Dissertation as a source for prosopography of scientific communities

Birutė Railienė & Giedrė Miknienė·

Abstract

This article deals with the bibliometric approach to the metadata of a dissertation record in a catalogue or database. As a document, a dissertation provides data on new knowledge, but also encodes important scientometric information. Research advising is one of main scholarly communication formal links (besides co-authorship and citation) in a social structure of science. PhD records may expose a scientific relation of research-advising both at the personal and institutional levels within a certain scientific area. This survey was inspired by bibliographical indexes of dissertation output in certain fields and during certain periods: 1793 to 1842, 1940 to 1993 in one country – Lithuania. The survey of selected scientific communities from 1945 to 1990 in Lithuania reveals a vast retrospective of scholarly communication, however, it is fragmented and not satisfactory for a comprehensive prosopographic mapping of scholarly communities within a science field or a country. After reviewing a tradition and style of electronic theses and dissertations (ETD) cataloguing and metadata creation, obstacles for prosopographic surveys and data mining in catalogues and databases of ETDs are revealed.

Keywords

Dissertation; Bibliography; Prosopography; Bibliometrics; Library catalogue; Thesis; Scientific school; Invisible college

Teses e dissertações como fontes para a prosopografia de comunidades científicas

Resumo

O presente artigo trata da abordagem bibliográfica dos metadados de registros de teses acadêmicas em catálogos ou bases de dados. Enquanto documentos, as teses acadêmicas trazem dados sobre novos conhecimentos, mas também codificam importante informação cientométrica. A orientação é um dos canais formais de comunicação acadêmica mais importantes (junto de coautorias e citações) na estrutura social da ciência. Os registros de teses podem evidenciar relações científicas de orientação da pesquisa em ambos os níveis pessoal e institucional dentro de uma certa área científica. O levantamento aqui realizado foi baseado em índices bibliográficos de teses acadêmicas em determinadas áreas e períodos: 1783 a 1842, 1940 a 1993, num país, Lituânia. O mapeamento de comunidades científicas selecionadas na Lituânia entre 1945 e 1990 revela um vasto panorama retrospectivo de comunicação acadêmica, porém fragmentária e insuficiente para um relevamento prosopográfico abrangente das comunidades científicas pertences a uma determinado área científica ou país. A avaliação da tradição e estilo de catalogação e geração de metadados teses e dissertações eletrônicas (TDE) revela obstáculos para mapeamentos prosopográficos e mineração de textos e catálogos e bases de dados de TDE.

Palavras-chave

Teses acadêmicas; Bibliografia; Prosopografia; Bibliometria; Catálogos de bibliotecas; Escolas científicas; Colégios invisíveis

·Wroblewski Library of the Lithuanian Academy of Sciences, Vilnius, Lithuania. ✉ railiene@mab.lt

Introduction

Prosopography of science is a long-term science history project in Lithuania that started in early 80s. Long before scientific communication moved to network, Nicholas Mullins indicated several links in science describing theory groups in sociology: discussion, co-authorship, curation and collegiality.[1] This description was adopted in the survey of social structures in science in Lithuania. A model of social communication in an invisible college was developed by Ona Voverienė and Nijolė Šaduikienė. It was based on bibliometric data of curation (dissertation research-advising) and collegial (co-authorship) links in sciences: [2] mathematics, biology, medicine, chemistry, physics, etc.

Survey of scientific schools was inspired by a rich experience of bibliographical tradition in Lithuania – Wroblewski Library of the Lithuanian Academy of Sciences, Vilnius University Library. A printed retrospective of PhD bibliography in medical sciences reaches 1793,[3] selective subject bibliographies cover the period from 1946 to 1990.[4] The fragmented picture of the bibliography is not satisfactory for a comprehensive report of the national output regarding dissertations. Nevertheless the survey of scientific communities in Lithuania from 1945 to 1990 reveals a vast picture on the history of scholarly communication in certain sciences: physics, biology, biochemistry, chemistry, technology, mathematics and medicine.[5]

Historical background

Research-advising is one of main scholarly communication formal links (besides co-authorship and citation) in a social structure of science. Dissertation records may expose a scientific relation of research-advising both at the personal and institutional levels within a certain scientific area.

---

[1]Nicholas C. Mullins, *Theories and Theory Groups in Contemporary American Sociology* (New York: Harper & Row, 1973), 193–4.

[2] Nijolė Šaduikienė, & Ona Voverienė, "Akademiko A. Jucio mokslinė mokykla," *Mokslas ir technika*, (1986): 30.

[3] Alfonsas Bielinis, ed., *Medicinos daktarų disertacijos, apgintos arba pripažintos Vilniaus universitete 1793–1842 metais : (bibliografinė rodyklė)* (Vilnius: LMA CB, 1958), 132.

