Monitoring Digital Library Activities – Status-, Trend- & Performance-Indicators based on analysis of OpenURL-requests

Peter Ahrens

Ex Libris Germany

Monitoring Digital Library Activities with OpenURL

Peter Ahrens* & Nettie Lagace+

* Ex Libris Deutschland GmbH, Gasstraße 18, Hs.2, 22161 Hamburg, Germany
(peter.ahrens@exl.de)
+ Ex Libris Inc., 313 Washington Street, Suite 308, Newton, MA 02458,) USA (nettie@exlibris-usa.com)

Summary: We present a method and practical examples of generating library service statistics, status parameters and performance indicators by means of analysing log-files of digital library service-request of OpenURL link resolvers. These statistics can reveal behavior of users as well as utilization of content that may otherwise be more difficult or impossible to detect or quantify. This can serve to support library's decisions and development.

Keywords: Library Usage Statistics; User behavior; OpenURL; Context Object; Open Access;

Introduction:

There has been a fundamental shift in the use of using scholarly literature which has resulted in totally new possibilities for the tracking of usage patterns. The academic community's usage of scholarly communications has, for the most part, shifted to publications available electronically. Over recent years, an increasing proportion of that electronic content has been interlinked using the OpenURL standard for context sensitive linking - physically sitting at disparate web locations (e.g. bibliographic source information on one server – fulltext on another target server). Many academic libraries today make use of these linking capabilities via their locally-maintained linking-server applications. SFX as a link resolver can record and classify every single click of the user's linking activity. Via a logging database and multiple predefined quesries these can be analysed in numerous ways. Thus, for the first time, it is possible for an institution to get, in an automatic way, an easy overview over the various aspects of the general usage patterns of their entire user community.

Method:

All user activities mediated by the SFX OpenURL Link Resolver of some universities were recorded as OpenURL 1.0 Context Objects using the NISO standard Z39.88-2004.

Among the data elements recorded about every link request are:

- Date and time
- User Identification information such as IP and possibly additional user group or faculty (the Requester)
- Type of request, e.g. "fulltext request" (Service Type)
- Source Information system that user was navigating, when issuing the link request (Referrer)
- Literature item requested (e.g. journal article) (the Referent)

Through use of the internal SFX KnowledgeBase about electronic objects and its relations and normalization steps a standardization of the incoming request is undertaken and additional information about the literature item is derived, such as whether the item requested is retrieved from specific vendor packages or interfaces and the fact whether a subscription is required or access is free.

Data was transferred daily into the SFX statistics module, an offline SQL logging database and evaluated using a subset of the 20 standard SFX-statistics queries and additional profiling tools and routines.

Data analysis was performed in a number of ways:
1) Counting parameters in a period (e.g. requests and clickthroughs); overall and for subgroups (e.g. sources, titles, institutes, services, targets)
2) Looking at ratios between parameters in a period (like service acceptance or fulltext availability rate) as performance indicators
3) Looking at changes of indicators between periods as trends

We present as examples preliminary results on some interesting aspects:

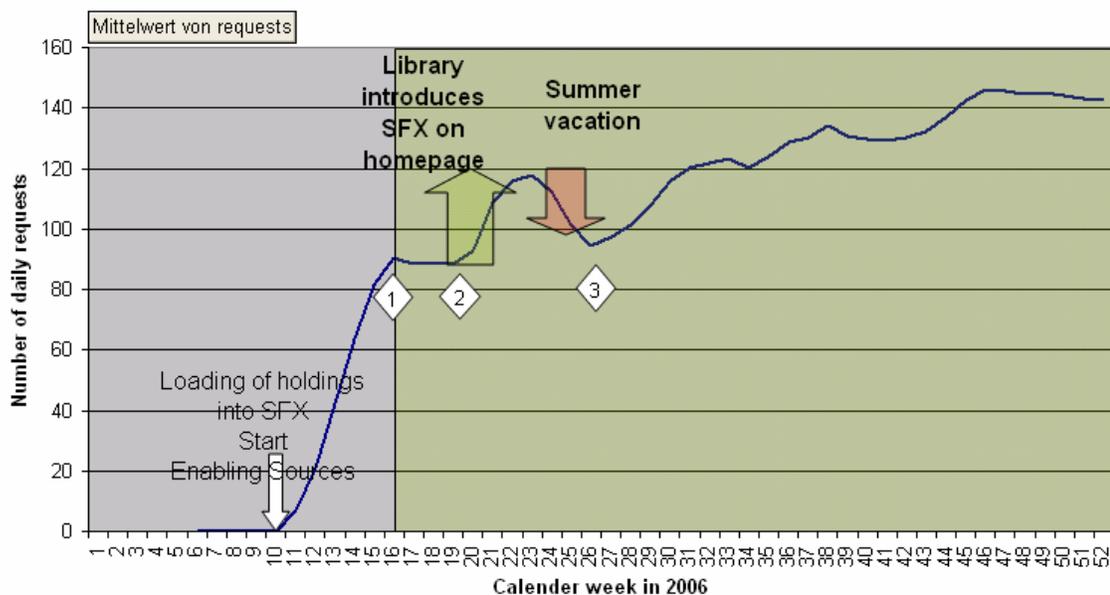**1.   Impact of promotional activities on usage**

