

The logo for xasia, with 'x' in orange and 'asia' in dark grey. The background features a complex, abstract graphic of overlapping lines and shapes in orange and grey, including a large 'x' and 'asia' rendered in a scribbled, wireframe style.

xasia

Was kommt nach der Lizenzierung: Daten zum Lesen und zum Arbeiten

CrossAsia Integrated Text Repository und Services

Ausgangslage im Kontext des FID Asien und CrossAsia

Sammlungen und die Herausforderungen der Digitalisierung

Sammlung zu und über Asien bereits seit der Gründung 1661

Ab 1951 Sondersammelgebiet Ost- und Südostasien mit klarem Verantwortungs-, Nutzungs- und Serviceprofil

2006: CrossAsia startet als Plattform, auf der alle Services gebündelt werden

Der FID Asien verantwortet Services aus und über die Regionen Ostasiens, Zentralasien und Südostasien

Die Sammlungen und Services werden von der DFG finanziell unterstützt



Was möchten wir mit der “Initiative” erreichen?

- **Verantwortung** für die Arbeit **mit** Daten erkennen und wahrnehmen
- Services, die auf Inhalte & Daten aufbauen, konzipieren, erproben und **produktiv setzen**
- **Abgestimmte Kooperationsmodelle** entwickeln
- Verantwortungen **verlässlich teilen**
- **Internationalisierung** und damit Unterstützung einer (international) vernetzten Wissenschaft



Heute?

Zugang organisieren/

Lizenzierung

und morgen ?!

Verantwortung / Daten / Werkzeuge/ Services



Gestern ?





Licensing

CrossAsia and Staatsbibliothek zu Berlin have more than twelve years of hands-on experience in licensing of electronic resources. Today the Library provides nationwide remote access to a wide range of licensed **Asia-related databases**. From the beginning, licensing had the goal to support the German-based researchers within the scope of the DFG-funded Special Research Collection Programme (SSG) and subsequently the Specialised Information Service Programme (FID). In addition, individual European consortia have already been realised. The goal is to further internationalise CrossAsia's licensing activities.

[Licensing](#)

Lizenzierung: integraler Teil der Bestandsentwicklung

Entwicklung eines eigenen und an den DFG Richtlinien orientierten Handlungsrahmens für elektronische Inhalte:

- **e-preferred Strategie**
- Zugang über die Stabi hinaus: Zugang zur überregionalen Nutzer:innengruppe aller asienbezogenen Wissenschaften in Deutschland (inkl. MWS oder MPI im Ausland)
- Metadaten mit persistenten und verlinkten Identifiers
- Archivierung - und Hostingrechte
- Text- und Dataminingrechte

Beyond Reading: Was kommt nach dem Lesen?

eine Antwort (unter mehreren möglichen):
gemäß unserer Richtlinien -

- “Verwalten” der lizenzierten Daten (Bilddaten und Textdaten)
- Daten Speicher / Inhalts Speicher
- Datensicherung für Daten und Inhalte
- Metadaten-Management
- Daten-Management
- QM, Datenkontrollen
- Einhalten der **FAIR** Prinzipien
- Entwickeln & Erproben neuer Services

- [Area Studies: China and Southeast Asia](#) (Adam Matthew Digital)
- [Area Studies: Japan](#) (Adam Matthew Digital)
- [Airiti ebooks](#)
- [Asian Studies](#) (ISEAS Publishing)
- [China America Pacific](#) (Adam Matthew Digital)
- [China Comprehensive Gazetteers : 中國綜合方誌庫](#)
- [China Trade & Politics](#) (Adam Matthew Digital)
- [The Chinese Student Monthly \(1906-1931\)](#)
- [CNKI eBooks](#)
- [道藏輯要](#)
- [敦煌史料](#)
- [Foreign Office Files China](#) (Adam Matthew Digital)
- [Foreign Office Files Japan](#) (Adam Matthew Digital)
- [古今圖書集成 : Qing Imperial Encyclopedia](#)
- [Historical Newspapers of China and South China Morning Post \(1832-1998\)](#) (ProQuest)
- [The Japan Chronicle \(1902-1940\)](#)
- [Local Gazetteers \(Erudition\)](#)
- [民國圖書數據庫 : Early Twentieth Century Book in China \(1912-1949\)](#)
- [Meiji Japan \(only metadata\)](#) (Adam Matthew Digital)
- [Missionary, Sinology, and Literary Periodicals \(1817-1949\)](#)
- [Mobilizing East Asia \(1931-1954\)](#)
- [National Palace Museum periodicals \(Airiti\)](#)
- [North China Daily News](#)
- [North China Herald](#)
- [North China Standard \(1919-1927\)](#)
- [清代史料](#)
- [Records of the Maritime Customs Service of China \(1854-1949\)](#)
- [人民日报 : People's daily](#)
- [日本古典書籍 : Classical Works of Japan](#)
- [四部備要](#)
- [四部叢刊](#)
- [四庫全書](#)
- [Ta Kung Pao 大公報 \(1902-1949\)](#)
- [續修四庫全書](#)
- [永樂大典](#)
- [正統道藏](#)
- [中國地方誌一集](#)
- [中國地方誌續集](#)
- [SBB digital : Western language Asia collection](#)
- [SBB digital : Asian language collection \(selection\)](#)
- [Fulltext search in print books \(sample set\)](#)
- [BuddhistRoad Papers \(Project\)](#)
- [The Maoist Legacy Database \(Project\)](#)

