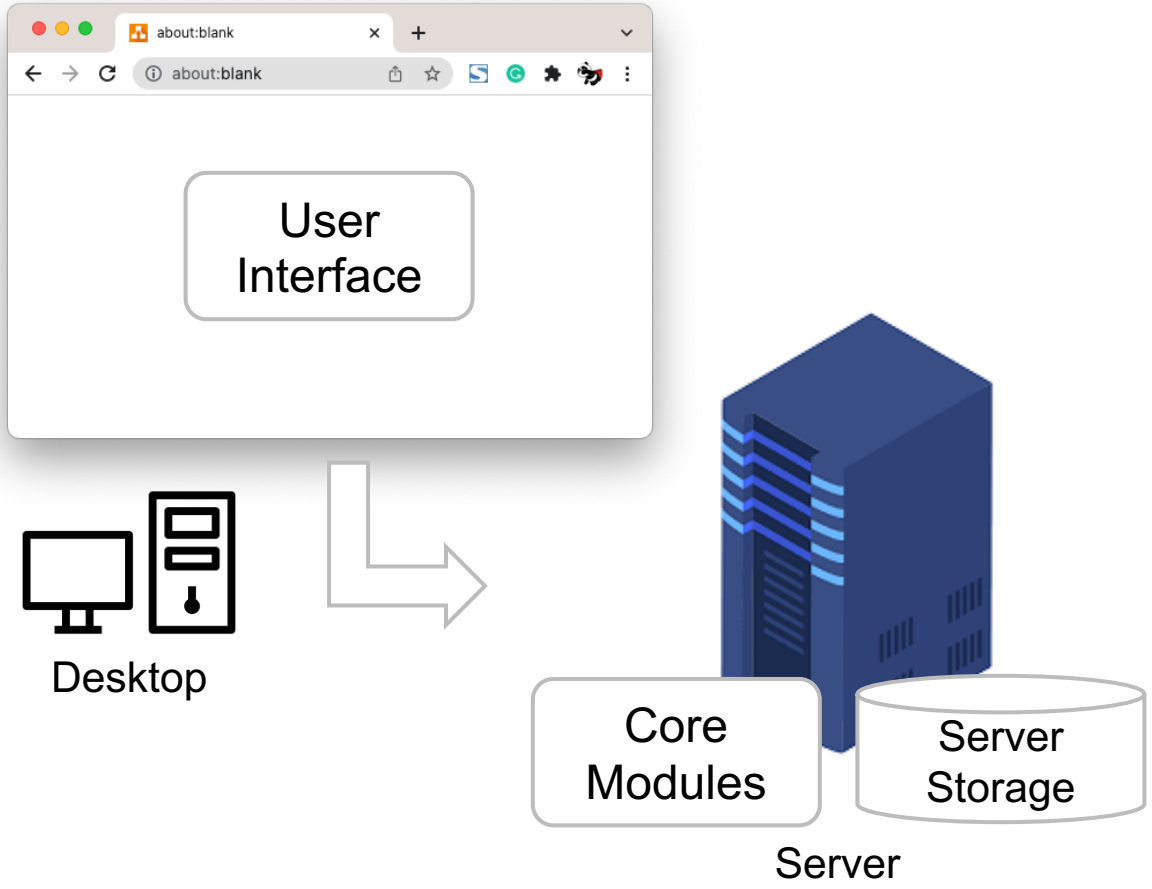
Text annotation Tool

The Architecture of Existing Tools

- Standalone tools

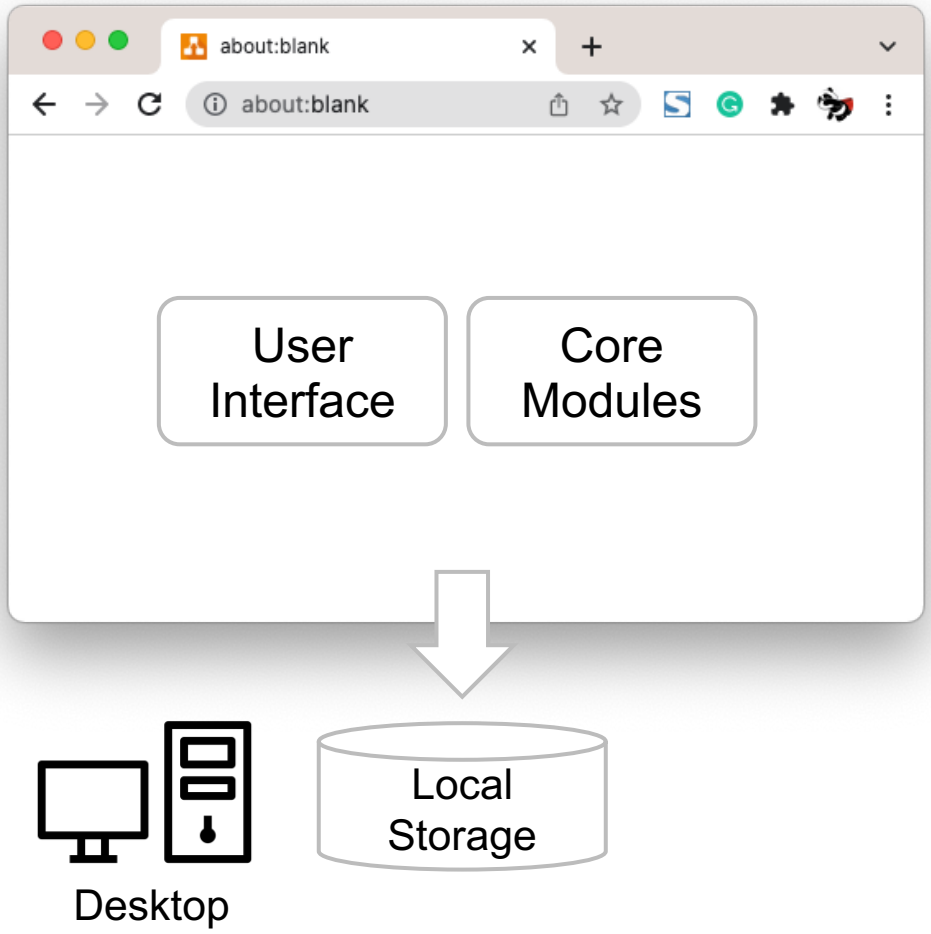


- Web-based tools



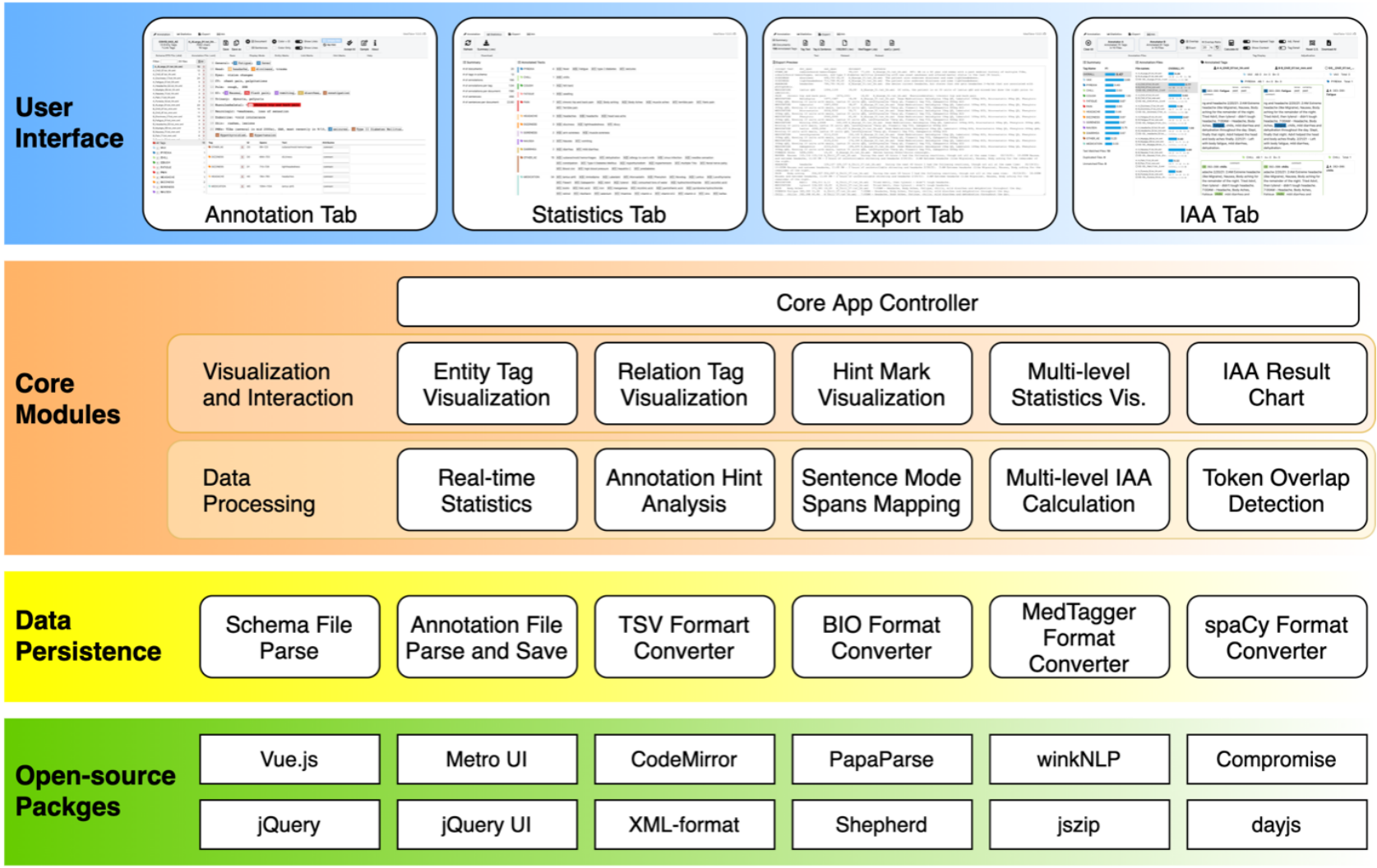
MedTator - Serverless Architecture

