Chapter 4
A Study of Big Data Analytical Frameworks in Research Data Management Using Data Mining Techniques

Madhavi Arun Vaidya
https://orcid.org/0000-0001-8356-1935
Vivekanand Education Society's College of Arts, Science, and Commerce, India

Meghana Sanjeeva
Vivekanand Education Society's College of Arts, Science, and Commerce, India

ABSTRACT

Research, which is an integral part of higher education, is undergoing a metamorphosis. Researchers across disciplines are increasingly utilizing electronic tools to collect, analyze, and organize data. This "data deluge" creates a need to develop policies, infrastructures, and services in organisations, with the objective of assisting researchers in creating, collecting, manipulating, analysing, transporting, storing, and preserving datasets. Research is now conducted in the digital realm, with researchers generating and exchanging data among themselves. Research data management in context with library data could also be treated as big data without doubt due its properties of large volume, high velocity, and obvious variety. To sum up, it can be said that big datasets need to be more useful, visible, and accessible. With new and powerful analytics of big data, such as information visualization tools, researchers can look at data in new ways and mine it for information they intend to have.

DOI: 10.4018/978-1-7998-3476-2.ch004

INTRODUCTION

Research which is an integral part of higher education is undergoing a metamorphosis. Researchers across disciplines are increasingly utilizing electronic tools to collect, analyze, and organize data. They are now producing, storing, and disseminating digital data in much larger volumes than the text. Vast quantities of born-digital data are being produced in a wide variety of forms at a rapid rate in universities and research institutes. This "data deluge" creates a need to develop policies, infrastructures and services in organisations, with the objective of assisting researchers in creating, collecting, manipulating, analysing, transporting, storing and preserving datasets for future use. This explosion of born-digital research data (data that are created in digital form) means that the era of BIG DATA has arrived. Along with this digital overload comes the growing need for intelligent and effective Research Data Management (RDM).

The continued existence and access of this data is concern since the data is not currently well organised and stored in libraries. Research is now conducted in the digital realm, with researchers generating and exchanging data among themselves. Sharing research data and scholarship is of national importance due to the increased focus on maximizing return on the government's investment in research programs. Research funders are proactive and encourage good practices and to achieve greater return on investment and value for the sponsored research and therefore they have mandated certain specific standards of data management and sharing to be followed by the researchers.

DEFINITIONS

Research data management (RDM) is about "the organization of data, from its entry to the research cycle through the dissemination and archiving of valuable results" (Whyte and Tedds, 2011).

Big data is an evolving term that describes a large volume of structured, semi-structured and unstructured data that has the potential to be mined for information and used in machine learning projects and other advanced analytics applications.

INTRODUCTION

Research data management concerns the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. Research Data Management is part of the research process, and aims to make the research process as efficient as possible, and meet expectations and requirements of the university, research funders, and legislation. Pinfield,Cox and Smith (2014) mentioned that RDM consists of a number of different activities and processes associated with the data lifecycle, involving the design and creation of data, storage, security, preservation, retrieval, sharing, and reuse, all taking into account technical capabilities, ethical considerations, legal issues and governance frameworks.

Data produced as part of research take a wide range of forms, from statistics and experimental results to interview recordings and transcripts (Borgman, 2012). Data could exist as physical records or files on a researcher's computer or terabytes of data on shared servers. This chapter is divided into various sections viz. section I elaborates on the need of data processing whereas section II, III and IV exemplifies the Literature Review, Need of research data management and Role of Big Data in research data management. Section V describes role of research management in context with Big Data in Library