Fig. 1 Uptake of OpenURL Linking usage by Users at University of Düsseldorf from March until December 2006

Fig. 1 shows number of daily of OpenURL requests via the institution's SFX server. A smoothing algorithm (30-days-average) was applied to filter out short term fluctuations in usage caused by weekends and bank holidays. Only requests from IP-authenticated library patrons (excluding staff) were taken into account.

Special dates are marked as <1> completion of Setup of SFX sources, <2> "Kick-Off" promotion of SFX where library actively introduces and explains this service on its homepage and additional intranet pages and <3> summer vacation. The growth curve shows a plateau at 90 daily requests in usage after initial setup (Week 10-15). Then following above mentioned promotional activities in week 20 there is a correlating (overshooting) peak with temporal increase of some 25 % additional usage. This is in turn followed by a dip of similar size due to the summer vacations <3>. Afterwards usage picks up again and continues to grow fairly steadily to about 145 requests towards the end of the first year (equivalent to some 60 % above usage level at "Kick-Off").

Discussion 1: The effect of this promotional activity can immediately be correlated with impact on usage parameters. This approach can not only be used to monitor overall effects but can be similarly applied to single information sources, targets or other entities (e.g. special user groups) during trials or to compare the relative evolution of use over several items over time.

Tracking of OpenURL usage over time helps to assess library services and activities. It gives additional – partly quantitative - insight to the extent and duration of effects. It should be taken care that the provision of any training classes on SFX be done preferably from within specially identified IP-ranges in order not to confound such demonstration usage with true usage trough "patrons".

2. Impact of free electronic journal access and use of backfiles?

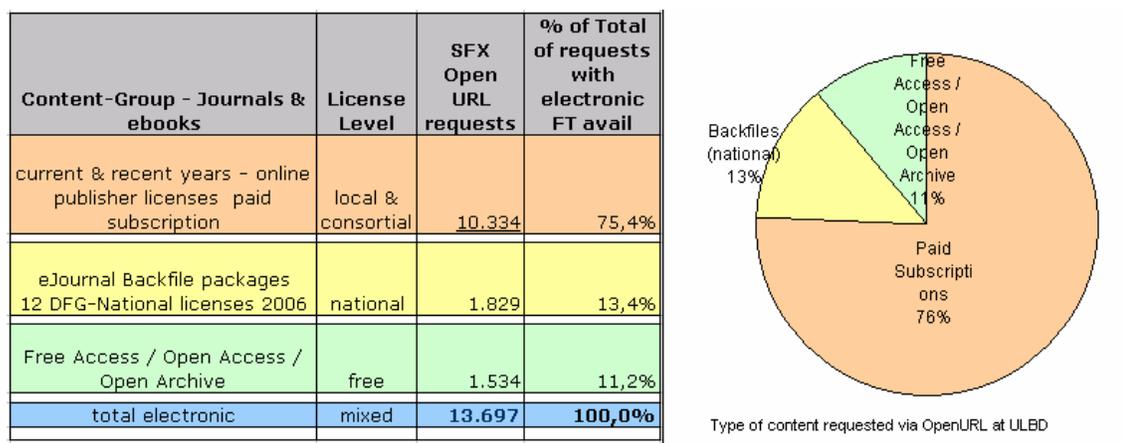| Content-Group - Journals & ebooks	License Level	SFX Open URL requests	% of Total of requests with electronic FT avail
current & recent years – online publisher licenses paid subscription	local & consortial	10.334	75,4%
eJournal Backfile packages 12 DFG-National licenses 2006	national	1.829	13,4%
Free Access / Open Access / Open Archive	free	1.534	11,2%
total electronic	mixed	13.697	100,0%

Fig. 2 Breakdown of content made available via different economic models

Data are based on subtotal numbers of end user requests to items available electronically via SFX at University of Düsseldorf from May to December 2006. These numbers were obtained using a standard SFX query by target. The intermediate target list was then further grouped for the three models of access using additional information about free/paid access in the SFX KnowledgeBase and a list of publishers and backfile packages of the German e-Journal backfile program[4].

