2019

Exploring Strategies to Integrate Disparate Bioinformatics Datasets

Charbel Bader Fakhry
Walden University

Walden University

College of Management and Technology

This is to certify that the doctoral study by

Charbel Fakhry

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Steven Case, Committee Chairperson, Information Technology Faculty
Dr. Jon McKeeby, Committee Member, Information Technology Faculty
Dr. Jose Feliciano, University Reviewer, Information Technology Faculty

The Office of the Provost

Walden University
2019

Abstract

Exploring Strategies to Integrate Disparate Bioinformatics Datasets

by

Charbel Fakhry

MS SE, Balamand University, 2012

BS BC, Holy Spirit University of Kaslik, 2009

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Information Technology

Walden University

October 2019

Abstract

Distinct bioinformatics datasets make it challenging for bioinformatics specialists to locate the required datasets and unify their format for result extraction. The purpose of this single case study was to explore strategies to integrate distinct bioinformatics datasets. The technology acceptance model was used as the conceptual framework to understand the perceived usefulness and ease of use of integrating bioinformatics datasets. The population of this study included bioinformatics specialists of a research institution in Lebanon that has strategies to integrate distinct bioinformatics datasets. The data collection process included interviews with 6 bioinformatics specialists and reviewing 27 organizational documents relating to integrating bioinformatics datasets. Thematic analysis was used to identify codes and themes related to integrating distinct bioinformatics datasets. Key themes resulting from data analysis included a focus on integrating bioinformatics datasets, adding metadata with the submitted bioinformatics datasets, centralized bioinformatics database, resources, and bioinformatics tools. I showed throughout analyzing the findings of this study that specialists who promote standardizing techniques, adding metadata, and centralization may increase efficiency in integrating distinct bioinformatics datasets. Bioinformaticians, bioinformatics providers, the health care field, and society might benefit from this research. Improvement in bioinformatics affects poistevely the health-care field which has a positive social change. The results of this study might also lead to positive social change in research institutions, such as reduced workload, less frustration, reduction in costs, and increased efficiency while integrating distinct bioinformatics datasets.

Exploring Strategies to Integrate Disparate Bioinformatics Datasets

by

Charbel Fakhry

MS SE, Balamand University, 2012

BS BC, Holy Spirit University of Kaslik, 2009

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Information Technology

Walden University

October 2019

Dedication

I want to dedicate this study to my late mother Antoinette for pushing me to pursue my doctoral degree, for her encouragement, support and for being proud that I am seeking a doctorate without being able to see me graduate and to my daughter Anthonia for being an inspiration.

Acknowledgments

I acknowledge and thank the following persons for helping me achieve my doctoral degree:

Dr. Steven V. Case for his unconditional support, patience, encouragement, guidance throughout my doctoral journey, and for making the completion of this degree possible.

Dr. Jon McKeeby, for his valuable comments that helped me improve the quality of my study and achieve my goals.

Dr. Jose Feliciano and Dr. Bob Duhaini for making my study more complete.

Walden University, the chief academic officier Dr. Jan Garfield, the professors and the staff for making this journey possible.

Dr. Joumana Dargham for her continuous support, motivation, and for sharing with me her wealth of knowledge unconditionally over the years.

Dr. Pierre Khoueiry, for his support, motivation, encouragement, and knowledge sharing that helped gain a more profound knowledge of my studied topic and achieve my goals.

The American University of Beirut (AUB) for opening the doors to me to finalize my study and interview specialists.

My Father Bader Fakhry for being by my side, helping me in many ways, and supporting me emotionally and financially.

My wife and daughter, for their support, understanding, patience, and missing time together.

Table of Contents

List of Tables

List of Figures

Section 1: Foundation of the Study

Background of the Problem

Computation helps refine biological data for pattern and knowledge extraction. It has been an essential part of structural biology since its early days, and the purpose of computation has significantly increased over the years (Samish, Bourne, & Najmanovich, 2015). Bioinformatics is an intersection between computer science, physics, chemistry, mathematics, statistics, engineering, and molecular biology used for analyzing biological data to develop algorithms and relations among different biological systems (Samish et al., 2015). In the early 1970s, bioinformatics began to grow significantly, introducing the ability to digitatize biological output and use computational power to analyze massive datasets (Marco-Ramell et al., 2018). Prior the advancement of the bioinformatics fields, biologists did not have access to a significant amount of data. Developing analytical methods for interpreting biological information were possible, yet it was challenging to share them quickly with other researchers. The technological advancement in computation, storage, and bandwidth has revolutionized the biological field, making it easier and cheaper to analyze biological information (Triplet & Butler, 2014). The technological advancement in computational power over the years has reduced the bottleneck of the costs of doing experiments in biological discovery (Miller Zhu & Bromberg, 2017).

Problem Statement

The bioinformatics fields lacks collaboration due to limited accessibility and availability of bioinformatics datasets (Machiela & Chanock, 2015). Web services retrieving information from a different location using different formats is inefficient and time-consuming (Machiela & Chanock, 2015). The need for additional integration datasets is due in part to the exponential growth of the datasets. For example, the 1000 Genome Project consists of an estimated 100 terabytes of data, and the follow on 1000 Genome Project consists of an estimated 10 petabytes of data (Merelli, Pérez-Sánchez, Gesing, & D'Agostino, 2014). The general information technology (IT) problem is that bioinformatics laboratories lack the means to analyze bioinformatics datasets using web services. The specific IT problem is that some bioinformatics specialists lack strategies to integrate distinct bioinformatics datasets.

Purpose Statement

The purpose of this qualitative case study was to explore the strategies used by bioinformatics specialists to integrate distinct bioinformatics datasets. The population of this study included bioinformatics specialists of a research institution in Lebanon that has strategies to integrate distinct bioinformatics datasets. The findings from the study may contribute to IT practice by identifying strategies to unify and integrate heterogeneous biological information from different locations and different structures. The study's findings may contribute to positive social change by positively impacting healthcare as a side effect of improvements to the bioinformatics.

Nature of the Study

I used a qualitative research method for this study. A qualitative study is conducted when the concept is immature due to lack of theory and previous research (Kahlke, 2014). The qualitative method was appropriate for this study because the strategies used to integrate bioinformatics datasets are limited. In a quantitative study, the problem is addressed by understanding what variables influence the outcome to confirm or disconfirm theoretical hypotheses (Fassinger & Morrow, 2013). The quantitative method is not appropriate for this study, as the research question was not used to confirm or disconfirm a hypothesis. The mixed method includes both qualitative and quantitative methods; it involves the process of collecting, analyzing and integrating quantitative and qualitative research designs (Noprisson et al., 2016). The mixed method is not appropriate for this study because this study did not apply quantitative methods.

I used a qualitative case study approach, which is used to relate patterns (Kahlke, 2014). Employing the case study design for this study allowed me to identify and relate patterns in the strategies used by bioinformatics specialists to integrate disparate datasets. The narrative design is appropriate when a study focuses on a specific person (Wiles, Crow & Pain, 2011); therefore, the narrative design was not selected because I focused on the strategies of bioinformatics specialists rather than strategies of one person. The phenomenological approach describes the common meaning for individuals of their lived experience of a phenomenon; a phenomenological approach is interested in the individual experiences of peoples throughout qualitative methods such as interviewing (Bevan, 2014). The phenomenological design was not selected because I focused on the strategies used in the case's organization rather than focus on the experience of the participants. In the ethnographic design, researchers describe the patterns, values, beliefs of a culture-sharing group (Wilson & Chaddha, 2009), which I did not choose because that was not the focus of the study.

Qualitative Research Question

What are strategies used by bioinformatics specialists to integrate disparate bioinformatics datasets?

Interview Questions

- What strategies do you use to integrate disparate bioinformatics datasets? Please explain.
- Have integrating disparate bioinformatics datasets helped you in creating a unified view for more efficiency in identyfing patterns? Please elaborate.
- What negative aspects of integrating disparate bioinformatics datasets have you experienced? Please explain.
- What are the complexities that specialists might face while integrating disparate biological datasets? Please explain.
- What are the challenges and difficulties that face you while retrieving information from different locations stored in different formats? Please elaborate.
- What strategies do you have for bioinformatics specialists to analyze biological data more efficiently? Please elaborate.

Conceptual Framework

For my bioinformatics study, I used the technology acceptance model (TAM) as a conceptual framework. The TAM was developed by F. D. Davis in 1986, and it is the prediction of user acceptance towards information systems (Davis, Bagozzi & Warshaw, 1989). The TAM approach addresses the relationship between the ease of use, usefulness, and user acceptance of technology (Davis et al., 1989). Identifying and measuring stimuli could predict the acceptance or rejections of the system; the two core beliefs that form the TAM are:

1. Perceived usefulness: a user's subjective probability that using a specific system/technology will increase his or her job. The user's attitude toward whether the new technology will help perform better affect the decision of adapting this technology.
2. Perceived ease of use: the degree to which a user expects the use of a system/technology to be free of effort. The system ease of use must outweigh the effort required in adopting it, and it is affected by three factors: the physical effort, mental effort, and the direct perception of how easy the system is to use. (Davis et al., 1989)

Social norms do not directly affect behavior or attitude in relation to system use; instead attitudes toward using a system is the function of perceived usefulness and perceived ease of use. There are six variables in the TAM that affect the user's acceptance toward a new technology. Three variables are latent that include perceived usefulness, perceived ease of use and attitude toward using, and there are three measured variables that include external variables, usage frequency and usage volume (Davis et al., 1989).

I utilized the TAM in my study to gain understanding of how perceptions of usefulness and ease of use affect the strategies used by bioinformatics specialists to integrate distinct bioinformatics datasets. The two core variables of the TAM, the perceived usefulness and the perceived ease of use, may contribute to the bioinformatics specialists strategies and affect the efficiency of bioinformatics specialists to extract knowledge from datasets.

Operational Definitions

Bioinformatics: Bioinformatics is an intersection of many disciplines to develop systems that analyze biological information for the experts to gain an understanding of the processed data (Samish et al., 2015).

Bioinformatics specialist: A bioinformatics specialist analyzes and studies the enormous amount of biological data using computers.

Data integration: Data integration is the action of merging disparate data into useful information (Lopes & Oliveira, 2015).

Data standardization: Data standardization is the process of storing disparate data in a standard format to improve the quality and consistency of the data to ensure efficiency in searching and analyzing the information (Micic et al., 2017).

High-performance computing: High-performance computing is the use of high computation capabilities and parallel processing techniques in computing (Miller et al., 2017).

Integrative iioinformatics: Integrative bioinformatics targets the issue of data integration in the life science field (Shah et al., 2005).

Next generation sequencing (NGS): NGS sequencing enables researchers to study a genome at a high level (Ma, Gong, & Jiang, 2017).

Assumptions, Limitations, and Delimitations

Assumptions

Assumptions are facts and expectations assumed to be true by researchers without evidence to support them (Kirkwood & Price, 2013). The following assumptions shaped the study. The first assumption is that the participants in the research are knowledgeable about techniques of retrieving, integrating and analyzing biological information. The second assumption is that the participants in this study provided trusted and honest information, and they shared their knowledge and expertise with transparency. The third assumption is that I understood and analyzed the responses of the participants and reliably answered the research questions.

Delimitations

Delimitations are the boundaries that influence the study (Marshall & Rossman, 2016). A delimitation of this study was the involvement of participants who are actively working in the bioinformatics field and have an acceptable experience. The second delimitation was the interview questions, which were limited to strategies to integrate distinct bioinformatics datasets. The third delimitation was the small sample size; a larger sample can take more time, effort, and resources. The fourth delimitation is that I considered a facility that is processing biological information using bioinformatics tools.

Limitations

Limitations are weaknesses that can affect the reliability and the findings of the study (Leedy & Ormord, 2013). The small population of the participants may have limited the research findings. A broader population sample might have given a more generalized finding. The strategies identified from the study to improve the quality and reliability of bioinformatics pattern extraction may not apply to all bioinformatics laboratories. Additionally, my ability to extract information from participants may have affected how the research questions were answered.

Significance of the Study

It is important to have biological data that is more available, accessible, and easy to analyze for pattern extraction. Having a web-based bioinformatics tool that stores associated data in a standardized manner is signficant for bioinformatics specialists. By providing strategies to eliminate the operation of formatting the data before analysis, the results of this study may provide bioinformatics experts the ability to share knowledge and experience anytime and anywhere. The results of the study may provide bioinformatics specialist with the capacity to access information retrieved from different locations and stored in the cloud in a standardized manner and share knowledge and expertise. This study also addresses the gap between biologists and computer science specialists because an open source web-based bioinformatics application can make the bioinformatics experts more involved in the development of an efficient web application.

The study’s findings may contribute to positive social change by enabling biologists and other professionals in the bioinformatics field to more efficiently analyze large, computationally complex datasets. Making use of computational power in the biological fields has had a positive effect (Claverie & Notredame, 2013). By enabling more efficient analysis of computationally complex datasets, the study’s findings may contribute to improvements in health care such as improving cancer diagnostics. Further, the bioinformatics field has developed methods and tools for understanding biological data; this acquired knowledge gives specialists a deeper understanding of patterns and evolutionary biology that affects the biological and health care fields positively.

A Review of the Professional and Academic Literature

The purpose of the qualitative case study was to explore strategies used by bioinformatics specialists to integrate distinct bioinformatics datasets. Bioinformatics has been utilized for a long time to extract useful information and patterns without being properly formatted and accessible from different locations. The bioinformatics field still lacks collaboration, accessibility, and availability of a bioinformatics specialists from various places.

This literature review contains 103 articles and journals. Ninety-one percent of the articles are peer-reviewed, and 86% of the artilces are published in the last 5 years. The research libraries where I searched for references included the ACM Digital Library, Sciences Direct, IEEE Xplore digital library, EBSCOhost Computers and Applied Sciences Complete, and ProQuest. I used Academic Search Complete throughout Walden Library, which is a multidisciplinary database containing peer-reviewed resources. I also used Google Scholar. I used Ulrich's Global Serials Directory to identify articles as peer-reviewed or not.

In this literature review, I outline the characteristics of bioinformatics to develop a conceptual framework for increasing the availability, accessibility, and pattern extraction. This literature review provides information regarding bioinformatics and data integration, integrating bioinformatics datasets, result analysis in bioinformatics, text mining, bioinformatics challenges, next generation sequencing (NGS), evolution of the TAM, and bioinformatics study. In this review, I also explore the strategies used by bioinformatics specialists to integrate distinct bioinformatics datasets. Additionally, I discuss the TAM and how it aligns with integrating bioinformatics datasets.

Bioinformatics Tools

The technological advancement in life science has resulted in collecting enormous amounts of data, and sequencing has become cheaper and faster (Heather & Chain, 2016; Triplet & Butler, 2014). For example, in the past it took 13 years and 2.7 billion dollars to sequence a human genome, and today sequencing a human genome takes a day and $1,000 US (Miller et al., 2017). With sequencing costs going lower, there is an unprecedented amount of data to significantly improve current models and tools (Miller et al., 2017). The exponential growth of biological data over the years has required the intervention of the information technology to store this data and process it for analytical purposes. For instance, in the past two decades, NGS and other high-throughput approaches have led to an explosion of data (Berger, Daniels, & Yu, 2016). The advancement in biological and biomedical fields has also led to increasing amounts of data from genomics and translational research (Zhao et al., 2015). The growth of biological information in sequences such as DNA, RNA, and protein require databases to store, manage, and retrieve the information (Zou et al., 2015). Bioinformatics pattern extraction from enormous amounts of data requires powerful software tools, which involve challenges to get those tools installed, running and produce results (Velloso, Vialle, & Ortega, 2015).

The increasing volume and complexity of biological information demand a workflow involving the multitude of steps for obtaining knowledge (Dubchak, et al., 2014). The increase of high-throughput omics experiments has led to better availability of biomedical data that need to be analyzed (Chen, Tripathi, & Mizuguchi, 2016). Integrating biological information from different sources helps detect biological signals and lower false discovery rates due to the evidence from multiple domains (Zhaoet al., 2015). However, the amount and complexity of biological information that is increasing make the process of extracting meaningful biological insights challenging (Chen et al., 2014). The increasing volume and complexity of genomic data make the process of analyzing variants challenging for researchers with limited bioinformatics skills (Alexander et al., 2017). Therefore, bioinformatics tools reduce the complexity of the high-throughput complex biological data, enhance the overall understanding of biological systems, and help to generate hypothesis (Emery & Morgan, 2017; Marco-Ramell et al., 2018). Centralized bioinformatics tools can eliminate complex installation procedures and high processing power. Applying bioinformatics and integrational methodologies to the increasing amounts of biomedical data enhance the process of generating knowledge (Bhuvaneshwar et al., 2016).

Bioinformatics software tools often require complex installation procedures and high processing power (Velloso, Vialle, & Ortega, 2015). Traditionally high processing computing was expensive and rarely reached the targeted use (Miller et al., 2017). However, advancement in high-performance computing is having a positive impact on bioinformatics tools (Miller et al., 2017). Advancement of high-performance computing aligned with new cloud computing solutions has created a new scope for the applications in today's science (Miller et al., 2017). High-performance computers with clustered processors, high internal bandwidth, and cutting-edge software are needed (Hong et al. 2013).

Data analysis is one of the most critical and challenging steps in the biomedical field, and specialists need tools for efficient pattern discovery. A user-friendly resource to visualize and analyze high-throughput data is a powerful medium for specialists to obtain meaningful output for better knowledge discovery (Chen et al., 2016). Bioinformatics tools using a friendly user interface can be beneficial and help analyze biological data and extract useful knowledge (Velloso et al., 2015). A user-friendly tool will help users with no programming skills to discover new hypotheses and patterns by performing complicated searchers to obtain the results in an easy to comprehend output format.

Many bioinformatics tools have emerged to deal with the increased volume and complexity of biological data. Bioinformatics tools interpret effectively and timely information from genomes. For example, pair wise search methods are used to detect distantly related homologues, and silico cloning is a recent low cost, high efficacy and easy operation method that is convenient for cloning novel gene (Bozgo, Hysi, & Hoda, 2017). A widely used methodology is functional enrichment or over-representation analaysis, which is used for performing analytic techniques that benefits from molecular pathway or network information to gain insights into a biological system by looking for descriptors in the sets of molecules of interest (Alexander et al., 2017). Additionally, Cluster-Blast tool is an automated approach identyfing related gene clusters (Soria-

Guerra et al., 2015). BLAST is tool at NCBI used to search homologues against the swiss prot protein database (Bozgo et al., 2017). Anothet tool is phylogeny, which is a web-based bioinformatics tool that allows the user to link multiple sequence alignment, tree building, and tree rendering to construct with high accuracy and rapidly phylagenic tree (Soria-Guerra et al., 2015). Further, quality control tools such as Qualimap are used for highliting the problems in the data and also integrating tools are also used as the field of NGS mature, and also a main goal of bioinformatics tools is to provide powerful visualization with a simple interface (Ewels et al, 2016). With the increasing number of bioinformatics tools, experts are eager for standardization for more efficiency (Lopes & Oliveira, 2015).

Bioinformatics History

Bioinformatics started with Gregor J. Mendel, whose work cross-fertilizing different colors of the same species of flowers (Thampi, 2009) helped establish the theoretical basis of today's biology (Schwarzbach et al., 2014). Mendel's discovery of the general law of unit character transmission across generations through reproductive cells containing unit factors is also considered the origin of genetics (Zhang, Chen & Sun, 2017). Mendel's seminal work set a foundation for the discipline of genetics (Hoßfeld et al., 2017); therefore, he can be considered as a founder of bioinformatics.

The understanding of genetics has advanced in the past 30 years, which started with Paul Berg on his work on recombinant DNA that led to him receiving a Nobel prize in 1980 (Berg, 2008; Dellureficia, 2015). Berg was involved in the making of the first recombinant DNA molecule (Thampi, 2009). Paul Berg, Herbert Boyer, Annie Chang and Stanely Cohen generated the first recombinant DNA in 1972 (Khan et al., 2016). But Berg's most famous work was gene splicing of recombinant DNA where he was the first to insert DNA from two different species into a molecule (Berg, 2008). Berg's research and contributions in the gene splicing techniques set a step toward the development of modern genetic engineering.

The work of Berg on recombinant DNA led to researchers like Boyer and Cohen advancing genetics. After Berg's 1971 gene splicing experiment, in 1973 Boyer and Cohen inserted recombinant DNA into bacteria so the foreign DNA can replicate (Russo, 2003). This provided evidence that DNA molecules can be cloned in foreign cells (Russo, 2003; Genetics and Genomics Timeline, 2004). This showed the possibility of transferring one genetic organism to another (Niosi, 2017). Thus, Boyer and Cohen's work had a considerable effect on the development of modern biotechnology.

The advancement of biotechnology allowed the emergence of biotechnology products. For example, Boyer and Itakura expressed a mammalian protein in bacteria before they constructed a plasmid that coded for human insulin in 1978 (Russo, 2003). These discoveries led to the development of Humulin, the first product of modern biotechnology (Niosi, 2017). In 1977, a method for sequencing data was invented, and Genentech, the first genetic engineering company, was founded (Thampi, 2009). The firm created recombinant human protein in 1977 and the second one in 1980, which was Humulin (Niosi, 2017). The ability to create biotechnology products such as Humulin has allowed patients with diabetes to live a better life.

In addition to these advancements, producing a map of the human genome has helped gain a deeper understanding of all the genes of human beings.The Human Genome Project is an important research project focused on mapping and identifying the genes of the human genome.The international organization of scientists involved in the Human Genome Project was founded in 1989 (Thampi, 2009). To protect the genome from mutagenics, the U.S. Department of Energy established an early genome project in 1987, and the National Institute of Health and Department of Education with the Congress funded coordinated research and technical activities related to the human genome in 1988 (“An Overview of the Human Genome Project,” 2016). Genethon, a human research center, produced a physical map of the human genome in 1993 (Thampi, 2009). In the year 2000, most of the human genome had been sequenced (“An Overview of the Human Genome Project,” 2016).

In response to projects like the Human Genome Project, automated DNA sequencing has replaced manual sequencing for more efficiency and greater ability to extract patterns. Manually sequencing DNA was made in the early 70s, but the following decades brought technologies like high-performance computers that enabled the automated sequencing of DNA and whole genome (Cyrus et al., 2015; Remmers & Siegel, 2015). Sequencing a whole human genome would not be possible without automated sequencing and without advancement in high-performance computing, storage, and bandwidth, which bioinformatics encompasses.

Following the evolution of sequencing DNA and with the technological advancement, bioinformatics tools have emerged to help predict the function of genes and extract meaningful information from biological data through analysis. Bioinformatics has created huge databases like GenBank, European Nucleotide Archive, and the DNA Data Bank of Japan. Today bioinformatics is involved in protein structure analysis and gene and protein functional information. Bioinformatics facilitates information sharing, knowledge management, and workflow tools. With IT, bioinformatics can digitize biological output and provide computational power to analyze massive datasets. Bioinformatics tools are efficient in synthetizing, analyzing, and extracting large volume of genetic information (Alansari et al., 2017). For example, bioinformatics can be used for obstructing signatures of disease, predicting diseases, proposing medicine, and figuring out disease mechanisms (Nussinov & Papin, 2016). Bioinformatics provides the data analysis tools to relate patterns and extract valuable information. Biological and biomedical questions are answered by computational biologists using computation in support of, or in place of, laboratory procedures to obtain answers at a reduced cost.

In bioinformatics, it is important for specialists to share knowledge and results as well as have work with technology to increase efficiency and discover knowledge (Hassani-Pak & Rawlings, 2017; Hoff & Bashir, 2015). Lack of collaboration in this area is weak due to limited accessibility and availability in bioinformatics datasets (Machiela & Chanock, 2015). Integrating heterogeneous bioinformatics data is challenging, yet it enables collaboration for bioinformatics specialists to benefit from the increasing biological information and the number of data types. Many challenges affect specialists' performance like the poor quality of the generated data, the sample size, false discovery, lack of novel algorithms for data integration, computational efficiency, data interpretation, and visualization (Zhao et al., 2015). Integrating biomedical data incomplete and unified format with sufficient metadata can help bioinformatics specialists improve efficiency in this field. Integration involves cross-referencing each dataset submitted to a bioinformatics database to the representation of the same or related biological entities in other databases.

Big Data

The advancement in high-throughput technologies has allowed biologists and other life science specialists to generate enormous amounts of data like genomic sequences. The continuous accumulation of increasing data has introduced new challenges on storing and analyzing this data and new techniques that involve big data that are used for knowledge extraction. Big data deals with enormous and complex datasets that traditional techniques cannot. Identifying patterns in large datasets through integrative analysis and fulfilling the commitment of big data in biology is necessary for solving biomedical problems (Greene et al., 2014). Biomedical informatics is one of the most active areas involved in big data analysis research (Chrimes & Zamani, 2017). However, mining data in the health informatics industry is complicated due to the size and nature of the data, which is why big data techniques are applied.

The significant increase of data in many fields over the past few years has made the experts turn to big data to contain this data. The big data concept was launched in 2000, but it did not get its popularity until 2010 when it was adopted by International Business Machines and Oracle (Chrimes & Zamani, 2017). Big data is targeted to solve problems that cannot be resolved using traditional techniques. Big data, particularly

NoSQL, were introduced to address the increased accumulation of data in traditional relational databases (Storey & Song, 2017). Big data became popular to store and analyze the vast and heterogeneous data to acquire valuable information. Big data started to be used by large companies to understand their businesses and support their decision making.

The V's of the big data define its structure and outcomes. The first "V" is the volume of data—usually terabytes of datasets; the second is the variety and represents the many formats of the big data; the third is the velocity it the speed of data processing; the fourth is the veracity and represents the uncertainty of the data; and the value of the data represents the worth of data being extracted (Hamilton & Kreuzer, 2018; Sharma, Panwar & Sugandh, 2018; Yao, 2017). The first four V's are concerned with collecting data, preprocessing, and transmission and storage. The fifth and last "V" is the process of extracting value from the data using analytical tools. The advancement in technology generated an unprecedented volume, velocity, and variety of data that is called big data (Ting et al., 2017). Big data is characterized by the high volume, high velocity, and high variety of information and extended to include two more V's the veracity and volume (Ang & Seng, 2016). Social media, mobile transaction, business transaction, and network sensors are generating thousands of heterogeneous information datasets that requires big data technology to be collected, stored, and analyzed (Kitchin & McArdle, 2016). Using big data specialists in various fields can solve problems that cannot be resolved using traditional methods.

Having enormous amounts of data in the biomedical field introduces many challenges from choosing the appropriate tools and algorithms, ensuring usability, and transforming data into knowledge. Big data challenges in the biomedical field include selecting appropriate computational methods to extract useful information, access it and share it efficiently (Chrimes & Zamani, 2017). Reusability in big data is paramount because data are being stored in significant amounts without being reused, reusing data for an analytical purpose has a significant positive impact on data-oriented fields (Chrimes & Zamani, 2017). Big data analytics require advanced technologies to deal with vast quantities of more massive, unstructured and complex datasets (Ting et al., 2017). Even though big data offer many solutions to deal with complicated huge amounts of data, the biomedical field struggle to reach efficiency while managing and analyzing biological information. To reduce challenges cutting-edge techniques and algorithms must be implemented to help specialists gain insights and acquire needed information.

