Data literacy for staff in healthcare and medical industry

Boonsit Yimwadsana

Faculty of Information and Communication Technology, Mahidol University

Abstract

Companies today collect a large amount of data thanks to the advancement in information technology and business process improvement. However, the collected data are often not used effectively. This is due to the lack of understanding about the benefits of data. Therefore, there is a world-wide effort led by OECD to push companies to use data more effectively and efficiently in order to improve business operation. Everyone now knows about the benefit of data, however, companies still feel that they are not making the best use of data they have. This problem is even more severe in healthcare and medical institutions since there are an immense amount of a variety of data that flow into the system every second. In this paper, we show that the staff in healthcare institutions can cope with this problem using the concept of research methodology which most staff do not have since it was not usually offered as a required course in undergraduate level. Research methodology covers all areas of data literacy. We identify that research methodology could only be provided through project-based or interactive learning and it cannot be learnt in a short period of time. From our findings, we also found that work environment greatly affected the ability to apply data literacy concept greatly as low-level staff did not have the opportunity to take part in the business problem establishment process. This discourage staff from learning and using data literacy concept. We also propose that companies should allow staff in different departments to have autonomy to conduct their own research and improvement strategy so that they can fully see the need of data literacy and apply it successfully.

Keywords: -

Revised: 10 May 2022, Revise: 25 July 2022, Accepted: 30 September 2022

Correspondence: Boonsit Yimwadsana, Faculty of Information and Communication Technology, Mahidol University, 999 Phutthamonthon Sai 4 Rd, Salaya, Phutthamonthon, Nakhon Pathom 73170, E-mail: boonsit.yim@mahidol.ac.th

Introduction

Companies in various industries are collecting more data than ever since they know the benefits of data once they are analyzed. Hidden knowledge that may not be clear to the executives and staff of the companies could be revealed and made use for great benefits such as how to better solve customer needs, improve their service and operation performance, and make better decisions about almost anything. We have fully entered the information age, all companies can collect and make use of the information effectively and efficiently due to various software tools available for them. The explosion of availability of and demand for information technology have skyrocketed since the introduction of the Internet in the 1990s, the smartphone evolution in the 2010s, and data science in 2020s. Companies try to find way to collect data and make the best use of available data. However, this is not so easy as just throwing data to a program and click a button. While data are available, and they become more influential than ever, there is still a big problem at hand: most executive and operational staff are not very good at interpreting and making sense of the data.

In recent years, there have a lot of suggestions and recommendations at the global level to tackle this problem. The COVID-19 crisis presents an obvious benefit of

using data to understand and tackle the pandemic. Since early 2020, the emergence of popular Covid-19 data sites, such as Worldometer [1], COVID-19 Map of Johns Hopkins Coronavirus Resource Center [2] and the New York Times Covid-19 Pandemic [3], have promoted interest and attention in data-driven tracking of the pandemic. Biologists, biochemists, pharmacists, physicians and researchers use data from respectable data sites such as the National Center for Biotechnology Information (NCBI) [4] to develop drugs and treatments. In addition to official data websites, social media and news outlets have pushed information to the public with trends, discussion, charts and graphs. This pandemic brings everyone in the world together through the sharing and dissemination of medical data, information and knowledge and this is a monumental incident of the widespread acceptance of data culture. In order to be a part of this, everyone must understand how to create and make use of the data. This gives rise to the "data literacy" initiative across the world.

The medical and healthcare industry is obviously the first industry that the public expect to move into the data-oriented business. However, in order to replicate the success of data websites, workers in the industry must not only have adequate domain knowledge in healthcare and medicine but also the knowledge in information technology and possibly specialized data analytics. Most of this knowledge are taught in higher education institutions in courses related to research methodology, information technology and applied statistics. However, these courses are not offered in most undergraduate programs. As workers who are working in healthcare and medical industry have already graduated long time ago and even new graduates do not have this knowledge, it is a challenge for the industry to provide learning resources to help their workers reach sufficient level of "data literacy".

This paper aims to conduct qualitative exploratory pilot study research to provide answers to two fundamental problems: which specific medical data literacy that is suitable for the staff in Thai medical and healthcare industry and how to provide data literacy to them. The paper is structured as follows: the background section discusses prior knowledge and information regarding medical and healthcare data literacy, the methods section discusses how we select sample groups and conduct our study, the results and discussion section discuss our findings and the conclusion section summarizes our findings.

