



Statistics usage by French academic librarians. A survey

Chérifa Boukacem-Zeghmouri, Joachim Schöpfel

► **To cite this version:**

Chérifa Boukacem-Zeghmouri, Joachim Schöpfel. Statistics usage by French academic librarians. A survey. Learned Publishing, Association of Learned and Professional Society Publishers, 2012, 25 (4), pp.271-278. 10.1087/20120406 . sic_00875862

HAL Id: sic_00875862

https://archivesic.ccsd.cnrs.fr/sic_00875862

Submitted on 22 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Research article

Statistics usage by French academic libraries – a survey

Chérifa BOUKACEM-ZEGHMOURI
Université Lyon 1

Joachim SCHÖPFEL
Université Charles-de-Gaulle Lille 3

Abstract

The article presents a survey on usage statistics management in academic libraries in France. The objective is to provide empirical evidence on real methods and processes, best practices and problems experienced by librarians. The survey adapted the Baker & Read (2008) methodology to the French context. One third of the French academic libraries answered to the survey. Results show that usage statistics are crucial for the librarians' daily work. COUNTER reports are mostly used, specifically the JR1 format. Two thirds of the libraries enhance the usage statistics with other data. Librarians face three major difficulties: missing time, lack of usage data produced by French vendors, and need for customized tools. The article contains a short literature review and ends with recommendations for further studies. Its originality is that it is the first published survey on usage data management in France and that it allows for comparison with results from other countries.

Introduction

The digital revolution has taken place. Processing e-resources became part of everyday life in academic libraries. Today, library management includes monitoring, assessment of usage, performance measurement. Stakeholders ask for reporting and return on

investment. How do academic librarians deal with usage statistics and log files? What do we know about their experience, feedback, needs?

Workshops and seminars¹, discussions on mailing lists such as lib-stats, SERIALST or lis-e-resources, and surveys² provide insight and empirical evidence on the reality of dealing with statistics. Yet, this knowledge is more or less limited to the UK and US library communities, and we know but little about other, non-English speaking countries such as France.

Seven years ago, when we published our first article on the COUNTER project³ and when the French institute for scientific information INIST translated the COUNTER code of practice⁴, French academic libraries generally had poor or no experience with usage statistics of online resources. Compared to the UK, uptake of digital resources was delayed because of language barrier, lack of larger scientific journal publishers and a rather fragmented landscape of scientific structures and libraries.

In 2012, the situation has changed. Small and medium-sized French universities merge into large and competitive organisations, and consumption of digital information increased fast and steadily. The French academic consortium COUPERIN established a working group on usage statistics. Two research projects evaluated usage statistics of academic libraries and open archives. France is partner of the European PEER project. The Lille 3 university organized an international conference on usage assessment and practice in 2009⁵; even if the context may be different between countries, the basic challenges and problems are not.⁶

These projects and initiatives reflect growing awareness and interest for the topic in France. Nevertheless, what do we know about real methods and processes, best practices, problems experienced by librarians?

Literature review

Compared to the growing amount of research based on results of the exploitation of usage statistics of digital resources in academic libraries, the small number of articles dedicated to the processing of usage statistics itself is surprising. These studies show that the statistics gained importance from about 2000 on. Yet in the beginning, this evolution slowed down by lack of committed human resources and technical competencies. The second difficulty libraries encountered was missing standards even if the COUNTER project proposed solutions from 2003-2004 on⁷. The following years 2005, 2006 and 2007 were a period of increasing automation of data collection, formatting, preservation and processing, with tools like SUSHI, ScholarlyStats or ERMS.⁸

Publications that are more recent often deal with two topics, technical problems⁹ and usage of statistics for decision-making. A recent UKSG seminar underscored the importance of usage statistics for decision-making. Paradoxically, the increasing availability of data, the continuous development of the COUNTER Code of Practice and the financial pressure by governments produce similar effects and foster exploitation of these data. Usage statistics contribute to decisions somewhere between “what could be cancelled” and “what should be cancelled”.¹⁰ The most recent articles focus on usage statistics as a tool for return on investment analysis (ROI).¹¹

So far, our study is the first in French-speaking scientific literature to analyze the uptake and usage of statistics in academic libraries.

Methodology

Our methodology adapts the 2008 Tennessee survey¹² to the French context. We sent the adapted survey in November 2010 to the digital resources librarians of 87 French academic libraries, in print format with three months to reply.

