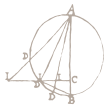


EASL, Oslo, 4-6 September 2019

RISE and SHINE: A distributed approach for text-centric research infrastructure

Sean Wang, Shih-Pei Chen

Max Planck Institute for the History of Science (MPIWG)

Current ecosystem for resource-centric research

Primary and **secondary** sources

Analog or **digital** (databases, e-books…)

External (commercial, freely available…)
or **internal** (produced by libraries themselves)
to the libraries

Libraries as mediators who acquire, negotiate, and/or provide access for scholars

Scholars

The trend in digital scholarship

Scholars increasingly use computational (and other more-than-read) methods to analyze resources.

This trend may be challenging for libraries to support…

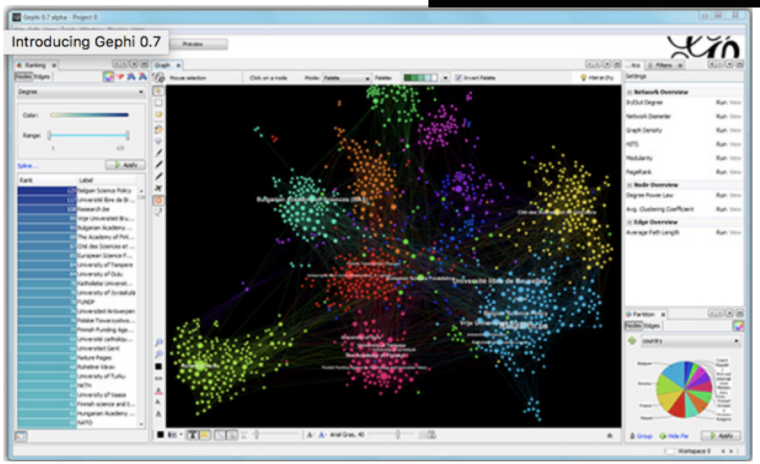
Technically, there are too many methods and tools, in different places and with different status, for libraries to provide in-house;

Legally, such usage often requires full access to large amount of computer-readable materials (e.g., “text-mining licenses” for commercial resources).

Existing research tools for general DH purposes

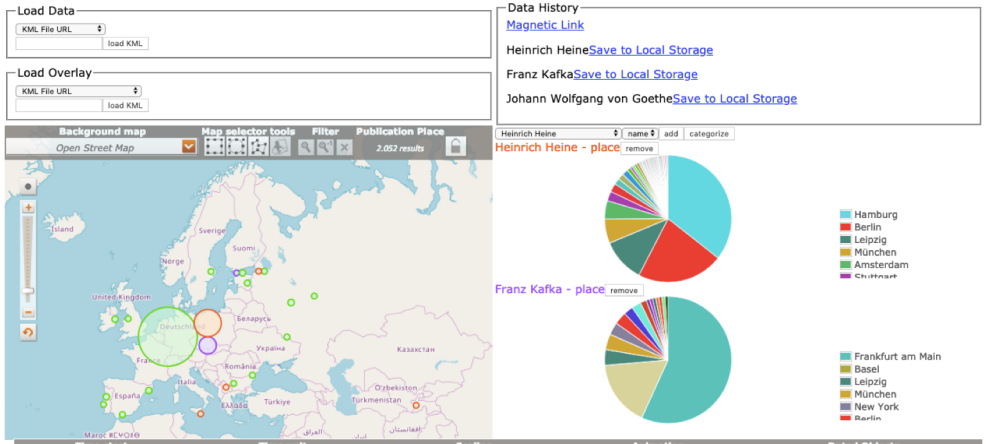
Network analysis

Gephi
makes graphs handy



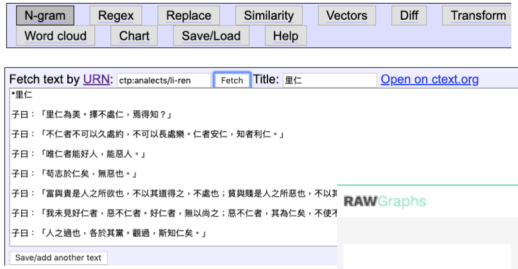
Mapping

GeoTemCo
TOPOI
PLATIN
Place and Time Navigator
MAX-PLANCK-INSTITUT FÜR WISSENSCHAFTSGESCHICHTE
DARIAH-DE

