# THE VALUE OF OUR DIGITAL COLLECTIONS AND SERVICES BASED ON LIBRARY STANDARDS

Rita Radó

**Evaluation is the best tool to learn and understand a system. Ideally, evaluation is incorporated into all the activities and processes of a library and it helps librarians rethink the traditional library principles in a fast-moving world. There are library standards that give us great help in this process. (ISO 11620: 2014 Information and documentation. Library performance indicators, and ISO 16439: 2018 Information and Documentation: Methods and Procedures for Assessing the Impact of Libraries.). Here, the Electronic Periodical Archives and Database (EPA) was analysed with the above-mentioned guides from the very specific viewpoint of evaluating its services and collection. EPA is edited, developed and operated by the E-Library Services Department of the National Széchényi Library (NSZL). It is an archive management system, equipped with source and location registration system of electronic periodicals, and also a national web site database operated by the NSZL. The main reason I used International Standards for this analysis is that the library standard supports the use of performance indicators regarding the quality of library services in libraries and to spread knowledge about how to conduct performance measurement. It also provides guidance on how to implement performance indicators in libraries where such indicators are not already in use. With these, long-term effects may appear on user satisfaction indicators, on library service policy, and it may also contribute to improved scientific research results for users, and can enhance their competence by providing them with a higher quality level of digital documents.**

**Keywords:** library standards, National Széchényi Library, Electronic Periodical Archives and Database, performance indicator, impact of libraries, evaluation, digital collection, digital services

rado.rita@oszk.hu
National Széchényi Library
Budapest, Hungary
http://www.oszk.hu/en

## EVALUATION AND PERFORMANCE MEASUREMENT

Evaluation is the best tool to learn and understand a system. However, evaluation and operational research have certain processes and methods. Ideally, the evaluation culture is incorporated into all the activities of the library. Efficiency increases dramatically when it becomes an integral part of quality management. The evaluation of a library or one of the library's services always begins with the evaluation of its system-level – broader or closer – environment in which the library is located. Measuring performance is part of the evaluation, so measuring performance or output is a must when evaluating libraries. Numerous procedures and indicators have been developed to measure librarian activities and services. Our metrics and methods that guide our decisions are increasingly sophisticated.

The library standards can be very helpful in this because the library standards are practice-oriented tools that serve as benchmarks in library workflows. Their basic function is to expressly communicate the relationship between library users and library services and to regularly apply the continnous change for better and more efficient operation. The standards provide a consistent basis for comparison.

The spread of these measurements can be attributed to several interrelated factors. For example,

- the more and more library researches,
- the increasing size and complexity of libraries,
- the increase in the number and quality of services,
- the increase in the support provided by the maintainer, and more control over their use, or
- for more efficient planning work.

## LIBRARY STANDARDS FOR MEASURING AND ANALYSING LIBRARY PERFORMANCE

Librarian Performance Indicators are addressed in *ISO 11620:2014 Information and documentation. Library performance indicators.*

*„The main purpose of this International Standard is to endorse the use of performance indicators regarding the quality of library services in libraries and to spread knowledge about how to conduct performance measurement. This International Standard specifies the requirements of a performance indicator for libraries and establishes a set of indicators to be used by libraries of all types. However, not all performance indicators apply to all libraries. It also provides guidance on how to implement performance indicators in libraries where such indicators are not already in use."* (ISO 11620:2014) The standard defines requirements and introduces a quantitative and qualitative metric system for library performance metrics that can be applied to any library type. Increasing focus on performance measurement in the library area is also due to the growing need for maintainers to demonstrate the benefits or uselessness of the institutions, and employers also want to see clearly the performance of their employees during working hours (Dippold 2008). However, performance indicators – in my opinion – can only fulfill their original goals if they come from the conclusions of correctly applied performance indicators. Only then they can be applied in the operation of the library, in the services and in their development.

Performance indicators also help you analyse the impact of libraries, which is another standard discussed in *the ISO 16439:2018 Information and Documentation. Methods and Procedures for Assessing the Impact of Libraries.*[1]

## PERFORMANCE INDICATORS FOR ELECTRONIC COLLECTIONS AND SERVICES IN ACCORDANCE WITH ISO 11620:2014 INFORMATION AND DOCUMENTATION. LIBRARY PERFORMANCE INDICATORS

In the standard ISO 11620:2014, the terms of electronic resources, electronic collections and the use of electronic services are surprisingly underrepresented. There is good reason to believe that the authors of the standard acknowledged that both the activities and the list of indicators have taken the possibility of future additions into account when designing the marking system.