More detailed than the Tennessee survey, our questionnaire adopts the same structure, with three main sections: the first section is on the library (size, patrons, and subjects); the second section is on the vendor-supplied usage statistics; the third section is on the data processing, tools, and objectives. Most questions are closed-ended but some are open and provide opportunity to add comments etc. important for this kind of exploratory study.

The questions cover the period 2009-2010.

The French official definition of “users” globally corresponds to the common “academic staff FTE” in the UK academic environment and includes academics, scientists, PhD (graduate) students, postdocs. For this survey, we added undergraduates.

Findings

Thirty-two academic libraries replied to the survey. The response rate (37%) is satisfying. The responding institutions represent the whole range of French universities classified in five groups following their dominant scientific domains: STM (sciences, technology, medicine), GMD (grand multidisciplinary) and SMD (small or medium-sized multidisciplinary), SS&H (social sciences and humanities, including arts), Law (law) and ECO (business, including economics and management). Table 1 shows their breakdown by type and size.

Insert Table 1: French university libraries break-down

| Size | No. of respondents | Per cent | STM | GMD | SMD | SS&H | Law | ECO |
|---------------|--------------------|----------|-----|-----|-----|------|-----|-----|
| <5,000 | 2 | 6% | | | 2 | | | |
| 5,000-9,999 | 6 | 19% | 1 | | 4 | | 1 | |
| 10,000-14,999 | 6 | 19% | | 4 | 1 | | 1 | |
| 15,000-19,999 | 7 | 22% | 1 | | 1 | 3 | | 2 |
| 20,000-24,999 | 6 | 19% | 2 | 2 | 1 | | | 1 |
| 25,000-29,999 | 2 | 6% | | 2 | | | | |
| 30,000-34,999 | 2 | 6% | 1 | 1 | | | | |

| | | | | | | | | |
|---------------|----|------|---|---|---|---|---|---|
| 35,000-39,999 | 1 | 3% | 1 | | | | | |
| 40,000-44,999 | | | | | | | | |
| 45,000-49,999 | | | | | | | | |
| ≥50,000 | | | | | | | | |
| Total | 32 | 100% | 6 | 9 | 9 | 3 | 2 | 3 |

The size of the participating universities ranges from 2,500 to more than 36,000 undergraduates, PhD students and scholars. 85% have less than 25,000 students and scholars. To allow direct comparison with Baker & Read (2008), we collapsed these groups into four: <10,000 (eight universities), 10,000-14,999 (six universities), 15,000-24,999 (thirteen universities), ≥25,000 (five universities). Undergraduates represent 85% of the overall population; PhD students represent 11% and scholars, 6%.

Publishers and resources

Together, the university libraries subscribed to 538,708 journals (print titles are not included) from 2071 publishers and aggregators (cumulated). The number of subscriptions ranged from 3,050 to 50,000 titles, with a median of 12,168. 81% of the libraries reported usage statistics beyond 50,000 annual downloads.

We asked the academic libraries for the number of publishers with usage reports. The libraries reported that only 25% of the publishers provide usage statistics. Their number ranged from zero to 55, with a median of 14 (Table 2).

Insert Table 2: Publishers providing usage statistics to libraries

| Size | No. of respondents | Per cent | No. of publishers | | |
|---------------|--------------------|----------|-------------------|---------|---------|
| | | | Median | Minimum | Maximum |
| <10,000 | 8 | 25% | 18 | 6 | 34 |
| 10,000-14,999 | 6 | 19% | 13 | 3 | 21 |
| 15,000-24,999 | 13 | 41% | 14 | 0 | 30 |
| ≥25,000 | 5 | 16% | 13 | 3 | 55 |
| Total | 32 | 100% | 14 | 0 | 55 |

68% of the respondents received monthly reports; annual (39%) or biannual (7%) reports were less frequent. 32% of the libraries received usage reports inconstantly and/or on demand.

Purpose for analyzing usage statistics

Why did academic libraries collect and exploit usage data? The most important reason was the reporting to the Ministry who, as the most important funding body, receives annual reports and standard statistics from all academic libraries. Nearly all respondents (97%) mentioned this reason. Another important purpose (90%) was the usage of statistics for acquisition policy decision-making, followed by justification of expenditures (71%). Less important were exploitation of data for training, communication and “knowledge of users” (19%).

