Measuring the impact of piracy and open access on the academic library services

MULLER, Floriane Sophie, IRIARTE, Pablo

Abstract
The University of Geneva library has seen, like others around the world, a slight but steady decrease in its document delivery service usage for a few years now and this decline also seems to affect the use of the electronic licensed collection. Indeed, 2016 was the first year where usage of licensed journals decreased. How could it be correlated with other actual trends, such as an ever growing open access corpus and an increased visibility of shadow libraries like Sci-Hub? To what resources and services, beside the offer of the library, can our scholars turn when they need an article? In this work we try to reveal and understand the mechanisms that operate behind this gradual disengagement of scholars in our library services. Working a reverse way, we checked in WoS all the papers published in 2015 and 2016 by authors affiliated in Geneva University STM faculties (3’833 and 3’989 articles respectively). Then we extracted all the references cited in those papers (364’445 references, 80% having DOIs) and enriched them with more identifiers, bibliographic data and access information, using bibliographic APIs from [...]
MEASURING THE IMPACT OF PIRACY AND OPEN ACCESS ON THE ACADEMIC LIBRARY SERVICES

ILDS conference, Paris, October 4th, 2017

Pablo Iriarte
Floriane Muller

Pablo.Iriarte@unige.ch
Floriane.Muller@unige.ch
AGENDA

1. Context
2. Methods: Data Analysis
3. Results & Discussion
4. Conclusions
CONTEXT

ILL/DD Trends
Shift from paper to e-resources
Open Access
Piracy
ILL/DD TRENDS IN SWISS UNIVERSITY LIBRARIES

ILL trends in Swiss University Libraries
(source: Federal Office of Statistics (OFS) 2017)

- Books borrowed from other libraries (OFS G61)
- Books loaned to other libraries (OFS G60)
- Articles (OFS G62)
ILL/DD TRENDS IN SWISS MEDICAL LIBRARIES

ILI Articles Trends:
some Swiss Libraries specialised in Medical Fields

[data sources: 1.) OpenILLLink software; 2.) Yearly report published by the library; 3.) OFS statistics for "bibliothèques de réseaux universitaires"]

- CHUV - Bibliothèque de médecine Lausanne (1)
- Bibliothèque de l'Université de Genève - Uni CMU (médecine) (1)
- Universitätsbibliothek Bern - Medizin und Maturwissenschaft (2)
- Universitätsbibliothek Basel - UB Medizin (2)
- Zentrum für Zahnmedizin Zurich (3)
- Schweizerisches Tropen- und Public Health-Institut (Swiss TPH) Basel (3)
- Vetsuisse-Fakultät Zurich (3)
- Kinderspital Zürich (3)
- Universität Zürich, Institut für Epidemiologie, Biostatistik und Prävention (3)
ILL/DD TRENDS

ILL Trends in DOCLINE & Loansome Doc

No data found for LoansomeDoc for FY2015 & FY2016

- ILL requests in Docline
- Loansomedoc Requests
PORTION OF ACQUISITION BUDGET DEVOTED TO E-RESOURCES

Proportion des frais d’acquisition de documents sur supports électroniques par rapport aux frais totaux d’acquisition de documents, 2004-2016

Bibliothèques universitaires

1 2006 la bibliothèque publique et universitaire de Genève perd son statut de bibliothèque universitaire pour devenir une bibliothèque publique. Dès cette année, elle ne fait donc plus partie de la catégorie des bibliothèques universitaires
2 2008 unification des finances de la bibliothèque et du réseau de bibliothèques de l’Université de Berne, création de la bibliothèque de l’Université de Genève à partir du réseau de bibliothèques de l’Université de Genève
3 2010 fusion de la bibliothèque centrale et du réseau de bibliothèques de l’EPFL

Source: OFS – statistique suisse des bibliothèques  © OFS 2017
Yet, for the first time in 2016

Full-text article requests at University of Geneva Library

2013: 1,679,037
2014: 1,713,114
2015: 1,935,068
2016: 1,900,846
Figure 2: Number of articles (left panel) and proportion of articles (right panel) with OA copies, estimated based on a random sample of 100,000 articles with Crossref DOIs

PIRACY


So...

When it comes to obtaining the full-text of an article...

...To what resources and services can/do our scholars turn?
DATA ANALYSIS

Our data sources
Data cleaning, processing & merge
OUR DATA SOURCES

- E-journals
- Serial collection
- SciHub
- oaDOI
- ILL/DD
- Institutional Repository

WOS SCI ed.
(articles cited by UNIGE members in 2015-6)
On March 3rd, 2017, extract all works published in 2015 and 2016, Affiliation enhanced = University of Geneva

Retrieval all citation data for those works

Deduplicate

DOI included in the reference?

Yes

Ndoi=228'092
(82%)

Use Crossref API to retrieve detailed metadata

223'283 Valid DOIs
32 duplicates
18'542 Bad DOIs

Not included in the reference?

No

Ndooi=48'521
(18%)

Use Crossref OpenURL web service to find the DOI from the available metadata

13'765 DOI found (28%)
34'756 Not found (72%)

Excluded references (missing data)

N=53'330

Valid DOIs

Dataset used for data analysis and merging N=223'283

Valid DOIs

Dataset used for data analysis and merging N=223'283

Invalid DOIs or duplicates

N=18'574

Excluded references

N=53'330
Export all anonymised orders meeting following criteria:
- Request date: 2015 & 2016
- User requesting doc is affiliated to UNIGE or Univ. Hospital (HUG)

Requested doc has DOI or PMID?

