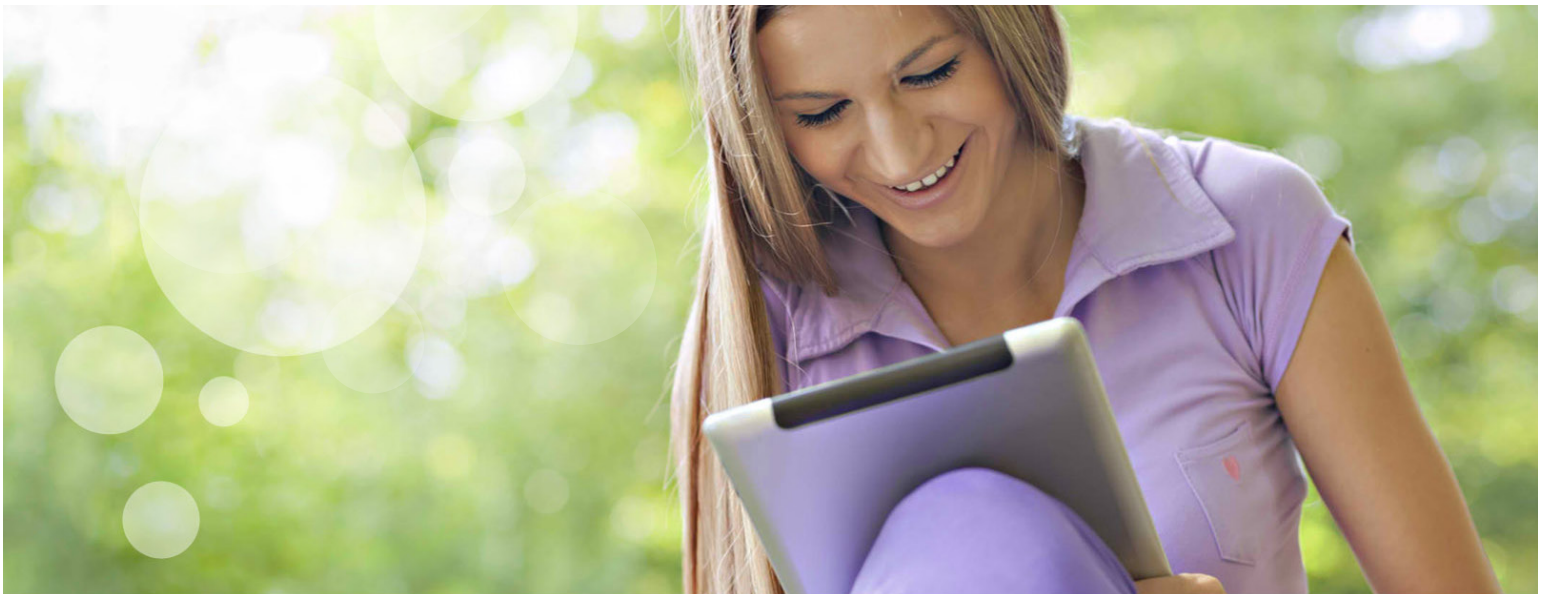# CASE STUDY: MAPPING RESEARCH TRENDS WITH PROQUEST DISSERTATIONS & THESES

The University of North Carolina at Chapel Hill

ProQuest®

The social sciences influence almost every aspect of daily life; from economic policy, to health and social policy, and from the environment and energy to technology and innovation. Understanding these trends and assessing the impacts on daily life takes skilled analysis — and given the diversity and complexities of the data, such analysis requires detailed modeling.

For Charles Kurzman, sociologist at The University of North Carolina at Chapel Hill, this was a fact known only too well. With published papers on major social issues including The Arab Spring Uncoiled, For Export Only: Diffusion Professionals and Population Policy and Islamic Studies in U.S. Universities, his latest research project is to understand the extent to which American social science has internationalized in the last 50 years. "Being able to look back at trends in the social sciences means that policy makers can understand what the important topics are and how to shape policy going forward," said Kurzman. "As you can imagine, the amount of data that has been published in social sciences over the last half century years is quite considerable. Our team has collected millions of journal articles, dissertations, and books from American academia, each of which have different subjects, headings etc. The aim of this project is to identify the geographical subjects for each piece of scholarship in order to make sense of this data and identify whether or not American social science is paying more attention to international subjects than it has in the past."

