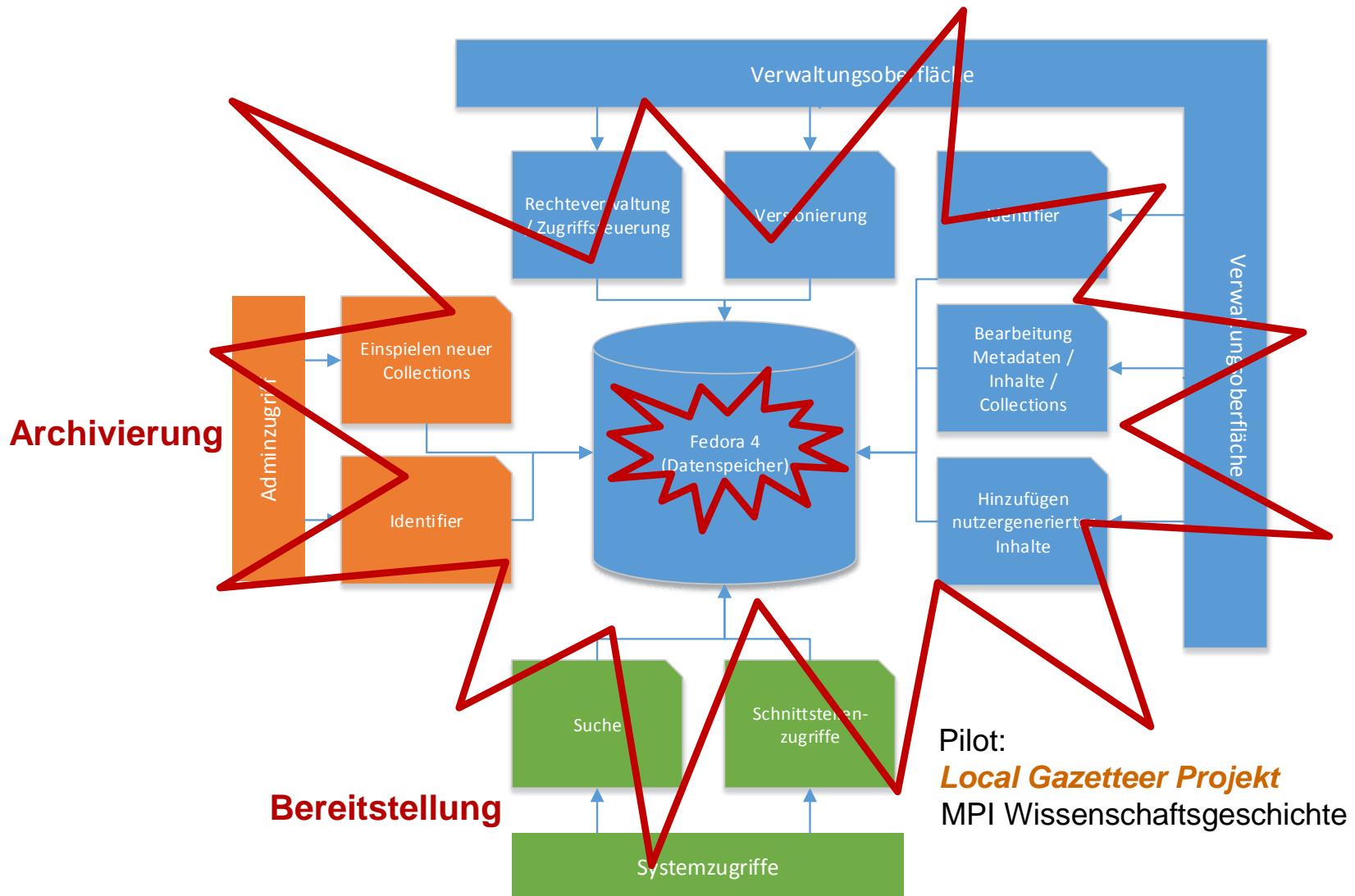


Fundament für die Forschung von morgen

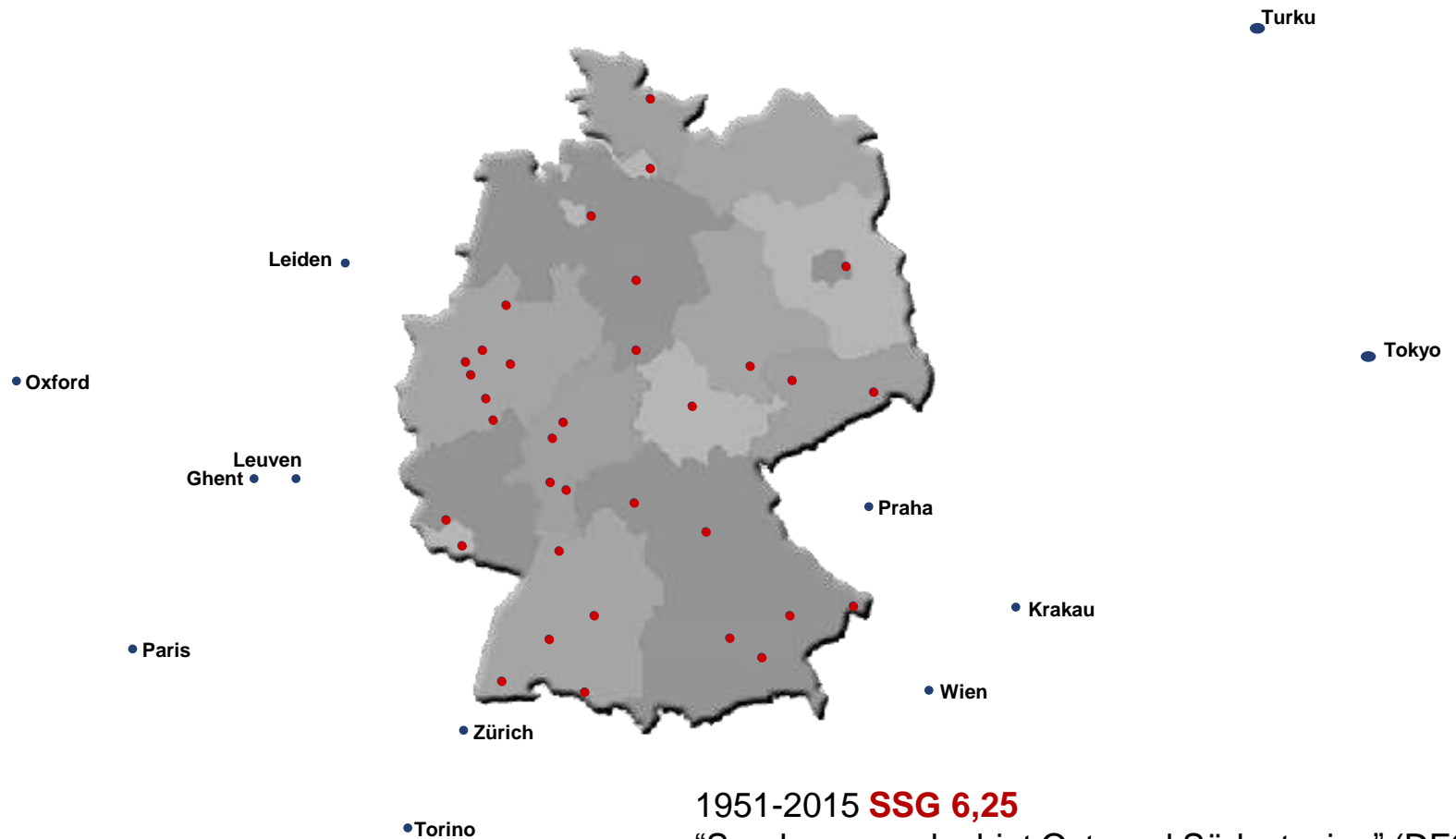
CrossAsia Integriertes Textrepositorium (ITR) zur
Archivierung und Bereitstellung von Objekten und Daten
für die Asienwissenschaften

