Web of Science™ as a Research Dataset

Katya Börner, 1 Valentín Pentchev, 1 Matthew Hutchinson, 1 James Pringle, 1 Jason Rollins, 2 Yadu N. Babuji, 5 & Eamon Duede 5

1Katy@indiana.edu CNS, SOC & Network Science Institute, Indiana University, Bloomington, IN, US
2maahutch@iu.edu & vpentchev@iu.edu CNS, SOIC & Network Science Institute, Indiana University, Bloomington, USA
3jason.rollins@clarivate.com & james.pringle@clarivate.com Clarivate Analytics, USA
5yadun@iu.edu & duedee@uchicago.edu Computation Institute, Knowledge Lab, University of Chicago, USA

Introduction
The Clarivate Analytics Web of Science (WOS) has served as a research dataset for more than 9,000 scholarly articles in the past 15 years alone—across a wide range of fields and disciplines. In this paper, researchers have developed a new generation of enclave supported, high performance, and commercial interests. Moreover, these environments enable researchers to collaborate on research projects in the fast paced, interactive style that has come to dominate research in the era of Big Data—namely, "discovery science".

WOS Data Enclave
The Indiana University Network Science Institute (IUNI) acquired the complete set of Clarivate Analytics’ data from ScienceXML, raw data (WOS) and knowledge version 5. The data was parsed and stored in a well-documented PostgreSQL database, see entity-relationship diagram, database schema, and data dictionary on https://iuni.iu.edu/resources/web-of-science. The Cloud Kotta WoS database schema can be found on GitHub (see: https://github.com/alexander-bellkov/wos). The Cloud Kotta database schema can be found on IUNI’s document page (http://iuni.iu.edu/resources/web-of-science). More about Cloud Kotta can be found at: http://cloudkotta.org.

New Computational Infrastructures
Research leveraging big, scholarly datasets like WOS presents researchers with challenges related to the size and nature of relationships, and collaborative (proprietary) nature. To overcome these challenges, researchers have developed a new generation of enclave supported, high performance, and commercial offerings that have been enabled and demanded in response to increasing reliance on the proprietary; 2) some recent findings that have been made possible by these infrastructures; and, 3) new commercial opportunities that can be pursued with the appropriate computational and analytical capabilities opens up a wide range of funding and subsequent publishing opportunities in high impact venues. Second, data providers can pursue new business opportunities, including novel and analytical capabilities opens up a wide range of funding and subsequent publishing opportunities in high impact venues. Second, data providers can pursue new business opportunities, including novel and analytical capabilities opens up a wide range of funding and subsequent publishing opportunities in high impact venues. Second, data providers can pursue new business opportunities, including novel and analytical capabilities opens up a wide range of funding and subsequent publishing opportunities in high impact venues.

IU International Co-affiliation Network, 2004-2013
CNS @ Indiana University 2016

Pushchino

Beijing

Vancouver

New Delhi

Beijing

New Delhi

Figure 1. IU Co-affiliation Network

Figure 2. Tracing Inventive Teams

New Commercial Offerings
The value of the Web of Science as a search and discovery tool is well established at thousands of research institutions worldwide. But the commercial opportunities for the use of its high-quality metadata outside of the platform for individual discovery and research are perhaps less clear. When researchers need to study broad-scale trends in science, technology, and innovation, they often turn to the Web of Science as the most comprehensive citation source to provide over 100 years of consistent, global publication data. Increasingly, user requests for this data take the form of custom reports, curated data sub-sets, and large-scale raw XML delivery.

Acknowledgements
This work was supported by and contributes to research for IBM, Facebook, Jump Trading, Arts and Social Sciences, the National Institute of Health under awards F31 DA039397 and 4U54CA198334 and the National Science Foundation under awards NSF-1538763, BES1546392, and NCF1553044. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This work uses Web of Science data by Clarivate Analytics.

References