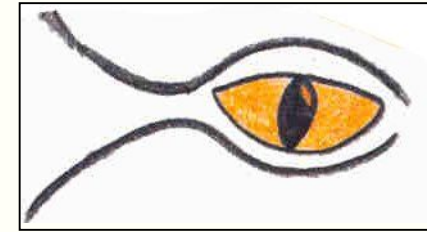




HTW Chur Sept. 2007 – Jan. 2008
Kurs STAN: Standards und Regelwerke



16) Google Books in Lausanne

Version 8. Juli 2008

Pierre Gavin



Das Projekt Google Books im Generellen

<http://books.google.com/>

<http://books.google.fr/intl/fr/googlebooks/about.html>

The screenshot shows the Google Books search interface. At the top left is the Google logo with 'Recherche de livres' and 'BETA' below it. A search box contains the text 'charles secrétan' and a button labeled 'Chercher des livres'. To the right of the search box are links for 'Rec' and 'Aid'. Below the search bar is a blue header with the word 'Livres' and navigation links: 'Tous les livres', 'Aperçu limité', and 'Affichage du livre entier'. The main content area displays three search results, each with a book cover thumbnail on the left and text on the right. The first result is 'La philosophie de la liberté' by Charles Secrétan, published in 1879, with 384 pages. The second result is 'Le principe de la morale' by Charles Secrétan, published in 1883, with 429 pages. The third result is 'La civilisation et la croyance' by Charles Secrétan, published in 1893, with 429 pages. Each result includes links for 'Affichage d'extraits', 'À propos de ce livre', 'Ajouter à ma bibliothèque', and 'Plus d'éditions'.

Google
Recherche de livres BETA

charles secrétan

Chercher des livres

Rec
Aid

Livres Tous les livres - Aperçu limité - Affichage du livre entier

La philosophie de la liberté
de Charles Secrétan - Metaphysics - 1879
Affichage d'extraits - À propos de ce livre - Ajouter à ma bibliothèque - Plus d'éditions

 Le principe de la morale
de Charles Secrétan - 1883 - 384 pages
Affichage d'extraits - À propos de ce livre - Ajouter à ma bibliothèque - Plus d'éditions

 La civilisation et la croyance
de Charles Secrétan - 1893 - 429 pages
Affichage d'extraits - À propos de ce livre - Ajouter à ma bibliothèque - Plus d'éditions



Stand des Projekts

Ziel: Digitalisierung von „all the world`s books“

Stand des Projekts: schon 1 Mio (?)

Bücher ohne Copyright: Full Text zugänglich

Bücher mit Copyright: nur einige Auszüge
zugänglich (oder gar keine)

Reaktionen

Von Seiten der Verlage:

Befürchtung die Rechte Betreffend

Von Seiten der Bibliothekare:

Befürchtung bezüglich der Monopolstellung (Jean-Noël Jeanneney) und bezüglich des obligatorischen Zugangs über Google (Suche und Konsultation)



Reaktionen

Von Seiten der Wissenschaftler:

1. Fehler beim Scannen
2. unvollständige oder falsche Metadaten
3. Zugang zum Fulltext von Texten ohne Copyright : manchmal unmöglich

<http://blog.historians.org/articles/204/google-books-whats-not-to-like>





Am Projekt teilnehmende Bibliotheken

- Bayerische Staatsbibliothek
- Harvard University
- New York Public Library
- Oxford University
- Princeton University
- Stanford University
- University of California
- etc. etc.