Background

Human society lacks data literate culture since ancient time. Until the introduction of the probability and statistics in the 17th and early 19th century respectively, human societies do not have a formal use of data. Despite the advancement in the fields of probability and statistics, people who can apply the knowledge in the fields were limited to mathematicians and scientific researchers. It was not until the last few decades that engineers and businesses start to use knowledge and techniques in probability and statistics in order to enhance their business operation thanks to the development of computer and information systems including the widespread adoption of the Internet. The world has already anticipated this change since the introduction of the "Information Age" era that we are all in. However, most people still do not fully understand that it means and how we can be a part of it. A major recent development is the reintroduction of a field called data science which combines business domain knowledge, computer science (mainly machine learning and data mining), and statistics together. Several recent successes of companies such as Amazon.com, Agoda.com, Facebook, etc. were supported mainly by data science.

The most recent important incident of Covid-19 pandemic make everyone realizes the power of data. All walks of life including government administrators, policy makers, scientists, healthcare workers, news outlets, and even social media influencers use data to help alleviate the Covid-19 pandemic. Given how data is used in the fight against the Covid-19 pandemic, medical and healthcare industry is expected to be transformed into a data-oriented industry. This requires the workforce in the industry to have knowledge about how to deal with data in the industry.

Data literacy is often defined, conceived of and understood differently depending on the context, business application, and a specific focus on data science which consists of 3 fields: business domain, statistics, and computer science. According to Gartner [5], data literacy is defined as the ability to read, write and communicate data in context,

including an understanding of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case, application and resulting value. Data literacy is one of the key components of digital dexterity, an employee's ability and desire to use existing and emerging technology to drive better business outcomes. Poor data literacy is ranked as the 2nd-biggest internal roadblock to the success of a company, according to the Gartner Annual Chief Data Officer Survey [6]. By 2023, data literacy will become an essential driving business value. This is demonstrated by the widespread adoption of data literacy in global enterprise across the world.

Every decision made is important, especially when it comes to life and death. Institutions in the healthcare and medical industry have to make important decisions every day. In order to make the best decision, only data is insufficient. The capability of the person who makes judgement based on the available data is equally important. In the real-world settings, not only one single person will make a decision, but several people are involved in making this decision. The focus now is therefore how healthcare and medical institutions can roll out data literacy training programs to their employees and staff so that an environment where learning data literacy skills is a part of the organizational culture.

According to OECD [7], data literacy is the ability to derive meaningful information from data, the ability to read, work with, analyze and argue with data, and understand "what data mean, including how to read charts appropriately, draw correct conclusions from data, and recognize when data are being used in misleading or inappropriate ways". Data literacy focuses on both the technical and social aspects of data. It covers a vast body of knowledge related to data such as data management, including data curation, data citation, data quality, data processing, data analysis and data presentation. After the data are cleaned, organized, structured, processed, interpreted, and presented so that they become meaningful or useful, the outputs of these actions are called information which convey a message for the information consumer to use it to improve business objectives. The information can be shared through various methods of communication and often the communication of information is included in the meaning of data literacy.

oday, most employees still do not fully understand the actions that convert data into information, hence they cannot perform these operations effectively and efficiently. The impact of industrial revolution is great, but it has a large impact on the mindset of employees. The concepts of pipeline systems, logistic and supply chain system, and modularized operational process allow great flexibility, efficiency and effectiveness in business operations. However, employees become focus on their own small world and fail to see the big picture of the business. They are unable to create or use the data outside what they are responsible. Only when a company employs data scientists and engineers, different data can be combined together and processed into information that will be useful to the business. However, data scientists can only do so much since they are not the one who operates the business. It is important that employees in key business operations understand and have ability to perform data literacy skills. Companies have to find ways to educate their employees regarding data literacy.

In healthcare and medical industry, the challenge is not only data literacy, but it is also data privacy. The recent introduction of data privacy act around the world is a major obstacle for sharing data between organizations and between organizations. When data are not shared, less people will work on a small amount of data. Since healthcare and medical data are often private data from patients and physicians, this problem is very severe in medical and healthcare industry. This major setback discourages employees from learning data literacy since they do not see the need for it.

