

## ***Bibliometrics: a Publication Analysis Tool***

Padrós-Cuxart, Rosa<sup>1</sup>; Riera-Quintero, Clara<sup>1</sup>, March-Mir, Francesc<sup>1</sup>

<sup>1</sup> Universitat Oberta de Catalunya rpadrosc@uoc.edu, crieraq@uoc.edu, fmarchm@uoc.edu

**Abstract.** Bibliometrics has become an increasingly necessary tool for studying and analysing an institution's scientific activity. Research librarians must be able to extract, interpret and represent bibliometric data in different ways depending on the user to whom they are to be presented. In this article, we describe the process of creating a bibliometric data management and consultation tool, called *Bibliometrics*, whose goal is to cater for the needs of the Universitat Oberta de Catalunya's different users. We describe the reasons that have led us to create this tool, the information sources from which the data are taken (both bibliographic and bibliometric), and the bibliometric indicators used to present the information.

**Keywords:** Bibliometrics, Evaluation, Quality, Research Libraries, Research Performance Measurement, Research Assessment, Information Retrieval

### **1 Introduction**

Researchers need bibliometric data when responding to calls for applications for accreditation, competitive projects, etc. Likewise, the University uses the data for strategic purposes when reporting to the public administration, managing project applications, accrediting doctoral programmes, analysing and assessing scientific output, setting research policies and disseminating its scientific activities. To provide an adequate response to all the information needs, as research librarians at the Universitat Oberta de Catalunya's (UOC) Virtual Library, we find ourselves in the difficult situation of having to consult a large variety of information sources and using many tools to manage them (the solution is often to download information from the databases and customize it to specific requests) [6]. We must be able to extract, interpret and represent the data in different ways [4]. This is a laborious, costly task that can absorb up to 80% of the workload of a bibliometric report. This is why the UOC's Virtual Library has developed an in-house database, called *Bibliometrics*, which enables bibliometric data to be managed and consulted and the impact of the publications and their authors to be measured. It stores the UOC's academic articles and lets users assess their quality based on national and international bibliometric indicators.

Requests for bibliometric data come from faculty and researchers alike, the vice president, deans and other administrative offices at the University, such as the Quality Office or the UOC Knowledge Transfer and Research Support Office (OSRT).

Although the Universitat Oberta de Catalunya is only 20 years old, it has 2 research institutes, 44 research groups and 262 teaching and research staff members. The most important research activity is related to social sciences and focuses mainly

on e-learning and the information and knowledge society. It is precisely here that we come up against a number of added difficulties, as the assessment of social sciences varies and is less standardized than other disciplines [8,3]: for example, many of the journals published in this area are not indexed in Web of Science [7] or Scopus, forcing us to use other national and international indicators, which makes them more difficult to assess. In Spain, a lot of weight is given to publication-associated indicators when assessing research and, to do this, a large number of national indicators have been devised to assess the social sciences and the humanities.

The UOC created its Library Services for Research team in 2011 [9] and, since then, our services portfolio has grown considerably to respond to researchers' emerging needs [3]. The bibliometric consultations service is one of the services that has grown most in recent years: from 37 consultations received in 2010-2011 to 145 in 2013-2014. In order to be able to offer this service, research librarians are required to monitor the existing sources of information on research, generate analysis, outlook and surveillance reports, and we must also offer training and counselling [10]. The bibliometric services we offer are the following:

- Evaluation of scientific publications: we produce assessment reports on scientific outputs at individual, research group, faculty, knowledge area or institutional level.
- Application call procedures for the teaching staff: we offer to the researchers a support service for finding bibliometric data (impact factors, citations, etc.) included in the publications section on standardized curricula vitae to be submitted with applications for accreditation by the Quality Assurance Agencies (AQU, ANECA, etc.). We also offer guidance on the assessment criteria established for scientific publications by quality assurance agencies' research committees.
- Calls for project proposals (publications of the IP): we provide bibliometric data for the project proposals, such as H2020, ERC, etc., in collaboration with the Project Management Department.
- Dissemination of scientific publications (strategic): we offer to the researchers guidance and/or reports on where to publish their articles. This strategy includes different issues concerning bibliometric data (the Q1 journals), number of issues per journal, the differences in publication and dissemination practices for each field of study, etc.
- External reports: several institutions at the central, state and local levels produce annual reports on the scientific productivity of public higher education institutions (number of articles and journal's impact factor). We send them the outcomes data of our university.