Out of a total of 13.697 requests for documents available electronically to the user in that period the vast majority of 75,4 % (10.334) were for paid scholarly publications (either local or consortial). 13,4 % (1.829) were for articles made available through the national backfile program. In 2006 this program covered about 2.500 titles from 11 major international publishers plus additional 7.000 of Chinese Academic Journals.

11,2 % (1.534) of the OpenURL requests went to freely accessible material such as free access, open access and open archive, which had all been activated through SFX using bulk activation. A fairly comprehensive backfile program seems to amount in this case to about 1 in 7 overall requests for electronically available articles. About 1 in 8 electronic articles requested via OpenURL were freely available.

Discussion 2:
These are interesting first time findings that could serve as a starting point for further research. It shows the relevance of open access and archive content to the user community. It also indicates a significant role the library and OpenURL resolving can play in facilitating the access to freely available literature and increasing the transparency about its usage in general and for specific communities. The standard reports that SFX provides today allow for further breakdown of free

journal targets to the title level. It will be interesting to compare results between different communities, over time and with more differentiation.

3.   What users are looking for, but don`t find in electronic full text
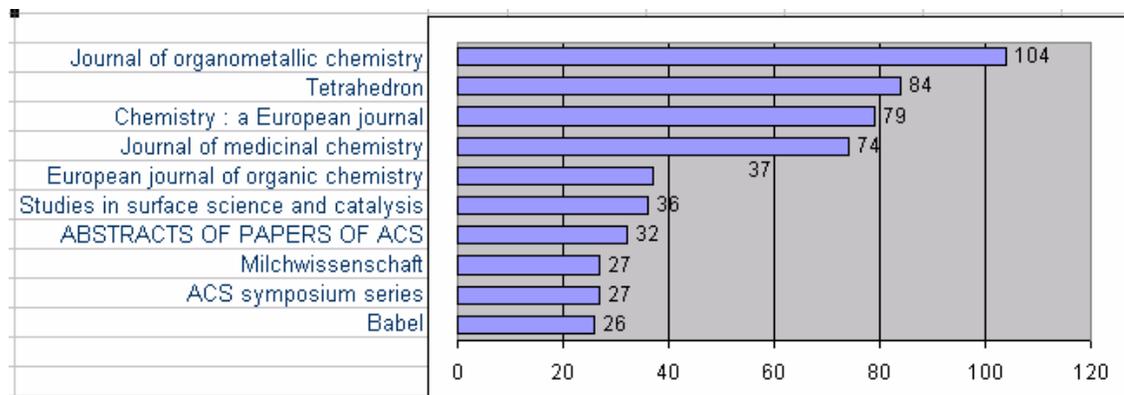

Fig. 3 Top ten Journals requested but not available electronically (from a German institution, January 2007)

While publisher's statistics (e.g. COUNTER) only count what has been electronically accessed (subscribed or complimentary) a library's OpenURL resolver can also detect and log their own user communities' attempts and interest to get to a document even if it is NOT subscribed to or available electronically (for whatever reason).

An example of such a top-10 rank journal titles sorted down by number of requests without electronic fulltext is shown above in fig. 3 (taken from a list of 7.900 requests without fulltext service going to 1.600 different journals) at an institution in one month..

Discussion 3: This shows data about titles that could not immediately be served electronically but were of highest popularity. Similar to interlibrary loan statistics, this can give an indication of current local demand. Compared to ILL however there are some differences:

a) the threshold for the user to generate an OpenURL-request is much lower (just a click) and therefore the overall sample obtained here is much broader and larger.
b) the inbuilt-logic of the SFX link resolving engine normalizes objects over issns and titles and produces consolidated lists out of the very heterogenous requests which are more readily comparable
c) In many cases the "Not available" electronically may be originating from requests to older back volumes, while there is a current online license present. In this case the report can be further correlated with the current volumes requests, and where appropriate aggregated for vendor backfile packages in order to get some indication if purchase of backfile volumes would be economical.

