Testing library catalog analysis as a bibliometric indicator for research evaluation in Social Sciences and Humanities

Abstract
This paper offers the results of an examination of library catalog analysis, a bibliometric approach in evaluating scientific research in Social Sciences and Humanities based on quantitative analysis of the presence of monographs in online library catalogues. The test was carried out in 2015-2016 in a large number of Italian and foreign libraries, highlighting the nonhomogeneous distribution of monographs in libraries. The survey demonstrated that calculating the presence of books in library holdings could not be considered a reliable indicator in research assessment. Moreover, the paper examines in depth recent tools for monograph search (discovery) and the Semantic Web standards for author identification.

Introduction
Research assessment is a scientific task that entails significant social aspects, as it aims at evaluating the scientific activity of each researcher, involving the activity of academic departments and the cultural policies of academic authorities.

The evaluation of research in Social Sciences and Humanities (SSH) is based on the analysis of books as the product of research, currently adopting the strategy of expert peer review. Henk Moed (2005) reported the problems that arise in SSH fields using bibliometric counts based on citations. Instead of computing the citations received by articles that scholars publish in journals, D. Torres-Salinas and H. F. Moed (2009), A. J. M. Linmans (2010), and H. D. White et al. (2009), autonomously proposed library catalog analysis\(^1\) as a bibliometric tool for research evaluation in SSH. Recently, Alesia Zuccala and Raf Guns (2013) again proposed the Libcitation count, previously suggested by White. Library catalog analysis is a new bibliometric approach that considers the number of copies of books in libraries using online catalogues, instead of counting the citations received by authors. It aims to evaluate the authors’ productivity in SSH and their scientific level.

The new bibliometric indicator for research evaluation in SSH proposed through library catalog analysis is a matter of importance in the knowledge organization (KO) field, as it involves library acquisition procedures and implementation of OPACs.

Objectives
The aforementioned in-depth analysis carried out covered a limited range of scientific fields: Economics (Torres-Salinas and Moed 2009), History and Literature (Zuccala and Guns 2013). Beyond these, the literature does not offer a wide empirical

\(^1\) According to Linmans, the correct definition is “Library Holdings Analysis”.
analysis of the potential and limitations of that strategy over a broad range of disciplines. Therefore, a deeper examination of scientific fields other than those tested by the authors mentioned is required.

This paper presents the concise results of an Italian survey on the diffusion in Italian and foreign library holdings of books published by Italian scholars in two scientific fields: history of books, bibliography, library science on one hand, and history of political institutions on the other. The survey, sponsored by the Italian Agency for the Evaluation of the University System and Research (ANVUR), was carried out in 2015 and 2016 with the purpose of verifying the possibility of using the results of analysis as an indicator in SSH research assessment.

Among the objectives of the research, was an analysis of the nature and operation of discovery tools in order to understand if these tools can affect the qualitative and quantitative outcomes of research into monographs. Discovery systems are software that appeared on the market around 2009 with features similar to the federated search tools created for electronic resources. Moreover, the survey aimed to highlight the possibilities offered by the new standards in the Semantic Web framework for the identification of authorships, which is a crucial point in research evaluation.

**Methods**

Authors who proposed the library catalog analysis mostly used unit catalogues such as WorldCat (Torres-Salinas and Moed 2009; Linmans 2010). In our survey, we preferred instead to verify the presence of books in a set of OPACs we chose as representative of different environments: 9 foreign libraries and 13 Italian libraries, including national libraries and libraries of prestigious European and US universities. Therefore, our work was not restricted to offering only quantitative results, but rather acquired a qualitative dimension. We set up a database of monographs published between 2000 and 2014 selecting them from the Italian SBN collective catalogue. The database includes 563 units: 279 in history of books, bibliography and library science, and 284 in history of political institutions.

Concerning the evaluation of discovery systems as tools for the research of monographs, a sample of monographs from the database used for our survey was also selected for testing discovery systems. The survey analyzed only a sample of the two scientific fields’ monographs: history of books, bibliography, library science and history of political institutions; then the results were compared with those of the same titles carried out in the respective libraries’ online catalogues.

**Main results**

Our survey produced several important results. First, the analysis of the holdings of Italian and foreign libraries allowed us to have a wide view and to compare different realities and make an articulate analysis. Our examination revealed a very mixed picture
and a substantial difference in the presence of monographs in the libraries in which the surveys were completed. Second, we stressed that key elements conditioning the inclusion of monographs in library holdings are numerous, vary according to the type of library and are often influenced by contingencies. Third, in evaluating the inclusion of books, we considered the incidence of gifts. None of the authors here mentioned, who proposed the library holdings as a reliable indicator of the scientific level of monographs, considered the incidence of gifts in library holdings.

