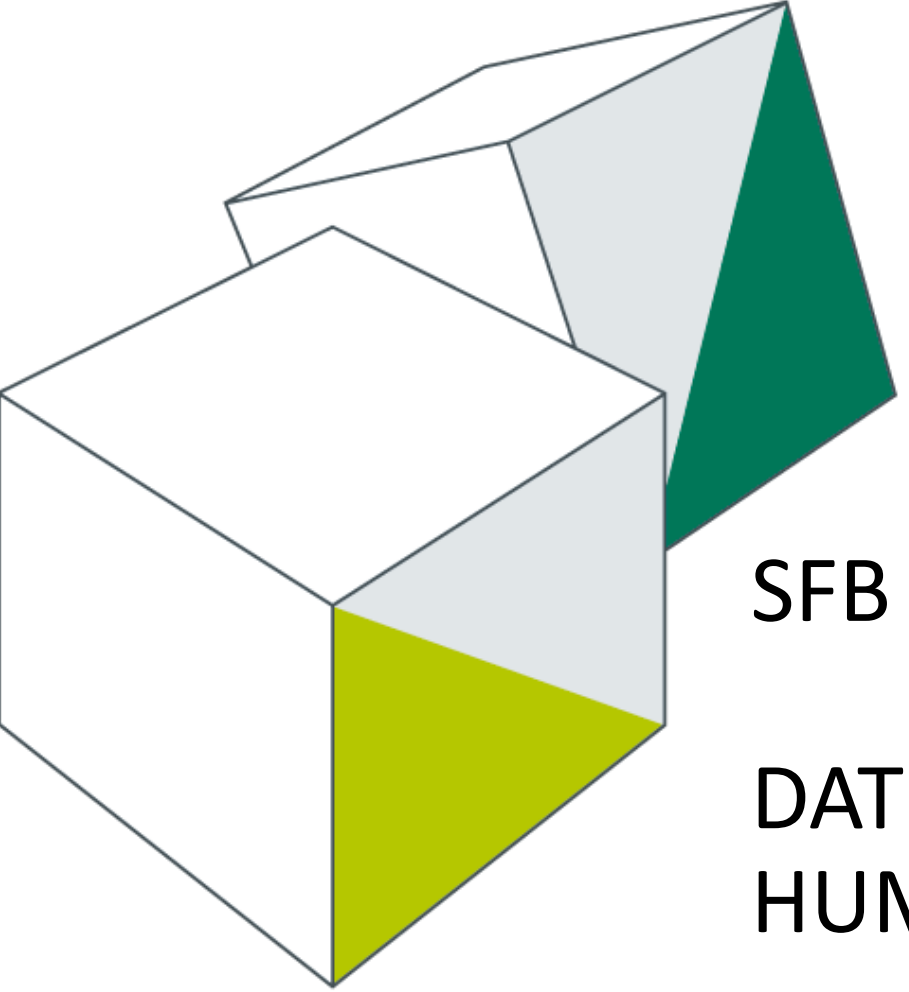


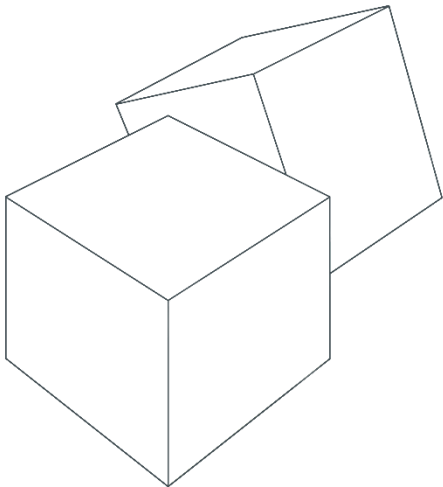
Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

SFB 1288:

DATENINFRASTRUKTUR & DIGITAL HUMANITIES

02. FEBRUAR 2018

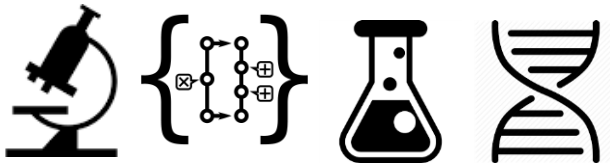


Kolloquium Wissensinfrastruktur

Vortragende: Madis Rumming, Johanna Vompras

WAS SIND „GEISTESWISSENSCHAFTLICHE“ FORSCHUNGSDATEN?

- Nicht so einfach zu definieren wie in den Naturwissenschaften



- Eher „Quelle“ oder „Forschungsliteratur“
- Charakter des ***Vorläufigen*** und ***Unfertigen***

- Begriffs „Forschungsdaten“ hängt elementar mit der Digitalisierung zusammen, gebunden an:

  (1) ***Prozessierbarkeit***: die Möglichkeit digitalen Arbeitens
  (2) ***Methode, bzw. Transformation***

Buzetti (2009): **Data is the representation of information in a form that can be processed by a machine.**

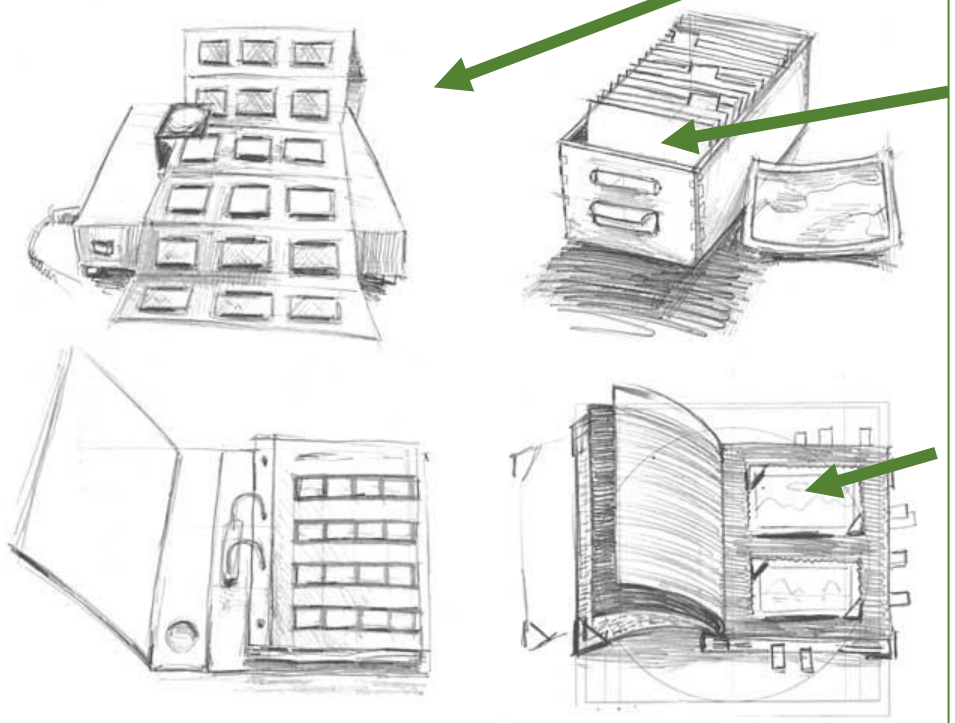
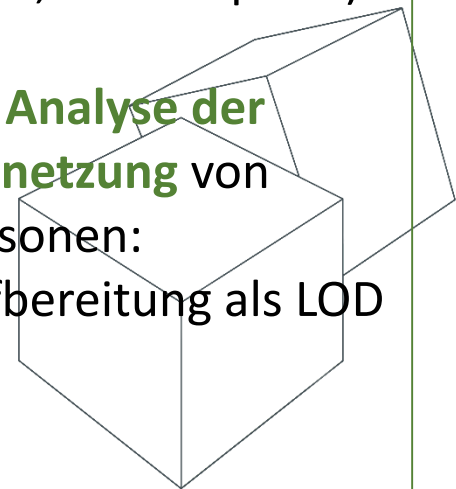


Bild Quelle: www.dietieredeswaldes.de / Copyright: Thomas Ries

Bildsuche: Herstellung eines digitalen Faksimiles (*image*)

linguistische Textanalyse: Text in strukturierter Form (OCR, Transkription)

die **Analyse der Vernetzung** von Personen: Aufbereitung als LOD



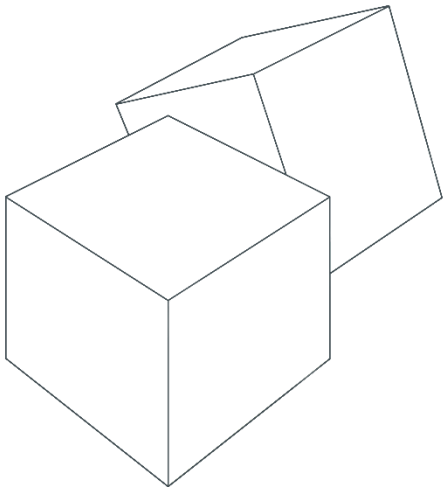
Quelle der Illustrationen: https://pixabay.com

DH-Aktivitäten (infrastrukturellen Belange)

- Vorgeschaltete Verfahren der Digitalisierung
- Speicherung von Forschungsdaten
- Entwicklung von Publikationsworkflows

DH-Verfahren („DH als Forschungsmethodik“)

- Ziel: *die Untersuchung der Daten für den Erkenntnisgewinn*
- Finden von Antworten auf Forschungsfragen durch digitale Verfahren



Quelle der Illustrationen: https://pixabay.com

DH is a research discipline that bridges the gap between engineering and humanities. DH is not simply about engineering software for humanities scholars. Instead, DH is where disciplines from engineering and humanities combine their expertise in dealing with information per se. (And obviously, one potential outcome of such a joint research may be tools that can be used in the humanities.)

Tim Weyrich

DIGITAL HUMANITIES

Digital humanities is the synergy of the humanities and digital technologies, resulting in potentially powerful new ways of researching, analysing, synthesising and presenting humanities scholarship.

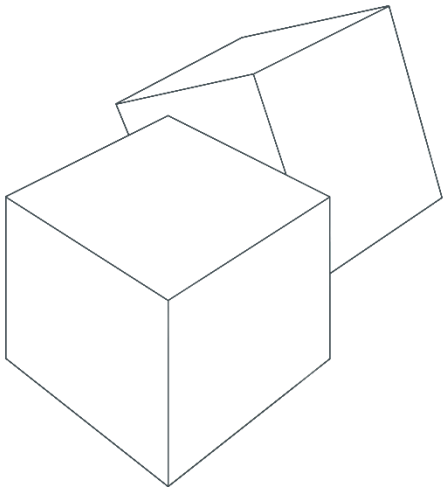
Rachel Murphy

DH is a term that covers all scholarly methods, practices, endeavors, and challenges lying at the intersection of Humanities and Information science and technology.

Aris Xanthos

DH are part of the Humanities. Humanists ask research questions. From time to time it makes sense to answer them by using digital means and methods. It's not a decision between digital or analog. Digital Humanities has to match with scope and focus of the research project

thomas



Quelle: http://whatisdigitalhumanities.com/

DH is a research discipline that bridges the gap between engineering and humanities. DH is not simply about engineering software for humanities scholars. Instead, DH is where disciplines from engineering and humanities combine their expertise in dealing with information per se. (And obviously, one potential outcome of such a joint research may be tools that can be used in the humanities.)

Tim Weyrich

DIGITAL HUMANITIES

Digital humanities is the synergy of the humanities and digital technologies, resulting in potentially powerful new ways of researching, analysing, synthesising and presenting humanities scholarship.

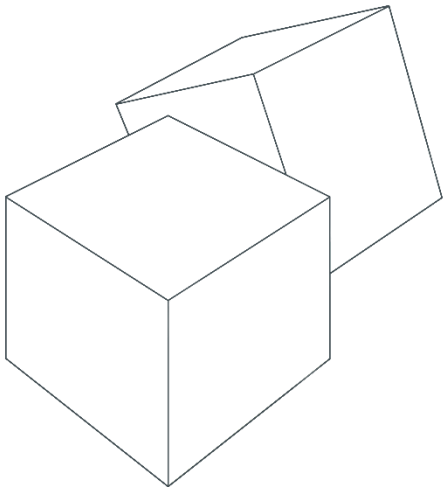
Rachel Murphy

DH is a term that covers all scholarly methods, practices, endeavors, and challenges lying at the intersection of Humanities and Information science and technology.

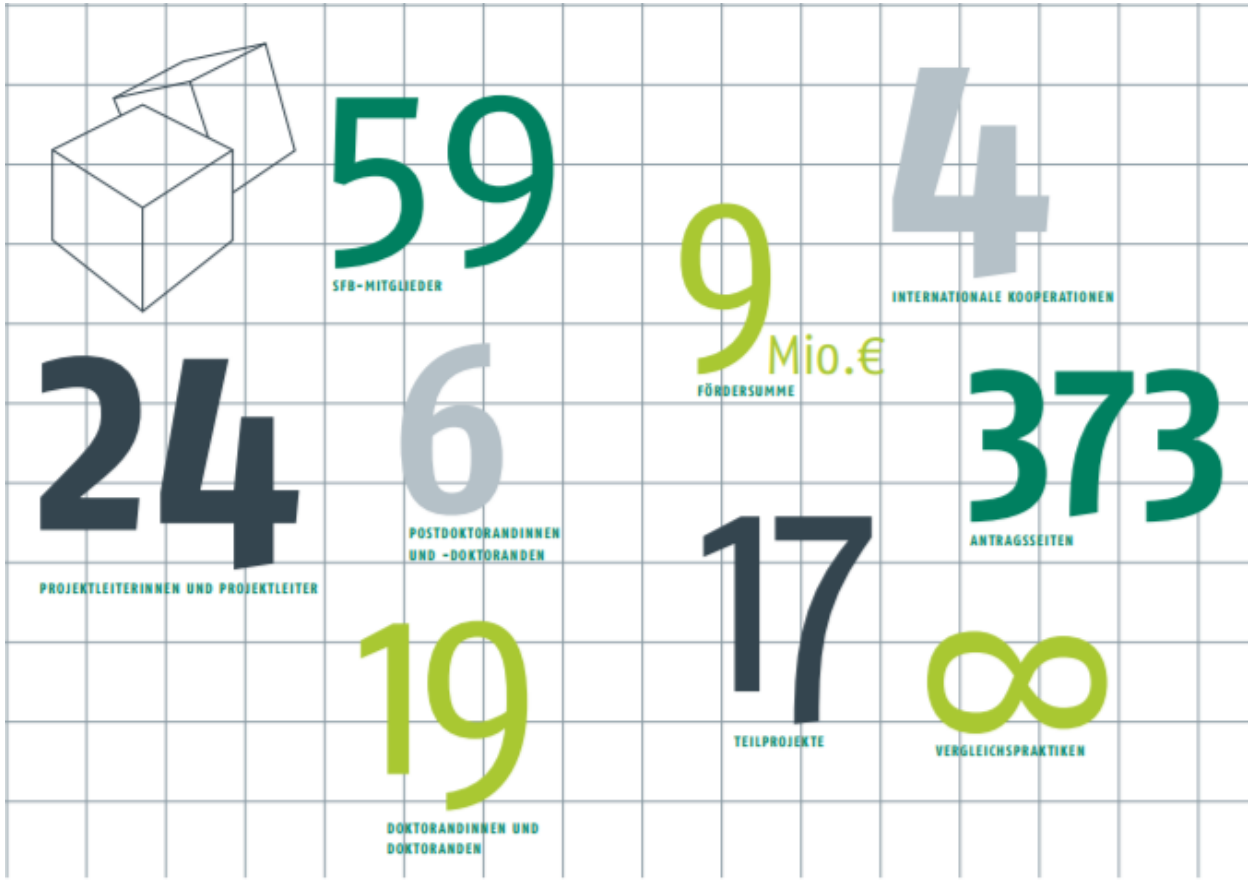
Aris Xanthos

DH are part of the Humanities. Humanists ask research questions. From time to time it makes sense to answer them by using digital means and methods. It's not a decision between digital or analog. Digital Humanities has to match with scope and focus of the research project

thomas



Quelle: http://whatisdigitalhumanities.com/



Universität Bielefeld

SFB 1288 PRAKTIKEN DES VERGLEICHENS

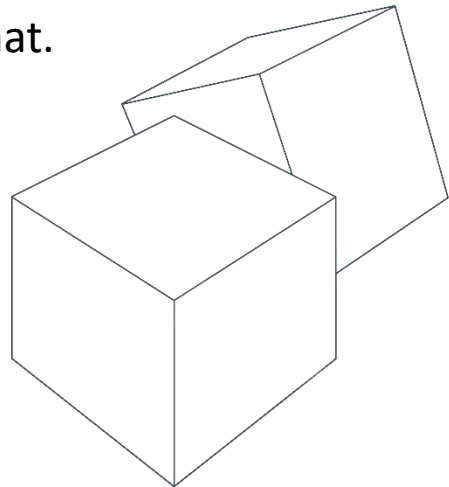
INF Projekt im SFB 1288
(seit Januar 2017)
„Praktiken des Vergleichens“

SFB 1288 nähert sich dem Vergleichen aus einer historischen Perspektive

Was tun Akteure, wenn sie vergleichen?

SFB will alles über die Anfänge heraus-finden und wie sich das Vergleichen im Lauf der Geschichte geändert hat.

Der interdisziplinäre, aus **Geschichts-** und **Literaturwissenschaft**, **Philosophie**, **Kunstgeschichte**, **Politik-** und **Rechtswissenschaft** bestehende Forschungsverbund, fragt wie sich die historisch variablen Praktiken des Vergleichens zu Routinen, Regeln, Habitus, Institutionen und Diskursen fügen und so Strukturen schaffen, aber auch Dynamiken mittlerer Reichweite oder übergreifenden Wandel anstoßen können.



