Glossary

**Aggregation** is a process of finding, gathering and merging data.

**Algorithm** is a mathematical formula that gives the computer a set of rules to follow to perform data analysis. Think of an algorithm as a set of directions or a recipe for combining data to get a solution.

**Anomaly detection** is the search for data in a data set that does not match a projected pattern. Anomalies are also known as outliers. They may provide critical and actionable information.

**Anonymization** is the process of making data anonymous; no ability to attribute the data to a specific individual; removing all data that could identify a person.

**Application program interface (API)** is a set of routines, protocols, and tools for building software applications. It specifies how the software components should interact and how to build the graphical user interface (GUI) so that it interacts with the software.

**Big data** is a term for data sets that are so large or complex that traditional data processing applications are inadequate. The data is characterized by volume, velocity, and variety. These very large data sets may be analyzed to reveal patterns and relationships, particularly about human behavior and interactions.

**Causality** is the relationship between cause and effect. This is often the goal of research.

**Classification analysis** is a process for obtaining information about data; also called metadata.

**Cloud computing/storage** is a distributed computing system over a network of remote servers hosted in the Internet rather than on a local device; used for storing data off site; saving a file to the cloud ensures access with any computer that has an Internet connection.

**Clustering analysis** is a statistical process for identifying objects that are similar to one another and to cluster them to reveal both similarities and differences.

**Commodification** is the transformation of data, ideas, services, and products into objects of trade. These data, ideas, services, and products become commodities in the marketplace.

**Comparative analysis** is a specified process of comparisons and calculations to detect patterns within very large data sets.
Complex structured data is composed of two or more complex, complicated, and interrelated parts that cannot be easily interpreted by structured query languages (SQL) and tools.

Computer generated data is simply data that is generated by a computer as it does its calculations, e.g. log files, time stamps, algorithm checking.

Correlation analysis is a statistical process to determine the relationship between variables; the relationships may be positive or negative.

Dashboard is graphical representation(s) of one or more analyses performed by an algorithm; only the results are shown, not the data or the calculations.

Data refers to a description of something that allows it to be recorded, analyzed, and reorganized; the observation and measurement of a phenomena created data.

Data analyst is someone who cleans/wrangles, analyzes, models, and processes data.

Data ethical guidelines guide an organization in making data management transparent. This is part of insuring privacy and security for the data.

Data feed is a live streaming of data. This is used by Twitter, news feeds, and RSS (really simple syndications) feeds.

Data governance is the management of the availability, usability, integrity, and security of the data owned by an organization. A program of data governance includes a governing body, policies, procedures, and plans to execute the procedures.

Data lake(s) refer to a massive, easily accessible data repository designed to retain all data attributes and built on relatively inexpensive computer hardware for storing big data.

Data modeling is the analysis of data objects using data modeling techniques (such as Unified Modeling language [UML]) to create insights concerning the data.

Data science. While there is no widely accepted definition of data science, several experts have made an effort. Loukides (2012) says that using data isn’t, by itself, data science. Data science is using data to create a data application that acquires value from the data itself and creates more data or a data product. Dumbill says that big data and data science create “the challenges of massive data flows, and the erosion of hierarchy and boundaries, will lead us to the statistical approaches, systems thinking and machine learning we need to cope with the future we’re inventing” (p.17, 2012). O’Neil and Schutt (2014) add the following skills for data science: computer science, math, statistics, machine learning, domain expertise, communication and presentation skills, and data visualization. A further distinction about data science is that the product of engaging in data science is creating a data product that feeds data back into the system for another iteration of analysis, a practical endeavor not traditional research.

Data scientist is a person who is able to search for data and develop algorithms to process the data. This may involve programming and statistics.

Data set is a collection of data.

Data visualization is the representation of data in a visual format that is a complex graph that includes many variables while remaining understandable and readable.
Data wrangling/cleaning/munging are synonyms for the process of reviewing and revising data to delete duplicates, correct errors, deal with missing data, provide consistency, and standardize formats.