Das Google Books-Projekt in Lausanne

„Googlos“ = Google – Lausanne (Losanna)

- Partnerschaft Google / Bibliothèque cantonale et universitaire de Lausanne (BCU-L)
- scan + OCR
- es handelt sich um Bücher des 17., 18. und 19. Jahrhunderts (ohne Copyright)

Arbeitsorganisation

1) Export von MARC-Aufnahmen aus Rero

- Auswahl nach diversen Kriterien (Daten, etc.)
- Export im MARC21 / ISO 2709

```
00699nam a22002297a
4500001001400000003000700014005001700021008004100038035001200079039004400091
0400009001350720015001441000032001592450042001912600026002333000020002595000
02500279949006400304949006400368990002400432999001300456_vtls000205162_REROV
D_20060609133000.0_970225s1799 fr ||||| ||||00|| |fre d_ _a0222845_
9_a200606091330_bVLOAD_y200208012122_zload_ _aRERO_7_as1fr_2rero_1_aBoileau-
Despréaux, Nicolas_10_aOeuvres /_cNicolas Boileau Despréaux_ _aParis,_can VII [1799]_
_a2 vol. ;_cin-18_ _aEdition stéréotype_
_AVIRTUAITEM_D10047_F1_X1000_61092273881_9[VOL. 1]_a1M 1127/1_
_AVIRTUAITEM_D10047_F1_X1000_61092273882_9[VOL. 2]_a1M 1127/2_00_a1M
1127_xla/bcu-/w/_ _aVIRTUA50
```





Arbeitsorganisation

2) Konvertierung ins MARC/XML-Format

```
</marc:datafield>
- <marc:datafield tag="072" ind1="" ind2="7">
  <marc:subfield code="a">s1fr</marc:subfield>
  <marc:subfield code="2">rero</marc:subfield>
</marc:datafield>
- <marc:datafield tag="100" ind1="1" ind2="">
  <marc:subfield code="a">Boileau-Despréaux, Nicolas</marc:subfield>
</marc:datafield>
- <marc:datafield tag="245" ind1="1" ind2="0">
  <marc:subfield code="a">Oeuvres /</marc:subfield>
  <marc:subfield code="c">Nicolas Boileau Despréaux</marc:subfield>
</marc:datafield>
- <marc:datafield tag="260" ind1="" ind2="">
  <marc:subfield code="a">Paris,</marc:subfield>
  <marc:subfield code="c">an VII [1799]</marc:subfield>
</marc:datafield>
- <marc:datafield tag="300" ind1="" ind2="">
  <marc:subfield code="a">2 vol. ;</marc:subfield>
  <marc:subfield code="c">in-18</marc:subfield>
</marc:datafield>
```



Arbeitsorganisation

3) XML Datenbank- Xindice

Die Aufnahmen werden gleichzeitig im Marc XML – Format und in einem vereinfachten XML-Format in der Xindice Datenbank gespeichert

```
dbitm - heap -
heap-data-001 Document 'heap-data-049'
heap-data-002
heap-data-003 <?xml version="1.0" encoding="UTF-8"?>
heap-data-004 <heap-data id="049" loc="10047" part="2/2" pfx="1G" size="586" tm-
heap-data-005 <item cat-id="0875796" cn1="1G 660" cn2="" itm-id="1094436462">
heap-data-006 <bcd>1094436462</bcd>
heap-data-007 <cn2/>
heap-data-008 <cn1>1G 660</cn1>
heap-data-009 <icl>3000</icl>
heap-data-010 <nit/>
heap-data-011 <npb/>
heap-data-012 <aut>Stahl, Friedrich Julius</aut>
heap-data-013 <tit>Der christliche Staat : Vortrag ...</tit>
heap-data-014 <yr2>1858</yr2>
heap-data-015 <zct>gw-</zct>
heap-data-016 <phy>1 vol.</phy>
heap-data-017 </item>
heap-data-018 <item cat-id="0332469" cn1="1G 666" cn2="" itm-id="1092414639">
heap-data-019 <bcd>1092414639</bcd>
heap-data-020 <cn2/>
heap-data-021 <cn1>1G 666</cn1>
heap-data-022 <icl>1000</icl>
```

Arbeitsorganisation

4) Benützung der XML-Datenbank

Die Produkte werden insbesondere via XSLT erhalten

Chariot t-14

Ce chariot contient livres, pour une somme estimée à \$.

Auteur	Titre	Date	Nb.vol./p.
Zurlauben, Beat-Fidel	Bibliothèque militaire, historique ...	1760	3 vol.
Zurlauben, Beat-Fidel	Bibliothèque militaire, historique ...	1760	3 vol.
Faesi, Johann Konrad,	Abhandlungen über wichtige ...	1763-1764	2 vol.
Faesi, Johann	Abhandlungen über wichtige ...	1763-1764	2 vol.