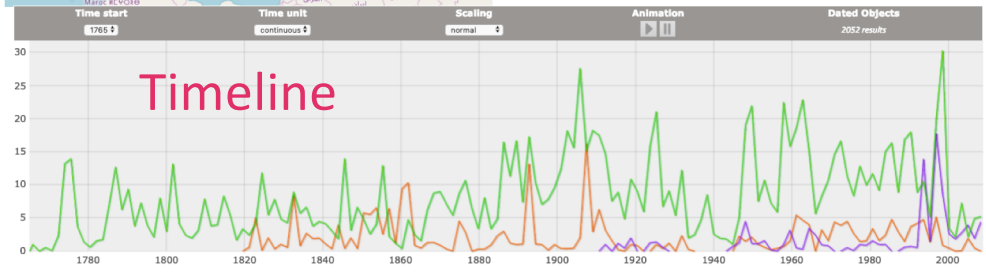


Text analysis

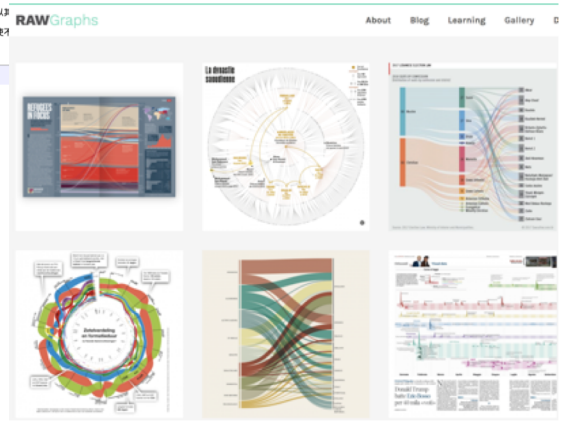
VOYANT
see through your text



Timeline



Visual analytics



Text markup

How could libraries better support digital scholarship?

Possible scenario #1: A non-reproducible solution to working with (commercial) resources

Publishers sometimes work with scholars on an *ad hoc* basis to provide downloadable 'raw data'. Some cultural institutions also provide data in this manner.

Challenging to replicate on a large scale, and makes reproducibility more difficult.

Risks losing track of content's lifecycle (e.g., versions) and further (re-)usage.

How could libraries better support digital scholarship?