Dr. Martina Siebert
Staatsbibliothek zu Berlin – Ostasienabteilung



CrossAsia 2005-
Virtuelle Fachbibliothek **CrossAsia**

Blauer Leihverkehr



1951-2015 **SSG 6,25**
“Sondersammelgebiet Ost- und Südostasien” (DFG)



CROSS

ASIA

 Deutsch

 Print

Home

About CrossAsia

Special collection
East/South-East Asia

Special Collection Altaic
Languages

Special Collection Non-
Conventional Material

Metasearch

Catalogues

Access Databases

Online Guide East Asia

Address

Impressum

CrossAsia

The Virtual Library East and Southeast Asia - **CrossAsia** - offers parallel access to scientific information in East and Southeast Asian studies. This includes the following:

- **CrossAsia** allows for a [simultaneous query](#) within a number of library catalogues. Some of these catalogues offer access to full text databases, others are combined with an order function.
- **CrossAsia** offers [access to bibliographic and full text databases](#). Many of them are free but some are subject to license. In the latter case, [registration](#) with the State Library Berlin is required.
- **CrossAsia** offers direct access both to the [Online Contents Database](#) presently evaluating some three hundred journals and the Electronic Journals Library which sometimes allows for access to the full text.
- The [Online Guide East Asia](#) (OGEA) will direct you to internet resources which are annotated and evaluated.

Please click the navigation bar at the left for further information about retrieval options and the sources utilized by **CrossAsia**.

The virtual special-topic-library is developed by the [East Asia Department](#) of the [Staatsbibliothek zu Berlin](#) and its [project partners](#) and is financially supported by the [German Research Foundation](#) (DFG).

Login

Name

Passwort

CrossAsia

Staatsbibliothek zu Berlin
Ostasienabteilung
Potsdamer Strasse 33
10785 Berlin

Tel.: +49-30-266 2402

Fax: +49-30-266 3613

[Email](#)

The grid contains the following flags and labels (from top-left to bottom-right):

- 上同 (Shōdō)
- 商通度印東 (Shōtōdo Inpō)
- 船号一 (Funagō Ichibu)
- 備白 (Bihaku)
- ガリュビムジエ (Galyubimujie)
- 旗小上同 (Hatai Kōshōdō)
- 船号二 (Funagō Nibu)
- 備音 (Bihin)
- 陣軍 (Jinjun)
- ジドニコ (Jidoniko)
- 旗用常 (Hatai Yōjō)
- 組手取 (Kumete)
- 商通度印東 (Shōtōdo Inpō)
- 船量測 (Funaryōso)
- 旗用常非 (Hatai Yōjō Hi)
- 船令使 (Funayōmei)

A search bar with the text 'xasia Suche' is overlaid on the grid.

xasia

Zur Anmeldung wählen Sie **CrossAsia.org (VHO)**. Sollten Sie ihr Konto bereits mit den Nutzerdaten Ihrer Heimatinstitution verknüpft¹ haben, so wählen Sie diese Institution bitte über das Auswahlfeld aus.

Vorherige Auswahl:
 CrossAsia.org (VHO)

Wählen Sie Ihre Institution aus...

Wählen Sie Ihre Institution oder suchen Sie nach ihr

- CrossAsia Registrierung
- Account verknüpfen¹
- Hilfe zu Login und Account-Verknüpfung

¹ Verknüpfen Sie Ihr CrossAsia-Konto mit den Nutzerdaten Ihrer Heimatinstitution. Dies ermöglicht die Anmeldung an CrossAsia.org über Ihre Heimatinstitution. Ist ihr CrossAsia-Konto noch nicht verknüpft, so wählen Sie bitte "CrossAsia.org (VHO)" und verwenden Ihre bisherigen Anmeldedaten.

Sicherheit: Wir empfehlen alle Browserfenster nach dem Besuch unserer Seiten zu schließen und nicht für weiteres Surfen im Internet zu verwenden. Dieser Hinweis gilt insbesondere dann, wenn Sie unsere Seiten von einem öffentlichen Ort aus nutzen (z.B. PC-Pool, Bibliothek oder Internet-Café).