In case of the resources, there is an index for the *Number of documents digitised per 1,000 documents in the collection.* (ISO 11620:2014 (B1.1.4)) This indicator is used to assess to what extent the library fulfils its task of making the documentary heritage publicly available in digitised format. The indicator can be used in each library, which includes the task of pre-

servation and promotion of documentary heritage. The high value of the indicator shows that the library is committed to making its collection available. However, in the case of further analyses, it should be taken into account that the indicator is influenced by the financing of local, regional or national heritage programs. There are two indicators for measuring the use of library resources and services for electronic resources. One examines the *Number of Content Units Downloaded per Capita* (ISO 11620:2014 [B 2.1.4]) and the other is the *Number of Downloads per Document Digitized* (ISO 11620:2014 [B 2.1.5]). In my opinion, both indicators, measure (indirectly) the usefulness and efficiency of the given collection.

In the first case, the indicator is used to assess whether users find items of interest in an electronic resource and in the second case, we can further investigate whether the library has digitised documents that are relevant and interesting to users. Both indicators may prove to be very useful for further development or reconsideration of the digitisation plan. However, the standard draws attention to the fact that the indicator can be influenced by a number of factors, many of which are outside the control of the library (for example user skill level, network access level, whether users are charged for access or download, and promotion of services.) Additionally, the number of downloaded content units can be highly influenced by the quality and effectiveness of the user's search strategy.

Among the benchmarks of this standard focusing on potentials and enhancements, two additional performance metrics can be found regarding electronic collections. One refers to the cost of electronic collections: *Percentage of Expenditure on Information Provision Spent on the Electronic Collection* (ISO 11620:2014 [B 4.1.1]) and the other *to Percentage of Library Staff Providing Electronic Services* (ISO 11620:2014 [B 4.2.1]).

Both of these indicators may be used to make comparisons between libraries, taking into account the differences in material coverage, in the collection building policy and in the socio-economic factors of the served audience. Based on these performance metrics it can be determined how committed the library is in creating electronic collections.

The above examples are very well reflecting the fact that quantitative and qualitative methods can be combined with surveys that can lead to the efficient operation of the institution and also the innovative development of digital services. The results of the measurement indicators in the standard are, by definition, quantitative data obtained quantitatively, but qualitative analysis of the values can make very serious impact assessments and development suggestions.

## PERFORMANCE INDICATORS FOR ELECTRONIC COLLECTIONS AND SERVICES IN ACCORDANCE WITH ISO 16439:2018 INFORMATION AND DOCUMENTATION. METHODS AND PROCEDURES FOR ASSESSING THE IMPACT OF LIBRARIES

However, the impact of library services on individuals or the communities served by the library is discussed in another standard, *the ISO 16439:2018 Information and Documentation. Methods and Procedures for Assessing the Impact of Libraries.* Libraries have developed and tested methods to validate their yields for users and society. The quality indicators used by libraries for this purpose show the impact of libraries on individuals and on society. These effects can be immediate (finding the information you need) or long-term effects (raising your literacy); far-reaching (changes affecting people's lives over the long term, such as well-being, satisfaction,) or limited (minor improvements in certain skills); intended or unintended effect; actual or potential (heritage conservation) benefits. Its impact can be individual, institutional or social.

# THE DIGITISATION WORKFLOWS OF THE NATIONAL SZÉCHÉNYI LIBRARY

The primary purpose of digitising the National Széchényi Library (NSZL) collection is to "*provide authentic source material for scientific research*" (Dancs, 2010). Secondary aim, however, is to "*develop other services based on authentic digital corpus*". (Dancs, 2010)

The introduction of digitised documents or born-digitised documents into the service system is always a difficult task, because either it needs to stay within the framework of the library's service system, or it may be necessary to develop a new service or service policy – which in many cases requires a re-

consideration of the strategy. It is a question of what technical means and which set of rules will provide credibility and authenticity. There are countless examples of such uniformly applicable rules and systems, these guidelines also help to analyse both quantitative and qualitative methods.