With the technological advancement in the biological field, enormous amounts of data were generated. This massive collection of data introduced challenges in storing and analyzing this data. Big data can be useful in the bioinformatics field because the use of big data showed efficiency in many areas where big data was used to solve challenges that cannot be solved using traditional techniques.

Data Quality

Data quality is paramount for organizations in all fields; it can help specialists analyze this data to acquire knowledge. Poor data quality can be misleading and can cause the firms tangible and intangible expenses. Poor data quality can cost firms high fees, can affect less tangible areas like decision quality and job satisfaction, and can affect the overall organization revenues (Hazen et al, 2014). Querying poor quality data can result in a loss of relevant information that can be helpful in many aspects (Hu et al., 2017; Veiga et al., 2017). Good data quality helps to enhance businesses and revenues throughout obtaining high-quality insights from analyzing this data. Analyzing and managing data that contain duplications and errors lead to many problems that affect organizations negatively.

Poor data quality is an obstacle to having a more efficient analytic based strategy. Executives and managers consider that having a high-quality data is a must to overcome many problems that poor data quality can cause (Hazen et al, 2014). Good data quality is paramount in finding, interpreting, and reusing data for practical anticipations. (Hu et al., 2017). High-quality data analysis provides executives with a clear view of their organization status, better expectations of revenues and also helps the executives to support their decision making with evidences form analyzed data. It is evident that any organization including the healthcare industry that correcting data is for the best of the organization.

To ensure that the data is of good quality, it must be accurate, not contain errors, and represent their real values. Data quality dimension include accuracy, timeliness, consistency, and completeness (Hazen et al, 2014). Redundancy, heterogeneity, inconsistency, and incompleteness are data quality issues (Hu et al., 2017). Also ensuring a good quality of data is done by eliminating redundancy. One of the significant problems that face data quality records duplication (Jones et al., 2017). Analyzing good quality data, helps organizations in many fields to enhance their businesses and reduce expenses. Cleaning raw datasets is becoming a priority for most experts to acquire knowledge without getting incorrect results from poor quality data.

Integrational Bioinformatics

In the biological and biomedical field, researchers witnessed advancement in the NGS and single cell technologies, enabling investigators to create massive amounts of data for genomics and translational research. These advancements made the process of data integration from multiple resources easier to enhance knowledge extraction and reduce false discovery rate (Dubchak et al., 2014). Integrative approaches include many advantages like low false discovery rates due to the evidence from multiple domains (Zhao et al., 2015). Researchers that are exploring bioinformatics patterns that contains faulty results affects the quality of decision making. The integration services and tools provided by multiple groups is essential for comprehensive data analysis (Dubchak et al., 2014). Having more than one source of knowledge enhances the accuracy of pattern and knowledge extraction from biological datasets, this means that if a researcher has access to multiple data sources he can have more evidences to support his decision. Extracting knowledge from multiple resources introduce new patterns that is hard to find analyzing one source of data.

Traditional integration solutions like data warehouse are considered a local integration solution. These solutions can enhance the resource sharing and collaboration inside an organization; however, these solutions fail to interact with each other to achieve global solutions, which are crucial for interdisciplinary integration (Chen, Tripathi &

Mizuguchi, 2016). Connecting various datasets in the bioinformatics field can have a positive impact, by making more valuable information available for specialists to analyze. Because connecting data can impact many fields, the large genomic data linked to phenotype and medical records can enable not only the discovery of biological features and regulations using genomic approaches, but also translate some of the findings for clinical practice (Zhao et al., 2015). Integrating biological data is paramount not only for the bioinformatics field but also to many life science areas.

Integrating bioinformatics datasets from various sources is a paramount topic in the bioinformatics field. Data integration has become common for life science in the past years (Lopes & Oliveria, 2015). Numerous efforts have been conducted to develop different types of frameworks and tools for integrating diverse biological data types, like grouping genes based on similarities in the biological annotations, providing pre-defined gene libraries to enhance analysis and using a standalone web interface not integrated to a data mining platform (Chen, Tripathi, & Mizuguchi, 2016). The biggest challenge in computational biology is putting together the available and disparate information (Nussinov & Papin, 2016), because of the heterogeneous nature of biological data, unifying the data types and the nature of collected data can be challenging. The biological data have grown exponentially, and these data are scattered over a number of repositories using various formats, which makes the process of analyzing data for knowledge extraction challenging.

Accessing heterogeneous datasets from disparate sources is a challenge for bioinformatics specialists. Processing data from different sources is a common task for scientists (Lopes & Oliveria, 2015). The quantity of biological information is increasing at a rapid pace, which introduces new challenges in software and hardware from disparate locations. The biomedical information is increasing alongside the number of data types, which is increasing the complexity of extracting knowledge and thus affecting the availability of the distributed data. It is essential to focus on data integration for more reliability in acquiring knowledge. Data integration is needed in bioinformatics with the increased amount of biological data to enable bioinformatics specialists to extract knowledge (Shah et al., 2005). Integrating information from disparate resources is paramount for specialists to have a more precise results from their analysis.

Biomedical database integration is classified into three main classes: federated, mediated, and warehouse style integration; federated integration, provides hyperlinks to join data; and mediated integration, provides unified query interface as well as collecting the results from various data sources (Ethier et al., 2015). Warehouse databases integrate data sources in one place (Ethier et al., 2015). The three main classes of integration in biomedical databases are approaches to integrate and unify data for more efficiency while extracting patterns. In past years, attention was given to integration in the bioinformatics field. Integrating biological information from multiple resources is efficient for more values while extracting knowledge. Integration of services and tools provided by multiple groups is essential for comprehensive data analysis (Dubchak et al., 2014). Gathering information from multiple sources using integration decreases the error rate while analyzing biological data (Hong et al., 2013). Because having multiple sources of data, for example different genomic analysis from two different locations, the specialist can have more confidence in his findings. Disparate datasets enrich the process of analyzing data in bioinformatics for more knowledge. Data integration and analyzing information from different resources has been helpful in finding new patterns and knowledge that were hidden before merging multiple information resources. Mining information, which is the process of discovering patterns in large datasets across different biological databases has the potential to lead to new knowledge (Hassani-Pak & Rawlings, 2017). Data integration has proved viable in the field of bioinformatics because more information is more knowledge and more efficient knowledge extraction.

Integrating biological data and integrating tools and services have become important. Integrating bioinformatics services and tools reduce time and effort for analyzing translational data and increases the efficiency of knowledge extraction (Dubchak et al., 2014). Also, these approaches enhance the data quality that is scattered on different technology platforms (Zhao et al., 2015). Integrating bioinformatics tools and services has its role in analyzing and comparing information from multiple sources of information.

Standardization of data types or unifying the format of data to increase collaboration, integrating tools and services are paramount while integrating data in the biomedical field to reduce complexity and enhance performance. The adoption of wide-scale biomedical ontologies and data standards is needed to ensure accurate data integration (Marti-Solano et al., 2014). Storing biological data in a standard fashion eases the complexity of interpreting this information (Zhao et al., 2015). The use of standards in collecting biomedical data in different types of metadata will facilitate data interoperability. The standardization of biological data is helpful to gain an understanding about the data before integrating them. It is paramount that data providers increase their use of ontologies and metadata standards to facilitate data integration (Hassani-Pak & Rawlings, 2017). Achieving a certain level of biological data unification throughout the process of integrating bioinformatics services and tools reduces time and effort for analyzing translational data and increases the efficiency of knowledge extraction (Bhuvaneshwar et al., 2016). Standardization and data integration eases the process of merging information. Bringing data into standard formats is critical, having same information stored in many forms increase complexity and ambiguity, standardizing the format of data enhance the quality of data making it simple to manipulate allowing collaborative research and sharing of complex methodologies.

Many principles in information technologies have been helpful in the process of integrating data from multiple sources (Hassani-Pak & Rawlings, 2017). Making more data available in a linked form will simplify data integration processes and improve most aspects of data origin. Using cutting edge technologies in information technology can be the solution for integrating disparate biological datasets.

Result analysis in bioinformatics. Biological information is increasing exponentially, making the process of analysis complicated. The increasing volume and complexity of biological information demands experts to create an analytical workflow involving multitude of steps for extraction of knowledge (Dubchak et al., 2014). The analytical tools increase the complexity of analyzing biological data. The tools are required in bioinformatics to help specialists to obtain results from the biological information. Specialists must efficiently use the biological data that are collected in large quantities at a rapid pace. Many approaches and techniques must be considered to achieve such a goal. With improvement in bioinformatics and life science, large amounts of biological data can be collected (Triplet & Butler, 2014). The amount of biomedical data being collected outweighs the amount of data that are being analyzed (Greene et al., 2016). Specialists have developed filtering strategies to produce quality positive datasets (Demirci & Allmer, 2017). Taking into consideration new IT techniques can be helpful to analyze these biological data. Researchers are using cloud computing, big data, Internet of things, or another recent technological trend to solve complex bioinformatics problems. (Alansari et al., 2017). Specialists who analyze bioinformatics data generate large amounts of scrambled data, which necessitates filters to enhance the quality of the generated data. An analysis of the positive data is needed to establish machine learning models.

The complexity of biological data and heterogeneity have introduced analysis challenges. Many tools exist for processing bioinformatics analysis; however, these tools require bioinformatics specialists to help with the analysis (Bhuvaneshwar et al., 2016). A one-step system that can handle all biological data including NGS and medical images without the need to switch from one system to another does not exist (Bhuvaneshwar et al., 2016). Those tools need to have a user-friendly interface to ease the complicated process of extracting knowledge from biological data (Velloso et al., 2015). Having only bioinformaticians to process analysis results is a drawback, and life science specialists will become dependent on the bioinformaticians with no ability to analyze and extract knowledge on their own. Analyzing bioinformatics data presents a bottleneck for laboratories having a lack of technical data and software specialists (Zou et al., 2017). It is difficult for bioinformatics laboratories to maintain the required number of bioinformatics experts with the growing complexity of bioinformatics tools (Williams et al., 2016). Developing bioinformatics tools that are useable by life science specialists will improve this field. Because high-performance computers are expensive, cloud computing is an alternative for analyzing biological information. Cloud computing allows specialists to share all available computational resources from both time and performance point of view. Life science specialists can use multifunctional platforms that are easy to use for extracting knowledge and patterns from biological data.

Text mining. It is the process of analyzing huge amounts of unstructured datasets. In the biomedical domain, the scientific community is producing huge amounts of scientific findings, which makes it challenging for scholars to find the required information in this large sea of knowledge (Basaldella et al., 2017). Text mining is paramount for knowledge extraction from vast amounts of information. Text mining technology can distil essential information from large quantities of biomedical literature (Przybyła et al., 2016). Text-mining developers use information retrieval technics, such as document classification and document retrieval, to select relevant documents (Huang & Lu, 2016). Using text mining technics, specialists will find relevant information in a fast manner. Most biomedical discoveries are written in scholarly publications. Extracting key information from free text and converting it into structured knowledge for human comprehension is crucial. Text mining technics are used to obtain relevant knowledge from the biomedical literature. Specialists can use article selection to narrow down the search space from the entire document to the knowledge of interest.

To extract relevant information using text mining technics, the specialist must be trained for such a complicated text because the data needs to be derived from vast amounts of heterogeneous data. For practical knowledge extraction, researchers must be skilled in the availability, suitability, adaptability, interoperability, and comparable accuracy of text mining resources (Przybyła et al., 2016). Many several sophisticated techniques are used in the text-mining process, such as using dictionaries and machine learning to recognize known entities (Basaldella et al., 2017). Results exctracted from biological data needs knowledgeable specialists effectively interpret the results (Przybyła et al., 2016). Entity recognition in text mining has switched from focusing on extracting a single entity type from scientific papers such as entity names to the use of terminological resources for more sophisticated text-mining paradigm (Basaldella et al., 2017). Text mining specialists must be trained in efficient information extraction using text mining technics. Stored information is increasing sequentially and in great amounts, discovering patterns out of this massive data is challenging and needs special technics and expertise (Inzalkar & Sharma, 2015). Despite the power of text mining technology, the inexperienced user finds text mining difficult, with an overload of resources, services, tools, and frameworks, a researcher find it overwhelming to use many methods to identify a certain pattern for his study. Without trained specialists in text mining, the bioinformatics field will not be able to effectively use the information hidden in the vast quantities of biological data.

Bioinformatics challenges. The heterogeneity of biomedical data is a significant challenge. A key determinant of data usefulness can be in many cases the availability of additional information to evaluate a particular datasets. Specialists must be able to track the data source and to retrieve information from the context in which the information was generated to determine if the data can be meaningfully combined (Marti-Solano et al., 2014). Even while combining the heterogeneous data, restrictions might be implemented on remote resources and security concerns during the transmission over the Internet must be dealt with (Shah et al., 2005).

Many challenges need to be addressed by bioinformaticians for more effectiveness in this field. Bioinformatics specialists experience problems like data quality and processes from different technology platforms, data inconsistencies, incomplete and inaccurate knowledgebases, false discoveries, a lack of novel algorithms for data integration, computational inefficiency, faulty data interpretation, and lack of visualization (Zhao et al., 2015). Other challenges might also be introduced like restrictions from remote servers, security of transmitting biological data over the Internet, and logistics for querying distributed resources (Shah et al., 2005). Understanding the nature of the problems in bioinformatics can be crucial in overcoming them for more efficiency in the bioinformatics research area.

Bioinformaticians must get their tools up and running, and cloud computing can reduce the complexity of such problems. Bioinformatics meet a range of difficulties to get tools locally installed, running, and producing results (Velloso et al., 2015). Specialists can use cloud computing target the problem of installing, running, and maintaining sophisticated bioinformatics tools directly (Souilmi et al., 2015). Specialists can centralize bioinformatics tools using a user-friendly interface and webservices to overcome the technical challenges in bioinformatics (Velloso et al., 2015). Bioinformatics field specialists find it difficult to get complicated tools and platforms running and locally installed. Centralization and webservices are helpful in installing a sophisticated bioinformatics tool. Bioinformaticians can use centralized tools and webservices to be more productive and avoid spending time on installation processes.

Due to the complexity of data, bioinformatics computational resources are required for analyzing the information. Bioinformaticians require a high-performance computer with a clustered processor, high internal bandwidth to fast storage, and software to carry out an elaborate multiple step workflow (Hong et al., 2013). Computational and storage limitations include the costs associated with keeping data, moving data, and analyzing data. Traditional high-performance computer resources are expensive both in purchase and maintenance (Miller, Zhu, & Bromberg, 2017). In an attempt to reduce costs, research labs use different techniques to deal with their computational needs. Some labs have their computational power, and others share machines across an institute or outsource their computing to collaborators (Miller et al., 2017). Sharing computing resources is an efficient cost-reduction solution because not all computational resources are used simultaneously. Computing nodes rarely reach the often-targeted use rates of 75%- 85% consistent workload; computing usage peaks only with a high priority project running on the cluster for a limited period (Miller et al., 2017). Distributing and sharing high-performance computing wisely may reduce cost and increase productivity while analyzing biological datasets.

Next generation sequencing (NGS). NGS has evolved enormously over the years from two perspectives efficiency and cost reduction, which resulted of generating enormous amounts of data. The increasing amount of data is accompanied with a fast advancement in the NGS technologies both in terms of increasing sequencing depth and decreasing cost of whole genome sequencing (Hong et al., 2013). The decreasing cost of sequencing, is increasing the number of sequencing projects and generated data is enormously increasing (Reddy et al., 2015). NGS has resulted huge amounts of data where bioinformatics stepped in. Bioinformatics used the generated data from NGS to relate patterns and extract knowledge. After sequencing the samples, bioinformatics implements storing management and interpreting the huge amounts of NGS data. The data generated by NGS is enormous and complex, to extract knowledge and related patterns from this data, a researcher needs bioinformatics best approaches. NGS is an extremely complicated process and many fields are involved. NGS is a complex integration of chemistry, biology, optical sensors and computer hardware and software (Hong et al., 2013). According to Ardeshirdavani et al. (2015) the analysis of a single whole genome is exceptionally complicated that the analysis process can take up to 50 Gb of collections of files as a result. The involvement of this number of areas is more than proof of the complexity of the NGS. NGS platforms generate intensity data that are determined from the image captured by their optical resources. Modern sequencing platform generates petabytes of data (Souilmi et al., 2015). The first bioinformatics task in a NGS project is to analyze the image files generated by an NGS instrument to deduce or conclude the individual bases from the intensity data, and this process is called base calling (Hong et al., 2013). An important aspect in NGS is accuracy. Increasing accuracy can be achieved using re-sequencing and combining multiple NGS runs. The accuracy of sequencing is improved by increasing the depth throughout re-sequencing the same DNA sample multiple times and combining data from multiple runs (Hong et al., 2013). Correlating multiple studies together is much more powerful then analyzing data from a single study (Ardeshirdavani et al., 2015). When dealing with complex and huge forms of data, to achieve accuracy is crucial and challenging, yet in NGS accuracy can be achieved throughout re-sequencing and combining data from multiple run, which is a trick process yet enhance NGS results.

A challenge that NGS faces is the huge amount of data generated from sequencing and needs to be stored; it may not be cost efficient to store this huge amount of data. In NGS retaining raw data in the future is not practical, as the data is increasing tremendously, it may become cheaper to sequence that to store (Hong et al., 2013). The current challenge in NGS is analyzing this large-scale data (Souilmi et al., 2015). Re-sequence from the original data is more practical and more cost-efficient than storing enormous amounts of sequenced data. Even though that the data storage cost is decreasing enormously, the NGS data storage and management remains a large portion of institutions budgets.

Raw data and mapped reads are large files occupying significant disk storage space. In the past sequencing was limited to a number of high importance organisms, with the decreasing cost of sequencing and the increasing technology associated with, sequencing nowadays is done on a higher scale resulting huge amounts of generated data (Reddy et al., 2015). The analysis of a single whole genome can take up to 50 Gb of collections of files as a result (Ardeshirdavani et al., 2015). This raises significant issues in term of computing, data storage and transfer.

Confidentiality and protecting private data is a great issue. NGS data analysis raises challenges like how to protect the confidentiality and privacy of personal genomic data during knowledge extraction (Ardeshirdavani et al., 2015). NGS data can be transferred over the Internet which raises security concerns about the confidentiality of the transferred data (Shah et al., 2005). Personal genome data is sensitive personal data, confidentiality must be ensured at all levels and all times, and only authorized researchers should have access to such personal data.

Another challenge is also related to data and management of the huge amounts of data generated from NGS which the bandwidth to store is, manage and access this data. High I/O bandwidth for storage and between IT components to keep up with data output from NGS instruments as well as to allow many users to access the data simultaneously is a huge challenge for NGS (Hong et al., 2013). When the data is massive it needs huge bandwidth to carry enormous amounts of data over the network. It also requires higher high-performance computing to analyze the exponentially growing biological data (Miller, Zhu & Bromberg, 2017). Managing NGS data at such scale, and especially that the amount of data will double every 2 years in an integrated environment will be an increasing challenge.

To overcome the flooded NGS generated data, cutting edge technics must be used. Improved data management technics and infrastructures is needed urgently, scientific community started to consider adopting cloud computing for NGS analysis, cloud computing in NGS analysis is extremely efficient such as distributing the time-consuming computational jobs on many cloud environments (Hong et al., 2013). Advancement of IT technologies, such as high-performance computing and new cloud computing solutions created new opportunities for computational biology (Miller, Zhu & Bromberg, 2017). NGS can benefit from technological advancements and benefit from a trending paradigm such as cloud computing to solve challenges in this field. Even that adopting cloud computing technology can have some drawbacks such as protection of proprietary data and unauthorized access to data stored in a public cloud. Cloud computing solution may reduce the cost by eliminating the costly in-house IT infrastructure development.

With the decreasing cost of sequencing, the number of sequencing projects and the amount of sequence data generated is increasing exponentially, storing metadata becomes inevitable for more meaningful data. The sequenced data are submitted resources or analysis platforms it becomes paramount to document the associated metadata in order to facilitate comparative analysis and hypothesis generation (Reddy et al., 2015). To ensure an effective reuse of data, it has to be enriched with relevant metadata, and converted into appropriate format for integrative knowledge management (Marti-Solano et al., 2014). Metadata give the users the ability to look at their data and analyze results from a whole different perspective (Reddy et al., 2015). When sequencing was expensive only limited number of high importance organism genomes were sequenced, maintaining the associated information in catalog format was sufficient. But since sequencing has become more affordable, it is now efficient to use data from multiple sources, it became paramount to collect common metadata to these samples.

Overcoming bioinformatics challenges. The exponential growth of biological data has introduced many challenges. Those challenges must be surpassed to advance in the bioinformatics field. The exponential growth in data poses significant challenges for researchers, because many bioinformatics applications requires the process to store, access and analyze large libraries of data (Berger, Daniels & Yu, 2016). With the extraordinary advancement in the IT, many paradigms exist to deal efficiently with storing, managing and manipulating huge amounts of data. Cloud computing is an approach to tackle these challenges. Google clouds and Amazon web services are commonly used for computational biology (Berger, Daniels & Yu, 2016). Cloud computing has many advantages, it solves high computational problems like high cost and complex installations. Cloud computing free the researchers from maintaining their own data centers, and provide cost saving benefits (Berger, Daniels & Yu, 2016). Cloud computing can be extremely beneficial in computational biology even though cloud computing does not truly address the problem of the exponential growth in omics data. The algorithms used to extract knowledge must evolve for more efficiency in analyzing huge amounts of data. The development of algorithms that leverage the structure of biological data that we can make sense of biology in light of evolution (Berger, Daniels

& Yu, 2016). The evolution of algorithms may eliminate some challenges and help bioinformatics to attain a more mature state.

Evolution of Technology Acceptance Model

Technology is evolving. According to Moore’s law, the complexity of technology will double every 24 months (Moore, 2006). Users may struggle to adapt to new technology in the rapid pace of technological advancement (Venkatesh et al., 2012). Leaders must predict user acceptance toward technologies, so they can start to identify variables that may affect the user acceptance for a particular technology. TAM was introduced by Davis (1989) where he determined that behavior intention and attitude toward using a system influence the actual use of the system.

Theoreticians have found that the performance of a user can be influenced by many factors. Davis studied the influence of external factors on a person’s performance that played a role in changing the perceptions and behaviors of people towards the use of a technology (Davis, 1989). Davis (1989) determined that the attitude and perceptions of a user will directly affect the system usability. According to the TAM, the user’s acceptance of a system is determined by two factors: perceived usefulness and perceived ease of use (Davis et al., 1989). TAM has been used to predict users’ acceptance toward a technology and the actual use of this technology (Venkatesh et al., 2012). Researchers modified TAM according to their research objectives (Masood & Lodhi, 2016). TAM is applicable for predicting acceptance and usage of new technologies in many fields (Rauniar, Rawski, Yang, & Johnson, 2014). TAM evolved into a leading model in explaining and predicting system use.

According to TAM, if a technology or innovation enhances the performance of a person without increasing mental and physical effort, it is considered useful and easy to use, and it is more likely to be adopted by users. TAM has two main factors: perceived usefulness and perceived ease of use (Davis et al., 1989). The user behavior to use a technology is influenced by the usefulness and easiness. The usefulness and easiness in the model is affected by external factors. When the technology is easy to use, the usefulness increases. According to TAM, perceived ease of use has a significant positive effect on perceived ease of use (Masood & Lodhi, 2016). Davis (1989) concluded that the user attitude toward using a system is a determinant of whether the user will use or reject the system. The attitude is influenced by two beliefs: the perceived usefulness and the perceived ease of use, with the perceived ease of use having a direct influence on the perceived usefulness (Marangunic & Granic, 2015). What makes a system useful is the ease of use of this system, if a user is using a system with relative ease the usefulness is increasing enormously.

In the original TAM, Davis (1989) explained that the user motivation is affected by three factors. Perceived usefulness refers to a user's subjective probability that using a particular system/technology improve the user's performance at work (Davis, Bagozzi, & Warshaw, 1989). The perceived ease of use is the degree that a user believes that using a particular system will enhance his/her job performance (Rauniar, Rawski, Yang, & Johnson, 2014). The perceived ease of use indicates the practicality of technology (Hui-Fei & Chi-Hua, 2017). The perceived ease of use is the user willingness to engage with new technology and the user's attitude toward the new technology includes his or her beliefs about whether this technology will help the user in performing his or her tasks more efficiently (Davis et al., 1989). (Venkatesh, 2002) determined that the perceived usefulness is a component in determining the user's acceptance toward a technological innovation. The usefulness of a particular system is determined by the system itself and by the user determination to use that system.

Using a system should be effortless, and the system should be free of unnecessary complexity, the effortless usability of a system is called perceived ease of use. Perceived ease of use refers to the degree to which a user expects the use of a system/technology to be free of effort (Davis et al., 1989; Venkatesh & Davis, 2000). Perceived ease of use indicates the effort required to learn to use technology perceived by an individual (Hui-Fei & Chi-Hua, 2017). perceived ease of use is defined by Davis (1989) as a system's utility. The system's ease of use must outweigh the effort required in adopting it. Perceived ease of use is classified into three clusters: the physical effort, mental effort, and the direct perception of how easy the system is to use. Attitude toward use is influenced by the perceived ease of use and perceived usefulness (Wu, Chou, Weng, & Huang, 2011), and it is the primary factor that determines the system usage. The perceived ease of use impacts the perceived usefulness, and all of these characteristics are affected by certain aspects characterized by X1, X2, X3 (Figure 1).

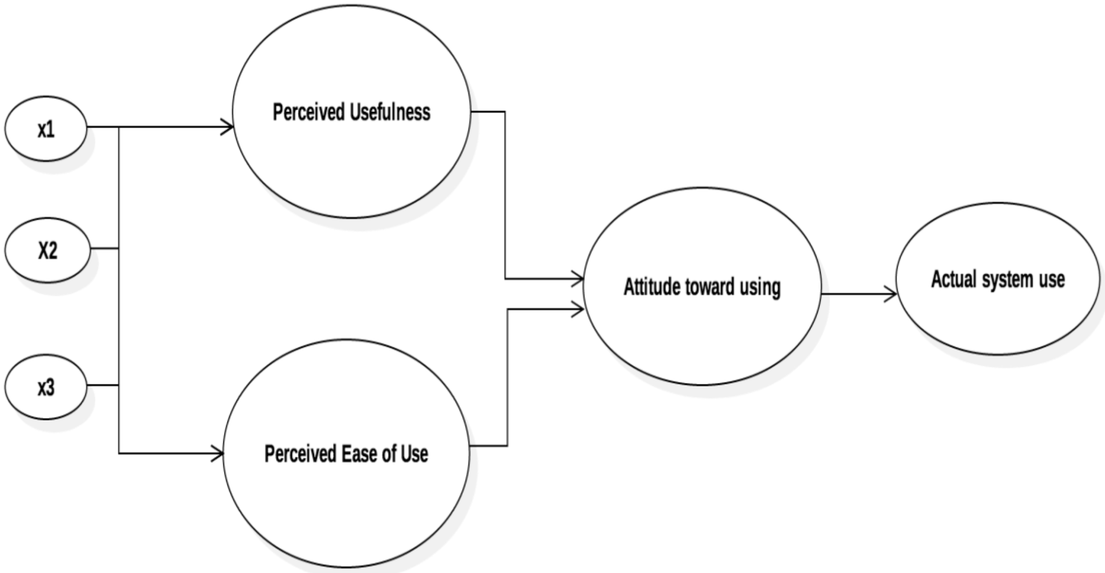


Figure 1. Davis original technology acceptance model. Reprinted from "Technology Acceptance Model: A literature Review from 1986 to 2013," by N. Marangunic´ and A. Granic´, 2015, *International Journal of Information Management*, 36(6, Part B), pp. 1248-1259. Copyright 2015 by SPRINGER-VERLAG. Reprinted with permission (Appendix C).