INF Projekt im SFB 1288:

„Dateninfrastruktur und Digital Humanities“

Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

1.3 Teilprojekt Informationsinfrastruktur

In einem Sonderforschungsbereich dient ein Teilprojekt Informationsinfrastruktur vor allem dem systematischen Management der im Kontext des Sonderforschungsbereichs relevanten Daten. Unter Forschungsdaten werden alle Ergebnisse und Bezugsquellen des Forschungsprozesses (u.a. auch Software oder Forschungsobjekte bzw. Proben) verstanden, die im Projekt erhoben, ausgewertet und/oder entwickelt werden. Ein auf diese Daten bezogener Einsatz sowie die Erprobung oder Entwicklung neuer wissenschaftlicher Kommunikationsformen ist ebenfalls möglich. Es können die Entwicklung

gemeinschaft
Postanschrift: 53170 Bonn
49 228 885-2777 · postmaster@dfg.de · www.dfg.de

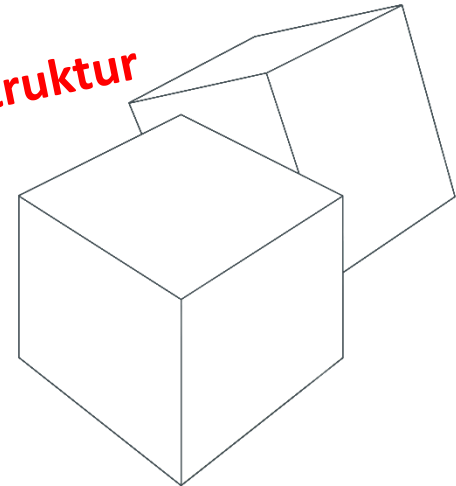
DFG

Damit soll erreicht werden, dass sich im Sonderforschungsbereich wissenschaftliche Synergien durch gemeinsame Datenplattformen und/oder Kommunikationsforen sowie eine effiziente Datennutzung einstellen.

Grundsätzlich wird ein professionelles Management der Daten, die im Sonderforschungsbereich erhoben, ausgewertet und/oder entwickelt werden, erwartet. In der Regel soll daher mit den einschlägig ausgewiesenen Informationseinrichtungen am Standort zusammengearbeitet werden (z.B. Bibliotheken, Rechenzentren oder Biobanken der antragstellenden Hochschule). Der Nutzung bereits vorhandener Repositorien, Werkzeuge und Techniken ist gegenüber der Entwicklung neuer Instrumente in der Regel der Vorzug einzuräumen, siehe auch:

www.re3data.org

Pilot für die Erweiterung der hochschulweiten Forschungsdateninfrastruktur



INF Projekt im SFB 1288:

Interdisziplinär besetztes Team, bestehend aus:

Dr. Silke Schwandt
Digital Humanities
(Fak. Für Geschichte)

Dr. Johanna Vompras
Dateninfrastruktur / FDM
(Universitätsbibliothek)

Anna Maria Neubert
Digital Humanities

Dr. Madis Rumming
Dateninfrastruktur
/ FDM

Helene Schlicht
Digital Humanities

Leonard Gödde
wiss. Hilfskraft

Ab 01. Feb 2018

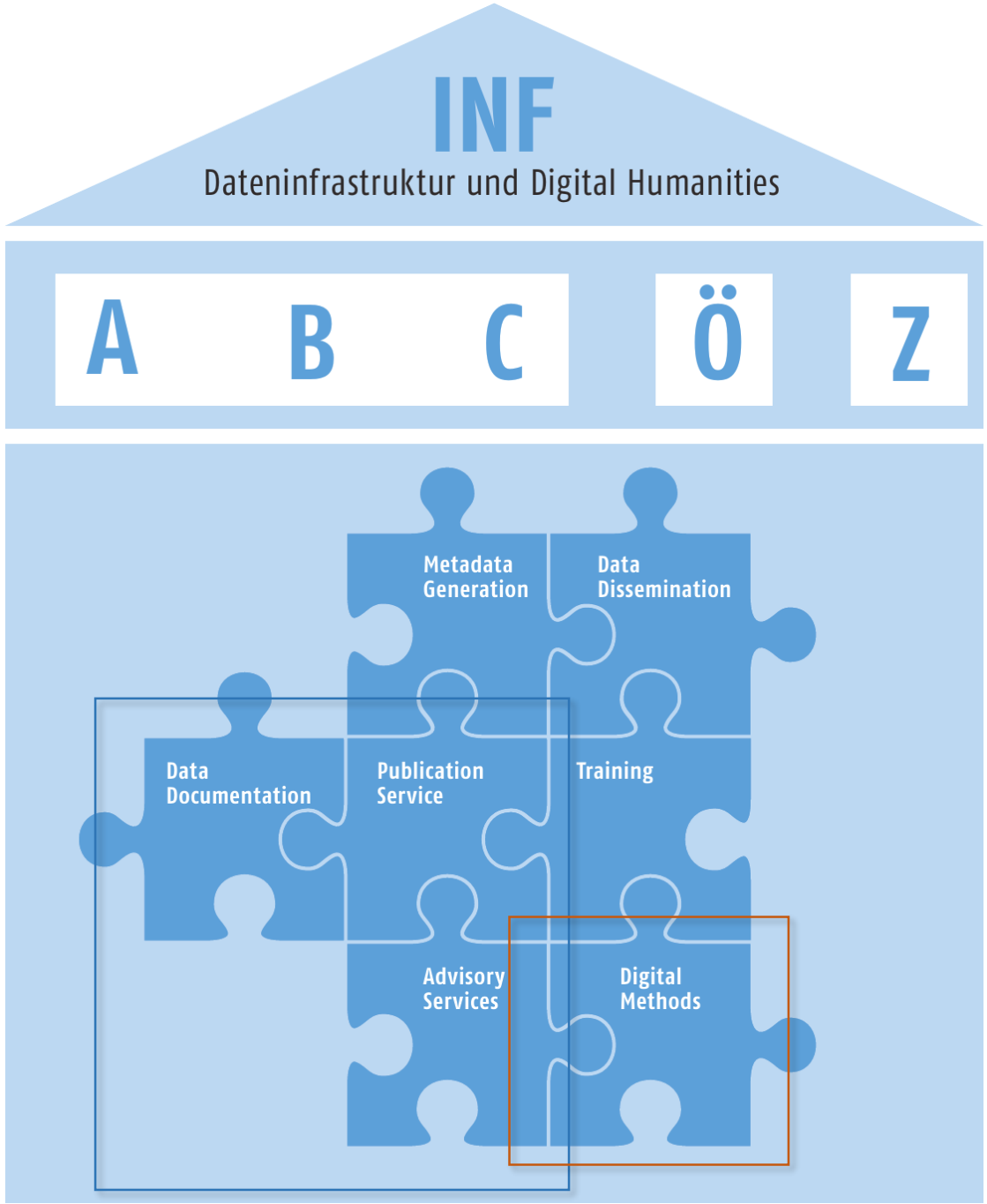
Stefan Porada
wiss. Hilfskraft

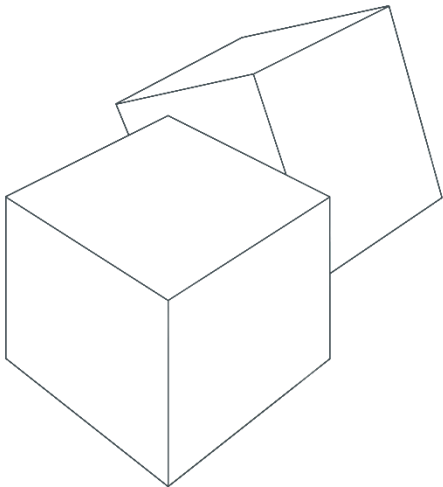
Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

INF Projekt im SFB 1288:

Dateninfrastruktur
und
Digital Humanities

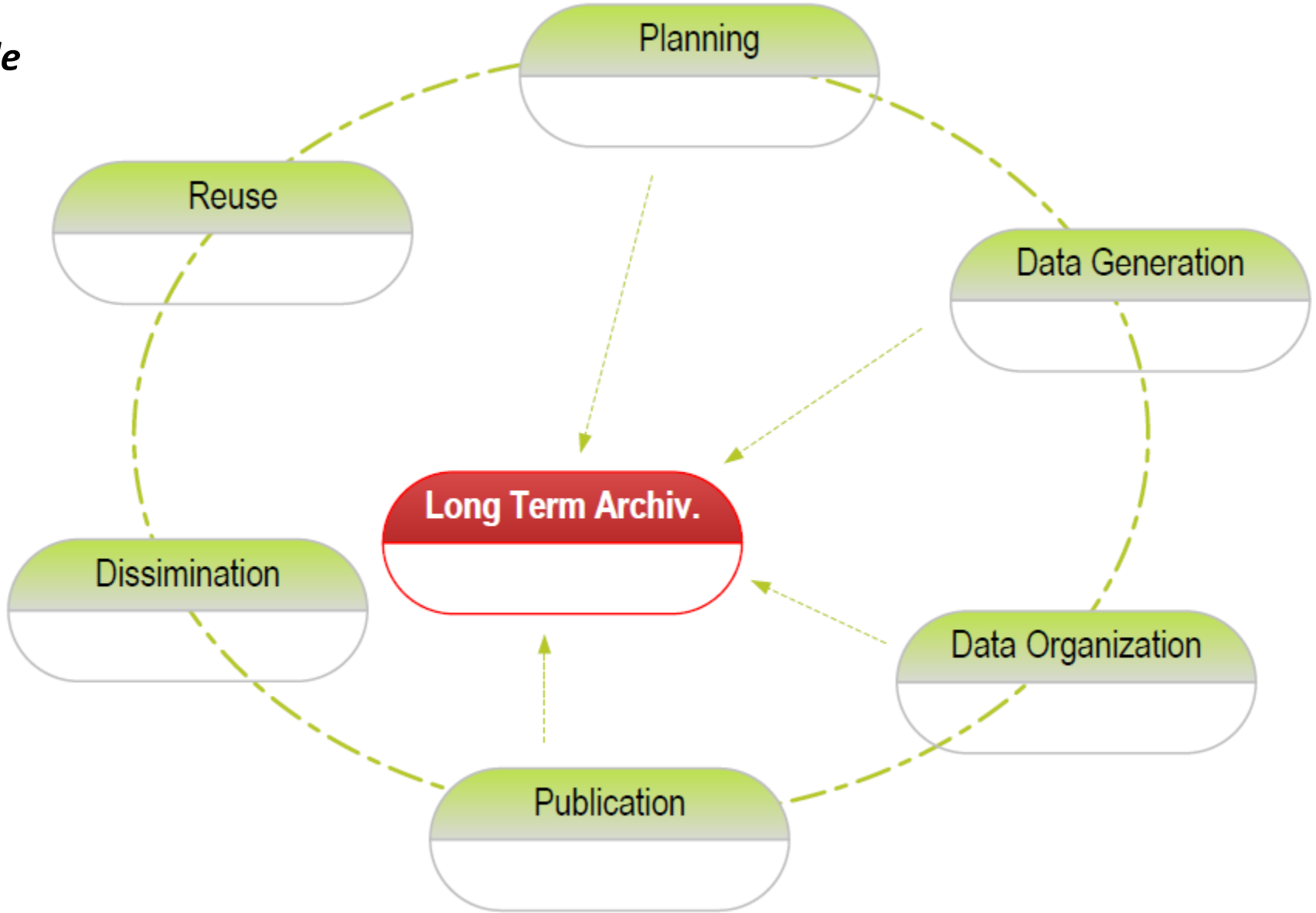


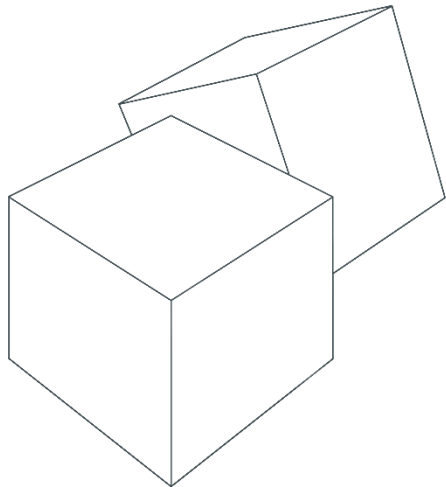


Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Generischer
Data Life Cycle

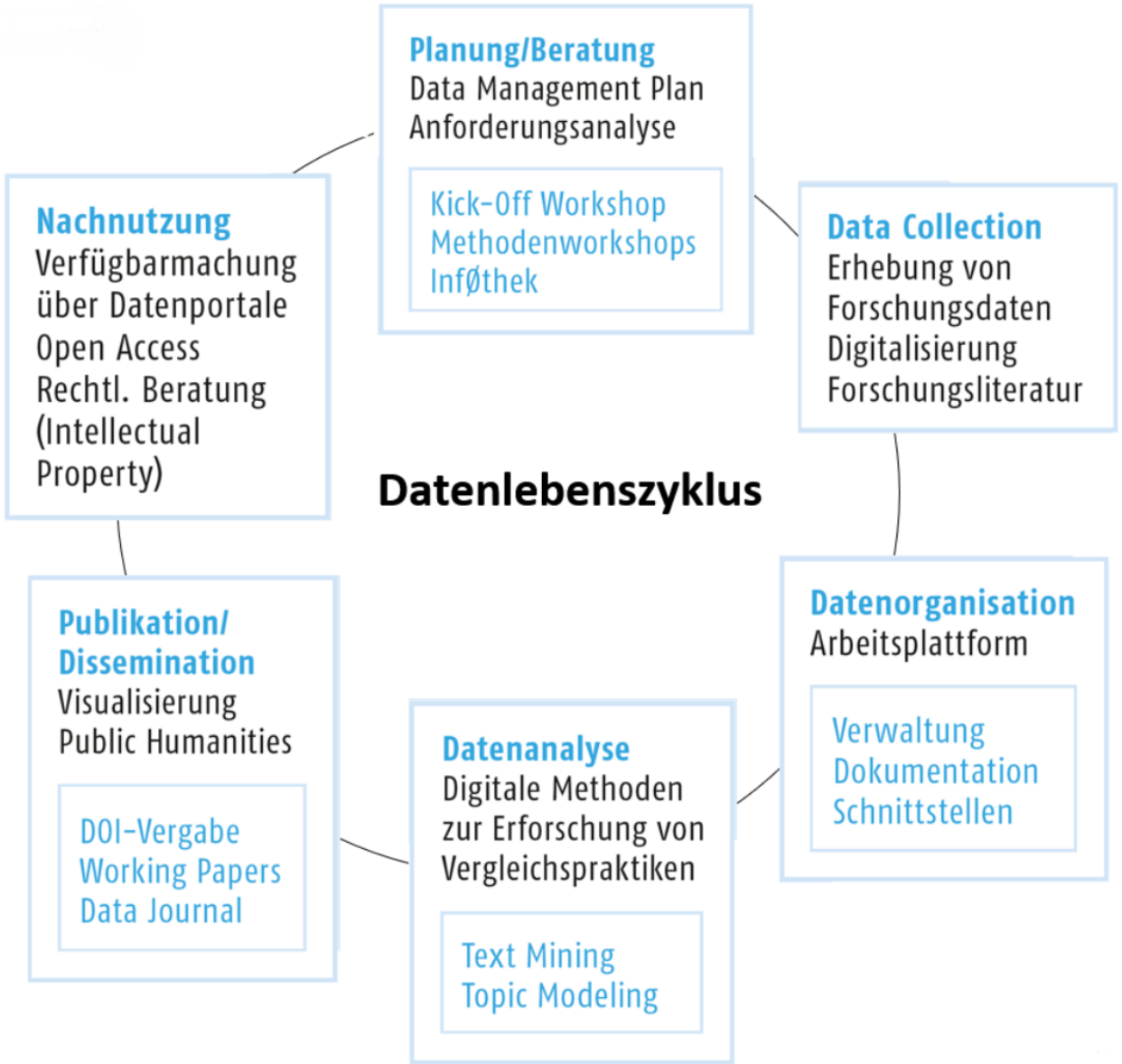


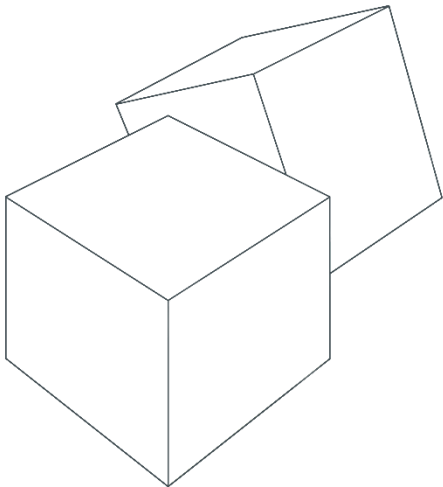


Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Support der zahlreicher Stadien des *Data Life Cycle* in den Digital Humanities

