

# Format Project Proposals Public-Private Center for Digital Humanities

## **Organization:**

(UvA, VU, KNAW)

UvA

## **Private partner(s):**

Elsevier

Contact person: Dr. Henk Moed / Dr. Gali Halevi

## **Title Project Proposal**

Evaluating the Humanities:

Citation Monitoring and University Press Prestige

## **Keywords**

Citation Analysis; Publisher/Academic Press prestige

## **Project description (maximal 1 page)**

(Motivation, Goal, Approach, Pilot study?)

Across a number of research subjects in the arts and humanities, the book or monograph is the sine qua non of scholarly output. Often it is a strong requirement for academic promotion and tenure. When writing a book, the humanities scholar is likely to give serious consideration to the academic/university press with whom he/she would like to publish. One of the deciding factors will be the 'prestige' of the press. In the end, what is important to the humanities scholar "is that the resultant publication from a prestigious university press is read, reviewed, and cited; it is purchased by libraries; it is taken seriously" (Gump, 2006, p. 76).

Where books appear in the citation/reference lists of journals, we are curious to know whether or not certain academic/university presses are cited more often than others. Due to the importance that humanities scholars place on choosing a press, one would expect to see the majority of citations to books published by those considered most 'prestigious'.

Our research objectives can be described in terms of four components. First, we are interested in determining the degree to which university press prestige is a mythical concept. If it is, whom does it serve, and is it possible to quantify this type of reputation? Specifically, we are interested in the following question: *How is university press 'prestige' reflected by natural journal citation practices in the humanities?*

Our approach is to focus on the book or monograph as a product of the university press/publisher. We are certain that a university press expects its monographs to be cited in journal articles, but we also recognize that the 'prestige' of the press need not necessarily be linked to citation counts. In fact journal-related citations may not even measure prestige, but they might shed some light on certain *expectations* surrounding prestige. In the meantime, we do not know of any study that has examined the journal-to-press citing-cited relationship using a large-scale dataset.

Our second objective is to provide some insight into the selection process of arts and humanities books for new book citation indices. The development of a book citation index should reflect, in part, citation exchanges occurring *between* journals and books (i.e., mutual rhetorical influences) but cited reference lists from books are not easy to obtain for citation analyses, thus it is more practical to start with a study of citing-cited behaviour *directed from* journal articles *to* books.

Our third objective is to map directed relationships from journals to university presses (via book reference strings). Here we hope to observe trends. *Do certain journals direct more citations to books published by [A] press, or will the distribution of citations demonstrate openness to many publishers?* Researchers from a certain subject might have the impression that publishing a book with [A] university press will lead to more recognition, but it may be true that books published by [B] university press receive more citations.

And finally, our fourth objective is to combine the results of our quantitative analyses using Scopus data with a qualitative analysis of the academic/university press publishers that we observe in the mapping procedure. There is value in using a combined quantitative and qualitative methodology in a study: each method can contribute answers to different questions. Bibliometric data conveys limited information about how a book published by a particular academic/university press is used; while qualitative data, which we plan to obtain with interviews, can illuminate underlying details associated with the academic/university press' publishing policies.

### Data specifications

We have been awarded a data grant of data from the Elsevier Scopus Arts & Humanities citation index for the proposed research. Table 1, below, outlines the data specifications:

Table 1. Elsevier data.

Scopus Arts & Humanities Citation Index	
<ul style="list-style-type: none"> <li>• 1,450 new titles (as of June 2009); 3,500 in total</li> </ul>	<ul style="list-style-type: none"> <li>• All records in <i>SciVerse Scopus</i> published since 1996 contain cited references.</li> </ul>
Journal article publication years: (1996-2000) and (2007-2011)	
Journal Classifications	
<ul style="list-style-type: none"> <li>• History</li> </ul>	<ul style="list-style-type: none"> <li>• ASJC 1202</li> </ul>
<ul style="list-style-type: none"> <li>• Literature and Literary Theory</li> </ul>	<ul style="list-style-type: none"> <li>• ASJC 1208</li> </ul>
Approx. 604 Journal Titles	
Data fields	
Scopus Document ID -- Citing Source Journal Title -- Citing Source ID -- Citing Publication Year -- Scopus Author Name + Id -- Citing Document Title -- Citing Volume Number -- Citing Starting Page -- Citing Document Type -- Cited Scopus Id (if available) --Cited Source Journal Article -- Cited Source ID -- Cited Author Name + Id (if available) --Cited Document Title -- Cited Volume Number -- Cited Starting Page -- Cited Document Type	

