CITATION SMOG
DATASETS FROM THE FUMES OF A SYLLABUS CONTENT ANALYSIS STUDY

CAROLYN HANK
UNIVERSITY OF TENNESSEE, KNOXVILLE

IDCC16 - AMSTERDAM
WHAT

INTENSIVE REVIEW OF DIGITAL CURATION SYLLABIS

FYI

COMPARING DIGITAL LIBRARY (DL) & DIGITAL CURATION (DC) SYLLABIS

WHAT EXTENT DO SYLLABIS OVERLAP?
WHY

CORE OF FREQUENTLY ASSIGNED DC “READINGS?”
FREQUENTLY ASSIGNED AUTHORS & JOURNALS
FREQUENTLY ASSIGNED CONTENT & FORMAT TYPES
OTHER SELECT SYLLABUS CHARACTERISTICS
AND WHERE BEING TAUGHT (PROGRAM-LEVEL)
WHO

5 students | 2 faculty

OVER MULTIPLE SEMESTERS
LACK OF CONSISTENCY
VARIATIONS IN QUALITY
HOW_1

DATA COLLECTION?

excel (access)

bad choice?
101 programs (ALA, SAA & iSchools)
93 English-language programs

Schedules:
71 course catalogs available but only 35 course schedules posted
17 for 2012-2013 academic year
33 for 2013-2014 academic year
46 for fall 2014 academic semester

Courses:
19 had courses in digital curation
then limited to “foundation” courses

Syllabi:
15 coded after removing for duplicates
11 complete (inc. reading outline)
4 partial (front matter only)
HOW

DATA COLLECTION?

excel (access)

bad choice?
729 unique citations

AFTER MUCH EFFORT (AND TIME) TO CLEAN-UP CITATIONS AND ASSESS FOR DUPLICATES
A very long tail of distribution

846 (98%)

Only assigned once among the digital curation syllabi

Just an FYI as an aside
HOW_3

FOR FREQUENCY OF CITATION BY TYPE:

24 types
IDENTIFIED (n=729)
MORE CONTEXT FOR DATA CODING/ MANAGEMENT

<table>
<thead>
<tr>
<th>TYPE</th>
<th># (CUMULATIVE %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>REFEREED JOURNAL ARTICLE</td>
<td>247 (33.9%)</td>
</tr>
<tr>
<td>ORGANIZATIONAL PUBLICATION</td>
<td>122 (50.6%)</td>
</tr>
<tr>
<td>PROJECT PUBLICATION</td>
<td>62 (59.1%)</td>
</tr>
<tr>
<td>CONFERENCE PROCEEDING</td>
<td>42 (64.9%)</td>
</tr>
<tr>
<td>TRADE JOURNAL ARTICLE</td>
<td>35 (69.7%)</td>
</tr>
<tr>
<td>NON-REFEREED JOURNAL ARTICLE</td>
<td>33 (74.2%)</td>
</tr>
<tr>
<td>WEB-BASED AV MATERIAL</td>
<td>24 (77.5%)</td>
</tr>
<tr>
<td>INDUSTRY PUBLICATION</td>
<td>20 (80.2%)</td>
</tr>
</tbody>
</table>

N=729 unique citations
<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book</td>
<td>14</td>
<td>82.2%</td>
</tr>
<tr>
<td>Book chapter</td>
<td>13</td>
<td>84.0%</td>
</tr>
<tr>
<td><strong>Software app</strong></td>
<td>12</td>
<td>85.7%</td>
</tr>
<tr>
<td>Personal blog</td>
<td>12</td>
<td>87.4%</td>
</tr>
<tr>
<td>Standards publication</td>
<td>12</td>
<td>89.0%</td>
</tr>
<tr>
<td><strong>Instructional material</strong></td>
<td>11</td>
<td>90.5%</td>
</tr>
<tr>
<td>Consumer magazine article</td>
<td>11</td>
<td>92.0%</td>
</tr>
<tr>
<td>Organizational blog</td>
<td>10</td>
<td>93.4%</td>
</tr>
<tr>
<td>Other (misc)</td>
<td>10</td>
<td>94.8%</td>
</tr>
<tr>
<td>Other (white paper)</td>
<td>9</td>
<td>96.0%</td>
</tr>
<tr>
<td>Newsletter or bulletin</td>
<td>8</td>
<td>97.1%</td>
</tr>
<tr>
<td>Newspaper</td>
<td>7</td>
<td>98.1%</td>
</tr>
<tr>
<td>Thesis, diss., student paper</td>
<td>5</td>
<td>98.8%</td>
</tr>
<tr>
<td>Reference work</td>
<td>4</td>
<td>99.3%</td>
</tr>
<tr>
<td>Unknown</td>
<td>3</td>
<td>99.7%</td>
</tr>
<tr>
<td>Monograph or paper series</td>
<td>2</td>
<td>100%</td>
</tr>
</tbody>
</table>

N=729 unique citations
HOW_4

INDEXING

PUBLICATION YEAR
JOURNAL (authority, e.g., name changes)
AUTHOR (authority; personal/organizational)
PROJECT (authority; also funding agency)
“VERSIONS”/EDITIONS (e.g., OAIS; “TDR”)
ITEMS IN A COLLECTION (e.g., DPM Workshop)
Data Deposit?

Data Sets or Data Tables?

DL Curriculum Project Dataset (wiki)
DONE IN FIVE MINUTES (HOPEFULLY!)  
THANKS!