Developing Metrics for the Humanities

Alesia Zuccala

Royal School of Library & Information Science
University of Copenhagen
Presentation:

1. Educate humanities scholars and involve them in the process of data curation and metric developments

2. The ‘humanistic’ focus is on the return to the ‘biblio’ in bibliometrics: books and measures of books

The Humanist

• In practice, the humanist may be defined as a scholar of philosophy, philology, theology, history, archaeology, media studies, and literature.

• Intellectually, they belong to a unique ‘tribe’ of specialists dedicated to examining, critiquing, and highlighting patterns of relationship across products of the human mind (Bod, 2013).

Faculty of Humanities:

“Interplay between people, information and technology”

Undergraduates to PhDs

Libraries, Public Administration & Private Companies
Bibliometric Education for Humanists?

CWTS Course 'Measuring Science and Research Performance'

Our one-week course provides participants with the bibliometric knowledge and skills they need to interpret bibliometric statistics properly and usefully.
Bibliometric Education Programs:

Library and Information Science


FIGURE 1. A map of information science.
A Humanistic Approach:

- Research Evaluation
- Education
- Feedback
- Humanistic Research
- The Metric-Oriented Humanist

Current Gap
Inciting the Metric Oriented Humanist:

- Empiricism in the Humanities
- Eschew false assessment ideals
- A return to the ‘biblio’ (books)
- Digital humanities versus humanistic bibliometrics
- Record keeping for a humanistic programme
- Explaining mathematical indicators (e.g. normalization)
- Problem-based learning
A Return to the "Biblio"

Biblio + Metron = Books and Measures
Denmark: BFI Performance Points

How BFI-points are calculated
BFI-points are calculated at the University level based on the following formula:

- BFI = B x C x F, where
  - B = Basic points
  - C = collaboration factor (1.25 if any external authors)
  - F = Local authors / All authors

Basic BFI-points:

<table>
<thead>
<tr>
<th>Publication type</th>
<th>Level 1 *</th>
<th>Level 2 **</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monograph with publisher</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Monograph in book series</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Article in series (journal, book series, conference serie)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Contribution to anthology with publisher</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>Doctoral thesis</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Patent</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Level 1 – Normal
Level 2 – Prestigious

The Book Citation Index:

Recent assessments of the BKCI point to numerous indexing problems, which can lead to over-counting or under-counting of books (e.g., Gorraiz et al., 2013; Leydesdorff & Felt, 2013; Torres-Salinas et al., 2013).
Metrics for Books as ‘Families’ of Works

Regardless of where or how publication counts or citations of books are collected:

It is essential to recognize that ‘books’ often belong to bibliographic families
Metrics for Books as ’Families’ of Works

We can examine bibliographic families both theoretically and empirically:

1. The family-oriented entity-relationship model: The Functional Requirements for Bibliographic Records (FRBR)

1. Collect and assess data from the Book Citation Index, Danish Pure Registry, OCLC-WorldCat, and Goodreads
Figure 1. Modified FRBR model (adapted from Tillet, 2001) of ‘bibliographic’ families for a scholarly work.
Model of a ’Complex’ Work:

I would consider the English [expression] as a kind of improved edition of the Dutch book. When the Dutch work was translated into English, I sent it to OUP [Oxford University Press] and incorporated the comments by the 5 OUP reviewers into the English version.... (R. Bod, personal communication, June 16, 2016)
Figure 2. Model of a complex ‘work’ with expressions and manifestations of a ‘new work’
Research:

• Do books currently indexed in the Book Citation Index (BKCI) have adequate metadata and data designed to reflect inherent familial components and relationships?
Method:

Extracted ISBNs for monographs (over 16,000):
• from the BKCI
• From the Danish Pure Repository

Matched with equivalent ISBNs and ‘family’ related ISBNs in:
• OCLC-WorldCat
• Goodreads

Added all ISBNs and metadata related to monographs to an SLQ database
Table 1. ISBN matching and retrieval results for total manifestations, expressions, and works.

<table>
<thead>
<tr>
<th></th>
<th>The Book Citation Index</th>
<th>The Danish Pure repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Number of ISBNs crawled</td>
<td>16,392 (35%)</td>
<td>8,604 (81%)</td>
</tr>
<tr>
<td>2 Number of overlapping ISBNs</td>
<td></td>
<td>35 (0.41%)</td>
</tr>
<tr>
<td>3 Extra related ISBNs found in OCLC-WorldCat and Goodreads</td>
<td>30,903 (65%)</td>
<td>2,042 (19%)</td>
</tr>
<tr>
<td>4 Total unique ISBNs in the dataset under study</td>
<td>47,295 (100%)</td>
<td>10,646 (100%)</td>
</tr>
<tr>
<td>5 ISBNs with distinct language and publication year</td>
<td>34,236</td>
<td>8,362</td>
</tr>
<tr>
<td>6 Total Expressions</td>
<td>20,284</td>
<td>7,844</td>
</tr>
<tr>
<td>7 Total Works</td>
<td>16,311</td>
<td>8,195</td>
</tr>
</tbody>
</table>

65% extra ISBNS found
On Average a monograph is usually published as 2 to 3 editions with 3 to 4 ISBNs

**Figure 3.** Frequency distribution of works with one or more ISBNs and published as one or more edition.
Figure 4. Indexing quality of the BKCI based on ISBNs per edition for publication years 1995-2015.

No pattern except increase in indexing of books, with poor recording of IBNS for multiple ‘work’ related editions.
Many are Danish texts published as one edition, or a translated edition that was not originally published in English by translating author.

**Figure 5.** Indexing quality of the Danish Pure Repository based on ISBNs per edition for publication years 1995-2015.
Indexing model for the BKCI for more accurate ‘metrics’

**Work ID:** Highest proposed target for citation count


**Expression ID:** Lowest proposed target for citation count

**Manifestations:** Include all series of ISBNs per edition
Conclusions:

• Book Citation Index is not an accurate index of monographs as complex works

• National data repositories (as in the Danish case) tend to be more accurate – but little overlap with the BKCI in terms of coverage

• Family-oriented entity-relationship model – the FRBR - is recommended for new indexing approach to monographs in the BKCI
Research:

• Zuccala, A. (2016) Inciting the metric oriented humanist: Teaching bibliometrics in a faculty of humanities. Education for Information 32(2), 149-164.

Thank you kindly for your attention!

E-mail: a.zuccala@hum.ku.dk