## 2 Methodology

When considering the creation of the tool we describe in this article, we found few examples and little literature that dealt specifically with bibliometric information consultation, management and retrieval tools. However, Nicolai Mallig's article, which describes in detail the process of creating a relational database that enables

bibliometric analyses to be performed, has been found to be particularly helpful [6]. Part of the inspiration for the tool's design has also been provided by the FUTUR portal [1] created by the Universitat Politècnica de Catalunya (UPC), although in our case we decided to take it a little further and visualize not only the productivity but also the impact of the University's and its researchers' scientific output.

## 2.1 Data model

In this section, we briefly define how the information contained in this database is structured and organized. *Bibliometrics'* main entities are:

- Author
- Article
- Journal
- Organization
- Subject area
- Indicator

The links between these entities are articulated through the following relationships:

- Authorship (it links an author with his or her article)
- Affiliation (it links an author with the organization)
- Publication (it links an article with a journal)
- Indicator value (it links an article, author, journal or subject area with an indicator)

## 2.2 Information about the authors

*Bibliometrics* includes the scientific output of the lecturers and researchers affiliated with the UOC and, for each one, provides a series of data, including start and end dates of the affiliation with the University, research centre, faculty or research group. The data are taken from the University's Current Research Information System (CRIS). The information stored for each author is as follows:

**Table 1.** Detail of the fields corresponding to the entity *Author*

Field	Definition
Citation name	Form of the name customarily used by the author
Given name	The author's first name
Surname	The author's surname
ORCID	ORCID identifier
Scopus author ID	The author's identifier in Scopus
Research ID	The author's identifier in WoS
Public profile	Indicates whether or not the author's detailed file will be viewable on the application
Affiliation	Indicates or not whether the author belongs (or has belonged) to the UOC
> From	Indicates the date on which affiliation with the UOC began
> To	Indicates the date on which affiliation with the UOC ended

### 2.3 Bibliographic data

The bibliographic data used by *Bibliometrics* come from our CRIS. The following bibliographic information is collected for each publication:

**Table 2.** Detail of the fields corresponding to the entity *Publication*

Field	Definition
Original title	The article's original title
English title	The title in English if the original title is in another language
AI Code	The article's unique code in the CRIS
WoS ID	The article's identifier in WoS
Scopus ID	The article's identifier in Scopus
DOI	The article's DOI identifier
Publication type	Type of publication
Publication status	The article's publication status
Language	The language in which the article is written
Journal	The journal in which the article is published
Volume	The journal's volume number
Issue	The journal's issue number
First page	The number of the article's first page
Last page	The number of the article's last page
Publication year	The year in which the article was published
Publication date	The date on which the article was published in date format
Collaboration	The article's field of institutional collaboration
Conference	Title of the conference
Open access	Indication of whether the article is published in open access
Authorship	The publication's authors, by order of appearance, the institution they belong to and, if they are affiliated with the UOC, details of the centre, faculty or research group to which they belong
Repository URL	Link to the full text of the article in the UOC's repository
Bibliographical reference (APA)	The article's bibliographical reference in APA format
Bibliographical reference (ISO)	The bibliographical reference in ISO-690 format

In view of the distinctive features of research in social sciences, during the process of creating this tool we felt that it was necessary to include certain document typologies beyond original articles and proceedings papers as, in this discipline, other typologies are also taken into account when assessing researchers' curricula [10,5]. Accordingly, the research outputs included in *Bibliometrics* are: original, review, proceedings paper, editorial material, book review, letter.