A Digitisation Centre will be set up under the *IT renewal project* in the NSZL. At present there is a development framework for digitisation. In this context, the *Digitisation Commission* in the library defined the quality requirements for image files produced by digitisation (or received from external institutions). For this we have taken the FADGI guidelines.[2]

The guidelines determine four quality levels of imaging according to the quality to be produced. (From 1 to 4). Higher star ratings are related to more consistent and concise image quality but at the same time require greater technical performance from the operator and imaging system and more digitisation expertise. A *one-star image* is sufficient for applications where the goal is to find only a reference to the original document, or the intention to render is textual, without reusing the content. The quality of these images is not suitable for optical character recognition or other information processing techniques.

*Two-star imaging quality* images are still only good for reference and are generally not suitable for OCR. We use this quality level when there is no reasonable expectation that they will be able to achieve three- or four-star quality performance.

*The three-star imaging level* defines a high-quality, professional-grade image that is suitable for almost any application.

*Four-star imaging* is the best imaging practice available today. Created on a four-star level, images represent the latest techniques in image capture and are suitable for any application.

Generally speaking, FADGI does not recommend digitising at a lower quality level than (three-star or-) triple imaging to avoid future re-digitisation and to avoid increased efforts and costs for digitisation projects. At the same time, this assumes the availability of good quality digitising equipment that meets the FADGI evaluation criteria and produces image files that meet the minimum quality requirements described in the Technical Guide. If the digitisers do not meet the evaluation criteria and/or are not able to produce image files of the appropriate quality, then according to FADGI's recommendation, it may be desirable to obtain better equipment or to contract with an external digitiser.

FADGI Star System is a very good example of how we can achieve higher quality results by developing (e.g. providing numerical values associated with different levels of quality) and applying quantitative units. With these, long-term effects may appear on user satisfaction indicators, on library service policy, and it may also contribute to improved scientific research results for users, and can enhance their competence by providing them with a higher quality level of digital documents.

## ANALYSIS OF THE ELECTRONIC PERIODICAL ARCHIVES AND DATABASE ACCORDING TO ISO 11620: 2014

The *Electronic Periodical Archives and Database* (EPA) is an archive managing electronic periodicals equipped with a source and location registration system; it is a separate collection of NSZL electronic periodicals, and also a national web site database operated by the NSZL provides publications with bibliography and article bibliographic data. There are currently close to 1025 archived publications, 100.000 issues, 612.000 articles. EPA is edited, developed and operated by the e-Library Services Department of the National Széchényi Library.[3]

In one hand, the EPA service seeks to comprehensively map Hungarian online and digitised electronic journals and press products, while on the other hand, archives major scientific and cultural journals. The Electronic Periodical Archives and Database is a multi-level system that essentially performs two functions: it records and archives. It also records the so-called distant items – these records describe the time interval at which an electronic periodical was published, and the URL (if) available.

Archived items aim to preserve electronic periodicals over the long term. The files can be accessed directly from the EPA server, so links do not point to a remote URL. A table of contents is made where necessary. In order to ensure the legal framework for archiving, a so-called User Agreement, in which publishers agree to archive e-journals in the NSZL and make them publicly available.

The offline items represent a special group. These are mostly digitised versions, which, however, are not available directly on the Internet, but can be used in

a manageable user format on local networks or media (such as CD-ROM editions or digitized versions provided from the local file system). (Uri-Kovács, J., 2017)

The social impact of The Electronic Periodical Archives and Databases is due, among other things, to the preservation nature of the service, as retrieving and archiving documents is of enormous help to users. Traceability is facilitated by the fact that objects in EPA can be individually searched, found, and visited. Not only on the web site of the maintainer's institution (epa.oszk.hu), but also by Google's search engine: they appear independently in their search results, thanks to the fact that Google's search engine's relatively quickly shortlists the content uploaded to EPA. This provides a qualitative increase in the usability of the database, which can be detected statistically.

