Available online at www.sciencedirect.com

ScienceDirect

Nurs Outlook 67 (2019) 39–48

Nursing Outlook

www.nursingoutlook.org

A model to evaluate data science in nursing doctoral curricula

Kimberly D. Shea, PhD, RN, Barbara B. Brewer, PhD, RN, FAAN, Jane M. Carrington, PhD, RN, FAAN, Mary Davis, PhD, RN, Sheila Gephart, PhD, RN, Anne Rosenfeld, PhD, RN, FAAN*

The University of Arizona, College of Nursing, Tucson, AZ

ARTICLE INFO

Article history:
Received 6 June 2018
Received in revised form 16 October 2018
Accepted 27 October 2018
Available online November 2, 2018.

Keywords:
data science
nursing
PhD education
curriculum
framework
nursing informatics

ABSTRACT

Background: Building on the efforts of the American Association of Colleges of Nursing, we developed a model to infuse data science constructs into doctor of philosophy (PhD) curriculum. Using this model, developing nurse scientists can learn data science and be at the forefront of data driven healthcare.

Purpose: Here we present the Data Science Curriculum Organizing Model (DSCOM) to guide comprehensive doctoral education about data science.

Methods: Our team transformed the terminology and applicability of multidisciplinary data science models into the DSCOM.

Findings: The DSCOM represents concepts and constructs, and their relationships, which are essential to a comprehensive understanding of data science. Application of the DSCOM identified areas for threading as well as gaps that require content in core coursework.

Discussion: The DSCOM is an effective tool to guide curriculum development and evaluation towards the preparation of nurse scientists with knowledge of data science.

Cite this article: Shea, K.D., Brewer, B.B., Carrington, J.M., Davis, M., Gephart, S., & Rosenfeld, A. (2019, January/February). A model to evaluate data science in nursing doctoral curricula. Nurs Outlook, 67(1), 39–48. https://doi.org/10.1016/j.outlook.2018.10.007.

Introduction

The ubiquitous presence of technology continually generates vast quantities of data that necessitate leaders within the health care system understand data science. As the foundation of the health care system, nurses must have a working knowledge of technology-generated data. Data science is more than just understanding numbers used for analysis. Data support evidence-based practice and nurse scientists must thoroughly understand the complexities of data science in order to use advanced analysis techniques to build knowledge for providers. In 2013, a "call to action" for colleges/schools of nursing to engage in educating nurses in data science was supported by the American Academy of Nursing (Clancy et al., 2014). The purpose of this paper is to discuss the development and use of the Data Science Curriculum Organizing Model (DSCOM) that can serve to guide mapping, organization, and insertion of data science constructs and concepts throughout an existing PhD nursing curriculum.

Data science has been defined by the National Consortium for Data Science, as "the systematic study and use of digital data in order to accelerate discovery,

* *Corresponding author:* Kimberly D. Shea, University of Arizona, 1305 N Martin Avenue, PO Box 210203, Tucson, AZ 85721.
E-mail address: kshea@email.arizona.edu (K.D. Shea).

https://doi.org/10.1016/j.outlook.2018.10.007

improve critical decision-making processes, and enable a data-driven economy." Within nursing, data science has been referred to as "a philosophy, a collection of methods and a suite of analytics that focuses on data storage, transport, and cleaning procedures in addition to visualization tools" (Broome, 2016, p. 113). Or more precisely, "Data science is turning data into action" (Herman et al., 2015, p. 21).

Data scientists are statistical specialists trained to evaluate digital data. Data scientists often normalize data when they are unfamiliar with the nuances of the raw data. Normalization without domain expertise can result in erroneous conclusions. Domain knowledge may be provided by practicing or research nurses. Nurse Data Scientists are nursing domain experts with data science training that enables accurate interpretation of data resulting in useful new knowledge for continuous improvement and innovation.

For many nurse scientists, their goal is to turn data into action. Data consist of text, images, video, and audio formats as well as traditional numbers. Nurses in all environments encounter data, ranging from cellular data generated by individuals to aggregated group metrics estimating disease prevalence in entire populations. Data from practice settings are initially "raw" and require processing for analysis. Nurses who obtain a PhD in nursing follow a curriculum that includes content covering many different kinds of data. Nurse scientists trained in quantitative methods often take courses that are data intensive. Existing nurse scientist curricula include quantitative and statistical content, which alone do not prepare them for evaluation of large data sets. Preparation of today's nurse scientists requires a paradigm shift affecting all core courses, which is the basis for threading data science content at strategic points in the curriculum. A guide, such as the DSCOM, can assure that data science constructs and concepts are included throughout an established PhD curriculum.