Aktuelle Nachrichten aus
China | Japan | Korea | Taiwan | Asien

Blog
Brill Chinese-English Dictionary Online

Forum
Apabi- Lexikon fragt nach login Daten

CrossAsia Datenbanken
Zugang zu den lizenzierten Datenbanken

CrossAsia 2005-
Virtuelle Fachbibliothek **CrossAsia**

CrossAsia – Fachinformationsdienst Asien

Projektlaufzeit: 1.2016 bis 12.2018



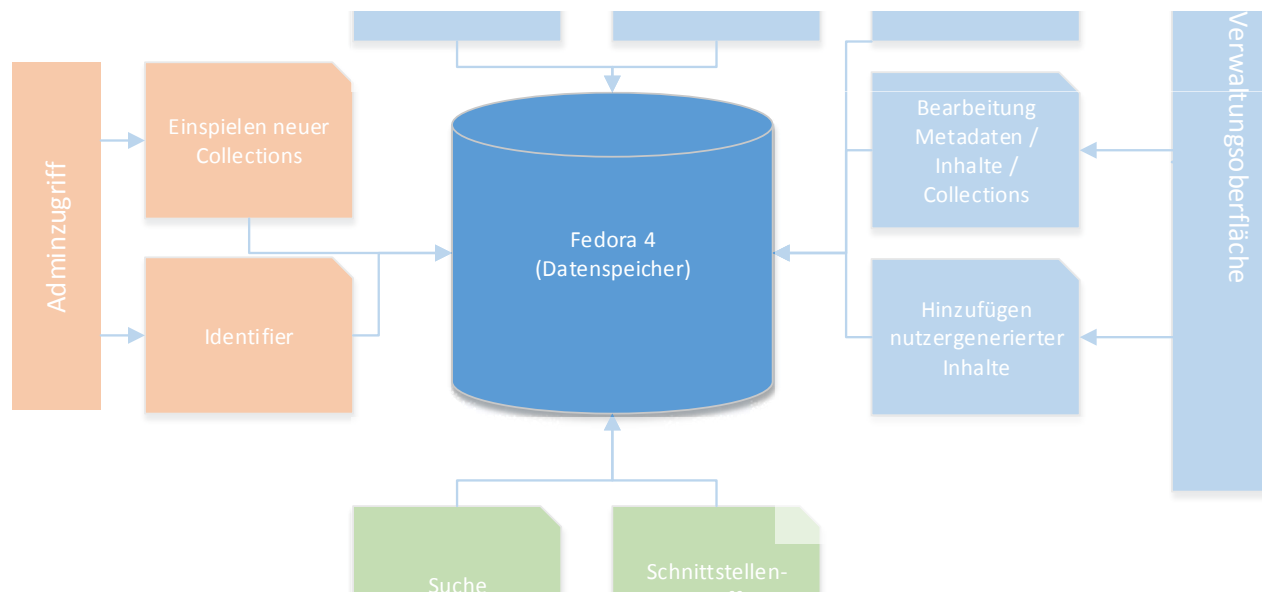
Content Management: Forschung und Entwicklung (Aktionsfeld 4)

- Infrastruktur für Metadaten und Volltexte (ITR)
Archivierung und **Bereitstellung** von digitalen Objekten und Daten (für und aus der Forschung) über Webseiten und definierte Schnittstellen

> Vorbereitung auf zukünftige Anforderungen im Kontext Digitaler Geistes- und Sozialwissenschaften

Archivierung ...

- > ... lizenzierter Materialien, für die Rechte für Hosting und Indexierung sowie u.U. weitere Rechte bestehen, die im Rahmen der Digitalen Wissenschaften erforderlich sind (text mining, caching = „non-consumptive use“ etc.)



- Akquise bzw. Extraktion der Content- und Metadaten
- Entwicklung von Ingest-Routinen und Metadaten-Mappings

IMG <
Metadaten <
Projektarbeit <

Bildnachlässe

Diaolong databases

- 續修 SKQS (5412 Buchtitel / ca. 3 Mio Seiten (**Metadaten/img/XML**))
- Daozang jiyao, Qingdai shiliao, SKQS (weitere rd. 5000 Titel)
- Weitere Diaolong Ressourcen

Brill eBooks
Airiti eBooks

- Asian Studies eBook 2007-2017 (ca. 11 T Titel)
- Liz. Airiti Titel
> **Metadaten/PDF mit Volltext**)