Perceived usefulness and perceived ease of use impacts user attitude. According to Davis (1989), the attitude of a user affects his or her intention to use or not a system or technology. The attitude of a user is the degree to which a user is interested in using a particular system, and it determines the behavioral intention that leads to actual system use (Tzafilkou & Protogeros, 2017). An individual attitude or intention toward using a technology is influenced by the perceived usefulness and perceived ease (Rauniar et al., 2014). An individual behavioral intention to use a new system is influenced by the attitude and perceived usefulness (Wu, Chou, Weng, & Huang, 2011). Researchers use

TAM to highlight external factors and internal beliefs that can explain system usage on the basis of perceived usefulness (PU) and perceived ease of use (PEU) (Persico, Manca & Pozzi, 2014). According to TAM, perceived usefulness has a direct effect on the attitude toward using a technology, and perceived ease of use has a direct impact on the perceived usefulness (Tzafilkou & Protogeros, 2017). User intention to use a technology will determine his or her behavior for using or not using that technology (Rauniar et al., 2014). The user behavior or intention, according to TAM, is a factor in determining the usability of a system, and the user behavior factor is determined by the perceived usefulness and the perceived ease of use. Perceived ease of use and perceived usefulness are factors that influence the user attitude toward using a technology, according to TAM.

Researchers use TAM to explain and predict IT user behavior. A user's personal intention to use and adapt to new IT is determined by the perceived ease of use and the perceived usefulness (Hui-Fei & Chi-Hua, 2017). In TAM, Davis (1989) illustrated how the attitude toward technology has a direct influence on system usage. The increasing progress of technological innovations makes a user's acceptance of technology a component of the success or failure of a technology (Wu et al., 2011). With the continuous technological development, and its overlapping into the user professional and private life, the decision regarding the acceptance or the rejection of the technology remains a dilemma because many factors influence the rejection or the acceptance of a new technology (Marangunic & Granic, 2015). If the users do not accept the new technology, the technology will fail, because the user acceptance and the usefulness of the system are the key factors that determines the success or failure of the system. The user behavior has a direct effect on the acceptance or rejection of a new technology.

TAM has been considered as one of the best theories to predict user acceptance toward a new technology it has been used by developers of new technologies and senior managers to predict user acceptance before introducing new technology to an organization. Among the theories that examinees the user acceptance of information technology systems, TAM is the most used theoretical model and the most cited (Polančič & Jošt, 2016). TAM has a strong prediction power and it has been widely used in investigating acceptance in various technologies (Jaehee et al., 2014). User acceptance of a new introduced technology in an organization is studied with care by senior managers because poor acceptance of a system by employees may lead to resources losses, and decreased productivity.

Theory of Reasoned Action and Theory of Planned Behavior

TAM evolved from the theory of reasoned action (TRA) and the theory of planned behavior. The TRA is used to predict the actual individual behavior (Fishbein & Ajzen, 1967) developed the TRA. The actual behavior of a person can be determined by his or her prior intentions along with the beliefs that the individuals has for the given behavior (Davis, Bagozzi, & Warshaw, 1989). TRA was used to explain and predict employee behavior (Wu et al., 2011). According to the TRA, attitude and subjective norm impact behavioral intention, where behavioral intention influences the actual behavior of an individual while using a technology (Tzafilkou & Protogeros, 2017). Attitude toward behavior is the person’s positive or negative feeling about performing the actual behavior (Hill, Fishbein, & Ajzen, 1977). The subjective norm is the person’s perception or opinion about what others believe the person should do (Hill et al., 1977). Proponents of TRA look at the behavioral intentions rather than the attitudes as the main predicators of behaviors (Marangunic & Granic, 2015). TRA has been used to understand the adoption of behaviors, technologies, or advice (Wallace & Sheetz, 2014). In the TRA, people form intentions to adopt a technology based on their beliefs about the consequences of adoption (Wallace & Sheetz, 2014). The immediate determinant of behavior is the individual intention to perform or not, and the intention is influenced by the attitude and subjective norm factors (Ajzen, 1991). The behavioral intention is determined by the attitude toward behavior and the subjective norm (Figure 2).

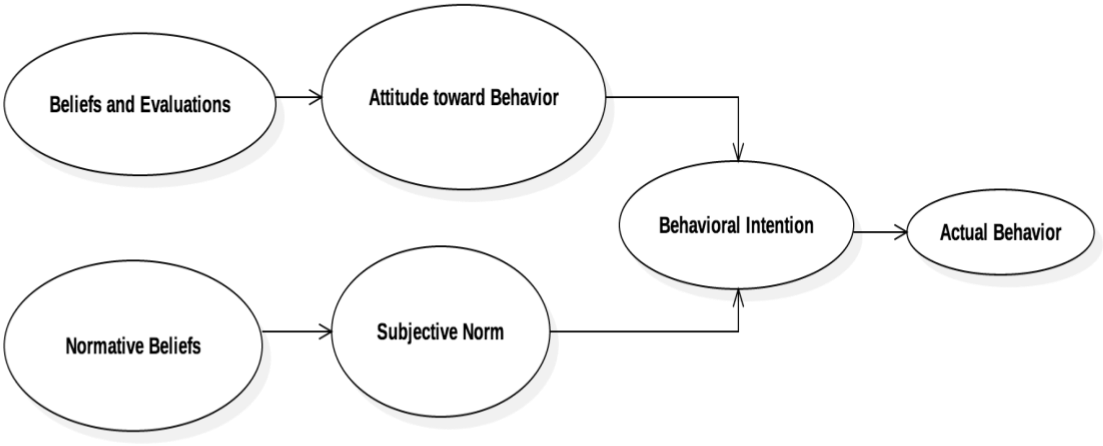


Figure 2. Reasoned action model.

The TRA model proved ineffective in predicting user behavior, and researchers found that new factors may enhance the user behavior. A user with positive attitude may not perform a certain behavior because of the lack of control over the individual activities. The TRA is constrained by the lack of appropriate opportunities, skills, and resources for a particular user (Tzafilkou & Protogeros, 2017). TRA has several limitations. One of the limitations of the TRA is that people have little power over their behavior and attitude (Marangunic & Granic, 2015). For this reason, the TRA was extended to include perceived behavioral control as a new variable (Tzafilkou & Protogeros, 2017). Ajzen (1991) added a new element called perceived behavioral control to the original TRA theory. This addition of the new element to TRA, created the theory of planned behavior (Marangunic & Granic, 2015).

The purpose of the theory of planned behavior is to explain volitional behavior. A factor of the theory of planned behavior is the user's intention to perform a behavior, which determines individual performance (Marangunic & Granic, 2015). The perceived behavioral control is the person's ability to perform a behavior (Ajzen, 1991). The theory of planned behavior components attitude, subjective norms, and perceived behavioral control are used to explain behavioral intention (Tzafilkou & Protogeros, 2017). TAM is more suitable to my study then theory of planned behavior, because in the theory of planned behavior self-efficacy has a direct effect on perceived usefulness but not on perceived ease of use (Jun, Lee & Jeon). Scholars use theory of planned behavior to examine how the decision-making process leads to the formation of attitudes that subsequently guide behaviors (Leeuw et al., 2015). The theory of planned behavior determines that an individual behavior is determined by its beliefs. An individual behavior intention is formed by the attitude toward behavior, perceived behavioral control and subjective norms.

Different Versions of the Technology Acceptance Model

According to the first modified version of TAM, in some cases a person might form a behavioral intention without forming any attitude about using a system (Davis, Bagozzi, & Warshaw, 1989). The first modified version of TAM is illustrated in Figure 3.

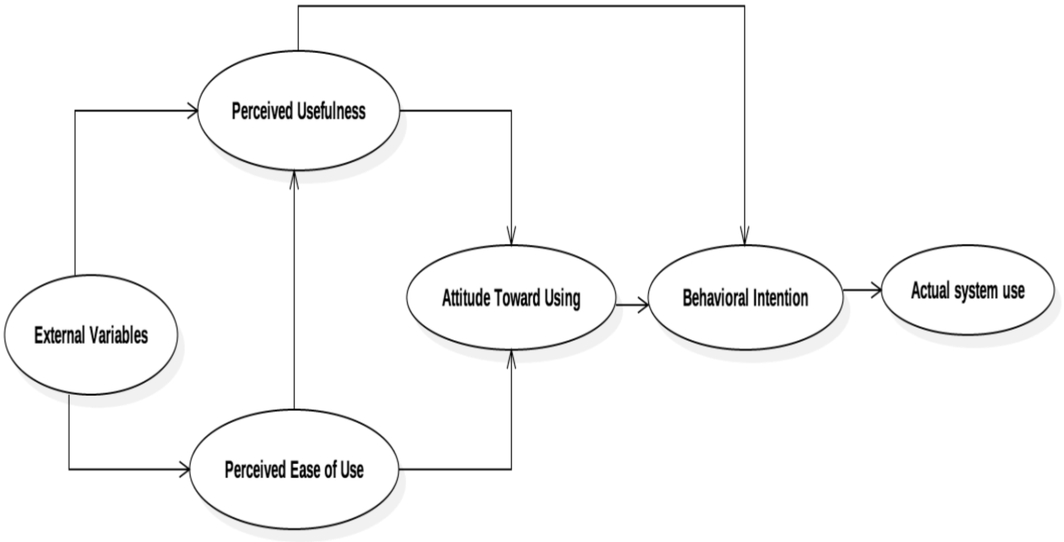


Figure 3. First modified version of the technology acceptance model. Reprinted from "User Acceptance of Computer Technology: A Comparision of Two Theortical Models," by F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, 1989, *Management Science*, 35(8), p. 985. Copyright 1989 by INFORMS. Reprinted with permission (Appendix C).

A later development of the TAM would eliminate the attitude toward using and introduce the behavioral intention as a variable that is directly affected by perceived ease of use (PEU) and the perceived usefulness (PU) (Davis, Bagozzi & Warshaw, 1989). Davis and associates found that the attitude toward using did not fully mediate the PEU and PU (Marangunic & Granic, 2015). So, Davis, Bagozzi & Warshaw (1989) removed the attitude from the model and they included the behavioral intention as a new variable which is directly influenced by the PEU. The design characteristics represented by X1, X2, and X3 (Figure 1) is being replaced by external variables. Both beliefs the PEU and PU are influenced by the system design characteristics, introducing external variables might influence the user intention toward a system (Marangunic & Granic, 2015). Many researchers have demonstrated that external variables have an influence on the process of adopting new technology by a user (Hamid et al., 2016; Hussein, 2017; Mortenson & Vidgen, 2016). In order to extend TAM, researchers have identified the perception of resources and support as a major external factor that affect the adoption of information technologies (Abdullah & Ward, 2016; Bach, Čeljo & Zoroja, 2016). External variables might affect the assumptions of a person towards a system; the external variables represent user training, user participation, system characteristics and the nature of the implementation process (Venkatesh & Davis, 2000). Brezavšček, Šparl, & Žnidaršič, (2014) modified the model with three additional external variables, they found significant positive effects of statistics learning self-efficacy and statistics learning value on perceived usefulness and perceived ease of use. TAM addressed PU, PEU, behavioral intention and system usage as major determinants that predict the acceptance of a new technology (Tzafilkou & Protogeros, 2017). Learning self-efficacy and subjective norms affected PU whereas system accessibility and learning self-efficacy significantly affected PEU (Hansen, Saridakis & Benson, 2018). PU and PEU has a significant effect on the intension to use technology. The final version of TAM is illustrated in Figure 4.

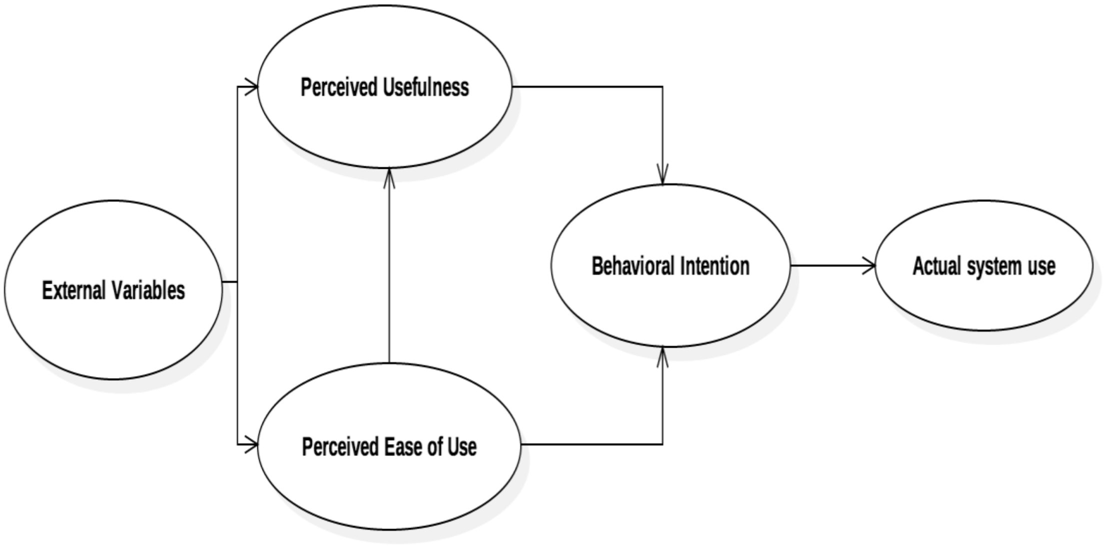


Figure 4. Final version of the technology acceptance model. Reprinted from "User Acceptance of Computer Technology: A Comparision of Two Theortical Models," by F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, 1989, *Management Science*, 35(8), p. 985. Copyright 1989 by INFORMS. Reprinted with permission (Appendix C).

The TAM is a convenient theory to explore strategies to integrate disparate bioinformatics datasets. TAM can help the researcher to identify the factors that influence the bioinformatics experts decision about what makes the pattern extraction more efficient and accepted. TAM could be a useful theoretical model to obtain a deep understanding of relationships among ease of use, usefulness, attitude toward using and intention about integrating disparate bioinformatics datasets.

Davis considered that the subjective norm effect on behavioral intention could be ignored; for that reason, in the TAM the variables of subjective norms was not included (Wu et al., 2011). The TAM had some limitations in explaining the reasons why a person would identify a system is used, so additional variables are added as antecedents to the perceived usefulness, this model is called by Venkatesh and Davis TAM 2 (Figure 5).

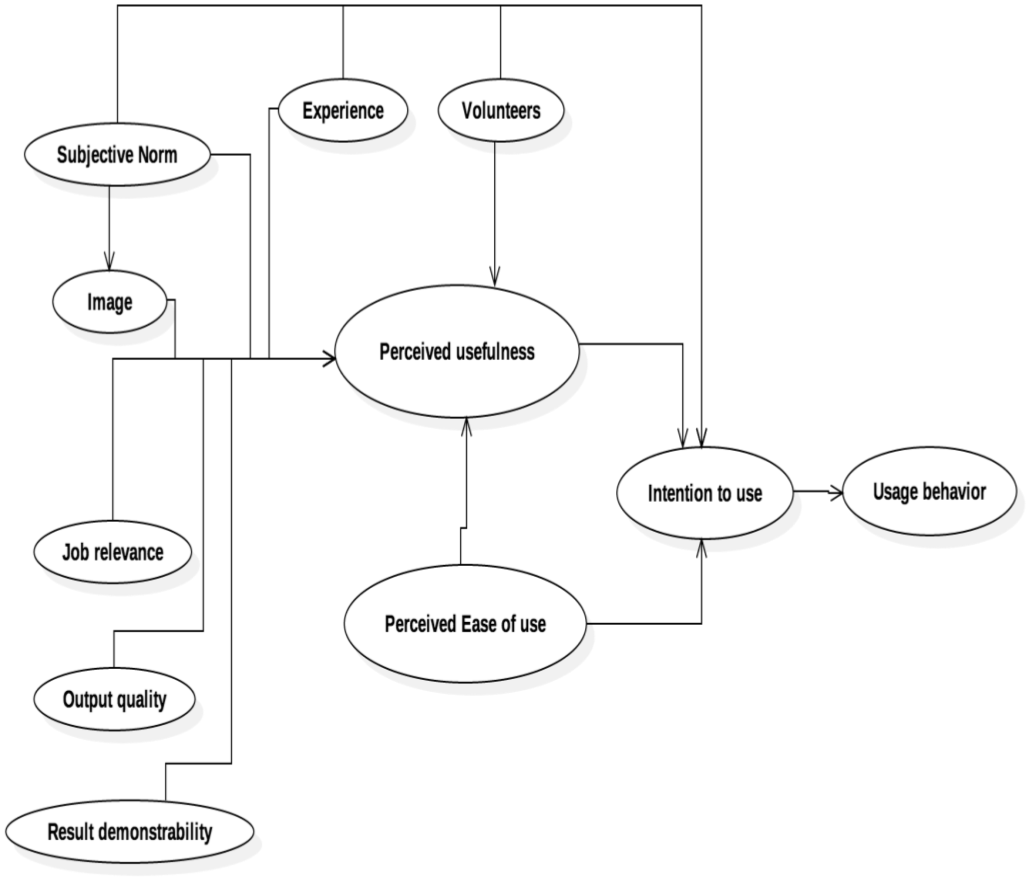


Figure 5. TAM 2. Reprinted from “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies,” by V. Venkatesh and F. D. Davis, 2000, Management Science, 46(2), p. 188. Copyright 2000 by INFORMS. Reprinted with permission (Appendix C).

Venkatesh and Davis in 2000 introduced TAM2, the new model introduced the social influence process that contains subjective norm and Image; and the cognitive instrumental process that includes job-relevance, output quality and result demonstrability (Wu et al., 2011). Extending the TAM to TAM2 was to provide a better understanding of the determinants of perceived usefulness and perceived ease of use by adding cognitive instrumental processes and social influence processes (Venkatesh & Davis, 2000). The subjective norm is the community attitude toward the user behavior (Wu et al., 2011); it is the influence of others on the user decides to use or not the technology (Venkatesh & Davis, 2000). TAM2 sought to identify the variables that influence the perceived usefulness (Marangunic & Granic, 2015). The Image, is the belief of a group relevant to the user, that the user behavior can enhance the performance of an organization (Wu et al., 2011). Voluntariness is the degree which a person believes that accepting the technology is not mandatory (Park et al., 2014). Voluntariness has a direct influence on the user intends to use a system (Riemenschneider et al., 2002). Job relevance, the degree to which the technology was applicable. It is the individual perception in which the system is targeting the user Job (Wu et al., 2011). Job relevance is the personal belief that technology applies to his job (Venkatesh & Bala, 2008). An individual is more likely to accept a technology if he believes that it is relevant to his work. Output quality, the extent to which the technology performed the required tasks. Output quality depends on job relevance (Venkatesh & Davis, 2000), and it is the degree which a user judges the effect of a new system; it is the user believes that a system can perform required tasks (Wu et al., 2011). The Result demonstrability, the production of tangible results, the users will have a positive attitude of the usefulness of a system if positive results are noticeable (Venkatesh & Davis, 2000).

Another evolution of the TAM was made by Venkatesh (2000), who included determinants for the perceived ease of use. Venkatesh identified two groups of antecedents for perceived ease of use. The Anchors, a general belief about computers and computer usage. The adjustments, a belief that are shaped based on direct experience with the target system. The model is represented in Figure 6.

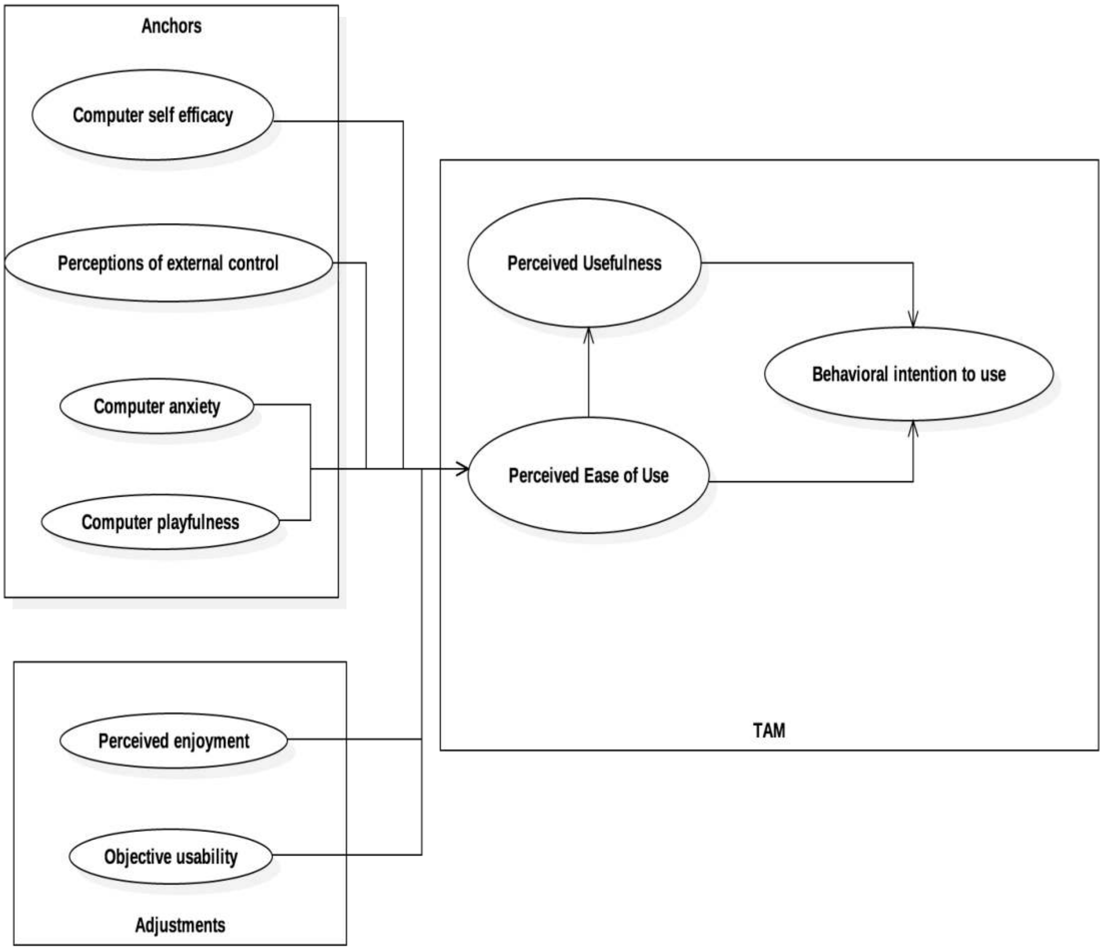


Figure 6. Extending TAM to include determinants for perceived ease of use. Reprinted from “Technology acceptance model: a literature review from 1986 to 2013,” by N. Marangunić and A. Granić, 2015, *International Journal of Information Management*, 36(6, Part B), pp. 1248-1259. Copyright 2015 by SPRINGER-VERLAG. Reprinted with permission (Appendix C).

Venkatesh and Bala (2008) updated the TAM 2 model and introduced TAM 3; they expanded the determinants for the perceived ease of use and perceived usefulness producing a positive behavior intention which provides the user behavior. The determinants that affect the PU are the subjective norm, image, job relevance, output quality and result demonstrability. The determinants that influence the PEU are the anchor variables which they include computer self-efficacy, computer playfulness, computer anxiety and perceptions of external control. Adjustments variables (perceived enjoyment and the objective usability) also influence the perceived ease of use. See Figure 7 for the illustrated the TAM 3 model.

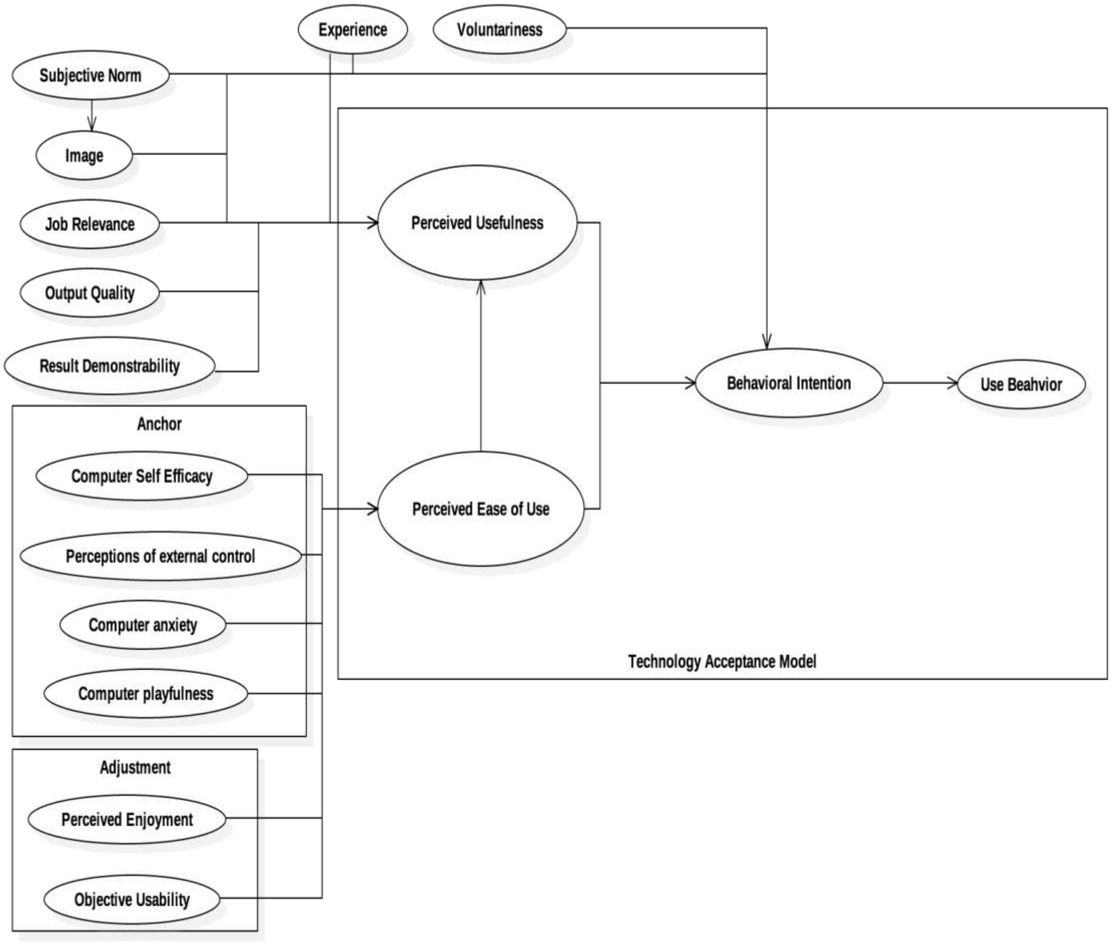


Figure 7. TAM 3. Reprinted from “Technology Acceptance Model 3 and a Research Agenda on Interventions” by V. Venkatesh and H. Bala, 2008, *Decision Sciences, 39(2)*, pp. 273-315. Copyright 2008 by SPRINGER-VERLAG. Reprinted with permission (Appendix C).