Roles of Nurses Using Data

In 2015, the American Association of Colleges of Nursing (AACN) and Manatt Health released a report, "*Advancing Healthcare Transformation: A New Era for Academic Nursing*" stating that baccalaureate and higher degree-granting colleges/schools of nursing have a lead role in educating nurses that will transition our health care system to one that learns (AACN & Manatt Health, 2016). The Institute of Medicine defines a Healthcare Learning System as an environment where "science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the care process . . . and new knowledge captured as an integral by-product of the delivery experience" (Executive Office of Health and Medicine, 2013, p. 2). Nurses need to be leaders in policy discussions about reform health care practices and delivery. Through partnering, bedside, and scientific nurses engage in ongoing dialogue about the roles and needs of newly educated nurses in the workforce. Colleges/Schools of nursing have an obligation to improve all three missions of academia: teaching, clinical care, and research by incorporating data science skills into the curriculum. By understanding how to use data, new evidence will be generated for nurses to transform health care delivery into a learning system.

The AACN report concludes that data generate knowledge for nurses. To enable data to be used for decision making to inform care, as well as economic, political, and health care processes, nurses must have skills to manage data at the point of collection and the point of care (AACN & Manatt Health, 2016). However, "There is a shortage of clinicians to support clinical trials and data integrity/analytics roles in Academic Health Centers, as well as shortages of researchers in data science and implementation science" (AACN & Manatt Health, 2016, p. 17).

The AACN addressed the shortage of nurse scientists trained in data science at the 2016 Doctoral Education Conference *"Enhancing Nursing Science and Improving Patient Care through Big Data"* (AACN, 2016). The primary objective of the conference was to "recognize the key roles that academic leaders, faculty, and students play when forming core data science learning content necessary for future nurse scientists and leaders in advanced nursing practice" (AACN, 2016, p. 1). Not all PhD prepared nurses will become "Nurse Data Scientists," however all nurse scientists would benefit from data science concepts embedded in core curriculum. Nurse scientists concerned with data science are characterized by three different educational tracks of preparation and expertise (Brennan & Bakken, 2015). Figure 1 displays the three tracks of educational of preparation in data science for nurse scientists. The first track represents the nurse scientist who gains a PhD in nursing and has principles of data science taught as part of the core curriculum. Many nurse scientists are characterized by this track. The second track consists of nurse scientists who pursue a PhD in nursing and choose courses in quantitative methods used for large data analysis to support their data intensive research program. These students may earn a minor in data science or pursue postdoctoral study in data science methods. The third track consists of nurses who want to become data scientists. Their primary focus is data science, so they would pursue a PhD in Data Science in contrast to a PhD in Nursing. Nurses with a PhD in Data Science are rare, so todays nursing colleges/schools do not have the faculty for this as a major within existing curricula. Currently, nurses as Data Science PhDs must enroll in Computer Science. Nurses wishing to get a PhD in Data Science would continue with a nursing minor, when possible, to give domain breadth to the methods and theories that they learn, apply, and study in a Data Science PhD

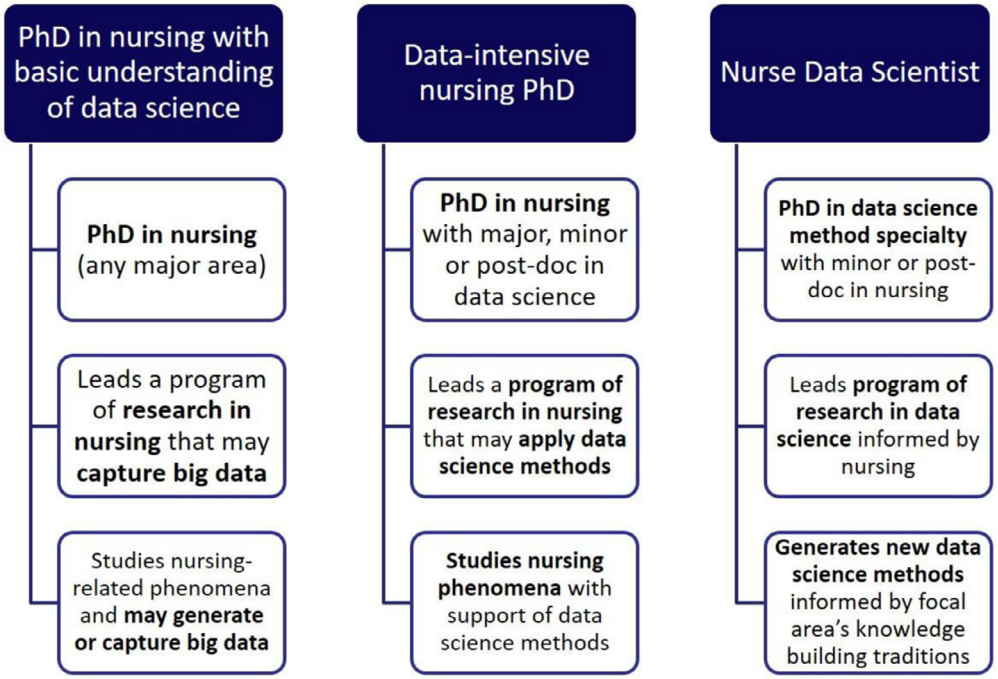

Figure 1 – Tracks of educational preparation in data science for nurse scientistis.

program. As Data Scientists, these nurse graduates will lead programs of research to build data science knowledge informed by their nursing lens. With hope that in the future, there would be faculty for a Nurse Data Scientist PhD within nursing colleges/schools. When the DSCOM is used to assure that data science concepts are threaded in appropriate areas throughout existing core curricula, each of the tracks of educational preparation in data science has exposure to knowledge and education to be scientists in a world that generates vast quantities of data.