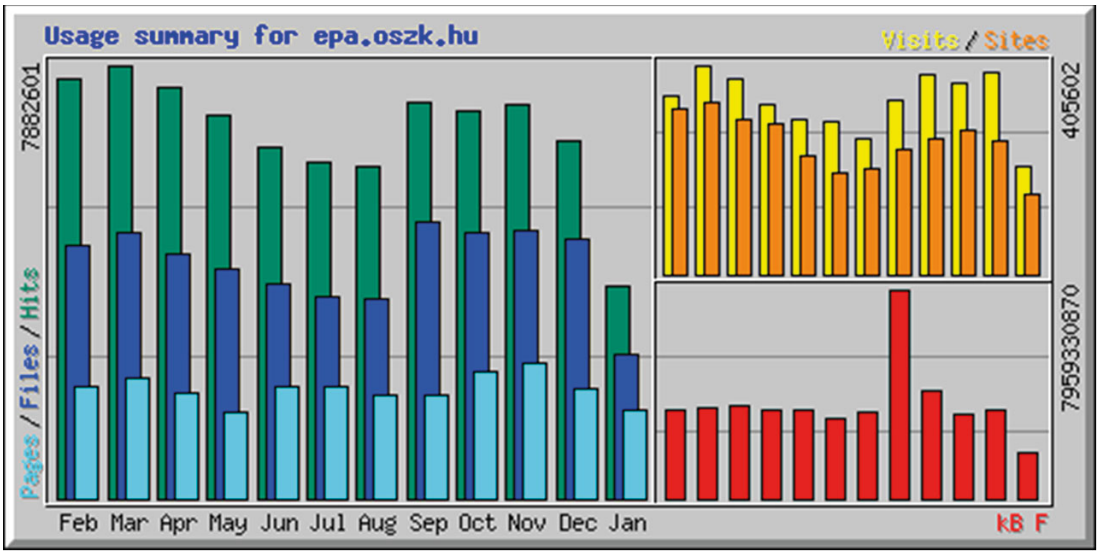


Figure 1 Web stats from epa.oszk.hu on 15. January 2020 (www.epa.oszk.hu)

Based on the diagrams illustrating The Electronic Periodical Archives and Database statistical data, the library can measure the popularity of the service (cf. ISO 11620:2014 B.2.1.4).

The first graph (Figure 1) shows the number of visitors, visited pages, files, and clicks, and the kB values of downloaded files in a monthly breakdown for 12 months. Table 1. below is a more detailed version of Figure 1. for deeper and more thorough analysis. Here we see monthly summaries with daily averages and full monthly values. *Hits* represent the total number of requests made to the server during the given period (month, day, hour, etc.). *Files* represent the total number of hits (requests) that actually resulted in something being sent back to the user. (Not all hits will send data, such as 404-Not Found requests and requests for pages that are already in the browser's cache). *Pages* are those URLs that would be considered the actual page being requested, and not all of the individual items that make it up (such as graphics and audio clips). Some people call this metric page views or page impressions and default to any URL that has an extension of .htm, .html or .cgi. *Visits* occur when a remote site makes a request for a page on your server for the first time. As long as the same site keeps making requests within a given timeout period, they will all be considered part of the same Visit. If the site makes a request to your server, and the length of time since the last request is greater than the specified timeout period (default is 30 minutes), a new Visit is started and counted, and the sequence repeats. Since only pages will trigger a visit, remotes sites that link to graphic and other non-page URLs will not be counted in the visit totals, reducing the number of false visits. The *Sites* is the number of unique IP addresses/hostnames that made requests to the server. Care should be taken when using this metric for anything other than that. Many users can appear to come from a single site, and they can also appear to come from many IP addresses so it should be used simply as a rough gauge as to the number of visitors to your server. *Kbytes* used to show the amount of data that was transferred between the server and the remote machine, based on the data found in the server log. (1 kilobyte).