Limitations of the Technology Acceptance Model

TAM has become nearly a law-like model, and it is even used in areas outside the technology adoption like marketing, advertising, information adoption and e-learning. But some researchers disagree about the overrated influence of TAM. Researchers argued that TAM is not progressive, it just provides alternative hypotheses when they face anomalies and parsimonious where there are variables other than PU and PEU that may have a significant influence on behavioral intention (Date, Ramaswamy & Gangwar, 2015; Islam et al., 2014). Organizations faill to accumulate benefits of investments if users are unwilling to use available systems (Brown, Venkatesh & Goyal, 2014). Personality traits such as extraversion and emotional stability could affect a person’s PU and behavioral intention to use technology. User might reject a system because of poor reliability even if he considers it useful (Fletcher, Sarkani & Mazzuchi, 2014). Also, external variables in the extended models of TAM are not clearly defined (Date, Ramaswamy & Gangwar, 2015). TAM has limited explanatory power; some aspects of TAM are hard to explain and to fully comprehend. Some other values of TAM are not practical and hard to implement. Researchers have extended TAM to include some 30 additional factors, which increases complexity (Fletcher, Sarkani & Mazzuchi, 2014). TAM can be considered limited because of its dependence on subjective and self-reported surveys rather than actual system use. At some level, TAM assumes that increased use of technology will improve performance, which is not always the case, and extensive use of a specific system will not necessarily affect the acceptance and TAM may have some arguably limitations like philosophical holes and constraints regarding missing variables.

The intention to use a new technology in TAM can be overrated, and training can enhance the user intention. In organizations where system usage is mandatory, intention to use a system loses its importance as an influential variable of TAM (Hwang et al., 2017). Training and educational programs enhance the feeling of self-efficacy where a user can develop a necessary experience to use systems efficiently with confidence (Gallego et al., 2015). A user with no programing skills might not develop a user intention to use a software that require so. But with training and educational programs the user acquire knowledge that might enhance his confidence and user intention to user a software that require little programming language. An employee might hold a negative attitude toward using a new technology, but the employee will eventually use the system because he simply has to. User acceptance is considered an important aspect in organizational success, and it is studied with care when a new technology is introduced. Developers of new technologies and senior managers are realizing that lack in user acceptance may lead to resources loss (Hwang et al., 2017). A big challenge for management is getting the users to maintain a positive attitude toward the adoption of new implemented systems.

The intention to use a system and the actual usage represents one aspect of the process, many other aspects affect the effectiveness and success of new technology when implemented. The intention to use a system does not take into consideration the pre-implementation contexts and does not provide a complete picture (Hwang et al., 2017). Some variables related to human and social change must be added to the original TAM form more efficiency when predicting user acceptance (Gallego et al., 2015). TAM is considered rigor in assessing the user acceptance toward new technologies, and scholars approved that TAM should be extended to include factors that explain PE and PEU (Jaehee et al., 2014). A drawback of the TAM is that the exited variables does not fully explain the user acceptance and therefore new variables needs to be included by extending the original TAM for more accuracy and effectiveness while predicting user acceptance when new technology is introduced in an organization.

Usage of the Technology Acceptance Model in IT and Other Areas

TAM has been used in many IT related areas to predict the user acceptance of a particular technology. TAM and the user intention is studied by researchers in many technological fields like websites, cloud computing, mobile banking, e-commerce and e-banking (Alkali & Mansour, 2017). In IT related areas the success and acceptance of a technology is highly related to the user acceptance, where the PU and PEU of use are paramount factors. TAM has been found useful in predicting user intention to use a technology (Alkali & Mansour, 2017). TAM has been applied to gain a better understanding of the quality techniques to improve software development practices (Holvitie, Lepparren & Hyrynsalmi, 2014); Also, TAM has been used to assess the user acceptance in many software development practices (Wallace & Sheetz, 2014). In a web acceptance model, users with high experiences are affected less by the TAM perceived ease of use than users with low experience in using the Internet (Gallego et al., 2015). In web development, an external factor which is the user experience has a direct influence on the user PEU and PU, where the user experience plays an integral part in the user acceptance of a website or web application. TAM has been engaged to assess the impact of PU and PEU in the e-Banking industry (Rodrigues, Oliveira & Costa, 2016). TAM is used to study the user acceptance of health-care mobile applications (Campbell et al., 2017). PU, PEU and investigating affecting factors are considered by managers when implementing health information technology systems (Garavand et al., 2016). The acceptance factors that influence the Google applications in a collaborative environment has been studied using TAM (Cheung & Vogel, 2013). TAM has been used to identify determinants that affect the behavioral intention to use YouTube for procedural learning (Lee & Lehto, 2013). TAM PU and PEU usefulness are relevant in predicting user intention in an e-training environment next to the strong influence of interactivity and trust (Alkali & Mansour, 2017). The adoption of e-government services has been explored using TAM (Rana et al., 2015). The PEU and PU of TAM alongside with environmental beliefs and normative beliefs affect the intention to use Green IT. Knowing that perceived usefulness has a more significant impact on the intention to use Green IT than perceived ease of use because people will use Green IT regardless of the convenience for the good of the environment (Yoon, 2018). TAM was used to predict elderly users' acceptance of new media entertainment technology where PU is translated into enjoyment and PEU as well as behavioral and psychological barriers then become direct or even indirect predictors of satisfaction and actual system use (Dogruel, Joeckel & Bowman, 2015). The attitude and the perceived ease of use of TAM are essential factors for adopting new technologies in online learning in higher education (Siegel, Acharya & Sivo). In an augmented reality tour-sharing application, the TAM perceived usefulness and perceived ease of use significantly affected the user intends to use the application because an application that is not user-friendly and complex to operate reduces user satisfaction (Hui-Fei & Chi-Hua, 2017). Another study suggested that when an enterprise adopts a mobile information device to share knowledge, the administration should take into consideration that the technology possesses an enhanced perceived ease of use, enhanced computer self-efficacy, enhanced computer playfulness and an enhanced perceived usefulness to ensure the user acceptance of the new technology (Yuan et al., 2017). Perceived ease of use, perceived usefulness and perceived playfulness which is the user expectation to obtain enjoyment from the use of technology, all affected the use of the social media application Snapchat positively (Makki et al., 2018). In information technology, the user acceptance is paramount for the adoption of a specific technology. For this reason, researchers have used TAM in many information technology related fields to predict factors that might influence the user acceptance of a particular technology.

The Technology Acceptance Model and the Bioinformatics Study

The lack of a theoretical foundation for this stream of research has limited the contribution of previous research and prevented organizations from understanding what are the measures that makes a bioinformatics system useful. Understanding the PEU and the PU of a system is necessary for the development of adoptable and practical measures that can lead to higher quality systems (Wallace & Sheetz, 2014). Theory based research on software adoption can provide reasons on why systems are used or not in practice.

In this PEU is defined as the degree to which the user believes that exploring strategies to integrate disparate bioinformatics datasets would be free from effort. PU is defined as the degree to which the user believes that exploring strategies to integrate disparate bioinformatics datasets would enhance the user performance. PE and PEU affects greatly the intention to use bioinformatics tools. PEU and PU are significant in explaining the intention of biologists to use bioinformatics tools (Shachak & Fine, 2008). Along with PE and PEU the specialists experience and knowledge toward bioinformatics tools cannot be excluded. Training enhance knowledge and skills that boost the awareness of bioinformatics tools. Bioinformatics is greatly related to other IT areas, and because TAM has showed great efficiency in many IT fields, bioinformatics is not different. Challenges in bioinformatics are similar to challenges in other IT areas, TAM showed efficiency in explaining the use of information systems in fields like office automation, software development, business application tools, telemedicine technology and digital libraries (Shachak & Fine, 2008). TAM was applied to explore strategies to integrate disparate biological datasets. TAM can be applied to software process improving initiatives, because the reason for taking new initiatives are similar to the reasons for introducing new technology. The PEU and the PU will be relevant when trying to anticipate who will adopt and begin using a bioinformatics system, therefore we feel that TAM is a relevant theory to examine bioinformatics system adoption. When bioinformatics specialists believe that using a particular bioinformatics system will increase the quality of pattern extraction, they will be more likely to use this system. Existing researches has shown that PU can predict user adoption (Wallace & Sheetz, 2014). Therefore, bioinformatics software must be perceived as useful, else ways life science experts will use them reluctantly or inappropriate.

Transition and Summary

Despite the continuous efforts in the bioinformatics field, this area stills lack strategies used by bioinformatics specialists to integrate disparate bioinformatics datasets. This literature review emphasized on the intersection of bioinformatics with IT, the essential role of bioinformatics, the vital part of computation and technological advancement in the bioinformatics field.

In this literature review, I highlighted a fundamental idea that illustrates the paramount role of technical progress in reducing the time and cost of sequencing biological information. Integrational bioinformatics took a big part in this literature review and how integrated disparate bioinformatics datasets enrich the process of knowledge and pattern extraction, the challenges that face the bioinformatics field like the complicated process of installing the sophisticated bioinformatics tools, massive computational resources and how cloud computing and efficient resources management can be beneficial. Additionally, this literature review described the process of sequencing and knowledge extraction, TAM and how TAM aligns with this study.

As addressed in my literature review concerning integrative bioinformatics, the advantages of integrating disparate bioinformatics datasets from different locations significantly outweigh the disadvantages. It allows bioinformatics specialists to have access to more information and extract knowledge and patterns more efficiently. I would encourage an in-depth look at the reaction of bioinformatics specialists to such a paradigm. This sort of study would help determining strategies used by bioinformatics specialists to integrate disparate bioinformatics datasets.

Section 2: The Project

Purpose Statement

The purpose of this qualitative case study was to explore the strategies used by bioinformatics specialists to integrate distinct bioinformatics datasets. The population of this study was bioinformatics specialists of a research institution in Lebanon that has strategies to integrate distinct bioinformatics datasets. The findings from the study may contribute to IT practice by identifying strategies to unify and integrate heterogeneous biological information from different locations and different structures. The study's findings may contribute to positive social change by impacting healthcare as a side effect of improvements to bioinformatics.

Role of the Researcher

In my role as a researcher, I conducted interviews with open-ended questions, transcribed the data, analyzed it, and presented the results and findings. Thus, as a researcher, I was the primary data collection instrument for this study (Marshall & Rossman, 2016; Sanjari et al., 2014). During my case study research, I strived for the highest standards by avoiding plagiarism, make sure that information included in the study are accurate to avoid deception, ensure accuracy in my research, and strive for credibility. Maintaining integrity and avoiding biases can be done when applying transparent and methodological research (Cronin, 2014). I performed a methodological research method by using an interview protocol, which is described in Appendix B. An interview protocol allows for uniformity of interview questions for all participants (Castillo-Montoya, 2016). I asked all participants the same questions in the same manner, which ensured equitable and reliable interviews. The uniformity of the interview questions also helped me to identify data saturation.

Interviews are one of the most important sources in a case study (Alshenqeeti, 2014). It is important for researchers to be able to conduct interviews that help answer the research questions while asking friendly and nonthreatening questions (Yin, 2014). A researcher develops good interview questions by avoiding bias, listening to the responses, and being flexible while conducting the interview (Roulston & Shelton, 2015). In addition to interviews, many sources can be used to collect evidence such as documents, direct observation, participant observation, archival records, and physical artifacts. As a researcher, my behavior changed to fit the world of the case and the participants to acquire knowledge.

My enthusiasm toward integrating data, software engineering, and bioinformatics drove me toward this study, but I have no personal or professional relationship with the participants. Having no affiliation with the organization ensures the conformability of findings. The conformability of conclusions means that the findings and the participants' data are aligned (Elo et al., 2014). I did not include my opinion while analyzing the data; I made sure that the collected data reflected the real knowledge of the participants and my intervention did not affect the quality of the study.

Ethical standards must be followed whenever a study involves human subjects. A researcher must show how he or she plans to protect his or her human subjects in his case study (Wolf et al., 2015). According to Belmont Report (U.S. Department of Health and Human Services, 1979), the researcher must follow certain ethical standards when conducting a study to ensure that participants are treated ethically and never exploited. These standards include the respect of human subjects, beneficence, nonmaleficence, and justice (Hammer, 2016). Additionally, the Belmont Report (U.S. Department of Health and Human Services, 1979) emphasizes basic ethical principles including autonomy and confidentiality and beneficence that maximize benefits and minimize harms of each participant and justice, which is the fairness in distribution for each participant. Further, informed consent is the participant's ability to choose what shall or not happen to them. I informed participants that they have their right to anonymity and they were able to withdraw at any time from the study even after the data collection process without any consequences. Respecting the participants from all aspects is a high priority for me as a researcher, I made sure that the participants freedom was ensured, and the conducted interview did not have a negative effect on the participants. As a researcher, I respected and protected the rights of the participants and followed the guidelines of the Belmont Report. I also completed the Protecting Human Research Participants training offered by the National Institutes of Health Office of Extramural Research.

To mitigate bias, I set aside my knowledge and biases and ensured transparency. The researcher should also go back to the data to make sure it is accurate and corroborated with other interviews (Elo et al., 2014). Asking participants to review the summary of the interview for accuracy also reduces bias (Yang & Banamah, 2014). I transcribed the interviews and conducted member checking interviews. I was transparent with the participants regardless of my professional skills and knowledge in information technology, software engineering, and data integration.

Participants

The participants for this study are bioinformatics specialists of a research institution in Lebanon that has strategies to integrate distinct bioinformatics datasets. Participants must meet certain criteria to ensure that they are eligible for the study and they have the needed knowledge and experience to answer the interview questions (Castillo-Montoya, 2016). Choosing the appropriate participants is critical for conducting a qualitative research study (DeFeo, 2013). To be eligible to participate, the bioinformatics specialists had to be at least 18 years of age, have a minimum of 2 years of experience in the bioinformatics field, and be currently employed with the research institution. Additionally, participants are useful if they have deep understanding about the researched topic (Reybold, Lammert, & Stribling, 2013). I included only participants that are knowledgeable in integrating distinct bioinformatics datasets. I tried to avoid deselection of participants and chose appropriate participants carefully because deselecting participants can cause potential harm (DeFeo, 2013).

I identified the experts that I needed to interview by navigating the institution bioinformatics department website. Determining whether the participants are eligible or is subjective and requires researchers to choose the participants that they think would add value to the research (Reybold et al., 2013). I got their direct information from a gatekeeper who works at the facility. Gatekeepers endorse researchers' work by using the trust they have with the participants, which facilitate access to participants (Peticca-Harris, deGama, & Elias, 2016). Gatekeepers also provide the researcher with insights on how to get access to the organization (Hoyland, Hollund, & Olsen, 2015). Recruitment can be optimized through communication with the participants. Obtaining access to the interviewee involve sending them an introduction to the study, study benefits, and convenience of the interview process (Hoyland et al., 2015). Obtaining approval from key stakeholders in the institution also helps the researcher to gain access to participants during the planning and designing phase of the study (Peticca-Harris et al., 2016). I explained the importance of the study and the possible outcomes that might positively affect a field that the participants are passionate about. Before conducting my study, I ensured that I had approval from the Walden University Institutional Review Board (IRB; approval no. 12-17-18-0554372.

Collaborating with the participants requires acquiring informed consent, organizing meeting locations and time, determining boundaries, and avoiding any surprise for the participants (Peticca-Harris et al., 2016). Obtaining informed consent from the participant is important for maintaining ethical standards and the quality of the research (Sanjari et al., 2014). I sent the participants an e-mail containing the consent form to inform them about all the aspects of the study. The specialists gave me more insights to explore strategies of integrating distinct datasets. The participants shared with me their expertise and knowledge that gave me more detailed information about my study. I followed up the e-mail invitation with a telephone call to discuss the research and answer any questions.

Research Method

I used a qualitative research method for this study to gain an in-depth knowledge in integration bioinformatics datasets and capture the participants’ views in a real-world setting. A qualitative study is conducted when the concept lacks previous research to obtain knowledge from experts through methods such as interviews and observations (Kahlke, 2014; Marshall & Rossman, 2016; Yin, 2014). A qualitative study is the process of studying a phenomenon throughout the experience of others in a natural setting (Yilmaz, 2013). A qualitative method permits open-ended responses, which is not possible in the quantitative approach (Marshall & Rossman, 2016). A qualitative approach also permits the researcher to analyze the studied phenomenon from the participants' perspectives without any beliefs imposed by the researcher (Kemparaj & Chavan, 2013). A researcher conducting a qualitative study aims to study a small sample of participants who can provide enough information to gain a deep understanding about the studied phenomenon (Yilmaz, 2013). This made the qualitative method appropriate for this study, as there is limited research on the strategies used to integrate distinct bioinformatics datasets.

In a quantitative study, the problem is addressed by understanding what variables influence the outcome to confirm or disconfirm hypotheses (Fassinger & Morrow, 2013). A quantitative approach is used by researchers to test a theory or experiment using statistical methods (Marshall & Rossman, 2016). Trustworthiness is difficult to evaluate in a quantitative study because of a focus on reporting the analysis of the study (Elo et al., 2014). A researcher conducting a quantitative study on a small population uses statistical analysis to test hypotheses and generalize to a large population (Barczak, 2015). The quantitative method was not appropriate for this study because the research question was not be used to confirm or disconfirm a hypothesis. The purpose of this study was to explore the strategies used by bioinformatics specialists to integrate distinct bioinformatics datasets.

The mixed method approach combines the process of collecting and analysis and integrates quantitative and qualitative research designs (Noprisson et al., 2016). The researcher should be aware of both qualitative and quantitative methods to implement mixed-method approach (Osborne & Jones, 2017). It covers collecting, analyzing, and interpreting qualitative and quantitative data in a single study (Barlow et al., 2018). When using the mixed-method approach, the likelihood of anticipated outcomes will be multiplied because the qualitative and quantitative are combined (Mao, 2014). The mixed-method approach was not applicable for this research, because the study did not rely on quantitative methods to define relationships between variables using statistical analysis.

The qualitative approach when using interviews allows researchers to gather and present rich data and provide detailed descriptions of participants in their settings (Alshenqeeti, 2014). This approach gave me the opportunity to answer my research question using a qualitative study. Acquiring information and knowledge using interview questions best suited this study, because the information was in the participants' own words.

Research Design

A qualitative case study approach was used to acquire in-depth knowledge in integrating distinct bioinformatics datasets. A case study is used to relate patterns (Kahlke, 2014). A case study helps the research gain an in-depth description of some social phenomenon (Alshenqeeti, 2014; Yin, 2014). The case study design was chosen in this study to gain an in-depth description of the phenomenon of integrating different datasets. The case study also helps the researcher expand the preposition of a group (Merriam, 2014). Employing the case study design for this study allowed me to identify and link patterns in the strategies used by bioinformatics specialists to integrate distinct datasets and acquire knowledge on the described case.

The narrative design is appropriate when a study focuses on a person (Wiles, Crow, & Pain, 2011). In my study, there was no focus on a person but rather a case. A narrative design is also used to emphasis on the live and culture of a person (White & Drew, 2011), but the life and culture of a person was not discussed in this study. Finally, the narrative design is used to learn biographical information about a person (Malagon-Maldonado, 2014). Therefore, the narrative design was not selected for my study because I was focused on the strategies of bioinformatics specialists rather than strategies of one person.

The phenomenological approach is used to explore the individual experiences of people through methods such as interviewing (Bevan, 2014; O’Gara, Tuddenham, & Pattison, 2018; Yüksel & Yıldırım, 2015). However, this study was not focused on the lived experience of the participants but instead the organizational perspective of the case being studied. In phenomenological designs the researcher describes the studied phenomenon (Sloan & Bowe, 2014), but integrating distinct biological datasets was not a phenomenon. Additionally, in phenomenological designs, understanding the phenomenon is done throughout the individuals experience with that specific phenomenon (Kruth, 2014; O’Gara et al., 2018; Sloan & Bow, 2014). However, exploring in-depth lived experience of the participants did not answer the research question of the study, so the phenomenological design was not selected for this study.

An ethnographic design is focused on the overall understanding of a cultural group including language, behavior, and beliefs (Draper, 2015; Edwards & Kaimal, 2016; Kruth, 2014). The ethnographic design was not selected because the study was not focused on the cultural values and beliefs of the participants. In the ethnographic design, the researchers describe the study from a holistic point of view including patterns, values, beliefs of a culture-sharing group (Draper, 2015; Edwards & Kaimal, 2016; Kruth, 2014). In this study, the aspects of the ethnographic design was not studied; instead, I studied the organizational perspective of integrating bioinformatics datasets. Significant time and effort is also required from the investigator when conducting an ethnographic study (Draper, 2015; Edwards & Kaimal, 2016; Kruth, 2014). Thus, using a case study through interviewing and observing bioinformatics specialists was less time- and effort-consuming.

I performed a qualitative case study to collect data and analyze it and answer the research question of my study. A case study is used to investigate the complexity of an event within a physical world situation and restricted boundaries (Yin, 2014). The case study was the most appropriate in my study because it allowed me to investigate in-depth the strategies of integrating distinct bioinformatics datasets. While conducting the case study, I ensured data saturation, which refers to the point that more sampling will not give any new information (Tran et al., 2016; Tran et al., 2017). I continued to interview participants until I reached data saturation to make sure that no new themes emerged.

Population and Sampling

The population of this study was bioinformatics specialists of a research institution in Lebanon that has strategies to integrate distinct bioinformatics datasets. The population included a group leader and six bioinformatics specialists who form the core unit of the bioinformatics group of the research institute. In a qualitative study, the population characteristics relate the subjective experience of the participants with the phenomenon of the study (Stern, Jordan, & McArthur, 2014).

I obtained the contact information of the participants from the publicly available directory. I applied for permission from the head of the organization via a letter of cooperation. This sampling method best suited me because the examined population is relatively small and having detailed information about all the participants helped answer my research question.

Because the total population that I examined is relatively small, I performed census sampling. Census sampling is studying everyone in a population, with a census providing detailed information about all or most of the participants (Killick et al., 2016). A census involves picking everyone in the population of the study (Lucas, 2014). Census sampling also involves collecting detailed and complete data from the studied population (Kish & Verma, 1986). When the population is small and finite and cost, and time are less of a concern, a census sampling is suitable (Jordan, 2013). The population of my study was small, so applying census sampling was the best to acquire required information about the study.

The richness of the sample is more important than its thickness or its heavy content. Rich data is multi-layered, detailed and nuanced (Fush & Ness, 2015). Having thick data doesn't mean rich, it is about the quality of the collected information. My main objective was to reach data saturation, to the point where no new data emerges. Data saturation is the process of adding new participants to the study to the point where the produced information has little or no effect to the study (Fush & Ness, 2015; Tran et al., 2016; Tran et al., 2017). Data saturation refers to the point that more sampling will not give any new information (Tran et al., 2016; Tran et al., 2017). The bioinformatics department in the research institution has a group leader, and six bioinformatics specialists, so my sample was seven bioinformatics specialists including the group leader. I interviewed all seven of them hoping to reach rich data saturation.

The interview setting should was a suitable, comfortable location away from distraction. For both participants and researchers, an ideal setting is where they could avoid interruption and make an adequate sound recording for their interview, so private place is more suitable than a public one (Goodell, Stage & Cooke, 2016). Background noise should be minimized in an interview setting to avoid participants distraction or interfering with audio recordings that might affect the data collection process (Dikko, 2016; Laura et al., 2016). It was paramount that no aspects within the organization environment affected the participants from sharing any information, like the eyes of their colleagues. Moving the participants away from the stressful work environment and provide for them a suitable pleasant environment may affect the quality of the interview and allow the researcher to acquire suitable knowledge for his study.

Ethical Research

Research that involves human participants raises complex issues; human subjects must be protected and respected during researches. Researchers who use human subjects in their research must follow approved ethical regulations and guidelines (Merriam, 2014). A researcher has the responsibility to maintain ethical standards during his study (Haahr, Norlyk, & Hall, 2014). Following a protocol while conducting a qualitative study is crucial to control any potential ethical challenges between the participants and the researcher (Sanjari et al., 2014). Ethical consideration was kept during my study to ensure the trustworthiness and reliability of the research. I followed a protocol while interviewing to protect the participants.

The consent form included details about the purpose of the study, criterias for participating, the procedure of participating, confidentiality, the voluntary nature of participation, compensation, the benefits of participating in the study, and risks of participation. Participants can withdraw from the study at any time for any reason, no compensation of any kind will be given to the participants, because the participation in this study is volountary.

Participants provided informant consent to participate, to make sure that they are participating in this study freely without any kind of obligations. An informed consent is a requirement in a qualitative study where the researcher must inform the participants about all the aspects of the study (Sanjari et al., 2014). The informed consent informs the participants about the study, risks, benefits, and confidentiality (Judkins-Cohn et al., 2014). Ethical standards require that participants must have free will (Marshal & Rossman, 2016). I obtained approval from the Walden University IRB before conducting interviews and collecting any data. Consent forms are required by IRB in any research where human subjects are involved, and the consent form should be at the reading level of the participants (Ferreria, Buttell, & Ferreria, 2015). I met with the participants to review their rights as outlined in the form of informed consent, explained for them any concerns they might have and finally I asked them to sign the informed consent form which indicates that they will participate in the study willingly and freely.

Before the interview, I made it clear for the participants that they have all the right to withdraw from the study at any time. All participants have the free will to participate or not in research (Marshal & Rossman, 2016). The withdrawal process can be done verbally or by submitting a written request. If any information is collected it would be destroyed and not to be used in the study. Knowing that the participants will face no consequences from withdrawing from the study.

I met with the participants for the interview in a private office inside the laboratory suitable for the interview. People may receive inducement that may influence their decisions and behaviors (Grady, 2012). I did not offer any incentives for the participants for conducting the interviews.

The adequate application of the basic ethical principles in the Belmont report which emphasize on (a) the respect for persons, including anonymity and confidentiality, (b) beneficence, maximizing benefits and minimizing harms of each participants and (c)

Justice, fairness in distribution for each participant (U.S Department of Health on Human Services, 1979). The university IRB guidelines will help me to follow ethical procedures for each participant during my study.

All the electronic information was stored on an encrypted thumb drive and stored in a locked cabinet at my home when not using it. All physical documents are locked in a cabinet. As required by Walden and recommended by Punch (2013), the retention period of the collected data for the study is five years. All electronic and physical information will be destroyed after five years.