"Big Data"

Data science training is a graduate education that provides the learner with the tools to collect, mangae, and analyze small and large datasets as well as "big data." Large datasets can be analyzed using traditional data processing software products, however these products are inadequate to deal with "big data." Brennan and Bakken (2015) advocate for a change in existing approaches to large dataset analyses by providing a stronger foundation to engage in the revolution of "big data." Opportunities for researchers with understanding of data science are flouishing. Big data are not only much larger than large datasets but also include other complexity dimensions. The current but evolving definition states "big data" are comprised of "5Vs": *volume, velocity, veracity, variety, and value* (Brennan & Bakken, 2015; Laney, 2001). Determining size or *volume* of data that qualify as "big data" is elusive to measurement. The capacity to collect data continues to evolve and data sets become bigger each day. It has been reported that due to the massive amount of information being entered into the Internet that there are 2.5 exabytes of data collected daily (Khoso, 2016). *Velocity* describes the ultra-fast rate of data being generated. *Veracity* refers to the potential that the data may not be accurate, truthful, and reliable due to the conditions in which they were collected. Since there are many *varieties* of data available, such as text, images, and video, it is difficult to assure the veracity. Emerging analyses are used to determine the *value* of data for answering questions about health care delivery systems. A familiar analysis commonly used to examine the value of obtaining and managing data is a cost-benefit analysis. Others of interest to nurse scientists include trending of longitudinal data, merging of personal data from the Internet of Things (i.e., sensors) with electronic heath record data, and identification of emergent disease within populations through analysis of tweets and other social media data.

Graduate colleges/schools of nursing are charged with providing education that is necessary to prepare nurse scientists for the future. Currently, PhD education does not address all the complexity that the "5Vs" present when using data for research. To-date, a model does not exists to assure that PhD nursing curricular content includes data science components that provide the necessary knowledge to address, in a comprehensive and organized manner, the challenging roles of working with big data. While nonresearch intense universities may find the use of such a model useful, DSCOM is intended primarily for PhD programs in colleges/schools of nursing that are considered "research-intensive." The comprehensive DSCOM can be used as a guide to evaluate nursing PhD curricula to

determine whether the recommended data science concepts are present. As a result of the evaluation, gaps in curriculum can be identified and filled. Our approach may be used broadly, however faculty with expertise in informatics or the statistical sciences may be best suited to apply this model to map and design curricular change.

Developing the Data Science Curriculum Organizing Model (DSCOM)

Our initial model was an adaptation of the many different Venn diagrams of data science models originally described by Conway (2010). Conway's model was developed out of contemplation and discussion about what a curriculum for an interdisciplinary data science program of study might contain. The model contained concepts such as hacking skills, math and statistical knowledge, and substantive expertise. After 2013, many similar versions of data science models were created to capture its complexity. Our intent was to adapt constructs and concepts from the existing models to fit the need of guiding core content in courses taught to PhD nursing students. These existing models displayed many overlapping constructs, such as those between traditional research, data processing, and machine learning. In the effort to use the new model as a guide to examine core curriculum for the PhD, our team grappled with how, and if, to include informatics in the model. The first challenge was to decide where informatics fit with data science. Our struggle was with concerns as to whether informatics contained data science concepts we identified, or if there was overlap between data science and informatics or if informatics was a separate and distinct construct from data science. Many different versions of our team's preliminary models were proposed to reconcile these questions. Ultimately, we decided that because health care informatics is a core survey course within the curriculum that did not cover all of the components of data science; we would divide components of our informatics course content into the constructs of our data science model. The next challenge was the lack of detail in data science constructs proposed by earlier models. This made it difficult to identify in which PhD nursing courses the constructs of data science could be embedded. In response, we deconstructed the constructs identified in prior data science models to organize them into concepts within our model. By having defined concepts, educators could more clearly determine where the concept fit within existing curriculum. The model was to be nonlinear, like a Venn diagram, but for purpose of curriculum review, overlaps in constructs were eliminated.

Given that the intent of our model is to organize nursing curriculum, we used Effken's (2003) Informatics Research Organizing Model (IROM) to provide several purposive similarities (Effken, 2003). The IROM's purpose is as a guide to organize research by providing constructs, just as the DSCOM is to organize curriculum. Similar to the System Development Lifecycle in the center of the IROM, the "5Vs" are in the center of the DSCOM and are to be considered within each construct. Finally, the IROM includes perspectives at multiple levels ranging from the individual, to the group and then the population. This multiple level perspective is also relevant to data science. By expanding on the constructs identified by early data science models and including the organizing nature of the IROM, the DSCOM was developed (Figure 2).