Given that they are willing to learn, the learning methods are quite challenging as different employees require different data literacy skills. It has been very difficult to provide education to all employees in an organization. Several training sessions and workshops have to be setup for different groups of employees. Thanks for Covid-19 pandemic, online meetings and learning systems are now popular and most workers are now familiar with it. Combining online learning systems such as Google Classroom [8], Microsoft Teams [9] and Moodle [10] with traditional on-site learning, could be beneficial to the learners.

Data literacy is a part of digital literacy skill which is one of the key bodies of knowledge for the 21st century skills promoted by the Organisation for Economic Co-operation and Development (OECD) [11]. Data literacy helps improve general thinking skills consisting of critical thinking, problem solving, and computational thinking. The ability to ask the right questions and be critical of concepts, claims, and arguments is essential to workers as well as general citizens. With the explosive increase of uncurated data created within an organization and outside the organization, problem solving is no longer the process of making simple logical decisions. It has evolved into a series of complex tasks such as statistics, computing algorithms, and business optimization, dealing with multiple layers and dimensions of data. In some cases, specialized knowledge and information technology and software tools must also be used to support the complex problem-solving tasks.

In order to provide data literacy education for an organization, education theory must be applied appropriately to fit the type of learners. Since data literacy education contains both knowledge and skills, the expected learning outcomes for data literacy spanning across the entire body of knowledge according to [12,13] are as follows:

1. Discuss what data are and the role of data in society.

• Students should be cognizant of data sources and be able to ascertain the validity of data that are presented to them. Students should know the role data plays in decision making at various levels of society.

2. Identify and select data appropriate to an inquiry

• Students should be able to clean, identify and select appropriate data for a given data inquiry.

3. Conduct analysis and interpretation of data in a critical manner.

• Students should be able to select an appropriate analysis technique for the data so that we can allow an appropriate interpretation of that data.

4. Communicate data-based insights.

• Students should be able to communicate in multiple media modes.

5. Differentiate between ethical and unethical use of data.

• Students as consumers and producers of data should be aware of the ethics of how the data they generate are used, ethical dimensions of data they consume, and legal issues surrounding data. An informed citizen of the 21st century should be aware of the ethical uses and misuses of data.

In order to successfully teach data literature to the learners, teachers should have the following teaching skills in data literacy [12]

1. Continuously: Using data as part of daily routines and on an ongoing basis, rather than as a one-time event

2. Effectively: Using data to inform improved and tailored instruction and other practices for the purposes of improving student learning

3. Ethically: Know and apply information with professionalism and integrity for intended uses only, and with consciousness of the need to protect student privacy

4. Access: Know the multiple types of data available (including but not limited to assessment data), understand which data are appropriate to address the question at hand, and know how to access the possible sources of data

5. Interpret: Analyze and synthesize data to make the information appropriate for addressing the given problem or question

6. Act: Take relevant information and apply it to generate further questions and/or apply it to decision making appropriate to the given question

7. Communicate: Share data points and the information synthesized from relevant data with stakeholders including parents, students, peers, principals, and other applicable, to generate further questions, inform decision making, or provide a better understanding of student learning. (p. 6)

Methods

We conducted our qualitative research by collecting interview feedbacks from interviews, consultation and training sessions for staff of several institutions in healthcare and medical business sector. According to Thailand's Personal Data Protection Act BE 2562, the information provided in this research study is only a summary from questions and feedbacks received during the author's consultation and training in order to protect participant's private information. The information collected will be analyzed for determining data literacy capacity building recommendation for healthcare and medical institutions. Additional information from external sources will be included to support the recommendation which will be discussed together with findings from the collected data.

Results and Discussion

According to the feedback collected during interview, consultation, and training, we observe that most workers are responsible for day to day operations at different levels of in healthcare and medical institutions. Most of the data collected were data requested by higher authority such as the Ministry of Public Health or the government. The majority of the staff, especially from the middle management level downward, were not involved in the process of selecting data to be collected. This means that they were not involved in the process of setting up research questions or objectives which are the most important process of data-oriented tasks. Participation or feeling of involvement usually bring motivation. Without the motivation, the workers do not pay attention to the data collection process. They just would like to collect the data to complete their assigned tasks. For a given data definition, staff might understand it differently. Thus, they collected and used the data differently. This affected the quality of data when data are combined from several sources or used in other areas. Different environment also played a key role since data may have a different definition depending on a context. Even though most staff are quite skilled in using information technology, without understanding the data and the purpose of data, the collected data often have poor quality. Work environment played a key role in the understanding of data and the purpose of the data. Since most staff were not involved in the process of establishing research questions or business problems, they would not analyze the data as they did not see the big picture. Anyone would not be able to analyze the data without clear definition and purpose. Most of the time, data were required to be collected from high authority. This makes the data sitting in repositories become useless and just taking up space.

