



Research Infrastructure for the Study of Eurasia (RISE)

**Towards a flexible and distributed digital
infrastructure for resource access via
standardized APIs and metadata**

Sean Wang Pascal Belouin

Shih-Pei Chen Hou Ieong “Brent” Ho

Max Planck Institute for the History of Science

Research Infrastructure for the Study of Eurasia (RISE): Towards a flexible and distributed digital infrastructure for resource access via standardized APIs and metadata

**Sean Wang, Pascal Belouin, Shih-Pei Chen, and Hou Jeong “Brent” Ho
Max Planck Institute for the History of Science**

Abstract

Digital humanities (DH) is a burgeoning field of research in sinology and Asian studies more broadly, and its diversity and maturity necessitate a digital research infrastructure fit for DH-focused scholars’ specific needs. In particular, the DH landscape evolved in a way that encourages fragmentation of both sources and tools, and these compartmentalized resources centered around disciplines and texts. **RISE** (formerly known as “Asia Network”) is our solution to address this fragmentation across disciplines. It is a pioneering approach for resource dissemination and emerging data analytics (such as text mining and other fair-use, consumptive research techniques) in the humanities. It is a language-agnostic software that facilitates the secure linkage between third-party research tools to different third-party textual collections (both licensed and open-access ones) via application programming interfaces (APIs). Put more simply, it reduces the distances among DH resources not by duplicating them in a central repository, but by linking them together via flexible APIs. It revolutionizes how scholars can work with textual sources by promoting a flexible, networked approach to digital infrastructure development. Crucially, RISE is a loosely-coupled software with flexible topologies; it can enable both federated or centralized linkages, and it can even “disappear” as long as its API and metadata standards remain in place to facilitate communications among distributed databases and tools in the back-end. Thus, unlike large-scale infrastructural projects, RISE actively lowers the profile of centralized infrastructure and instead promotes existing tools and resources by enabling their interoperability in a flexible and distributed manner. As a result, it allows scholars to fully leverage the potential of material digitization and digital research tools without re-creating silos of resources in the digital realm. We believe that RISE, coupled with developing novel licensing models suited for digital research methods (including consumptive research like text mining), would significantly improve the infrastructure behind DH scholarly research in sinology and beyond.

Keywords

digital humanities, sinology, cyberinfrastructure

1. Introduction

Digital humanities (DH) is a burgeoning field of research in sinology and Asian studies more broadly. DH research techniques, including various databases from digitization efforts and growing numbers of digital research tools, have had an impact on Sinologist research communities globally. Stanford University, for example, has held an annual “Digital Humanities Asia” conference since 2016¹, and this venue is the ninth International Conference of Digital Archives and Digital Humanities. There are also collaborations across multiple regions, as evidenced by the enthusiastic participants at the International Conference on Cyberinfrastructure for Historical China Studies this March at Harvard Center Shanghai.² Such events demonstrate the diversity and maturity of DH in sinology globally.

Outside of sinology, DH has been grappling with issues such as long-term sustainability and interoperability. In response, many have proposed that DH needs basic infrastructures behind research projects to ensure its long-term success. In Europe, for instance, CLARIN³ and DARIAH⁴ are two such large-scale research infrastructures for humanities. While they have done a tremendous job in centralizing available digital resources, much of their infrastructures remain at the administrative level, and their generic coverage across the entire humanities meant that their utility for a specific discipline like sinology is limited. How can we, as DH scholars and Sinologists, design a cyberinfrastructure fit for our specific needs, taking past experiences with these large-scale infrastructural projects into consideration?

“RISE” (formerly known as “Asia Network”) is our answer to this question.⁵ It is a pioneering approach for resource dissemination and emerging data analytics (such as text mining and other fair-use, consumptive research techniques) in the humanities. It is a language-agnostic software that facilitates the secure linkage between third-party research tools to different third-party textual collections (both licensed and open-access ones) via application programming interfaces (APIs). It revolutionizes how scholars can work with textual sources because, under the current condition, it is impossible for scholars to use digital research tools to analyze licensed textual collections without downloading or scraping the full texts, which violates licensing terms. The RISE software can securely pass through these licensed texts to digital research tools, thus allowing scholars to work in a legal manner and ensuring commercial publishers the safety of their collections. Such flexible, networked approach to e-infrastructure development avoids re-creating silos of resources in the digital realm and allow scholars to fully leverage the potential of material digitization and digital research tools. Crucially, RISE is a loosely-coupled software with flexible topologies; it can enable both federated or centralized linkages, and it can even “disappear” as long as its API standards remain in place to facilitate communications among databases and tools in the back-

¹ <http://dhasia.org/>

² <https://projects.iq.harvard.edu/cbdb/international-conference-cyberinfrastructure-historical-china-studies>

³ <https://www.clarin.eu/>

⁴ <https://www.dariah.eu/>

⁵ See <https://asia-network.mpiwg-berlin.mpg.de/> for the web user interface for RISE’s beta prototype.

end. Thus, unlike large-scale infrastructural projects, RISE actively lowers the profile of centralized infrastructure and instead promotes existing tools and resources by enabling their interoperability.

“RISE” stands for Research Infrastructure for the Study of Eurasia, and this name retains a degree of regional specificity even though RISE’s technical set-up works with resources and tools in all languages. RISE’s core developers have backgrounds in sinology and DH, and the ideas behind it grew out of disciplinary challenges there (especially with text mining and licensed textual resources). RISE’s functionalities, however, address common infrastructural issues across the DH landscape regardless of disciplines, and RISE’s development has also expanded to include multilingual resources and tools that pilot users at our institute work with in their research.