We will be working with papers from journals published in the years 1996 to 2000 and also during the later period of 2007 to 2011. The selected journals (n=604) shall fit under the *All Science Classification Codes (ASJC)* of *History (1202)* and *Literature and Literary Theory (1208)*. From each journal article record we would like to isolate reference strings to books. These references may be works that have been published in English as well as other languages. The two separate publication periods for the journal articles will be used in the development of directed network maps, based on journal title links to book publishers

### **Planning (6-9 months)**

(indicate which concrete research will be carried out)

The plan is to hire a Database Developer / Junior Bibliometrics Researcher who will assist us for a period of 8 months with the development of a MySQL database. We will receive the data granted from the Elsevier Bibliometrics Research Program (<http://ebrp.elsevier.com/aboutUs.asp>) sometime in June 2012. The starting date for the research assistant is likely to be July 1<sup>st</sup> 2012.

The programmer will work on the project four days per week with the principle investigator, Dr. Alesia Zuccala.

**Month 1-3:** Optimize the MySQL relational database for citation-based searches. Isolate reference strings to books in journal article citation lists.

**Month 4-6:** Identify and input the university press/academic publisher names of cited books by linking known data from the Scopus2012 dataset (author, date, book title) to data from book catalogs or book thesauri – i.e., WorldCat® library catalog (API - <http://www.worldcat.org/affiliate/tools?atype=wcapi>).

**Month 7-8:** Conduct pilot citation analyses with the dataset. Data will be prepared and exported to software programs for bibliometric mapping. Software tools may include: Network Workbench and Sci2 (see links below).

NWB Team. (2006). *Network Workbench Tool*. Indiana University and Northeastern University, <http://nwb.slis.indiana.edu>.

Sci2 Team. (2009). *Science of Science (Sci2) Tool*. Indiana University and SciTech Strategies, <http://sci2.cns.iu.edu>.

### **Budget:**

(Costs researcher, contribution private partner(s) in kind or financially)

- Costs programmer: 0.8 fte for 8 months: 24.000 euro
- In-kind contribution private partner: Scopus data

### **Societal value, impact**

The research is novel since little is known about the relationship between the 'prestige' of certain university presses and publishers, and how their books are actually used or cited in the Humanities Journal literature. More specifically we are interested in examining citations to books from certain presses/publishers in journal articles related to History and Literature and Literary Theory.

### **Risks and dependencies**

(e.g. Intellectual property, license)

The major risk is that we cannot automatically match the reference lists to books. This can be dealt with semi-automatically/manual searching.

### **Feasibility**

(Can the project be completed within allotted time and budget?)

As shown by the planning, the project is likely to be completed within an 8 months period. We have already have been granted the data.

### **Organizations involved:**

(Mention all organizations and their role)

- ILLC-FGw: Dr. Alesia Zuccala
- Vacancy: programmer
- Elsevier: as supplier of the data

### **Relation with existing research programme?**

Dr. Alesia Zuccala has recently been hired within the FGw-spearpoint Digital Humanities. Her background in bibliometrics and recent research record fits with the program theme: Evaluating the Humanities.

### **Deliverables/concreet product**

(Which deliverables? Open Access?)

Specialized data set with information on citations to university presses /academic publishers (monographs)

### **Techniques**

(Standards, open source)

SQL, database development, search algorithms/APIs

The search algorithm software created by the programmer will be open source.

The original Scopus data granted by Elsevier is not open source.