- Similar to standalone tools:
  - Data is saved in local storage
  - UI and core run togather
- Similar to web-based tools:
  - UI runs in web browser
- Advantages
  - No need to install



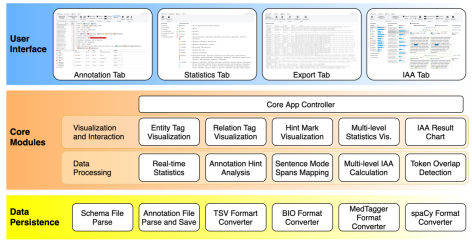
Functionalities

- Text annotation
- Statistics
- Adjudication
- Export



Implementation

Single HTML file



index.html

GitHub Pages

JS libraries

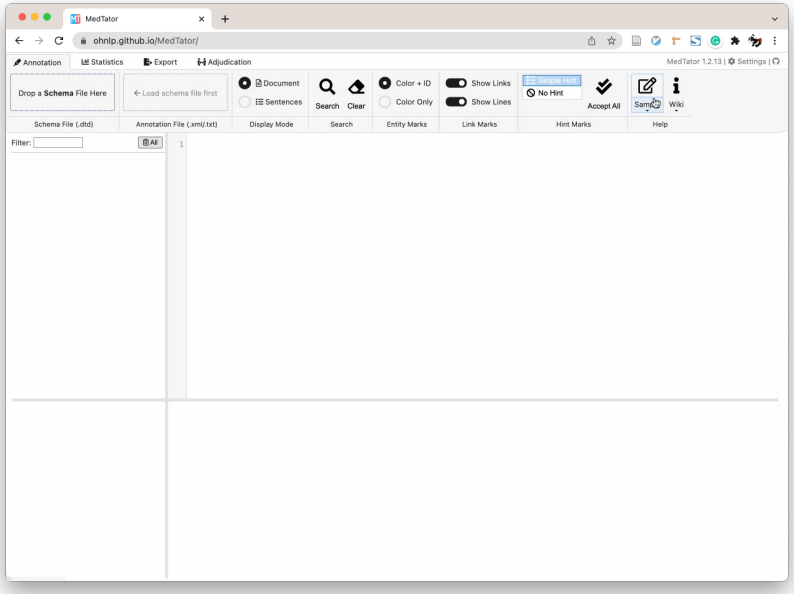
Vue.js	Code Mirror
Metro UI	jQuery
PapaParse	xlsx
jszip	D3.js
dayjs	FileSaver
wink-NLP	. . .