Possible scenario #2: Big silos

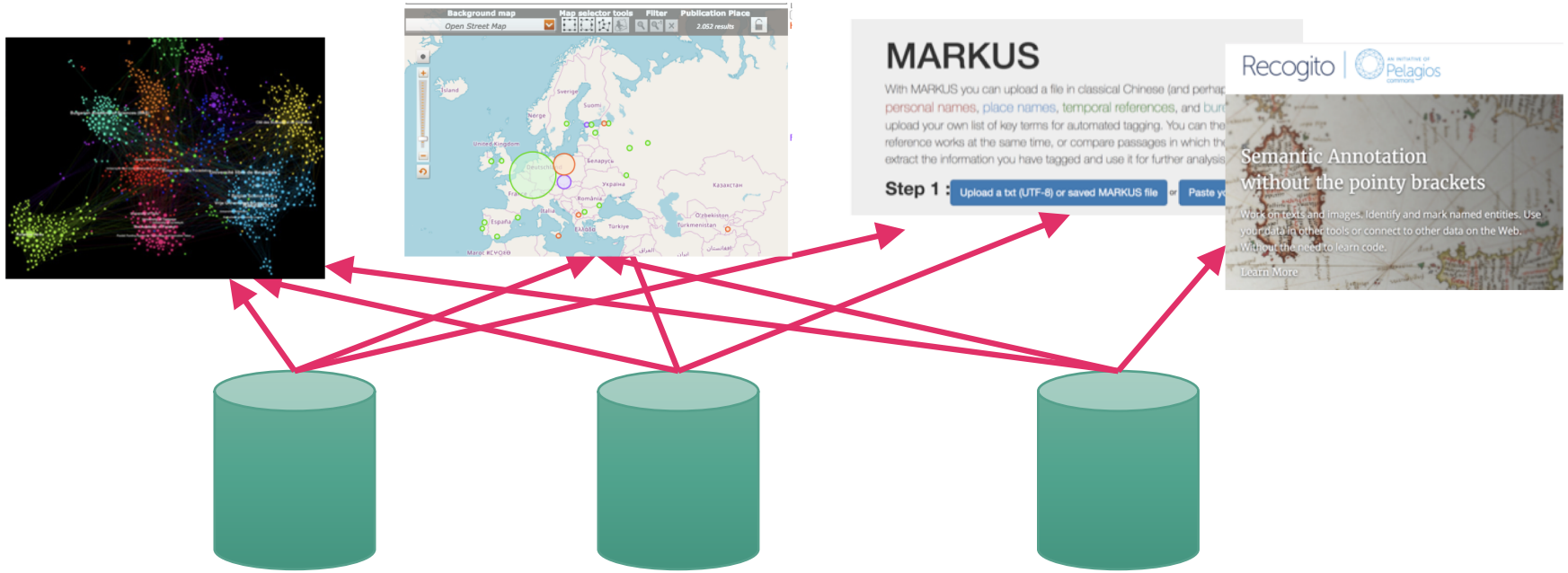
Some large research infrastructures provide encapsulated environments for scholars to use computational methods on open-access or commercial resources (e.g., HathiTrust, TextGrid)

⇒ Open platforms cannot work with commercial resources.
⇒ Closed platforms (such as encapsulated environments) suffer from limited tool coverage and challenging licensing negotiations.

For scholars, such silos are limiting in terms of both available resources and tools.

Our proposal: a distributed network of resources and tools

Linking contents with research tools in place, creating a *distributed network* of resources and tools.



Scholars can *choose resources from any databases* and *send them to any tools of choice* to analyze, if they have proper rights.

Our proposal: RISE and SHINE

https://rise.mpiwg-berlin.mpg.de

A common *API* (a language that everyone talks) for transferring texts between databases and tools => SHINE

A secure and trust-worthy *guard in between every transfer* to make sure a text is transferred to an authenticated user from an authorized institute => RISE

The RISE infrastructure

A suite of software packages covering the needs of different stakeholders

A ***middleware*** that catalogs all linked resources, and authenticates and authorizes text transfers

Management interfaces for libraries / resource providers, research institutes, and tool developers to set privileges

A suite of ***JavaScript libraries*** to allow easy integration with the SHINE API for software developers

A ***Resource Provider*** software package that allows resource providers (database owners, archival institutions, or even scholars) to share resources in a protected, SHINE-compatible way

Challenges: how to make the RISE middleware work for libraries?

How to work with libraries' internal database management systems automatically?

- To avoid duplication of efforts!

How to make authentication and authorization seamless?

- Options beyond Shibboleth and RISE's own user registry?

How to technically represent the entire spectrum of licensing rights, from fully open to completely protected?

Call for collaboration

This should be a network built *by* the community

- Work together to define this network
- Call for collaboration with libraries to test this concept!

RISE will allow libraries to offer digital scholarship with both licensed and open resources!

Check our website for detailed documentations, API, & available toolkits

- **https://rise.mpiwg-berlin.mpg.de**