Table 1 Figure 1 in numbers (www.epa.oszk.hu)

| Summary by Month | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Month | Daily Avg | | | | Monthly Totals | | | | |
| | Hits | Files | Pages | Visits | Sites | kB F | Visits | Pages | Files |
| Jan 2020 | 258228 | 173516 | 106433 | 13870 | 154403 | 1727630472 | 208050 | 1596504 | 2602747 |
| Dec 2019 | 210148 | 152808 | 63975 | 12620 | 258592 | 3381044210 | 391233 | 1983255 | 4737072 |
| Nov 2019 | 238642 | 162484 | 81865 | 12384 | 280754 | 3215310470 | 371531 | 2455968 | 4874544 |
| Oct 2019 | 226740 | 156299 | 74964 | 12444 | 262813 | 4062454065 | 385777 | 2323901 | 4845276 |
| Sep 2019 | 240278 | 167148 | 62856 | 11260 | 240679 | 7959330870 | 337825 | 1885705 | 5014456 |
| Aug 2019 | 195095 | 117056 | 60700 | 8471 | 206185 | 3291771844 | 262623 | 1881730 | 3628761 |
| Jul 2019 | 197735 | 118143 | 65866 | 9531 | 198185 | 3015689289 | 295481 | 2041873 | 3662435 |
| Jun 2019 | 212990 | 130364 | 67788 | 10036 | 228083 | 3403764589 | 301102 | 2033658 | 3910946 |
| May 2019 | 224330 | 134717 | 50380 | 10678 | 293665 | 3375931026 | 331045 | 1561780 | 4176251 |
| Apr 2019 | 249708 | 147702 | 64117 | 12639 | 300639 | 3536506786 | 379178 | 1923516 | 4431086 |
| Mar 2019 | 254277 | 156164 | 70497 | 13083 | 334739 | 3467445859 | 405602 | 2185418 | 4841090 |
| Feb 2019 | 272065 | 164681 | 72209 | 12283 | 319169 | 3377333776 | 343937 | 2021873 | 4611084 |
| Totals | | | | | | 43814213256 | 4013384 | 23895181 | 51335748 |

Based on the content evaluation charts (cf. ISO 11620: 2014 B.2.1.5), it can also be measured what are the most interesting documents for users. All this contributes greatly to the development of a proper service policy.

