TEXT MINING RESEARCH JOURNALS:
THE WHOLE IS GREATER THAN
THE SUM OF THE PARTS

UKSG FORUM, London
16 November 2016
Colleen Campbell
IN THE SPIRIT OF THE FORUM THEME, THE ORIGIN OF FOOTBALL

Stradanus. Gioco del calcio in Piazza Santa Maria Novella (Game of Football). 1562-72. All images in this presentation are taken from the Artstor Digital Library.
ITHAKA

ITHAKA is a not-for-profit organization that helps the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways.

JSTOR is a not-for-profit digital library of academic journals, books, and primary sources.

Ithaka S+R is a not-for-profit research and consulting service that helps academic, cultural, and publishing communities thrive in the digital environment.

Portico is a not-for-profit preservation service for digital publications, including electronic journals, books, and historical collections.

Artstor provides 2+ million high-quality images and digital asset management software to enhance scholarship and teaching.
TEXT ANALYSIS: AN EARLY EXAMPLE
DATA FOR RESEARCH (LAUNCHED IN 2008)
GROWTH OF DIGITAL HUMANITIES SCHOLARSHIP

Analyzing trends in academic language

Isolating key terms for whole disciplines

Mapping author networks

Creating visualizations of datasets

Tracing concepts across disciplines

Support for large-scale or specialized datasets that may include:

- **Article metadata**
- **Article full-text**
- **Word frequencies**
- **N-grams (bigrams, trigrams, quadgrams)**
- **Reference citations**

**Process:**

- *Researcher completes a brief online request form, specifies dataset query*
- *JSTOR prepares a brief letter agreement, research or library signs*
- *JSTOR processes dataset, delivers researcher AWS link to download custom custom dataset*
Service and datasets are free, permitted under Terms & Conditions.

Letter agreement is required for full-text or large-scale requests, and may be established with individuals and/or libraries.

Dataset content is not limited by licenses or institutional affiliation.

Support is available for scholars worldwide.

JSTOR SUPPORTS DATA MINING – WHY?

• Supporting new types of scholarship is part of our mission
• Opportunities to build beneficial partnerships
• Increasing value of publications; corpus in and of itself has value as a scholarly tool
EXAMPLES OF SUPPORTED PROJECTS


- Tracks the use of “Enlightenment”, in historic and language context.
- Explores questions on the way scholars present research and the “terminological woods” in naming an historical age.
GENDER COMPOSITION OF SCHOLARLY PUBLICATIONS (1665 - 2011)

Date range: 1665 - 2011 | 1665 - 1989 | 1990 - 2011 | top papers

Legend:  Female authors  Male authors

JSTOR  Law  US constitutional law

http://www.eigenfactor.org/projects/gender/
PRESS COVERAGE

“Women still are not publishing...in the same proportion as they are present in academe as professors.”

http://www.chronicle.com/article/The-Hard-Numbers-Behind/135236


Thank you!

Vink, J. Diego Maradona celebrating Argentina victory of World Cup at Azteca stadium. 1986.

COLLEEN CAMPBELL
DIRECTOR, INSTITUTIONAL PARTICIPATION & STRATEGIC PARTNERSHIPS EUROPE
Colleen.Campbell@ithakainternational.org