DSCOM

The DSCOM includes six constructs anchored by the "5Vs" in the center. The six constructs in the model: Domain, Ethics, Theory, Technical, Analytics, and Dissemination, are data science constructs to which doctoral students must be exposed, if they are to be successful in a data ubiquitous health care world. Using the model to guide embedding DSCOM content into a course requires an a priori understanding of whether the concepts are intended to be applied to individual, group, and/or population levels. The constructs are the same at each level, only the perspective that the data represent in the course may be different. For example, when applying data science theory, the data can represent each nurse, the unit in a hospital, or the hospital system. Each of the "5Vs" is to be considered every time data science concepts are taught. By considering the "5V's," nurse scientists are reminded of issues that may be present within the data to challenge the integrity of the construct such as ensuring the data are valid representations of the construct. Exposure to the DSCOM constructs, level of application and consideration of the 5Vs of data are valuable to every nurse scientist. As students begin to specialize after core courses, advisors will recommend additional training specific to the needs of a data-intensive nursing PhD and Nurse Data Scientist. However, through exposure during the core elements of a PhD education, students will have a stronger foundation in data science concepts. As stated earlier in the development of the model, by design the constructs do not overlap so that the constructs can be deconstructed and embedded within a course's curriculum. The six constructs of the DSCOM with their underlying concepts will be described below.

Domain is a specified sphere of knowledge, influence, or activity (Merriam-Webster, 2018a). Domain is a critical construct that makes it specific to a profession such as nursing. Health care professionals such as nurses, physicians, pharmacists, and social workers (to name a few) are domain experts in the application, understanding, and discovery of patient-centered services. The importance of including expertise in nursing domain knowledge is preeminent when