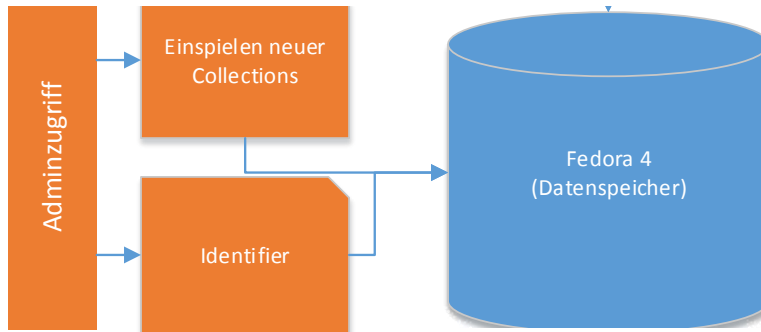
Local Gazetteers
地方志

- 2000 Titel
(195 T Kapitel,
2,5 Mio Seiten)
> **Metadaten/XML**

pro
batch

Volltext von
Printbüchern

SCAN > IMG > OCR



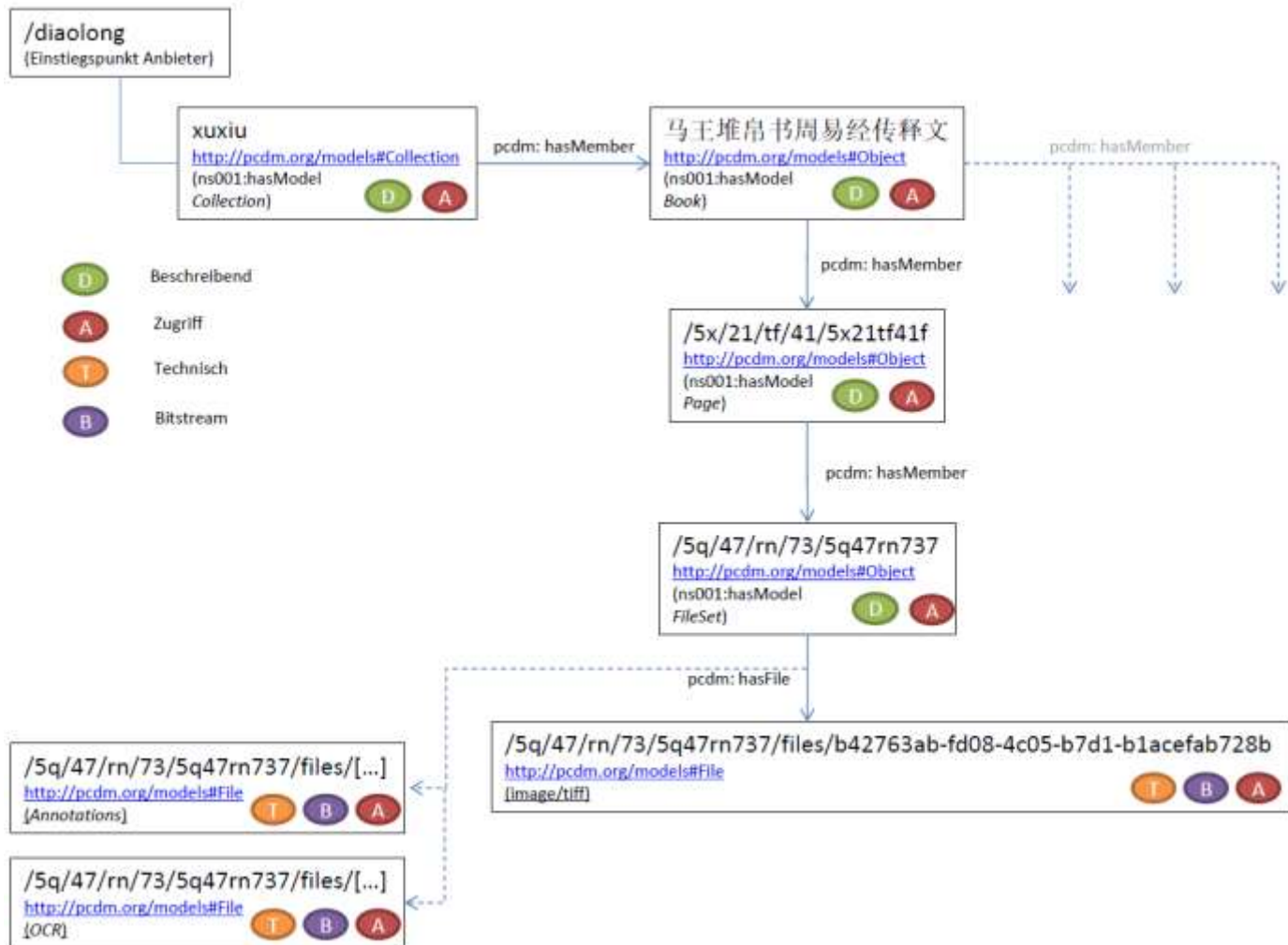
People's Daily 人民日报
(1946-2012)

- ca. 2 Mio Artikel (**TXT**)

Adam Matthews Digital

- 12.500 bibliographische Einheiten / **XML + 1,3 Mio Img**
(Foreign Office Files China, Meiji Japan, China Trade & Culture, China, America, Pacific)

Fedora-Struktur: Beispieldatenmodell (Diaolong)



```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<dataExport name="AD">
```

```
- <document nodeId="1250374">
```

```
<ImageDirectory>115149</ImageDirectory>
<Title>Supply of aircraft to China</Title>
<Collection>Foreign Office Files for China, 1949-1956</Collection>
<TopLevelCollection>Foreign Office Files for China, 1919-1980</TopLevelCollection>
<Reference>FO 371/115149</Reference>
<Date>1955</Date>
<SearchableDate>1955</SearchableDate>
<Description/>
<Notes/>
<DepartmentCode>FO</DepartmentCode>
<Office>Foreign Office</Office>
<Countries>China</Countries>
<Region>East Asia</Region>
<Places>Beijing; Burma; East Asia; Guangzhou; Hong Kong; Shanghai; United Kingdom; United States</Places>
<People>Addis, Sir John; Zhou Enlai</People>
<Topics>aircraft; embassy; industry; propaganda; trade</Topics>
<Organisations/>
<Library>The National Archives</Library>
<Copyright>Crown Copyright documents © are reproduced by permission of The National Archives London, UK</Copyright>
<Url>http://www.archivesdirect.amdigital.co.uk/Documents/Details/115149</Url>
```

