MTMT: The Hungarian Scientific Bibliography

András Holl, Gábor Makara

Library and Information Centre of the Hungarian Academy of Sciences
Arany J. str. 1., Budapest, Hungary
holl.andras@konyvtar.mta.hu, makara.gabor@mtmt.hu

András Micsik, László Kovács

Institute for Computer Science and Control, Hungarian Academy of Sciences
Lagymanyosi u. 11. Budapest Hungary
{andras.micsik,laszlo.kovacs}@sztaki.mta.hu

Abstract:

The Hungarian Scientific Bibliography (Magyar Tudományos Művek Tára, MTMT) is a comprehensive national bibliographic database of scientific publications and citations. Since 1999 MTMT collects data from researchers and institutions, containing almost 5 million records presently. The database is used by all major centres of research: universities and the research institute network of the Academy. Planned legislation states that registering all publications from research using taxpayer money will be required in the future. Data in MTMT can be used for supporting evidence-based management in science, fund allocating for institutions, projects and individuals. This database promotes transparency with displaying scientific output, in the form of statistics, but also enabling access to the articles themselves - the latter residing either at the publisher or at institutional repositories. MTMT supplies data to the electronic proposal management system of the Hungarian Scientific Research Fund and other information systems run by universities or the government.

Keywords: open science, open data, bibliography, scientometrics

1   Introduction

The Hungarian Scientific Bibliography (Magyar Tudományos Művek Tára, MTMT)[1] is a comprehensive national bibliographic database of scientific publications and citations. Since 1999 MTMT collects data from researchers and institutions, containing almost 5 million records presently. The bibliography is run by the Library and Information Centre of the Hungarian Academy of Sciences. Its goal is to create a comprehensive national scientific bibliography of all researchers working in Hungary, affiliated to Hungarian institutions, funded by Hungarian grants or participating in the scientific life of the country in other ways. It contains metadata about the publications and their citing papers. Currently, the MTMT contains cca. 1.2 million publications from 36,000 researcher and 3.7 million citations.

2   About the MTMT

The creation of MTMT was launched in 2009 by the founders, the Hungarian Accreditation Committee, the Hungarian Rectors Conference, the Hungarian Scientific Research Fund (OTKA) and the Hungarian Academy of Sciences (MTA). It was based on a similar system of research institutes and scientific associates of the MTA, which was expanded and improved in order to be able to cover the scientific output of all Hungarian scientists and their host institutions. At the moment, the host institutions of about two thirds of the Hungarian scientists are affiliated with MTMT and the list is growing. At about the same time similar but slightly different systems have been built in the Netherlands (NARCIS)[2] and some other European institutional scientific bibliographies have been launched.

MTMT's underlying principles are the following:

A. It is open to all Hungarian scientists and their host institutions.

B. Participating institutions sign a contract with MTMT to provide funds to partially cover the costs of the centralized part of the system and accept responsibilities for uploading their recent scientific bibliography into the centralized database.

C. Responsibility for the validity of the personal and institutional part of the bibliography is shared by the institution and the individual scientist.

D. A quality control system is being developed to ensure validity using centralized as well as distributed components.

E. MTMT aims to collect data on all scientific work of the scientists, but accepts data on related but non-scientific writings of the scientists involved, so that both a scientific and a full bibliography of the scientists are available from the system.

The organization involves institutional members who delegate administrative duties to a small selection of library-informatics experts as well as a central support system providing software development, bibliographical help and expertise, courses and instructional material to all the member institutions.

For the research institutes of the Academy the MTMT already covers the years 1992-present and for the higher education system the recent data are being collected for 2007-present. Coverage is rapidly expanding. Member institutions can compile various metrics of their scientific output so that the MTMT can also be used as an institutional scientific management information system as well as a continuously updated publication and citation list of the individuals involved. Being public and controlled the publication lists in the MTMT are accepted by practically all Hungarian scientific bodies.

The MTMT works in cooperation with the Repository of the Academy's Library (REAL) [3] and other institutional repositories where the full text of publications can be deposited. The MTMT user interface provides an easy way to upload the full text into the selected repository while recording bibliographic data. The uploaded file is forwarded to the selected repository via the SWORD protocol, and the link to the deposited file is stored in MTMT. Using this metadata MTMT not only provides links to the full text of publications, but also collects statistics on the status of open access in Hungary.

2.1 Legal background

The Academy's presidential commitment 35/2009. (VI. 23.) established MTMT which, since then, has been involved in the 2011./CCIV. Act on National Higher Education and in the 1167/2014. (III. 25.) government resolution. The latter defines support from the government to establish a national bibliography database funded by the state and run by the Hungarian Academy of Sciences. It also emphasizes that scientific results reached with public funding must be registered in the bibliography, and the voluntary registration of other results should also be accepted.

The open access to scientific publications has some other regulations in Hungary; since 2007 PhD theses should be publicly available, since 2008 papers funded by the Hungarian Scientific Research Fund (OTKA) should be publicly available. The president of the Hungarian Academy of Sciences issued a mandate on making the scientific results of MTA open access starting from 2013. This mandate also regulates that links to the open access versions of publications have to be collected in the MTMT.

