Can bibliometrics be used to evaluate research in the social sciences and humanities?

Professor Ben R. Martin

SPRU – Science and Technology Policy Research,
The Freeman Centre, University of Sussex

Brighton, BN1 9QE, UK

(B.Martin@sussex.ac.uk)

Presentation at the International Workshop on 'Evaluation of Research in Social Sciences and Humanities: Problems and Perspectives', Ca' Foscari University of Venice, 16-17 March 2012

Contents

Introduction – Study for ESF et al. in 2009/10

Recent developments in databases, indicators etc.
- Developments in WoS and Scopus
- Role of indicators in research assessment
- Existing SSH databases/lists

Main issues in creating a SSH database
- Underlying considerations
- Operational issues
- Strategic options for development

Potential approaches

Recommendations

Introduction

Aim of ESF study

- to explore the possibility of developing a bibliometric database for capturing the full range of research outputs from Social Sciences & Humanities (SSH) to help assess impact

Coverage

- not just international (WoS) journal articles
- also national journals, books/chapters, 'enlightenment literature', 'grey literature'
- plus non-textual research outputs (if possible)

Definition

- use the term 'bibliometric' to cover the full range of research outputs from SSH and their impacts
- i.e. not just WoS journal articles and citations

Background context

Growing pressure for 'accountability', performance indicators, 'value for money' etc.

- Established indicators for sc not appropriate for SSH

Developments in databases & publishing

- 'Open access' publications
- Improved coverage of WoS & Scopus
- Emergence of Google Scholar/Books
- National/disciplinary bibliographic databases
- Institutional repositories of research outputs

➔ What is the potential for developing an inclusive database for assessing research output and impact in SSH?

Recent bibliometric devlpts in SSH

WoS (Thomson-Reuters – previously ISI)
- Increased from 1700 to 2400 SSH journals (including 1200 'regional')

Scopus (Elsevier)
- Increased from 2050 to 3500 SSH journals
- Begun to add data on highly cited SSH books

Google Scholar
- Not (yet) systematic or rigorous in coverage
- But covers books, chapters, reports etc.
- New source of citation data

i.e. shift from ISI monopoly to competition
- Opens up new opportunities

Role of bibliometric indicators in research assessment

Research assessment growing

Often relies on WoS (or Scopus) for bibliometric indicators

- But ignores non-WoS journals, books/chapters etc.

Bibliographic databases

- e.g. ECONLIT, Sociolog Abstracts, Psychinfo
- Often wider coverage
- Currently not suitable for bibliometric analysis (Moed et al.)
    - Author/institution names not standardised
    - Lack of cited references
    - Differing quality criteria for inclusion
- Need standardised database structure & criteria

Role of bibliometric indicators in research assessment

Norwegian reference list
- Covers all sc, soc sc & humanities
- Includes national as well as international journals
- Classified into 2 categories (to avoid Australian problem)

European Reference Index for Humanities (ERIH)
- Covers humanities research in international & national journals in English & other languages
- Journal lists peer-reviewed

Australian ERA HCA
- 19,500 journals
- Single quality rating
- List peer-reviewed

Moed et al. and Hicks and Wang analyses
- Pros & cons of above approaches
- Above databases include some non-refereed/non-scholarly literature

Creating a SSH bibliometric database

1. Underlying considerations

- Need to raise awareness among research funders, policy-makers and others of the significant time required for development of a SSH bibliometric database
- Allow flexibility in terms of coverage
  - Start with scholarly articles & books
  - Then add other published outputs
  - Then non-published research outputs like artwork, exhibitions, excavation reports and photos
- Build on bibliographic lists of institutional & national repositories, but need
  - standardised database structure
  - similar quality criteria for inclusion

Creating a SSH bibliometric database

2. Operational issues

Different options

- Top-down approach – creating European database or strong coordination of national organizations
- Bottom-up approach – producers of existing national bibliographic databases etc. working together to develop common rules, procedures etc.
- Hybrid approach – e.g. European group develops a 'bibliometric manual' on requirements for a SSH research output database
    - Definitions, data & format, criteria for inclusion, database structure