The researcher must assign codes to each participant instead of using their names (Babbie, 2015). Each participant received a code such as X1, X2, Xn, only identifiable by the researcher, which will protect the names of the participants. Using codes to identify participants means that the participant's real identity is only revealed to the researcher and their identity is protected from public exposure. Using identifiers to determine the participants will helped ensuring the confidentiality and anonymity of the participants because the identifying information of the participants was removed such as name and addresses.

Data Collection

Instruments

The researcher is the primary data collection instrument (Marshall & Rossman, 2016). In addition to being the primary data collection instrument, I used semi-structured interviews guided by open-ended questions to obtain detailed responses from participants. Interview question is a very important method for collecting data in a qualitative study (Brinkmann, 2014). Using open-ended questions in semi-structured interviews allows participants to more freely and openly express their feelings and ideas (Patton, 2015). The purpose of conducting semi-structured interviews is to obtain responses from the participants that are subjective regarding a specific phenomenon following a detailed interview protocol (McIntosh & Morse, 2015). All of the interview sessions was audio-recorded. Audiotapes provide more accurate transcription of the interview than taking notes (Speer & Stokoe, 2014). Interviews was held to question the participants about strategies to integrate disparate bioinformatics datasets. Researchers in a qualitative study should have a list of questions for face-to-face interviews to answer the research questions (Brinkmann, 2014). The interview was audio recorded, and then I transcribed the recordings and conducted a member checking after the interviews to ensure accuracy. The interview response are stored using NVivo software to analyze the data and find insights. I coded the text to gain a deeper understanding of the patterns and themes (Merriam, 2014).

The secondary data collection instrument is direct observation, it is considered as an important source of evidence in doing a case study research, it helps in understanding the actual use of technology, any problems that might occur, it provides additional information about the studied phenomenon, it helps indicates the culture of the organization, and the researcher also throughout direct observation can assess the occurrence of behaviors in the field (Yin, 2014). Direct observation enables the researcher to widen the focus on the observed phenomenon of the study (Hahlweg et al., 2017). Direct observation gives the researcher the opportunity to record and analyze participants behavior while they occur (Manan & Várhelyi, 2015). Qualitative research methods such as observation is a primary means to understand the interviewee experience of the studied phenomenon (Salloch, 2014). After interviewing the bioinformatics specialists, I visited the laboratory and shadowed the specialists at their work location while they are using the bioinformatics artifacts, analyze datasets and extracting knowledge.

My tertiary data collection instruments were organizational documents, a variety of documents was considered like documents, minutes of meetings, written reports, administrative documents, formal studies and articles appearing in the mass media. Software documentation can provide detailed instructions about the used bioinformatic tool, its features, and the software capability; Bioinformatics scientist also documents their work by recording their work into a local wiki or a notebook, having access to such documents can be informative. Analyzing organizational documents that might involve written or recorded materials provide further understanding about the studied phenomenon (Lincoln & Guba, 1985). Analyzing organizational documents is an important method for studying a research phenomenon in a qualitative study (Islam, 2014). I asked participants to provide me with any documents, logs, engineering journals, multimedia sources and historical documents that might help me to explore strategies to integrate disparate bioinformatics datasets. Selecting a wide variety of resources ensure a deep understanding of the data (Dunne, Pettigrew, & Robinson, 2016). The researcher should assess the quality of the document and choose to adopt or not. The quality of the document depend on the reliability, trustworthiness, and accuracy of the document

(Donaldson, 2016). I also collected copies of the organizational policies and procedures documents and other organizational documents from the organization that might help me to explore strategies to integrate disparate bioinformatics datasets. I kept an annotated bibliography of the organizational documents collected from the participants to facilitate the process of storing and retrieving the documents.

During the interview, participants was asked identical questions in the same order to ensure reliability and trustworthiness. Asking the participants, the identical questions increase the logic, accuracy and fairness (Tucker, Yeow & Viki, 2013). Another approach to ensure reliability and validity is to perform an expert review to my interview questions. Expert review enhances the validity and reliability of the interview questions (Jacob & Ferguson, 2012), it ensures that the interview questions are concise (Castillo-Montoya, 2016). It is a great approach to make sure that my interview questions are adequate to answer my research question. I used member checking to enhance the reliability and validity of my collected data.

Member checking requires from the interviewee to review the summary of the interview for accuracy (Chronister et al., 2014). Member checking is important for assessing the validity of the study. Member checking allows the interviewee to assess the collected data which creates dependability and reliability (Lub, 2015). The process of member checking is where the researcher asks the interviewee to confirm the accuracy of the information collected by the researcher and ask the participants follow up questions for additional clarification (Birt et al., 2016). This process will increase credibility, validity, dependability, and reliability to my study by confirming my interpretation after the interviews, because the collected information is double checked by the participants for validity. I used member checking in my study to increase credibility, reliability, and dependability. I performed member checking throughout a scheduled follow up interviews after my interview with the participants. I e-mailed the participants my analysis of their data before the follow-up interview. I asked the participants during the follow-up meeting to confirm whether my study reflects their actual viewpoints. I iteratively used member checking until my analysis of the collected data indicates the participant's real views, and the participant confirms all my interpretations.

The interview setting where I asked the participants the interview questions is a suitable, comfortable location away from distraction. An ideal interview location should be a quiet place away from distraction (Goodell, Stage & Cooke, 2016). Interview questions are identical and in the same order. Asking the participants, the same questions in the same order enhance the equity and logic of the interview (Tucker, Yeow & Viki, 2013). Observation is another important source of knowledge in a case study (Alshenqeeti, 2014). I performed observation in the laboratory of the research center, to obtain a clear view on how the process of extracting knowledge from biological datasets is done.

Data Collection Technique

For a qualitative case study, many data collection techniques exist. There are six sources of evidence for conducting a case study: (a) documentation, (b) archival records, (c) interviews, (d) direct observation, (e) participant observation and (f) physical artifacts (Yazan, 2015). I conducted semi-structured face-to-face interviews guided by open-ended structured questions. Semi-structured interviews are based on planned structured that are open and flexible to allow open discussion (Wahumi, 2012). The interviews were audio-recorded for more accuracy; audio-recordings helps the researchers catch the original words of the participants while transcribing (Speer & Stokoe, 2014). Observation is also used as a data collection technique for the study. Observation is a great data collection technique for a case study, it allows the researcher to observe the interviewee body language and affect next to the participant verbal answers (Marshall & Rossman, 2016).

Interviews are a great source of knowledge. It is the most important source of evidence when conducting a case study (McIntosh & Morse, 2015). Interviews allow the interviewee to express himself, it is likely to be fluid rather than rigid (Hanna et al., 2016). Interviews are flexible enough to allow open discussion (Wahumi, 2012). Open ended questions during semi-structured interviews allows the researcher to clarify any ambiguity and explore new topics that might emerges during the interview (Laura et al., 2016). A knowledgeable participant provides crucial insights into the discussed topic provide shortcuts for certain situations and helps the researcher identify other sources of evidences (O’Keeffe et al., 2016). On the other hand, participant’s reports are subject to bias, poor recall and inaccurate articulation and a researcher should corroborate the collected interview data with other sources of evidence for reliability and trustworthiness.

Observation is a great evidence for providing additional information next to interview. Observation is helpful for understanding the actual use of a technology, yet for more reliability multiple observations might be required (Marshall & Rossman, 2016). Using interviews and observations as my main sources of evidence helped me gain a deeper understanding about strategies used by bioinformatics specialists to integrate disparate bioinformatics datasets.

A researcher uses a pilot study to refine the planned research questions and procedure to be later used in the formal case. Pilot study is used in a case study to identify issues with the interview questions, it is a great way to ensure validity in a study (Dikko, 2016). I used the pilot study to determine if the interview questions target information viable to answer the research question.

Validating the accuracy of the interview throughout member checking will enhance the trustworthiness of the study. Checking with the participants if the transcribed responses are accurate throughout member checking enhance reliability and validity (Chronister et al. 2014). It is a quality control strategy (Birt et al., 2016). After the interviews, I asked each participant for any documents that can be related to the interview questions. I asked the Gatekeeper for any materials that he can give me access to and can be beneficial for my study. I also conducted member checking by asking the interviewees to interpret their responses and check the transcribed interviews for accuracy. The member checking process enhanced the trustworthiness and credibility of the study.

Data Organization Techniques

The reliability and trustworthiness of the data is enhanced while using organizational technics (Elo et al., 2014; Roer-Strier & Sands, 2015). During the data collection process, I used a reflective journal to record observation to collect deep meaning from the knowledge the participants shared. Reflective journal improved the quality and validity of qualitative data (Vicary, Young & Hicks, 2016).

Labeling is crucial to organizing data and protecting the participant’s identity (Davis, 2013). Using codes by the researcher facilitates labeling and comparison of the collected data (Vaismoradi et al., 2016). Codes were assigned for each participant to keep track of the data and protect the participant’s identity. I transcribed the interview questions into a Microsoft word documents, then the data was uploaded and cataloged into NVivo software which is a computer-aided qualitative data analysis software tool (CAQDAS). The researcher using the NVivo software can code themes, evaluate findings and interpret them (Zamawe, 2015). NVivo helps identifying themes to answer the research question (Robins & Eisen, 2017). I used the NVivo software to continuously evaluate the participants respond data. The study database elements for this study included the interview questions, participant’s responses analysis results from NVivo software, and hand writing notes.

All physical artifacts are held in a locked cabinet at my home. All electronic information are stored on an encrypted thumb drive and locked in a cabinet when not in use, for a period of five years. I will permanently delete all electronic information and destroy and physical artifacts after five years.

Data Analysis Technique

Data analysis is the process of examining, categorizing, tabulating, and testing collected data to produce findings (Noble & Smith, 2014). The data analysis helps the researcher to understand the data by reducing the amount of data throughout grouping into categories (Bengtsson, 2016). Data analysis is the process of grouping the collected data in a meaningful manner for more insights (Noble & Smith, 2014). After data collection, the data needs to be analyzed to retrieve themes and answer the research question. Because I am conducting a case study I used triangulation. Triangulation is the merging of data from different sources to determine the reliability and trustworthiness of findings and to build confidence in the study result (Hussein, 2015). A single method of data analysis is not adequate to sufficient to represent a research phenomenon (Fush & Ness, 2015). Four types of triangulation exist, (a) data triangulation, (b) investigator triangulation, (c) theory triangulation and (d) methodological triangulation (Patton, 2015). Data triangulation consists of using multiple sampling strategies (Carter et al., 2014). Investigator triangulation is when multiple researchers are involved in the analysis process (Hussein, 2015). Theoretical triangulation is when more than one theoretical position is used to analyze the data (Modell, 2015). Methodological triangulation is when multiple data sources are involved (Wahuni, 2012). I found that methodological triangulation is the most appropriate because I am the only data collector, which makes theoretical and investigator triangulation not applicable. Methodological triangulation involves multiple data collection and analysis techniques to add to the studied phenomenon more understanding, reliability, trustworthiness, validity, and accuracy (Fush & Ness, 2015; Hussein, 2015). I used methodological triangulation to analyze the information collected from interviews, reflective journal, and observation. I transcribed the colledcted data using a third party service, validate the transcriptions using a transcription software, and load all the data into NVivo software to define codes and identify themes.

The data was analyzed to answer the research question; the data analysis was based on interview responses and observations. Codes are assigned for each participant to protect their identity; there was an alphanumeric code range that represents each participant and their response to the interview question. In a qualitative study codes are a word or short phrase that symbolically refers to a captured data (Marshall & Rossman, 2016). It is often necessary in a qualitative study to arrange the data in a Computer-assisted qualitative data analysis software CAQDAS, because CAQDAS provide better results than manual analysis (Moylan, Derr, & Lindhorst, 2015). Analysis of the database helped in collecting data; analyzing the data explored the strategies use by bioinformatics specialists to integrate disparate datasets. In a case study the analysis involves preparation, comprehension and interpretation of the collected data (Merriam, 2014). Unlike statistical analysis, a case study analysis depends highly on the researcher's thinking and ability to analyze and discover findings.

NVivo is a CAQDAS that is a recommended by qualitative researchers such as (Yin, 2014). The data which is open ended responses will be systematically analyzed. Open-ended responses are subject to analysis to uncover themes (Wahymi, 2014). Computer software are paramount for analyzing big amounts of data, but the researcher has the important role for defining codes and interpret any observed patterns, computerized software's cannot substitute the general analytic strategy (Yin, 2014). NVivo can be used to visualize data and display it as graphs, reports and maps to audit the emerging themes (Edwards-Jones, 2014). I loaded the collected data into NVivo and visualized the information to find emerging themes. I analyzed the data using NVivo software repeatedly until I found major themes that are aligned with my research questions. The researcher has the main role to collect, interpret and comprehend the data in a case study (Merriam, 2014). I searched for recurring themes to find correspondence between bioinformatics, data integration, collaboration, strategies and TAM.

After collecting the data and analyzing it, the findings was compared with Davis's (1986) TAM. Davis detailed that the technology acceptance is determined by the perceived usefulness, perceived ease of use and attitude toward using. Where perceived usefulness is how much the use of a technology will increase user's productivity, perceived ease of use is the degree of using a technology is free of effort and attitude toward using is influenced by the previous two factors and determines the system usage (Davis, Bagozzi & Warshaw, 1989). TAM has been used widely over the years to predict user's acceptance thanks to the strong relationship between user's perception toward a technology and the actual use of this technology (Venkatesh, Thong & Xu, 2012). Therefore, the collected data from each participant was analyzed to explore factors for efficiently integrating disparate bioinformatics datasets.

Reliability and Validity

Validity is proving that you as a researcher are measuring what you said you are going to measure in your study without researcher bias. It is how well the researcher studies what is intended (Kruth, 2014). A research study needs to be systematic, ethical, and conducted in a rigorous manner (Merriam, 2014). Reliability is the degree of repeating the study in different contexts or settings and having the same outcomes. Qualitative researchers generally reject the validity framework that is commonly accepted by quantitative researchers, and they argue that there are different standards for judging the research quality. Four types of validity exist: (a) construct validity, (b) external validity, (c) internal validity, and (d) reliability (Howleg & Helo, 2014). Yet Guba and Lincoln proposed four criteria's for judging the validity of a qualitative research, credibility, transferability, dependability and confirmability.

Credibility determines if the results of the study are believable, and how accurate they are (Leung, 2015). It ensures that the provided information is accurate (Bengtsson, 2016). The main objective of qualitative research is to gain a deep understanding of the studied phenomenon (Marshall & Rossman, 2016); it is paramount for the study results to be credible and trustworthy, to ensure that the acquired knowledge is viable. Member checking, triangulation, and data saturation are used to ensure credibility in a qualitative study (Houghton et al., 2013). I ensured credibility in the study throughout the process of asking topic related interview questions that answers the research question. The interviewee are specialists in the studied area and they should add credibility to the study. Data saturation, was achieved throughout prolonged engagement with the participants and interview participants until answers are redundant and no new information are introduced. Data saturation is achieved when the collected data does not continue to inform the research question (Kruth, 2014), or when the researcher identifies no new knowledge from the collected data (Houghton et al., 2013). Member checking to make sure that all collected information's are accurate and reflects the opinion of the participants. Member checking is the process of asking the participants to review the summary of the interviews and give feedback regarding accuracy. Confirmation by the participants give the study objectivity and accuracy bringing credibility to the study (Houghton et al., 2013). Triangulation is also paramount to ensure the credibility of the study, I used methodological triangulation to analyze the information collected from interviews, reflective journal and observation.

Transferability is the degree which the study results can be generalized and transferable to a different contexts, situation or settings (Cope, 2014). The transferability of a study requires an in-depth description of the background of the study, the population of the study and the generated results, for other researchers to determine the study transferability to different context with different participants (Connelly, 2016). It is the consistency of the procedures when conducting the same research (Leung, 2015). Transferability can be enhanced if the researcher provide detailed descriptions about the study context allowing the readers to take a decision if the research is transferable or not (Lincoln & Guba, 1985). To determine the transferability of the study the researcher must include adequate details (Yin, 2014). In a qualitative study richly describing the context, situation, setting and participants achieve transferability (Houghton et al., 2013). To ensure transferability in my study I included thick adequate details, so the reader can easily determine if the results are transferable to their research.

Dependability is when the qualitative researcher needs to describe the ever-changing context within his research; in a qualitative study a researcher needs to explain the changes in a setting richly and how those changes affect the results of the study (Houghton et al., 2013). A study achieves dependability if it can be reproduced with equal participant in an alike context (Lincoln & Guba, 1985). Dependability can be obtained when the researcher clarifies the research processes and methods (Yilmaz, 2013). Dependability was addressed by collecting relevant information, asking participants the same questions in the same order, member checking, analyzing the information reliably, establishing a chain of evidence, and including reflexive journal and observation data

Confirmability is the degree which the study results could be confirmed by other researchers and that the participants experience is reflected and not the researcher's interpretation (Wahuni, 2012), it is an indicator of the accuracy and objectivity of the data (Houghton et al., 2013). Confirmability is ensured when chain of evidence, audit trails, and reflexive journal are included (Leung, 2015). To ensure confirmability I included reflexive journals, member checking, observation data, and create a chain of evidence by recording when and from which participant I collected data.

Transition and Summary

In Section 2, details about the study were provided, indicating that the purpose of the study is to explore the strategies used by bioinformatics specialists to integrate disparate bioinformatics datasets. In Section 2 I included details and justification for the chosen research method and design. I discussed sampling plan, data collection instrument, collection procedure and data analysis. All data collection technics was described acting me as a primary data collection instrument. All ethical guidelines as mentioned in the Belmont report were respected when writing the section two, even that no protected groups was targeted as participants. In Section 2 I explained that I used a qualitative case study and gatekeeper sampling to identify participants. I also mentioned in section two that data was collected from interviews and observations and I used NVivo to organize and analyze collected data. I also said that I used methodological triangulation to ensure saturation and plentitude. I mentioned that I maintained validity and reliability throughout including sufficient details, member checking, and reflexive journal. Section three will present study results and recommendations. I will add details about the findings, application to professional practice, implications for social change, conclusions and recommendations for future research.

Section 3: Application to Professional Practice and Implications for Change

Overview of Study

The purpose of this qualitative case study was to explore strategies to integrate distinct bioinformatics datasets. The findings showed methods and tools that the bioinformatics specialists used to encourage knowledge sharing, participation, and best practices to improve pattern extraction from biological data. The data are generated from bioinformatics specialists and organizational documents from an institution located in Beirut. I interviewed six bioinformatics specialists in a research institution in Lebanon, performing member checking and collecting 27 organizational documents to acquire knowledge about the strategies to integrate distinct bioinformatics datasets. Interviewing the bioinformatics specialists helped me acquire a deeper knowledge about the strategies used to integrate distinct bioinformatics datasets, how this process can add efficiency to bioinformatics pattern extraction, and the challenges while integrating distinct bioinformatics datasets.

Presentation of the Findings

The research question that I sought to address was the following: What are the strategies used by bioinformatics specialist to integrate distinct bioinformatics datasets? Answering the research question addresses the problem that some bioinformatics specialist lacks strategies to integrate distinct bioinformatics datasets. For this study, I used semistructured interviews to collect data on the perceived ease of use and perceived usefulness for exploring strategies to integrate distinct bioinformatics datasets. I also reviewed organizational documents related to bioinformatics data integration, pattern extraction, and knowledge sharing. Five main themes emerged from analyzing the collected data. I will review the themes based on the data analysis process in the following sections.

Theme 1: Focus on Integrating Bioinformatics Datasets

Focusing on integrating the data was one of the prominent themes that emerged. The concept was that the submitters of bioinformatics data should unify the format of the data and store them in a centralized location. Homogenizing the data is the biggest problem that most of the interviewees faced. To make sense of the data, bioinformatics specialists need to integrate and normalize the data against each other for more efficiency. All six participants indicated that focusing on integrating the bioinformatics data will ease the process for locating the required datasets and analyze the datasets more efficiently.

My review of the organizational documents confirmed the importance of integrating bioinformatics datasets. Five of the 27 corporate documents supported the theme (see Table 1 for information source metrics). The five organizational documents contained integrated biological data and analysis results from heterogeneous data sources, with the relevant information from different biological databases retrieved to conduct a certain experiment. Two documents contained analysis of metabolic diseases with datasets from different locations using different format.

The process to integrate the downloaded datasets needed to be done before performing the dataset analysis, which emphasizes the theme for more efficieny in the process of retrieving integrating and analyzing distinct bioinformatics datasets. Data integration systems integrate bioinformatics datasets from various resources for more enriched data analysis process. For example, a document represented data integration in a system called Ondex, and a document represented data integration from a system called MultiDataSet. The data integration using these two systems is complicated; the user must be knowledgble with the software to perform the integration to a unified format. There is also a document that represents inconsistencies in data integration, where inconsistencies have been obsereved while performing data integration that affected the quality of data. The process of using sophisticated bioinformatics software and the in-depth knowledge that a user needs to acquire to integrate distinct bioinformatics datasets support the theme that more robust methods are required to locate and integrate diverse bioinformatics datasets.

Table 1

Major Themes of Focus on Integrating Bioinformatics Datasets

Major/Minor Theme	Participant count	Document count
Focus on integrating bioinformatics datasets	6	5
Challenges while integrating bioinformatics datasets	6	3
Strategies used to integrate disparate bioinformatics datasets	5	5
Success while integrating disparate bioinformatics datasets	3	4

Integrating bioinformatics datasets is related to user experience; a less experienced user may not incorporate data properly and may lead to false result discovery while analyzing the data. Participant X1 asserted that when he first started to integrate disparate bioinformatics datasets, the data became more challenging to analyze and extract useful information, but with the experience, he gained knowledge on how to efficiently integrate datasets. Having the datasets standardized in a specific format will eliminate the challenge for less experienced users to successfully integrate bioinformatics datasets. Efficiency in locating and integrating disparate bioinformatics datasets will add more reliability and efficiency while analyzing bioinformatics datasets for results extraction.

Unifying data formats from different locations is also a challenge while integrating bioinformatics datasets. Many bioinformatics databases store the information in a specific format. To work with data retrieved from different locations, the user must standardize the data formats. Participant X1 stated that once the datasets have been downloaded, the datasets from different locations are stored using different formats, which is a huge problem. Participant X4 also stated that the different bioinformatics datasets are stored in different formats in different locations, which increases the complexity of integrating these datasets. Further, Participant X3 stated that integrating bioinformatics datasets is not an easy process because of the different formats and different information about the data; various tools exist, so it is not a straightforward process. To be able to work with all the downloaded files, they need to be converted to a unique format. Thus, standardizing the bioinformatics datasets will eliminate this problem. Participant X2 added that the issues that they face in bioinformatics are normalizing data, regulating the data type and files, and normalizing the data against each other to constrain all the variables and make sense of the collected data.

Theme 1 complements the literature review supporting that the biggest challenge in computational biology is putting together the available and diverse information (Nussinov & Papin, 2015). Additionally, the literature supports the finding that many challenges prevent bioinformaticians from performing effectively like poor quality of the generated data, the sample size, false discovery, lack of novel algorithms for data integration, computational efficiency, data interpretation and visualization (Zhao et al., 2015). Increasing volume and complexity of genomic data makes the process of analyzing variants challenging for researchers with limited bioinformatics skills (Alexander et al., 2017). A bioinformatics user needs to locate the required information scattered in various databases then perform the integration process. Achieving a certain level of biological data unification throughout integrating bioinformatics services and tools reduces time and effort for analyzing translational data and increases the efficiency of knowledge extraction (Bhuvanesh et al., 2016; Dubchak et al., 2014). Integrative approaches enhance the data quality that is scattered on different technology platforms (Zhao et al., 2015). To achieve efficient integration, the user should have expertise, which allows bioinformatics specialists to focus on the analysis and result extraction. Thus, Theme 1, which pertains to a focus on integrating datasets, aligns with the literature review.

Focusing on standardizing bioinformatics datasets also aligns with the conceptual framework of the study regarding ease of use, the usefulness, and user acceptance (Davis et al., 1989). The collected data support the TAM component of ease of use. Perceived ease of use indicates that using a particular system will enhance the user job performance

(Rauniar et al., 2014). Participant X2 talked about how over the years, the biological databases are upgrading their user interfaces to be used by less experienced users. Perceived usefulness indicates that using a system improve performance (Davis et al., 1989). Five of the participants stated that integrating distinct bioinformatics datasets by unifying the data format, files, and location will increase efficiency in the bioinformatics field. Finally, the data supported the TAM component of usefulness; five of the participants emphasized the importance of focusing on standardizing the bioinformatics datasets to enhance usability while retrieving results.

Theme 2: Adding Metadata with the Submitted Bioinformatics Datasets

Adding metadata with the submitted bioinformatics datasets was another prominent theme. The central concept of this theme is that leading organizations in the bioinformatics field must ensure that the submitted data has enough metadata. Bioinformatics metadata ensure clarity, so the bioinformatics users have no misleading information that might affect their experiments. The bioinformatics data providers should increase their use a standardized set of terms to facilitate data integration; the availability of metadata will help the bioinformatics specialist to add more efficiency while integrating distinct bioinformatics datasets. Having a controlled description of the submitted sample starting from the acquisition of raw data to the publications of the result provides reliable metadata. Annotating the submitted bioinformatics data with reliable metadata is an important factor for its structuring, interpretation, and reusability.

This theme was supported by participant responses. Three of the six participants indicated that providing reliable metadata with the submitted bioinformatics datasets would improve the process of locating the scattered information, and with additional information the bioinformatics specialists can integrate the downloaded data more efficiently. Participant X6 asserted that most of time bioinformatics software packages are not supported with a controlled description such as contextual data or part of a PhD or post-doc, so the user gets misleading information while analyzing the data. Having metadata and well-structured documentation will allow the user to efficiently use the information. Participants X3 mentioned that the users of bioinformatics datasets do not understand the submitted data; a submitter might think that a particular filed is self-explanatory, but the user might miss-understand it and the user will interpret it differently. Participant X4 also asserted that the lack of metadata with the submitted data leads to misinterpretation while analyzing the datasets.

My review of the organizational documents confirmed the importance of this theme regarding metadata. Three organizational records support the theme. For example, these documents contain bioinformatics datasets that lack metadata that should be provided by the submitters such as development and growth conditions, genotype, tissues of biological objects, and environmental conditions, which makes these data hard to understand and may lead to misinterpretation while analyzing. But the metadata information is sometimes missing or is described using different vocabularies. Three of the documents contained samples of sequenced bioinformatics datasets with no controlled description or metadata about the data that can be considered essential like information on the samples from collection to sequence generation plus contextual data such as clinical observation and environmental conditions. The submitters must describe sequenced details with these data for more efficiency during analysis and comparisons with other projects. Due to the lack of well-controlled metadata, a lot of time is spent on analyzing incorrect data.

The literature review also supports this theme. Research has indicated that it is important that data providers increase the use of metadata to facilitate data integration (Hassani-Pak & Rawlings, 2017). The associated metadata need to be documented to facilitate comparative analysis and hypothesis generation (Reddy et al., 2015). Additionally, to ensure an effective reuse of data, it has to be enriched with relevant metadata and converted into appropriate format for integrative knowledge management (Marti-Solano et al., 2014). Metadata give the users the ability to look at their data and analyze results from a whole different perspective (Reddy et al., 2015).