Since our inception in May 2017, RISE has progressed to the beta development stage and we plan to release it publicly by the end of this year. At the time of writing, RISE is linked via APIs to the following resources: Chinese Buddhist Electronic Texts (CBETA)⁶, the Taiwan History Digital Library⁷, the Kanseki Repository⁸ (Kanripo), the Chinese Text Project (CText)⁹, a small set of Staatsbibliothek zu Berlin’s classical Chinese collections, and Perseus Digital Library (open-access Greek and Latin materials).¹⁰ The only linked research tool is MARKUS¹¹, though DocuSky¹², Recogito¹³, LERA¹⁴ and other tools are on our immediate development horizon. It is important to note that RISE’s current linked resources span the entire spectrum in terms of license and copyright restrictions. Some, like Kanripo and Perseus, are completely open-access. Others, like the Chinese Text Project, are generally open-access but require a license subscription for advanced functionalities. And then there are proprietary resources, whose licenses (regardless of read-only or text-mining) are very expensive to acquire. We are committed to the principle of open access, but we also recognize that it is our current reality, perhaps more so in sinology, that many resources are held in private hands. While RISE alone cannot solve this issue (and licensing restrictions are also not the main focus of this paper), we believe RISE’s technical set-up provides a useful alternative for scholars to work with resources across various licensing restrictions and could induce some private publishers and database vendors to implement these technical standards.

Here in this paper, we outline (1) the current landscape of DH in sinology and what we see as the main challenges it presents to researchers; (2) a basic summary of RISE’s functions and

⁶ <http://www.cbeta.org/>

⁷ <http://thdl.ntu.edu.tw/index.html>

⁸ <https://www.kanripo.org/>

⁹ <https://ctext.org/>

¹⁰ <http://www.perseus.tufts.edu/hopper/>

¹¹ <https://dh.chinese-empires.eu/markus/beta/>

¹² <https://docusky.digital.ntu.edu.tw/DocuSky/ds-01.home.html>

¹³ <https://recogito.pelagios.org/>

¹⁴ <https://lera.uzi.uni-halle.de/?lang=en>

features design; and (3) a call for collaborators to develop common API and metadata standards for DH in sinology.

2. Landscape of DH in sinology

Digitization of historical materials has dramatically transformed how Sinologists gather research sources and approach research questions. As more and more archives, libraries, and other research institutions embrace digital technologies, DH-focused projects and initiatives in sinology would only continue to proliferate. This growth, however, cannot be assumed as a foregone conclusion, as the current landscape of DH in sinology is incredibly fractured and already presents many roadblocks to seamless access and sustainability. In this section, we briefly survey this fractured landscape and show how our RISE infrastructure bridges the fault lines within it.

The current landscape is fractured both geographically and thematically. It is not a hyperbole to say that sinology is a global discipline today, as China studies departments and research centers exist in many countries. However, a core-periphery relationship among the roles and foci of these nodes of global sinology research community persists. In Mainland China, where much of the primary sources still reside in various archives and institutions, the commercialization of resource digitization reigns supreme. Despite the fact that many of these sources originate in the public domain (or have long passed their copyright protection terms), commercial publishers build proprietary databases from the digitization and charges high royalties for access. Such commercial models are being copied by university libraries and presses as well. While open-access movements have been gaining steam in recent years and the Chinese DH community has grown dramatically, as seen in the third DH symposium at Peking University¹⁵, this sector has continued to rely on this commercial model and shows little signs of movement. In particular, subscription access often does not include provisions for text mining, full-text access, and other standard DH techniques today, as the pricing model still prioritizes read-only access. It should be noted that we are not advocating for eradicating commercial database vendors from this landscape; rather, it is to point out that their existing business model (and the accompanying lack of technical improvements) make it difficult to leverage the full potential of their digitized materials, even for researchers who have subscription access to their materials.

In Taiwan, DH's first strong foothold in the Chinese-speaking world, intersecting scholarly expertise in humanities and computer science resulted in a strong environment for the development of research databases and research tools development. Many of these databases are driven by thematic interests, such as collections of Buddhist texts or Taiwanese historical documents, just to name a few. Also of note is the DocuSky platform developed by the National Taiwan University, which allows individual users to organize and work on their own set of materials in one place. While there are certainly monetizing tendencies in Taiwan as well, it is notable that many of these databases and tools are built on open-access principles

¹⁵ <https://www.lib.pku.edu.cn/portal/cn/news/0000001622>

and, if there is a fee, it is usually only for high-usage clients and/or to cover basic maintenance costs. These practices encourage development and sharing of new tools and resources, and Taiwan hosted the first Asia-focused DH international conference in 2009.¹⁶ Since then, the International Conference of Digital Archives and Digital Humanities (DADH) has become an annual event, and the Taiwanese Association for Digital Humanities became a constituent organization of the global Alliance of Digital Humanities Organizations in 2018.¹⁷

Elsewhere in the world, especially in North America and Europe, thematic DH research projects dominate the landscape. Close alliances between scholars and librarians, aided by international funding bodies like the Andrew W. Mellon, Chiang Ching-kuo, and Luce Foundations create diverse projects that largely consume primary sources for producing research results and presentations. Nonetheless, long-term preservation and sustainability remain serious issues, as is the linking of individual silos of project repositories and making them interoperable. In the context of sinology-focused DH projects, Harvard's China Biographical Database (CBDB) exemplifies both the success and the pitfalls, as its continued growth over almost two decades of open-access development was threatened by funding uncertainty and, just this year, sold its distribution rights in Mainland China to a commercial publisher.¹⁸ In Europe, centralized governmental funding agencies like the European Research Council provides more stability, but similar stories exist as well. Nonetheless, many top research projects and tools (such as MARKUS, Ten Thousand Rooms¹⁹, and Ming Qing Women's Writings²⁰) come from Sinologists based in North America and Europe and continue to base at libraries and research institutions there.