[4] Vladas Abramavičius, Konstancija Čepienė, comp., *Lietuvos TSR Mokslų akademijos rankraštinių darbų ir disertacijų bibliografija, 1946–1956* (Vilnius : LMA CB, 1958), 158; Zita Petrauskienė, & Petronėlė Valentėlienė, comp., *Lietuvos TSR mokslininkų disertacijos, 1945–1968 : bibliografija* (Vilnius: LMA CB, 1971), 490; Teresė Dragūnienė, compiler, *Lietuvos medicinos mokslų daktarų ir kandidatų disertacijų sąrašas, 1896–1968* (Vilnius: Valst. moksl. medicinos b-ka, 1970), 90.; Vida Stankuvienė, & Ala Zozulia, compilers, *Lietuvos TSR medicinos mokslų daktarų ir kandidatų disertacijų sąrašas, 1969–1977* (Vilnius: Valst. moksl. medicinos b-ka, 1981), 78; Birutė Railienė, comp., *Geologijos-mineralogijos disertacijos Lietuvoje, 1953–1993 : bibliografijos rodyklė* (Vilnius: Vilniaus univ. l-kla, 1994), 29; Olga Vilkina, compiler, *Lietuvos bibliotekininkų, knygotyrininkų ir informatikų disertacijos, 1960–1990: (bibliografija)* (Vilnius: LNB, 1992), 15.

[5] Ona Voverienė, & Birutė Railienė, Giedrė Sasnauskaitė, Vaida Vanagaitė, *Lietuvos mokslinės mokyklos (1945–1990)* (Vilnius: Mokslo aidai, 2002), 310. [*Scientific schools (invisible colleges) in Lithuania (1945–1990).*

Before making final conclusions on research advisors names one should consider how to deal with homonymous and synonymous names. The analysis of a personal profile and authority files is one of the solutions, though involving additional efforts; [6] also the data for national authority files are of help. Not of lesser relevance is the issue of historical background in science. During the second Soviet occupation (1944–1990) Lithuanian science developed in coordination with soviet standards, including requirements for the style of PhD abstracts: the language used in abstracts was Russian in most cases (exception were made for the humanities, when a monograph on the dissertation subject was published). A research advisor was assigned according to strict rules of the Supreme Commission on Attestation, in Moscow, which gave rise to a tradition of national science in Lithuania.

As a document, dissertations provide data on new knowledge, but also encode important scientometric information: Research advisor; Official opponents; Dissertation committee; Degree granting institution; Institution of research; and International classification number.

All the above might be used for mapping scientific communities, invisible colleges or other social structures in science with special emphasis on the teacher-student scholarly link on the personal level and science curation on the institutional, national and even international ones. Research advising was the criterion we applied in our prosopographic project.[7]

Data for prosopography of science in a dissertation record

Records about dissertations maintained before 1996 are available in published bibliographies or library catalogues. Data are also registered by research and academic institutions at the national level; information is available from the *Lithuanian Academic Electronic Library - eLABa*[8] (data from 1996 onwards, 2637 records), the Database on Nostrificated Dissertations (data for 2003–2008),[9] and Database of Maintained Dissertations (run by Research Council of Lithuania, data from 2006 onwards, 4533 records).[10] Martynas Mažvydas Lithuanian National Library holds almost 9000 records in the Online Public Access Catalogue (OPAC), though one should note that the result set includes all records

[6] Clara Calero, Renald Buter, Cecilia Cabello Valdés, & Ed Noyons, "How to Identify Research Groups Using Publication Analysis: An Example in the Field of Nanotechnology," *Scientometrics* 66, no. 2 (2006): 365–76.

[7] More about prosopography in Koenraad Verboven, Myriam Carlier, & Jan Dumolin, *A Short Manual to the Art of Prosopography"*, available from http://prosopography.history.ox.ac.uk/images/01%20Verboven%20pdf.pdf last access on October 13 2016.

[8] *Lithuanian Academic Electronic Library - eLABa*, http://www.lvb.lt/

[9] *Lietuvos mokslo potencialas*, http://mokslas.lmt.lt/

[10] Lietuvos mokslo taryba, "Disertacijų gynimų duomenų bazė", http://www.lmt.lt/lt/paslaugos/disertacijos/d-db.html

with Universal Decimal Classification (UDC) auxiliary form no. 043.3.[11]

Dissertations and theses are indexed in *Isis Cummulative Bibliography*; the data were made available for scientometric surveys on a new platform, *IsisCB Explore*, where a service for thesis search was launched in 2015. Records are provided with the names of the research advisors and doctoral committee members. These names are linked to the records in which they are indicated.[12]

Commercial company *ProQuest* markets a product names *ProQuest Dissertations and Theses Global (PQDT Global)*, which holds a collection of 1,7 million full-text graduate works. *PQDT Global* was designated as an official offsite repository for the US Library of Congress. Comprehensive historic and ongoing coverage from North American universities is supplemented with records from Europe and other countries. The search options include research advisor and degree granting institution.[13]

Since its establishment in 1993, the *Commission on Bibliography and Documentation* (CBD) of the International Union of History and Philosophy of Science and Technology/Division of History of Science and Technology (IUHPST/DHST) strives to foster the development and utilisation of bibliographic and archival documentation to assist historians in their study of science and technology. CBD also looks forward to professional advice and advocating from key specialists in bibliography and archives.[14] The idea to investigate scholarly communication documented as research advising (supervising) with degree-granting institutions as a contributors became one of CBD interests since 2013.