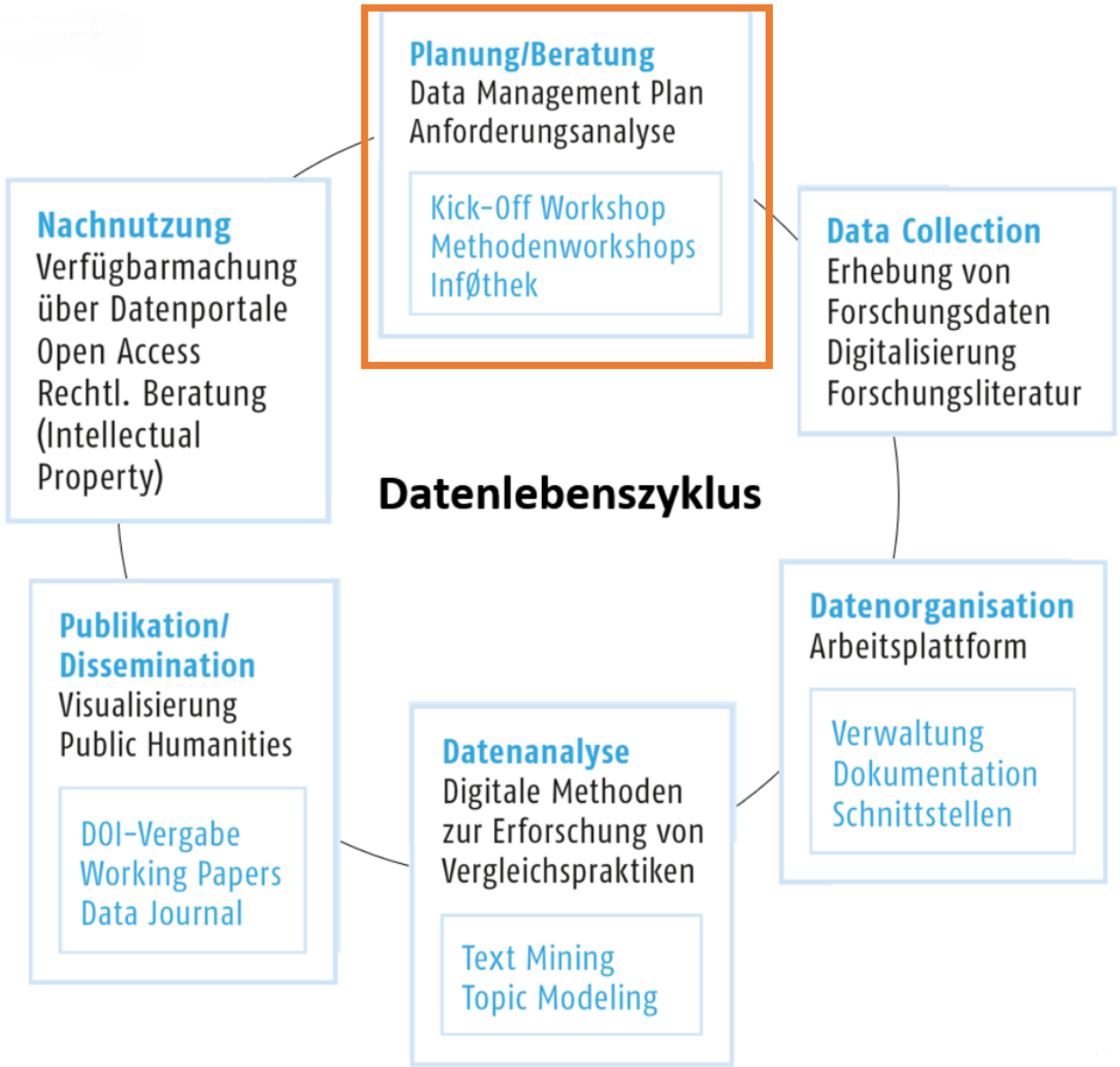


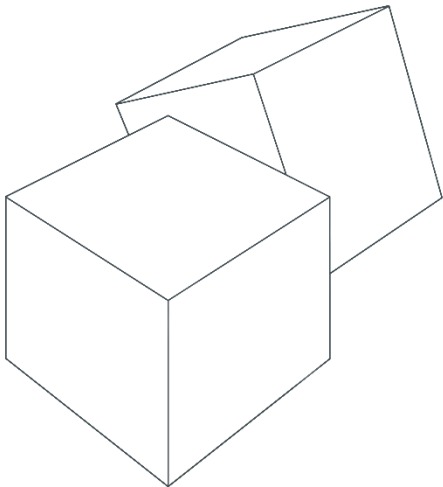


Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Support der zahlreicher Stadien des *Data Life Cycle* in den Digital Humanities





Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

INF Projekt im SFB 1288:
Formen der Zusammenarbeit

„*INFØthek*“: bilaterales Vorgehen

- Austausch über dateninfrastrukturelle Belange & Nutzung von Tools
- Analyse Ihrer ***Forschungsprozesse*** und ***Arbeitsweisen*** im Projekt
- Evaluation potenzieller ***Schnittstellen zu INF*** und ***Anforderungen an digitale Methoden***

Planung/Beratung
Data Management Plan
Anforderungsanalyse

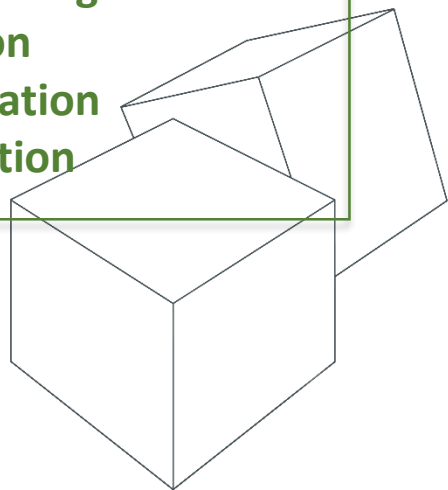
Kick-Off Workshop
Methodenworkshops
Infθthek

„Vergleichen“ im *Digital Age*

- Discovering
- Annotating
- Comparing
- Referring
- Sampling
- Illustrating
- Representing

Infrastruktursicht

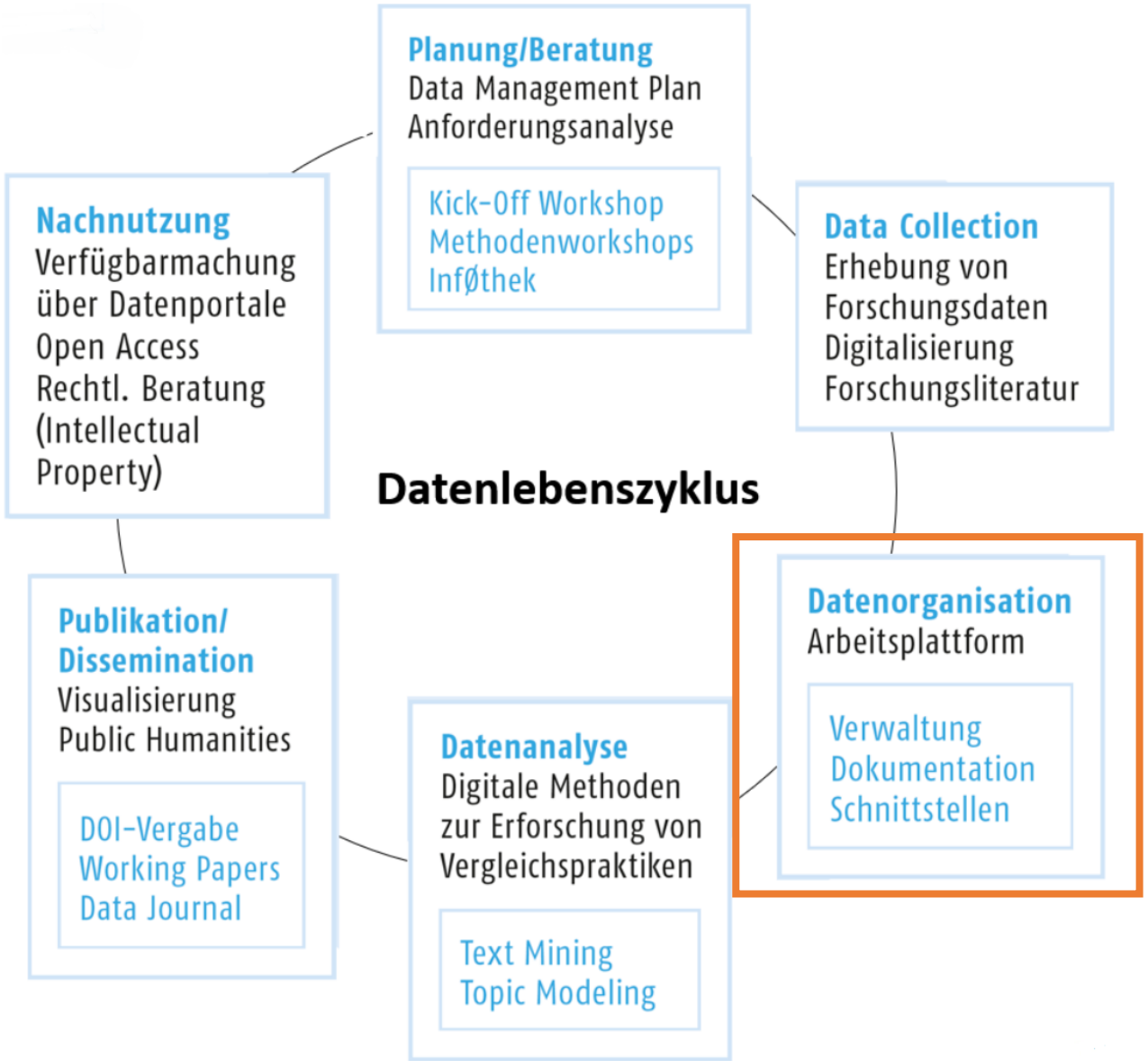
retrieval
metadata enrichment
similarity computation
entity linkage
selection
visualization
publication

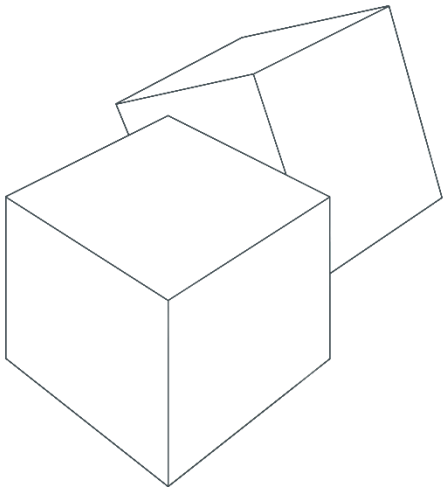


Universität Bielefeld

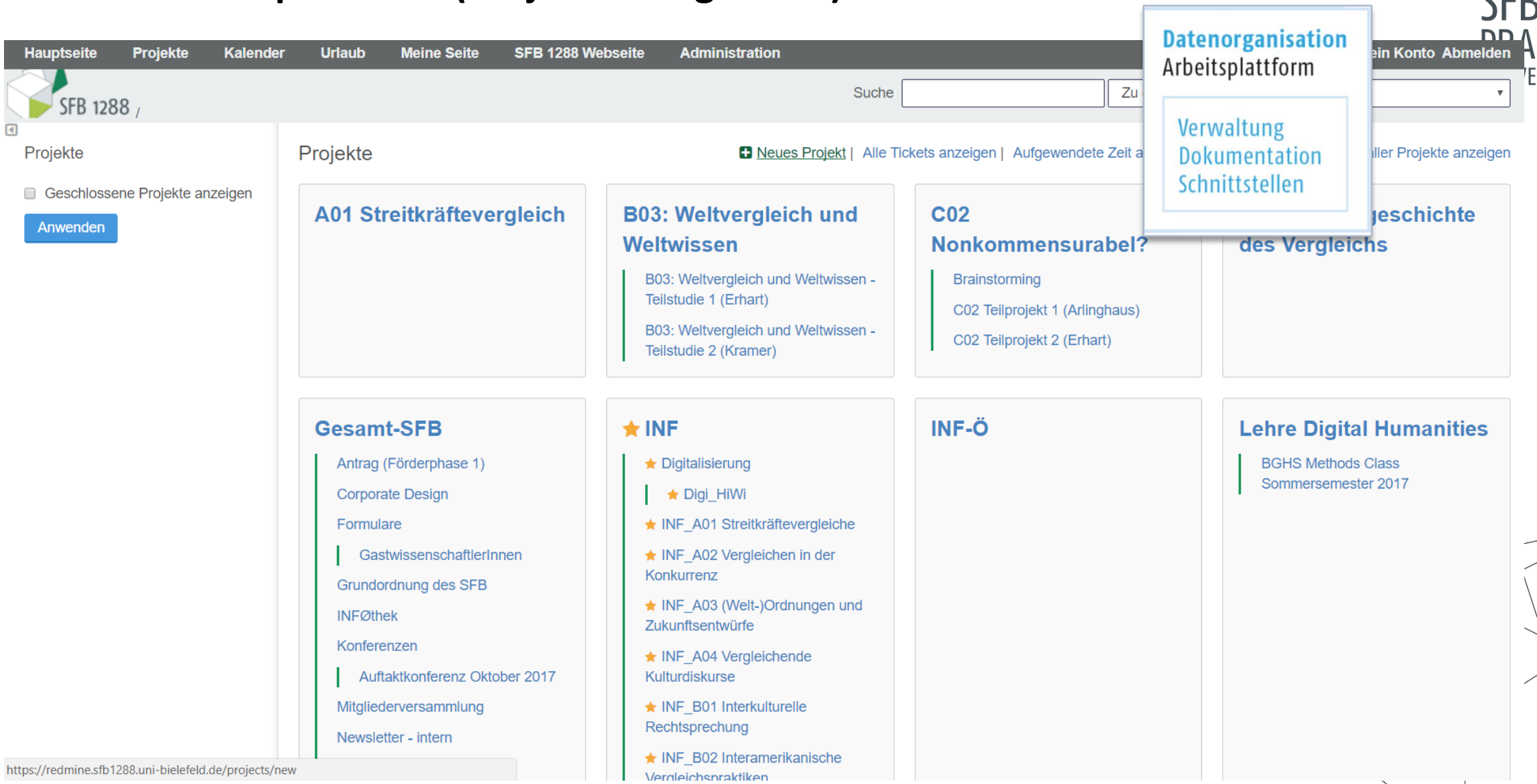
SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Support der zahlreicher Stadien des *Data Life Cycle* in den Digital Humanities





Kollaborationsplattform (Projektmanagement):



Kollaborationsplattform (Projektmanagement):

Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Datenorganisation
Arbeitsplattform

Verwaltung
Dokumentation
Schnittstellen

Hauptseite | Projekte | Kalender | Urlaub | Meine Seite | SFB 1288 Webseite | Administration | Angemeldet als **admin** Mein Kon

SFB 1288 /

Suche | Zu einem Projekt springen...

Projekte

Geschlossene Projekte anzeigen

Anwenden

Projekte

Neues Projekt | Alle Tickets anzeigen | Aufgewendete Zeit aller Projekte anzeigen | Aktivitäten aller Proj

A01 Streitkr

Gesamt-SFB

Antrag (Förderph
Corporate Desig
Formulare
Gastwissens
Grundordnung d
INFØthek
Konferenzen
Auftaktkonfer
Mitgliederversam
Newsletter - inte

https://redmine.sfb1288.uni-bielefeld.de/projects/new

Hauptseite | Projekte | Kalender | Urlaub | Meine Seite | SFB 1288 Webseite | Administration

SFB 1288 / INF

Suc

+ | Übersicht | **Aktivität** | Tickets | Gantt-Diagramm | Kalender | News | DMS | Wiki | Foren | Dateien | Konfiguration

Aktivität

- [x] Tickets
- [x] Changesets
- [x] News
- [x] Dokumente
- [x] Dateien
- [ ] Wiki-Bearbeitungen
- [x] Forenbeiträge
- [ ] Benötigte Zeit
- [ ] Dokumentezugriffe
- [ ] Dokumenteversion
- [x] Unterprojekte

Anwenden

Aktivität

von 16.09.2017 bis 15.10.2017

10.10.2017

14:13 Digi_HiWi - Digitalisierung #89 (Neu): C01 Arbeitspaket#1

Die zu bearbeitenden Dateien des 1. Arbeitspakets (C01) finden Sie hier: https://uni-bielefeld.sciebo.de/index.ph...
Madis Rumming

12:39 Digi_HiWi - Digitalisierung #88 (Neu): A04 Arbeitspaket#1

Die zu bearbeitenden Dateien des 1. Arbeitspakets (A04) finden Sie hier: https://uni-bielefeld.sciebo.de/index.ph...
Madis Rumming

12:38 Digi_HiWi - Digitalisierung #87 (Neu): B03 Arbeitspaket#1

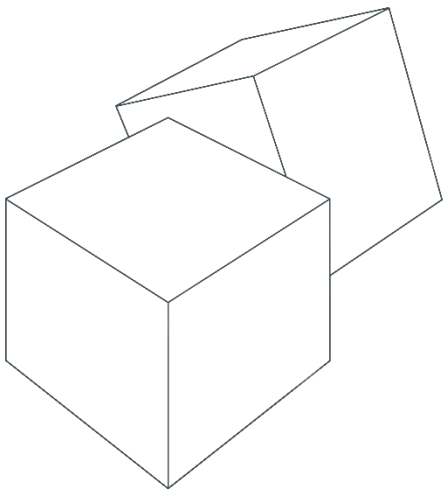
Die zu bearbeitenden Dateien des 1. Arbeitspakets (B03) finden Sie hier: https://uni-bielefeld.sciebo.de/index.ph...
Madis Rumming

12:29 Digi_HiWi - Digitalisierung #85 (Neu): B01 Arbeitspaket#1

Die zu bearbeitenden Dateien des 1. Arbeitspakets (B01) finden Sie hier: https://uni-bielefeld.sciebo.de/index.php/s/...
Madis Rumming

07.10.2017

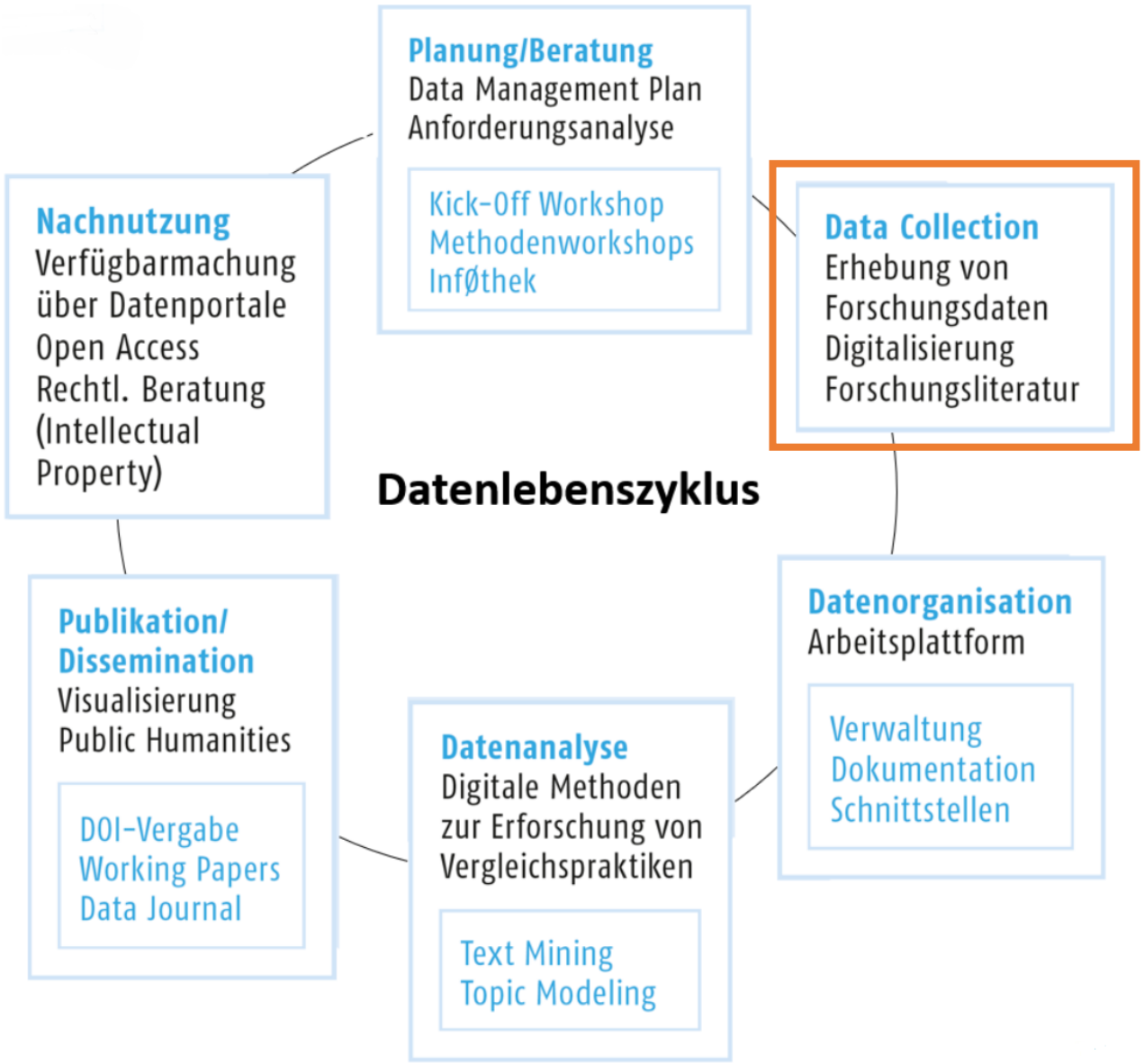
10:18 Digitalisierung - Termine_INFDigi #10 (Erledigt): Nachfrage Sabrina: WHKs und Oxygen
Madis Rumming