To do this, the team developed an algorithm that looks at the different pieces of information to try and determine where trends and patterns exist. It was with this analysis that the ProQuest Digital Dissertations team was able to help. "As I mentioned, the research includes thousands of dissertations, as well as other data sources. As your readers may know, ProQuest is the Library of Congress' official repository for all dissertations published in the United States. So I approached them and told them about my project. They were incredibly helpful and agreed to provide the metadata for the dissertations in their vault."

The ProQuest dissertations vault contains more than 2 million full-text dissertations and theses and 3.5 million A&I records in electronic format. "Receiving this metadata has been invaluable," continued Kurzman. "Without this access, we would have had to attempt to scrape the information from library catalogs, the web and other ProQuest databases to which our library subscribes— a not inconsiderable task!"

The first task was to provide the metadata for the dissertations that were available in electronic format at that time. The metadata was exported in excel spreadsheets which were then stored on a secure part of the institution's server. For the last decade, ProQuest has undertaken digitization of the older, paper-based dissertations and theses stored in their vaults. As such, as new files were digitized their metadata was also uploaded providing even more valuable sources to support Kurzman's research. "ProQuest supplied individual files for each ETD (electronic dissertation or theses) so after we received the data, we then had to combine the separate files into a single spreadsheet."

In total, more than one million files were submitted for analysis and the team's algorithms were then run across the data to analyze trends. As a result of access to the metadata, it was possible for geographic locations and subjects to be identified for more than 50% of the dissertations supplied. "ProQuest effectively provided the information we needed for more than half of the dissertations in the study, which is fantastic!"

In addition to exploring the internationalization of American social science, access to the dissertations metadata has helped the team explore new and innovative avenues for research. As Kurzman commented, "The analysis has opened up many new potential avenues to explore. There is growing interest in social sciences to begin to break down big data, and the tools are there to help this. Access to the dissertations metadata

allows for all sorts of topic modeling to be available and in this respect, ProQuest is in a unique position to help move forward future research in social science.

> **"When we have completed our analysis, we will be producing a report for U.S. social science organizations to help inform social science policy going forward."**
>
> Charles Kurzman, The University of North Carolina at Chapel Hill

There are many circumstances under which an academic would be interested in the historical direction and emerging trends of their fields. For social sciences, analysis of the dissertations data would allow for deeper understanding of what the hot topics were historically and how this might impact interest in those topics, and therefore policy, in the future."

As well as the dissertations data, the team was also given access to the International Bibliography of Social Sciences (IBSS) which formed a big part of the analysis on scholarly journals. "The richness of the data in ProQuest's dissertations and IBSS is really unparalleled. For example, IBSS comprised over half of the number of records we were looking at for journal articles. Plus, the abstracts available in IBSS — which ProQuest creates to add value to the full text of articles — helped greatly in obtaining a more valuable analysis."

Of course, with such rich data at their fingertips it was perhaps inevitable that the team would see new opportunities to further their research. As Kurzman concurred, "It is my intention to follow through different avenues of research although as yet I cannot say what forms this will take — and I will need to check with ProQuest that data could be used for other projects! By way of example, I would like to re-do our current research across the full text of ProQuest's dissertations as this would capture much more information. Access to a centralized database such as ProQuest Dissertations and Theses allows for all sorts of different research projects could be considered. I think in the future, research will begin to look at citation patterns and networks – for example who is citing who and how this plays out over time. If we can process the full text of dissertations and articles, this will help exponentially with this type of research as it will enable the identification and extraction of citations."

But for now, the focus is to complete the analysis of over half a century of social science research and assess the extent to which internationalization of American social science has occurred. "When we have completed our analysis, we will be producing a report for U.S. social science organizations to help inform social science policy going forward. ProQuest has been instrumental in helping us to achieve this. The teams in the U.S. and the U.K. (who provided access to IBSS) were extremely efficient and understanding of the needs of this project. They did what they could to provide seamless access to the data required and have enabled the research to move ahead much more quickly than would have been possible otherwise. I truly believe that by doing this, ProQuest is living up to its responsibility of being the central repository for U.S. dissertations and I look forward to working with them on other projects in the future."