Then producers of existing national bibliographic databases etc. invited to supply such data
    - Analogy with 1963 OECD 'Frascati Manual' for measuring R&D

Creating a SSH bibliometric database

2. Operational issues

- Bibliographic databases/lists need to be able to demonstrate that they include *high-quality research outputs* validated by experts
- Establishment of *basic threshold criteria* for determining which SSH research outputs of sufficient quality/ importance to merit inclusion e.g.
  - scholarly articles in peer-reviewed national & international journals
  - scholarly books that have been subject to a peer-review process
  - other SSH research outputs that have been subject to some quality-control process
- Need to carefully *monitor consequences* (both intended and unintended) on research process
  - e.g. use of publication counts in Australian funding formula
    ➔ proliferation of articles in lesser journals

Creating a SSH bibliometric database

3. Strategic options for development

- Whether new SSH database be developed by a European agency or national bodies
- Whether WoS, Scopus or Google Scholar be asked to assume responsibility
- Whether to support further development of digital repositories with common standards & data formats
- Whether to build on existing initiatives e.g. DRIVER
- Whether to build a collaboration of European research councils, or seek funding from a European source

Potential approaches for consideration

Synthesis of suggestions by Moed et al., and
Hicks & Wang ➔ 6 options

1. Create more comprehensive national bibliographic systems through *development of institutional repositories*
   - Existing digital repositories only cover ~10% of published output ➔ considerable scope for coverage to be extended
   - Some countries/institutions will need help in capability-building
   - Need to coordinate repositories to capture full range of research outputs in standardised form
   - Encourage repositories to begin capturing cited reference lists

   Implication – need to develop
   - relevant capabilities
   - institutional repositories

Potential approaches for consideration

2. Enhance and build upon existing national documentation systems through the development and standardisation of *institutional research management systems*

   - Build upon an existing research information system (e.g. METIS in the Netherlands)
   - Expand through development and application of interfaces to bibliographic lists that include books and monographs
   - Or build on e.g. the DRIVER initiative
   - Link institutional repositories to chosen research information system

   Implications
   - Establish a minimum threshold criterion
   - Investigate possibility of adapting/combining existing systems

Potential approaches for consideration

3. Create a new SSH database *from publishers' archives & institutional repositories*, adding data on enlightenment literature and non-textual outputs
(cf. Spanish initiative)

   - Create new database including publication and citation data obtained from publishers
   - Identify *enlightenment* books & periodicals, categorise and assign levels
   - List and assign levels for *non-textual outputs* agreed by national experts

   Implication

   - Cost & complexity of creating & maintaining such a database large ➔ probably not suitable to kick-start SSH database initiative

Potential approaches for consideration

4. Take advantage of competition between *commercial database producers* (WoS, Scopus, Google Scholar) to strengthen coverage of SSH research outputs

    - Decide who should explore whether a deal might be negotiated
    - Then approach publishers re expanding their coverage

    Implication

    - Need someone with (i) extensive knowledge and (ii) necessary authority to negotiate with publishers

Potential approaches for consideration

5. *Integrate specialised SSH bibliographic lists* into one comprehensive bibliographic database
   - Move towards agreed standardisation of database structure among main producers
   - Examine existing selection criteria and how these might be standardised
   - Add in books etc.