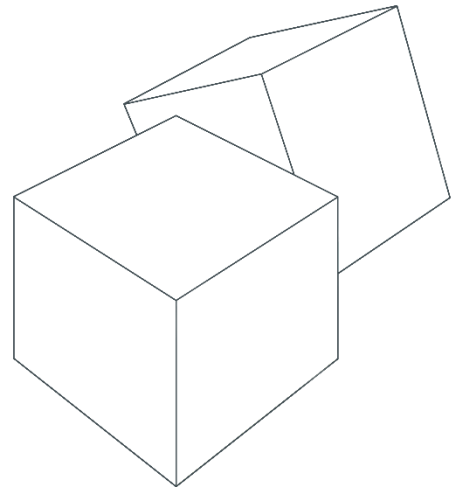


Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Support der zahlreicher Stadien des *Data Life Cycle* in den Digital Humanities

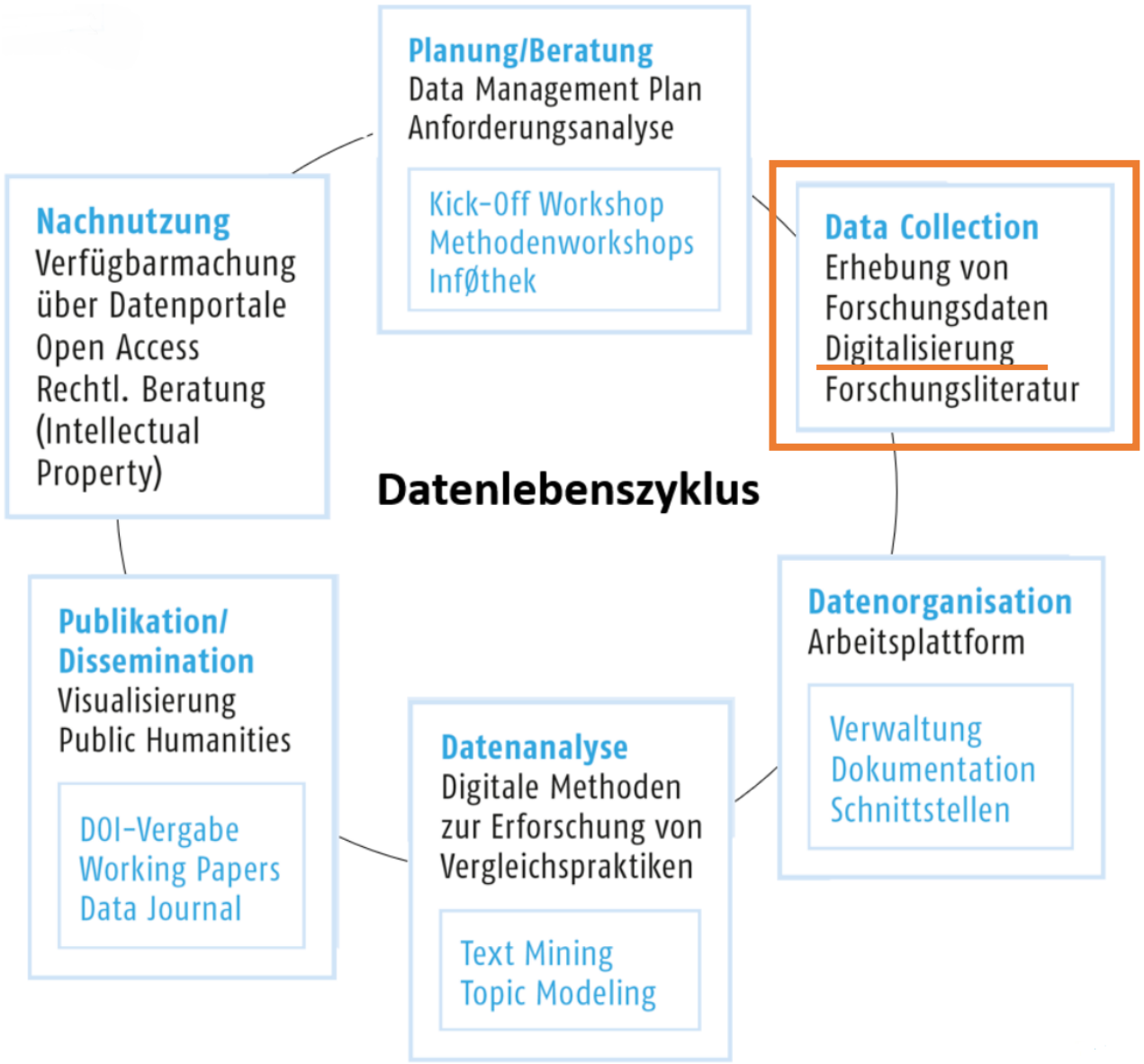


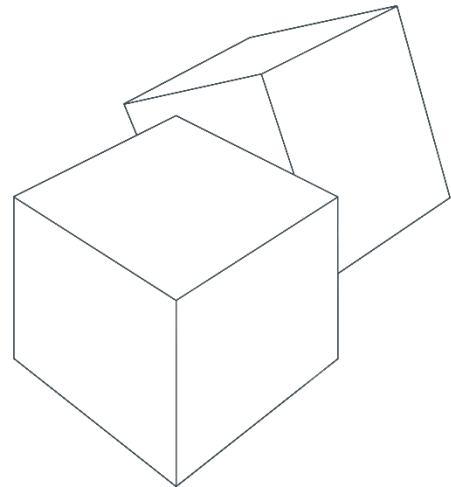


Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Support der zahlreicher Stadien des *Data Life Cycle* in den Digital Humanities





Data Collection
Erhebung von
Forschungsdaten
Digitalisierung
Forschungsliteratur

Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Digitalisierung

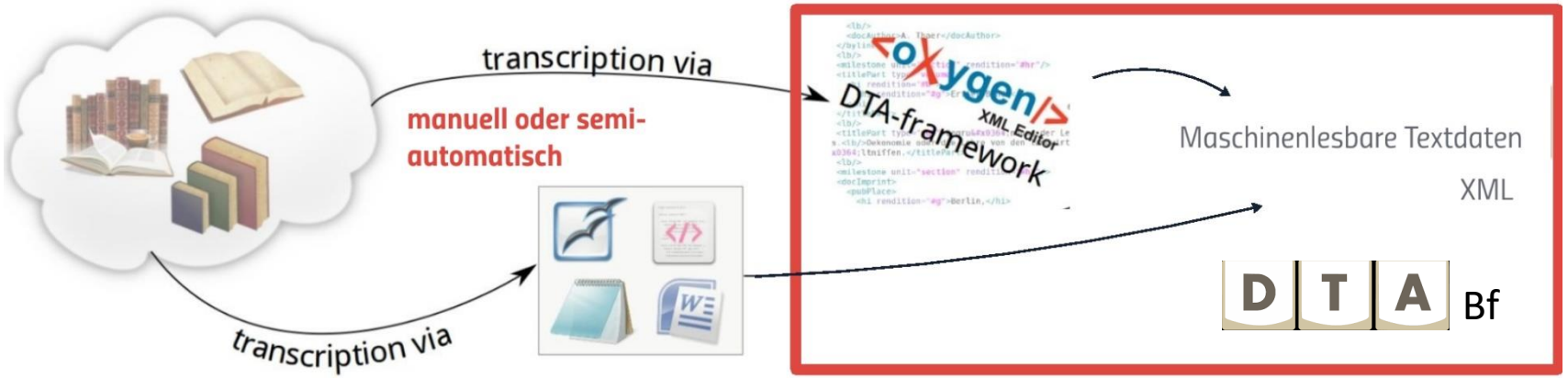
Etablierung von Workflows

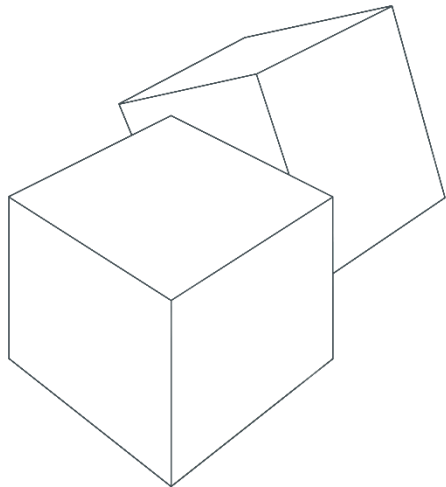
- OCR
- Transkription / Richtlinien
- Dokumentationssprache

Etablierung von Tools

Software (z.B. PoCoTo)
Oxygen Framework
DTA-Bf (TEI-P5-basiert)

TEI Standard: Text Encoding Initiative

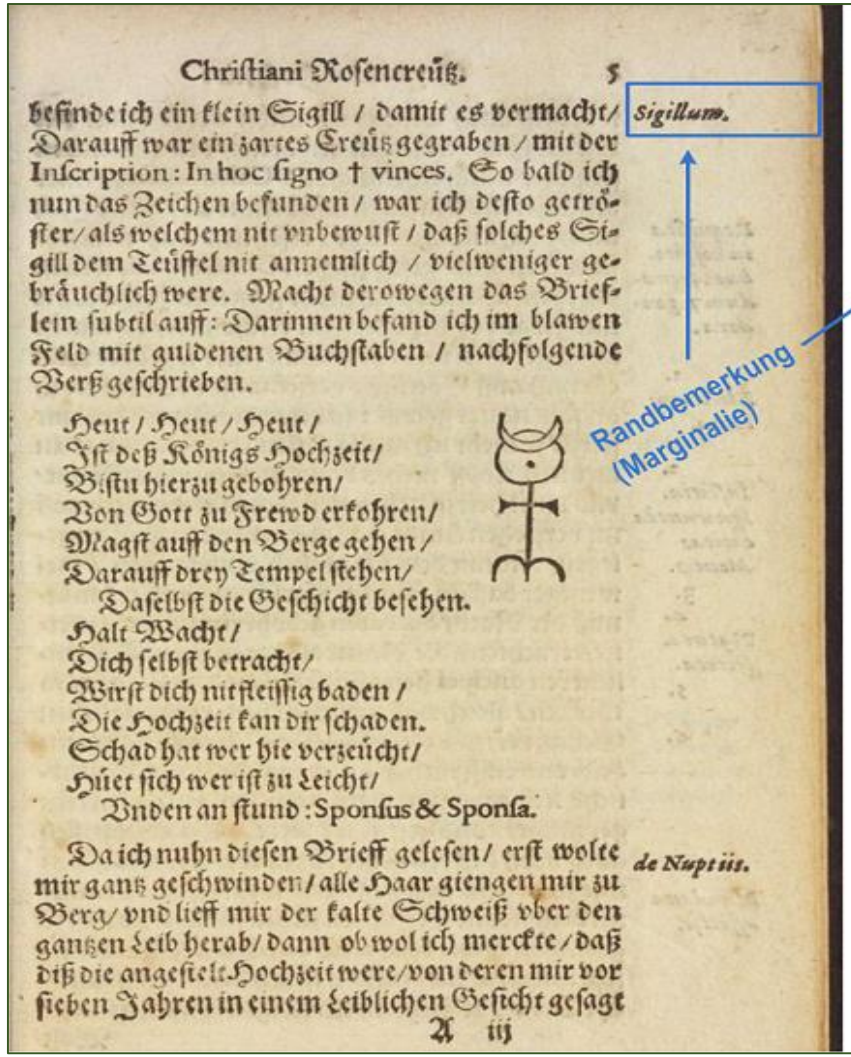




Teile der Bildes entnommen aus dem Workflow des Kurationsprojekts 1 der Facharbeitsgruppe 1 Deutsche Philologie (www.clarin-d.net)

Digitalisierung

Textauszeichnung mit DTA-Bf (umfasst 133 TEI-P5 Elemente, XML-basiert)



Christiani Rosencreütz. 5

befinde ich ein klein Sigill / damit es vermacht/ Darauff war ein zartes Creütz gegraben / mit der Inscription: In hoc signo † vinces. So bald ich nun das Zeichen befunden / war ich desto getröster/ als welchem nit vnbewust / daß solches Sigill dem Teüffel nit annemlich / vielweniger gebräuchlich were. Macht derowegen das Brieflein subtil auff: Darinnen befand ich im blawen Feld mit guldenen Buchstaben / nachfolgende Verß geschrieben.

Sigillum.

Heut / Heut / Heut /
Ist deß Königs Hochzeit/
Bistu hierzu gebohren/
Von Gott zu Frewd erkohren/
Magst auff den Berge gehen /
Darauff drey Tempel stehen/
Daselbst die Geschicht besehen.
Halt Wacht /
Dich selbst betracht/
Wirst dich nitfleissig baden /
Die Hochzeit kan dir schaden.
Schad hat wer hie verzeücht/
Hüet sich wer ist zu Leicht/
Vnden an stund: Sponsus & Sponsa.

Randbemerkung (Marginalie)

Da ich nuhn diesen Brieff gelesen/ erst wolte mir gantz geschwinden / alle Haar giengen mir zu Berg/ vnd lieff mir der kalte Schweiß vber den gantzen Leib herab/ dann ob wol ich merckte / daß diß die angestelt Hochzeit were/ von deren mir vor sieben Jahren in einem Leiblichen Gesicht gesagt

de Nuptiis.

A iij

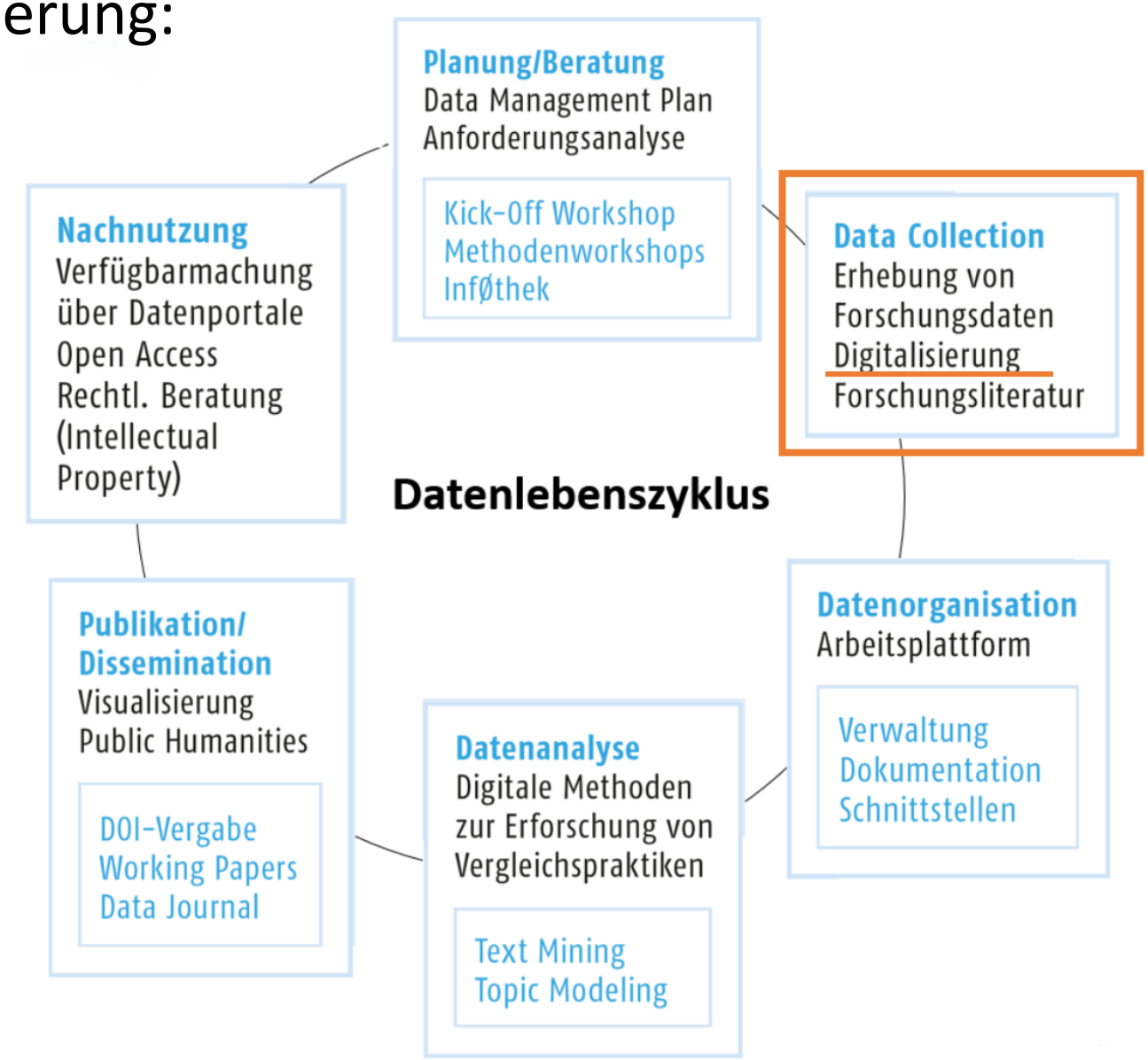
```
<TEI>
  <text>
    <body>
      <div n="1">
        <div n="2">
          <p><pb facs="#f0009" n="5"/><fw place="top" type="header"><hi
rendition="#aq">Chri&#x017F;tiani</hi> Ro&#x017F;encreu&#x0364;tz.</fw><lb/>
befinde ich ein klein Sigill/ damit es vermacht/<note place="right"><hi
rendition="#i"><hi rendition="#aq">Sigillum.</hi></hi></note><lb/>
Darauff war ein zartes Creu&#x0364;tz gegraben/ mit der<lb/><hi
rendition="#aq">In&#x017F;cription: In hoc &#x017F;igno &#x2020; vinces</hi>.
So bald ich<lb/>
nun das Zeichen befunden/ war ich de&#x017F;to getro&#x0364;-<lb/>
&#x017F;ter/ als welchem nit vnbewu&#x017F;t/ daß &#x017F;olches Si-<lb/>
gill dem Teu&#x0364;ffel nit annemlich/ vielweniger ge-<lb/>
bra&#x0364;uchlich were. Macht derowegen das Brief-<lb/>
lein &#x017F;ubtil auff: Darinnen befand ich im blawen<lb/>
Feld mit guldenen Buch&#x017F;taben/ nachfolgende<lb/>
Verß ge&#x017F;chrieben.</p><lb/>
          <lg type="poem">
            <lg n="1">
              <l>Heut/ Heut/ Heut/ <figure/></l><lb/>
              <l>I&#x017F;t deß Ko&#x0364;nigs Hochzeit/</l><lb/>
              <l>Bi&#x017F;tu hierzu gebohren/</l><lb/>
              <l>Von Gott zu Frewd erkohren/</l><lb/>
              <l>Mag&#x017F;t auff den Berge gehen/</l><lb/>
              <l>Darauff drey Tempel &#x017F;tehen/</l><lb/>
              <l>Da&#x017F;elb&#x017F;t die Ge&#x017F;chicht be&#x017F;ehen.
</l>
            </lg><lb/>
            <lg n="2">
              <l>Halt Wacht/</l><lb/>
              <l>Dich &#x017F;elb&#x017F;t betracht/</l><lb/>
              <l>Wir&#x017F;t dich <choice>
<sic>nitflei&#x017F;&#x017F;ig</sic><corr>nit flei&#x017F;&#x017F;ig</corr>
</choice> baden/</l><lb/>
              <l>Die Hochzeit kan dir &#x017F;chaden.</l><lb/>
```