In addition to the motivation, the capability of staff, mostly at the middle management level downward, is quite lacking in several data literacy areas, in particular data analysis, data interpretation and data presentation. These skills required one to understand all three areas of data science including business knowledge, statistics and data science. A hidden knowledge and skill not often discussed is called research methodology encompasses all areas of data literacy. Since most undergraduate education do not offer research methodology as a learning outcome, it is only natural that most staff do not possess data literacy skills.

The research methodology skill can easily be applied to business to become business research which was a very well-known term in the past. In healthcare and medicine, there are a lot of technical terms and specialties which make research methodology very complex. For a large healthcare service provider, communication about business aim is very important as well as the distribution of responsibilities and freedom to departments so that they can act autonomously. This will drive staff in each department to think about the best way to conduct research to improve their service and operation instead of waiting for orders from top-level management which they may not understand.

Knowing that research methodology and scientific methods are important, sub-topics within this area of knowledge include problem solving, data quality, statistics, experimental design, database management system, data communication, and data presentation in order to achieve the learning outcomes outlined in the Background section. Research methodology and scientific method are best taught using project-based approach or interactive lecture that allow students to create their own business problems and reasonable solutions though scientific methods. Healthcare staff can come up with healthcare-related problems within their responsibilities and try to find a way to solve the problem. For instance, a staff in HIV care unit could think about how to reduce HIV infection in the vicinity of the health center. The staff can use data to answer a slightly different problem which is how to make sure that the patients of the unit will regularly take anti-HIV drugs. The staff can collect data such as number of drug request, dates that visits for drug refill, etc. These data can be analyzed so that the behavior of patients can be summarized using statistics and presented to the unit and top-level managers. Only by doing so, the staff will be able to learn from their jobs and increase their data literacy.

Conclusion

In this paper, we introduce the concept of data literacy and the knowledge that staff in healthcare institutions needed. We identify that research methodology and scientific methods skills are needed. The ability to work with the data alone is not as important as to understand why the data have to be collected and used in a certain way and how to do just that. Research methodology covers sub-unit of skills often mentioned in data literacy such as data collection, data understanding, data quality management, data analysis, problem solving, data communication and data presentation By understanding research methodology, the staff would be able to work with data comfortably. However, teaching research methodology is not simple. A person cannot learn how to do things without prior experience. Since research is not something that a person does every day, knowledge in research methodology should therefore be provided with the aim to practically train a person for the skill. Therefore, project-based or interactive learning are necessary to provide data literacy to the staff.

References

[1] Coronavirus Cases – Worldometer (https://www.worldometers.info/coronavirus)

[2] COVID-19 Map – Johns Hopkins Coronavirus Resource Center (https:/coronavirus.jhu.edu/map)

[3] The Covid-19 Pandemic – The New York Times (https://www.nytimes.com/news-event/coronavirus)

[4] SARS-CoV-2 Data, National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/sars-cov-2/)

[5] A Data and Analytics Leader's Guide to Data Literacy, Gartner (https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guide-to-data-literacy)

[6] Gartner 5th Annual CDO Survey: Ways to Achieve Focus and Drive Impact (https://www.gartner.com/en/webinars/3983427/gartner-5th-annual-cdo-survey-ways-to-achieve-focus-and-drive-im)

[7] OECD Future of Education and Skills 2030, Core Foundations for 2030 (https://www.oecd.org/education/2030-project/)

[8] Google Classroom (https://classroom.google.com)

[9] Microsoft Teams (https:/www.microsoft.com/en-us/group-chat-software/)

[10] Moodle (https://moodle.org)

[11] Ananiadou, K., Claro, M., "21st Century Skills and Competences for New Millennium Learners in OECD Countries", OECD Education Working Papers, No. 41, 2009 (https://dx.doi.org/10.1787/218525261154)

[12] Teaching and Assessing Data Literacy: Resource Guide for Supporting Pre-Service and In-Service Teachers, Northern Arizona University, 2020

[13] Data Literacy, Office of Instruction, University of Georgia (https://ovpi.uga.edu/initiatives/data-literacy/)