UNPKG

JSDELIVR CLOUDFLARE®

Content Delivery Network (CDN)

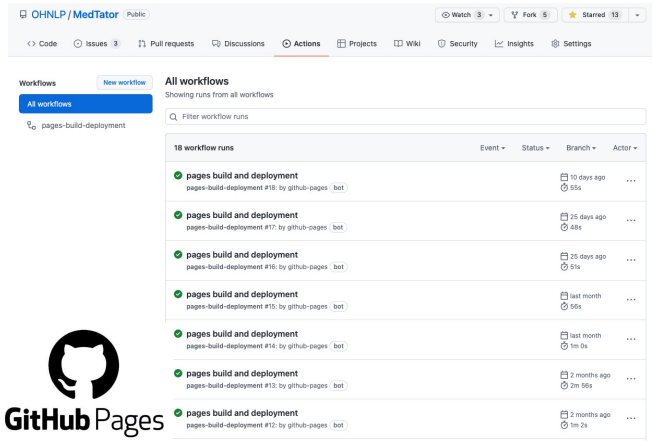
Web Browser



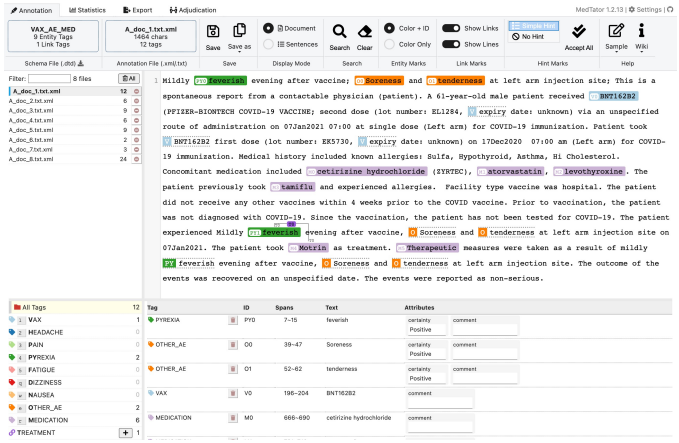
Chromium-based browser

Advantages

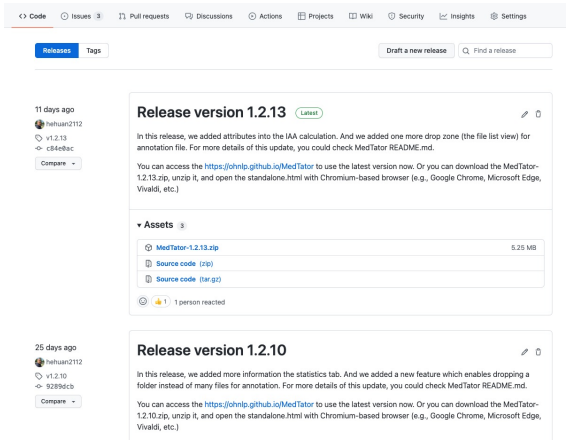
- Researchers are not concerned with capacity planning, management, maintenance, or scaling of virtual machines, or physical servers.
- **Free** public services: GitHub Pages, GitLab Pages, BitBucket Pages…
- Easy to deploy, “install”, and update



