

DEFINING DATA LITERACY: AN EMPIRICAL STUDY OF DATA LITERACY DIMENSION

JEONGHYUN (ANNIE) KIM, LINGZI HONG, AND SARAH EVANS
UNIVERSITY OF NORTH TEXAS, TX, USA

INTRODUCTION

With the advent of big data, data literacy has become an essential skill set for everyone. The importance of data literacy is increasingly recognized, but there remains a lack of agreement regarding the concept and scope of data literacy across disciplines. Further, the constituent dimensions of data literacy remain disputed.

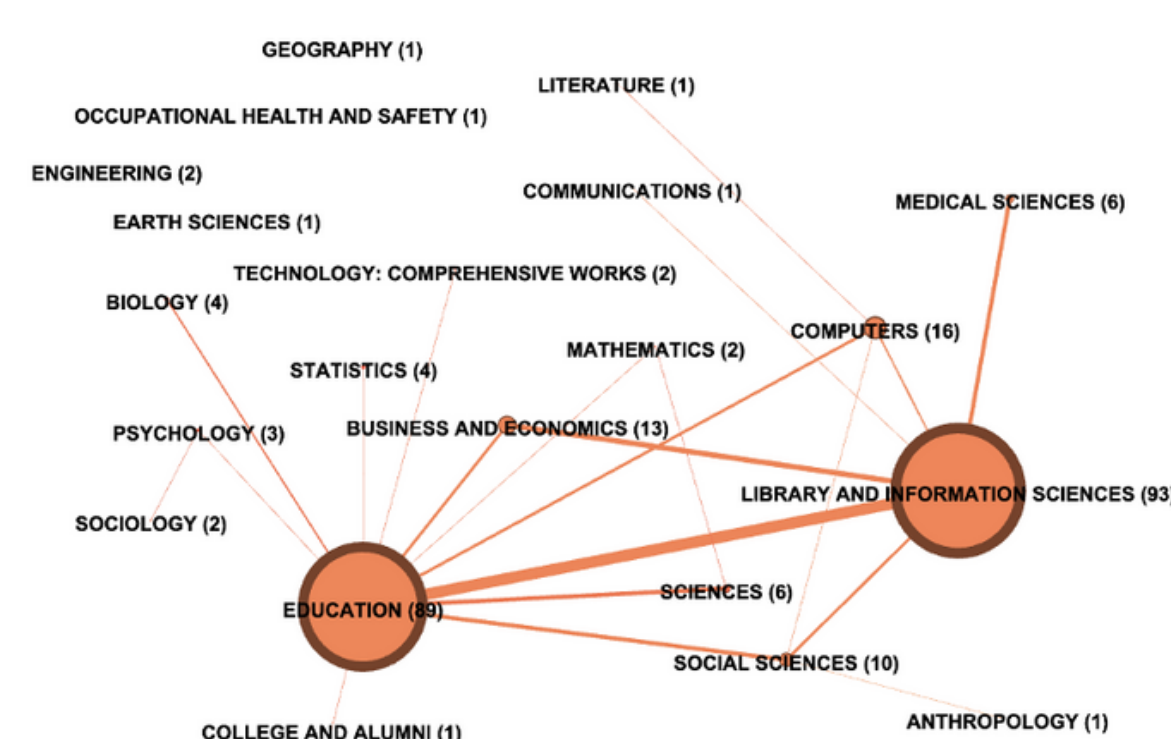
OBJECTIVE

This preliminary work examines the literature published on data literacy and reveals the key aspects of data literacy publication trends. Through the review of scholarly publications, we attempted to capture the emerging topics and trends reflected over 20 years covering 2002-2021.