Struktur 1: Buch und Seiten

Datenlieferung XML und Images

- Extraktion
Titel-Metadaten

```
- <images>
```

```
- <image imageFile="035846.jpg" id="1">
```

```
<imageUrl>http://www.archivesdirect.amdigital.co.uk/Documents/Images/115149/0</imageUrl>
```

```
- <imageText>
```

```
<![CDATA[FO 371/115149 Supply of aircraft to China (1955)]]>
```

```
</imageText>
```

```
</image>
```

```
- <image imageFile="035847.jpg" id="2">
```

```
<imageUrl>http://www.archivesdirect.amdigital.co.uk/Documents/Images/115149/1</imageUrl>
```

```
- <imageText>
```

```
<![CDATA[115149]]>
```

```
</imageText>
```

```
</image>
```

```
- <image imageFile="035848.jpg" id="3">
```

```
<imageUrl>http://www.archivesdirect.amdigital.co.uk/Documents/Images/115149/2</imageUrl>
```

```
- <imageText>
```

```
<![CDATA[mmmm n n n n p u u u C FAR EASTERN DEPARTMENT K FROM teHihv <1/ F ^#f>oavn ov Dated sjo /Tfu^U ^ /4qjX. ^WO A6iiv &, ^
001^4 Received in /\ Registry- ^jUKL U References to former relevant papers fCU /I . MINUTES Vm/" L5u[6 . - (Print) 115149 (How disposed of) U-***
cu* t-vw- ,v*n H. .&*b . ^,'Tcp/ rt u<, /vLJU ^ ^ ^ (Action completed) ^il7 d (Index) ,IM^( 1 1 Vs? References to later relevant papas TW r<<<
iw-U« toj %P uLJ$*Kr VIAf fhht AdKz (4*Ot/Js tr*n\S. m^m/UI Jlr J* **L fa MtJZ*/ /n, 1* . *** A^o CjuU***** w 44875]]>
```

```
</imageText>
```

```
</image>
```

```
- <image imageFile="035849.jpg" id="4">
```

```
<imageUrl>http://www.archivesdirect.amdigital.co.uk/Documents/Images/115149/3</imageUrl>
```

```
- <imageText>
```

```
<![CDATA[y * . y y " Arvx*-^*^*^* ^ »?»>»*«»« tIvV ZtUvUCy («..t.i) H,* r.tcn . A .A. Number of copies 3 Number-of -copies' of enclosures .T3V..
```

have, been trying to Wo ok down thmk that » ia me ^ unlikely that the ai
imported from the U.S.A. since 1950; the U.S. security export controls are
enforced and we do not believe that the aircraft could have slipped through
authorities being aware of]#». it. Nevertheless it would be useful if you co
about their date of C%iK<n«*k«>L manufacture/ In this connexion we ar
the Convairs owned by the Nationalists were flown^to Hong Kong in 1949
wao humdul uver to the * i athnri ttt*? r " 3. We have also considered the
these Convairs might have been mistaken for IL 1 2j]. Convairs apparentl
these aircraft. Perhaps you could bear this in mind if you have any further
investigate the existence of these alleged Convairs. Yours ever, FAR EAST

</imageText>

:/image>

→ image imageFile="035860.jpg" id="15">

→ <imageUrl>http://www.archivesdirect.amdigital.co.uk/Documents/In
</imageText>

<![CDATA[VU i FILS (SO 1386/3) VOBZGHf OmOE, 3«Vi. . QepteIbBey 5
Ohnnoopy, Please refer to your telegram Mb. 312 of August 27 about civil
2. We were very interested to hear that the* Indian Cultural Delegation h
by Convsire. We have been trying to discover how the Chinese have these
We think that it la unlikely that the aircraft can have been imported from
since 1950 1 the United States security export controls are very strictly er
not believe that the aircraft could have slipped through without the United
being aware of it, nevertheless It would be useful if you could find anythin
manufacture or acquisition by the Chinese Communists, In this connexion
that all the Convairs owned by the nationalists were flown out to Hong Ko
have also considered the possibility that these Convniro might hove been
12*3, Oonvniro apparently look very similar to these aircraft, Poxtiape yo
mind if you have any further opport unity to Investigate the existence of t
Convairs, Yours ever, PAR KAFITBHN DEPARTMENT, The Chancery, British
X]]>

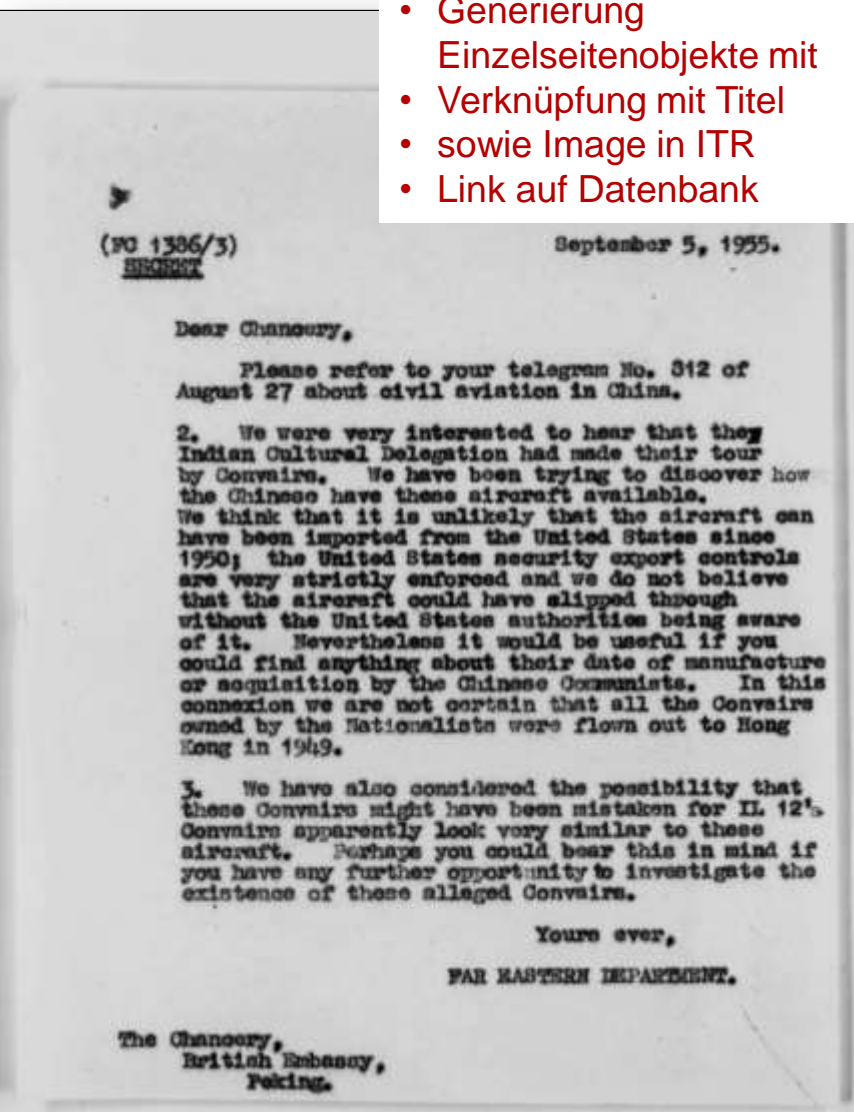
</imageText>

:/image>

image imageFile="035861.jpg" id="16">

<imageUrl>http://www.archivesdirect.amdigital.co.uk/Documents/In
</imageText>

<![CDATA[n n n u u FAR EASTERN DEPARTMENT CHINA FROM jU. . Qct
n /* \$ Dated Received in Registry- ^ i References to former relevant pape
disposed of) (Action , , x completed) V*®" Va- References to later relevant papers u 1