This information, which is easily available on an ongoing basis, can serve as suggested by STENGEL[6] as additional input and complimentary feedback about users' information seeking for use in rational collection development decisions.

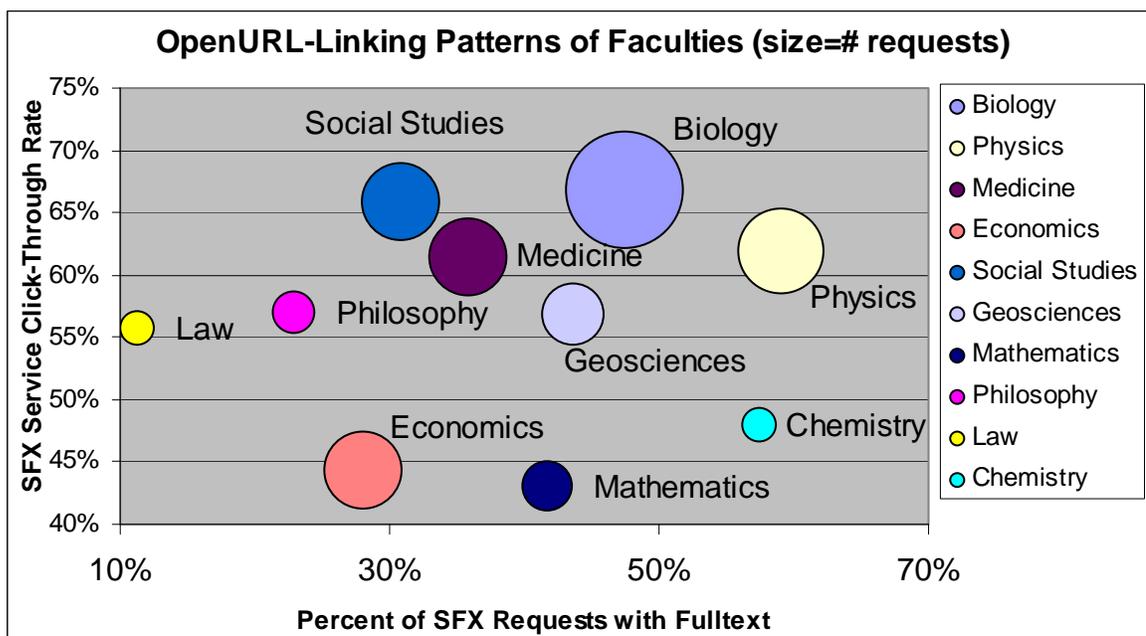**4.   Multi-Criteria analysis – e.g. Usage by faculty**

Fig. 4 – 3-Dimensional representation of different Link Resolver usage-patterns at different faculties of University of Göttingen in 2006.

User-IPs of University of Göttingen were allocated to their respective faculty using the SFX IP-institute feature. Size of circles represents relative number of requests per faculty. The circles are placed in a scatter diagram, where X represents increasing availability of full text per request in percent. Y reflects the clickthrough rate, i.e. the percentage by which user has followed through the link service menu by clicking on a particular service offered.

This experimental illustration was compiled by correlating data derived from different standard SFX reports. It gives a compact overview of the different degrees that the same SFX server is used at different faculties at a university at the same time interval. The graph combines the full text availability rate (X) as a measure of the ability to deliver direct electronic services with the clickthrough-rate (Y) a measure of the overall acceptance by the user of the linking services offered.

The graph reveals and quantifies a strong difference in full text availability via OpenURL across the different faculties (12 to 60%) with Law being lowest and Physics and Chemistry highest and also quite a variation in the clickthrough rates / acceptance (43–67 %) with Economics and Mathematics lowest and Biology highest. There seems to be some correlation between both, indicating that better full text availability produces higher acceptance of linking services provided.

Discussion 4: The difference in full text availability via OpenURL in different subjects (especially in a non-native-English-speaking country) is not surprising; however it is interesting to see how differences can be revealed quantitatively using this approach. However, these results are preliminary and should be interpreted carefully, as they are derived from a case where SFX is still in early implementation and not all possible sources have been activated for all faculties. This kind of statistic may be useful for institutions who wish to undertake rational profiling of their different user communities in order to better understand their behaviour, usage patterns and demands. It

may also help in discussions with database vendors about OpenURL-enabling. As this approach is entirely based on analysing data that can be collected routinely and continuously in the background, and analysed using standard routines it seems to be well-suited for periodical follow-ups on the development of a university's communities and library portfolio.