Methodology

Since communication is transforming through the use of information technologies very rapidly, historians of science have but to master instruments to investigate. In a study of scholarly communication Christine L. Borgman and Jonathan Furner suggested bibliometrics as a powerful set of methods and measurements for studying the structure and process of scholarly communication. A study of links in science enables to produce maps or networks

> "[…] of individuals positioned in a way that demonstrates their relatedness to one another. Such maps may then be used (a) to inform descriptions and explanations of the historical and contemporary structure and direction of communication in particular organizations, domains, or geographical areas, and to assist in the making of inductive predictions of

[11] *Lithuanian National Martynas Mažvydas Library*, http://www.lnb.lt/

[12] *IsisCBExplore*, http://data.isiscb.org/

[13] *ProQuest Dissertations and Theses Global*, http://www.proquest.com/products-services/pqdtglobal.html

[14] Birutė Railienė, "Bibliographical Control for History of Science Commission on Bibliography and Documentation of the IUHPST/DHST", *Acta Baltica Historiae et Philosophiae Scientiarum, vol. 1, no 2 (Autumn)* (2013): 91–96; DOI: 10.11590/abhps.2013.2.08.

future trends, and (b) in information retrieval (IR) systems, to assist information seekers in identifying probably relevant documents."[15] (p. 5).

All libraries in Lithuania use UDC for cataloguing documents. Thesis records are allocated common auxiliary form number 043.3. Automatic search in electronic catalogues returns many more documents – surveys of theses, manuals, historical studies of scholarship, etc. A search in Lithuanian national library catalogue returned over 9,000 items – including all pertinent items. We agree on that a UDC filtering in a search is essential, but additional criteria for thesis search in library catalogues should be applied, and only specialised databases with relevant records should be used in a prosopographic survey of scholarly links and social structures in science.

By comparing dissertation records from different datasets one may conclude that the decision to include data, as mentioned above, still depends on the institutional policy for cataloguing. Although no one argues title and authority data, the indication of research advisors and degree-granting institutions still needs advocacy.

Metadata in a database record inherited a tradition of library cataloguing, sometimes its restrictions too. A library catalogue record covers only basic information about a document according to cataloguing rules. Machine-Readable Cataloguing (MARC) opened vast possibilities to enter additional data of a document. MARC field 700 was designed for personal names, including thesis/dissertation advisors. Unfortunately, these possibilities, or even enthusiasm, are not always met. Much information on dissertations or dissertation abstracts is not event entered into the records, but is considered to be 'optional'. Even NDLTD, an international organization devoted to promoting the adoption, creation, use, dissemination and preservation of electronic theses and dissertations (ETDs) consider research advisors and degree-granting institutions as optional data.[16] Also the *Guidelines for Description of Theses and Dissertations* issued by world reputed Online Computer Library Centre (OCLC) recommend omitting these data,[17] thus impoverishing the scientometric surveys of scientific communities and the history of scholarly communication in general.[18]

Library consortia of Lithuania – eLABA - issued a rule to indicate research advisors in the ETD records,[19] therefore opening a possibility to investigate research clusters, identify

---

[15] Christine L. Borgman, & Jonathan Furner, "Scholarly Communication and Bibliometrics," *Annual Review of Information Science and Technology* 36, (2002) doi: 10.1002/aris.1440360102

[16] NDLTD, *ETD-MS v1.1: An Interoperability Metadata Standard for Electronic Theses and Dissertations*, available from http://www.ndltd.org/standards/metadata, last access on October 13 2016.

[17] OCLC. *Guidelines for Description of Theses and Dissertations,* available from http://oclc.org/bibformats/en/specialcataloging.html, last access October 13 2016.

[18] In private correspondence to author B. Railienė, from August 26 2016, Glenn E. Patton, a director of WorldCat Quality Management, OCLC, stated: "(...) when we rewrite section 3.1, to allow the use of authorized access points for the thesis advisor and the degree-granting institution if the appropriate relationship designator ("degree supervisor" and "degree granting institution") and/or relator code ("dgs" and "dgg", respectively) is included as part of that access point. These updated instructions should be available in the next few months."

[19] "ETD pateikimo instrukcija", available from

scientific schools (invisible colleges) and map the development of science in a country. Our project to edit dissertation records now has an information technology (IT) solution, and only human resources will set the limits to a complete index of dissertations database.

Conclusions

Thesis/ dissertation records are considered complete when they include data on research advising and institutional affiliation. Only then they become a source able to indicate **formal** links in a scholarly communication

Theses/ dissertations as documents provide data on new knowledge, but also encode important scientometric information.

Thesis/ dissertation records expose a scientific relationship of research advising both at the personal and institutional levels within a certain scientific area.

Indexes of research advisors and degree-granting institutions provide data for science maps (both local and international). If applied with a coordinate of time, a map becomes both a source for science history and a research development plan.

https://etd.elaba.lt/web/docs/ETD_pateikimo_instrukcija_studentams.pdf, last access on October 13 2016.