So, how should an individual researcher in sinology approach this fractured landscape? If one is interested in primary sources, there are many individual databases that must be searched one by one. While there are open-access, full-text databases like the Kanseki Repository, the majority remain proprietary and read-only, making the usage of digital research tools on full texts incredibly difficult, as well as synthesis of sources from mixed copyright origins. In the rare cases where digital analyses and manipulation are possible via tools like MARKUS or LoGaRT²¹, sharing of results is challenging tool. If one is interested in integrating research products from many DH projects, many employ static silos of project websites or repositories without appropriate technical linkages that enable interoperability. It is heartening that DH in sinology has progressed to such a point where a critical mass of research and researchers meant that we must consider cyberinfrastructure, and our "RISE" is a proposed solution that bridges these complex fault lines in this fractured global landscape.

3. Primary issues to be addressed

¹⁶ <http://www.dadh-record.digital.ntu.edu.tw/Scope.php?LangType=en&His=D09k2>

¹⁷ <https://www.adho.org/announcements/2017/adho-welcomes-new-organizations-0>

¹⁸ <https://projects.iq.harvard.edu/cbdb>

¹⁹ <https://tenthousandrooms.yale.edu/>

²⁰ <http://digital.library.mcgill.ca/mingqing/>

²¹ <https://www.mpiwg-berlin.mpg.de/research/projects/logart-local-gazetteers-research-tools>

The current landscape of DH in sinology and beyond, in our opinion, severely restricts scholars' ability to conduct digital research. To summarize, we identified three primary issues that any basic cyberinfrastructure must address.

3.1. Unconnected resource silos

Many research resources in DH are thematic, and some have digital research tools specifically developed to match or fit particular texts. While this research method may be the most efficient way to address a particular research question, at a large scale it re-creates the same fractured, unconnected silos of resources from the physical world to cyberspace. For example, many museums and cultural heritage collections have digitized their materials and put them online. But even in cases where there are common metadata standards, there often still is not for resources to be accessed or shared across multiple collections. In effect, we have buckets and buckets of valuable resources that could only be accessed by scholars one by one.

3.2. Heterogeneous access and exchange data formats

We certainly do not pretend to be the only ones to recognize the previous issue, and many attempts have been made to develop common metadata formats to aid resource access and exchange. For texts, TEI has been one such popular format²², and more recently IIIF has become widely used for images.²³ Since resources necessary for scholarly research are often held in different locations (both physically and digitally), such formats have greatly improved compatibility with a variety of databases and file systems. The introduction of these formats and their derivatives, however, is not fully exempt from the issue they are trying to solve, however. Multiplication of sub-formats and metadata fields developed for specific disciplines continue to proliferate. While we are certainly not advocating for a 'one-size fits all' approach to developing metadata standards, we believe that data formats for basic cyberinfrastructure must be sufficiently generic and flexible to enable common technical work.

3.3. Data import into (browser-based) digital research tools

For DH-focused sinologists, the tools they use are predominantly text-based. While some do use command line to code their analyses, most use some form of pre-made (browser-based) digital research tools for textual analysis. This research methodology necessitates importing (or uploading) the texts into the research tool itself, and this innocuous act of loading the text actually involves a number of thorny issues.

Our survey of various research tools available for DH research made apparent the fact that importing textual data is very complicated. Common tools such as MARKUS or Recogito provide the user with options to upload some text file or to 'cut and paste' into a textbox available on the browser. This methodology has several drawbacks, however. The scholars

²² <http://www.tei-c.org/>

²³ <https://iiif.io/>

often must manually enter metadata of the text they upload or import into the tools. This input method is extremely problematic when working with large bodies of text. In cases where the texts are copyrighted, this method is next to impossible *even when* the user has a text-mining license. To circumvent copyright protections, many scholars simply do what they must in order to complete their analyses; that is, they do so illegally. In select situations, license holders permit the use of particular research tools in a closed environment by installing local copies of the tool and the texts together, such as the premises of a particular institutional library. In any case, these are imperfect (and occasionally not strictly legal) solutions to a serious problem to DH research.

Licensing issues are complex and require stakeholders beyond researchers and librarians to address comprehensively. We recognize that new licensing models must be developed to fit the cutting-edge DH research scholars are conducting every day, even though licensing issues are not the main focus of this paper. Instead, we think from our sinology-specific situation, where the proportion of copyrighted or protected texts is relatively high and sold at a high price, to develop technical solutions that could ‘bridge the gap’ until more comprehensive licensing solutions could be implemented. Indeed, working with current licensing restrictions in a legal manner is one of the main impetuses behind RISE.