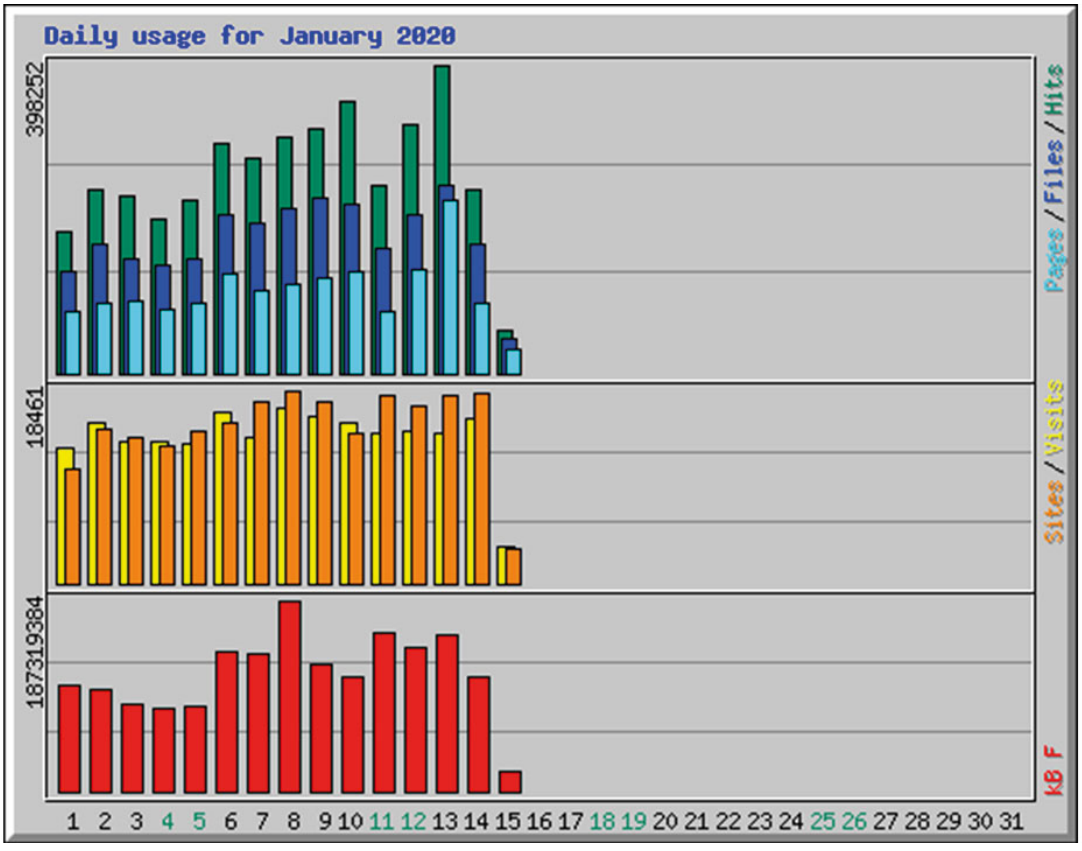


Figure 2 Graphic representation of EPA daily usage statistics in January 2020 (www.epa.oszk.hu)

There are also statistics broken down by day and time in the same categories. These values are less suitable for impact assessment, but they help to analyse the efficiency of the service, show the time distribution of the daily usage, and the service can be further developed according to the needs of the users.

| Top 200 of 786185 Total URLs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| # | Hits | | kB F | | kB In | | kB Out | | URL |
| 1 | 172821 | 2.09% | 757935 | 0.02% | 0 | 0.00% | 0 | 0.00% | /html/vgi/kardexlap.phtml |
| 2 | 132735 | 1.60% | 43738200 | 0.95% | 0 | 0.00% | 0 | 0.00% | /webalizer/usage_201708.html |
| 3 | 63354 | 0.76% | 708870 | 0.02% | 0 | 0.00% | 0 | 0.00% | /webalizer/ |
| 4 | 45891 | 0.55% | 244033 | 0.01% | 0 | 0.00% | 0 | 0.00% | /html/vgi/boritolapuj.phtml |
| 5 | 41104 | 0.50% | 122939 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/ujepa_keres/key_eredmeny_eng.phtml |
| 6 | 38094 | 0.46% | 1036309 | 0.02% | 0 | 0.00% | 0 | 0.00% | / |
| 7 | 34459 | 0.42% | 64059 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/kereses/key_eredmeny.phtml |
| 8 | 28461 | 0.34% | 120101 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/expand.phtml |
| 9 | 26283 | 0.32% | 43245 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/index_xml.php3 |
| 10 | 25842 | 0.31% | 1679 | 0.00% | 0 | 0.00% | 0 | 0.00% | /robots.txt |
| 11 | 25621 | 0.31% | 117667 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/expand_long_eng.phtml |
| 12 | 25583 | 0.31% | 29961 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/kapcsolat.phtml |
| 13 | 25422 | 0.31% | 98749 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/ujepa_borito.phtml |
| 14 | 25360 | 0.31% | 114272 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/expand_long.phtml |
| 15 | 25171 | 0.30% | 73344 | 0.00% | 0 | 0.00% | 0 | 0.00% | /html/vgi/expand_links.phtml |

Figure 3 EPA's 200 most popular URLs-extract (www.epa.oszk.hu)

The statistical measurements presented so far contained all numerical values. The following tables, on the other hand, measure usage on the basis of quantitative and content data that allows for more complex analysis.

Figure 3 shows the 200 most popular URLs from the 440027 URL in the EPA based on the number of "clicks" and the number of kBs downloaded. (I only show the first 15 due to space restrictions.). The following figure illustrates the content of searches:

| Top 20 of 436 Total Search Strings | | | |
|---|---|---|---|
| # | Hits | | Search String |
| 1 | 16 | 3.14% | https://epa.oszk.hu/00000/00022/nyugat.htm |
| 2 | 6 | 1.18% | hajnali reszegseg |
| 3 | 4 | 0.79% | boldog szomorĂş dal |
| 4 | 4 | 0.79% | kĂĄvĂŠhĂĄz presszĂł saly |
| 5 | 4 | 0.79% | verebelyi galuska |
| 6 | 3 | 0.59% | az albatrosz |
| 7 | 3 | 0.59% | childe harold bĂşcsĂşja |
| 8 | 3 | 0.59% | csend van csend van |
| 9 | 3 | 0.59% | edgar allan poe a hollĂł |
| 10 | 3 | 0.59% | evolĂşciĂł bĹąn |

Figure 4 The top 10 most common search queries from 436-extract (www.epa.oszk.hu)

Based on the first figures showing the EPA's statistical data, the library can measure the popularity of the service based on users and usage (KB, click, etc.) (cf. ISO 11620:2014 B.2.1.4). This is very good feedback on when you may need to change your service policy, examine your preferred periods, develop a technical background, or allow for a "quick response" in the event of a more prevalent user downgrade. However, the latter two figures are suitable for evaluating content (cf. ISO 11620: 2014 B.2.1.5). It shows what documents are really interesting for users and helps to assess whether the library has actually digitised documents that are of interest to users.