Final Discussion

A few practical applications of using OpenURL statistics in specific context have been presented above. Numerous others exist but because of space, can't be mentioned in detail here.

The enormous potential of this approach stems from the OpenURL and resolver design, whereby the institution's resolver acts as a central hub for connecting digital library content and services. It covers a very broad range of information items with multiple characteristics, which serve as "atoms" for flexible further aggregation and analysis.

The key advantages lie in:
- A comprehensive and generic approach for the handling of bibliographic and similar service requests and associated logging.
- The fact that OpenURL requests cover almost the entire scholarly document landscape -- not only including commercial publishers but equally open access publications and digital repositories; as of April 2007 many hundreds of data sources are OpenURL compliant; some 50.000 ejournals and hundreds of thousand ebooks are targets for SFX.
- The institution's SFX usage logging database, the unique place where all consolidated usage information is stored in a standardized format and logically accessible.

As a consequence of this wide scope of OpenURL and NISO-standardization it has – now in its 7th year of existence – found broad adoption amongst academic libraries. Many institutions report on their usage within their intranets. An increasing number of institutions find it useful to regularly communicate their basic SFX usage openly on the web – be it annually[3], quarterly, monthly or even weekly[1]. The number of parameters monitored varies; besides requests and clickthroughs, the full text availability rate seems to be generally considered as particularly valuable for patrons and management equally. The California Digital Library, for example, shows a considerable continuous quarterly increase of requests but also of full text availability over the last 3 years[2]. We feel that these objective performance indicators have the potential to become standard parameters for library's annual reporting and benchmarking. Management and sponsors should be aware of them. Changes in them readily indicate trends.

SFX also completes the picture of usage patterns across resources and their "connectivity" through combined source-target statistics[5]. Institutions begin to do "data-mining" with the logging information or use it for specific questions like head-to-head comparisons of different interfaces for similar database against each other over time. Institutions mention quite consistently, that looking at source statistics reveals data-sources that were much more or less frequently used than expected.

SFX usage logging is institution- and user-centric – in contrast to vendor statistics (e.g. COUNTER). OpenURL does not record the direct accesses to full text through homepage links and browsing. But vendor statistics, where and when available, are always selective and aggregated. They provide what is used usually on a journal level. Whereas SFX stats also tell institutions a lot more about the complete context, such as where the user are coming from and what they want that they may not get (in a more complete way than ILL stats do,).

The relative value of SFX linking statistics has been increasing over the years because the statistics reflect a rapidly growing number and proportion of the electronic activities of users and thus can generally be considered a robust sample, particularly for comparative analysis. Harvard University Libraries report an increase in proportion of "attributable deep-links through SFX" by a factor of

about ten to approx. 20% of all library logins in the period from 2003 to 2006[3]. Similar to Harvard other institutions plan to introduce integrated vendor and SFX usage reporting on title level.

To summarize: OpenURL usage statistics can give management an objective instrument to determine the status and evolution of their electronic library services in a comprehensive and easily understandable form. Regular and timely availability of this kind of information provides valuable additional guidance for management decisions on collection development.

As an outlook to possible future studies we would like to mention a few aspects of interest: On a "macro-level" comparing normalized usage data between libraries as parameters for benchmarking of library performance and "source gap analysis" to help libraries extend their services offerings and mutually benefit from free information sources. On a "micro-level": the rich information in the context objects can – based on existing infrastructure - reveal an interesting granularity of information well below journal level down to individual articles.

Ex Libris is determined to continue with investments into R&D in this field to improve and extend context sensitive services for the scholarly community.

References

1.   http://www.cdlib.org/inside/projects/uc-elinks/
2.   http://www.cdlib.org/inside/projects/uc-elinks/overall.trends.xls
3.   http://hul.harvard.edu/ois/systems/portal/stats.html
4.   http://www.nationallizenzen.de
5.   J KREISCHE – 2006, Die Messung von Vernetzung, 9. InetBIB-Tagung,, 6.-8. September 2006, Münster
6   STENGEL, M.G., Using SFX to Identify Unexpressed User Needs, Collection Management, 2005, May 24, pp. 7.14 (DOI: 10.1300/J105v29n02_03)