`git push` to GitHub Pages to deploy on free domain name https://ohnlp.github.io/MedTator/



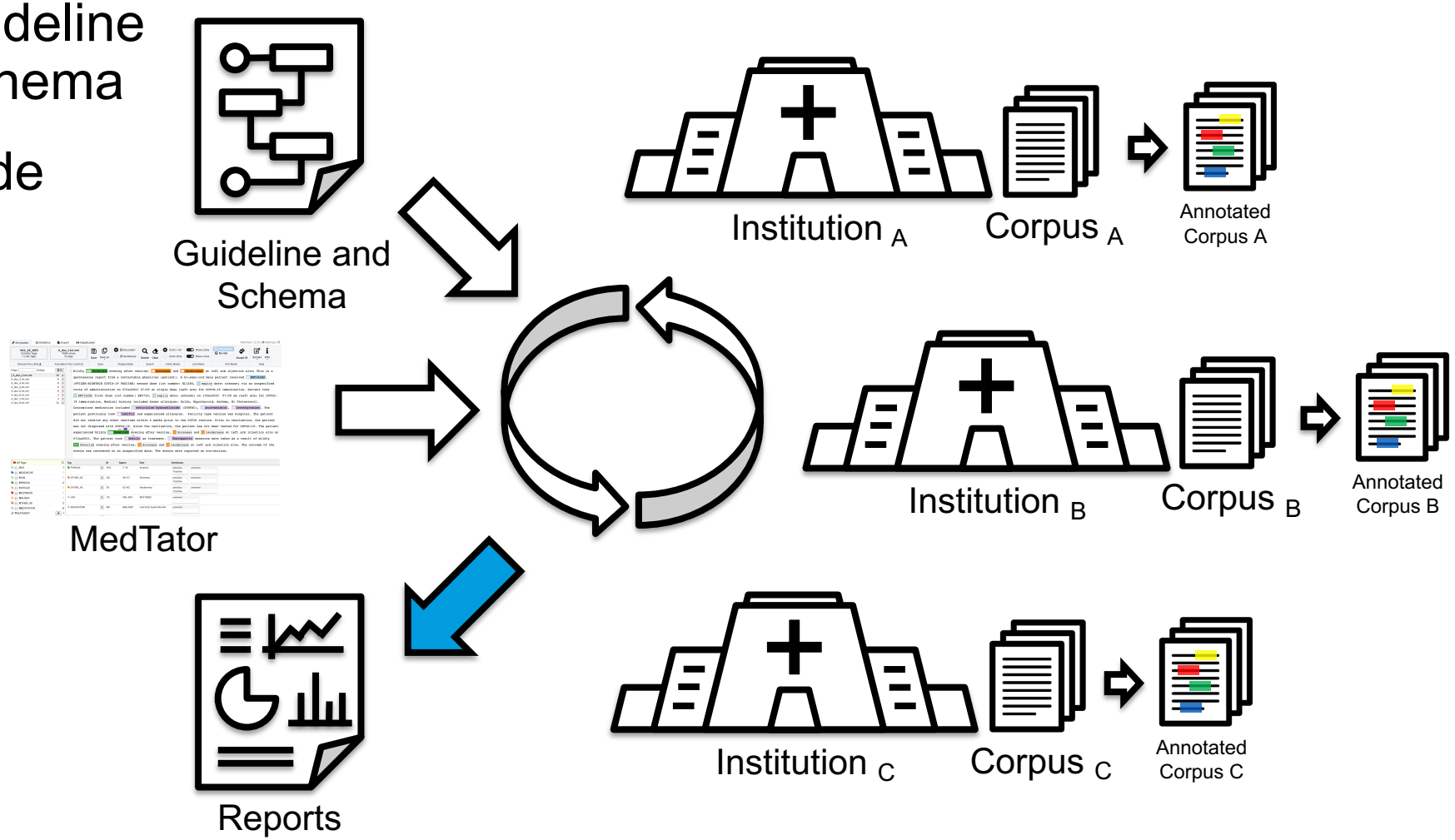
Global access with fast public CDN
No need to install
(Or run as a local static webpage)