4. “RISE”: a basic cyberinfrastructure for DH research in sinology

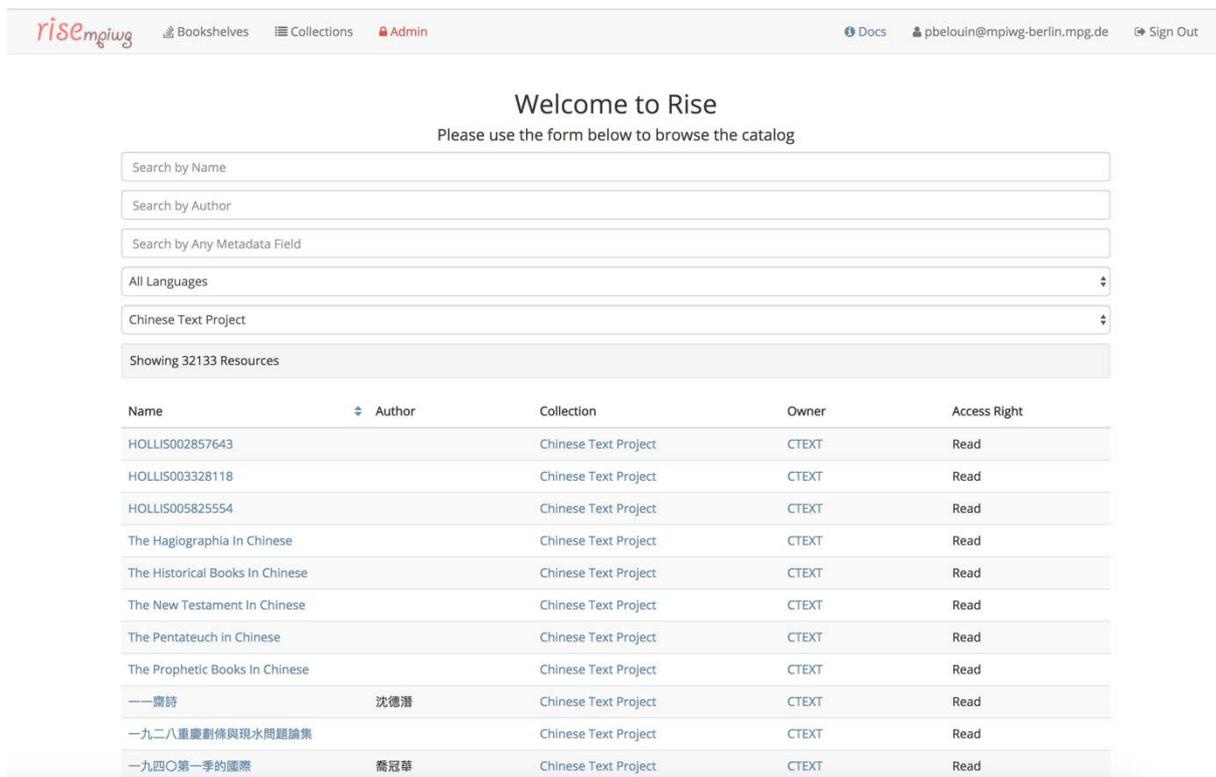


Figure 1. RISE's browser landing page

RISE is an infrastructure for DH projects (e.g., databases, tools, research platforms) to link with one another. Based on a collaborative process involving experienced stakeholders like

system developers, DH researchers in sinology, and librarians, our goal is to produce a general, reusable APIs that can cover common DH projects activities, such as log-in mechanism, contents discovery, tools discovery, contents and tools matching, and personal online research workspace (linked to researchers' individual storage). While there is a front-end web user interface that we have developed in-house, it is not essential for the API-linked ecosystem to function. This loosely-coupled infrastructural design and its flexible topologies are RISE's distinguishing feature from other large-scale, centralized research infrastructures. RISE's main instance is a codebase built on Ruby on Rails, and Figure 2 describes its primary components.