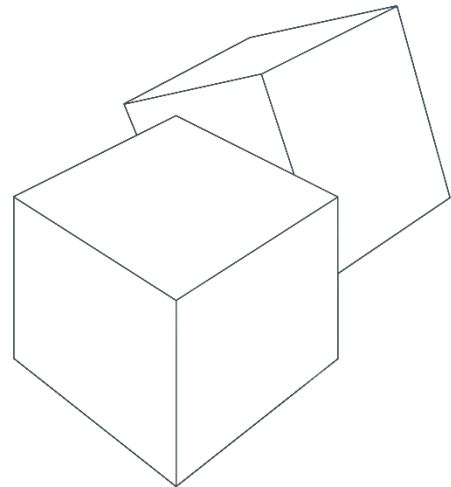
Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Pilotphase Digitsalisierung:

Vorbereitung für maschinengestütze Korpusanalyse





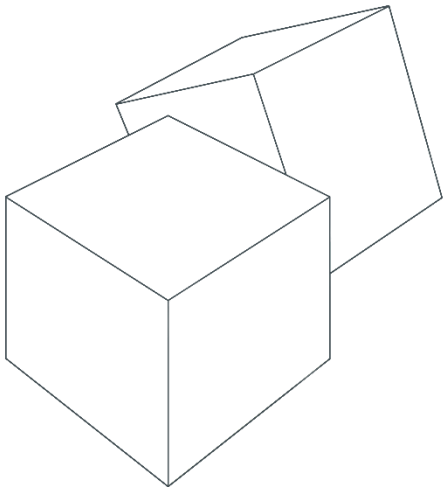
Pilotphase Digitsalisierung: Vorbereitung für maschinengestützen Korpusanalyse

Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

Korpus: Sammlung von (repräsentativen) Texten oder sprachlichen Äußerungen zum Zwecke wissenschaftlicher Untersuchungen.

- Digitale Methoden zur Verarbeitung textueller Daten setzen Maschinenlesbarkeit voraus
- Ziel: Erzeugung maschinenlesbarer Textformate durch Texterkennung in Digitalisaten

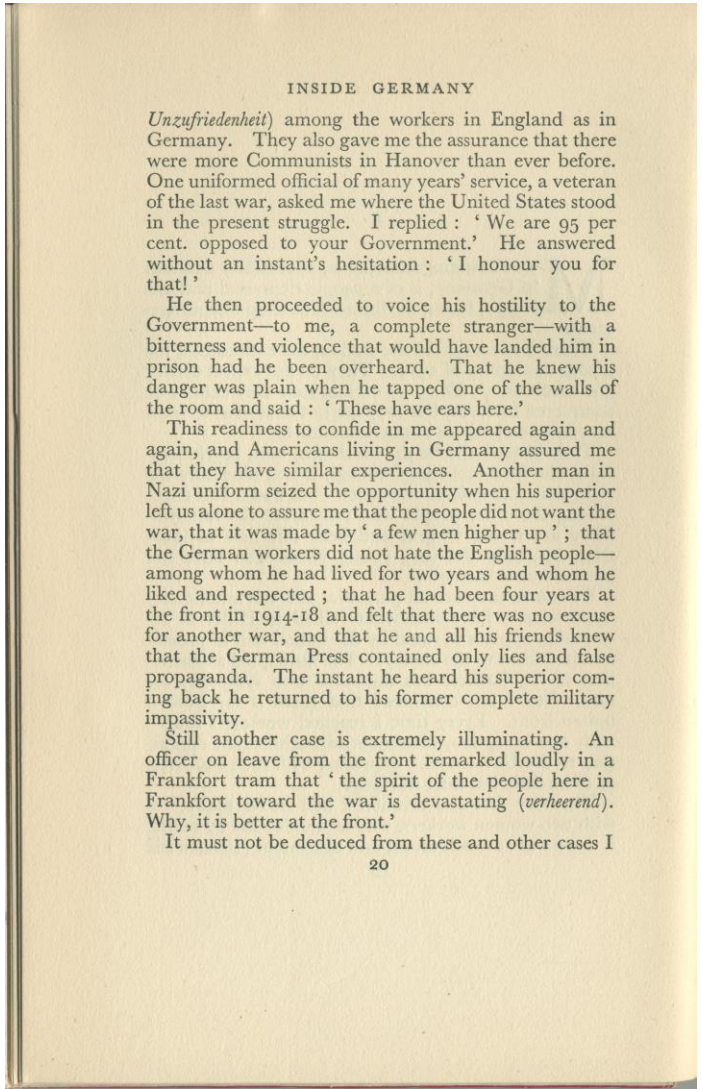


Vorbereitende Data Collection zur Datenanalyse und (Daten)Publikation

Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

- Digitalisat bisher: Digitale Fotographien oder Scans
  - Einzelne Bilddateien
  - Eingebettet in eine PDF
- Digitalisat in den Digital Humanities: Annotiertes Textformat, vorzugsweise XML-basiert
- Digitalisat im SFB1288: Markup gemäß DTAbf oder TEI P5



INSIDE GERMANY

Unzufriedenheit) among the workers in England as in Germany. They also gave me the assurance that there were more Communists in Hanover than ever before. One uniformed official of many years' service, a veteran of the last war, asked me where the United States stood in the present struggle. I replied : 'We are 95 per cent. opposed to your Government.' He answered without an instant's hesitation : 'I honour you for that!'

He then proceeded to voice his hostility to the Government—to me, a complete stranger—with a bitterness and violence that would have landed him in prison had he been overheard. That he knew his danger was plain when he tapped one of the walls of the room and said : 'These have ears here.'

This readiness to confide in me appeared again and again, and Americans living in Germany assured me that they have similar experiences. Another man in Nazi uniform seized the opportunity when his superior left us alone to assure me that the people did not want the war, that it was made by 'a few men higher up' ; that the German workers did not hate the English people—among whom he had lived for two years and whom he liked and respected ; that he had been four years at the front in 1914-18 and felt that there was no excuse for another war, and that he and all his friends knew that the German Press contained only lies and false propaganda. The instant he heard his superior coming back he returned to his former complete military impassivity.

Still another case is extremely illuminating. An officer on leave from the front remarked loudly in a Frankfort tram that 'the spirit of the people here in Frankfort toward the war is devastating (*verheerend*). Why, it is better at the front.'

It must not be deduced from these and other cases I

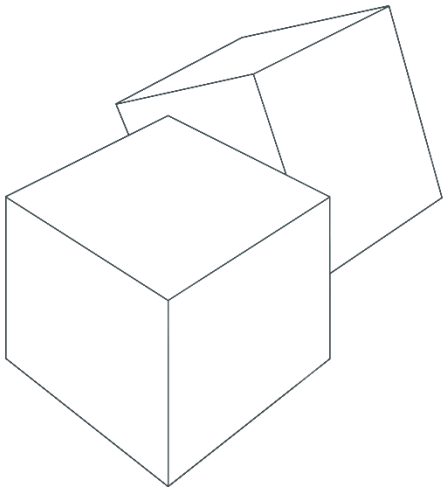
20

```
    <pb n="20" facs="0023.tif"/>
    <fw type="header" place="top">INSIDE GERMANY</fw>
    <lb/>
    Unzufriedenheit) among the workers in England as in<lb/>
    Germany. They also gave me the assurance that there<lb/>
    were more Communists in Hanover than ever before.<lb/>
    One uniformed official of many years' service, a veteran<lb/>
    of the last war, asked me where the United States stood<lb/>
    in the present struggle. I replied : 'We are 95 per<lb/>
    cent. opposed to your Government.' He answered<lb/>
    without an instant's hesitation: 'I honour you for<lb/>
    that!'<lb/>
</p><lb/>
<p><lb/>
    He then proceeded to voice his hostility to the<lb/>
    Government-to me, a complete stranger-with a<lb/>
    bitterness and violence that would have landed him in<lb/>
    prison had he been overheard. That he knew his<lb/>
    danger was plain when he tapped one of the walls of<lb/>
    the room and said : 'These have ears here.'<lb/>
</p><lb/>
<p><lb/>
    This readiness to confide in me appeared again and<lb/>
    again, and Americans living in Germany assured me<lb/>
    that they have similar experiences. Another man in<lb/>
    Nazi uniform seized the opportunity when his superior<lb/>
    left us alone to assure me that the people did not want the<lb/>
    war, that it was made by 'a few men higher up' ; that<lb/>
    the German workers did not hate the English people-<lb/>
    among whom he had lived for two years and whom he<lb/>
    liked and respected ; that he had been four years at<lb/>
    the front in 1914-18 and felt that there was no excuse<lb/>
    for another war, and that he and all his friends knew<lb/>
    that the German Press contained only lies and false<lb/>
    propaganda. The instant he heard his superior com-<lb/>
    ing back he returned to his former complete military<lb/>
    impassivity.<lb/>
</p><lb/>
<p><lb/>
    Still another case is extremely illuminating. An<lb/>
    officer on leave from the front remarked loudly in a<lb/>
    Frankfort tram that 'the spirit of the people here in<lb/>
    Frankfort toward the war is devastating (verheerend).<lb/>
    Why, it is better at the front.'<lb/>
</p><lb/>
<p><lb/>
    It must not be deduced from these and other cases I<lb/>
    <lb/>
```

DTAbf: Deutsches Textarchiv Basisformat

- Entwickelt vom Deutschen Textarchiv
  - Deutschsprachige Texte von 1600-1900
  - 3319 annotierte Volldigitalisate
- DTAbf: Erweiterung von TEI P5
  - Eliminierung vonTEI P5 Ambiguität
  - Erfassung von Inhalt und formaler Struktur


```
nicht ſagen/ das Chriſtus allein im Himmel vnd nicht bey vns<lb/>
<cit xml:id="bibl9" next="#quote12">
<bibl>
<note place="left">Pſal. 139.</note>
</bibl>
</cit>
auff Erden ſey/ ſintemahl er ſitzet zu der Rechten der krafft Got-<lb/>
tes/ welche ſich nicht theilen leſt/ ſondern allenthalben iſt/ vnd alles<lb/>
erfüllet. Deñ alſo ſtehet geſchrieben im 139 Pſalm.
<cit xml:id="quote12" prev="#bibl9">
<quote>Wo<lb/>
ſoll ich hingehn für deinem Geiſt/ vnd wo ſol ich hinfliehen für<lb/>
deinem Angeſicht? Fuhre ich gehn Himmel/ ſo biſtu da/ bettet<lb/>
ich mir in die Helle/ ſiehe ſo biſtu auch da/ Nehme ich flügel der<lb/>
Morgenröte/ vnd bliebe am euſſerſten Meer/ ſo würde mich doch<lb/>
deine Handt daſelbſt führen/ vnd deine Rechte mich halten:</quote>
</cit>
```

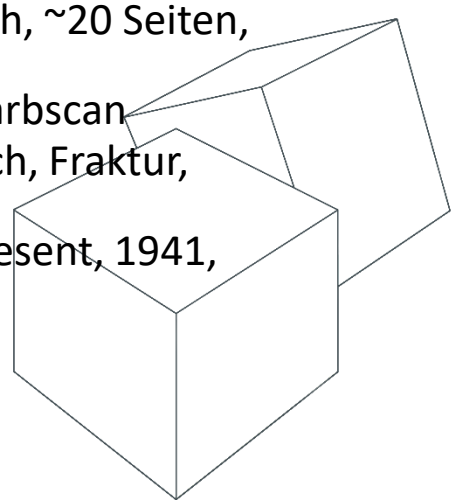


Universität Bielefeld

SFB 1288 PRAKTIKEN DES VERGLEICHENS

Teilnehmende Projekte

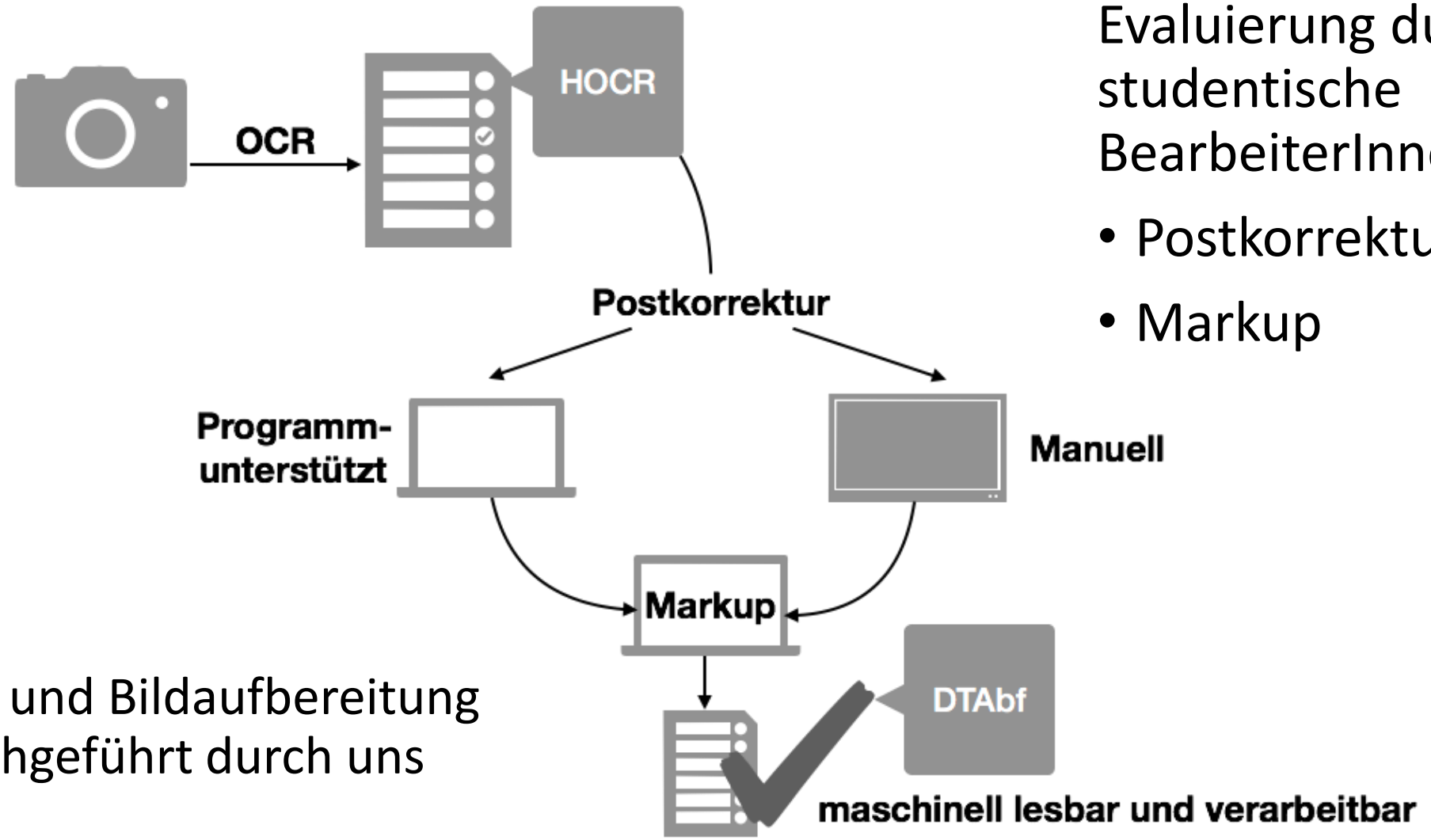
- A04 — Kulturvergleiche durch Zeitschriften
  - Die Gegenwart 1915-1917, deutsch, Fraktur, 300+ Seiten, Microfiche (2003)
- B01 — Welches Recht gilt für wen?
  - Mémoire a consulter, et consultation, 1768, französisch, 75 Seiten, Farbfotografien
  - Voyage dans les mers de l'Inde, 1779, französisch, 143 Seiten, Farbscan
  - Mœuers et coutumes des indiens, 1987, französisch, 129 Seiten, SW-Scan
- B03 — Travel is the school of comparison
  - Entdeckungs-Reise in die Süd-See und nach der Herings-Straße zur Erforschung einer nordöstlichen Durchfahrt, 1821, deutsch, Fraktur, ~290 Seiten, SW-Scan
- B05 — Der englische Roman als Labor
  - Castle Rackrent, an hiberian tale, 1880, englisch, 224 Seiten, SW-Scan
- B06 — Rechtliche Vergleichsverbote
  - Ausgewählte UN-Protokolle, 1972-2012, englisch/ französisch, ~300 Seiten, SW-Scans (Schreibmaschine, PDF als Officeexport)
- C01 — Das vergleichende Sehen
  - An essay on the theory of painting, 1725, englisch, ~290 Seiten, Farbscan
- C03 — Eine Begriffsgeschichte des Vergleichens
  - Moskau 1937 Ein Reisebericht für meine Freunde, 1937, 152 Seiten, Farbscan
  - One year of hitlerism, 1934, englisch, ~20 Seiten, Farbscan
  - Inside Germany, 1939, 90 Seiten, Farbscan
  - Das System Mussolini, 1924, deutsch, Fraktur, ~140 Seiten, SW-Scan
  - Black Record: Germans past and present, 1941, englisch, ~70 Seiten, Farbscan



Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

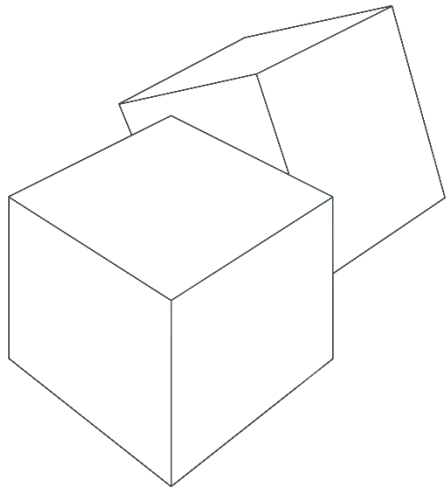
Digitalisierungspipeline



Evaluierung durch studentische BearbeiterInnen:

- Postkorrektur
- Markup

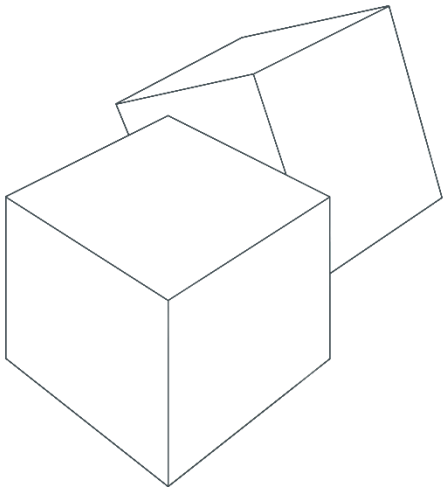
OCR und Bildaufbereitung durchgeführt durch uns



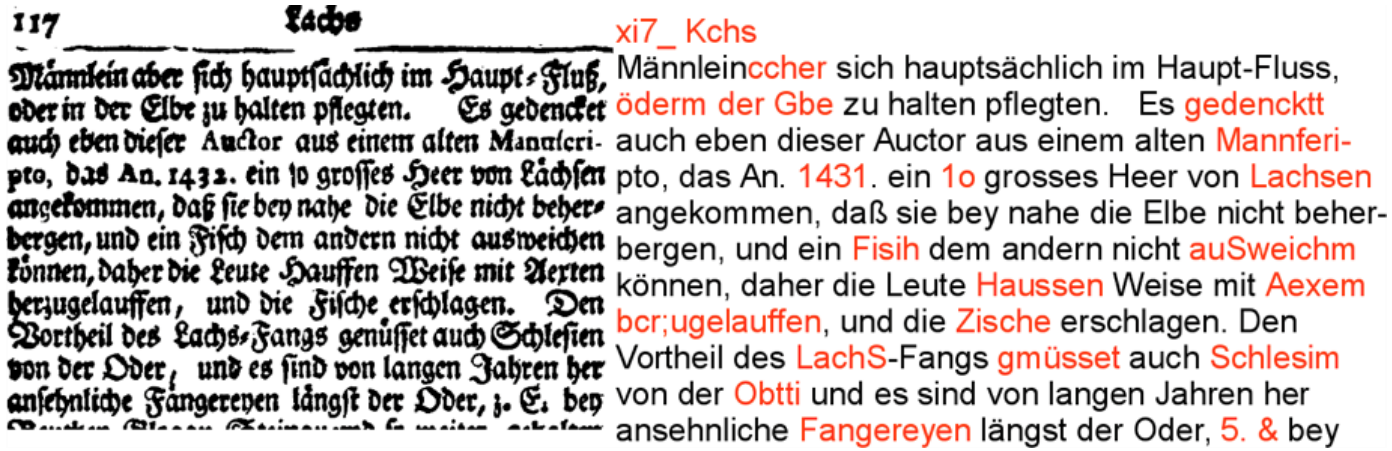
Ziele der Pilotphase Digitalisierung: Ursprungsmedium

Welche Eingabeformate eignen sich für eine tiefergehende Digitalisierung?
Welche Mindestanforderungen müssen erfüllt sein?

- Auflösung, Detailfülle
- Farbe, Schwarz-Weiss
- Maximale Verzerrung im Bild, Verschmutzungsgrad



Ziele der Pilotphase Digitalisierung: OCR



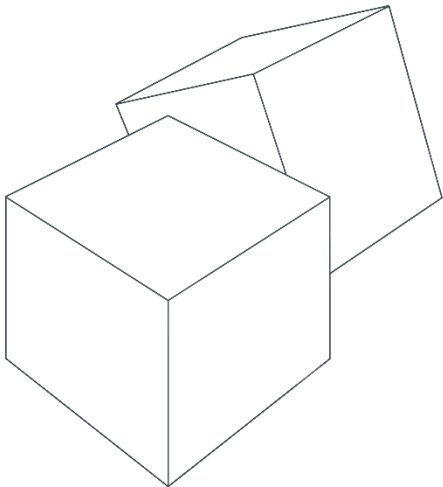
117 Lachs

Männlein aber ſich hauptſächlich im Haupt-Fluß, oder in der Elbe zu halten pflegten. Es gedencket auch eben dieſer Auctor aus einem alten Manuſcripto, das An. 1432. ein ſo groſſes Heer von Lachſen angekommen, daß ſie bey nahe die Elbe nicht beherbergen, und ein Fiſch dem andern nicht ausweichen können, daher die Leute Hauffen Weiſe mit Aexten herzugelauffen, und die Fiſche erſchlagen. Den Vortheil des Lachs-Fangs genüſſet auch Schleſien von der Oder, und es ſind von langen Jahren her anſehnliche Fängereyen längſt der Oder, z. E. bey

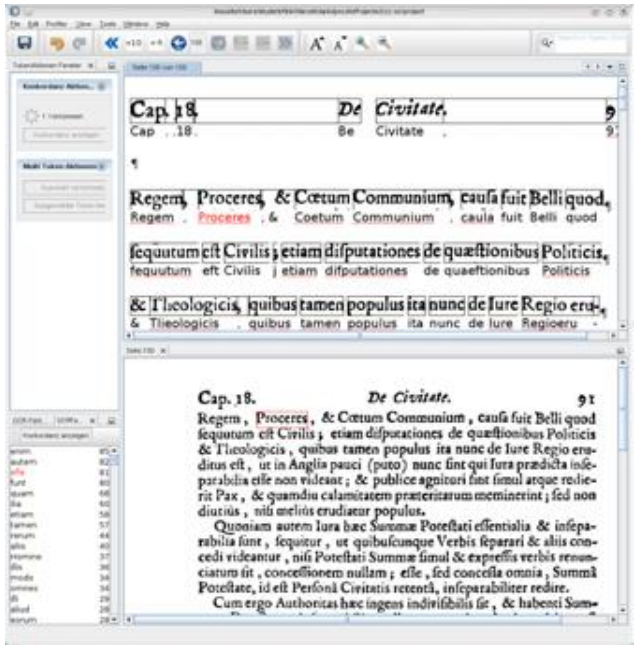
xi7_ Kchs
Männleinccher sich hauptsächlich im Haupt-Fluss, öderm der Gbe zu halten pflegten. Es gedencktt auch eben dieser Auctor aus einem alten Mannferi- pto, das An. 1431. ein 1o grosses Heer von Lachsen angekommen, daß sie bey nahe die Elbe nicht beher- bergen, und ein Fisih dem andern nicht auSweichm können, daher die Leute Haussen Weise mit Aexem bcr;ugelauffen, und die Zische erschlagen. Den Vortheil des LachS-Fangs gmüsset auch Schlesim von der Obtti und es sind von langen Jahren her ansehnliche Fangereyen längst der Oder, 5. & bey

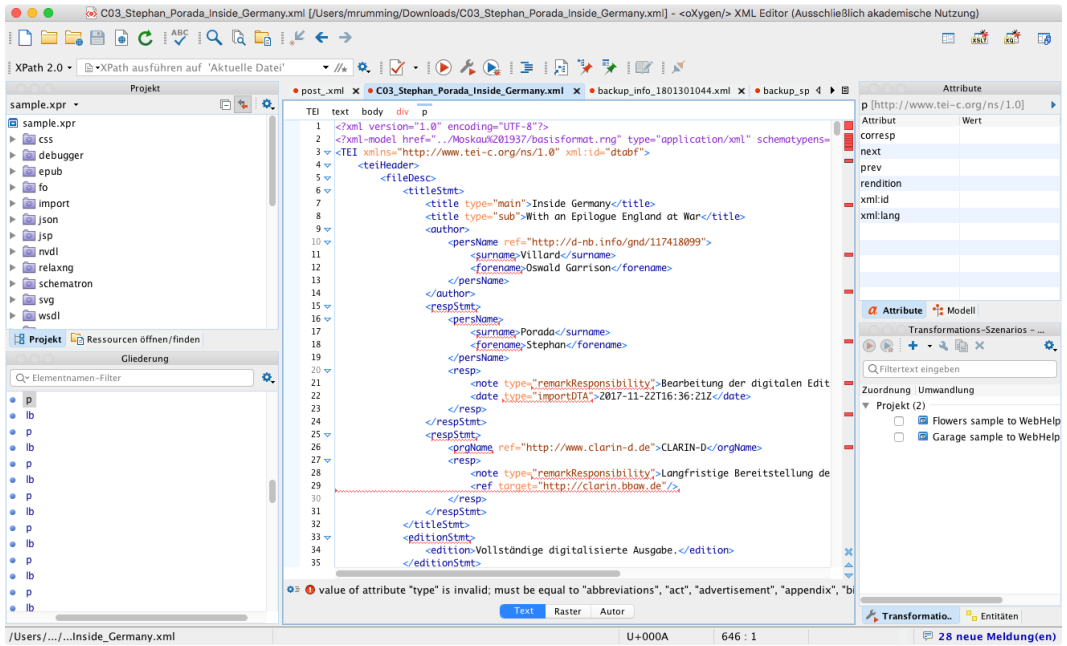
Welche open source OCR-Engine liefert die besten Ergebnisse?
Welches freie Erkennungsmodel ist am besten?

- ocropy: en, french_balzac, de, de_frak(2x), antiqua
- tesseract: deu, deu_fra, fra, frm, eng, enm

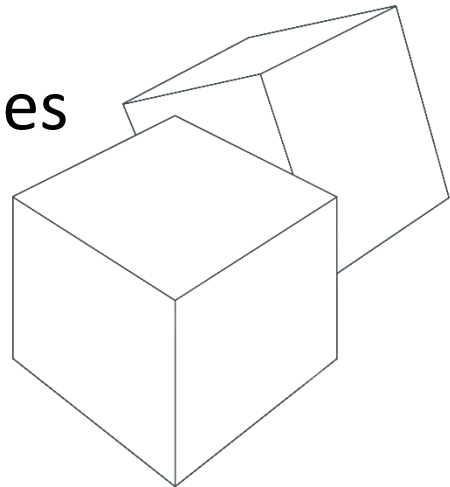


Ziele der Pilotphase Digitalisierung: Postkorrektur und Markup



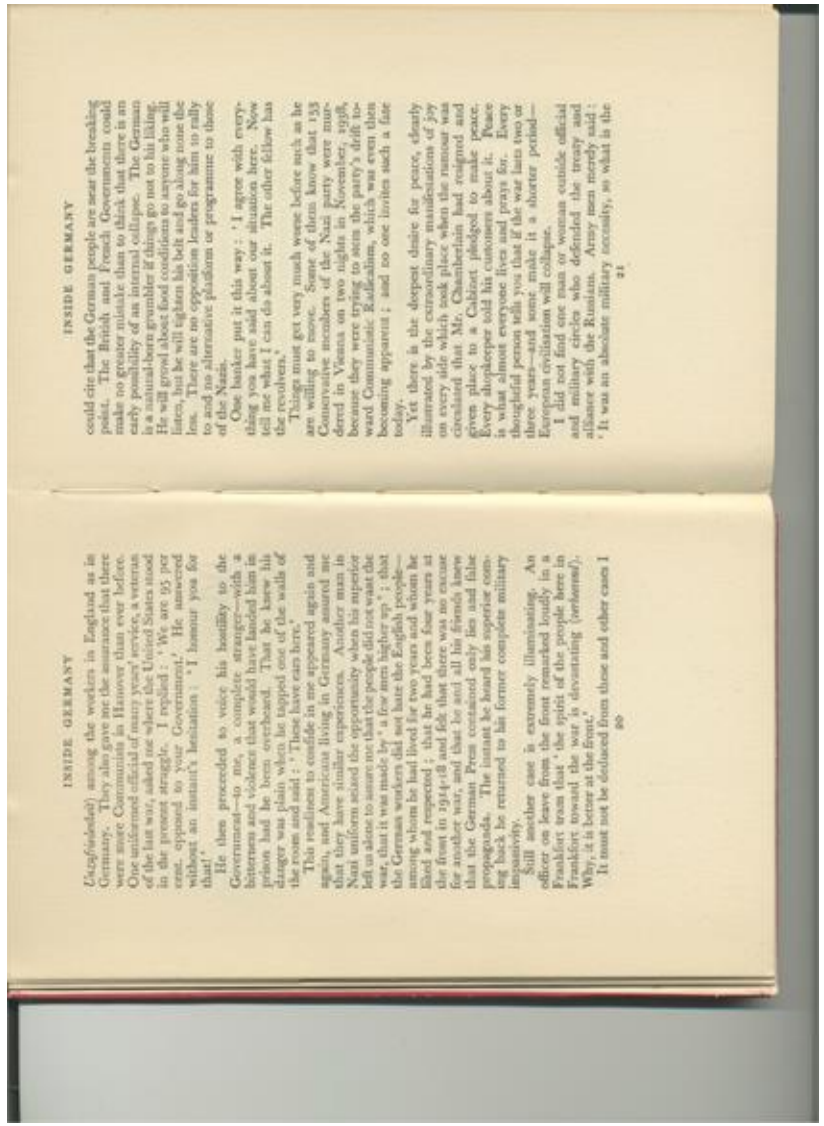


- Programmunterstützte Postkorrektur und danach anschließendes Markup
- Postkorrektur und Markup direkt in einem Schritt
- „Abtippen“ und Markup schneller als vorheriges OCR?



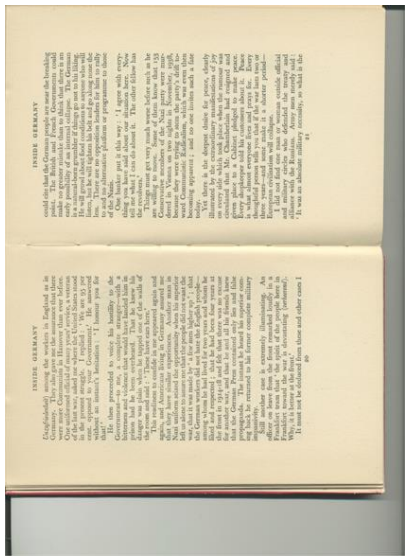
C03: Eine Begriffsgeschichte des Vergleichens

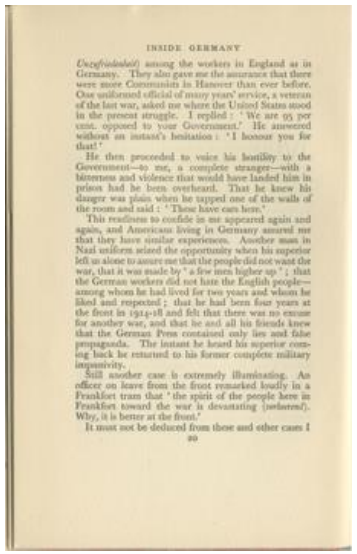
- Doppelseitiger Farbscan
- Einheitlicher Bildausschnitt ermöglicht:
  - automatisches Drehen
  - automatisches Zuschneiden

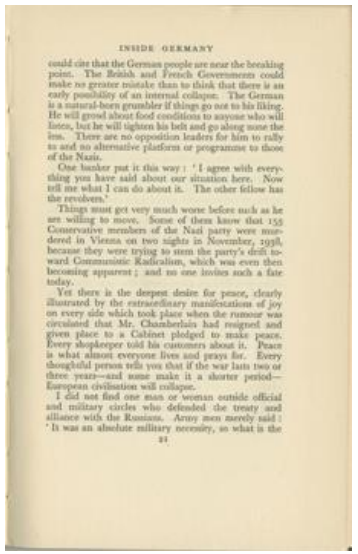


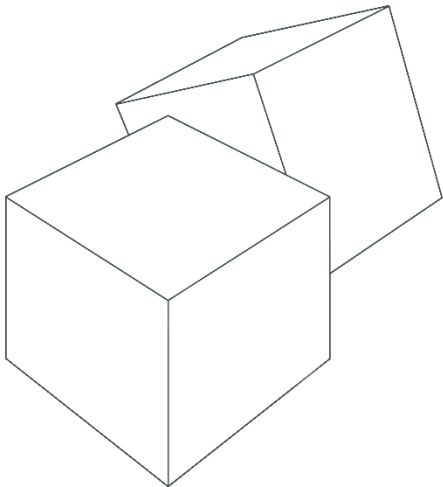
C03: Eine Begriffsgeschichte des Vergleichens

- Doppelseitiger Farbscan
- Einheitlicher Bildausschnitt ermöglicht:
  - automatisches Drehen
  - automatisches Zuschneiden









A04: Kulturvergleiche durch Zeitschriften

- Microfichefotografie
  - Kontrastarm
  - Bildkontamination
- Zweispaltiges Seitenlayout
- „schiefe“ Bilder
- Frakturschrift