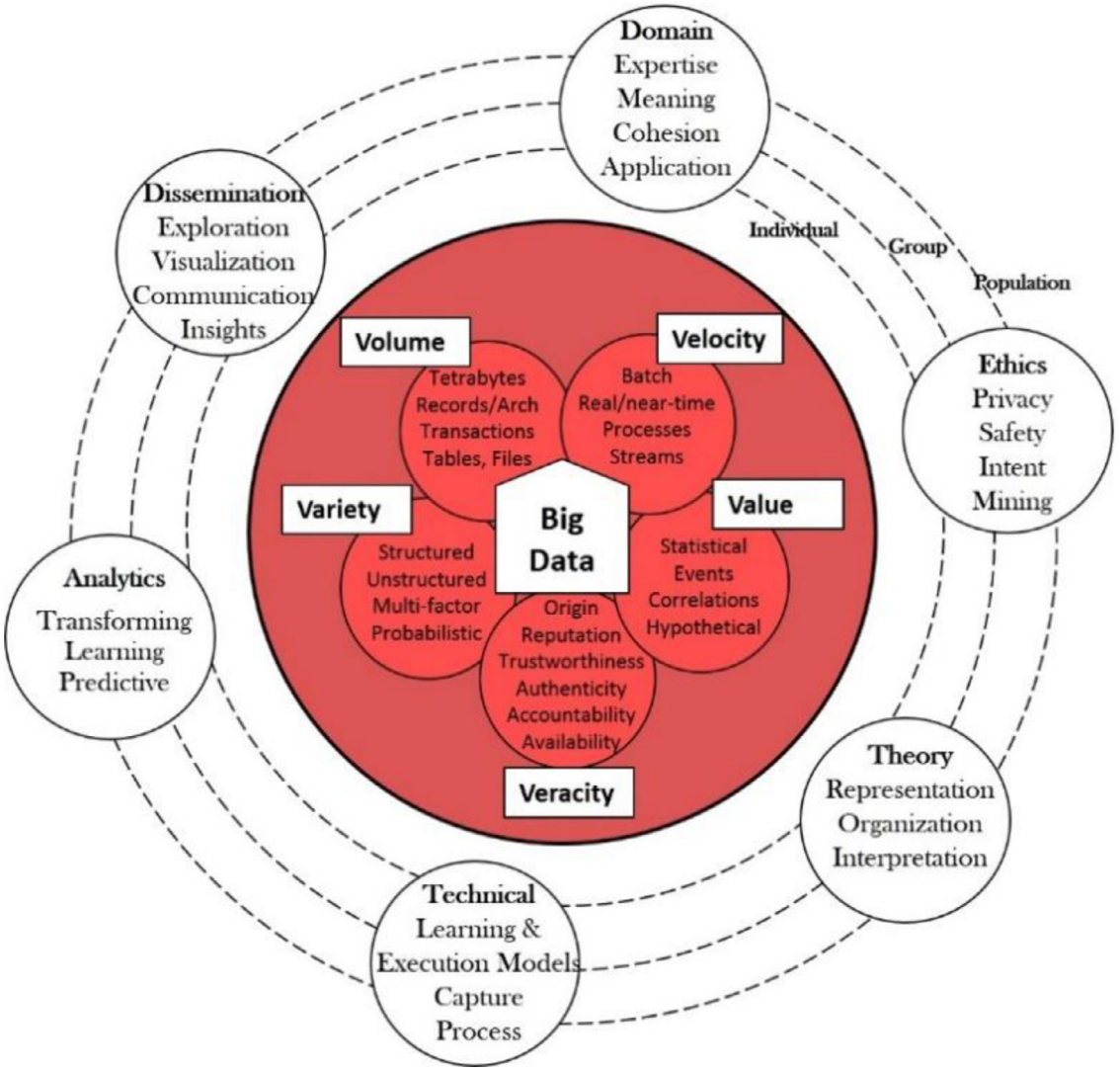

Figure 2 – Data Science Curriculum Organizing Model (DSCOM).

computers have the capability of collecting vast amounts of data. Simply searching for statistical significance through mining big data with software programs can find highly correlated relationships among variables. Domain knowledge is necessary to interpret if there is a basis for the meaning, cohesiveness, and application of the relationships among those variables. When examining the 5Vs of Big Data in raw data, the expertise that the nurse scientist represents enables her to examine the cohesiveness and applicability of findings, resulting in meaningful conclusions.

Concepts identified within the construct of *Domain* are:

- Expertise—understanding of nuances derived from exposure and knowledge of the profession.
- Cohesion—continuity of data, findings and context.
- Application—ability to relate data driven findings to real world context.
- Meaning—interpretation of information applicability within a context.

Ethics are moral principles that govern a person's behavior (Merriam-Webster, 2018b). Ethics are common to any professional curriculum and data science is no exception. Not only do ethics govern data management but provide insight to the veracity of data collection and reporting. Nurses involved with data need to participate in structured guidance to critically reflect on the management of data and the allocation of responsibilities that are associated with data collection and evaluation (Leonelli, 2016).

Concepts identified within the construct of *Ethics* are:

- Privacy—determination of what will be shared with a third party.
- Safety-assure that each human subject or clinical investigation has a system for appropriate oversight and a process for monitoring.
- Intent—an a priori reason for collecting data to address an identified question and consideration of the potential impact of contribution.

- Manipulation—falsification, misrepresentation or fabrication of data or the scientific message.

Theory in nursing is defined as "a creative and rigorous structuring of ideas that project a tentative, purposeful, and systematic view of phenomena" (Chinn & Kramer, 2018, p. 190). Nursing theory supports domain knowledge. Hypotheses for research outcomes are based on a well-developed theory. Nursing is a profession that deals with human challenges. Introspective and prospective philosophies and theories are valuable to anchor and guide practice and research. Theory knowledge is a foundational component of nursing stretching the continuum from grand, middle-range, and situation-specific theories. Datum in isolation inherently has no meaning, within the context of a theory; the data are represented, organized, and interpreted. A theory enables data interpretation to have a profound influence on knowledge and application. Although problem-driven analysis of data can "put out the fires" in health care systems, analysis and application of big data that is not theory driven is rarely generalizable to the bigger problems that exist (Verran, 1997).

Concepts identified within the construct of *Theory* are:

- Representation-varying aspects of *how data* classify all possible *data* situations according to a few united concepts.
- Organization—the way you categorize and determine relationships that will impact the effort required for answers.
- Interpretation—making sense of qualitative or quantitative data that have been collected, analyzed, and presented.

Technical is defined as having special and usually practical knowledge, especially of a mechanical or scientific subject (Merriam-Webster, 2018c). Technical knowledge is not commonly found within the domain

Table 1 – Matches of DSCOM Concepts to Material in Core Course Syllabi

Constructs and Concepts From DSCOM	# Matches With Syllabi Material	Course Numbers (# of Matches)
Domain		
Expertise	1	N646 (1)
Meaning	45	N705 (43), N731(2)
Cohesion	1	N646
Application	8	N731(1), N646 (7)
Total	**55**	
Ethics		
Privacy	2	N731(1), N646 (1)
Safety	0	n/a
Intent	0	n/a
Mining	0	n/a
Total	**2**	
Theory		
Representation	3	N646 (3)
Organization	1	N705 (1)
Interpretation	33	N731(1), N646 (2), N706 (30)
Total	**37**	
Technical		
Learning & Execution Models	0	n/a
Capture	2	N646 (2)
Process	3	N731 (1), N646 (2)
Total	**5**	
Analytics		
Transforming	14	N730 (14)
Learning	0	n/a
Predictive	0	n/a
Total	**14**	
Dissemination		
Exploration	0	n/a
Visualization	5	N646 (5)
Communication	1	N646 (1)
Insights	4	N646 (4)
Total	**10**	
Total Entries	**123**	

Note. DSCOM, data science curriculum organizing model; N631, advanced statistics for the health sciences; N646, healthcare informatics: theory and practice; N705, philosophy of nursing science and practice; N706, theory development and evaluation; N730, quantitative methods in nursing research; N731, qualitative methods in nursing research.