Integrated Text-Repository




CrossAsia N-gram Service

ical, linguistic and
ram datasets, i.e.

splitting the texts into fragments - for Chinese texts into fragments with one, two or three character combinations - and presenting only the frequency of the respective fragments in the corpus, they can be downloaded as unrestricted N-gram datasets.

Here you can download various datasets and explore them on your own computer and with your own tools. Currently, three datasets of Chinese text collections with N-grams per book are available, each with uni-, bi-, and trigrams. In addition, we will publish further datasets as well as some online services for analysing the N-grams, soon.

N-gram datasets

Collection	Description	N-gram
Xuxiu Siku Quanshu 續修四庫全書	'Sequel to the Siku quanshu' of the late 18th century with more than 5,000 titles.	
Local Gazetteers	Chinese geographical works from Tang dynasty to the Republican Era with about 8,000 titles.	
Daozang Jiyao 道藏輯要	'Essentials of the Daoist Canon', collection of Daoist texts with about 300 titles.	



The datasets are released under a [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](#).



Recommended citation: Please follow the suggestions provided in the repository.


Integrated Text-Repository – manage & archive licensed content in Berlin


« 1 2 3 4 5 6 7 8 9 »


List of titles


China International ... (484/838 pages)
 


China Opium and drug... (327/615 pages)
 


Imperial Maritime Cu... (253/927 pages)




Imperial Maritime Cu... (235/972 pages)



Imperial Maritime Cu... (234/798 pages)


Imperial Maritime Cu... (233/884 pages)




The Maritime Customs... (232/960 pages)


The Maritime Customs... (230/912 pages)


China Kwangtung opiu... (224/280 pages)
 

Imperial Maritime Cu... (223/664 pages)


« 1 2 3 4 5 6 7 8 9 »



opium  

About 99k pages founds in 25.4k titles in CrossAsia full-text search



China Opium and drug traffic (1917-1922).
FO 671/452, The National Archives.
Collection:Foreign Office Files China

327 of 327 pages of this title contain the search term(s) "opium"



« 1 2 3 4 5 6 7 8 9 »

p.103  



of the Company's Fleet, Dear Sirs, **OPIMUM SMUGGLING**. With a view to preventing the smuggling.... of **Opium** or ensuring that if **Opium** is on board it is discovered and handed over.... of the **Opium** and we wish every encouragement to be given to informers, whose identity must.... not be disclosed. Any **Opium** found on board is to be handed over to the Captain who will claim.... and no precautions already taken to prevent **Opium** smuggling are to be relaxed. Yours faithfully

p.104  

- 102 I. 2. **OPIMUM SMUGGLING**. **Opium** smuggling is a very bad thing to do.... everyone must try to stop it. Everyone must look out for **opium** and if he can find it he..... The Captain will claim the reward for finding **Opium** and every man on board will share.... to find the **Opium**. # 5. The rewards paid by the Customs are on the following scales..... : - CUSTOMS' **OPIMUM SEIZURE REWARDS**. 3- 4. ft per 100 Taels weight **Opium**, Raw, Foreign.... or Native **Opium**, Prepared, Foreign or Native **Opium**, Dross, Foreign or Native Morphia.... of Prepared **Opium** is found the Informer will receive Sg60-and Crew will have S960-divided

p.130  

, Seizure of **Opium** on board the .a.- -Smyang" at Snan^hai. The sample of prepared **opium**.... shewn to me was "French **Opium**" i.e. the output of the Monopoly of Indo-China, **Opium**...., not as a general .rule to compete in Hongkong with the ' **opium** of the Government Monopoly.... seizures of "French **Opium**" for 191? were 25 ,000 taels out of a total of 27,000 taels.... prepared **opium** and for the first half of the current year 18,677 taels out of a total

p.183  

any new Bye-Law as to **Opium** by embodying it in an International Regulation in **Opium** and as to the