Type and format of usage reports

90% of the participants received COUNTER compliant statistics, in particular the Journal Report 1 (JR1) but also other type of reports (JR1a, JR5, JR2 or JR3, database reports). Nearly half of them (39%) also received usage data that were non-compliant with the COUNTER code (downloads, sessions, requests, views...).

The usage reports were delivered in different formats, mostly as an Excel spreadsheet (94%) but also in PDF (58%) or in a database (29%) or text format (13%) (see Table 3).

Non-compliant reports were most often supplied in PDF.

Insert Table 3: Format of usage reports

| Format | Nb | % |
|--------|----|-----|
| xls | 29 | 94% |
| pdf | 18 | 58% |
| csv | 9 | 29% |
| html | 6 | 19% |
| txt | 4 | 13% |
| xml | 2 | 6% |
| doc | 1 | 3% |

| | | |
|--------|---|----|
| others | 2 | 6% |
|--------|---|----|

Most of the participants reported two or more different formats. Just 19% received their statistics in only one format, e.g. Excel.

Time spent on usage data management (2009-2010)

Retrospective estimation of workload probably is not very reliable and difficult to validate. Yet, it may be interesting to gather some global data, especially on relative time allocated to different tasks in processing usage statistics (Table 4).

Insert Table 4: Number of hours spent processing vendor usage reports

| Size | No. of respondents | Per cent | No. of hours | | |
|---------------|--------------------|----------|--------------|---------|---------|
| | | | Median | Minimum | Maximum |
| <10,000 | 6 | 22% | 96 | 16 | 166 |
| 10,000-14,999 | 5 | 19% | 72 | 20 | 280 |
| 15,000-24,999 | 11 | 41% | 32 | 18 | 300 |
| ≥25,000 | 5 | 19% | 112 | 20 | 400 |
| Total | 27 | 100% | 42 | 16 | 400 |

Librarians reported between 16 and 400 hours spent in 2009 on processing of usage data, with a median of 42 hours. The maximal workload increased with the size of the institution. They spent roughly the same amount of time (27-29%) on downloading, manipulating and analyzing of data and less time on reformatting (16%).

Combining data

Two third (67%) of the participants declared that they combine usage data from publishers with other data, especially with three different types of data: local usage data derived from log files, for instance for non commercial items (scanned items in repositories); information from the licence (number of authorized users, total amount,

co-funding, it's duration); and data related to the accessed serials (impact factor, subject/discipline).

The purpose was to obtain detailed metrics on preferred items and collections, financial metrics, and data on other than commercial resources.

Useful statistics and tools

The most important and useful statistics were the COUNTER Journal Report 1 (JR1) “Number of Successful Full-Text Article Requests by Month and Journal”. Other statistics were less frequently mentioned, such as number of downloads and sessions, or the hit parade of most often viewed documents (Table 5).

Insert Table 5: Useful statistics in libraries

| | Librarians | | | Answers | |
|---------------------------|------------|-----|-----------------|---------|-----|
| JR1 | 25 | 81% | COUNTER (total) | 35 | 66% |
| COUNTER | 6 | 19% | | | |
| JR5 | 2 | 6% | | | |
| DB1 | 1 | 3% | | | |
| JR1a | 1 | 3% | | | |
| Sessions | 6 | 19% | Traffic total | 13 | 25% |
| Téléchargement | 4 | 13% | | | |
| Connexions | 2 | 6% | | | |
| Consultations résumés/TOC | 1 | 3% | | | |
| Hitparade | 2 | 6% | Others total | 5 | 9% |
| Origine des consultations | 1 | 3% | | | |
| Affichage document | 1 | 3% | | | |
| Back files | 1 | 3% | | | |
| | 53 | | | 53 | |

No participant used an electronic resource management system (ERMS) for the handling of usage statistics (one library was planning to do so), and no one retrieved usage reports via the SUSHI protocol. One library subscribed to a service that delivered their vendor statistics through a single point of access but was not satisfied with the result.

Three libraries only analysed local usage data (log files) from their EZ proxy server or gathered usage data with web analytics software, with significant assistance from their IT department.

Differences between different types of universities

Related to the disciplinary category of their university, the academic libraries have unequal access to usage statistics. French-speaking publishers in social sciences, humanities, law and economics often don't provide usage data and even when they do, the statistics are seldom COUNTER compliant. So without surprise, the SS&H libraries have less access to data, and only 33% of them spend some amount of time on comparing and enrichment of data.