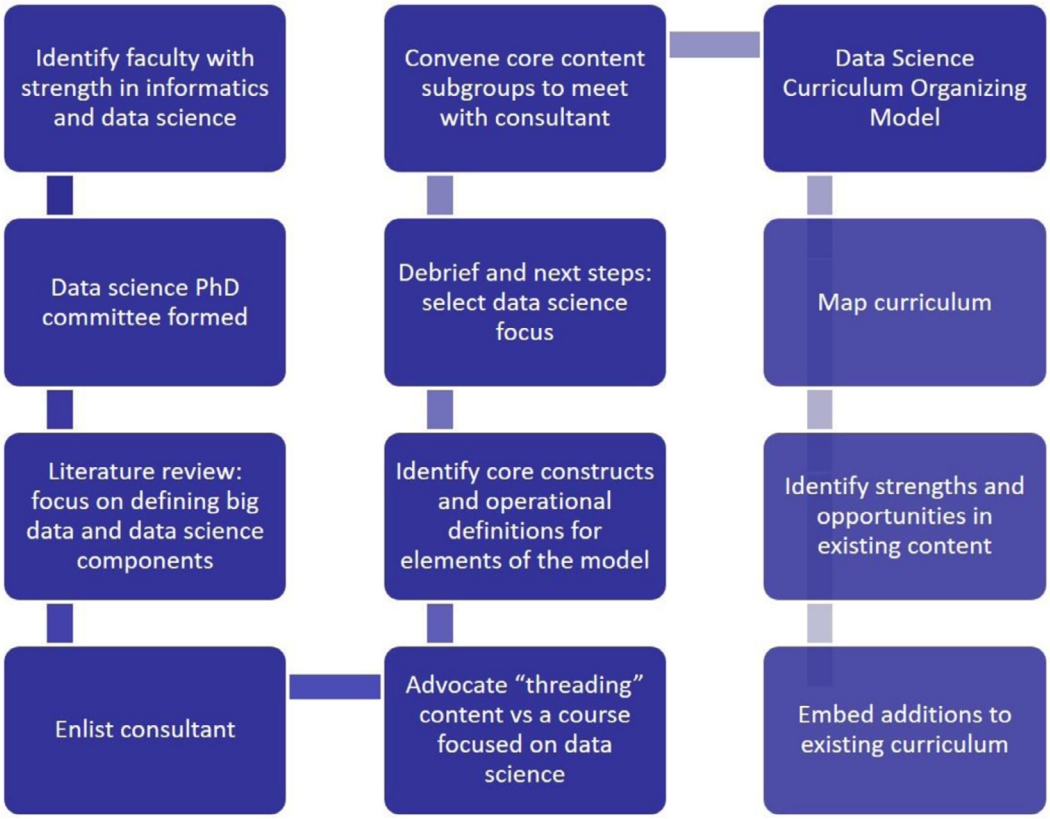

Figure 3 – Process for the application of the DSCOM. DSCOM, data science curriculum organizing model.

knowledge of nurses. "Computer Science" is the construct used in the many different originating versions of the Venn diagrams of data science (Conway, 2010). The necessary hardware and software programs involved with data collection, management and analysis are the domain knowledge of computer engineers and Information Technology (IT) specialists. In the DSCOM, we transformed "Computer Science" to "Technical" knowledge because this knowledge would encompass a necessary basic understanding of what is required to accommodate the "5Vs." Technical concepts such as data capture and processing are vital to understand if data veracity is maintained. An intimate knowledge of algorithm components and contributions to learning and execution models requires extensive education. Today's nurses make decisions based on technological functions that are part of everyday practice. Computer science courses are not a usual part of nurse PhD curricula, so technical concepts may be added to Informatics courses. However, to acquire knowledge needed for the Nurse Data Scientist, additional courses outside the nursing college/school may be required.

Concepts identified within the construct of *Technical* are:

- Models—learning (how the analytics are used for judgments that change over time) and execution (how data are manipulated to be embodied by an execution framework such as programming language or a distributed computing system).
- Capture—input of data using technology such as barcode scanners, optical readers, voice recognition.
- Process-collecting, transforming raw data into a machine-readable form.

Analytics are techniques that give value to structured or unstructured data. Analytic methods are part of every data science model. Without analytics, it would be nearly impossible to derive meaning from large datasets and big data. Three classes of analytic techniques are transforming, learning, and predictive (Herman et al., 2015), which are the concepts within the DSCOM analytics construct. Nurse scientists inevitably ask questions that can be answered using one of these techniques. By having a basic understanding of each of these classes of analytics, nurses can choose which techniques will derive the greatest value from the data.