Datafication turns a phenomenon into a quantified format so it can be tabulated and analyzed. The earliest foundation of datafication is the measuring and recording which facilitated the creation of data.

Database is a digital collection of data stored using specified techniques depending on the type of database. Databases can be hierarchical, relational, object, graphical, or a hybrid.

Digitization makes analog information readable to computers; makes it easier to store and process. Digitization is the process of converting analog information into the zeros and ones of binary code so computers can handle it.

Distributed file system is a system that stores, analyzes, and processes data from many sites.

Electronic Health Record (EHR) is a longitudinal record, stored in a database, of a patient’s health information from their encounters in all care settings. It includes demographics, health history, problem list, medications, progress notes, check lists, immunizations, laboratory and diagnostic tests, images, vital signs, consultations, and therapies received by the patient.

Epigenetics is the study of inheritable changes (either mitotically or meiotically) that alter gene expression and phenotypes, but are independent from the underlying DNA sequence.

Exploratory data analysis was proposed by John Tukey in 1977. The procedure describes the data and finds its main characteristics. It also finds patterns in the data without standard procedures or methods.

Exposome describes the complementary environmental component of the gene-environment interaction indicative of complex traits and diseases.

Extract, transform, and load (ETL) is a database process that identifies and moves a set of data from one database to another. It is also used for the same purpose in data warehouses.

Fault-tolerance design is used to design computer systems that will continue to work if part of the system fails.

Funding opportunity announcements (FOAs) are announcements posted by the federal government, foundations, or other funding bodies. These FOAs solicit program or research proposals for specified target areas of research or services.

Genetic risk scores (GRS) are developed using algorithms about genetic risk (based on big data) to predict the risk for a specific individual. Genetic risk is the probability that a trait will occur in a family. The probability is based on the genetic pattern of transmission.

Grid computing is connecting different computer systems from different locations. The connection is often done via the Cloud.

Hadoop is an open-source Java-based programming framework that supports the processing of large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation.
**HBase** is a NoSQL database designed to work with Hadoop when the volume of data exceeds the capacity of a relational database.

**Hadoop distributed file system (HDFS)** is a file system designed to work with Hadoop. HDFS is a file system that stores data on multiple computers or servers. The design of HDFS facilitates a high throughput and scalable processing of data.

**Health disparity** is a particular type of health difference that is closely linked with social, economic, and/or environmental disadvantage.

**Health equity** is the attainment of the highest level of health for all people. Achieving health equity requires valuing everyone equally with focused and ongoing societal efforts to address avoidable inequities, historical and contemporary injustices, and the elimination of health and health care disparities.

**In-memory database** is a database management system that stores data in the main memory of the computer instead of on a disk. This characteristic facilitates very fast processing, storing and loading of data.

**Internet of Things (IoT)** are devices with sensors that connect to the Internet. The devices generate data and can be analyzed for relationships.

**Interoperability** is the ability of health information systems to share data and information within and across organizational boundaries to promote effective health care.

**Location data** is the data generated by Geo-Positioning Satellites (GPS). The data is recorded in longitude and latitude and describes a geographical location.

**Log file** is a file that is generated by a computer that documents all events taking place in the computer while it is operational.

**Machine data** is data created by machines by sensors or algorithms.

**Machine learning** is a subset of artificial intelligence. Through algorithms machines learn from what they are doing and become more efficient over time. Machine learning is a key component of data science and is used in big data analysis.

**MapReduce**, invented by Google, is software for processing very large amounts of data. The MapReduce algorithm is used to divide a large query into multiple smaller queries. Then it sends those queries (the Map) to different processing nodes and then combines (the Reduce) those results back into one query.

**Massively parallel processing (MPP)** uses numerous processes, located in many separate computers, to perform computational tasks at the same time.

**Mathematical model** is an abstract model that uses mathematical language to describe the behavior of a system.

**Metadata** is data about data; giving information about the characteristics of the data.

**News feed(s)** refer to continuous transmission of data (consisting of news updates) to websites through a syndicated news service provider. Subscribers receive the news feed(s) or web feed(s) as summaries or links to the original news source.