## 2.4 Information about bibliometric indicators

Each article is assigned the applicable national and international bibliometric indicators, depending on the journal they are published in, the academic discipline or the citations received. It should be remembered that the indexes applied in any given area of knowledge and speciality field vary as the criteria are different for each discipline: we cannot show data only from Journal Citation Reports knowing that the WoS database's coverage is modest in disciplines such as sociology, political science, anthropology, education science or humanities [7]. Thus, we have also added national indicators in order to cater for all the disciplines.

Bibliometric indicators included in *Bibliometrics*:

- International indicators: Journal Citation Reports, Scimago Journal Rank, Erih Plus and Latindex.
- National indicators: MIAR, Carhus+ and DICE.
- Citations: ISI WoS, Scopus, Google Scholar.

## 2.5 Data maintenance

Adequate data maintenance requires cross-functional collaboration among all the University departments responsible for data maintenance and quality in order to guarantee their reliability: Personnel and Social Responsibility, Planning and Quality, and the UOC Knowledge Transfer and Research Support Office (OSRT).

- Author maintenance: affiliations, name variants, unique identifiers.
- Indicator maintenance: indicators and their databases.
- Publication maintenance: publication types, status.
- Organization maintenance: UOC faculties, research centres, research groups, institutions.
- Subject area maintenance: academic disciplines, fields of study, research areas.
- Term list maintenance: countries, languages, nature, publication types, publication status, collaboration, occupational categories, research group roles, database types.
- Database maintenance: coverage, academic disciplines, UOC faculties.

## 2.6 Consultation environment

*Bibliometrics* has defined 3 user roles: public, author (this inherits the public role's permits but can also edit its profile) and administrator (this can edit and create data). The consultation environment is common to all of the tool's users, irrespective of their role.

The tool enables searches to be made by publication (original title, title in English, AI code, DOI, WoS ID, Scopus ID), author, (citation name, IDP Code) and journal (journal title, ISSN). Filters can also be applied to these three entities, namely:

**Table 3.** Filters for the entities *Publications*, *Authors* and *Journals*

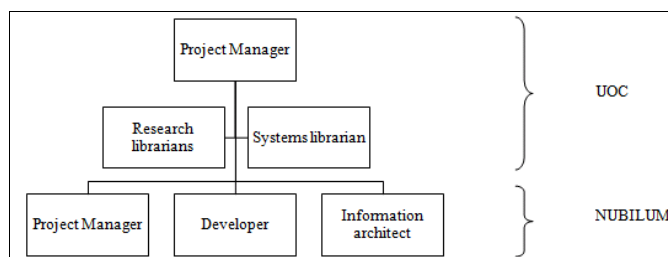
Entity	Filter
Publications	UOC faculty
	Research centre
	Research group
	Academic discipline
	Field of study
	Author
	Publication type
	Author profile
	Peer review
	Open access
	Indexed
	Publication year
	Date range
	Authors
UOC faculty	
Research centre	
Research group	
Academic discipline	
Field of study	
Active	
Journals	Open access
	Peer review
	Indexed
	Language

Furthermore, the data displayed on the results pages can be downloaded with the following formats:

- Publications: CSV, RIS, ISO-690, APA
- Authors: CSV
- Journals: CSV

## 2.7 Project organization chart

To create this tool, research librarians, a systems librarian and, from an outsourced company (Nubilum), developers and an information architect have worked together.

**Fig. 1.** Project organization chart

## 2.8 Technology

The application is compatible with the following operating environments:

- Ubuntu LTS 12.04 y Red Hat Enterprise 6.3
- Java EE JDK 1.6
- JBoss AS 7.2
- MySQL 5.5

Frameworks and libraries that have been used to develop the tool:

- Hibernate 4.2 / 4.3
- Spring Framework 4.0
- Maven 3.2
- JQuery 2.x

The graphics library used is *Highcharts*, as it allows numeric values to be displayed inside the charts, which is a feature the other tools do not offer.