- Yes
  - N=4'737 (63%)
- No
  - N=2'772 (27%)

Which one?
- DOI
  - N=9
- PMID
  - N=4'728

Use NLM E-utils service to obtain equivalent DOI?

- Yes
  - DOI obtained = ?
  - No DOI obtained = ?
- No
  - DOI obtained = ?

Use Crossref OpenURL web service to find the DOI from the available metadata?

- Yes
  - 1'523 DOI found
  - N=1'523
- No
  - No DOI found
  - N=3'781

Dataset used for data analysis and merging
- N=3'728 (50%)

Excluded references (missing data)
- N=3'781 (50%)

Very small because our OpenILLink Tool had a DOI problem

Orders\textsubscript{2015-2016} = 7'509

Orders\textsubscript{2015-2016} = 7'509
Our process to verify accessibility / usage

For each article (with a valid DOI and corresponding metadata) cited in UNIGE 2015-6 publications in WOS SCI, check if

ISSN is amongst print / online subscription resources and corresponding year match?

DOI is present in 2015-2016 ILL/DD requests

DOI is amongst our institutional repository and if yes, what is full-text status (Green OA)

Query DOI in OaDOI API and get answer regarding status: OA or closed?

DOI is present in SciHub requests from Switzerland IP address (2015-16 partial data)

Dataset usable for data analysis and merging \( N_{DOI} = 223'283 \) (80%)

Keep only articles: 220'527

N=3'728 (50%)

Usable dataset N=33'876

Usable dataset

Nprint= 9'671
Nonline=79'758

Cleaned dataset N=33'993

Dataset usable for data analysis and merging \( N_{DOI} = 223'283 \) (80%)
**Final Matrix**

```python
# Export final matrix
df2015ex.to_csv('df2015ex.tsv', sep='\t', encoding='utf-8', index=False)
df2016ex.to_csv('df2016ex.tsv', sep='\t', encoding='utf-8', index=False)
```

```python
In [60]: df2015ex

Out[60]:
```
<table>
<thead>
<tr>
<th></th>
<th>DOI</th>
<th>OA</th>
<th>AOU</th>
<th>SHCH</th>
<th>OI</th>
<th>EJ</th>
<th>PJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.1074/fp01241</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>10.1074/fb0386713</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>10.1016/j.pharmthera.2014.05.002</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>10.1010/j.jclinepi.2007.11.008</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>10.1103/physrevb.54.15800</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>10.1093/mnras/stu1423</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>10.1093/mnras/stt1945</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>10.1093/glycob/cwq188</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>10.1016/j.nmd.2005.10.001</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>10.1000/19346182.2009.9648483</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>10.1371/journal.pone.0077691</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>10.1034/j.1500-0501.2001.012001019.x</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>10.1007/bf01337700</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>10.1007/bf01333110</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>10.1002/pssa.2210210220</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>10.1016/j.cytogrt.2010.10.003</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>10.1063/1.2788435</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```
SOME LIMITATIONS

- Imperfect data sources
  - Sci-hub data is partial and we don’t know all about it
SOME LIMITATIONS

• Imperfect data sources
  o Sci-hub data is partial and we don’t know all about it
SOME LIMITATIONS

• Imperfect data sources
  o Sci-hub data is partial and we don’t know all about it
  o SFX / catalog data : one-time 2017 snapshots.
  o Not all UNIGE papers are in WOS
  o WOS citations data are not so clean (missing metadata, no DOI or DOI not valid)
  o Half of ILL requests could not be analysed

• Imperfect Unique identifiers
  o No PMID to DOI easy reliable conversion tool
  o Undeclared DOI ?

• Short time-frame
  o Publication takes time. Researchers may cite articles they have gotten long ago (so not in our 2015-6 radar)

• Other factors
  o Eg. Free ILL/DD may have a role is trends
  o Researcher may resort to many other article providing « sources »
  o Do they cite unread articles ?

• Only journal articles
RESULTS & DISCUSSION
CITED IN 2015 ARTICLES AVAILABILITY WITHIN DATA SOURCES

Possible sources of articles cited in UNIGE publications on 2015

- Unknown
- E-Journals
- OA
- ILL/DD
- Sci-Hub CH
- Print Journals

Scale = 5000 citations
CITED IN 2016 ARTICLES AVAILABILITY WITHIN DATA SOURCES

Possible sources of articles cited in UNIGE publications on 2016
THE HIDDEN FIGURES

Unknown

2015: 18’595
2016: 14’412
DISTRIBUTION OF CITED ARTICLE BY YEAR PUBLISHED

Distribution of the 115’045 different articles cited in 2015 publications
Distribution of the 127’738 different articles cited in 2016 publications
Shift between 2015 & 2016?

Percentage of articles cited by UNIGE researchers in their 2015 & 2016 publication that are present in

- Institutional repository: 4% (both years)
- Print journals: 25% (2015) & 24% (2016)
- ILL/DD requests: 0% (both years)
- Sci-hub (CH): 0% (both years) & 1%
- None of those: 16% (2015) & 11% (2016)
CONCLUSIONS

• Need for articles should still be present as our researchers cite more and more
• Confirmed some feelings we had:
  ✓ OA is growing faster than the rest
  ✓ The library offers nice coverage.
  ✓ Piracy is not so high in Switzerland
  ✓ We don’t know all sources (and practices)
• But : we can’t explain the slight 2016 usage decrease of our e-journals.
• More research needed (survey ? direct observation ? …)
• Repeat for all resources provided by university and develop as a promotion tool for library services incl. ILL/DD?
Selected References

THANK YOU VERY MUCH!

Pablo.Iriarte@unige.ch
Floriane.Muller@unige.ch