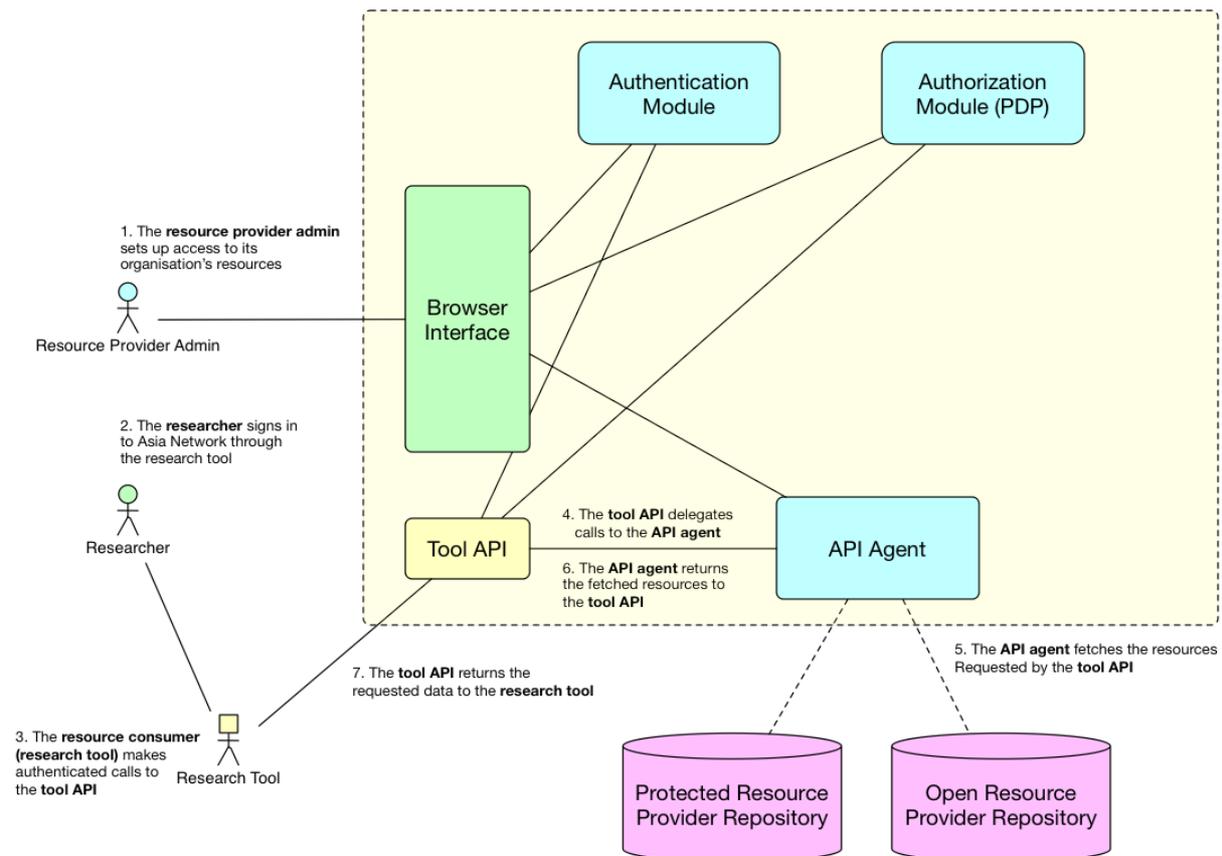


Figure 2. RISE's architecture overview

Despite its name, the RISE software can handle resources in all languages. Many projects and infrastructures have proposed similar ideas (including many European Commission-funded e-infrastructures), creating complex new initiatives like CLARIN and DARIAH. Ours, by comparison, is a modular solution that works, adapting and growing with research projects. Research and structural design remain intimately connected. This demonstrates the significant returns from our early investment into DH research in sinology.

The flexible topologies based on APIs enable diverse DH tools and contents development because they allow decentralized, role-based collaborative growth; said more simply, as long as individual stakeholders implement common API standards, everyone can just focus on their specific tasks knowing that results will be interoperable. This allows each DH project to focus on its own critical, unique contributions. At the same time, individual researchers can

still speed up their research using their existing research tools to search across multiple databases without necessarily going to a centralized portal. This workflow based on APIs apply to DH projects based on computational (rather than text-based) methodologies as well. The RISE APIs are designed to be flexible for different topologies (*ad hoc*, centralized, federated/distributed). In all cases, RISE maintains critical activities (i.e., audit, security check, transaction monitoring and encryption) for interactions among licensed materials and research tools.

The current API definition was designed to connect resource providers and tools developers. It requires resource provider to implement a minimum number of API endpoints in order to make their resources indexable and reachable through the RISE middleware. Resources can be protected by requiring the API client to provide a RISE-API-TOKEN authorization header. This allows resource providers to limit and monitor access to protected resources on a per-affiliation basis. The API allowing research tools to connect to the resources provided by the RISE middleware is similar to the resource provider API, but provides a number of extra features. These two sets of API endpoints adhere loosely to the REST standard and follows the data model presented in the following parts of this document. Below are a few examples of RISE API endpoints, and those who are interested in implementing our standard can find the full list online.²⁴

²⁴ See all three tabs at https://asia-network.mpiwg-berlin.mpg.de/pages/doc_for_resource_providers for our current API definition.

[root]/collections/ Lists the collections available to the client	Lists the sections that belong to a particular resource
[root]/collections/[collection_uuid]/resources Lists the resources that belong to a particular collection	[root]/resources/[resource_uuid]/metadata Returns the metadata for a particular resource
[root]/resources/[resource_uuid]/sections	[root]/sections/[section_uuid]/content_units Lists the content units that belong to a particular section

In an *ad hoc* topology, resource provider and tool developer must directly interact with each other's APIs. However, in most general cases (centralized and federated/distributed), RISE functions as a hub to maintain the most up-to-date API standards and to facilitate interactions among hooked-up resources and tools. Audit and other authorization actions are done via RISE's web interface as well. This infrastructure enables speedy back-end integration among existing DH resources and tools and does not reinvent the wheel at a large scale. It also enables researchers to freely manipulate and analyze resources they have access to in different research tools without violating licensing terms.