Concepts identified within the construct of *Analytics* are:

- Transforming—techniques for aggregation (summarization and plotting), enrichment (adding data) and processing (data cleaning, preparation, and separation).
- Learning—techniques for regression, clustering, classification, and recommendation.
- Predictive—techniques for simulation, and optimization to maximize utility of data.

Dissemination is wide sharing or spreading of information. Big data have become the basis for growing businesses through decisions and marketing. Marketing and dissemination go hand-in hand, especially when sharing knowledge is a main consideration. Exploration, visualization, communication, and insights are concepts that promote data interpretation. In accordance with the pending decision and decision makers, dissemination ensues. Through dissemination, findings from data science analysis that produce credible evidence spread knowledge. An understanding of concepts such as exploration of best places to disseminate to reach the target audience; visualization to provide easy and quick transformation of information into practice; communication methods to ensure that a connection is being made between report senders and receivers; and insights into how data will provide value. Without dissemination, data cannot be translated into action or practice.

Concepts identified within the construct of *Dissemination* are:

- Exploration—investigation of where the information will do the most good for fulfilling the research purpose and advancing nursing science.
- Visualization—images that contain information.
- Communication—a sender and receiver of information.
- Insights—observation of the impact of the information.

DSCOM Application

A Data Science Task Force was formed, including Nurse Informatics and Systems faculty experts. Initial research by the Task Force identified the national data science expert who would understand the challenges with academic course revisions. Consideration for faculty workload was of the utmost concern. Based on consults with the expert, the decision was made to thread changes throughout the PhD core curriculum courses (i.e., theory, informatics, quantitative methods, qualitative methods, and statistics). Benefits to threading the DSCOM concepts throughout courses are that small changes for each course are more doable and no additional credit hours are required for the PhD plan of study. Faculty buy-in was achieved through the process of a general faculty meeting with a national nursing consultant on data science that was open to all faculty with purposive inclusion of core course faculty. The general faculty meeting was followed-up with small meetings with each course chair, the national consultant and a Data Science Task Force member. The process of course mapping to identify data science concepts was reviewed and discussed with each chair. Faculty who taught the core courses for academic year 2015 to 2016 agreed to have their syllabi evaluated using DSCOM. All concerns and questions were addressed and buy-in was achieved. After the national expert left, materials such as the course syllabi and content outlines were sent to the Task Force for review. The review included two Task Force members identifying DSCOM concepts in the material. In the case of the "informatics" core course, the course chair was a member of the Task Force and not permitted to review her own materials. If clarification was required course chairs were consulted. The review information was recorded in a spreadsheet for analysis. The spreadsheet contained each DSCOM concept and the core course number. The course materials were reviewed using the DSCOM concept definitions and when a topic within the course was identified as consistent with the definition of the concept, the course was noted. In the analysis, the course was counted as a potential match for embedding content about data science. For example, in a course entitled, *Philosophy of Nursing Science and Practice,* the "structure of nursing knowledge" was identified as an opportunity to embed the "Organization" of data science from within the DSCOM *Theory* construct. Table 1 displays a basic frequency analysis of the number of entries from course material identified as matches with concepts.

In total, there were 123 matches for opportunities to embed data science concepts. Construct match frequencies were *Domain* (55), *Ethics* (2), *Theory* (37), *Technical* (5), *Analytics* (14), and *Dissemination* (10). When there were concept matches to two or more courses, discussions with course chairs provided insight as to where the embedding was most likely. There were seven concepts that did not have any matches, representing gaps in the course content. To our surprise, three content areas within *Ethics* (safety, intent, and mining) had no matches, so additional content was added to the video that all PhD students are required to watch. Two concepts within *Analytics* (learning and predictive) and one within *Technical* (learning and execution models) were not matched. Currently, students who would like to move in the direction of Nurse Data Scientist can be offered an outside course, such as computer science, that covers these gaps in concepts from the DSCOM until a new course can be developed. Figure 3 displays the process before and after the application of the DSCOM. All core faculty chairs were receptive to additional content for their course, as long as there was a fit. A basic overview of DSCOM concepts is all that will be added to core courses. An academic year following the inclusion of data science threaded into the core curriculum concluded in 2018. Curricular changes involving Data Science will be ongoing due to the rapidly evolving nature of this science and are now the responsibility of each course chair. A re-evaluation of the DSCOM is required to determine the depth of knowledge within the concepts to be part of core data science knowledge as compared to depth that would be contained in specialty courses for Nurse Data Scientists. Feedback is to be gathered from faculty about the experience of embedding concepts into existing curriculum is in process. Natural Language Processing has also been used to analyze the matches between course material and the DSCOM concepts. The findings have been presented nationally (Carrington et al., 2017) and a manuscript is in process.