Release new version online

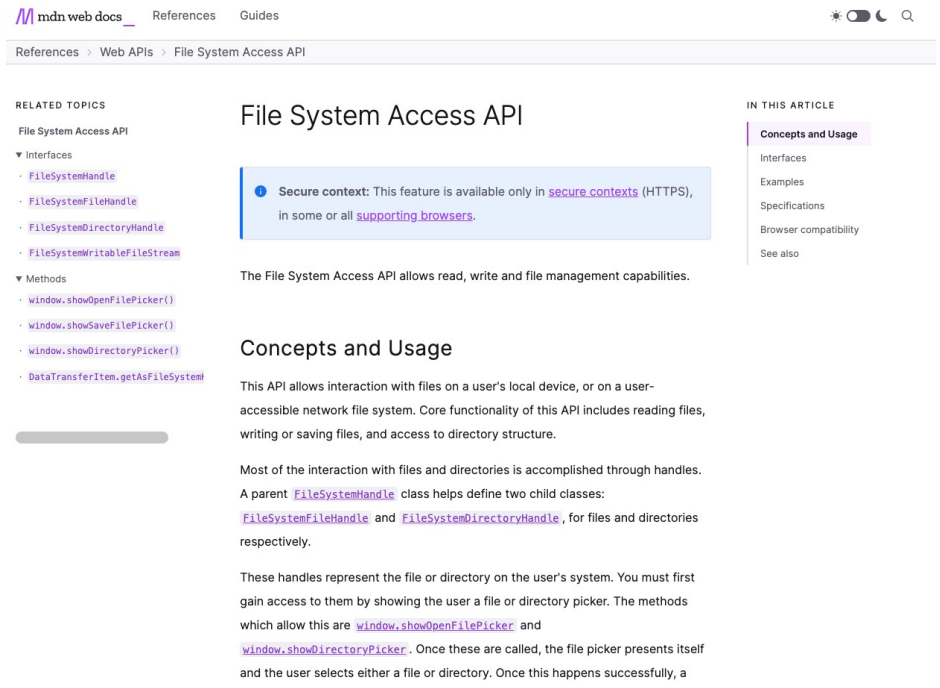
Advantages - Federated annotation

- Only share the guideline and annotation schema
- No centralized node
- No server cost
- No PHI data transfer
- Same annotation user experience



Limitations - File management in HTML5

- File system access API
  - Not supported by all browsers



Browser compatibility

Report problems with this compatibility data on GitHub

	Chrome	Edge	Firefox	Internet Explorer	Opera	Safari	WebView Android	Chrome Android	Firefox for Android	Opera Android	Safari on iOS	Samsung Internet
FileSystemHandle ⚠	✓ 86	✓ 86	✗ No	✗ No	✓ 72	✓ 15.2	✗ No	✓ 86	✗ No	✗ No	✓ 15.2	✓ 14.0
isSameEntry ⚠	✓ 86	✓ 86	✗ No	✗ No	✓ 72	✓ 15.2	✗ No	✓ 86	✗ No	✗ No	✓ 15.2	✓ 14.0
kind ⚠	✓ 86	✓ 86	✗ No	✗ No	✓ 72	✓ 15.2	✗ No	✓ 86	✗ No	✗ No	✓ 15.2	✓ 14.0
name ⚠	✓ 86	✓ 86	✗ No	✗ No	✓ 72	✓ 15.2	✗ No	✓ 86	✗ No	✗ No	✓ 15.2	✓ 14.0
queryPermission 🧪 ⚠	✓ 86	✓ 86	✗ No	✗ No	✓ 72	✗ No	✗ No	✓ 86	✗ No	✗ No	✗ No	✓ 14.0
requestPermission 🧪 ⚠	✓ 86	✓ 86	✗ No	✗ No	✓ 72	✗ No	✗ No	✓ 86	✗ No	✗ No	✗ No	✓ 14.0

✓ Full support ✗ No support 🧪 Experimental. Expect behavior to change in the future.