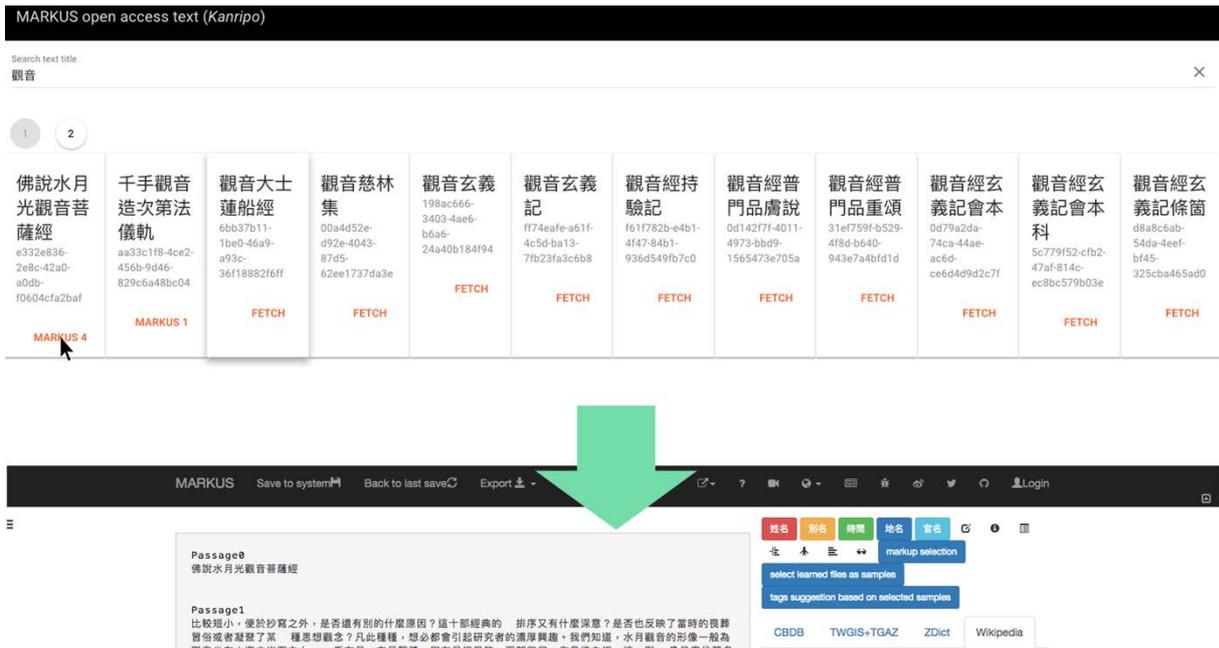


Figure 3. An example of direct link between resources and tools

4.1. Main concepts

To successfully design and implement the RISE software, we defined the following main concepts within the domain following requirements elicitation. Here we list their technical definitions by group. Furthermore, these main concepts are illustrated in Figure 4.

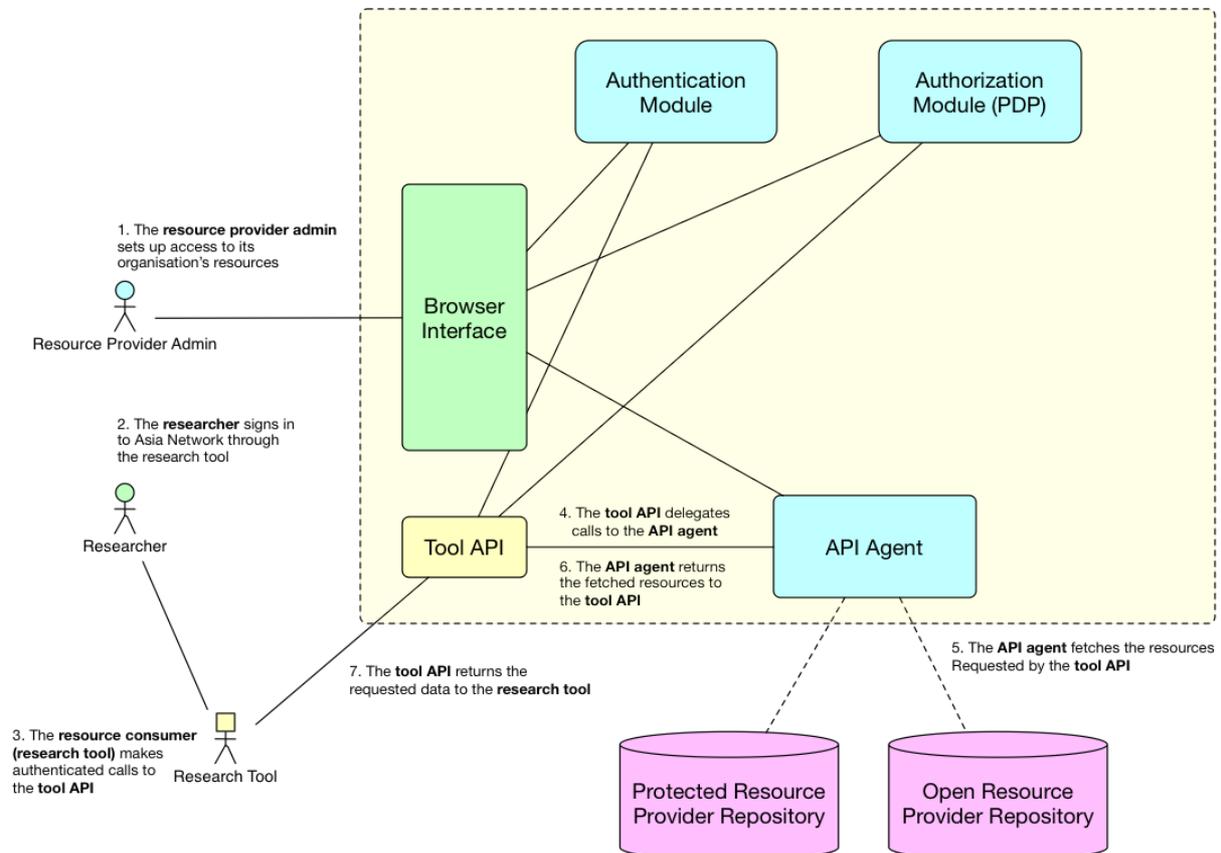


Figure 4. RISE's main concepts

4.1.1. Users

A user is a physical person who interacts with the RISE architecture, either through the browser-based interface or through the use of a research tool (in the case of the ‘researcher’ role). Authenticated users can manage their accounts via the browser interface and access resources according to the access rights defined in RISE’s authorization scheme. There are three possible user roles within RISE.