## 3 Results

### 3.1 What does *Bibliometrics* offer?

What does *Bibliometrics* offer to its users?

- Bibliographic and bibliometric information and high-quality, normalized data in terms of researchers, research groups, research centres, faculties or the University as a whole.
- Graphic representation of the data.
- Comparative bibliometric analysis of the UOC's researchers, research groups, research centres and faculties.
- Data extraction in different formats for exporting.
- Faceted search.

### 3.2 Indicators

*Bibliometrics* provides four levels of indicators to represent the data:

#### **Productivity**

The number of publications can be consulted by year of publication, author and type of publication.

As regards consultations of publications by year of publication, the user can compare the publications between faculties, research centres, research groups, academic disciplines or field of study. The calculation is performed as follows: for the research groups, research centres or faculties, it counts the number of publications whose authorship is assigned to the UOC research group, research centre or faculty. For the knowledge areas, it counts the number of publications that have an author affiliated with the knowledge area.

One of the calculations offered by *Bibliometrics* in this section is the *Author's productivity*, and the information that the user can access is the following:

**Table 4.** Calculations of the indicator *Author's productivity*

Indicator	Calculation
PDI (research and teaching staff)	This counts the total number of UOC members who are active within each organization unit or knowledge area.
% PDI	The percentage of the PDI with respect to the total number of active UOC members.
Publications	The total number of publications whose authorship is assigned to the organization unit or knowledge area.
% Publications	The percentage of publications by the organization unit or knowledge area with respect to the total number of publications.
Productivity by PDI	The number of publications by the organization unit or knowledge area divided by the PDI of the organization unit or knowledge area.

**Visibility**

This enables the articles' quality to be analysed in terms of the journal they are published in, using international (WoS, Scopus) and/or national (MIAR, Carhus+, etc.) bibliometric indicators. It is true that sometimes an article's visibility has been measured on the basis of whether or not it has been cited (the most cited articles are considered more visible than those that have not been cited) [2], but we have applied this interpretation to the impact indicator.

As regards the possibility of comparing data, the tool offers the user the possibility of comparing only between Journal Citation Reports and Scimago Journal Rank, as they are the only two indicators that apply to all the knowledge areas in which research is performed at the UOC.

**Impact**

This enables the citations received by publication in WoS, Scopus and Google Scholar to be analysed. The user can consult the number of citations received by year of publication and, in addition, by way of summary, he or she can see how many publications are indexed, what is the % of publications cited and what is the total number of citations received (the results will vary depending on the search carried out by the user, depending on whether or not the data have been filtered by an organization). Lastly, the user can also compare the citations received between research centres, faculties or research groups.

**Table 5.** Calculations of the indicator *Comparative citations*

Indicator	Calculation
Publications	For the UOC research groups, research centres or faculties, it counts the number of publications whose authorship is assigned to the UOC research group, research centre or faculty. For the knowledge areas, it counts the number of publications that have an author affiliated with the knowledge area.
Total citations	It adds the total number of citations received in each of the databases.
Total cited	It counts the number of publications that have at least one citation in each of the databases.



Not cited	It counts the number of publications that are not cited in any of the databases.
% Cited	The percentage of cited publications with respect to the total number of publications by the UOC research group, research centre or faculty.

### Collaboration

This enables the level of co-authoring of publications to be analysed in terms of the authors' affiliation (institution and country): international, national, inter-university or without collaboration. The user can compare the collaboration among the publications' authors between faculties, research centres, research groups, academic disciplines or field of study. The calculation is performed as follows:

- The percentage of publications from each organization unit or knowledge area by level of collaboration (UOC, national, international), with respect to the total number of publications.
- For UOC research groups, research centres or faculties, it counts the number of papers whose authorship is assigned to the organization.
- For knowledge areas, it counts the number of publications that have an author affiliated with the knowledge area.

For collaboration between institutions, the user can access data on the mean number of citations received at Scopus for publications co-authored with one member of the institution (number of citations/number of publications). The same thing happens with collaboration between countries, as the user can access data about the mean number of citations received at Scopus for papers co-authored with an institution in the country (number of citations/number of publications).