**NoSQL databases** are used when the volume of data exceeds the capacity of a relational database.
Nurse data scientists are educated as nurses and then pursue a research doctorate or post-doctorate in a data science field (data analytics, computer science). The primary research focus of a nurse data scientist is on methods and analytics as opposed to specific health and illness concerns of individuals, families, communities or populations.

Nursing informatics is defined by the American Nurses Association as the specialty that combines nursing science, computer science, and information science to manage and communicate data, information, knowledge, and wisdom in nursing practice.

Oomics is the application of powerful high through-put molecular techniques to generate a comprehensive understanding of DNA, RNA, proteins, intermediary metabolites, micronutrients and so forth involved in biological pathways resulting in phenotypes.

Ontology, from a computer science perspective, is created to represent knowledge as a set of concepts and their relationships with one another within a domain. Ontologies limit complexity and organize information, thus they can be used to solve problems.

Outlier is a piece(s) of data that deviate significantly from the other data in the data set. It is important to detect these during data wrangling/cleaning and exploratory analysis since it might indicate something useful happening.

Pattern recognition is identifying patterns within the data using algorithms. It is used to make predictions about new data coming from the same source.

Population health refers to the management and improvement of health outcomes for a group of individuals, including the distribution of such outcomes within the group. Managed Care under Population Health organizes populations and panels under the care of delivery systems, practices and physicians with accountability for the health of all enrollees and for the resources and costs of providing this care.

Portability is the ability of different types of hardware that allow software to operate on a variety of platforms employing different operating systems.

Precision medicine/personalized healthcare is a medical model that proposes to customize healthcare by incorporating medical decisions, practices, and products that are based on individual variability in genes, environment, and lifestyle.

Precision nursing uses big data from diverse sources; genetic records, medical and insurance records, data from social media, and wearable sensors are effectively harnessed to outline a detailed picture of the patient and offer a customized healthcare solution. Using big data customized treatment plans to specific individuals based on their preferences are provided. Precision nursing enhances the nurse’s ability to detect complex nursing problems during initial stages which is imperative for effective and successful treatment and to offer treatment for lifestyle-related diseases by intrinsically analyzing data pertaining to lifestyle patterns of patients.

Public health refers to the function of state and local governments to provide services for preventing epidemics, containing environmental hazards, and encouraging healthy behaviors. The Future of the Public’s Health in the 21st Century calls for significant movement in “building a new generation of intersectoral partnerships...
that draw on the perspectives and resources of diverse communities and actively engage them in health action.”

Python is a general purpose programming language created in the late 1980s, and named after Monty Python. It is considered to be the optimal language used in data science by people with a computer science background.

Quantified self is a movement to use applications to track an individual’s every move (activity) during the day to better understand one’s behavior and health.

Query is asking a question of the data to gain information to answer a question. It is done through a query language, e.g. SQL, Hive, or Pig.

R is a programming language for statistical computing and graphics. It is supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

Re-identification is a process for re-identifying an individual from an anonymized data set.

Radio Frequency Identification (RFID) is a sensor that uses a wireless non-contact radio-frequency electromagnetic field to transfer data.

Real-time data is data that is created, processed, stored, analyzed, and visualized in milliseconds.

Schema-on-read is a data analysis strategy in new data-handling tools like Hadoop and other more involved database technologies. In schema-on-read, data is applied to a plan or schema as it is extracted out of a stored location, rather than as it is entered.

Schema-on-write has been the standard in relational databases. Before any data is entered, the structure of that data is strictly defined, and that metadata stored and tracked. Irrelevant data is discarded and data types, lengths and positions are all defined and enforced with constraints.

Secondary use of health data applies to patient data used, not for the delivery of care, but for other purposes. These purposes may be for research, quality and safety measurement, public health, billing/payment, provider credentialing, marketing, and other entrepreneurial applications.

Semi-structured data is a form of structured data that does not have a formal structure like structured data (using a set of standards or terminology to specify meaning), but does have tags, metadata, or other markers to enforce a hierarchy of records.

Sentiment analysis uses algorithms to determine how people feel about certain topics.