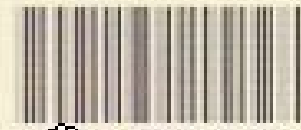
```
- <html>
- <head>
  <title>Test Résumé de notices</title>
</head>
- <body>
  <h1>Test Résumé de notices</h1>
- <table width="100%">
  - <colgroup>
    <col width="35%" />
    <col width="50%" />
    <col width="15%" />
  </colgroup>
  - <tr>
    <th>Auteur</th>
    <th>Titre</th>
    <th>Date</th>
  </tr>
  - <xsl:for-each select="//marc:record">
    <xsl:call-template name="rsm-record" />
  </xsl:for-each>
</table>
</body>
</html>
</xsl:template>
- <xsl:template name="rsm-record">
- <tr>
- <td>
  <xsl:value-of select="marc:datafield[@tag='100']" />
</td>
- <td>
  <xsl:value-of select="marc:datafield[@tag='245']" />
```



Arbeitsorganisation

5) Laufzettel

Druck von Laufzettel



92399023

AA 653+1

Florus, Lucius Annaeus
Lucii Annaei Flori Rerum romanarum ...
1684
366 p.

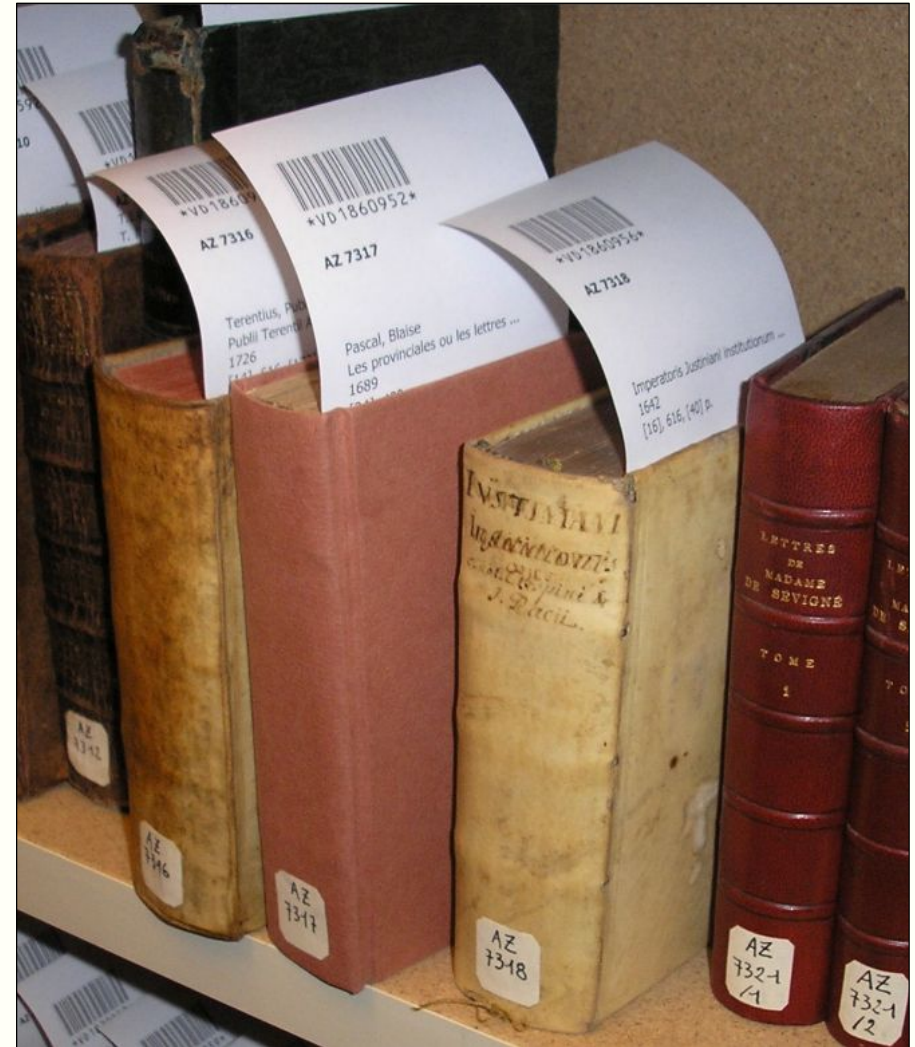
- absent
- incomplet
- mauvais état
- hors dimensions
- hors date
- recueil
- reliure serrée
- marge insuffisante
- item(s) à revoir
- numérisation BCU
- à couper