Theme 1 also aligns with TAM. For example, PEU is using a particular system will enhance the user job performance (Rauniar et al., 2014). Submitting metadata also aligns with the perceived usefulness concept of TAM, where the user will perform the data analysis more reliably. Perceived usefulness is when using a particular system will improve performance (Davis et al., 1989). Participants suggested that adding metadata for the submitted bioinformatics datasets improves the efficiency in the bioinformatics field. This also aligns with TAM that the efficient use of a system improves performance (Davis et al., 1989). Three of the participants argued that while they are working with bioinformatics datasets, they needed more information on some files and they needed to contact the submitters for the detailed information and that submitting metadata, additional information, and well-structured documentations can increase efficiency while analyzing bioinformatics datasets, which aligns with TAM.

Theme 3: Centralized Bioinformatics Database

Having a centralized bioinformatics database is another theme. Four participants argued that finding reliable datasets for an experiment will require them to visit multiple bioinformatics databases, which is time and effort-consuming and requires individual skills and experience, and having a centralized database can help specialist to locate the bioinformatics data. Standardizing can be done throughout contribution between bioinformatics provider to establish a centralized bioinformatics database where all submitters can submit their information based on defined criteria. This process can help the bioinformatics specialists to easily locate the information. I found a similar emphasis on the theme of centralized bioinformatics database in the organizational documents. Five of the 27 organizational records support the theme (see Table 2). The documents contain the analysis of biological information retrieved from different databases. For example, one of the documents contains biological information for a certain mouse mutation retrieved from the databases BRENDA, IntAct, JASPAR, TRANSPATH and EMBL-Bank, which shows that the bioinformatics specialists have visited multiple bioinformatics databases to download the needed datasets to perform a reliable experiment. The data are divided into different categories then integrated into a standardized format in a complicated process using the R programming language for analysis.

Table 2

Minor Themes of Centralized Bioinformatics Database

Major/Minor Theme	Participant count	Document count
Centralize bioinformatics database		
Challenges in retrieving disparate bioinformatics datasets	4	3
Strategies for having a centralized bioinformatics database	4	5

Participant responses further support this theme. Participant X1 asserted that a specialist working on a particular mutation needs to navigate at least four or five different locations to get the information for those mutations, and each database describes the variance differently, which is frustrating and difficult. Participant X2 mentioned the use of different publicly available data from different databases like ENCODE and BATCH projects, which is a complicated and time-consuming process. Participant X4 also asserted that comparing different datasets is tough because each database stores its information using a specific format. Participant X6 added that every package from a specified location has its own data structure and to comprehend and analyze this data, its composition should be unified. The solution to different details and locating is having a centralized database, so the specialist can locate a needed dataset without navigating to any location for the same purpose.

The theme corroborates with the literature review. Processing data from different sources is a common task for scientists (Lopes & Oliveria, 2015). The growth of biological information requires specialized databases to store, manage, and retrieve data efficiently (Zou et al., 2015). The biggest challenge in computational biology is putting together the available and disparate information (Nussinov & Papin, 2016).

The conceptual framework aligns with the participant’s thoughts from the perspectives of easy to use and usefulness. Five of the six participants agreed that the process of navigating several bioinformatics databases to download information, unify the formats, and integrate the data is challenging and time-consuming. This aligns with TAM that perceived usefulness is the process of using a precise technology to improve performance (Davis, Bagozzi & Warshaw, 1989); Perceived ease of use is using this technology will enhance the user job performance (Rauniar et al., 2014). Perceived usefulness is a component in determining the user acceptance toward a technological innovation (Venkatesh, 2002). The perceived ease of use indicates the practicality of technology (Hui-Fei & Chi-Hua, 2017). Having the participants navigating many locations to retrieve information and then integrate the downloaded datasets formats is not practical reducing the efficiency of the bioinformatics in analyzing datasets by performing complicated effort and time-consuming tasks to integrate disparate bioinformatics datasets which aligns with TAM. The efficient use of a system or technology improves performance (Davis, Bagozzi & Warshaw, 1989).

Theme 4: Resources

Another theme that emerged is resources, resources are crucial in the bioinformatics field because of the huge bioinformatics datasets. The need of high-performance computing to analyze the datasets and decent Internet bandwidth to download and upload large files. Most of the participants discussed, their struggle with the Internet bandwidth, the need of expnsive high process computing, and that storage is a limitation for their work. Participants report that resources are a paramount part of integrating disparate bioinformatics datasets and three organizational documents supported the theme (see Table 3).

Table 3

Minor Themes of Resources

Major/Minor Theme	Participant count	Document count
Resources	3	3
High-performance computing	4	3
Internet Bandwidth	5	2

My review of the organizational documents confirmed the importance of resources. The organizational documents emphasize on the importance of high-performance computing to perform highly complicated analytical process and the importance of bandwidth to manage massive biological datasets. The documents that contains analysis of metabolic diseases emphasize on the size of the analyzed datasets and such datasets requires huge Internet bandwidth to download, high-performance computing to perform analysis and huge storage space to save on disk.

The Internet bandwidth is seen as a big problem; the bioinformatics laboratory is located in a location where the Internet bandwidth is low even for everyday usage. Participants X1, X3, X4, and X6 asserted that from our area, we face a huge challenge in resources, biological datasets are huge, and you need a decent Internet bandwidth to download the datasets. Participant X2 said that in his location, the resources are a massive problem like hardware problems, availability of high computing platform, Internet, and electricity issues. Participant X1 talked about how having a user-friendly centralized bioinformatic tool in the cloud can eliminate the problem of high expensive computational power that is expensive and the complicated installation process. Downloading massive biological datasets is a huge challenge and time consuming for bioinformatics specialists. Five of the participants argued that the Internet is a big problem for them; the Internet breaks while downloading a re-download is required. To download a 50GB file, for example, you need to leave it over the weekend wherein other locations it might take only 30 min, it is time-consuming and delays the specialists progress. The Internet speed is a challenge; a reliable bandwidth is mandatory to support the downloads and uploads of the huge biological files. The study is made in a specific geographical location with reduced Internet bandwidth. Because the Internet bandwidth problem is not a common issue in all places the occurrence of poor Internet bandwidth problem did not often occur in the literature review. It is an essential factor in the bioinformatics field; high-performance computers with clustered processors, high internal bandwidth, and cutting-edge software's are needed (Hong et al. 2013).

High-performance computing in bioinformatics is essential; dealing with massive datasets requires computationally intensive tasks to analyze huge datasets in a timely manner efficiently. Four of the participants stated that high-performance computers are expensive, not available to everyone but essential to perform bioinformatics tasks. The thoughts of the participants complement the literature review in many occurrences; advancement of high-performance computing, aligned with new cloud computing solutions created a new scope for the applications in today's science (Miller, Zhu &

Bromberg, 2017). Advancement in high-performance computing is having a positive impact on bioinformatics tools (Miller, Zhu & Boomberg, 2017). Bioinformaticians require a high-performance computer with a clustered processor, high internal bandwidth to fast storage, and software to carry out a multiple step workflow (Hong et al., 2013). Traditional high-performance computer resources are expensive both in purchase and maintenance (Miller, Zhu & Bromberg, 2017).

The theme of resources aligns with TAM from the perspective of usefulness, where the availability of resources increases the efficiency of integrating biological datasets. The perceived usefulness is using a system or technology to improve the performance (Davis, Bagozzi & Warshaw, 1989). The perceived usefulness is a component in determining the user acceptance toward a technological innovation (Venkatesh, 2002). Perceived usefulness and the perceived ease of use affect the user attitude which is the degree to which a user is interested in using a particular system and it determines the behavioral intention that leads to actual system use (Tzafilkou & Protogeros, 2017). Lack of resources align with TAM from the perspective of efficiency and usefulness, struggling with resources will decrease the job performance of the bioinformatics users' which will affect their attitude and finally their efficiency in integrating disparate bioinformatics datasets and extract information.

Theme 5: Bioinformatics Tools

Another theme that emerged is the bioinformatics tools. The concept was that the tools of bioinformatics are incredibly complicated and needs a lot of time and effort to be mastered by the bioinformatics specialists. Bioinformatics tools require a significant amount of computational power and a complicated installation procedure. Tools that facilitate the communication amd collaboration between bioinformatics specialists can add some efficiency. Most of the bioinformatics tools require in-depth knowledge and countless hours of practice to be mastered by a user. Developing friendly and simple user interfaces to allow less experienced users to work with complex, disparate datasets and to integrate those datasets into their research increases the efficiency in the bioinformatics field. Having centralized bioinformatics tools on the cloud can eliminate some of the complexities like installation and local computational power.

All of the six participants described their use and struggle with bioinformatics tools to integrate disparate bioinformatics datasets Participant X1, X2, X3 and X6 asserted that to integrate disparate bioinformatics datasets they use online scripts, or they develop their own script using different programming languages like Perl, Python, R and shell scripting in Linux. Also, participant X4 asserted that he uses a number of local or publicly available software's' to efficiently integrate, interpret and to perform analysis. Which indicates that the process is not straightforward, and a specialist needs expertise and knowledge in many programming languages and software's to perform such a task. Participant X6 said that the use of a single bioinformatics tool doesn't provide you with all the needs to perform integration and analysis, the use of multiple tools is a must. Participant X1 asserted that he and his colleagues develop some UI's that can allow other researchers to integrate complex bioinformatics datasets into their work. Participant X3 said that different tools for integrating disparate bioinformatics datasets exists, and the use of different tools might give you different results. Participants talked about bioinformatics open-source and commercial tools and how commercial tools are best suited for bioinformatics specialists. Participant X5 asserted that bioinformatics tools availability is doubted, commercial tools are expensive and open-source tools require a lot of training and skills, a bioinformatics specialist needs to find a trade between complexity and user-friendliness. Participant X5 also argued that the bioinformatics users' needs to be trained for every emerging bioinformatics tool which is a limitation. Participant X6 cited that using commercial tools is a plus, it has a closed environment, no heavy lifting, it is easier and more direct to use than open-source software.

Three of the six participants argued that having more friendly user interfaces will add more efficiency in the bioinformatics field. Participant X2 talked on how over the years, the biological databases are upgrading their user interfaces to be used by less experienced user's. Participant X4 asserted that a platform that facilitates the dialog between the submitters and users could overcome many challenges that we might face as bioinformatics specialists when they are not provided with sufficient metadata and well-structured documentation to completely understand all the aspects of the submitted bioinformatics datasets. I found similar acknowledgment on the importance of bioinformatics tools in the revised organizational documents. Three organizational records support this theme. These documents contain the analysis of bioinformatics datasets using sophisticated tools, those tools are not user-friendly, and to perform a simple task requires significant expertise from the users.

Having user-friendly tools aligns with the literature. The main goal of bioinformatics tools is to provide powerful visualization with a simple interface (Ewels et al, 2016). With the increasing number of bioinformatics tools, experts are eager for standardization to improve efficiency (Lopes & Oliveira, 2015). Specialists can centralize bioinformatics tools using a user-friendly interface and web services to overcome the technical challenges in Bioinformatics (Velloso et al., 2015). Bioinformatics software tools often require complex installation procedures and high processing power (Velloso, Vialle & Ortega, 2015). Users face many challenges to get those tools installed, running and produce results (Velloso, Vialle & Ortega, 2015). A user-friendly resource to visualize and analyze high-throughput data is a powerful medium for specialists to obtain meaningful output for better knowledge discovery (Chen, Tripathi & Mizuguchi, 2016). Bioinformatics tools using a friendly user interface can be extremely beneficial and ease up the specialists' job to analyze biological data and extract useful knowledge (Velloso, Vialle & Ortega, 2015).

The theme of bioinformatics tools aligns with TAM in that having an easy to use bioinformatics tool that allows a user to perform his task in a free of effort complication free environment aligns with the perceived ease of use of TAM. PEU is using a particular system will enhance the user job performance (Rauniar, Rawki, Yang & Johnson, 2014). Having a user-friendly bioinformatics tools that does not require programming skills or advanced knowledge in the software will improve the user performance which aligns with the perceived usefulness of TAM. PU is when using a particular system will improve performance (Davis, Bagozzi & Warshaw, 1989).

Applications to Professional Practice

The specific IT problem that formed the basis of this research was that some bioinformatics specialists lack strategies to integrate disparate bioinformatics datasets. Participants in this research provided strategies that bioinformatics specialists and bioinformatics providers and databases could apply to increase efficiency in the bioinformatics field. The majority of the participants stated that exploring new approaches to integrate disparate bioinformatics datasets and using centralized techniques could eliminate many challenges that might face while locating and analyzing bioinformatics datasets for result extraction. After evaluating the collected data, I identified four primary themes: Focus on integrating bioinformatics datasets, centralized bioinformatics database, adding metadata with the submitted bioinformatics datasets, and resources.

The findings have confirmed that there is a significant problem identifying, locating, and downloading bioinformatics datasets needed for a particular experiment, because of the significant number of bioinformatics databases that exist. Even after downloading the required datasets, each dataset is stored in a format and contains different variances. The datasets need to be integrated into a unique format to be analyzed for result extraction. Most of the participants argued that this is a complicated, frustrating and time-consuming process and exploring strategies to integrate disparate bioinformatics datasets will eliminate those challenges and help the bioinformatics specialists to focus their effort and time no analyzing datasets and not to be bothered by the complicated process of locating, downloading and integrating bioinformatics datasets.

Integrating disparate bioinformatics datasets is a complicated IT practice that bioinformatics needs to struggle with software and coding to perform. Exploring new strategies to integrate can enhance the efficiency of this process and improve IT practice. Bioinformatics providers should organize the documentation from the raw data to the submitted results to eliminate ambiguity for more efficiency in analyzing datasets and reduced faulty results. Bioinformatics providers should consider a communication and collaboration platform to ease the communication and knowledge sharing between the submitters of bioinformatics datasets and the users to eliminate any confusion about the submitted data. Bioinformatics leaders should focus on providing tools that concentrate on standardizing bioinformatics datasets. This step can enhance the efficiency of integrating bioinformatics datasets from different databases and eliminated the sophisticated and frustrating in-house integration process for bioinformatics specialists.

Bioinformatics leaders should focus on building a centralized database that stores the bioinformatics datasets using a unified format and details. This can eliminate the challenge of locating the information, because of the significant number of bioinformatics databases that exist. It will also reduce the complex process of integrating datasets into a unified format because the datasets will be stored in a centralized database using a blended format. It will also help the bioinformatics software builder on focusing on building their software for a particular format and not include many formats. The centralized bioinformatics database is an integral part of the overall strategies to integrate disparate bioinformatics datasets.

Implications for Social Change

My expectation for this research on social change would include bioinformaticians, bioinformatics providers, the health care field, and society. Exploring the strategies to integrate disparate bioinformatics datasets has a wide range of implications. Bioinformatics specialist that integrate different bioinformatics datasets will realize more efficiency, reduced workload, less frustration while integrating data and they will be able to focus more on their work which is extracting results from the biological datasets. Eliminating the complex process of locating and integrating disparate bioinformatics datasets can increase the efficiency for the specialists to center their effort on the analysis process which will have positive implications on the bioinformatics field. Bioinformatics providers will realize a reduction in costs of both developing bioinformatics solutions throughout focusing on analyzing standardized datasets and less time spent on testing and finalizing the product as well as the productivity increase associated with increased user efficiency. Having disparate bioinformatics datasets stored in different locations and using different formats requires the bioinformatics providers a lot of resources, effort and time to be able to develop efficient solutions that integrate those datasets and extract valuable information.

The impact of social change exists outside the bioinformatics field. The effect is extended to the biology and health-care fields. Bioinformatics has the primary purpose of understanding biological data throughout the analysis and interpretation of these data. Producing meaningful information from biological data had an essential part in many areas of biology. In the genetics field bioinformatics had its role in sequencing genomes and observed mutations, bioinformatics played a significant role in organizing querying biological data. Also, bioinformatics helped in understanding molecular biology. Any improvement in the bioinformatics field affect the biology field and the health care field as a side-effect, and positive influence on healthcare has a positive impact on society because it directly affects a significant number of patients.

Recommendations for Action

I explored strategies to integrate disparate bioinformatics datasets. The study findings showed an environment that promotes well-structured documentation and metadata with the submitted bioinformatics data to reduce data ambiguity and allow the users of the bioinformatics datasets to use the datasets efficiently without the need to contact the providers for extra information on some fields. I recommend bioinformatics leaders to focus work on tools that concentrate on standardizing bioinformatics datasets for more efficiency in integrating disparate bioinformatics datasets. The findings of the study also emphasized on the challenging process of locating bioinformatics datasets that are scattered over various locations and stored using many formats. Leaders should focus on building a centralized database that stores bioinformatics datasets in a unified manner, so a bioinformatics specialist can quickly locate bioinformatics dataset and start focusing on analyzing the bioinformatics datasets for result extraction without wasting time on effort on the complicated process of finding and integrating disparate bioinformatics datasets.

In general, this study might be beneficial to bioinformatics specialists, bioinformatics leaders, and health-care. It is essential for bioinformatics leaders to define roles, responsibilities, and expectations for bioinformatics specialists while locating and integrating bioinformatics datasets for more efficiency in this field; and to allow the bioinformatics specialist to focus solely on extracting information from bioinformatics datasets for their experiments and lab work.

Recommendations for Further Study

Several limitations of this study included recalling further research. The chosen methods, design, participants, data collection, the institution, and other aspects of the research enforced limitations on the results. One of the limitations is the potential influence of bias due to the subjective nature of qualitative studies. Another limitation is that the investigation is limited to a single organization in one location. I recommend additional qualitative studies in different organizations in different places and to compare the findings with the findings of this research. The study findings were restricted to bioinformatics specialists because of the limited participant criteria of the study population. I recommend additional qualitative researchers to include software developers, database administrators, and others that might be involved in developing bioinformatics software's and data integration techniques. Another limitation is that the data collection was limited to some techniques I recommend additional qualitative studies with expanded data collection to include more methods. Finally, this study was limited to a single case institution restricting the generalizability outside the case of the institution. I recommend studies to determine if the findings of this study are generally outside the studied case.

Reflections

The doctoral journey was a hard journey filled with difficulties, but it was also filled with sophistication, literacy, and knowledge. At every obstacle, I focused on the light at the end of the tunnel, and I was able to overcome the challenge and expand my expertise throughout the process. Conducting a doctoral degree thought me how to do academic research, write academically, recruit participants, collect qualitative data, analyze research, and benefit from other knowledge to support my research. The doctoral journey changed me as a person from being a software developer not too interested in research to a passionate individual in conducting academic studies. Finally, I did my best to ensure the credibility and reliability of the study. I have learned a lot from the participants, and I hope my study contributes in the academic literature and open perspectives for other studies that are related to my findings.

Summary and Study Conclusions

Bioinformatics specialists require a significant amount of collaboration to integrate disparate bioinformatics datasets efficiently. Well-structured documentation should be added with the submitted bioinformatics datasets. Bioinformatics leaders should focus on techniques to incorporate diverse bioinformatics datasets and build a centralized database where the biological data are stored in a normalized manner. Acquiring those methods can help bioinformatics specialist to focus on the process of extracting results from biological information without being bothered with the complicated process of integrating disparate bioinformatics datasets.

References

An Overview of the Human Genome Project. (2016). Retrieved from https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/

Abdullah, F., & Ward, R. (2016). Developing a general extended technology acceptance model for e-learning (GETAMEL) by analysing commonly used external factors. *Computers in Human Behavior, 56*, 238–256. doi:10.1016/j.chb.2015.11.036

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. Retrieved from https://www.elsevier.com

Alkali, A. U., & Mansor, N. N. A. (2017). Interactivity and trust as antecedents of e-training use intention in Nigeria: A structural equation modelling approach. *Behavioral Sciences (2076-328X), 7*(3), bs7030047–bs7030047. doi:10.3390/bs7030047

Alexander, J., Mantzaris, D., Georgitsi, M., Drineas, P., & Paschou, P. (2017). Variant ranker: A web-tool to rank genomic data according to functional significance. *BMC Bioinformatics*, 181–189. doi:10.1186/s12859-017-1752-3

Alshenqeeti, H. (2014). Interviewing as a data collection method: A critical review. *English Linguistics Research, 3*(1). doi:10.5430/elr.v3n1p39

Ang, L.-M., & Seng, K. P. (2016). Big sensor data applications in urban environments. *Big Data Research, 4*, 1–12. doi:10.1016/j.bdr.2015.12.003

Ardeshirdavani, A., Souche, E., Dehaspe, L., Van Houdt, J., Vermeesch, J. R., & Moreau, Y. (2015). NGS-Logistics: data infrastructure for efficient analysis of NGS sequence variants across multiple centers. *BMC Bioinformatics, 16*(2), A10.

doi:10.1186/1471-2105-16-S2-A10

Babbie, E. R. (2015). *Observing ourselves: Essays in social research.* Long Grove, IL: Waveland Press.

Bach, M. P., Čeljo, A., & Zoroja, J. (2016). Technology acceptance model for business intelligence systems: Preliminary research. *Procedia Computer Science*, *100*, 995–1001. doi:10.1016/j.procs.2016.09.270

Barczak, G. (2015). Publishing qualitative versus quantitative research. *Journal of Product Innovation Management, 32*(5), 658. doi:10.1111/jpim.12277

Barlow, E., Öztürk, D. T., Revie, M., Akartunalı, K., Day, A. H., & Boulougouris, E. (2018). A mixed-method optimisation and simulation framework for supporting logistical decisions during offshore wind farm installations. *European Journal of Operational Research, 264*(3), 894–906. doi:10.1016/j.ejor.2017.05.043

Basaldella, M., Furrer, L., Tasso, C., & Rinaldi, F. (2017). Entity recognition in the biomedical domain using a hybrid approach. *Journal of Biomedical Semantics, 8*, 51. doi:10.1186/s13326-017-0157-6

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open, 2*, 8–14. doi:10.1016/j.npls.2016.01.001

Berg, P. (2008). Moments of discovery. *Annual Review of Biochemistry, 77*(1), 15–44. doi:10.1146/annurev.biochem.76.051605.153715

Bevan, M. T. (2014). A method of phenomenological interviewing. *Qualitative Health Research, 24*(1), 136-144. doi:10.1177/1049732313519710

Berger, B., Daniels, N. M., & Yu, Y. W. (2016). Computational biology in the 21st century: Scaling with compressive algorithms. *Communications of the ACM, 59*(8), 72-80. doi:10.1145/2957324

Bhuvaneshwar, K., Belouali, A., Singh, V., Johnson, R. M., Song, L., Alaoui, A., . . . Madhavan, S. (2016). G-DOC Plus – an integrative bioinformatics platform for precision medicine. *BMC Bioinformatics, 17*, 193. doi:10.1186/s12859-016-1010-0

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member checking: A tool to enhance trustworthiness or merely a nod to validation? *Qualitative Health Research, 26*(13), 1802–1811. doi:10.1177/1049732316654870

Bozgo, V., Hysi, L., & Hoda, A. (2017). Characterization of human CRB gene product by the use of bioinformatic tools. *Albanian Journal of Agricultural Sciences*, 347–351. Retrieved from https://sites.google.com/a/ubt.edu.al/rssb/

Brinkmann, S. (2014). *Interview*. In T. Teo (Ed.), *Encyclopedia of critical psychology* (pp. 1008–1010). New York, NY: Springer.

Brezavšček, A., Šparl, P., & Žnidaršič, A. (2014). Extended technology acceptance model for SPSS acceptance among Slovenian students of social sciences. *Organizacija, 47*(2), 116-127. doi:10.2478/orga-2014-0009

Campbell, J. I., Aturinda, I., Mwesigwa, E., Burns, B., Santorino, D., Haberer, J. E., . . . Siedner, M. J. (2017). The technology acceptance model for resource-limited settings (TAM-RLS): A novel framework for mobile health interventions targeted to low-literacy end-users in resource-limited settings. *AIDS and Behavior, 21*(11), 3129–3140. doi:10.1007/s10461-017-1765-y

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum, 41*(5), 545–547. doi:10.1188/14.ONF.545-547

Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. *The Qualitative Report, 21*(5), 811–830. Retrieved from http://nsuworks.nova.edu

Chen, Y.-A., Tripathi, L. P., & Mizuguchi, K. (2016). An integrative data analysis platform for gene set analysis and knowledge discovery in a data warehouse framework. *Database*, 2016, baw009–baw009. doi:10.1093/database/baw009

Chen, Y.-A., Tripathi, L. P., Dessailly, B. H., Nyström-Persson, J., Ahmad, S., & Mizuguchi, K. (2014). Integrated pathway clusters with coherent biological themes for target prioritisation. *Plos One, 9*(6), 1–11. doi:10.1371/journal.pone.0099030

Cheung, R., & Vogel, D. (2013). Predicting user acceptance of collaborative technologies: An extension of the technology acceptance model for e-learning. *Computers & Education, 63*(0), 160–175. doi:10.1016/j.compedu.2012.12.003

Chrimes, D., & Zamani, H. (2017). Using distributed data over HBase in big data analytics platform for clinical services. *Computational & Mathematical Methods In Medicine*, 1-16. doi:10.1155/2017/6120820

Chronister, K. M., Marsiglio, M. C., Linville, D., & Lantrip, K. R. (2014). The influence of dating violence on adolescent girls' educational experiences. *The Counseling Psychologist, 42*(3), 374–405. doi:10.1177/0011000012470569

Connelly, L. M. (2016). Understanding research: Trustworthiness in qualitative research. *MedSurg Nursing, 25*(6), 435–436. Retrieved from https://www.amsn.org/

Cope, D. G. (2014). Methods and meanings: Credibility and trustworthiness of qualitative research. *Oncology Nursing Forum, 41*(1), 89–91. doi:10.1188/14.ONF.89-91

Cronin, C. (2014). Using case study research as a rigorous form of inquiry. *Nurse Researcher, 21*(5), 19–27. doi:10.7748/nr.21.5.19.e1240

Date, H., Ramaswamy, R., & Gangwar, H. (2015). Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. *Journal of Enterprise Information Management, 28*(1), 107–130. doi:10.1108/JEIM-08-2013-0065

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. Retrieved from https://www.misq.org

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982. doi:10.1287/mnsc.35.8.982

DeFeo, D. J. (2013). Toward a model of purposeful participant inclusion: Examining deselection as a participant risk. *Qualitative Research Journal, 13*(3), 253–264. doi:10.1108/qrj-01-2013-0007

Dellureficia, A. J. (2015). A biography of Paul Berg: The recombinant DNA: Controversy revisited. *Quarterly Review of Biology, 90*(4), 420–421. Retrieved from https://www.journals.uchicago.edu

Demirci, M. D. S., & Allmer, J. (2017). Improving the quality of positive datasets for the establishment of machine learning models for pre-microRNA detection. *Journal of Integrative Bioinformatics, 14*(2). doi:10.1515/jib-2017-0032

Dikko, M. (2016). Establishing construct validity and reliability: Pilot testing of a qualitative interview for research in Takaful (Islamic Insurance). *Qualitative Report*, *21*(3), 521–528. Retrieved from http://nsuworks.nova.edu

Donaldson, D. R. (2016). The digitized archival document trustworthiness scale. *International Journal of Digital Curation*, *11*(1), 252–270. doi:10.2218/ijdc.v11i1.387

Dogruel, L., Joeckel, S., & Bowman, N. D. (2015). The use and acceptance of new media entertainment technology by elderly users: development of an expanded technology acceptance model. *Behaviour & Information Technology*, *34*(11), 1052-1063. doi:10.1080/0144929X.2015.1077890

Dubchak, I., Balasubramanian, S., Wang, S., Meyden, C., Sulakhe, D., Poliakov, A., . . . Gilliam, T. C. (2014). An Integrative Computational Approach for Prioritization of Genomic Variants. *Plos One*, *9*(12), e114903. doi:10.1371/journal.pone.0114903

Dunne, B., Pettigrew, J., & Robinson, K. (2016). Using historical documentary methods to explore the history of occupational therapy. *British Journal of Occupational Therapy*, *79*(6), 376–384. doi:10.1177/0308022615608639

Edwards-Jones, A. (2014). Qualitative data analysis with NVIVO. *Journal of Education for Teaching*, *40*(2), 193–195. doi:10.1080/02607476.2013.866724

Edwards, J., & Kaimal, G. (2016). Using meta-synthesis to support application of qualitative methods findings in practice: A discussion of meta-ethnography, narrative synthesis, and critical interpretive synthesis. *The Arts in Psychotherapy*, *51*, 30–35. doi:10.1016/j.aip.2016.07.003

Emery, L. R., & Morgan, S. L. (2017). The application of project-based learning in bioinformatics training. *PLOS Computational Biology*, *13*(8), 1–8. doi:10.1371/journal.pcbi.1005620

Elo, S., Kaariainen, M., Kanste, O., Polkki, T., Utriainen, K., & Kyngas, H. (2014). *Qualtiative content analysis: A focus of trustworthiness*. Sage Open, 1-10. doi:10.117/2158244014522633.