From the quantitative indicators presented so far – which is being registered by the EPA in great detail and cautiousness – we can see that this is a popular and vital service, the impact of which is very diverse. Among its effects on individual users, we must highlight convenience. Users can reach documents even from home, with a couple of clicks. Moreover, the

online use of the database is not subject to an NSZL enrollment or membership.

In addition to the above, the social impact of the EPA is mainly due to its archiving policy. Archiving and retrieving documents is invaluable in scientific research. There are countless examples of different organizations contacted the EPA to help them preserve the digitised copies of their magazine and involve them in services, mostly because their previous service providers went out of business (e.g. School Culture, Journal of Swedish-speaking Hungarians).

The outdated service area of EPA has been criticized because it was old-fashioned, lagged behind today's technical and digital expectations. However, there was a user group that preferred this simplicity: for the visually impaired users' reading-software this kind of interface was very ergonomic and well managed (e.g. serial display of text instead of columns; no graphics; the interface can be used without a mouse, etc). All of these feeds later lead to the implementation of an (unforeseen) user-friendly development and have launched the VMEK service for blind and partially sighted people, another service of the same department, the Hungarian Electronic Library (MEK).

We may conclude, that although the appearance of the site does not meet today's expectations, the easy-to-understand usage, the user-friendly search interface, the high proportion of responses to searches, the large amount of data available, and the ever-expanding database make the EPA a very popular service of the National Library. Maintainers and developers pay great attention to the values and feedback of their performance indicators and operate their services accordingly.

## SUMMARY

There is a wealth of methods for effective evaluation and measurement, but I think that the above examples also demonstrate that the combined use of quantitative and qualitative indicators can bring very effective and innovative results. By analysing and evaluating the actual and underlying content of the data obtained from the surveys and measurements, a library can indeed perform better, which serves both the users and the maintainer.

Digital collections and digital services for libraries are an opportunity and a task at the same time. We should not miss the very complex and multi-faceted path that needs to be traveled for an efficient and highly user-friendly electronic service. These are all aspects that need to be analysed individually, both quantitatively and qualitatively, to create the right operation and user-friendly services.

## NOTES

[1] The adoption of both standrads was initiated and funded by the National Széchényi Library in the framework of National Library System project. The National Library System Project is the ongoing IT Development Project of the National Széchényi Library.

[2] Federal Agencies Digitisation Guidelines Initiative (FADGI). Started in 2007, this is a collaborative effort by federal agencies to define common guidelines, methods, and practices for digitising historical content.

[3] www.epa.oszk.hu

## REFERENCES


DANCS, Sz. (2010). Authentic digital copy and bibliographic management of copies – Part 1: Authenticity on a standard basis. [in Hungarian]. *Könyv, Könyvtár, Könyvtáros*, No.10, 7 – 15.

DIPPOLD, P. (2008). Virtual tour at the Quebec IFLA conference [in Hungarian]. *Könyvtári Figyelő*, 3, 463 – 468.

Electronic Periodical Archives and Database Digital Journals – http://efolyoirat.oszk.hu/static/politikank.html [2020.01.31.]

EPA web-statistics http://epa.oszk.hu/webalizer/ [2020.01.31.]

ISO 11620:2014 Information and documentation. Library Performance Indicators

ISO 16439:2018 Information and documentation. Methods and procedures for assessing the impact of libraries

RADÓ, R. (2018). Library Organization Development in the Age of Digitization – Opportunities and Problems. Proceedings of at the K2. *University Technology Research Conference, Budapest, National Széchényi Library*.

URI-KOVÁCS, J. (2017). *The National Database of Electronic Periodicals – EPA – in the NSZL*. Proceedings of at Networkshop, Szeged.