A04: Kulturvergleiche durch Zeitschriften

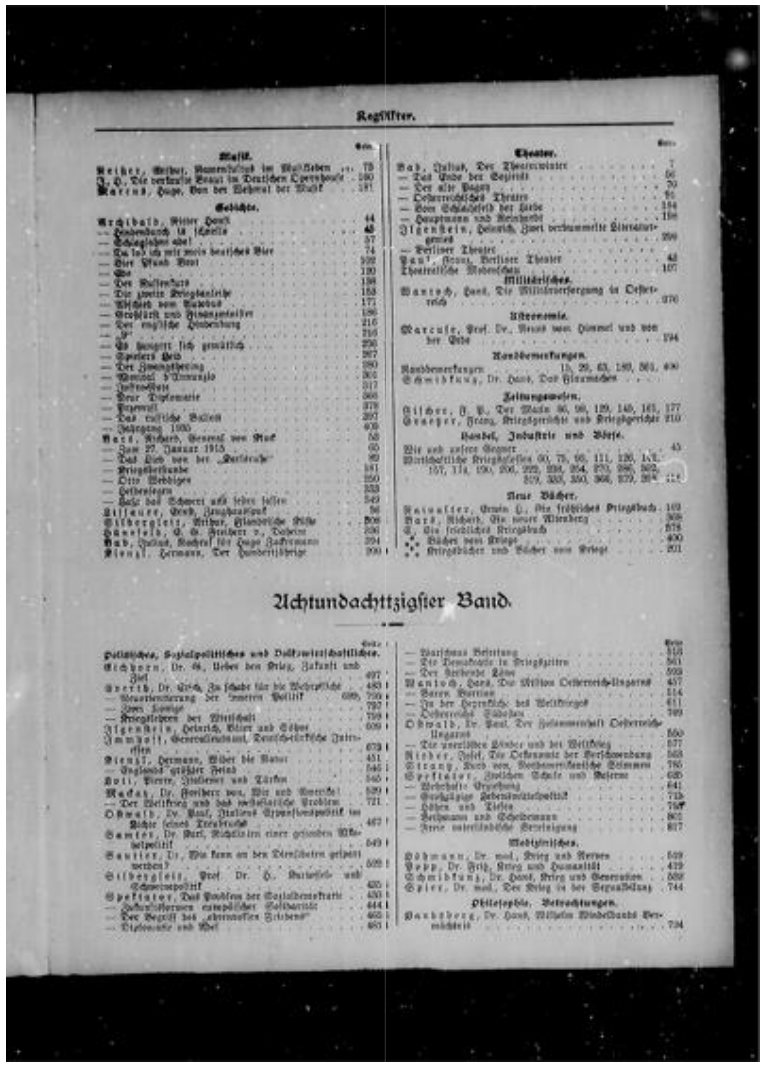
- Microfichefotografie
  - Kontrastarm
  - Bildkontamination
- Zweispaltiges Seitenlayout
- „schiefe“ Bilder
- Frakturschrift

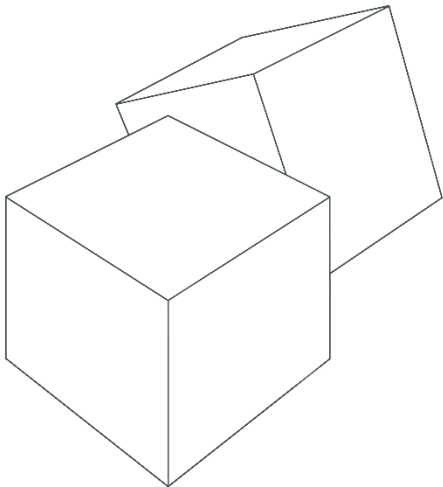
Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

A04: Kulturvergleiche durch Zeitschriften

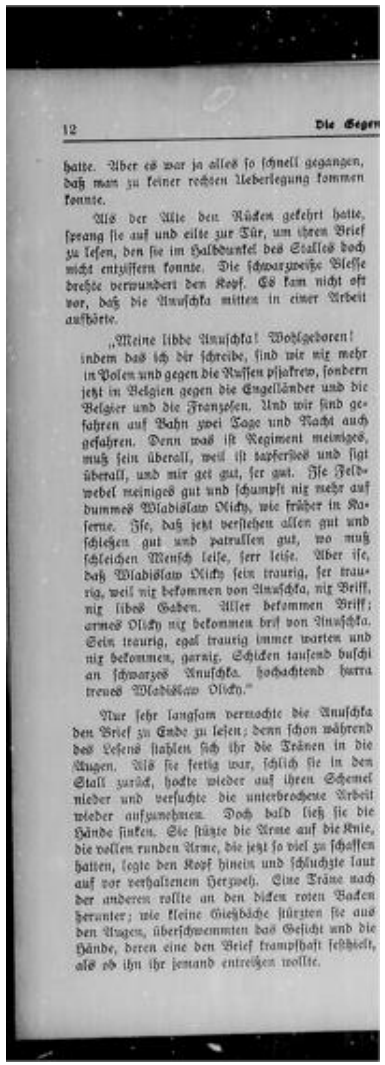
- Microfichefotografie
  - Kontrastarm
  - Bildkontamination
- Zweispaltiges Seitenlayout
- „schiefe“ Bilder
- Frakturschrift





A04: Kulturvergleiche durch Zeitschriften

- Microfichefotografie
  - Kontrastarm
  - Bildkontamination
- Zweispaltiges Seitenlayout
- „schiefe“ Bilder
- Frakturschrift



12 Die Gegen

hatte. Aber es war ja alles so schnell gegangen, daß man zu keiner rechten Ueberlegung kommen konnte.

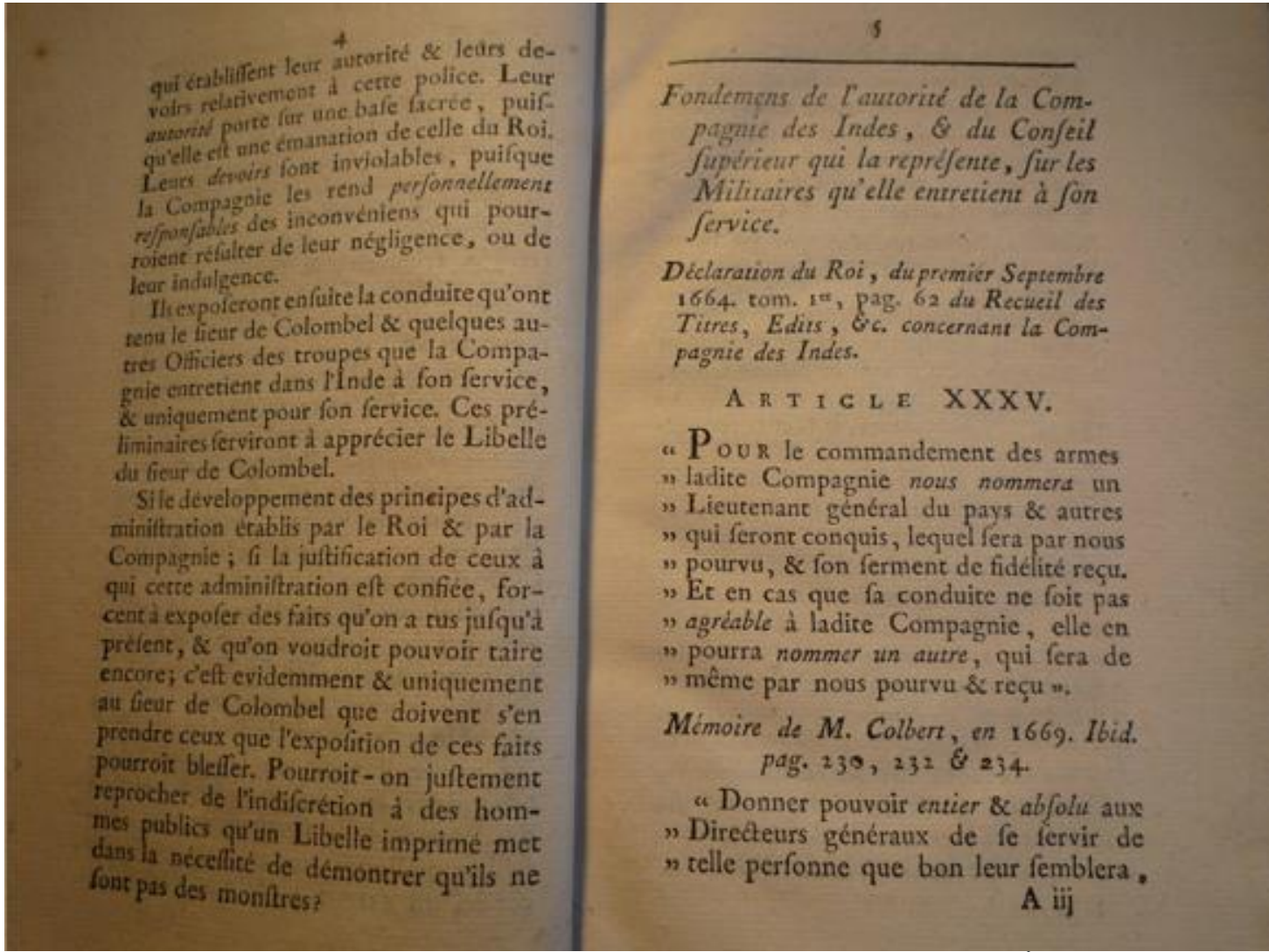
Als der Alte den Rücken gekehrt hatte, sprang sie auf und eilte zur Tür, um ihren Brief zu lesen, den sie im Halbdunkel des Stalles doch nicht entziffern konnte. Die schwarzweiße Bliesse drehte verwundert den Kopf. Es kam nicht oft vor, daß die Anuschka mitten in einer Arbeit aufhörte.

„Meine libbe Anuschka! Wohlgeboren! indem daß ich dir schreibe, sind wir nig mehr in Polen und gegen die Russen pjiakrew, sondern jetzt in Belgien gegen die Engelländer und die Belgier und die Franzosen. Und wir sind gefahren auf Bahn zwei Tage und Nacht auch gefahren. Denn was ist Regiment meiniges, muß sein überall, weil ist sapferdes und sigt überall, und mir get gut, ser gut. Ise Feldwebel meiniges gut und schumpft nig mehr auf dummes Wladislaw Nicky, wie früher in Kaserne. Ise, daß jetzt verstehen allen gut und schlesen gut und patrullen gut, wo muß schleichen Mensch leise, ser leise. Aber ise, daß Wladislaw Nicky sein traurig, ser traurig, weil nig bekommen von Anuschka, nig Briff, nig libes Gaben. Aller bekommen Briff; armes Nicky nig bekommen brif von Anuschka. Sein traurig, egal traurig immer warten und nig bekommen, garnig. Schicken tausend buschi an schwarzes Anuschka. hochachtend hurra treues Wladislaw Nicky."

Nur sehr langsam vermochte die Anuschka den Brief zu Ende zu lesen; denn schon während des Lesens stahlen sich ihr die Tränen in die Augen. Als sie fertig war, schlich sie in den Stall zurück, hockte wieder auf ihren Schemel nieder und versuchte die unterbrochene Arbeit wieder aufzunehmen. Doch bald ließ sie die Hände sinken. Sie stützte die Arme auf die Knie, die vollen runden Arme, die jetzt so viel zu schaffen hatten, legte den Kopf hinein und schluchzte laut auf vor verhaltenem Herzweh. Eine Träne nach der anderen rollte an den dicken roten Backen herunter; wie kleine Gießbäche stürzten sie aus den Augen, überschwemmten das Gesicht und die Hände, deren eine den Brief krampfhaft festhielt, als ob ihn ihr jemand entreißen wollte.

B01: Welches Recht gilt für wen?

- Doppelseitige Farbfotografie
- Ausschnitt von Bild zu Bild variierend
- Text ist stark verzerrt



4

qui établiſſent leur autorité & leurs de-
voirs relativement à cette police. Leur
autorité porte ſur une baſe ſacrée, puiſ-
qu'elle eſt une émanation de celle du Roi.
Leurs *devoirs* ſont inviolables, puiſque
la Compagnie les rend *perſonnellement*
reſponſables des inconvéniens qui pour-
roient réſulter de leur négligence, ou de
leur indulgence.

Ils expoſeront enſuite la conduite qu'ont
tenu le ſieur de Colombel & quelques au-
tres Officiers des troupes que la Compa-
gnie entretient dans l'Inde à ſon ſervice,
& uniquement pour ſon ſervice. Ces pré-
liminaires ſerviront à apprécier le Libelle
du ſieur de Colombel.

Si le développement des principes d'ad-
miniſtration établis par le Roi & par la
Compagnie ; ſi la juſtification de ceux à
qui cette adminiſtration eſt confiée, for-
cent à expoſer des faits qu'on a tus juſqu'à
préſent, & qu'on voudroit pouvoir taire
encore ; c'eſt évidemment & uniquement
au ſieur de Colombel que doivent s'en
prendre ceux que l'expoſition de ces faits
pourroit bleſſer. Pourroit-on juſtement
reprocher de l'indiſcrétion à des hom-
mes publics qu'un Libelle imprimé met
dans la néceſſité de démontrer qu'ils ne
ſont pas des monſtres ?

5

*Fondemens de l'autorité de la Com-
pagnie des Indes, & du Conſeil
ſupérieur qui la repréſente, ſur les
Militaires qu'elle entretient à ſon
ſervice.*

*Déclaration du Roi, du premier Septembre
1664.* tom. 1^er^, pag. 62 *du Recueil des
Titres, Edits, &c. concernant la Com-
pagnie des Indes.*

ARTICLE XXXV.

« POUR le commandement des armes
» ladite Compagnie *nous nommera* un
» Lieutenant général du pays & autres
» qui ſeront conquis, lequel ſera par nous
» pourvu, & ſon ſerment de fidélité reçu.
» Et en cas que ſa conduite ne ſoit pas
» *agréable* à ladite Compagnie, elle en
» pourra *nommer un autre*, qui ſera de
» même par nous pourvu & reçu ».

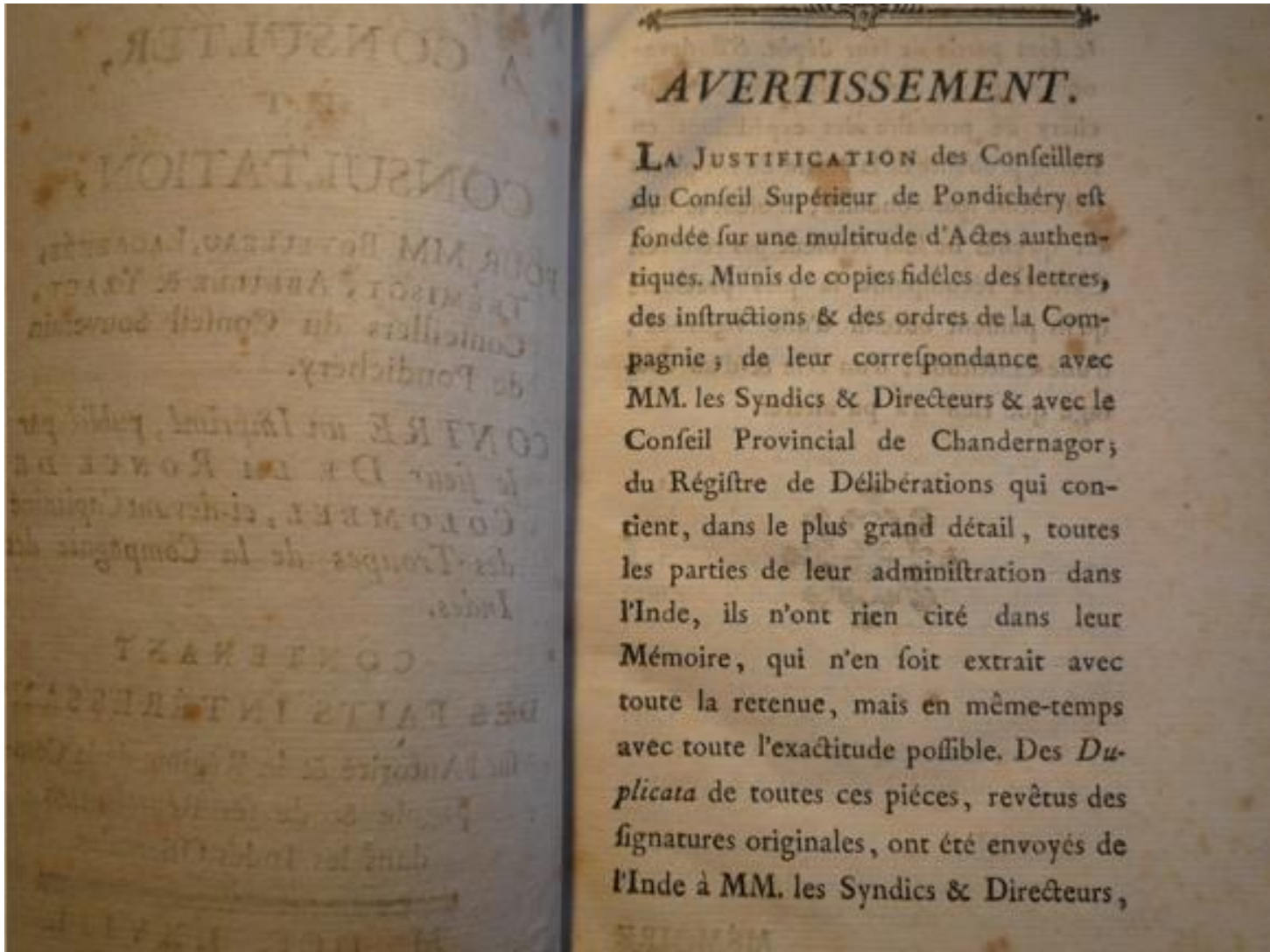
*Mémoire de M. Colbert, en 1669. Ibid.
pag. 230, 232 & 234.*

« Donner pouvoir *entier* & *abſolu* aux
» Directeurs généraux de ſe ſervir de
» telle perſonne que bon leur ſemblera,

A iij

B01: Welches Recht gilt für wen?