Subject

Find filter item

- Essay (2619)
- Ostasiatica (2229)
- China (1585)

Date

Find filter item

- 1924 (440)
- 1925 (402)
- 1930 (379)

Language



- English (6276)
- Chinese (808)
- Dutch (518)

Author

Find filter item

- 著者不詳 (55)
- Rewi Alley (43)
- Institute Of Southeast

Collection

- North China Daily News (7566) 
- The North China Herald Online (4351) 
- Missionary, Sinology, And Literarv Periodicals (1817-

Integrated Text-Repository

Der CrossAsia ITR Explorer bietet eine andere Perspektive auf die CrossAsia ITR-Ressourcen und ist eine Ergänzung zur CrossAsia Volltextsuche. Der ITR Explorer ermöglicht es, Suchergebnisse zu kombinieren, sie zu vergleichen und sie zu visualisieren, indem ihre Überschneidungen oder Verteilung über die Zeit genutzt werden. Die Ergebnisse beziehen sich im Allgemeinen auf das Buch oder die Ausgabe, nur für zwei chinesische Zeitungen beziehen sie sich auf die Articlebene. Die Ressourcen sind dieselben wie in der CrossAsia-Volltextsuche; die CJK-Zeichenzuordnung ist wählbar.

CrossAsia ITR Explorer (Beta version)

1. Select source(s) ⓘ
Chinese Students Monthly (1906-1947) (selected) X

2. Enter search term or phrase ⓘ
search term or phrase Q

☑ CJK Mapping ☑ Phrase search

3. Show and combine result sets ⓘ
Result sets: X

opium X 鴉片 X
Resources (2708 hits) Resources (2284 hits)
Combine with ... Combine with ...

4. Visualize your result sets ⓘ
VIEW (OVERLAP) STACKED LINE CHART (TIME RANGE)
Show sets over time, define time range. Please select sets for visualization. X

From 1700-JAN to 1979-JAN ⚙

opium (2708)

Legend: opium (2708)

Timeline: 1700 1800 1900 2000 1979

Und wie unterstützen wir digital scholarship:

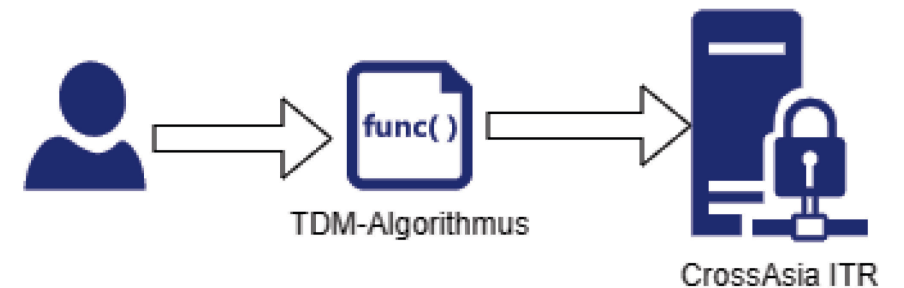
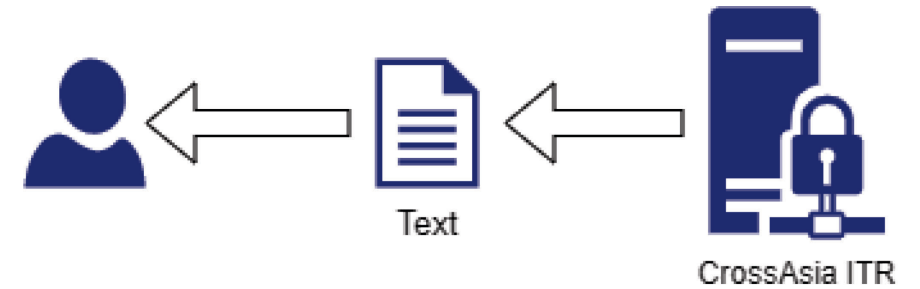
ein explorativer und wahrscheinlich auch innovativer Proof-of-Concept:

Einsatz von Compute-to-Data innerhalb von **Gaia-X**

Im Rahmen vom FID Asien und CrossAsia verwaltete Ressourcen als Datenobjekte im ITR (derzeit ca. +75 Mio. überwiegend in Englisch und Chinesisch)

Aktuelle Dienste:

- ITR-Volltextsuche, ITR-Explorer, TDM-Rechte vorhanden
- Compute-To-Data: Algorithmus geht zu den Daten (nicht andersherum!)
- Ausführung des Algorithmus in einer geschützten Umgebung Die Daten selbst werden nicht weitergegeben
- Nur die Ergebnisse des Algorithmus stehen zum Download bereit Zusammenführung über die Datendrehscheibe



WERBEBLOCK

Hands-on Lab
Konferenzraum 15



111.BIBLIO **CON**

Vorwärts nach weit

23. - 26.05.2023 | Hannover

Donnerstag, 25. Mai, 16:30 - 18:00

Einstieg in die Gaia-X-Infrastruktur

Mit Gaia-X entsteht eine neue europäische Cloud-Infrastruktur, die sich maßgeblich von bestehenden Cloudanbietern wie Google, Dropbox oder Amazon AWS unterscheidet. Sie ist im Kern dezentral und offen angelegt, so dass Abhängigkeit von einem Anbieter vermieden wird und sich grundsätzlich jede Institution an dem Netzwerk beteiligen kann. Die Nutzung neuer Technologien wie Decentralised Identifier (DID) oder dem Ocean Protocol eröffnet auch Bibliotheken und anderen Wissenschaftsinfrastruktureinrichtungen vollkommen neue Wege, digitale Inhalte für die Digital Humanities wie auch für das Machine Learning bereitzustellen. Die Staatsbibliothek zu Berlin hat in einem Proof-of-Concept im Rahmen des Integrierten Textrepositoriums (ITR) von CrossAsia erste Erfahrungen mit dem Gaia-X-Netzwerk sammeln können, wie im begleitenden Vortrag "Was kommt nach Lizenzierung: Daten zum Lesen und Arbeiten" dargestellt wird.

In dem 90-minütigen Hands-On-Lab sollen zunächst die Grundlagen für die Nutzung des Gaia-X-Netzwerks vermittelt und die unterschiedlichen Zugriffswege für Download und Compute-To-Data erläutert werden. Danach sollen die Teilnehmenden anhand praktischer Beispiele selbst auf Daten im Gaia-X-Testnetzwerk zugreifen und dort Inhalte veröffentlichen, sowohl für den Download als auch für Compute-To-Data.

Notwendig für die Teilnahme sind ein Laptop mit einem aktuellen Browser.

[Sitzungsleiter:in: Gerrit Gragert \(Staatsbibliothek zu Berlin, Berlin\)](#)

Heute & Morgen?

Bei allen Projekten ist der Bezug zu unseren Sammlungen und Services entscheidend

Unsere Erfahrung:

Zu den klassischen und bisherigen Aufgaben wie Buch- und Medienauswahl und -bearbeitung kommen neue Aufgaben im Bereich des Daten- und Content-Managements hinzu

Skalierung ist dann Herausforderung

Grundsatz:

in einem **vertrauenswürdigen internationalen Netzwerk** einigen sich die Partner auf gemeinsame Datenstandards und auf kontrollierte Zugriffsmodi

teilen sich die Partner Aufgaben,

tauschen Ergebnisse untereinander aus, um Doppelarbeit zu vermeiden und

unterstützen (durch neue **Betriebsmodelle**) die digitale Wissenschaft über ihre eigene Institution hinaus

... auf einen Blick:

CrossAsia

<https://crossasia.org/>

CrossAsia N-Gram Service

<https://crossasia.org/service/crossasia-lab/crossasia-n-gram-service/>

Volltext Suche (ITR)

<https://crossasia.org/service/crossasia-lab/crossasia-fulltext-search/>

ITR Explorer

<https://crossasia.org/service/crossasia-lab/crossasia-ittr-explorer/>

Vielen Dank!

Matthias Kaun

Staatsbibliothek zu Berlin

Ostasienabteilung

matthias.kaun@sbb.spk-berlin.de

<http://staatsbibliothek-berlin.de/>

<https://crossasia.org/>