



**Staatsbibliothek
zu Berlin**
Preußischer Kulturbesitz

Chasing Unicorns and Vampires in a Library

Visuelle Exploration der Digitalisierten Sammlungen

Dr.-Ing. David Zellhöfer, Informations- und Datenmanagement
Berlin, 04.04.2018

Motivation



**Staatsbibliothek
zu Berlin**
Preußischer Kulturbesitz

Die Staatsbibliothek zu Berlin ist die größte wissenschaftliche Bibliothek des deutschsprachigen Raums.

In 350+ Jahren wurden mehr als 13 Millionen verschiedene Medien angesammelt

133.837 digitalisierte Werke mit ca. 16 Millionen Einzelseiten unter <http://digital.staatsbibliothek-berlin.de/>

Die Staatsbibliothek zu Berlin



**Staatsbibliothek
zu Berlin**
Preußischer Kulturbesitz

Großteil der Werke lief durch die OCR, damit sind
Bildpositionen bekannt





**NUR WAS IST AUF DEN
BILDERN DARGESTELLT?**

Beispiele



Staatsbibliothek
zu Berlin

Preußischer Kulturbesitz

23

Vierter Streich.

Also lautet ein Beschluß:
Daß der Mensch was lernen muß. —
Nicht allein das A-B-C
Bringt den Menschen in die Höh';
Nicht allein in Schreiben, Lesen
Nicht sich ein vernünftig Wesen;
Nicht allein in Rechnungsachen
Soll der Mensch sich Nütze machen;
Sondern auch der Weisheit Lehren
Muß man mit Vergnügen hören. —



Daß dies mit Verstand geschah,
War Herr Lehrer Lämpel da. —
— Max und Moritz, diese beiden,
Wochten ihn darum nicht leiden;
— Wer hies böse Streiche macht
— Auf den Lehrer Acht.



Der Vampyr (Vespertilio, Vampyr).

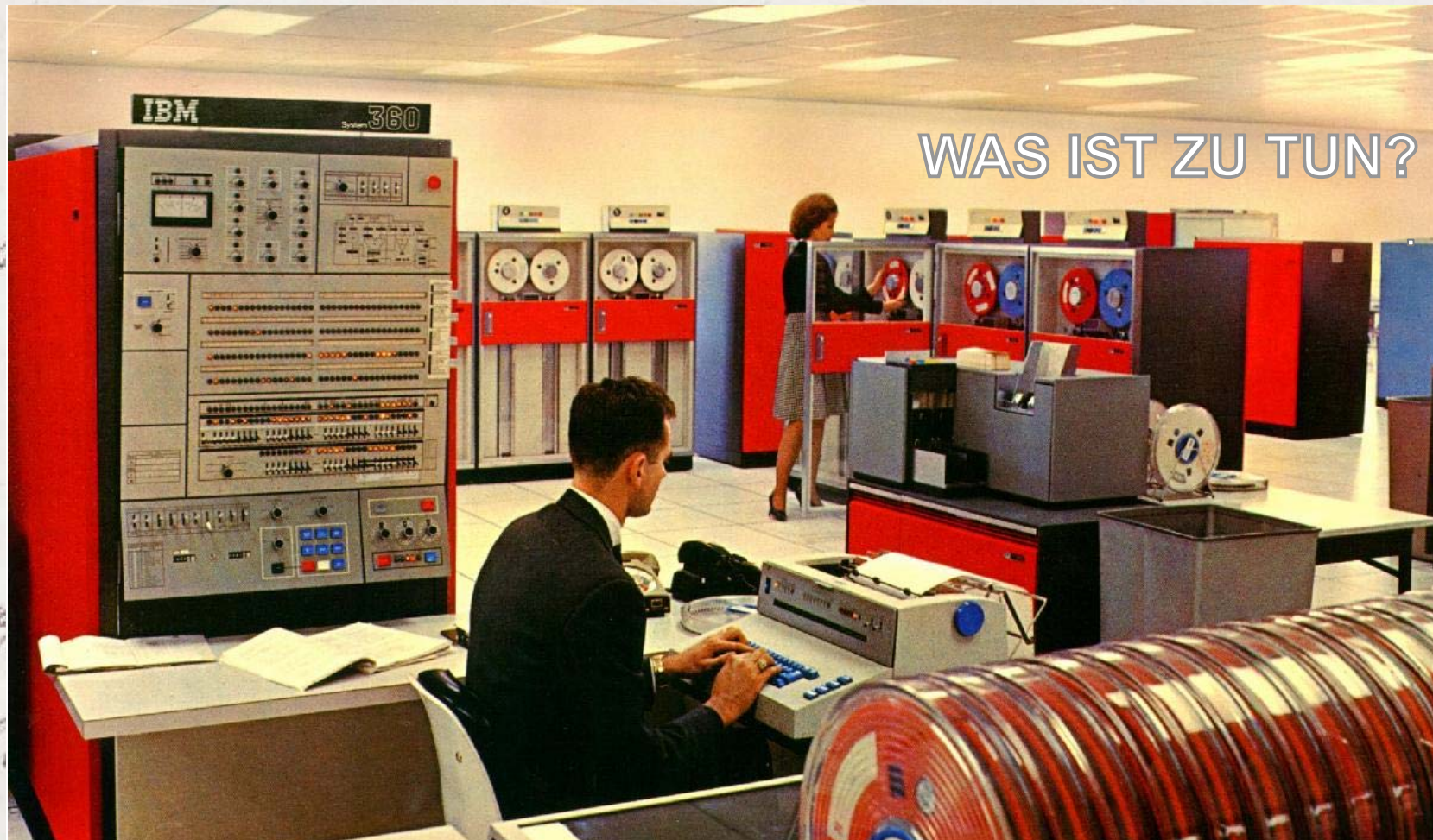
Der Vampyr, oder fliegende Hund, ist eine entfernte Fledermausart, welche schlafenden Menschen das Blut ausaugt und in Amerika und Ostindien, wiewohl mangelhaft, die Stelle der Vampyr von Marfchner.