Insert Table 6: Library access to statistics by disciplinary categories

| | Number of libraries | Comparative analysis | Added value |
|----------------------|----------------------------|-----------------------------|--------------------|
| SMD | 9 | 88% | 66% |
| STM | 6 | 83% | 83% |
| Law & ECO | 5 | 75% | 75% |
| GMD | 9 | 50% | 50% |
| SS&H | 3 | 33% | 33% |

In comparison, 88% of the SMD libraries (small or medium-sized multidisciplinary) and 83% of the STM libraries (sciences, technology, medicine) make comparative analyses between the usage reports from different vendors.

Following the survey results, the estimated median time the libraries spent per year on acquisition and processing of usage data shows a significant variation. The SS&H, law, economics and business libraries spend only 37 hours per year on this activity, STM and small multidisciplinary campus libraries spend 70 hours per year while the libraries of

big and medium size multidisciplinary campus say spending 210 hours per year on acquisition and processing of usage reports.

Summary of findings - Discussion

Only 25% of the publishers are reported to provide statistics. Such a low percentage is not linked to the publishers' size but is explained by the fact that the majority of French and francophone publishers – often in humanities and social sciences – do not dispatch statistical data. This problem mainly concerns the libraries dedicated to law, letters and social sciences. The STM libraries are less affected, since they subscribe to the most important international scientific publishers. This phenomenon induces great inequalities in the French university library network.

Academic libraries generally accepted the COUNTER Code of Practice as a standard; especially the JR1 format is valued as an essential indicator. Librarians consider the fact to deal with standardized and comparable data as very helpful for their work.

Nevertheless, they also criticize the occasionally poor quality of vendors' statistics, e.g. errors such as wrong customer ID or unexplained abnormal activity figures, unstable title bundling, forbidden access to usage reports etc. Some of them even start to question the reliability of COUNTER data produced by some platforms.

This may partly explain why at least half of them collect and process non-COUNTER compliant statistics, such as reports on sessions, information seeking characteristics such as form of navigation or search approach. Sometimes, the analysis of local data and log files is meant to diminish the library's dependence from vendors and, furthermore, to control their data.

The survey confirms Baker & Read's findings on lack of time spent on usage statistics. In average, French librarians say spend less time than their American colleagues do (1-2 hours per week). This lack of time – even if the estimated time may be subjective and

biased - is cause of frustration and dissatisfaction, especially in STM and multidisciplinary academic libraries. Not only have the usage statistics increased the librarians' workload but also additionally, they feel that they cannot carry out the data analysis as they would like to or should do.

The quantitative methodology of our survey provides tendencies while interpretation is not always easy. Even so, most of the results seem to confirm the findings of Baker & Read (2008). Furthermore, the delay of the survey – three months – allowed for a large number of detailed answers to open questions and comments to closed questions. Yet, some answers obviously need more investigation, for example, the number of subscribed journals because the given figures appear over-estimated or may include other type of documents.

Conclusions

Six years after the first publication on COUNTER in France, the findings of our survey seem to be encouraging as they clearly show that the usage statistics became part of the daily work of most of the academic librarians. However, they also show the limits and indicate what need to be done to progress:

1. French academic librarians are short of usage statistics for a significant part of the subscribed resources. Especially many French publishers do not provide data. This partial availability of information limits the interest of metrics and dashboards.
2. French academic librarians often feel isolated and complain of missing opportunity for training and exchange. So far, there are no workshops or seminars on usage statistics, no discussion list or website such as the NESLI and JISC Journal Usage Statistics Portal initiative¹³.

3. The lack of time, the “do-it-yourself” character of their work and their dependency on the vendors’ data heavily affects the value they can add to the statistics. Globally, French academic librarians feel frustrated with too much work for a rather unsatisfying outcome.
4. Tools, procedures and data formats should be adapted to the French situation, in particular they should take into account the specific needs expressed by the French Ministry of Higher Education for the academic libraries’ annual activity reports.

Recommendations for further studies

Our findings provide first empirical evidence on the handling of usage statistics in French academic libraries. In order to gain further insight, we need qualitative studies on work organisation, job skills and expertise, best practices.

We also need studies on other digital resources than e-journals, such as e-books, institutional repositories or datasets, and they should distinguish between commercial resources and non-commercial information, e.g. grey literature and documents available through open access.