- Generierung Einzelseitenobjekte mit
- Verknüpfung mit Titel
- sowie Image in ITR
- Link auf Datenbank

Struktur 1: Buch und Seiten

```
{
  "id": "125_9789864370139",
  "book_id": "9789864370139",
  "position_isi": 125,
  "text_tcsi": "第六講 性情詞 107 /",
  "has_model_ssim": "Page",
  "_version_": "1585208763214725120",
  "timestamp": "2017-11-27T08:54:43.",
  "score": 0.58736765,
}
```

Seite

vom Volltexttreffer
> zu den Buch-Metadaten
und (authentifiziert)
> zur Seite in der Datenbank

Buch

```
"system_modified_dtsi": "2017-08-08T10:49:57Z",
"has_model_ssim": "Book",
"id": "2z10ws213",
"accessControl_ssim": ["997ed1f9-674a-4ed9-a6fb-8956fbfc5fc3"],
"depositor_ssim": ["x-asia@sbb.spk-berlin.de"],
"depositor_tesim": ["x-asia@sbb.spk-berlin.de"],
"title_tesim": ["唐宋詞名篇講演錄"],
"creator_tesim": ["王兆鵬"],
"description_tesim": ["走進《唐宋詞名篇講演錄》的明朗春光時，你沐浴的是唐時風；宋時雨。《唐"],
"publisher_tesim": ["Airiti Press Inc."],
"subject_tesim": ["語言學",
  "文學 - 中國文學",
  "type=\中國圖書分類法\" 820.9104"],
"language_tesim": ["chi"],
"identifier_tesim": ["http://www.airitibook",
  "type=\ISBN\" 9789864370139",
  "type=\CrossAsia Link\" http://erf.sbb.",
"col_id_tesim": ["airiti_ebooks_licenced"],
"title_transcription_tesim": ["tang song ci",
  "唐宋詞名篇講演錄"],
"creator_transcription_tesim": ["wang zhao",
  "王兆鵬"],
"publication_place_tesim": ["[s.l.]"],
"issued_isim": [2015],
"genre_tesim": ["book"],
"physical_description_tesim": ["電子文獻",
  "electronic"],
"isbn_tesim": ["9789864370139"],
"keywords_tesim": ["中國古典文學"],
"thumbnail_path_ss": "/assets/work-ff055336041c3f7d310ad69109eda4a887b16ec50"]
```

Datenlieferung
Metadaten und
Volltext-PDF

- Metadaten transformieren
- Text seitenweise extrahieren

第六講 性情詞

這節課我們講性情詞。

唐宋詞作為一種抒情文學，抒發的不僅僅是愛情、兒女私情，也經常表現詞人豐富多彩的性情。性情，是指人格個性和人生態度。不同的人格個性，在詞中會展現出不同的藝術風貌。從共時性的角度來看，詞表現出了各式各樣的性情。共時性，就是我們通常說的橫向地看，是把唐宋詞放在同一時間進行整體觀照。但從歷時性的角度看——歷時性就是縱向地考察，唐宋詞表現愛情和性情，又有一個變化過程。唐五代和宋初的詞，主要是表現愛情、兒女私情、私密化的情感。到蘇軾以後，才注重表現創作主體的個性、懷抱，由歌唱比較單一、狹窄的愛情轉變為表現更豐富闊大的精神世界。詞中的藝術世界和心靈世界，到蘇軾手中才變得日益豐富、日益複雜。我們這裡不是講詞史，所以只能簡單地提示一下。當然，在蘇軾之前，詞表現性

Struktur 2: Buch, Kapitel, Seiten

```
"id": "1552584page",
"page_id": 1552584,
"hasModel": "Page",
"book_id": 35403,
"position": 495,
"text": ["張一龍一五袴廂人任黎平府推官"],
"chapter_id": [110490,
110491,
110492],
"_version_": 1582405972952875008},
```

Seite

```
"id": "35403book",
"hasModel": "Book",
"book_id": 35403,
"title": ["藤縣志 22卷"],
"author": ["(清)逸其晉",
(清)胡毓璠"],
"title_transcription": ["teng xian zhi",
藤縣志"],
"creator_transcription": ["bian qi jin",
hu yu fan",
邊其晉",
胡毓璠"],
"medium": ["电子文献",
electronic"],
"issued": 1908,
"date": 1875,
"edition": "清光緒三十四年鉛印本",
"temporal": "清 同治 (1911-1875)",
"admin_level_1": "廣西",
"admin_level_2": "梧州",
```

Buch

```
"latitude": 23.37312,
"longitude": 110.90428,
"language": "chi",
"url": "http://erf.sbb.spk-berlin.de/han",
"responsibility": ["(清)逸其晉修 (清)胡毓璠"],
"_version_": 1582222500133076992}]
```

Kapitel

```
"id": "110465chapter",
"chapter_id": [110465],
"title": ["卷之十"],
"hasModel": "Chapter",
"book_id": 35403,
"pageStart": 303,
"pageEnd": 325,
"value": 1,
"_version_": 15822224941985562624},

"id": "110466cha",
"chapter_id": [110466],
"title": ["軍政志"],
"hasModel": "Chapter",
"book_id": 35403,
"pageStart": 303,
"pageEnd": 325,
"value": 2,
"_version_": 15822224941985562624},
```