   Implication
   - Need for a group of bibliometric/library science experts to spearhead process of standardisation

Potential approaches for consideration

6. Encourage further *development of Open Access approach* to overcome barriers of accessibility and enhance visibility of smaller journals/publishers
    (cf. US initiative; also some European university presses)

    - Build and maintain an electronic full-text SSH journal infrastructure
    - Include peer-reviewed journals not on-line and not indexed by WoS or Scopus
    - Build upon OAPEN digital library and include more European book publishers
    - Integrate above through development of appropriate interfaces
    - Agree a set of metrics

    Implications

    - Potential redundancy of effort
    - Potential conflict of interest with current database publishers

    Each of above approaches has various advantages and disadvantages (see Box 1 on pp.26-28 of SPRU report)

Recommendations

Three main recommendations

For each, we propose a **hybrid approach** combining top-down and bottom-up actions

- top-down to ensure necessary coordination and 'clout'
- extensive bottom-up involvement to build on existing expertise in production & development of bibliographic databases

Recommendations 1 and 2 may be undertaken in parallel to save time and to 'test' which is likely to prove more effective

Decided not to pursue other options because of cost &/or practicality

- Open Access approach
- integration of specialised SSH bibliographic lists
- creation of a new database of SSH research outputs from publishers' archives and institutional repositories

Recommendations

1. Define criteria for inclusion of SSH articles & books, and establish a standardised database structure for national bibliometric databases
   Top-down
   - Small number of Res Councils to take initial lead ('Lead RCs')
   - Appoint standard-setting body of ~6 experts (bibliometric, library sc etc)
   - Consult with SSH scholars & others re SSH research outputs, quality & impact criteria, appropriate 'book metrics' etc.
   - Establish minimum criteria for inclusion in SSH bibliometric databases
   - Seek inputs from publishers, repositories etc.
   - Seek funds

   Bottom-up – national institutions, repositories etc.
   - apply inclusion criteria – transform databases from bibliographic to bibliometric
   - identify high-quality journals & books
   - implement standardised database structure
   - monitor effects

Recommendations

2. Explore option of involving a commercial supplier in the construction of a single international SSH bibliometric database

   Top-down – standard-setting body to
   - consult with those who have dealt with Thomson-Reuters, Elsevier & Google
   - decide whether these publishers be asked to 'clean up' existing data, or invited to construct new database
   - approach and obtain quotes

   Bottom-up – national institutions, repositories etc. to
   - develop bibliographic databases to input into eventual SSH bibliometric database
   - consult with broad range of SSH researchers to ensure quality & validity of data; also to monitor effects on research behaviour

Recommendations

3. Longer-term expansion and enhancement of the SSH bibliometric database to include other SSH research outputs
   Top-down
   - Decide who is to be responsible for maintaining SSH bibliometric database
     - Will require collective funding from RCs or European Union
     - Then issue 'Invitation to tender'
   - Standard-setting body to
     - consult with SSH scholars etc, then decide what other SSH research outputs to include e.g. 'grey' & 'enlightenment' literature
     - seek advice on criteria etc. from leading HEIs experienced in producing bibliographic databases & data on non-textual outputs
     - consult with commercial suppliers, bibliometric experts etc.

   Bottom-up – national institutions, repositories etc. to
   - include other SSH research outputs as identified above
   - apply agreed inclusion criteria

References

- Burnhill, P.M. and M.E. Tubby-Hille (2003) 'On Measuring the Relation between Social Science Research Activity and Research Publication'. *Research Evaluation*, 4, 3, 130-152.
- CWTS (Centre for Science and Technology Studies), Leiden University, (2007) *Scoping study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions.* A Report to HEFCE.
- Hicks, D. and J. Wang (2009) *Toward a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project.*
- Kyvik, S (2003) 'Changing trends in publishing behaviour among university faculty, 1980-2000'. *Scientometrics*, 58, 1, 35-48.
- Martin, B.R. et al., 2010, *Towards a Bibliometric Database for the Social Sciences and Humanities – A European Scoping Project*, A report produced for ESF, ANR, ESRC, DFG and NOW, Brighton: SPRU.
- Moed, H. et al. (2009) *Options for a Comprehensive Database of Research Outputs in Social Sciences and the Humanities*.
- White, B.D. (2006) Examining the claims of Google Scholar as a serious information source. *New Zealand Library & Information Management Journal*, 50 no 1, pp. 11-24.