Ethier, J.-F., Curcin, V., Barton, A., McGilchrist, M. M., Bastiaens, H., Andreasson, A., … Burgun, A. (2015). Clinical Data Integration Model. *Methods of Information in Medicine*, *54*(1), 16–23. doi:10.3414/ME13-02-0024

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. doi:10.1093/bioinformatics/btw354

Ferreria, R., Buttell, F., & Ferreria, S. (2015). Ethical considerations for conducting disaster research with vulnerable populations. *Journal of Social Work Values and Ethics*, *12*(1), 379–384. Retrieved from http://jswve.org/

Fletcher, J., Sarkani, S., & Mazzuchi, T. A. (2014). A Technology Adoption Model for Broadband Internet Adoption in India. *Journal of Global Information Technology Management*, *17*(3), 150–168. doi:10.1080/1097198X.2014.951294

Fush, P.I., & Ness, L.R. (2015). Are we there yet? Data saturation in qualitative research.

The Qualitative Report, 20(9), 1408-1416. Retrieved from http://nsuworks.nova.edu

Garavand, A., Mohseni, M., Asadi, H., Etemadi, M., Moradi-Joo, M., & Moosavi, A. (2016). Factors influencing the adoption of health information technologies: a systematic review. *Electronic Physician*, *8*(8), 2713-2718. doi:10.19082/2713

"Genetics and Genomics Timeline". (2004). Retrieved from http://www.genomenewsnetwork.org

Goodell, L. S., Stage, V. C., & Cooke, N. K. (2016). Practical Qualitative Research Strategies: Training Interviewers and Coders. *Journal of Nutrition Education and Behavior*, *48*(8), 578–585.e1. doi:10.1016/j.jneb.2016.06.001

Green, E. D., Watson, J. D., & Collins, F. S. (2015). Twenty-five years of big biology. *Nature*, *526*(7571), 29–31. doi:10.1038/526029a

Greene, C. S., Tan, J., Ung, M., Moore, J. H., & Cheng, C. (2014). Big data bioinformatics. *Journal Of Cellular Physiology*, *229*(12), 1896-1900. doi:10.1002/jcp.24662

Greene, A. C., Giffin, K. A., Greene, C. S., & Moore, J. H. (2016). Adapting bioinformatics curricula for big data. *Briefings In Bioinformatics*, *17*(1), 43-50. doi:10.1093/bib/bbv018

Hahlweg, P., Härter, M., Nestoriuc, Y., & Scholl, I. (2017). How are decisions made in cancer care? A qualitative study using participant observation of current practice. *BMJ Open*, *7*(9). doi:10.1136/bmjopen-2017-016360

Hammer, M. J. (2016). Informed consent in the changing landscape of research.

Oncology Nursing Forum, *43*(5), 558. doi:10.1188/16.onf.558-560

Hamid, A. A., Razak, F. Z. A., Bakar, A. A., & Abdullah, W. S. W. (2016). The Effects of Perceived Usefulness and Perceived Ease of Use on Continuance Intention to Use E-Government. *Procedia Economics and Finance*, *35*, 644–649. doi:10.1016/S2212-5671(16)00079-4

Hamilton, S. P., & Kreuzer, M. P. (2018). The Big Data Imperative: Air Force Intelligence for the Information Age. *Air & Space Power Journal*, *32*(1), 4-20. Retrieved from http://www.airuniversity.af.mil/ASPJ/

Hanna, K., Anna-Maija, P., Martin, J., & Mari, K. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954–2965. doi:10.1111/jan.13031

Hansen, J., Saridakis, G., & Benson, V. (2018). Risk, trust, and the interaction of perceived ease of use and behavioral control in predicting consumers' use of social media for transactions. *Computers in Human Behavior*, 80, 197–206. doi:10.1016/j.chb.2017.11.010

Hassani-Pak, K., & Rawlings, C. (2017). Knowledge Discovery in Biological Databases for Revealing Candidate Genes Linked to Complex Phenotypes. *Journal of Integrative Bioinformatics*, *14*(1). doi:10.1515/jib-2016-0002

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72–80.

doi:10.1016/j.ijpe.2014.04.018

Hill, R. J., Fishbein, M., & Ajzen, I. (1977). Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. *Contemporary Sociology, 6(2*), 244. doi:10.2307/2065853

Hoßfeld, U., Jacobsen, H.-J., Plass, C., Brors, B., & Wackernagel, W. (2017). 150 years of Johann Gregor Mendel's "Versuche über Pflanzen-Hybriden." *Molecular Genetics & Genomics*, pp. 1–3. doi: 10.1007/s00438-016-1254-4

Hoff, K. A., & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407-434. doi:10.1177/0018720814547570

Holvitie, J., Leppanen, V., & Hyrynsalmi, S. (2014). Technical debt and the affect of agile software development practices on it - An industry practitioner survey. *2014 Sixth International Workshop on Managing Technical Debt*. doi:10.1109/MTD.2014.8

Hong, H., Zhang, W., Shen, J., Su, Z., Ning, B., Han, T. & Tong, W. (2013). Critical role of bioinformatics in translating huge amounts of next-generation sequencing data into personalized medicine. *Science China. Life Sciences, 56(2)*, 110-118. doi:10.1007/s11427-013-4439-7

Hoyland, S., Hollund, J. G., & Olsen, O. E. (2015). Gaining access to a research site and participants in medical and nursing research: A synthesis of accounts. *Medical Education*, *49*(2), 224–232. doi:10.1111/medu.12622

Hu, W., Zaveri, A., Qiu, H., & Dumontier, M. (2017). Cleaning by clustering:

methodology for addressing data quality issues in biomedical metadata. *BMC Bioinformatics*, *18*(1), 415. doi:10.1186/s12859-017-1832-4

Huang, C.-C., & Lu, Z. (2016). Community challenges in biomedical text mining over 10 years: success, failure and the future. *Briefings in Bioinformatics*, *17*(1), 132–144. doi:10.1093/bib/bbv024

Hui-Fei, L., & Chi-Hua, C. (2017). Combining the Technology Acceptance Model and Uses and Gratifications Theory to examine the usage behavior of an Augmented Reality Tour-sharing Application. *Symmetry (20738994), 9*(7), 1-22. doi:10.3390/sym9070113

Hussein, A. (2015). The use of Triangulation in Social Sciences Research: Can qualitative and quantitative methods be combined? *Journal of Comparative Social Work*, *4*(1). Retrieved from https://doaj.org

Hussein, Z. (2017). Leading to Intention: The Role of Attitude in Relation to Technology Acceptance Model in E-Learning. *Procedia Computer Science*, *105*, 159–164. doi:10.1016/j.procs.2017.01.196

Houghton, C., Casey, D., Shaw, D., & Murphy, K. (2013). Rigour in qualitative case-study research. (2013). *Nurse Researcher*, *20*(4), 12–17. doi:10.7748/nr2013.03.20.4.12.e326

Inzalkar, S. M., & Sharma, J. (2015). A Survey on Text Mining- techniques and application. *International Journal of Research In Science & Engineering*. Retrieved from http://ijrise.org

Islam, M. M. (2014). Dealing with Qualitative Methods in Bangladesh: The Potentials and Pitfalls of Research Design and Fieldwork. *Oriental Anthropologists*, *14*(1), 13-26. Retrieved from http://www.printspublications.com/.

Islam, A. K., Azad, N., Mäntymäki, M., & Islam, S. M. (2014). TAM and E-learning Adoption: A Philosophical Scrutiny of TAM, Its Limitations, and Prescriptions for E-learning Adoption Research. *IFIP Advances in Information and Communication Technology*, 164-175. doi:10.1007/978-3-662-45526-5_16

Jones, K. K., Zenk, S. N., Tarlov, E., Powell, L. M., Matthews, S. A., & Horoi, I. (2017). A step-by-step approach to improve data quality when using commercial business lists to characterize retail food environments. *BMC Research Notes*, 101-12. doi:10.1186/s13104-016-2355-1

Jordan, H. S. (2013). Maximizing sampling efficiency. *Applied Mathematics*, *4*(11), 1547. doi:10.4236/am.2013.411209

Judkins-Cohn, T. M., Kielwasser-Withrow, K., Owen, M., & Ward, J. (2014). Ethical principles of informed consent: Exploring nurses' dual role of care provider and researcher. *Journal of Continuing Education in Nursing*, *45*(1), 35–42. doi:10.3928/00220124-20131223-03

Jun, C.-J., Lee, J.-H., & Jeon, I.-S. (2014). Research about factor affecting the continuous use of cloud storage service: user factor, system factor, psychological switching cost factor. *Journal of Society for E-Business Studies*, *19*(1). doi:10.7838/jsebs.2014.19.1.015

Kahlke, R. M. (2014). Generic Qualitative Approaches: Pitfalls and Benefits of Methodological Mixology. *International Journal of Qualitative Methods, 13*(1), 37-52. doi:10.1177/160940691401300119

Kemparaj, U., & Chavan, S. (2013). Qualitative research: A brief description. *Indian Journal of Medical Sciences*, *67*(3-4), 89–98. doi:10.4103/0019-5359.121127

Khan, S., Ullah, M. W., Siddique, R., Nabi, G., Manan, S., Yousaf, M., & Hou, H. (2016). Role of Recombinant DNA Technology to Improve Life. *International Journal of Genomics*, 1-14. doi:10.1155/2016/2405954

Killick, L., Hall, H., Duff, A. S., & Deakin, M. (2016). The census as an information source in public policy-making. *Journal of Information Science*, *42*(3), 386–395. doi:10.1177/0165551516628471

Kish, L., & Verma, V. (1986). Complete censuses and samples. *Journal of Official Statistics*, *2*(4), 381–395. Retrieved from http://www.degruyter.com

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 2053951716631130. doi:10.1177/2053951716631130

Laura, D., Maura, D., Philip, L., & Kathy, M. (2016). Sensitive Interviewing in Qualitative Research. *Research in Nursing & Health*, *39*(6), 480–490. doi:10.1002/nur.21743

Lee, D. Y., & Lehto, M. R. (2013). User acceptance of YouTube for procedural learning: An extension of the Technology Acceptance Model. *Computers & Education*, *61*, 193–208. doi:10.1016/j.compedu.2012.10.001

Leeuw, A. de, Valois, P., Ajzen, I., & Schmidt, P. (2015). Using the theory of planned behavior to identify key beliefs underlying pro-environmental behavior in high-school students: Implications for educational interventions. *Journal of Environmental Psychology*, *42*, 128–138. doi:10.1016/j.jenvp.2015.03.005

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of Family Medicine and Primary Care*, *4*(3), 324–327. doi:10.4103/2249-4863.161306

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury, CA: Sage Publications.

Lopes, P., & Oliveira, J. L. (2015). An automated real-time integration and interoperability framework for bioinformatics. *BMC Bioinformatics*, *16*(1). doi:10.1186/s12859-015-0761-3

Lub, V. (2015). Validity in Qualitative Evaluation: Linking Purposes, Paradigms, and Perspectives. *International Journal of Qualitative Methods*, *14*(5), 1–8. doi:10.1177/1609406915621406

Lucas, S. R. (2014). Beyond the existence proof: Ontological conditions, epistemological implications, and in-depth interview research. *Quality & Quantity*, *48*(1), 387–408. doi:10.1037/a0038087

Ma, R., Gong, J., & Jiang, X. (2017). Novel applications of next-generation sequencing in breast cancer research. *Genes & Diseases*, *4*(3), 149–153. doi:10.1016/j.gendis.2017.07.003

Makki, T. W., DeCook, J. R., Kadylak, T., & Lee, O. J. (2018). The Social Value of Snapchat: An Exploration of Affiliation Motivation, the Technology Acceptance Model, and Relational Maintenance in Snapchat Use. *International Journal of*

Human–Computer Interaction, 34(5), 410–420. doi:10.1080/10447318.2017.1357903

Malagon-Maldonado, G. (2014). Qualitative research in health design. *HERD : Health Environments Research & Design Journal*, *7*(4), 120-34. Retrieved from https://www.healthdesign.org

Manan, M. M. A., & Várhelyi, A. (2015). Exploration of motorcyclists' behavior at access points of a Malaysian primary road – A qualitative observation study. *Safety Science*, *74*, 172 – 183. doi:10.1016/j.ssci.2015.01.005

Mao, J. (2014). Social media for learning: A mixed methods study on high school students' technology affordances and perspectives. *Computers in Human Behavior*, *33*, 213 – 223. doi: 10.1016/j.chb.2014.01.002

Marangunic, N., & Granic, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81-95. doi:10.1007/s10209-014-0348-1

Marco-Ramell, A., Palau-Rodriguez, M., Alay, A., Tulipani, S., Urpi-Sarda, M., Sanchez-Pla, A., & Andres-Lacueva, C. (2018). Evaluation and comparison of bioinformatic tools for the enrichment analysis of metabolomics data. *BMC Bioinformatics*, 191-11. doi:10.1186/s12859-017-2006-0

Marshall, C., & Rossman, G. B. (2016). *Designing qualitative research (6th ed.).* Sage.

Marti-Solano, M., Birney, E., Bril, A., Della Pasqua, O., Kitano, H., Mons, B., Xenarios, I. & Sanz, F. (2014). Integrative knowledge management to enhance pharmaceutical R&D. Nature Reviews. *Drug Discovery*, *13*(4), 239-240.

doi:10.1038/nrd4290

Masood, A., & Lodhi, R. N. (2016). Determinants of behavioral intentions to use SPSS among students: Application of Technology Acceptance model (TAM). *FWU Journal of Social Sciences*, *10*(2), 146–152. Retrieved from https://thescipub.com/journals/jss

McIntosh, M. J., & Morse, J. M. (2015). Situating and constructing diversity in semi-structured interviews. *Global Qualitative Nursing Research*, *2*, 1–12. doi:10.1177/2333393615597674

Merriam, S. B. (2014). *Qualitative research: A guide to design and implementation*. Chichester, United Kingdom: Wiley & Sons.

Miller, M., Zhu, C., & Bromberg, Y. (2017). clubber: removing the bioinformatics bottleneck in big data analyses. *Journal of Integrative Bioinformatics*, *14*(2). doi:10.1515/jib-2017-0020

Micic, N., Neagu, D., Campean, F., & Zadeh, E. H. (2017). Towards a Data Quality Framework for Heterogeneous Data. *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (pp. 155–162). doi:10.1109/iThings-GreenCom-CPSCom-SmartData.2017.28

Modell, S. (2015). Theoretical triangulation and pluralism in accounting research: a critical realist critique. *Accounting, Auditing & Accountability Journal*, *28*(7), 1138-1150. doi:10.1108/aaaj-10-2014-1841

Moore, G. E. (2006). Progress in digital integrated electronics [Technical literature, Copyright 1975 IEEE. Reprinted, with permission. Technical Digest. International Electron Devices Meeting, IEEE, 1975, pp. 11-13.]. *IEEE Solid-State Circuits Society Newsletter*, *11*(3), 36–37. doi:10.1109/N-SSC.2006.4804410

Mortenson, M. J., & Vidgen, R. (2016). A computational literature review of the technology acceptance model. *International Journal of Information Management*, 36(6, Part B), 1248–1259. doi:10.1016/j.ijinfomgt.2016.07.007

Moylan, C. A., Derr, A. S., & Lindhorst, T. (2015). Increasingly mobile: How new technologies can enhance qualitative research. *Qualitative Social Work*, *14*(1), 36–47. doi:10.1177/1473325013516988

Niosi, J. E. (2017). Bioinformatics: The Convergence between Biotechnology and ITC Impacts on the Productive Sector. *Comparative Approaches to Biotechnology Development and Use in Developed and Emerging Nations*, 18–44. doi:10.4018/978-1-5225-1040-6.ch002

Noprisson, H., Husin, N., Utami, M., Rahayu, P., Sucahyo, Y. G., & Sensuse, D. I. (2016). The use of a mixed method approach to evaluate m-Government implementation. *2016 International Conference on Information Technology Systems and Innovation (ICITSI)*. doi:10.1109/icitsi.2016.7858195

Noble, H., & Smith, J. (2014). Qualitative data analysis: a practical example. *Evidence Based Nursing*, *17*(1), 2–3. doi:10.1136/eb-2013-101603

Nussinov, R., & Papin, J. A. (2016). Computing Biology. *PLoS Computational Biology*, *12*(7), e1005050. doi:10.1371/journal.pcbi.1005050

O'Gara, G., Tuddenham, S., & Pattison, N. (2018). Haemato-oncology patients' perceptions of health-related quality of life after critical illness: A qualitative phenomenological study. *Intensive and Critical Care Nursing*, 44, 76–84. doi:10.1016/j.iccn.2017.09.011

O'Keeffe, J., Buytaert, W., Mijic, A., Brozović, N., & Sinha, R. (2016). The use of semi-structured interviews for the characterisation of farmer irrigation practices. *hydrology and earth system sciences*, *20*(5), 1911–1924. doi:10.5194/hess-20-1911-2016

Osborne, T., & Jones, P. I. (2017). Biosensing and geography: A mixed methods approach. *Applied Geography*, *87*, 160–169. doi:10.1016/j.apgeog.2017.08.006

Patton, M. Q. (2015). *Qualitative research and evaluation methods: integrating theory and practice (4th ed.)*. Thousand Oaks, CA: Sage.

Persico, D., Manca, S., & Pozzi, F. (2014). Adapting the Technology Acceptance Model to evaluate the innovative potential of e-learning systems. *Computers in Human Behavior*, *30*, 614–622. doi:10.1016/j.chb.2013.07.045

Peticca-Harris, A., deGama, N., & Elias, S. R. S. T. A. (2016). A dynamic process model for finding informants and gaining access in qualitative research. *Organizational Research Methods*, *19*(3), 376–401. doi:10.1177/1094428116629218

Przybyła, P., Shardlow, M., Aubin, S., Bossy, R., Eckart de Castilho, R., Piperidis, S., McNaught, J. & Ananiadou, S. (2016). Text mining resources for the life sciences. *Database*, *2016*(0), baw145–baw145. doi:10.1093/database/baw145

Park, N., Rhoads, M., Hou, J., & Lee, K. M. (2014). Understanding the acceptance of teleconferencing systems among employees: An extension of the technology acceptance model. *Computers in Human Behavior*, *39*, 118–127. doi:10.1016/j.chb.2014.05.048

Punch, K. (2014). *Introduction to social research quantitative and qualitative approaches.* Los Angeles: Sage.

Rana, N. P., Dwivedi, Y. K., Williams, M. D., & Weerakkody, V. (2015). Investigating success of an e-government initiative: Validation of an integrated IS success model. *Information Systems Frontiers*, *17*(1), 127-142. doi:10.1007/s10796-014-9504-7

Rauniar, R., Rawski, G., Yang, J., & Johnson, B. (2014). Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. *Journal of Enterprise Information Management*, *27*(1), 6-30. doi:10.1108/jeim-04-2012-0011

Reddy, T. B. K., Thomas, A. D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., … Kyrpides, N. C. (2015). The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification. *Nucleic Acids Research, 43*(D1), D1099–D1106. doi:10.1093/nar/gku950

Reybold, L. E., Lammert, J. D., & Stribling, S. M. (2013). Participant selection as a conscious research method: Thinking forward and the deliberation of “Emergent” findings. *Qualitative Research*, *13*(6), 699–716. doi:10.1177/1468794112465634

Riemenschneider, C. K., Hardgrave, B. C., & Davis, F. D. (2002). Explaining software developer acceptance of methodologies: A comparison of five theoretical models.

IEEE Transactions on Software Engineering, *28*(12), 1135–1145. doi:10.1109/TSE.2002.1158287

Rio Deiros, D., Gibbs, R. A., & Rogers, J. (2016). DNAism: exploring genomic datasets on the web with Horizon Charts. *BMC Bioinformatics*, 1749. doi:10.1186/s12859-016-0891-2

Robins, C. S., & Eisen, K. (2017). Strategies for the Effective Use of NVivo in a Large-Scale Study: Qualitative Analysis and the Repeal of Don't Ask, Don't Tell. *Qualitative Inquiry, 23*(10), 768–778. doi:10.1177/1077800417731089

Rodrigues, L. F., Oliveira, A., & Costa, C. J. (2016). Does ease-of-use contributes to the perception of enjoyment? A case of gamification in e-banking. *Computers in Human Behavior*, *61*, 114–126. doi:10.1016/j.chb.2016.03.015

Roer-Strier, D., & Sands, R. G. (2015). Moving beyond the "official story": when "others" meet in a qualitative interview. *Qualitative Research*, *15*(2), 251–268. doi:10.1177/1468794114548944

Roulston, K., & Shelton, S. A. (2015). Reconceptualizing bias in teaching qualitative research methods. *Qualitative Inquiry*, *21*(4), 332–342. doi:10.1177/1077800414563803

Russo, E. (2003). Special Report: The birth of biotechnology. *Nature*, 421(6921), 456–457. Retrieved from http://www.nature.com

Sanjari, M., Bahramnezhad, F., Fomani, F. K., Shoghi, M., & Ali Cheraghi, M. (2014). Ethical challenges of researchers in qualitative studies: The necessity to develop a specific guideline. *Journal of Medical Ethics & History of Medicine*, *7*(14), 1–6.

Retrieved from http://jmehm.tums.ac.ir/

Samish, I., Bourne, P., & Najmanovich, R. (2015). Achievements and challenges in structural bioinformatics and computational biophysics. *Bioinformatics*, *31*(1), 146–150. doi:10.1093/bioinformatics/btu769

Salloch, S., Ritter, P., Wäscher, S., Vollmann, J., & Schildmann, J. (2014). Medical Expertise and Patient Involvement: A Multiperspective Qualitative Observation Study of the Patient's Role in Oncological Decision Making. *The Oncologist*, *19*(6), 654–660. doi:10.1634/theoncologist.2013-0268

Schwarzbach, e., Smýkal, p., Dostál, o., Jarkovská, m., & Valová, s. (2014). Gregor J. Mendel - Genetics Founding Father. *Czech Journal of Genetics & Plant Breeding*, *50*(2), 43-51.

Shah, S. P., Huang, Y., Xu, T., Yuen, M. M., Ling, J., & Ouellette, B. F. (2005). Atlas – a data warehouse for integrative bioinformatics. *BMC Bioinformatics*, *6*, 34. doi:10.1186/1471-2105-6-34

Sharma, N., Panwar, A., & Sugandh, U. (2018). Big Data Analytics in Health Care: A Literature Survey. *International Journal of Recent Research Aspects*, *5*(1), 127-132.

Siegel, D., Acharya, P., & Sivo, S. (2017). Extending the Technology Acceptance Model to Improve Usage & Decrease Resistance toward a New Technology by Faculty in Higher Education. *Journal Of Technology Studies, 43*(2), 58-69.

Sloan, A., & Bowe, B. (2014). Phenomenology and hermeneutic phenomenology: the philosophy, the methodologies, and using hermeneutic phenomenology to investigate lecturers’ experiences of curriculum design. *Quality and Quantity*, *48*, 1291–1303. doi:10.1007/s11135-013-9835-3

Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D. O., & Rosales-Mendoza, S. (2015). An overview of bioinformatics tools for epitope prediction: Implications on vaccine development. *Journal of Biomedical Informatics*, *53*, 405 – 414. doi:10.1016/j.jbi.2014.11.003

Souilmi, Y., Jung, J.-Y., Lancaster, A., Gafni, E., Amzazi, S., Ghazal, H., Wall, D., Tonellato, P. (2015). COSMOS: cloud enabled NGS analysis. *BMC Bioinformatics*, *16*(2), A2. doi:10.1186/1471-2105-16-S2-A2

Speer, S. A., & Stokoe, E. (2014). Ethics in action: consent-gaining interactions and implications for research practice. *The British Journal of Social Psychology*, *53*(1), 54–73. doi:10.1111/bjso.12009

Storey, V. C., & Song, I.-Y. (2017). Big data technologies and Management: What conceptual modeling can do. *Data & Knowledge Engineering*, *108*, 50 – 67. doi:10.1016/j.datak.2017.01.001

Thampi, S. M. (2009). Introduction to Bioinformatics. *arXiv*. Retrieved from http://arxiv.org

Ting, P. L., Chen, S., Chen, H., & Fang, W. (2017). Using Big Data and Text Analytics to Understand How Customer Experiences Posted on Yelp.com Impact the Hospitality Industry. *Contemporary Management Research*, *13*(2). doi:10.7903/cmr.17730

Tran, V.-T., Porcher, R., Falissard, B., & Ravaud, P. (2016). Point of data saturation was assessed using resampling methods in a survey with open-ended questions. *Journal of Clinical Epidemiology*, *80*, 88–96. doi:10.1016/j.jclinepi.2016.07.014

Tran, V.-T., Porcher, R., Tran, V.-C., & Ravaud, P. (2017). Predicting data saturation in qualitative surveys with mathematical models from ecological research. *Journal of Clinical Epidemiology*, *82*, 71–78.e2. doi:10.1016/j.jclinepi.2016.10.001

Triplet, T., & Butler, G. (2014). A review of genomic data warehousing systems. *Briefings in Bioinformatics*, *15*(4), 471–483. doi:10.1093/bib/bbt031

Tucker, D. A., Yeow, P., & Viki, G. T. (2013). Communicating During Organizational Change Using Social Accounts: The Importance of Ideological Accounts. *Management Communication Quarterly*, *27*(2), 184–209. doi:10.1177/0893318912469771

Tzafilkou, K., & Protogeros, N. (2017). Diagnosing user perception and acceptance using eye tracking in web-based end-user development. *Computers in Human Behavior*, *72*, 23–37. doi:10.1016/j.chb.2017.02.035

U.S. Department of Health and Human Services, Office of the Secretary, The National Commission for the Protection of Human Subjects of Biomedical and Behavioral 105 Research. (1979). *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research (45 CFR 46)*. Retrieved from http://www.hhs.gov

Vaismoradi, M., Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice*, *6*(5), 100-110. doi:10.5430/jnep.v6n5p100

Veiga, A. K., Saraiva, A. M., Chapman, A. D., Morris, P. J., Gendreau, C., Schigel, D., & Robertson, T. J. (2017). A conceptual framework for quality assessment and management of biodiversity data. *Plos ONE*, *12*(6), 1-20. doi:10.1371/journal.pone.0178731

Velloso, H., Vialle, R. A., & Ortega, J. M. (2015). BOWS (bioinformatics open web services) to centralize bioinformatics tools in web services. *BMC Research Notes*, 8. doi: 10.1186/s13104-015-1190-0

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157–178.