- Doppelseitige Farbfotografie
- Ausschnitt von Bild zu Bild variierend
- Text ist stark verzerrt

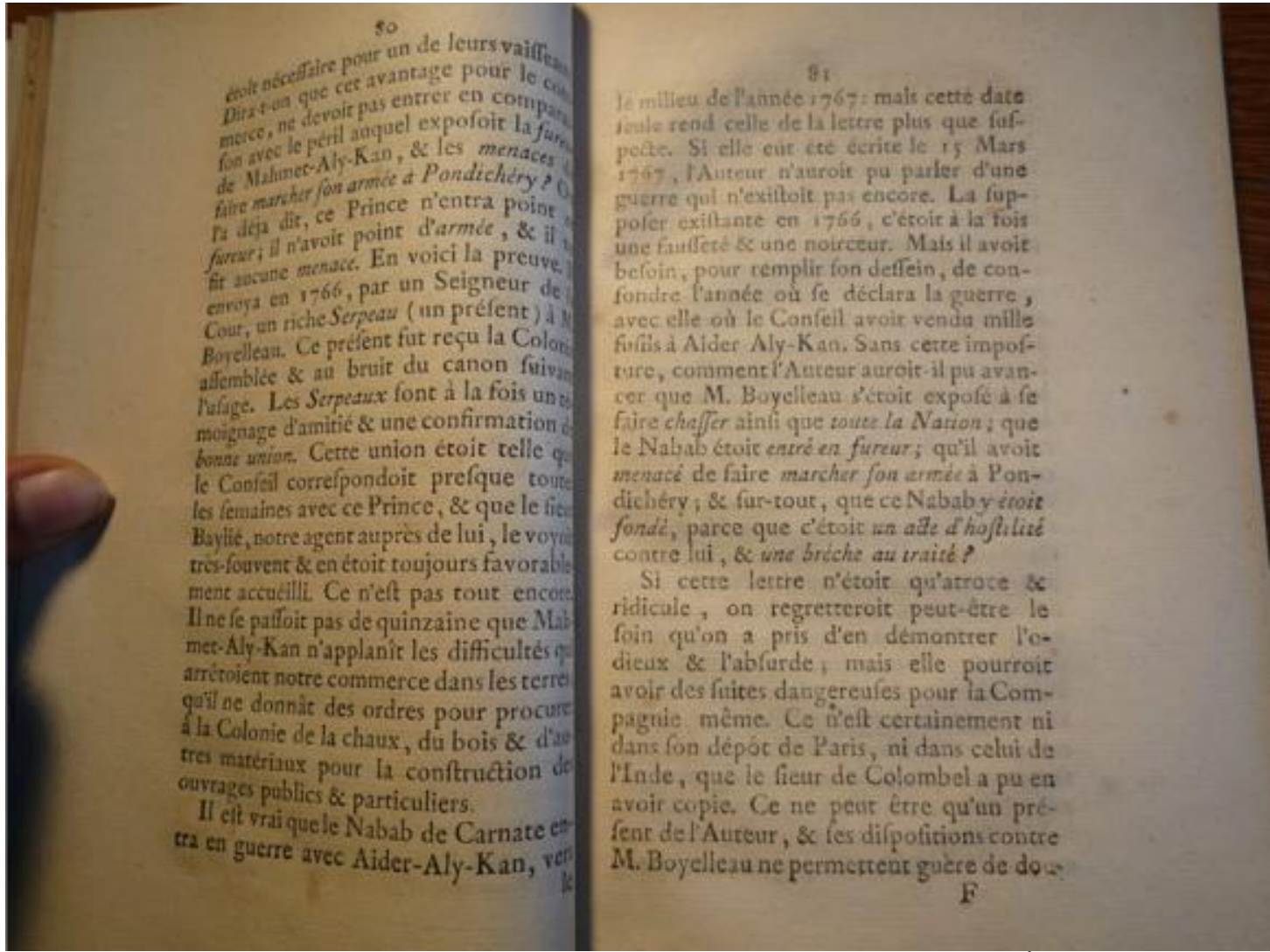


AVERTISSEMENT.

LA JUSTIFICATION des Conſeillers du Conſeil Supérieur de Pondichéry eſt fondée ſur une multitude d'Actes authentiques. Munis de copies fidéles des lettres, des inſtructions & des ordres de la Compagnie ; de leur correſpondance avec MM. les Syndics & Directeurs & avec le Conſeil Provincial de Chandernagor ; du Régiſtre de Délibérations qui contient, dans le plus grand détail, toutes les parties de leur adminiſtration dans l'Inde, ils n'ont rien cité dans leur Mémoire, qui n'en ſoit extrait avec toute la retenue, mais en même-temps avec toute l'exactitude poſſible. Des *Duplicata* de toutes ces piéces, revêtus des ſignatures originales, ont été envoyés de l'Inde à MM. les Syndics & Directeurs,

B01: Welches Recht gilt für wen?

- Doppelseitige Farbfotografie
- Ausschnitt von Bild zu Bild variierend
- Text ist stark verzerrt



Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

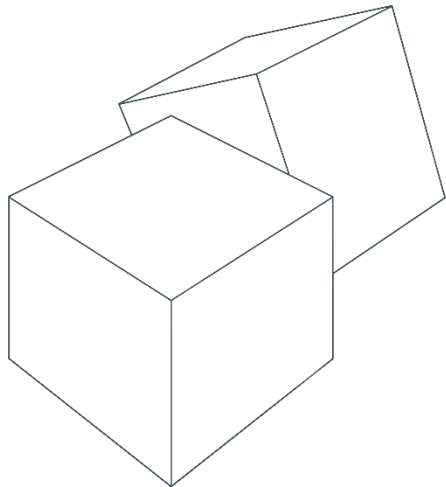
Ergebnisse: OCR

OCR-Pipeline: Kombination aus ocropy und tesseract

- ocropy: Binärisierung der Eingabebilder
- tesseract: Texterkennung

Texterkennungsmodelle:

- Fraktur kein Problem
- Englisches Sprachmodell besser als deutsches (Type: Antiqua)
- Französisch problematisch
- Handschriften nicht untersucht



Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

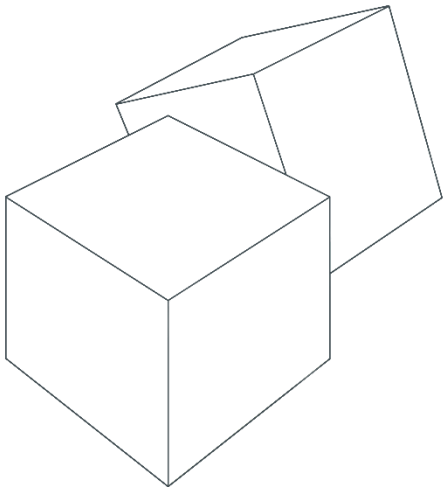
Ergebnisse: Postkorrektur und Markup

PoCoTo fehlerbehaftet

- Support wurde eingestellt
- Ersatz wird entwickelt (webbasiert)
- Ausgabeformate fehlerhaft

Steile Lernkurve für Markup

- Im weiteren Verlauf sehr flüssige Umsetzung
- Tabellen und Listen weiterhin aufwendig umzusetzen



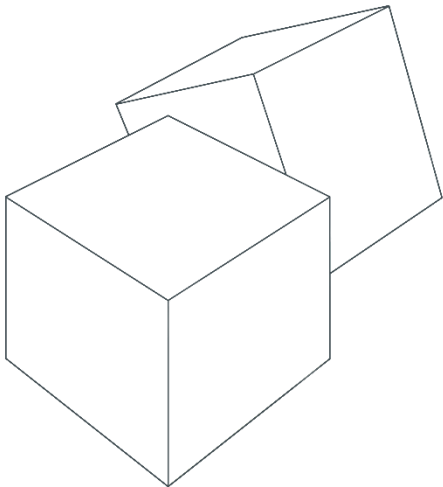
Fazit: Eingabeformate

Bildformate: TIFF, PNG, JPEG (möglichst verlustfrei komprimiert)
oder
PDF mit Bildern

Auflösung nicht aussagekräftig, sondern Punktdichte

- 300ppi
- Große Zeichen (wenig Inhalt pro Seite): 150ppi
- Viel Bildinhalt (A04): 450ppi/ 600ppi

Möglichst gleichbleibender Bildaufbau ohne große Verzerrungen



Fazit: Workflow

Postkorrektur

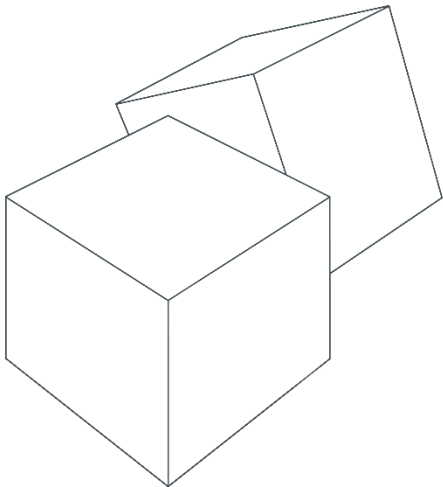
- Bei Fehlertoleranz: Unnötig für rein quantitative Analyse
- Direkt im reinen vorausgezeichneten Textformat

Notwendigkeit für Markup je nach Anwendungsfall

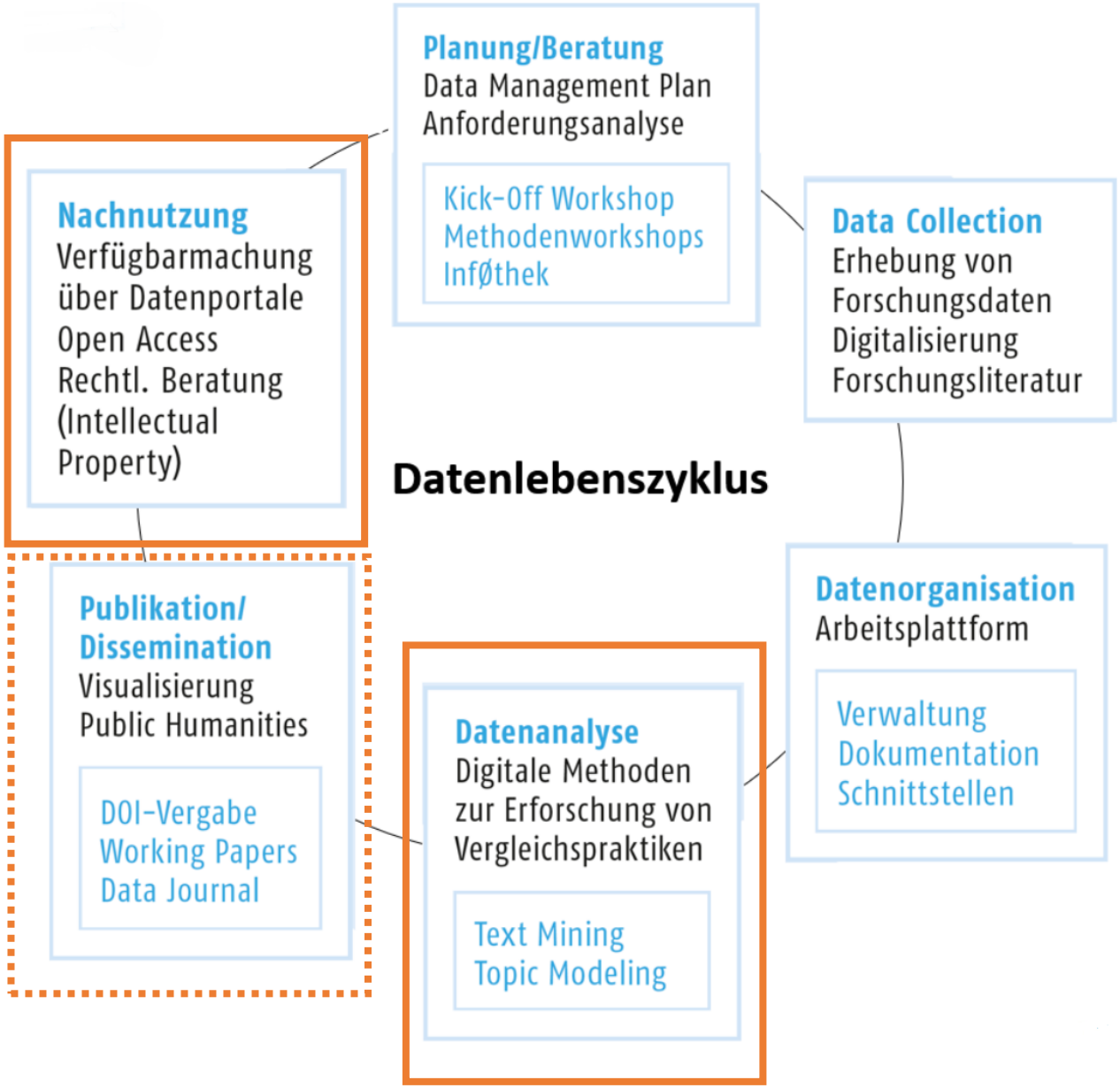
- Korrektur (vorausgezeichnetes Textformat) und Markup in einem Arbeitsschritt möglich
- Kosten-Nutzen-Frage

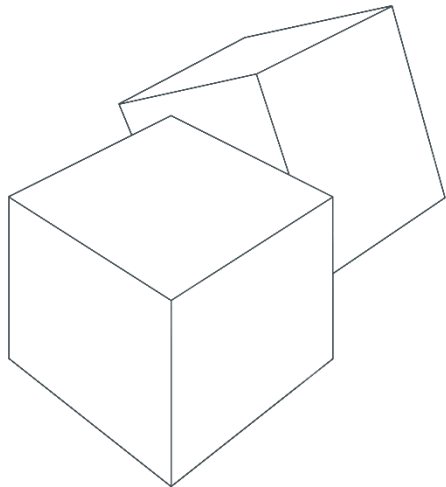
Arbeitsaufwand (Normseite):

- Postkorrektur: 10-15 min
- Markup: 5 min

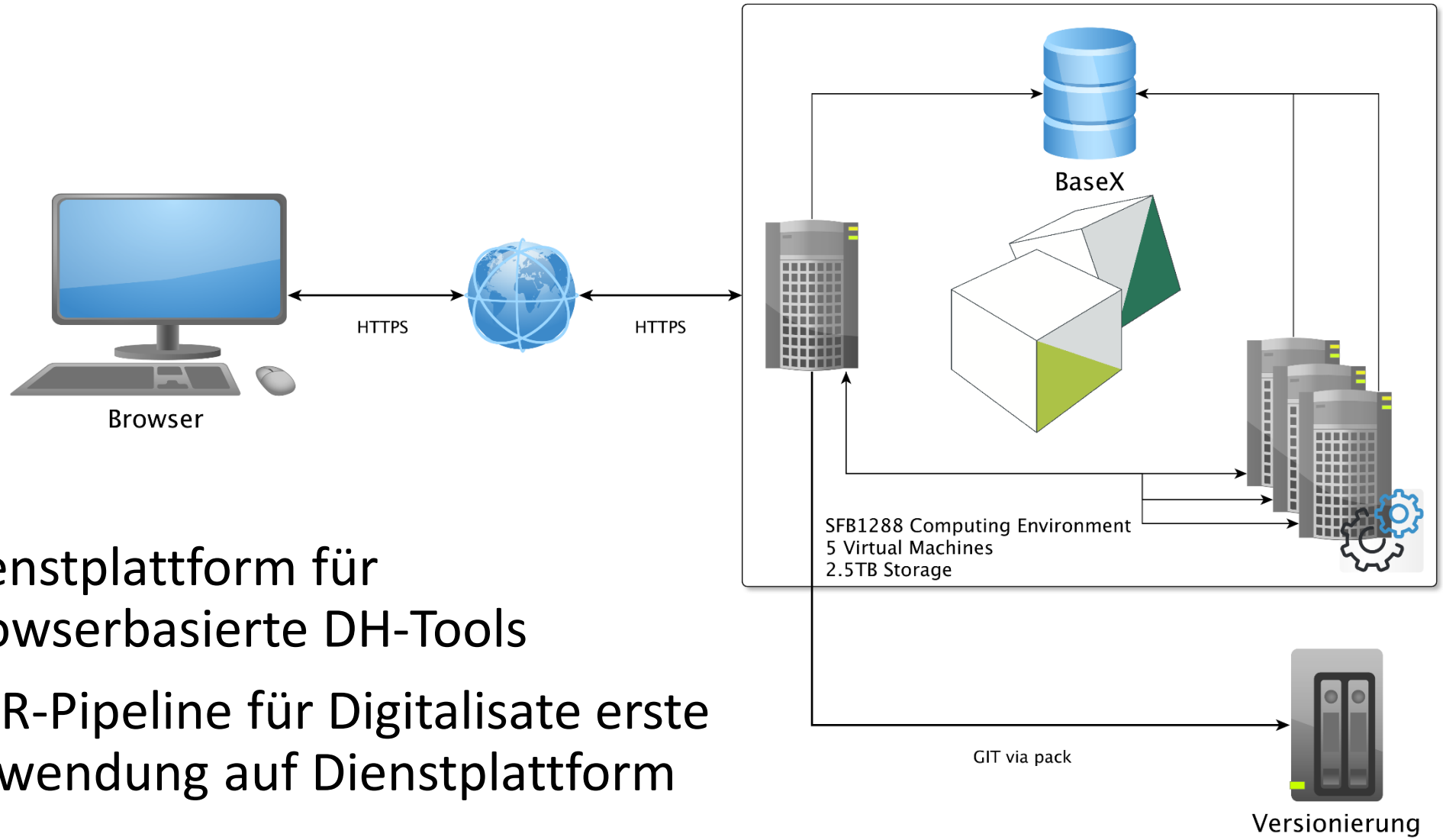


Ausblick


Planung/Beratung
Data Management Plan
Anforderungsanalyse
Kick-Off Workshop
Methodenworkshops
Infθthek
Data Collection
Erhebung von
Forschungsdaten
Digitalisierung
Forschungsliteratur
Nachnutzung
Verfügbarmachung
über Datenportale
Open Access
Rechtl. Beratung
(Intellectual
Property)
Datenlebenszyklus
Datenorganisation
Arbeitsplattform
Verwaltung
Dokumentation
Schnittstellen
Publikation/
Dissemination
Visualisierung
Public Humanities
DOI-Vergabe
Working Papers
Data Journal
Datenanalyse
Digitale Methoden
zur Erforschung von
Vergleichspraktiken
Text Mining
Topic Modeling



Ausblick



- Dienstplattform für browserbasierte DH-Tools
- OCR-Pipeline für Digitalisate erste Anwendung auf Dienstplattform

Universität Bielefeld

SFB 1288
PRAKTIKEN
DES VERGLEICHENS

VIELEN DANK!