Table 1 presents an aggregate view of the distribution of monograph publications in Italian and foreign libraries. Table 2 gives the detailed distribution of Italian monographs in national and academic foreign libraries, while Table 3 shows the detailed distribution of the same in national and academic Italian libraries. It is to be noted that Italy has 9 national libraries. In Table 3 includes details of 7 national libraries, as the central national libraries in Florence and Rome were not included, as they are mandatory repositories according to the Italian legal deposit act. Regarding the presence of monographs by Italian scholars in national and academic foreign libraries, the Harvard University libraries revealed the highest level of presence.

Table 1: Presence of monographs in the Italian and foreign libraries

<table>
<thead>
<tr>
<th>DISTRIBUTION OF PUBLICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITALIAN UNION CATALOGS SBN</td>
</tr>
<tr>
<td>FOREIGN ACADEMIC</td>
</tr>
<tr>
<td>FOREIGN NATIONAL</td>
</tr>
<tr>
<td>ITALIAN ACADEMIC</td>
</tr>
<tr>
<td>ITALIAN NATIONAL</td>
</tr>
</tbody>
</table>

Table 2: Presence of Italian monographs in national and academic foreign libraries

<table>
<thead>
<tr>
<th>DISTRIBUTION IN NATIONAL AND ACADEMIC FOREIGN LIBRARIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNE</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>69</td>
</tr>
</tbody>
</table>
Table 3: Presence in national and academic Italian libraries

<table>
<thead>
<tr>
<th>CS</th>
<th>PZ</th>
<th>TO</th>
<th>NA</th>
<th>BA</th>
<th>VE</th>
<th>MI</th>
<th>CAG</th>
<th>VEA</th>
<th>RMS</th>
<th>UTO</th>
<th>UMC</th>
<th>USM</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>13</td>
<td>18</td>
<td>30</td>
<td>90</td>
<td>32</td>
<td>45</td>
<td>100</td>
<td>62</td>
<td>127</td>
<td>68</td>
<td>58</td>
<td>139</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>10</td>
<td>32</td>
<td>5</td>
<td>32</td>
<td>61</td>
<td>58</td>
<td>46</td>
<td>216</td>
<td>153</td>
<td>153</td>
<td>329</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

CS: Cosenza National library; PZ: Potenza National library; TO: Torino National and University library; NA: Napoli National library; BA: Bari National library; VE: Venezia National library; MI: Milano National library; CAG: Cagliari University library system; VEA: Venezia University library system; RMS: Roma Sapienza University library system; UTO: Torino University library system; UMC: Macerata University library system; USM: Milano State University library system.

In our survey, we found that occurrences are not always the result of conscious choices, but are influenced by several factors, such as cooperative policies for acquisition of publications, purchasing plans, approval plans and patron driven acquisition or particular forms of giving. Approval plans are a widespread acquisition method in libraries, based on the development of collections’ profiles and formal agreements signed between libraries and service providers. Title choices and purchase decisions are also often guided by users through a service called patron driven acquisition by means of which orders are generally suggested using next generation online catalogs.

The distribution of monographs in libraries analyzed in our survey is not homogeneous, whether in Italian or in foreign libraries. With respect to the field of history of books, bibliography, library science, one of the most significant results of our test was evidence that monographs on the history of books are preeminent in Italian and foreign libraries.

Another important result arising from our survey was its study of the operation of discovery systems as tools for research into monographs. A discovery system is a
centralized index of contents and metadata resulting from multiple different sources, which is combined with a multi-functional research system (Christensen 2013). It is presented as the solution to the problem of separately querying multiple indexes of collections that libraries give access to. The latest discovery tools perform a federated search (Wanga, Mi 2012) among all bibliographic library collections, even though their main feature is a central index which contains bibliographic records metadata, digital content in full text, metadata of institutional collections and online resources which the library gives access to through an indirect subscription, since they are included in the discovery system licence.

Web scale discovery systems thus allow users to query the various resources that libraries make available to them through a single access point. Discovery tools are index-based systems: all database contents, both local and remote, are re-indexed by discovery, including those subscribed to by the library with commercial agreements and catalogue data. Indexing and retrieval systems in the research phase are not clear, and there are no standards regulating this process, which remains, together with relevance criteria, completely outside librarians’ control (Breeding 2015).