⚠ Non-standard. Check cross-browser support before using.

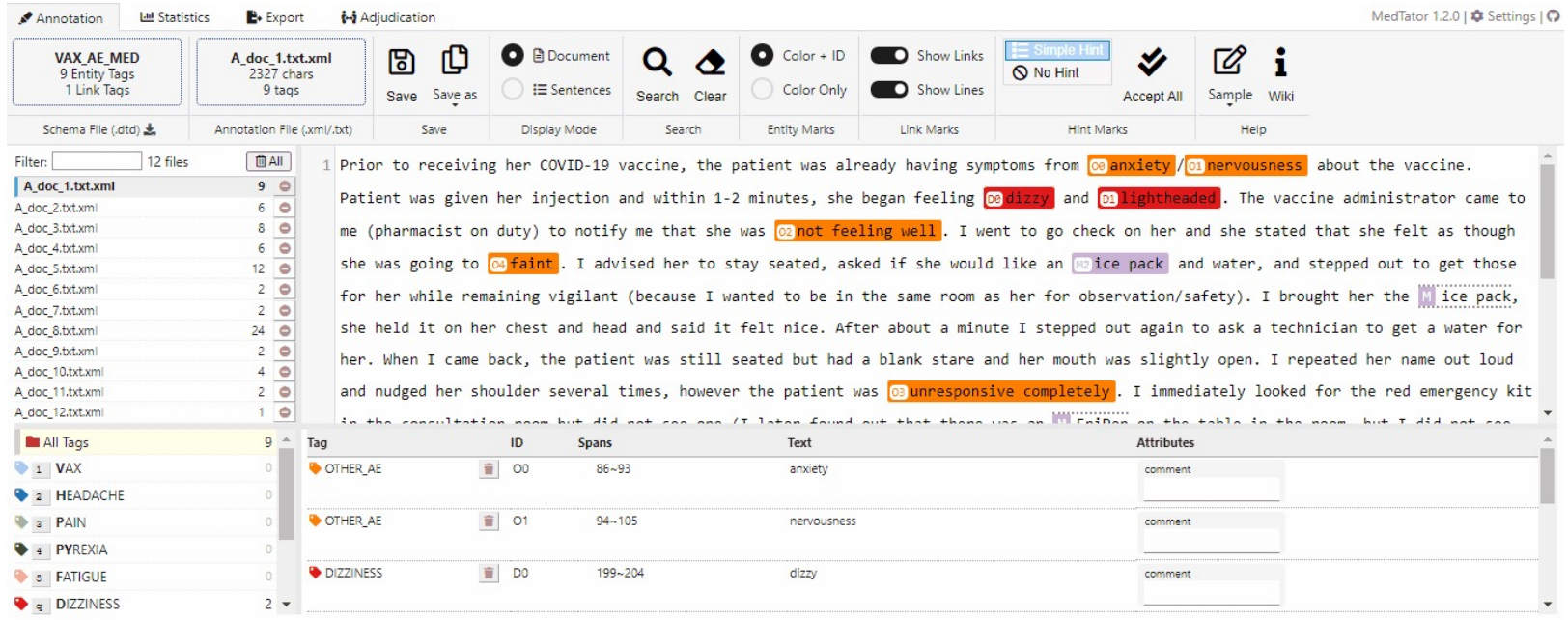
Limitations - Functionality (for science)

- Limited support in scientific computing libraries for advanced features

	JavaScript	**Python**
Computation	NumJS	NumPy
	Danfo.js	Pandas
	std.js	SciPy, statsmodels
Machine Learning / NLP	tensorflow.js compromise, winkNLP	scikit-learn, tensorflow spaCy, flair
Data Visualization	D3.js, p5.js ECharts Plotly	Matplotlib, seaborn

Demonstration

- Live demo with sample datasets: https://ohnlp.github.io/MedTator/



Thank you!

Email: Hongfang Liu (Liu.Hongfang@mayo.edu)
Huan He (He.Huan@mayo.edu)

MedTator: https://ohnlp.github.io/MedTator/