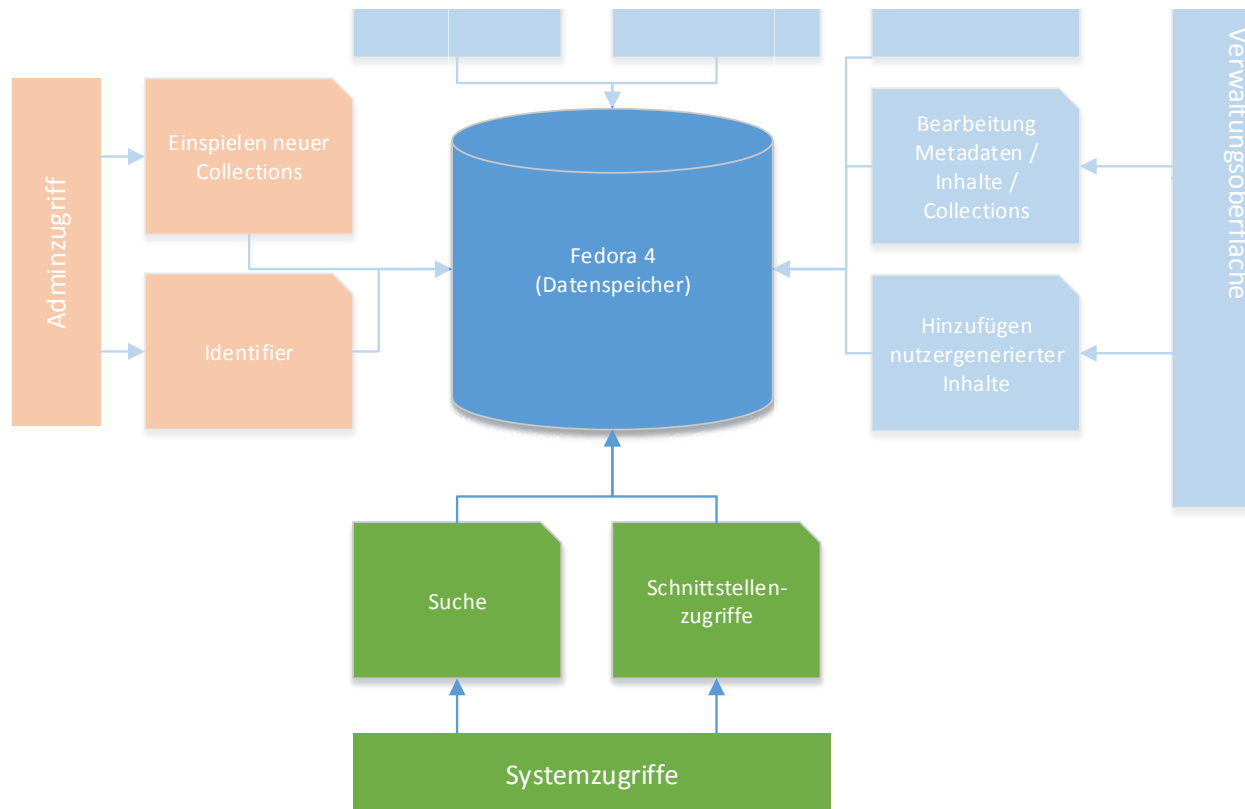
```
"id": "110467chapter",
"chapter_id": [110467],
"title": ["兵制"],
"hasModel": "Chapter",
"book_id": 35403,
"pageStart": 303,
"pageEnd": 307,
"value": 3,
"_version_": 15822224942002339840},
```

Extraktion aus Datenbank Metadaten und Text

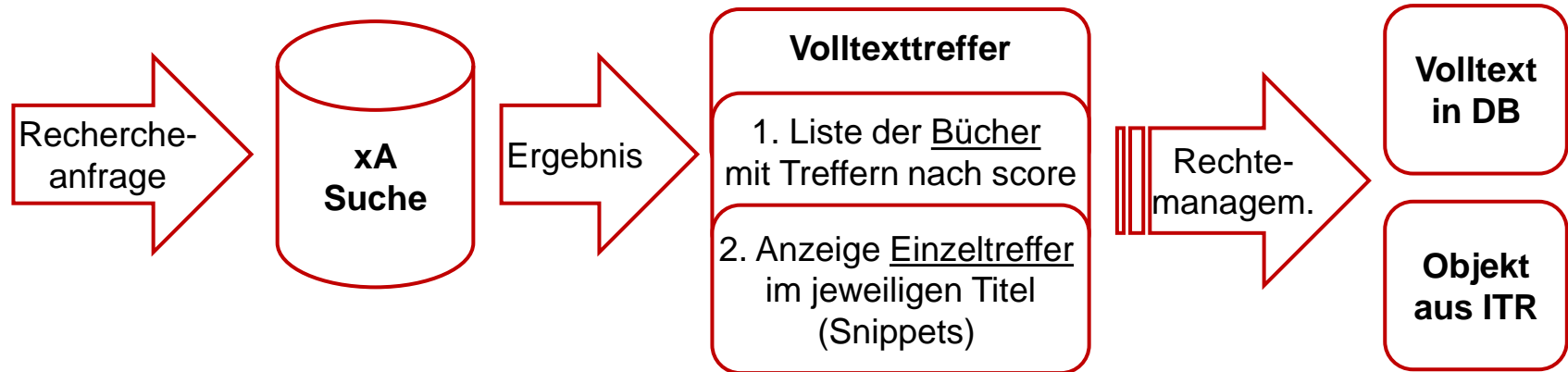
- Metadaten transformieren
 - Text seitenweise extrahieren,
 - Kapitel identifizieren
- (letzteres Projektarbeit MPIWG)

Bereitstellung ...