Vampyr benennt man kurz und gut,
Was sich ernährt von Menschenblut.
Zu dieser Ordnung zählt daher
Vor Allem auch der Vampyr.
Den, wenn er nimmt nur 10 Prozent,
Gewöhnlich Privatje man nennt.



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz



WAS IST ZU TUN?

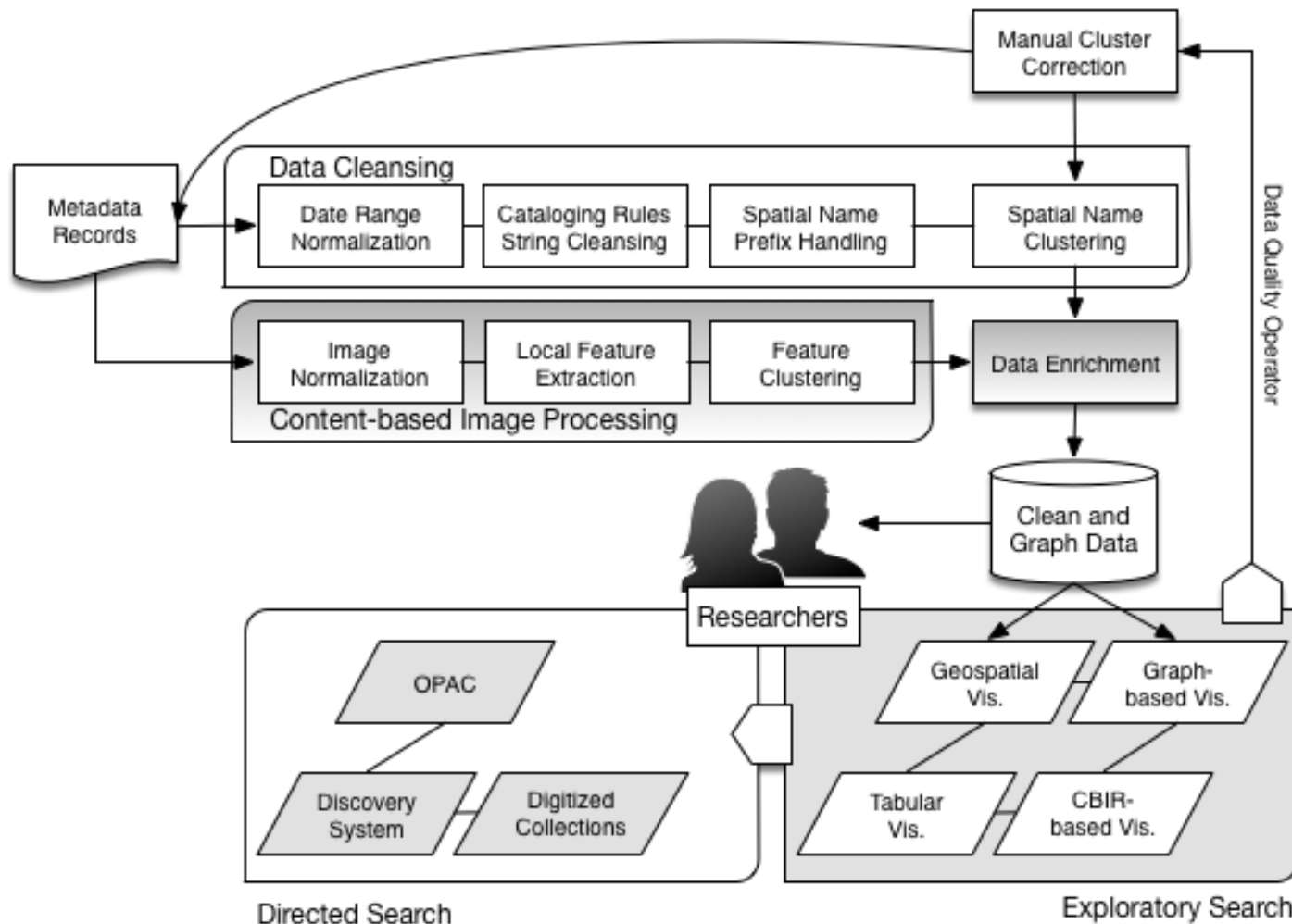
IBM

Der Workflow



Staatsbibliothek
zu Berlin