4.1.1.1. Researcher

Researchers are the default user role in the RISE architecture. Researchers can access the resources they have access to through RISE’s authorization scheme. They can also use research tools made available by the research organization they belong to.

4.1.1.2. Research Organization Administrator

Research Organization Administrators of research organizations can manage the collections and resources their research organization subscribes to via the browser interface. They can also manage the pool of users affiliated with their own organization and monitor their activities.

4.1.1.3. Resource Provider Administrator

Resource Provider Administrators moderate access to the resources their organization provides access through RISE’s browser-based interface.

4.1.2. Organizations

Although using RISE as a middleware by unaffiliated and even unauthenticated users is possible, the authorization mechanism relies on the fact that users and the resources they access belong to certain organizations. Therefore, there are also three types of organizations within RISE.

4.1.2.1. Research Organization

Research organizations have affiliated users. Their administrators must configure resource access and manage affiliated users via the browser interface. Research Organization Administrators also must make sure that resource access rights are kept up to date within RISE’s authorization mechanism.

4.1.2.2. Resource Consumer

A Resource Consumer is a software entity that consumes RISE-compatible resources by calling a RISE-compatible API, such as typically a research tool. However, we foresee that certain resource consumers such as NLP parsers would also make use of the resources provided by the RISE API and generate output in a certain format, which could then be in its turn consumed by another resource consumer.

4.1.2.3. Resource Provider Organization

A Resource Provider is an organization that makes protected or open access resources available to the RISE middleware or RISE-compatible resource consumers through a set of API endpoints. The RISE middleware software is capable of digesting and converting non-RISE-API-compatible API endpoints through the use of custom API mapping modules, which are components of the RISE middleware codebase.

4.1.3. Domain objects

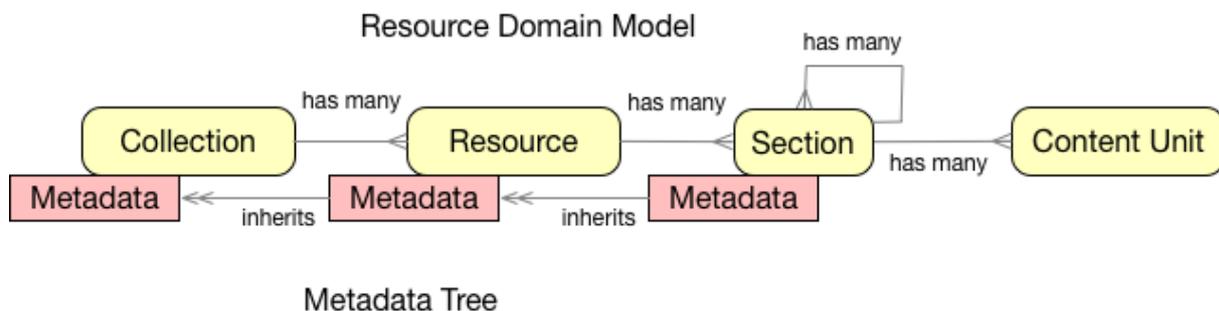


Figure 5. RISE's resource domain model

Texts and other resources accessed via RISE are modeled in a generic and flexible format to account for the heterogeneous types of resources and metadata standards provided by different resource providers. RISE's format is built in a hierarchical model and can be customized to fit different resource providers. It includes the following components below.

4.1.3.1. Collections

Collections are, as their name indicates, collections of resources. Access to these collections can be moderated by their owners on an organization-to-organization basis, so as to mirror licensing agreements made between institutions.

4.1.3.2. Resources

Resources represent items that can be accessed through the system (e.g. text, images or scanned pdfs, tables). These resources can be accessed through a unique uniform resource identifier by users according to access rights implemented in RISE's middleware. It is important to note that a particular resource (for example, a book) can be represented —and

therefore accessed— in various formats. For example, a resource’s content can be made available both in computer-readable text format or as images (page scans).

4.1.3.3. Sections

Sections are the part of the RISE resource domain model that are used to represent the hierarchical structure often found in resources such as books in the form of chapters, subchapters, etc. The meaning of different section levels for a particular resource is expressed as part of the metadata tree.

4.1.3.4. Content units

Content units are the units of text and form the base layer of our resource domain model. In practice, a content unit can for instance represent a page or a line. What content units represent for a particular resource is expressed as part of the metadata tree.

4.1.3.5. Metadata tree

Metadata is made available by resource providers through their API and is reflected in the RISE resource domain model at the collection, resource and section level. This metadata is inherited through this hierarchical model; however, if a metadata value is set both at the collection and resource levels, the resource-level metadata definition supersedes the definition set at the collection level.

4.2. Progressive software architecture and development

We developed RISE’s key features through a series of requirements elicitation with stakeholders, including DH researchers, sinologists, historians, research tool developers, and resource providers. We settled on a prototype middleware solution, and its accompanying browser interface was started in June 2017. As the development progressed, further feedback was gathered from these stakeholders and fed into the development cycle.

The heterogenous nature of resources and resource providers, combined with the fact that some providers do not have the ability nor the incentive to implement RISE’s API standard, means that our current middleware solution needs to be flexible enough to adapt to existing API endpoints of resource providers. While this may not be a long-term solution, we now provide custom API mapping modules for select providers such as CText and Perseus.