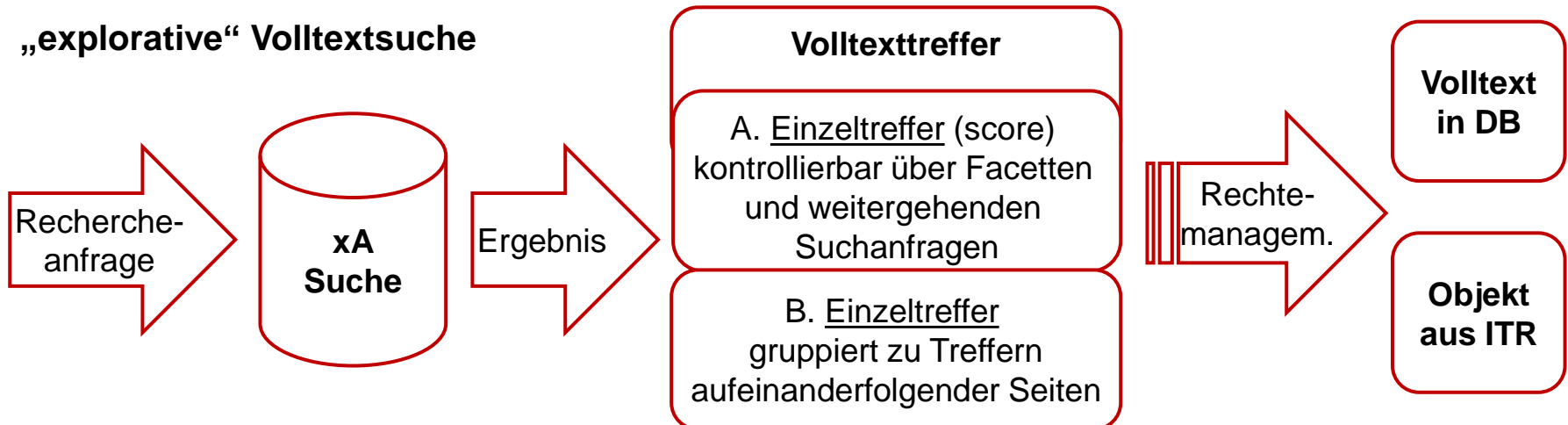
- > ... Schnittstellen zur Nutzung und Datenübertragung (**SRU**)
Entwicklung und Evaluierung prototypischer Volltextrecherchen („explorativ“ vs „geführt“)
und Einbindung in die CrossAsia Suche (**Ajax-Solr**)



„geführte“ Volltextsuche



„explorative“ Volltextsuche



Ostasien OPAC (crossasia.stabikat.de)

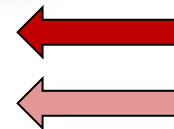
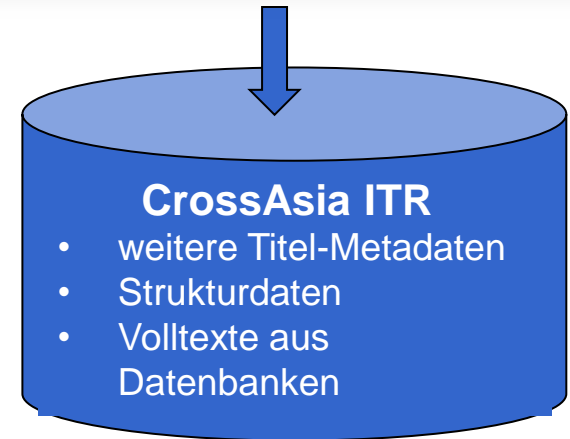
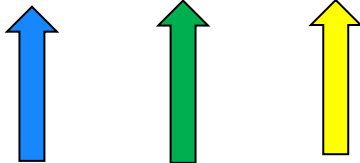
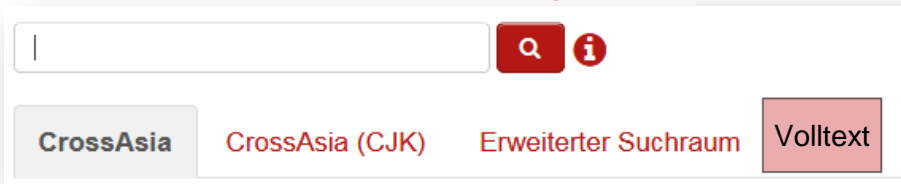


← Titelimport

CrossAsia Datenbankportal



CrossAsia Suche (crossasia.org)



Schnittstellen zu externen Ressourcen (CiNii, DBpia, Duxiu); Solr Indexe mit regelmäßig geharvesteten Daten (NDL, CRL); PDA Titel (Lieferung)

Aufbrechen der Grenze gedruckt vs. elektronisch

Erweiterte Metadatensuche (inkl TOC) > „Volltextsuche“ im Druckbestand

 **巴拉第的汉学研究**. 陈开科著, 2007. ▲

Autor: 陈开科著
Reihe: 列国汉学史书系
Sprache: Chinesisch
Themen: K207.8
历史、地理->中国史->通史->研究、考订、评论

Ausgabe: 北京：学苑出版社
ISBN 9787507729092

Datum: 2007

Beschreibung: 本书是俄罗斯著名汉学家巴拉第在中国宗教、译注蒙古史等方面的研究成果。

[Sample pages]

[Im Katalog ansehen >](#)

[CrossAsia Lizenz >](#)

Weg 1: Matching StaBi-Bestand mit Titeln, die in Duxiu im Volltext indexiert wurden
> Abfrage über vereinbarte Schnittstelle

Aufbrechen der Grenze gedruckt vs. elektronisch

Volltext-Suche

> Snippet
公缓过神道：“启禀皇上，在库房捉着一个小太监，饿了两天两夜，说是怕被杀了躲在那里，胡言乱语的。给他灌下碗粥才清醒，问他为什么躲在那里，说是怕被人杀了。”岚琪听得心惊肉跳，沉声问：“他原先在哪儿的？谁要杀他？”公公继续说，那孩子原在内务府当差，那日躲着偷懒时，瞧见有人来找他的师父，不知说了什么话不合，那人竟当

Probeseiten
online



Papier
ausleihen

[Im Katalog ansehen >](#)
[Bestellen via Blauen Leihverkehr >](#)
[Bestellen via SBB Ausleihsystem >](#)

Weg 2: Scannen und OCR-Verarbeitung von Erwerbungen > Aufbau Solr-Index

Danke!

A service of



In cooperation with



Supported by