Arbeitsorganisation

6) Büchersuche im Magazin

Bücherkontrolle nach verschiedenen Scan-Kriterien





Arbeitsorganisation

Einige Probleme

- ungeschnittene Bücher
- falsche oder fehlende Exemplarsätze
- Dubletten
- zu dicke, zu grosse oder unmöglich zu öffnende Bücher
- fehlende Bücher
- zusammengebundene Werke
- etc.



Arbeitsorganisation


7) Bücherwagen

Jeder Bücherwagen wird mittels einer Nummer und eines Barcodes identifiziert; die Laufzettel werden gescannt um die XML-Datenbank à jour zu halten



Arbeitsorganisation

8) Druck von Begleitlisten für die Zollformalitäten

 **Chariot BCUL014 du lot BCUL_20071107**

Ce chariot contient 17 livres, pour une somme estimée à 10 000.

Ce rapport comporte 17 lignes, dont 6 décrivent en principe une partie d'un recueil (marquées (*)) et 1 marquent l'emplacement d'un sigillet pas encore catalogué.

Il existe environ 9 erreurs de codifications (marques (*) manquantes).

Auteur	Titre	Date	Pays	Vol.	Cote
Pictet	Une course à Chamounix : conte ...	1838	fr		AZ 3926
Thierry	Histoire de la conquête de ...	1856	fr	[VOL. 1]	AZ 3946/1
Thierry	Histoire de la conquête de ...	1856	fr	[VOL. 2]	AZ 3946/2
Thierry	Histoire de la conquête de ...	1856	fr	[VOL. 3]	AZ 3946/3
Thierry	Histoire de la conquête de ...	1856	fr	[VOL. 4]	AZ 3946/4
	Vie de Saint Bernard de Menthon, ...	1862	fr		AZ 3949
Tocqueville	De la démocratie en Amérique ...	1835	be	[VOL. 1]	AZ 3951/1
Tocqueville	De la démocratie en Amérique ...	1835	be	[VOL. 2]	AZ 3951/2
Lecchi	Piano della separazione : ...	[1762]	it		AZ 3954
Bonaparte	Economie politique	[ca 1864]	fr		AZ 3956
Bonaparte	La question italienne sous ...	[ca 1862]	fr		AZ 3958
Leresche	Les rudimens, ou les premiers ...	1791	sz		AZ 3962
Pasor	Etyma nominum priorum, ...	[ca 1633?]	sz		AZ 3964
Pasor	Georgii Pasoris lexicon Graeco-Latinum ...	1633	sz	<i>Recueil (*)</i>	AZ 3964
Delavigne	Théâtre	1863	fr	[VOL. 1]	AZ 3969/1



Arbeitsorganisation

9) Der Google-LKW transportiert die Wagen von Lausanne nach xxx (?)

10) Die Metadaten werden zu Google geschickt

11) Google scannt die Bücher + OCR und schickt einen Bericht zurück (nicht gescannte Bücher, etc.)

12) Der Google-LKW bringt die Bücher zurück



Arbeitsorganisation

13) Die zurückgebrachten Laufzettel werden gescannt, zur Kontrolle

14) Die Links zu Google Books werden integriert in den RERO Katalog.



Arbeitsorganisation

Ein Spezialprogramm
(Gonin/DIDOC)
ermöglicht die
Verwaltung aller
Operationen