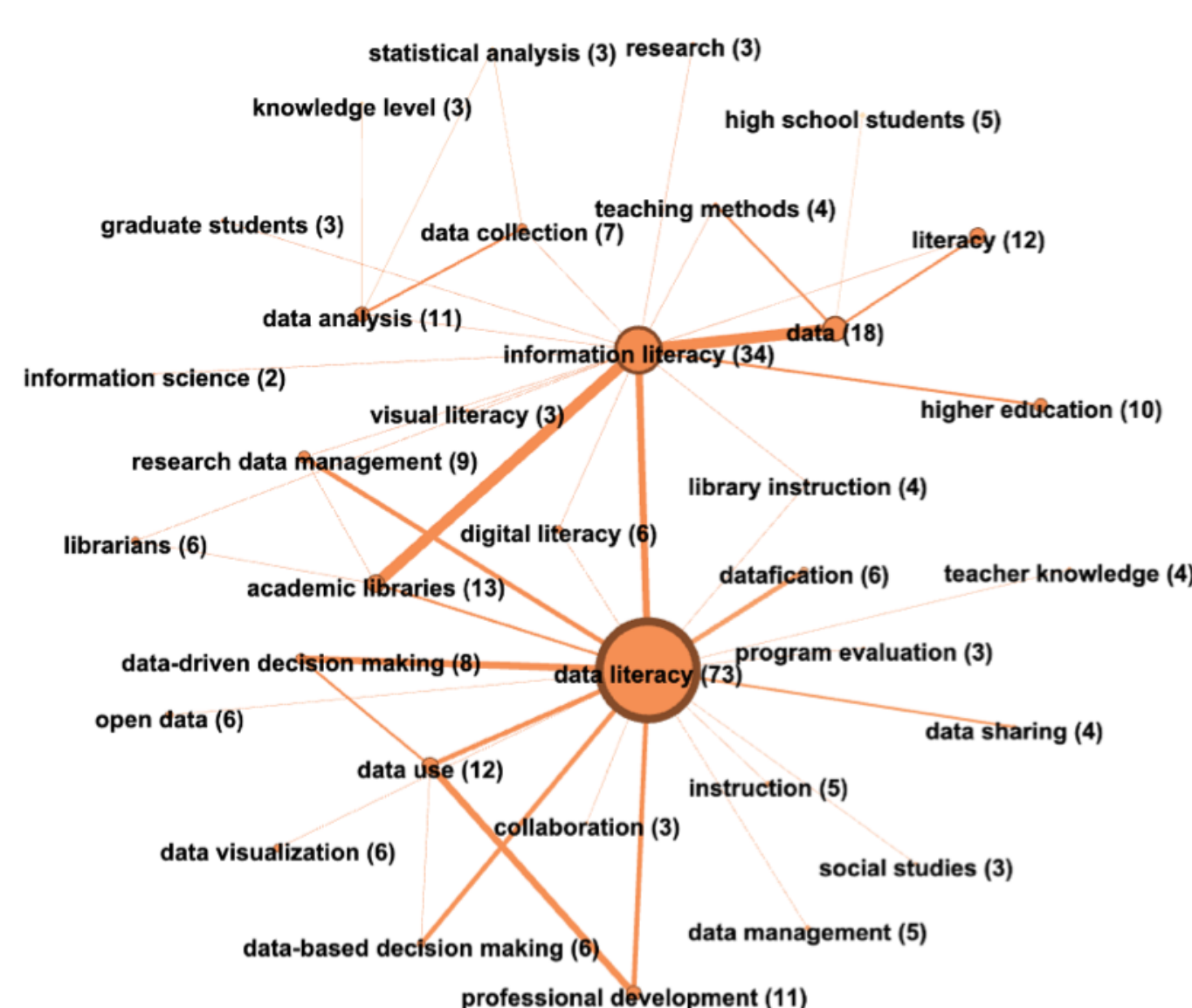
METHODOLOGY

We derived the data from Scopus, ScienceDirect, EbscoHost, and Web of Science. A keyword search was performed to identify literature on the topic to locate publications published from 2002 to 2021 that contained the words and phrases “data literacy,” “data literate,” or “data” AND “literacy” in their title and/or keyword and topic fields. A total of 427 results, including duplicates, were retrieved. Out of those results, 206 were peer-reviewed journal articles. We exported the library from RefWorks as a CSV file for analysis. We analyzed the metadata in each data frame, including publication year, journal title, article title, article abstract, and keywords, to examine trends in publication. We added the discipline of the publication source, which was coded based on Ulrich’s journal subject classification. We applied topic modeling to the article abstracts, a type of text analysis that identifies frequently co-occurring terms and latent topics. We used Gephi to visualize the network of co-occurring terms.

RESULTS



We observed consistent growth in literature on data literacy from 2015 to 2021. Research on data literacy is strongly connected with education, library and information science, computers, and business. Both education and library and information science are principal contributors to published research outcomes.



The most frequent keywords include “data literacy,” “information literacy,” “data,” “academic libraries,” “data use,” “data analysis,” “professional development,” and “higher education.” The keyword network shows the knowledge structure of data literacy scholarly work published in 2002-2021. Although the dominant keyword in the map is always “data literacy,” other keywords linked to “data literacy” are increasingly added in a series of subnetworks.

CONCLUSION

Although many higher education institutions have not yet fully embraced data literacy as a foundational literacy, the scholarly publication patterns we found in this study show a significant research activity in the field of data literacy during the study period and an exceptional growth from 2015 onward.



We identified three distinct domains: 1) Data literacy as part of digital competency; 2) Teaching data literacy in higher education; and 3) Academic libraries’ support of research data literacy. The first domain discusses digital competency assessment frameworks, models, and methods, and data literacy is often included as one of the competency areas. The second domain investigates the impact and effectiveness of a data literacy intervention program, which is often implemented in pre-service teacher education. The last domain focuses on data literacy instruction as part of research data services, which is a campus-wide program that provides academic libraries with the expertise, tools, and infrastructure necessary to manage and steward research data.