Chapter 20
Emerging Approaches to Data Management for a New Geospatial Science Research: The Data Management Optimization Perspective

Joyce Gosata Maphanyane
University of Botswana, Botswana

Read Brown Mthanganyika Mapeo
University of Botswana, Botswana

Sethunya Simela
University of Botswana, Botswana

ABSTRACT

This chapter is a continuation of themes discussed in Chapter 19. It draws attention to the newly emerging fields and the growth they injected in geospatial science research procedures. It analytically examines the new fields' role in data management optimization perspectives that emanate from the history of their developments and applications. A robust and rigorous data science methodology framework necessary for the success of a geospatial science research has been submitted, its components and challenges thereof are scrutinized. The overall analyses indicate increased growth in the collaborative efforts and a quantum leap in geospatial science technological development. The superior ICT tools: the Internet, communication networks; high performance computer infrastructure and sophisticated software algorithms: Big Data and cloud computing; increase in accuracy for geo-referencing tools: the GPS and other systems like CORS; and lastly increase in availability of geospatial data including the satellite images and hyperspectral data.

DOI: 10.4018/978-1-5225-3440-2.ch020

SECTION C: EMERGING FIELDS AND DATA MANAGEMENT OPTIMIZATION PERSPECTIVE

Section C is a continuation of the previous sections A and B in Chapter 19. It develops the present status of information communication technologies (ICT) and analytically discussed their impact in changes made in geospatial science research tools. Data quality management and procedures also form a vital part of this chapter.

Emerging Fields in Science and Information Communication Technologies (ICT)

ICT is the underlying drive to the newly emerging fields of science that has a total uplift of the geospatial science research. The ones reviewed include data science and its methodologies, visualization, big data, hyperspectral images, data and its management models, warehouses, metadata, cloud computing, World Wide Web, artificial intelligence, neural and communication networks. What is interesting is that all these technologies can only function in an internet platform – they are digital. The review also includes their advantages and disadvantages when applied to geospatial science research environment. Their highlights and challenges are put forward. An attempt is made by the suggesting of possible remedial measures to give them a better fit for their application in geospatial science endeavours (Table 2).

Data Science

Data Science is a process of utilizing various types of data to gain knowledge about any aspect of the world. It is a new wave of technologies which scientists use by employing statistical programs, advanced analysis, interpretation and exceptional communication skills to transform raw information into actionable insights. The Data Scientists research whole host of questions between data relationships and by putting those through regression analysis to give quantified answers of their magnitude. The quantified findings and their magnitude are then communicated to the decision makers for expert policy or strategy formulation within a certain specific knowledge domain which relies on the questions modelled. The final result is storytelling with data. Its success is dependent upon the initial planning and conceptualization of the final deliverable. This includes the execution and the thinking through of a scheme for all data needed and its analytics. A scientific Data reporting procedure always follows the same method; that is the cover page, abstract, table of contents, executive summary, detailed contents, acknowledgments, references, and appendices.

Data science has developed today because of a number of events. First, there are so much data available: It could be structured data that has being created by machines without human intervention. For example, the Global Positioning System (GPS) data or human-generated data created by humans and input into the computer. It could also be unstructured data like that of the Earth's surface captured by satellite. There are volumes such data available. Second, unlike in the past, where there were very little algorithms; now algorithms are available: These has led to the availability of software at less cost or for free (open source software). Third, today computer infrastructures have become more affordable, effective, and efficient and work at great speeds: large amounts of data could be processed with little cost and at no time. And fourth, possibility to store voluminous data cheaply: the ability to store large amounts of data cheaply in the cloud or within Big Data facilities.