Signal analysis is the extraction of information from complex signals in the presence of noise, generally by conversion of the signals into digital form followed by analysis using various algorithms. This is important when testing sensor data that uses time or other varying physical quantities.

Social determinants of health are conditions that shape a person’s health: where they are born, grow, live, work and age, including the health system, and distribution of resources at global, national and local levels.
Structured data is data that is identifiable as it is organized in rows and columns. The data resides in fixed fields within a record/file (as in a relational database).

Transactional data is dynamic data that changes over time.

Transparency is a process that data owners provide consumers that generate the data to inform the consumer how the data is being used.

Unstructured data is data that is usually text in nature, though numbers and dates may be included. There is no known location for the data as there is for structured data.

Value is generated from data by the decision made and the products produced from that data.

Variability occurs when the meaning of the data can change rapidly. For example, in the same tweet a word can have more than one meaning.

Variety indicates the many different formats that data has in the big data world. The data is not ordered, due to its source or collection strategy, and it is not ready for processing. Even the data sources are highly diverse: text data from social networks, images, or raw data from a sensor. Big data is known as messy data with error and inconsistency.

Velocity is the speed at which data is created, stored, analyzed and visualized. Data flows into systems and is processed in batch, periodic, near real time, or real time.

Veracity is the correctness or integrity of the data. This should be established before analysis is performed.

Volume is the amount of data. Data volume is quantified by a unit of storage that holds a single character, or one byte.

Wisdom and clairvoyance, in big data, is the ability to predict and correct before a user knows something is wrong. Traditionally, wisdom is the ability to think and act using knowledge, experience, understanding, common sense, and insight.

YARN (Yet Another Resource Negotiator) is a management system that keeps track of CPU, RAM, and disk space and insures that processing runs smoothly.
Index

A
ABC codes, 121
Academic/corporate partnerships
innovation and management of intellectual property, 165–168
OptumLabs, 165–168
predicting adverse events prescribing statins using OLDW, 172–179
Academic nursing. See also Education, nursing
big data and, 374–375
impact of data proliferation on, 189–191
Accountable Care movement, 337
Accountable Care Organizations (ACO) demonstration projects, 16
Accreditation
of graduate health informatics programs, 391–396
international accreditation programs, 431–432
ACOG App, 320
Acoustic stethoscope, 69
Active learning, 280
Activity monitors, 72–73
Acute care
identifying direct nursing cost per patient episode in acute care case study, 359–363
mobile apps for nurses, 321–322
Acute Mental Status Change (AMSC)
VA’s VINCI database case study, 287–298
Administrative data, 81
Advanced Health Informatics Certification (AHIC), 388–389
Advocate Cerner Collaboration (ACC), 364
Advocate Health Care, 364
Affordable Care Act (ACA)
incentives to leverage healthcare data, 230–231
Agency for Healthcare Research and Quality (AHRQ), 231
Alarm management case study, 345–351
Alert design, 291
Algorithms
genetic algorithm, 89–90, 279
machine learning and, 163
Alert fatigue, 345
Alzheimer’s patients
remote monitoring systems and, 75
Amazon Web Services, 4
American Association of Colleges of Nursing (AACN)
information management as core competency, 375
nursing research defined, 190
American Congress of Obstetrics and Gynecology
mobile app from, 320
American Medical Informatics Association (AMIA)
informatics certification, 385–389
American Nurses Association (ANA), 7
NDNQI clinical data repository, 143
nursing informatics specialty, 377–378
standard nursing languages, 120–121
American Nurses Credentialing Center (ANCC), 122, 401, 402
American Recovery and Reinvestment Act, 121
American Society of Statistics, 40
Analytics
big data in nursing, 6
causal discovery modeling methods, 272–275
curse of dimensionality, 266
data mining, 83–84 (see also Data mining)
defined, 12
feature selection, 178, 270–271
growing demand for, 5
inpatient nursing unit example of, 125–127
kernel-based transformation, 269–270
machine learning, 85–86 (see also Machine learning)
need for analytic infrastructure with HCS, 261–262
overfitted models, 266
predictive modeling, 85, 265–269
retailer use of, 12–13
text mining, 84–85 (see also Text mining)
Analytics Institute, 261–262
Apache Spark, 5
Apple, 315, 316
Applied Evidence-Based Practice Model, 246–247
Apps. See Mobile apps, for nursing
AquaData Studio® (ADS), 175
Articles
on big data, 48
on data science, 50–51
on data visualization, 52
Artificial knee replacement registry, 231
Artificial Neural Networks (ANNs), 279
Assessment
within defined limits (WDL)
assessments mapped to standardized terminology, 248
Association of American Universities (AAU), 190
Association of Research Libraries (ARL), 190
Asthma, mobile apps for care management, 324
Audio files, as data source, 82
Augmented cognition, 74