Conclusions

Leaders in nursing and advocates for big data have provided a strong impetus for changing the current insufficient educational dataset-inquiry content in nursing schools (Brennan & Bakken, 2015). Data science application is the future of health care and nurses must be exposed to the many concepts to maintain their roles as change agents. The DSCOM is a model that is based on the evolution of prior models designed for data science curriculum evaluation. The evolved model contains broad constructs: *Domain, Ethics, Theory, Technical, Analytics, and Dissemination*, with concepts compiled for use in nurse scientist curricula. The existing content covered in PhD nursing education is extensive and provides courses that have topics where content covering data science concepts could be readily embedded. Therefore, using the DSCOM to guide mapping of data science content within each core course is imperative. When the constructs in the DSCOM are threaded throughout the PhD nursing core curricula, nursing colleges/schools will fulfill the obligation to educate nurse scientists to teach, build clinical knowledge and conduct research by effectively using the abundance of available data. When the DSCOM is used as a guide, keys concepts are indicators as to the comprehensiveness of a PhD nursing curriculum to provide data science preparation for nurses to be strong leaders in the reform of health care practice and delivery. Including data science in PhD curricula, nurse scientists will be prepared to organize, collect, manage, analyze, and most importantly apply data to continually improve patient care delivery using data based decision-making and policy development to transform health care. An unexpected consequence was that faculty advisors now have a delineation for discussing how in-depth each PhD student would like to become with data science studies. Nurses who want to become more intense in data science will be fully prepared to take additional courses as needed. Mapping of DSCOM concepts in existing course content is challenging process and refinement is needed. Using the DSCOM as a foundation for course revision provides faculty with information that will foster change in curricula that is inclusive of data science. Further applications of the DSCOM are encouraged to strengthen and continue the evolution of data science as a component of doctoral education for nurse researchers to guide care delivery for individuals, groups, and populations.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Dr. Gephart received training support from the Agency for Healthcare Research and Quality (K08HS022908) and the Robert Wood Johnson Foundation Nurse Faculty Scholars Program (72, 114). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Robert Wood Johnson Foundation or the Agency for Healthcare Research and Quality.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.outlook.2018.10.007.

REFERENCES

AACN. (2016). *Enhancing nursing science and improving patient care through Big Data: PhD-DNP*. Retrieved from http://www.aacn.nche.edu/downloads/meetings/2016/doc/2016-Doctoral-Schedule.pdf.

AACN and Manatt Health. (2016). *Advancing healthcare transformation: A new era for academic nursing*. Retrieved from https://www.manatt.com/Insights/White-Papers/2016/Advancing-Healthcare-Transformation-A-New-Era-for-Academic-Nursing.

Brennan, P. F., & Bakken, S. (2015). Nursing needs big data and big data needs nursing. *J Nurs Scholarsh, 47*(5), 477–484, doi:10.1111/jnu.12159.

Broome, M. E. (2016). Big data, data science, and big contributions. *Nurs Outlook, 64*(2), 113–114. http://dx.doi.org/10.1016/j.outlook.2016.02.001.

Carrington, J. M., Brewer, B., Davis, M., Gephart, S. M., Shea, K. D., & Rosenfeld, A. R. (2017). Use of NLP to analyze core course content and weave in data science. *Paper presented at the AMIA inspire educational conference.*

Chinn, P., & Kramer, M. (2018). *Knowledge development in nursing: Theory and process* (Tenth Edition). St. Louis, MO: Elsevier, Inc.

Clancy, T. R., Bowles, K. H., Gelinas, L., Androwich, I., Delaney, C., Matney, S., & Westra, B. (2014). A call to action: Engage in big data science. *Nurs Outlook, 62*(1), 64–65. http://dx.doi.org/10.1016/j.outlook.2013.12.006.

Conway, D. (2010). *The Data Science Venn Diagram*. Retrieved from http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram.

Effken, J. A. (2003). An organizing framework for nursing informatics research. *CIN: Comput Inform Nurs, 21*(6), 316–323.

Executive Office of Health and Medicine. (2013). Roundtable on value & science-driven health care. *Paper presented at the leadership consortium for a value & science-driven health system.*

Herman, M., Rivera, S., Mills, S., Sullivan, J., Guerra, P., Cosmas, A., & Kim, M. (2015). *The Field Guide to Data Science* (2nd edition). Retrieved from https://www.boozallen.com/s/insight/publication/field-guide-to-data-science.html.

Khoso, M. (2016). *How much data is produced every day?* Retrieved from http://www.northeastern.edu/levelblog/2016/05/13/how-much-data-produced-every-day/.

Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety*. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Leonelli, S. (2016). Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems. *Philos Trans R Soc*, A374(20160122). http://dx.doi.org/10.1098/rsta.2016.0122.

Merriam-Webster. (2018a). https://www.merriam-webster.com/dictionary/domain.

Merriam-Webster. (2018b). https://www.merriam-webster.com/dictionary/ethic.

Merriam-Webster. (2018c). https://www.merriam-webster.com/dictionary/technical.

Verran, J. (1997). The value of theory-driven (rather than problem-driven) research. *Semin Nurse Manag*, 5(4), 169–172.