Analysis of discovery tools use show that they are consulted above all as a catalogue, so as to look for publications held by a library; they are also widely used to download journal articles already known and frequently searched using Google Scholar (Moore 2016). However, the accuracy characteristics may not be the same as those obtainable by querying the search interfaces of the individual databases connected to the discovery. The discovery system offers less precision even than online catalogues and searching for a known document, that is to say a publication whose bibliographic identity is known can be problematic. In the same way, researchers may prefer to query native database indexes directly because they need precise answers, which do not require exploratory research (Ellero 2013).

A sample of monographs of the database used for our survey was also selected for testing discovery systems. Accordingly, we carried out a comparison between the outcomes of using the discovery system and the results of queries from libraries’ online catalogues.

The issues taken into consideration show a percentage of misalignment between the data from the catalogue and the discovery system. It is not possible to define the causes of this mismatch between data, which could depend on the software adopted and the settings defined for the attribution of relevance criteria to publications, for example.

The results showed that searches through online catalogues are to be preferred for quantitative and qualitative assessments, in particular for the comprehensibility of the document retrieval techniques, entirely managed in the field of library and information science.

Moreover, our survey highlighted the currently inaccurate management of identifiers
of entities registered in bibliographic records. Authority control is a crucial point for purposes of scientific evaluation. It enables precise identification of each entity of interest (author and publication) ensuring the correct attribution of works to their respective authors. The traditional authority files provide standardized access points for the authors of monographs, but often overlook the authors of articles in scientific journals, partly for economic choices, partly due to factors such as the preference given to new discovery tools for research (Martin 2015). This makes the task of philologically reconstructing the scientific production of an author very complex, because of the difficulty in obtaining results from a catalogue that associates monographs with contributions in periodical publications or with individual contributions in monographs. This condition could change with the implementation of the new bibliographic standards and protocols designed specifically for the Semantic Web: RDA and BIBFRAME. Resource Description and Access (RDA 2013) application requires a systematic process of ‘entity identification’, since every bibliographic data is univocally identified by the URIs and linked to other documents in the Web of data. Identifying entities with dereferenceable elements is thus the premise for creating a network of relationships between entities. The BIBFRAME model\(^2\) aims to replace MARC with a new extensible format, flexible and integrated with the Web and part of the Web of data. This huge change has important implications for controlled vocabularies (Southwick et al. 2015). Disambiguation and identification of authors are relevant issues in the Semantic Web where a different model is adopted from our bibliographic entity-relation model (E-R), such as the Resource Description Framework (RDF) model of triples and graphs. RDF application for structuring and modelling bibliographic information radically changes the identification process, which will be able to represent entities that appear in a variety of names in various resources (Peponakis 2016).

The collaborative nature of the Semantic Web introduces a new distributed vision of authority control, which no longer depends on national agencies, but involves subjects of various different types.

Conclusions

In our survey we analysed the presence of monographs in a range of unit catalogues and OPACs belonging to national and university libraries in Italy and abroad. Thanks to a deeper analysis of different library types, it has been possible to offer a broader picture and to better verify the possibility of using that indicator. We demonstrated also the effect of monographs presented as gifts in library holdings, verifying the number of monographs received as donations in particular by the Italian academic libraries of our sample. Moreover, our analysis was extended to include an accurate examination of the

\(^2\) https://www.loc.gov/bibframe/.
issues that arise using discovery systems in place of online catalogues, and the problems connected to the unambiguous identification of entities, which could be resolved with the implementation of Semantic Web standards. The latter analysis constitutes an important part of the testing procedure with regard to library catalog analysis, as we updated the study carried out by Torres Salinas – Moed, White and Linmans, and broadened our examination to include the modern tools (discovery and Semantic Web technologies) that in 2008-2009 were not yet available.

A comparison between the two scientific fields, history of books, bibliography, library science on one hand, and history of political institutions on the other, highlights that the latter shows a less widespread distribution compared to the former. This difference also persists in Italian libraries.

The results of our survey highlighted that the criterion of the presence of monographs in libraries cannot be adopted as a reliable indicator to establish the scientific value of monographs and, therefore, of their authors. However, it may be reasonable to claim that a check on the distribution of monographs in prestigious libraries, selected on the basis of well-balanced criteria, could constitute a means of reinforcing a positive or negative evaluation formulated following a well-managed peer review.

References


Linmans, Adrianus J. M. (2010). Why with bibliometrics the Humanities does not need to be the weakest link: indicators for research evaluation based on citations, library holdings, and productivity measures. Scientometrics, 83(2): 337-354.


Moed, Henk F. (2005). Citation analysis in research evaluation. Dordrecht: Springer.