B
Bayesian methods, 278
Beautiful Soup, 41
Beers criteria app, 324
Betty Irene Moore School of Nursing (SON) nursing’s role in pSCANNER, 223
Bias
machine learning and, 164
observational studies, 168–169
predictive analysis and, 21
Big data
academic nursing and, 374–375
books on, 48
case studies of (see Case studies)
certificates and training in, 54–55
challenges of, 90–91
characteristics of, 37–39, 80, 256
compared to data science, 116
conferences on, 47
data sources for, 81–82
definition of, 4, 8, 12, 33–34, 62, 80, 116, 409
doctoral education in nursing, 407–423
to eliminate health disparities, 240–241
emerging team research roles for, 167, 168
evaluating technology of, 36–37
future possibilities in, 441–461
global society and, 427–439
in healthcare (see Healthcare, big data in)
historical perspective of, 34–36
impact of, 4
integrating nursing data into, 235–236
nursing informatics and, 334–335
professional associations for, 54
rapid growth in, 188–189
research and, 185–187
resources for, 47–49
retailer use of, 12–13
as team sport, 43–45
trends in, 4–6
ubiquity of, 408
value of, for nurse leaders, 119–120
variety, 38, 80, 256
velocity, 39, 80, 256
volume, 33–34, 37–38, 80, 256, 332–333
Big data analytics. See Analytics
Big Data to Knowledge (BD2K) program, 188–189, 191, 192–193, 232–233
BigQuery, 5
Bloom’s Taxonomy, 55–56
Books
on big data, 48
on data science, 50–51
on data visualization, 52
Boosting, 278
Bootstrapping, 276
Bourne, Phil, 192
Boyer, E. L., 190
Burroughs, Thomas, 386
Index