The standardization of usage data needs more communication and promotion, and this promotion should include the development of efficient tools and procedures. That is why the French academic consortium COUPERIN recently launched a more detailed follow-up study. Their first results largely confirm our own. Furthermore, COUPERIN works on a French solution of a journal usage statistics portal¹⁴, together with MIMAS¹⁵ and Cranfield University. The project should be completed in 2012.

¹ Bucknell, T., Taylor-Roe, J., Killick, S., and Thompson, S. 2011. Usage statistics for decision making. *Serials* 24:17-20.

² Baker, G. and Read, E. J. 2008. Vendor-Supplied usage data for electronic resources: a survey of academic libraries. *Learned Publishing* 21:48-57.

³ Boukacem-Zeghmouri, C. and Schöpfel, J. 2005. Statistiques d'utilisation des ressources électroniques en ligne : le projet COUNTER. *Bulletin des Bibliothèques de France* 50:62-66.

⁴ <http://counter.inist.fr/>

⁵ C. Boukacem-Zeghmouri (ed.) 2010. L'information scientifique et technique dans l'univers numérique. Mesures et usages. ADBS Editions, Paris.

⁶ Schöpfel, J. and Boukacem-Zeghmouri, C. 2010. Assessing online usage. *Research Information* p. 25.

⁷ Shepherd, P. 2006. The COUNTER code of practice for books and reference works. *Serials* 19:23-27.

⁸ Bhatt 2007, Farb 2007

⁹ Ruddock 2009, Xu 2010

¹⁰ Killick in Bucknell et al. (2011), see note 1.

¹¹ Borin 2011

¹² Baker & Read (2008), see note 2.

¹³ <http://www.jusp.mimas.ac.uk/>

¹⁴ <http://jusp.mimas.ac.uk/>

¹⁵ <http://mimas.ac.uk/>

Appendix

Survey on electronic journals usage statistics

I – The University Library

1) Your Library belongs to which disciplinary category ?

- STM
- SS&H
- Law & Economics
- Small and medium cross-curricular section - GMD
- Large cross-curricular section - SMD

2) What is the scope of your audience?

- Undergraduate and graduate students
- Postgraduate students
- Academics

3) How many research laboratories are located on your campus (including combined laboratories) ?

.....
...

4) The library currently subscribes to how many e-journal titles?

- : titles

5) What is the current FTE at your institution?

- From 0 to 5000:
- From 5,000 to 10,000:
- From 10,000 to 15,000 :
- From 15,000 to 20,000 :
- From 20,000 to 25,000 :
- From 25,000 to 30,000 :
- From 30,000 to 35,000 :
- From 35,000 to 40,000 :

-
- From 40,000 to 45,000 :
 - From 45,000 to 50,000 :
 - 50 000 or more:

II) Usage Data (for e-journals)

6) Number of vendors to which your library is subscribed?:

.....

7) Out of the previous number - see 6) - How many vendors provide reports for e-journals?:

.....

8) How often?

- Weekly
- Monthly
- Quarterly
- Semestrially
- Annual
- Other (specify)

9) Which reports are provided ?

.....
.....

10) In which format (PDF, txt, xls,...) ?

.....
.....

III) Manipulating and Analyzing data

11) For what purpose are you collecting usage data and analyzing it? :

- Local and national reports (ESGBU, ERE...)
- Collection management (subscriptions, cancellations)
- To justify expenditures
- Other (Specify) :

.....

12) Which reports are the most useful to you and why?

.....
.....

13) Is there data not included in the reports that you would like to obtain? For what purpose?

.....
.....

-
- 14) The professionals workers involved in the following tasks belong to which categories (A, B or C)?
- Ccontactin vendors:.....
 - Downloading :
 - Consolidation and analysis:.....
- 15) In 2009, how many hours did you spend for:
- Downloadingh
 - Reformatting.....h
 - Consolidation and manipulatingh
 - Analyzing.....h
- 16) In order to accomplish these tasks, did you use the following tools ?
- SUSHI
 - COUNTER
 - SCHOLARLYSTAT
 - ERMS
 - Other (Specify) :
.....
- 17) Do you sometimes combine other types of data to those supplied by vendors?
- Yes No
- If so, which data?
.....
.....
- 18) Do you sometimes combine data from different sources in order to have a comparative outlook?
- Yes No
- If so, which data did you compare?
.....
.....
- 19) What are the biggest challenges (difficulties, biases) you faced in making effective use of vendor usage statistics ?
.....
.....