2.2 Cooperation with other services

The grant application at the Hungarian Scientific Research Fund (OTKA) is aided by a web-based tool. Applicants need to supply a list of their recent papers, which will be provided by MTMT automatically. The articles regarded most relevant should be marked by the applicant. The reports required during an awarded grant are collected similarly from MTMT using the grant identifiers stored in the bibliographic records. Grant reports and associated papers should be stored in the REAL repository according to the open access mandate of OTKA.

The Hungarian Doctoral Council (HDC) is an organization created by the universities offering PhD programs. HDC publishes the PhD topics and positions of the doctoral schools with the condition that supervisors' scientific output should be imported from MTMT (with a relatively easy, guided process).

The Hungarian Academy of Sciences (MTA) awards a D.Sc. degree. Applicants need to submit a thesis, and certain scientometric parameters are required. Certified bibliometric data are extracted from the MTMT. The submitted theses, together with other documents of the process are made public in the REAL repository.

These examples show that MTMT can serve several needs of various organizations and the public:
- the need for validated bibliographic data for individuals, organizational units and whole organizations,
- the need for statistical and scientometric data regarding individuals, sponsors, organizations, etc.,
- the need for information and overview about the state of research domains in Hungary,
- and the need to collect the open access availability of Hungarian scientific output.

Our further plan is to use ORCIDs [4] for all of our authors in order to help data harvest and author disambiguation. ORCID membership and connection to the ORCID API could offer amenities for MTMT users to export/import their publication lists to/from their ORCID profile.

The database of MTMT can be used to analyse the scientific output of the country, and get quick access to scientific results.

2.3 Data collection

The MTMT aims at collecting data from the broadest range of sources, but also to ensure the good quality of imported metadata. Therefore, MTMT works on establishing direct connections with global and national publishers and scientific services. Negotiations are initiated with Web of Science and Scopus in order to collect publication and citation metadata from these services. Unfortunately, while the majority of scientific outcome in life sciences is recorded by 2-4 global services, in social sciences, arts and humanities the needed metadata is scattered over many small sources.

MTMT will also try to connect directly with Hungarian publishers and harvest metadata from them.

With these activities MTMT would like to move from manual metadata upload towards automated harvesting of validated metadata. MTMT also plans to apply automated methods for the extraction of metadata and for the correction of manual input. A further plan is to implement metadata import from the database of CrossRef, similarly to the DOI import function implemented in EPrints.

Currently, the authors or local administrators can add their bibliographic data by filling web forms. Metadata inserted goes through two levels of verification: first inside the organization and second by the MTMT support centre.

2.4 Data consumption

Anyone can use the public MTMT portal to search and browse bibliographic data and statistics. Information about items found can be exported as Word, RTF, CSV, RIS and other formats. Harvesting of data is only possible for member organizations and contracting partners. They are also able to import and export metadata using the specific MTMT XML format. The use of data for commercial purposes is currently not permitted.

The main beneficiaries of MTMT are scientists, as the database provides a discovery service, and as it enables prudent distribution of research grants. The taxpayer profits from such a service as well, as research grants will be spent better. Scientists and citizens benefit from greater transparency as well. Furthermore, there is another group of beneficiaries - located between the scientists and the layman citizen - these are the teachers, students, hobby scientists. The portal functions of the MTMT can lead them to scientific results - publications residing in repositories - they can understand and use in their studies.

One should keep in mind that the direct consumers of such information services need not be humans any longer. Linked Open Data capabilities can open up databases like MTMT for machine processing - thus further increasing the visibility of national scientific output.

Analysis of the bibliographic mesh (publications interconnected by citations) could provide insights of the structure of a discipline and the position and connection network of a research unit. In sciences, commercial tools and services are available for this purpose, but alas, no such possibilities exist in social sciences, arts and humanities. MTMT will have the potential to facilitate such analysis in those areas publishing in local journals, mainly in Hungarian.

3 Conclusion and outlook

MTMT is on the way to become the central registry of Hungarian scientific output. It already stores the majority of the results achieved in the last years, and its coverage is growing further rapidly. MTMT uses a hybrid model for data ingest and curation: automatic harvest, expert control and the possibility of researcher intervention all play a role.

The IT infrastructure is constantly improving, and the whole software framework will soon be renewed with the help of European and national funding. The new software will offer more possibilities for data access via an open API, and will exploit more connections to external services via open protocols. The use of globally unique identifiers for publications (such as DOIs) is also encouraged by MTMT.

MTMT is an authentic source of scientific achievements in Hungary, and in a few years it may become a public scientific portal as well as a major source of open data and an important proxy for access open documents.

Data stored in MTMT could facilitate deep analysis of information in the fields of science via text mining and other methods for exploring the content and metadata of publications. Investigation e.g. the latent, social structures of sphere for better understanding the dynamics of science could result better decisions of the future.

4 References

[1] MTMT, the Hungarian Scientific Bibliography, http://www.mtmt.hu

[2] National Academic Research and Collaborations Information System (NARCIS), http://www.narcis.nl/?Language=en

[3] REAL, the Repository of the Academy's Library, http://real.mtak.hu

[4] Open Researcher and Contributor ID (ORCID), http://orcid.org/