As we link with more and more resource providers and make necessary alterations to our standard to cater for an ever-growing range of heterogenous resources, we hope that both the network effect and the efficiency of our standard will create a strong enough incentive for resources providers to adopt our API standard.

We see the middleware solution currently being developed as a necessary yet temporary ‘placeholder’ to facilitate the development of RISE’s API standard. Indeed, the ultimate goal of this centralised middleware instance is to eventually vanish, replaced by a comprehensive standard allowing for the seamless integration of resource providers and resource consumers.

4.3. Authentication and access control

Many resource providers require access right management that is very granular, yet robust and simple, to handle access control to their contents. This management requires control not just at the collection level but also at the resource, and even sometimes even section, levels. In order to provide an adequate solution to this problem, our authorization scheme relies on the hierarchical inheritance of access rights across our domain model as illustrated below.

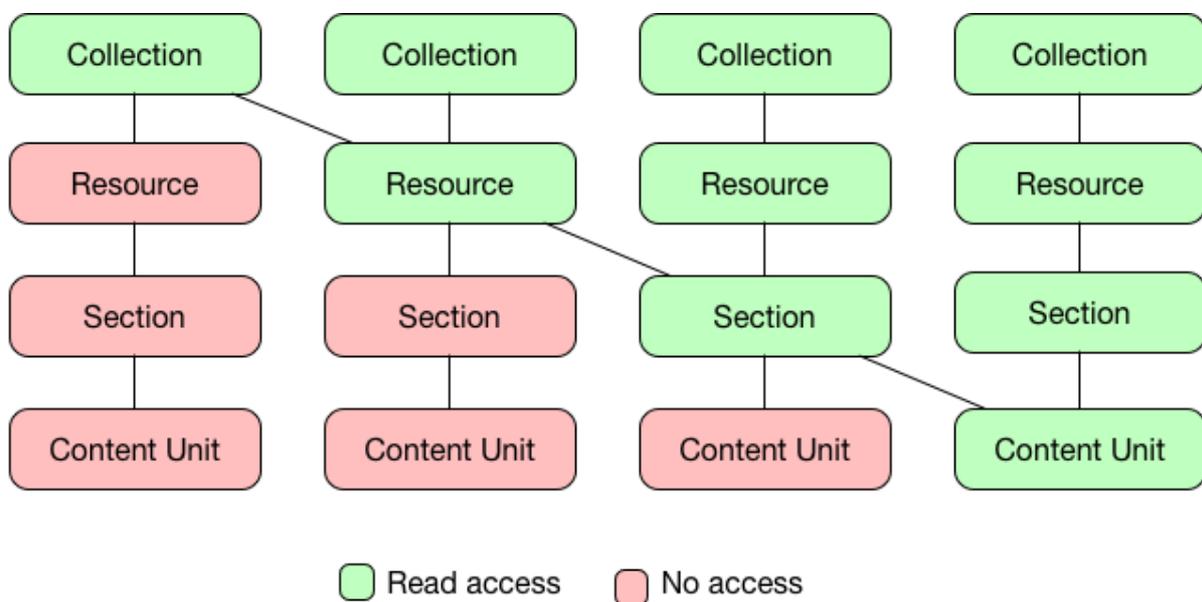


Figure 6. Granular access control

4.4. Dealing with various resource formats

In order to represent the resources linked via RISE as well as their associated metadata, we devised a multi-level data model made available as JSON objects by the middleware through RESTful API endpoints. However, some research tools prefer to access textual resources in richer formats than plain text, such as the ubiquitous TEI format and its derivatives. To cater for this issue, we also make available various formats through our resource REST API endpoints *if* the resource providers provide those formats themselves.

5. Conclusion

In RISE, we have built a middleware solution to shorten the distance between texts and research tools, provided secure linkage to facilitate digital research with licensed texts, and developed a generic yet flexible API standard for exchange between resource consumers

(often research tools) and resource owners that we hope to popularize. At the moment, RISE maintains a clearinghouse of access rights of different resources and institutions so that authentication and authorization could be done properly. In the future, we hope that adoption of RISE would gradually encourage the sinology community – especially the private vendors – to introduce novel licensing mechanisms (e.g., on-demand or consumptive text-mining) for the long-term sustainability of DH research. One of the driving ideas behind RISE is that even though such innovative licensing models do not exist yet – and the RISE team also works in parallel with our collaborators at the Staatsbibliothek zu Berlin on that – the technical mechanisms to implement these new models have now been put in place.

While open-access remains the ideal end goal in any DH endeavors whenever possible, the reality is that many digitized resources in the humanities are still sold by publishers or private vendors. In sinology especially, a fractured landscape as described above creates difficult conditions for DH scholars. We have had to navigate this complex licensing terrain during our everyday work, and RISE is now a prototype primed for transforming DH in sinology. Besides our core collaborators at Leiden University and the Staatsbibliothek zu Berlin, we have now linked up with additional collaborators in the Germany, United States, Taiwan, Japan, Mainland China, Singapore, and beyond. We look forward to launching RISE later this year, and we encourage collaborative development and constructive feedback from all those who wish to contribute to building basic cyberinfrastructure in sinology.

6. Acknowledgments

We thank Dagmar Schäfer, Matthias Kaun, and Hilde de Weerd for their ongoing work and support for this project. We also thank all of our collaborators. Finally, we thank our student assistant Vitaly Lyapunov for his invaluable work in RISE's ongoing technical development.