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences, 39(2)*, 273-315. doi:10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186. doi:10.1287/mnsc.46.2.186.11926

Vicary, S., Young, A., & Hicks, S. (2016). A reflective journal as learning process and contribution to quality and validity in interpretative phenomenological analysis. *Qualitative Social Work*. doi:10.1177/1473325016635244

Wallace, L. G., & Sheetz, S. D. (2014). The adoption of software measures: A technology acceptance model (TAM) perspective. *Information & Management*, *51*(2), 249–259. doi:10.1016/j.im.2013.12.003

White, J., & Drew, S. (2011). Collecting data or creating meaning? *Qualitative Research Journal, 11, 3–12*. doi:10.3316/QRJ1101003

Wiles, R., Crow, G., & Pain, H. (2011). Innovation in qualitative research methods: a narrative review. *Qualitative Research*, *11*(5), 587-604. doi:10.1177/1468794111413227

Williams, C. L., Sica, J. C., Killen, R. T., & Balis, U. G. J. (2016). The growing need for microservices in bioinformatics. *Journal of Pathology Informatics*, *7*(1), 45. doi:10.4103/2153-3539.194835

Wolf, L. E., Patel, M. J., Williams Tarver, B. A., Austin, J. L., Dame, L. A., & Beskow, L. M. (2015). Certificates of Confidentiality: Protecting Human Subject Research Data in Law and Practice. *Journal of Law, Medicine & Ethics*, *43*(3), 594-609. doi:10.1111/jlme.12302

Yang, K. & Banamah, A. (2014). Quota sampling as an alternative to probability sampling? An experimental study. *Sociological Research Online, 19*. doi:10.5153/sro.3199

Yao, S. (2017). Big data in clinical research. *Australian Medical Student Journal*, *8*(1), 22-25.

Yazan, B. (2015). Three Approaches to Case Study Methods in Education: Yin, Merriam, and Stake. *The Qualitative Report*, *20*(2), 134-152. Retrieved from http://nsuworks.nova.edu

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal* *of Education*, *48*(2), 311–325. doi:10.1111/ejed.12014

Yin, R. K. (2014). *Case Study Research: Design and Methods (5th ed.)*. Sage.

Yoon, C. (2018). Extending the TAM for Green IT: A normative perspective. *Computers in Human Behavior*, *83*, 129 – 139. doi:10.1016/j.chb.2018.01.032

Yuan, Y.-H., Tsai, S.-B., Dai, C.-Y., Chen, H.-M., Chen, W.-F., Wu, C.-H., … Wang, J. (2017). An empirical research on relationships between subjective judgement, technology acceptance tendency and knowledge transfer. *PLoS ONE, 12*(9), 1 – 22.

Yüksel, P., & Yıldırım, S. (2015). Theoretical Frameworks, Methods, and Procedures for Conducting Phenomenological Studies in Educational Settings. *Turkish Online Journal of Qualitative Inquiry*, *6*(1), 1–20. doi:10.17569/tojqi.59813

Zamawe, F. C. (2015). The Implication of Using NVivo Software in Qualitative Data Analysis: Evidence-Based Reflections. *Malawi Medical Journal*, *27*(1), 13–15. doi:10.4314/mmj.v27i1.4

Zhao, Z., Jin, V. X., Huang, Y., Guda, C., & Ruan, J. (2015). Frontiers in Integrative Genomics and Translational Bioinformatics. *BioMed Research International*, 2015725491. doi:10.1155/2015/725491

Zhang, H., Chen, W., & Sun, K. (2017). Mendelism: New insights from Gregor Mendel's lectures in brno. *Genetics*, *207*(1), 1-8. doi:10.1534/genetics.117.20143.

Zou, Q., Hu, Q., Guo, M., & Wang, G. (2015). HAlign: Fast multiple similar DNA/RNA sequence alignment based on the centre star strategy. *Bioinformatics*, *31*(15), 2475–2481. doi:10.1093/bioinformatics/btv177

Appendix A: Interview Protocol

Interview: Exploring strategies to integrate disparate Bioinformatics datasets.

A. Introduce myself to the participant and thank them for participating.

B. Verified receipt of the consent form and answer questions that the participant might have and remind the participant that the interview will be audio-recorded.

C. Explain the purpose of the study to the participant.

D. Describe the reason for participating and that the information the participant provide will support my study.

E. Describe the benefit of participation by mentioning that the information provided by the participants could add to the academic and professional body of knowledge.

F. Discuss ethics, the participant right to privacy, and request permission to audio-record the interview.

G. Turn on the recording device and tell the participant their identification code, the date and time of the interview.

H. Discuss confidentiality by informing the participant that he can refuse to answer any question or stop participating at any time, and all information provided by the participant is confidential, and I will not disclose it to anyone.

I. Start interviewing by asking interview questions in order. Give the participant time to respond to each question and ask follow-up questions when necessary.

1. What strategies do you use to integrate disparate bioinformatics datasets? Please explain.
2. How successful have you been in integrating disparate bioinformatics datasets? Please elaborate.
3. What negative aspects of integrating disparate bioinformatics datasets have you experienced? Please explain.
4. What are the complexities that specialists might face while integrating disparate biological datasets? Please explain.
5. What are the challenges and difficulties that face you while retrieving information from different locations stored in different formats? And how exploring strategies for integrating disparate bioinformatics datasets can overcome those challenges? Please elaborate.
6. What strategies do you have for bioinformatics specialists to analyze biological data more efficiently? Please elaborate.
    - Ask the participant if they are aware of any secondary data or artifacts that might be relevant to the topic.
    - Explain the concept of member checking and schedule a follow-up meeting to review my analysis.
    - Turn off the audio recording device and thank the participant for their participation.

Appendix B: Observation Protocol

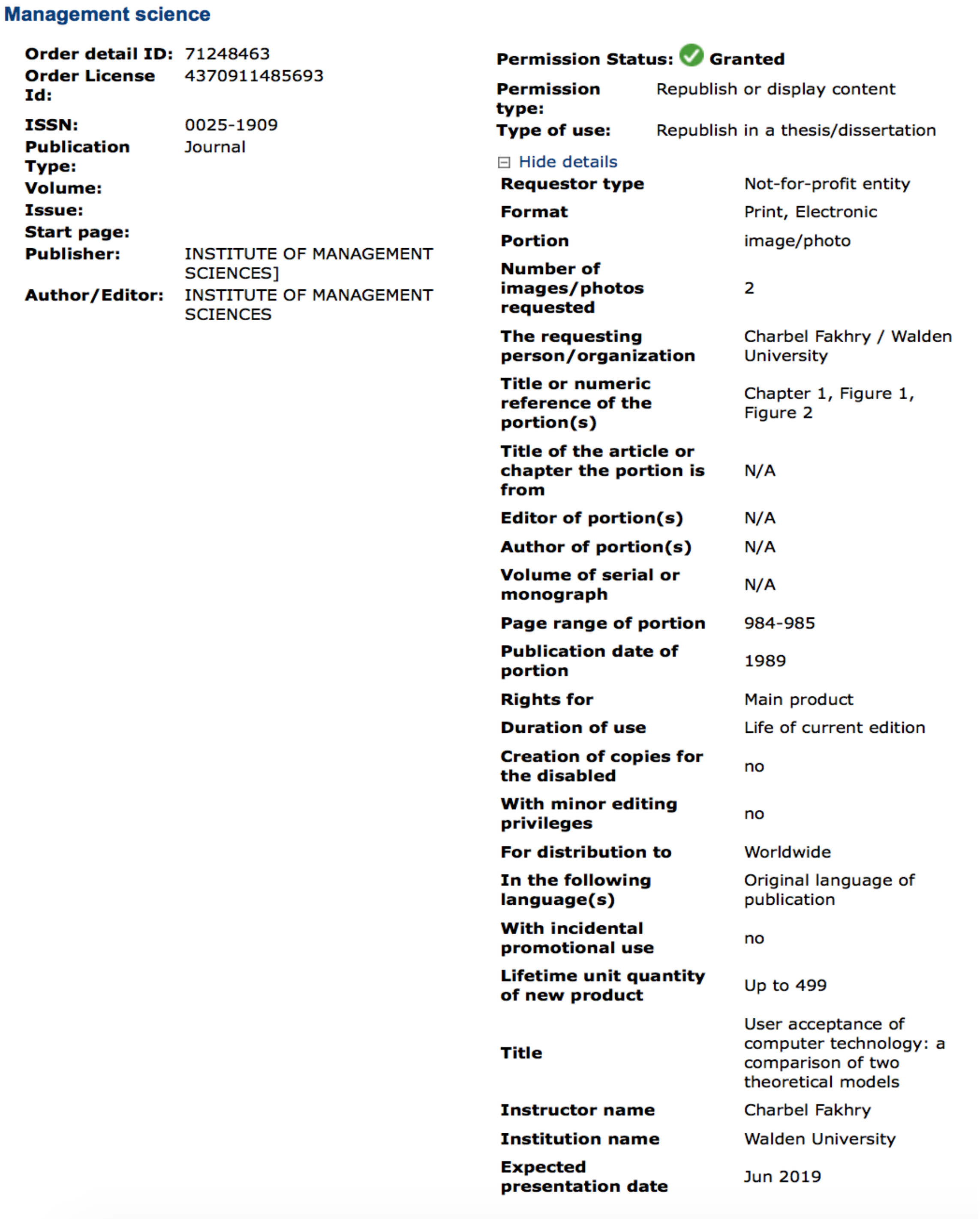
The purpose of this observation protocol is to help the researcher to focus on the technicalities while observing the work environment.

Directions: When starting the observation, the researcher must follow the table below to write a clear description of the work environment, the date and time of the observation and notes describing the work process or any details that the researcher may consider relevant. After the observation, the researcher will find concepts based on the records that may help the researcher in the data analysis process.

Name of the Researcher	
Tentative Schedule	
Date	
Work Environment Background Describe in great details the setting	
The position The distance of the researcher from performed work	
The Action What happened in detail during the work process.	
Observation Type (Direct or Participant)	
Areas the specialists focused on	

Time	**Observation notes**

Appendix C: Permissions to use Figures



Management science

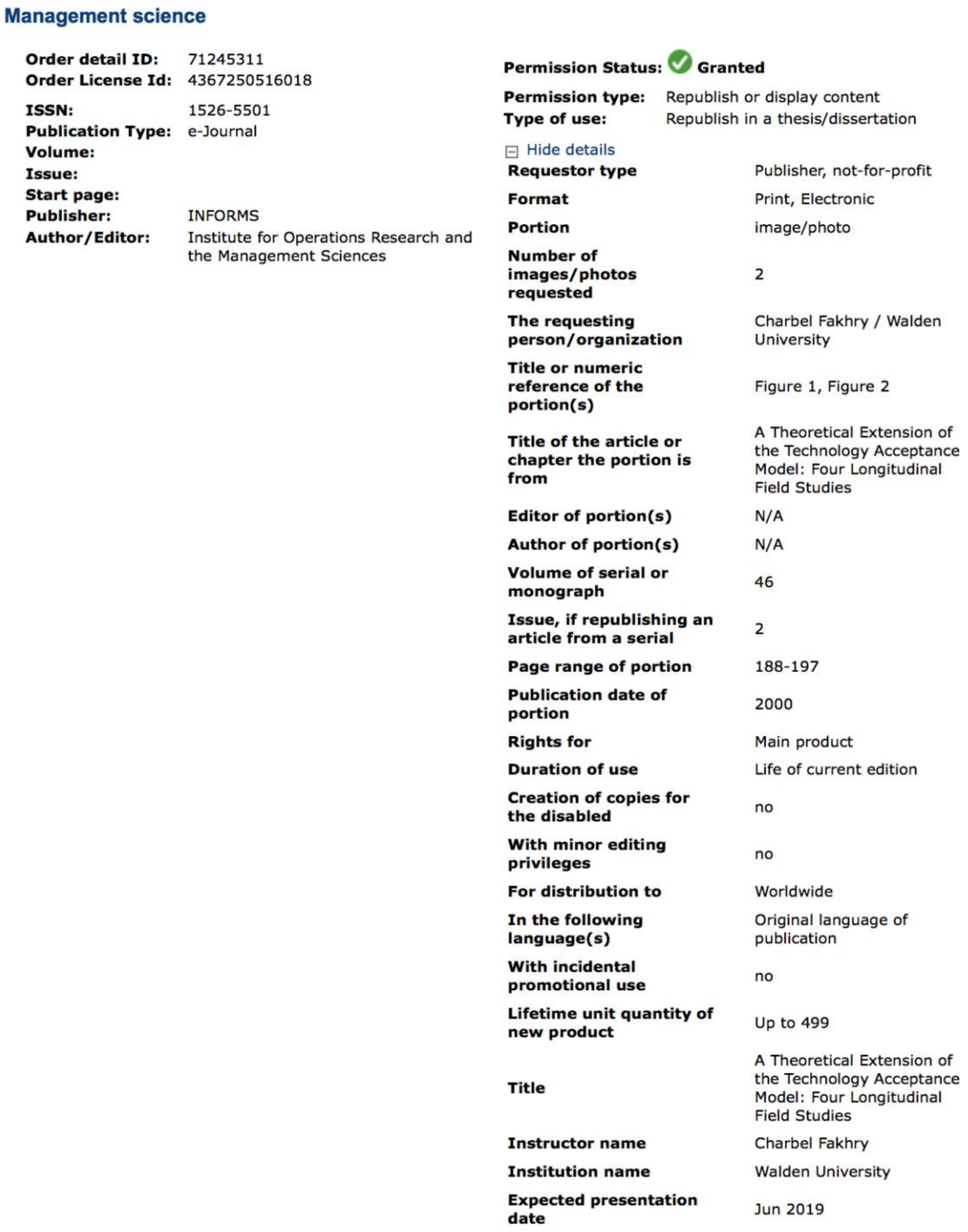
Order detail ID: 71248463
Order License Id: 4370911485693
ISSN: 0025-1909
Publication Type: Journal
Volume:
Issue:
Start page:
Publisher: INSTITUTE OF MANAGEMENT SCIENCES]
Author/Editor: INSTITUTE OF MANAGEMENT SCIENCES

Permission Status: ✓ **Granted**

Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation

⊟ Hide details

Requestor type	Not-for-profit entity
Format	Print, Electronic
Portion	image/photo
Number of images/photos requested	2
The requesting person/organization	Charbel Fakhry / Walden University
Title or numeric reference of the portion(s)	Chapter 1, Figure 1, Figure 2
Title of the article or chapter the portion is from	N/A
Editor of portion(s)	N/A
Author of portion(s)	N/A
Volume of serial or monograph	N/A
Page range of portion	984-985
Publication date of portion	1989
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	User acceptance of computer technology: a comparison of two theoretical models
Instructor name	Charbel Fakhry
Institution name	Walden University
Expected presentation date	Jun 2019



Management science

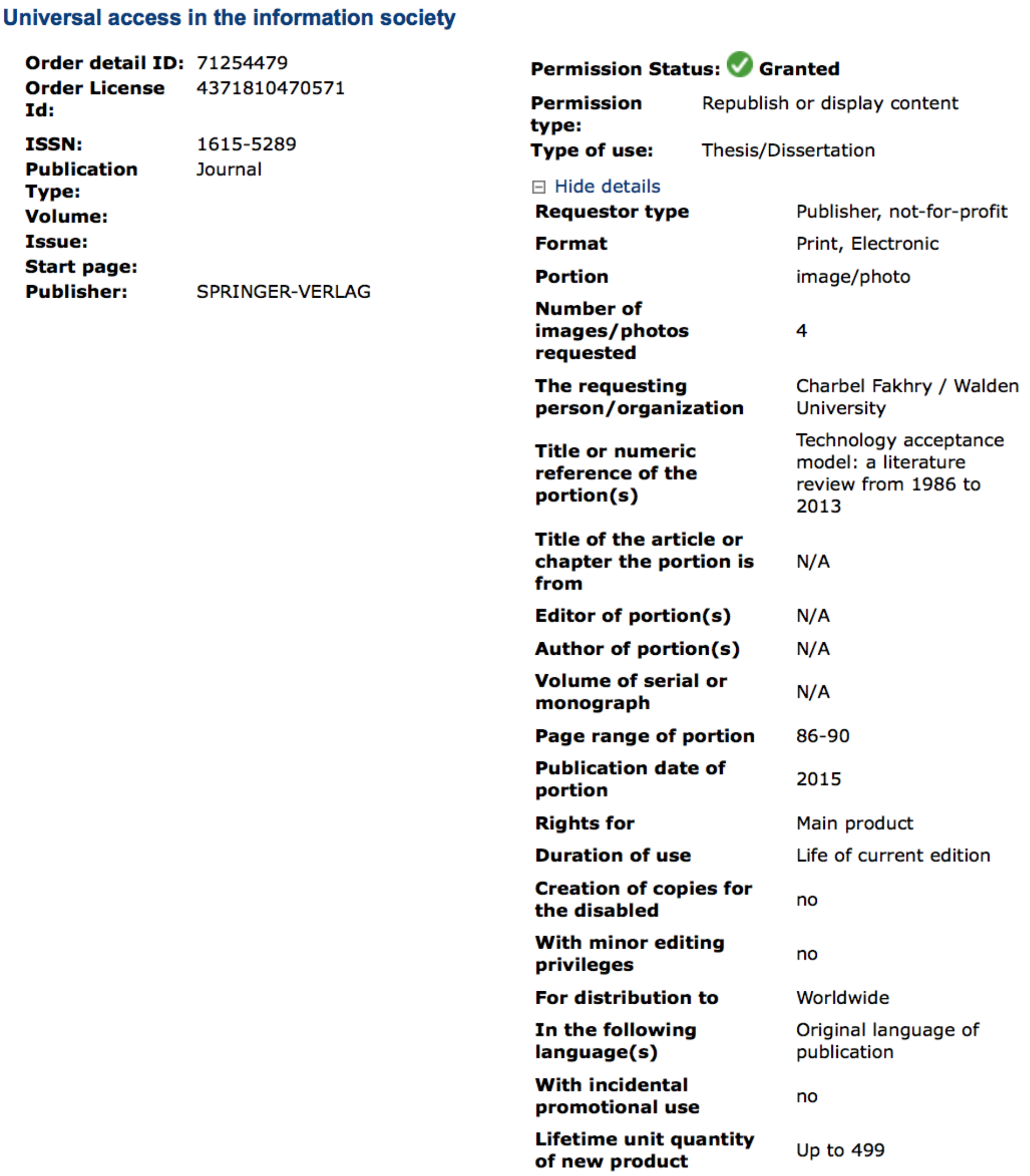
Order detail ID:	71245311
Order License Id:	4367250516018
ISSN:	1526-5501
Publication Type:	e-Journal
Volume:	
Issue:	
Start page:	
Publisher:	INFORMS
Author/Editor:	Institute for Operations Research and the Management Sciences

Permission Status: ✓ **Granted**

Permission type:	Republish or display content
Type of use:	Republish in a thesis/dissertation

⊟ Hide details

Requestor type	Publisher, not-for-profit
Format	Print, Electronic
Portion	image/photo
Number of images/photos requested	2
The requesting person/organization	Charbel Fakhry / Walden University
Title or numeric reference of the portion(s)	Figure 1, Figure 2
Title of the article or chapter the portion is from	A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies
Editor of portion(s)	N/A
Author of portion(s)	N/A
Volume of serial or monograph	46
Issue, if republishing an article from a serial	2
Page range of portion	188-197
Publication date of portion	2000
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies
Instructor name	Charbel Fakhry
Institution name	Walden University
Expected presentation date	Jun 2019



Universal access in the information society

Order detail ID:	71254479
Order License Id:	4371810470571
ISSN:	1615-5289
Publication Type:	Journal
Volume:	
Issue:	
Start page:	
Publisher:	SPRINGER-VERLAG

Permission Status: ✓ **Granted**

Permission type:	Republish or display content
Type of use:	Thesis/Dissertation

⊟ Hide details

Requestor type	Publisher, not-for-profit
Format	Print, Electronic
Portion	image/photo
Number of images/photos requested	4
The requesting person/organization	Charbel Fakhry / Walden University
Title or numeric reference of the portion(s)	Technology acceptance model: a literature review from 1986 to 2013
Title of the article or chapter the portion is from	N/A
Editor of portion(s)	N/A
Author of portion(s)	N/A
Volume of serial or monograph	N/A
Page range of portion	86-90
Publication date of portion	2015
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499

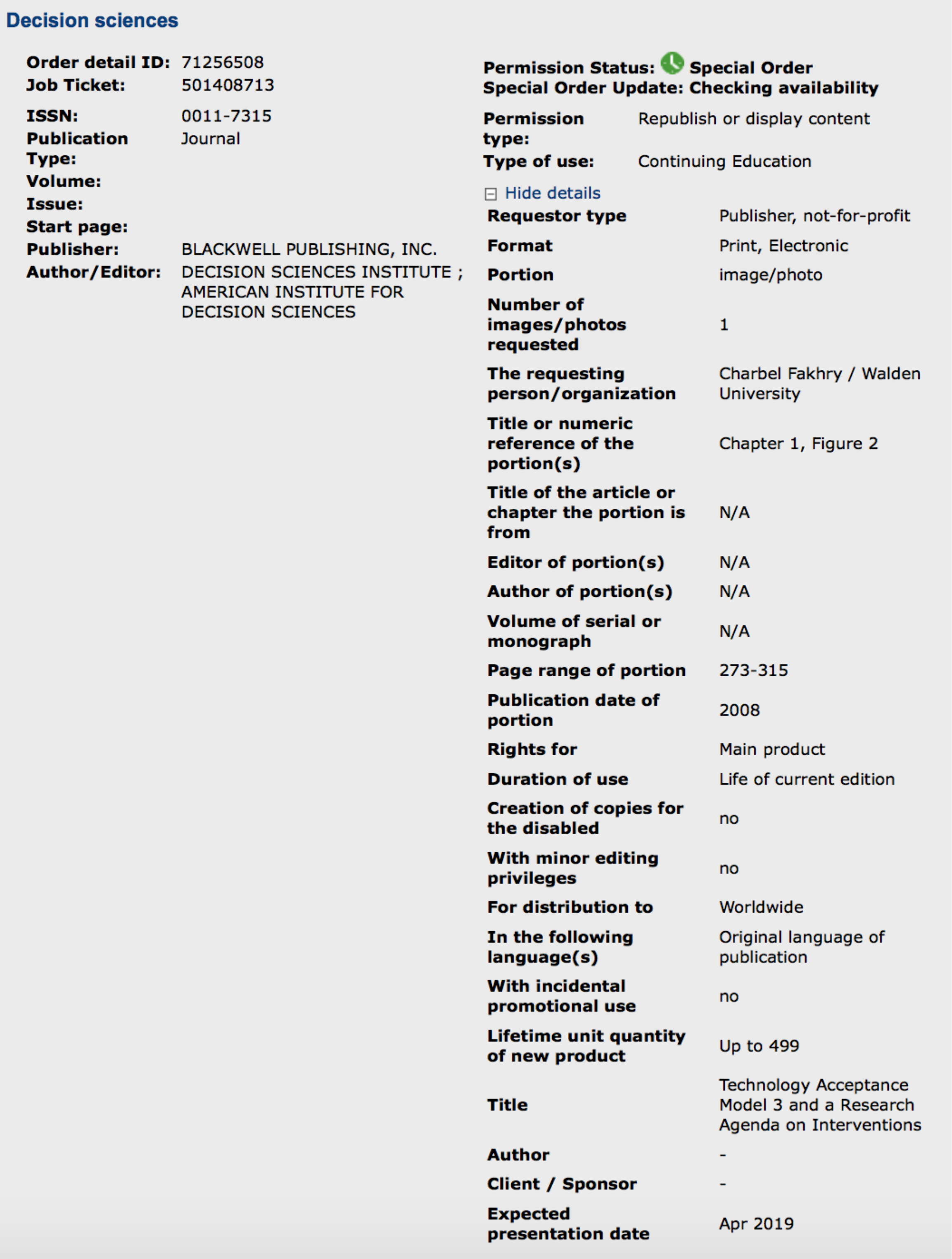
Journal of enterprise information management

ISSN:	1741-0398
Publication Type:	Journal
Volume:	
Issue:	
Start page:	
Publisher:	EMERALD GROUP PUBLISHING LIMITED

Permission type:	Republish or display content
Type of use:	Republish in a thesis/dissertation

⊟ Hide details

Requestor type	Publisher, not-for-profit
Format	Print, Electronic
Portion	image/photo
Number of images/photos requested	1
The requesting person/organization	Charbel Fakhry / Walden University
Title or numeric reference of the portion(s)	Chapter 1, Figure 1
Title of the article or chapter the portion is from	N/A
Editor of portion(s)	N/A
Author of portion(s)	N/A
Volume of serial or monograph	N/A
Page range of portion	6-30
Publication date of portion	2014
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	Technology acceptance model (TAM) and social media usage: an empirical study on Facebook
Instructor name	-
Institution name	Walden University
Expected presentation date	Mar 2019



Decision sciences

Order detail ID: 71256508
Job Ticket: 501408713

ISSN: 0011-7315
Publication Type: Journal
Volume:
Issue:
Start page:
Publisher: BLACKWELL PUBLISHING, INC.
Author/Editor: DECISION SCIENCES INSTITUTE ; AMERICAN INSTITUTE FOR DECISION SCIENCES

Permission Status: **Special Order**
Special Order Update: Checking availability

Permission type: Republish or display content
Type of use: Continuing Education

⊟ Hide details

Requestor type	Publisher, not-for-profit
Format	Print, Electronic
Portion	image/photo
Number of images/photos requested	1
The requesting person/organization	Charbel Fakhry / Walden University
Title or numeric reference of the portion(s)	Chapter 1, Figure 2
Title of the article or chapter the portion is from	N/A
Editor of portion(s)	N/A
Author of portion(s)	N/A
Volume of serial or monograph	N/A
Page range of portion	273-315
Publication date of portion	2008
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	Technology Acceptance Model 3 and a Research Agenda on Interventions
Author	-
Client / Sponsor	-
Expected presentation date	Apr 2019