## 4 Conclusion

This bibliometric data management and consultation tool created for the University is the outcome of the realization that, if the bibliometrics units at the research libraries are to offer a better service, universities need to invest in tools that support research assessment needs. Accordingly, in order to assure successful creation of such tools, the different areas of institutions (all those involved in data management, maintenance and quality) must work together across functional divisions.

*Bibliometrics* supports administrative staff and research librarians in generating bibliometric reports. However, on an individual level, it also provides information on the impact of an author's publications or, at vice president, centre, faculty or research group level, assistance in decision making. It is therefore a strategic tool for the University.

The research activity carried out at the University is focused on social sciences. On the basis of the knowledge we have acquired concerning this discipline's assessment casuistry and its differences from other disciplines as regards its publication culture [11], we have seen a need to expand the tool in the near future to include other document typologies that could be beneficial for authors performing research in this field: books and book chapters. Once these document typologies have been included in the

tool, we will be able to offer data on productivity, impact and collaboration, but not on visibility (these data are linked to the journal in which the articles are published and, in the case of books and book chapters, it would be necessary to define how data on the publisher's standing will be included, for example). It is important to be aware of the assessment differences between disciplines when interpreting the data extracted from *Bibliometrics*. Without this awareness, it is easy to come to mistaken conclusions and make incorrect interpretations.

Lastly, we would like to highlight that the response received from the users of the UOC Virtual Library's bibliometric consultations service has always been satisfactory, not just in the cases of specific requests for support in faculty assessment processes but also in terms of the real perception that researchers have gained on the impact of their own publications or those of a specific department.

## 5 References

1. Cacho, Y., Clavero, J., Prieto, A.J., Martínez, D., Rovira, A., Serrano, J.: FUTUR: el nou portal de la producció científica de la Universitat Politècnica de Catalunya. Ítem. 57, (145-156 (2013)
2. Cronin, B., & Sugimoto, C. R.: Beyond bibliometrics: harnessing multidimensional indicators of scholarly impact. The MIT Press, Massachusetts (2014)
3. Drummond, R., & Wartho, R.: RIMS: The Research Impact Measurement Service at the University of New South Wales. Australian Academic and Research Libraries. 40(2), 76-87 (2009)
4. Gumpfenberger, C., Wieland, M., & Gorraiz, J.: Bibliometric practices and activities at the University of Vienna. Library Management. 33(3), 174-183 (2012)
5. Huang, M., & Chang, Y.: Characteristics of research output in social sciences and humanities: From a research evaluation perspective. Journal of the American Society for Information Science and Technology. 59(11), 1819-1828 (2008)
6. Mallig, N.: A relational database for bibliometric analysis. Journal of Informetrics. 4(4), 564-580 (2010)
7. Moed, H. F.: Implications for the Use of the ISI Citation Indexes in Research Evaluation. In: Citation analysis in research evaluation, pp. 137-143. Springer, Dordrecht (2005)
8. On the proper use of Bibliometrics to evaluate individual researchers. Académie des Sciences. Paris: Académie des Sciences, Institut de France.
9. Riera Quintero, C., Padrós Cuxart, R., & Zuñiga Ruiz, A.: Research librarians in a virtual environment : new skills, new services (2012), <http://openaccess.uoc.edu/webapps/o2/handle/10609/14221>
10. Torres-Salinas, D., & Jiménez-Contreras, E.: Hacia las unidades de bibliometría en las universidades: modelo y funciones. Revista Española de Documentación Científica. 35(3), 469-480 (2012)
11. van Leeuwen, T.: The application of bibliometric analyses in the evaluation of social science research. Who benefits from it, and why it is still feasible. Scientometrics. 66(1), 133-154 (2006)
12. van Leeuwen, T., Costas, R., Calero-Medina, C., & Visser, M.: The role of editorial material in bibliometric research performance assessments. Scientometrics. 95(2), 817-828 (2012)