Preußischer Kulturbesitz

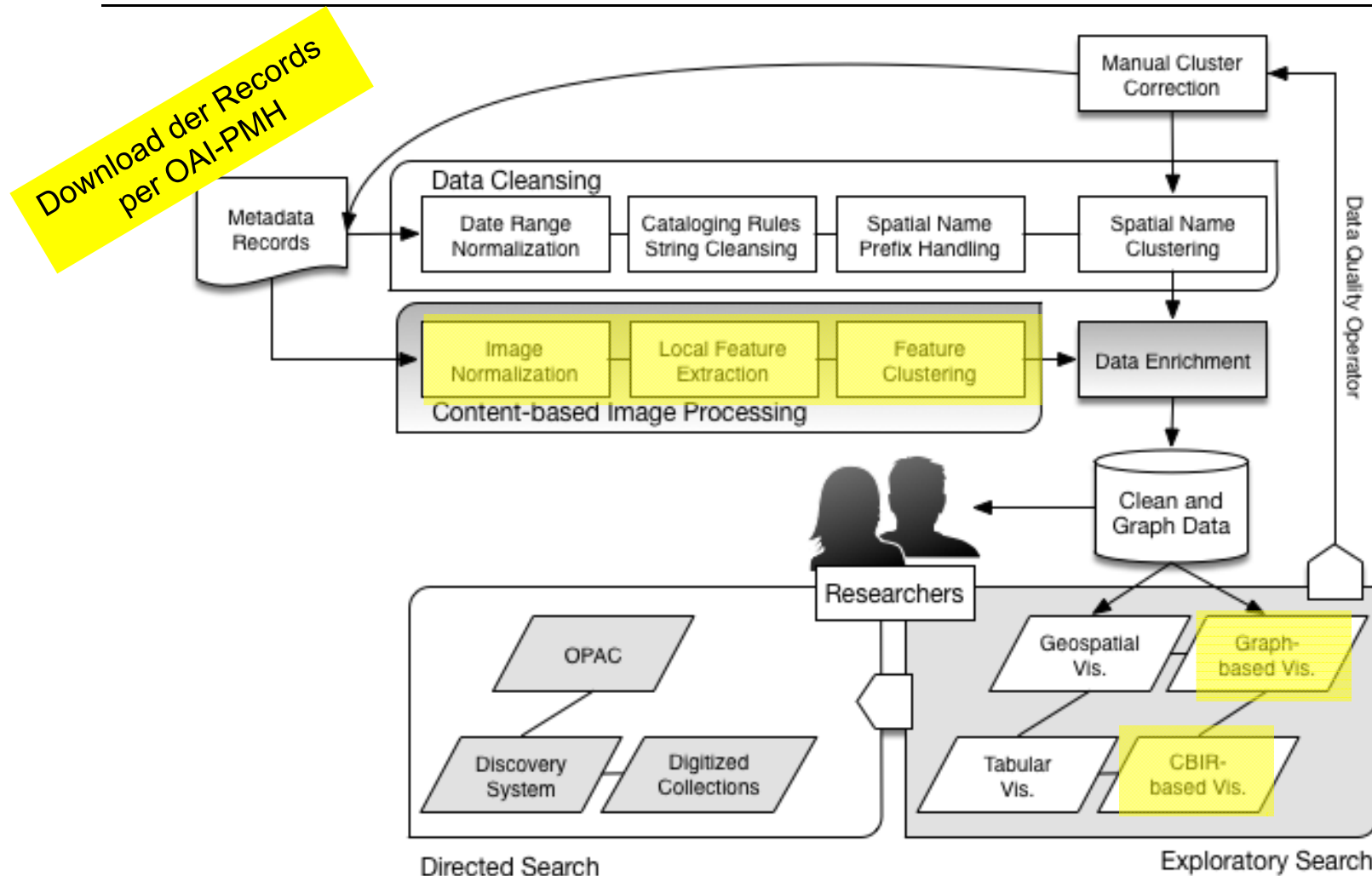


Der Workflow



Staatsbibliothek
zu Berlin

Preußischer Kulturbesitz

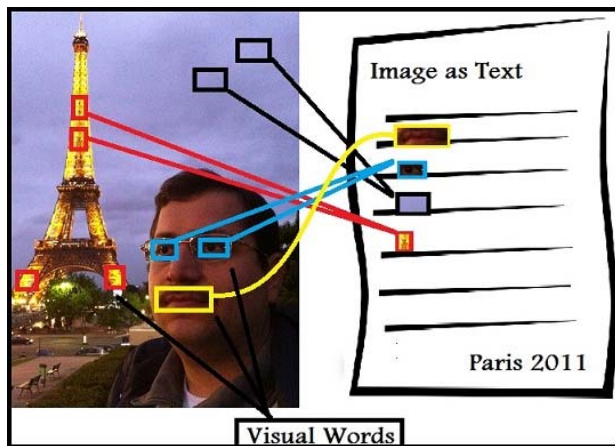


Content-Based Image Processing

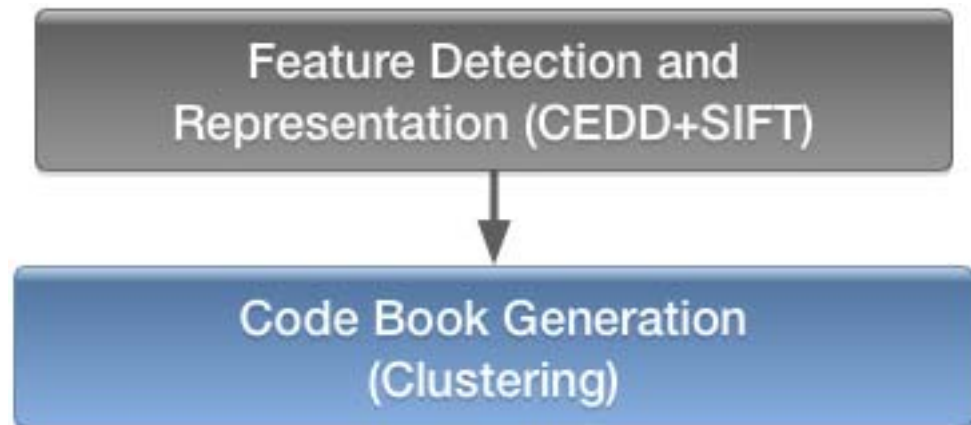
Download als TIFF oder JPEG

Normalisierung mit ImageMagick

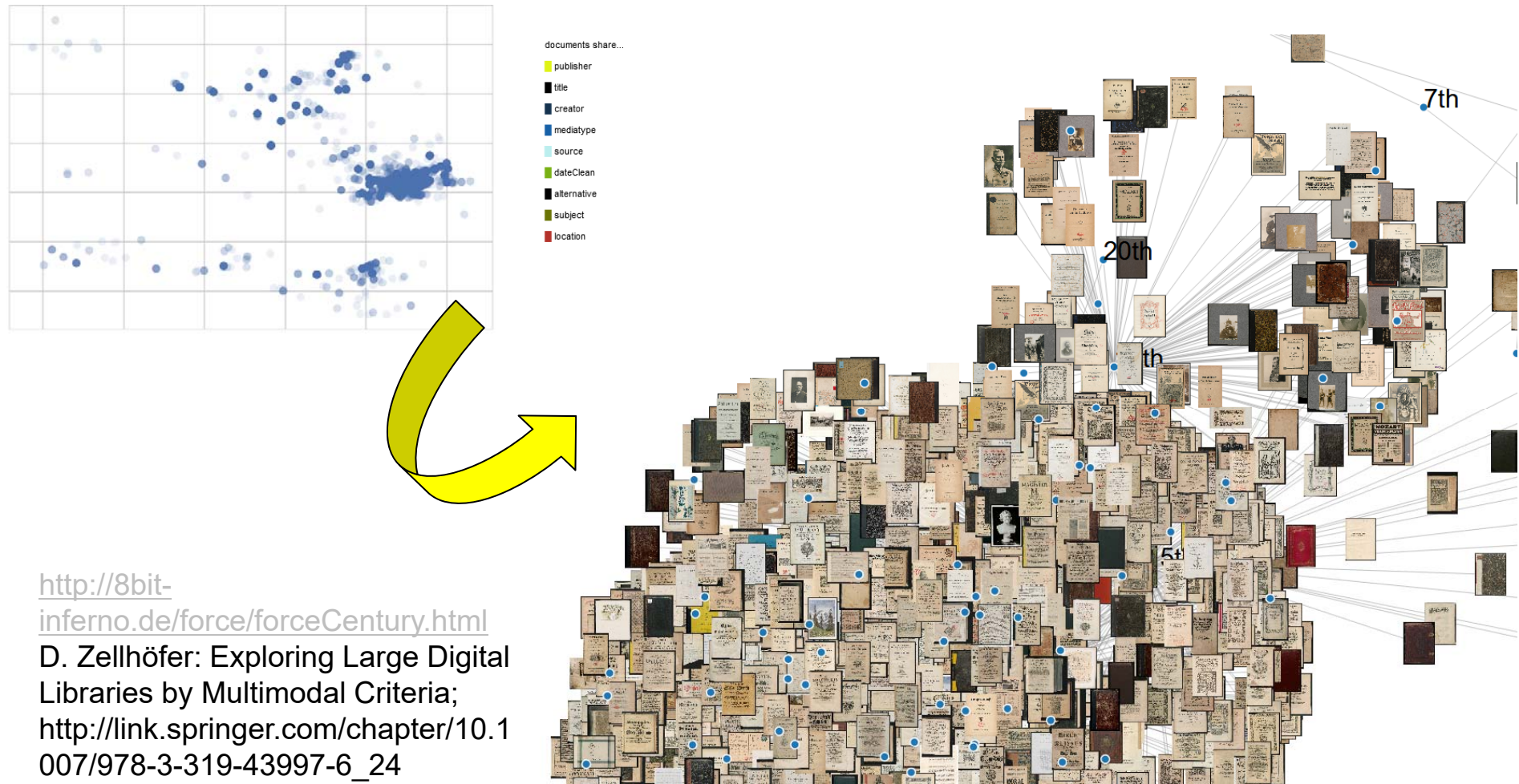
Feature-Extraktion, z.B. mit LIRE



By Masterwaw - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=19645418>



Content-Based Image Processing





**Staatsbibliothek
zu Berlin**
Preußischer Kulturbesitz

Tech Stack

Tech Stack



**Staatsbibliothek
zu Berlin**
Preußischer Kulturbesitz

ImageMagick (Shell)

LIRE (Java)

Grundlegender Workflow (Python)

Proof of Concept (JS+D3JS)

Tech Stack



Staatsbibliothek
zu Berlin
Preußischer Kulturbesitz

ImageMagick (Shell) ✓

LIRE (Java) ✓

Grundlegender Workflow (Python) ✓

Proof of Concept (JS+D3JS) ✓

liegt als Jupyter-Notebook auf GitHub

Treffen einmal pro Woche, sonst Internetz

Support durch Stabi-Team

VIELEN DANK.

Werdet Teil des am besten bewerteten Projekts des
SS 2017!

VIELEN DANK.