C
CareAware AlertLink, 345
Care coordination, mobile apps for nurses, 323–324
Care coordinator role, 66
Case studies
  academic/corporate partnership
    prescribing statins using OLDW, 172–179
  accreditation of graduate health
    informatics programs, 391–396
  alarm management, 345–351
  AMIA informatic certification, 385–389
  big data resources, 46–57
  CPM framework for clinical practical
    interoperability and comparative
    analysis, 244–250
  delirium and VA’s VINCI database, 287–298
  general to specific and failure to diagnose
    case study, 136
  home healthcare sector, 23–25
  identifying direct nursing cost per patient
    episode in acute care, 359–363
  improving nursing care through Trinity
    Health System Data Warehouse, 129–136
  Interdisciplinary Plans of Care (IPOCs)
    case study, 132–134
  Kaiser Permanente interprofessional EHR
    data to prevent delirium, 304–310
  NHS Spine, 25–28
  nursing time in EHR case study, 352–357
  pressure ulcer case study, 134
  readmission prevention with learning
    health system, 364–369
  Type 2 diabetes and big data, 197–206
  value-based nursing care model
    development, 95–100
  venous thromboembolus (VTE) advisory
    case study, 135
  Veterans Health Administration (VHA), 18–23
  Catheter associated urinary tract infections
    (CAUTI) study, 126–127
  Causal discovery modeling methods, 272–275
    causal probabilistic graphs, 273–274
    overview of, 272–273
  Causal probabilistic graphs, 273–274
  Causation, big data and shift away from, 35–36
  CellTrak, 322
Centers for Disease Control and Prevention
  (CDC)
    claims data as dataset, 188
    flu outbreak prediction, 4, 35
  Public Health Genomics Knowledge,
    233–234
Centers for Medicare and Medicaid
  Services (CMS)
    incentives to leverage healthcare data, 230–231
    Oncology Care Model, 162
    registry data for policy decisions, 231
    reimbursement reform based on quality
    outcomes, 15–16
Centers of Excellence for Big Data
  Computing, 192
Cerner Corporation, 364
Certification
  AMIA informatic certification, 385–389
  quality patient outcomes and credentialing, 399–405
Chief Nurse Executive Big Data Checklist,
  123–125
Chow, M., 122
Chronic disease indicators, 237
Claims data
  nursing studies using, with linked EHR
    data, 160–164
  predictive model of diabetes, 164–165
  research databases examples, 158–159
  strengths/characteristics of, 159, 161
  used in national clinical research
    network, 213
Clancy, T., 119, 123
Classification, of delirium patients
  with NLP, 292–295
ClinGen, 203
Clinical Assessment, Reporting, and
  Tracking (CART) system, 22
Clinical Care Classification System
  (CCC), 121
Clinical Commissioning Group
  Outcomes Indicator Set, 28
Clinical data
  used by nurse leaders, 117, 119
  working toward sharable and
    comparable data, 121–123
Clinical data repositories (CDR)
  flowsheet data from EHRs, 140
  overview of, 140
  purposes of, 142–143
  sharing data between for research
    purposes, 339
Clinical data repositories (CDR) (cont.)
standardizing patient data for,
144–145, 147, 150
structure of, and querying data, 143–144
Clinical data research network
characteristics of, 215
PCORNet, 215–218
pSCANNER, 218–223
Clinical Data Research Networks (CDRNs), 339
Clinical decision support (CDS), 247
interdisciplinary approach to delirium prevention, 308–309
Clinical effectiveness research (CER),
214–215
Clinical indicators, need for standardized,
432–433
Clinical practice guidelines
mapped to standardized terminology, 248
Clinical Practice Model (CPM) framework
coding to standardized clinical terminology, 247
CPG pressure ulcer-risk for, 249–250
within defined limits (WDL) assessments
mapped to standardized terminology,
248
organization of, 245–247
origins of, 244–245
Clinical terminology, coding to standardized,
247–248
Clinical translation, accelerating research finding to, 164–165
Clinical Translational Science Awards (CTSAs), 7
goals of, 451–452
members of consortium, 141
mission of, 140–141
Clustering, 277–278
Coalition for Network Information (CNI), 190
Collins, Francis, 192
Columnar databases, 39
Common Data Model (CDM)
for PCORNet, 216–217
Common Fund LINCS program, 192
COMMONS, 233
Commonwell Health Alliance, 339
Communication, in data science process,
41–42
Compendium of Federal Datasets Relevant to Disparities Research and Programs, 239
Complex Adaptive Systems (CAS), 279
Computerized Patient Record System (CPRS), 18–22
Conferences
big data, 47
data science, 50
data visualization, 52
Confusion Assessment Method (CAM) score,
304, 308
Connecting for Health (Markel Foundation),
16–17
Consolidated Clinical Data Architecture (CCDA), 144
Consumer Privacy Bill of Rights Act, 45
Content interoperability, 247
Conway, D., 40
Coproduction of healthcare, 260
Corporate Data Warehouse (CDW)
goal of, 290–291
overview of, 19–21
Corporate partnerships. See Academic/
corporate partnerships
Correlation, big data shift towards, 35–36
Costs
direct nursing cost per patient episode in
acute care case study, 359–363
County health educator, 238
CPU (central processing unit) speed, 175–176
Credentialing
quality patient outcomes and, 399–405
Critical Capabilities, 37
Cukier, K., 4
Curse of dimensionality, 266

D
Dartmouth-Hitchcock (DH), Analytics Institute, 261–262
Data
challenges to accessing for clinicians,
305–306
claims, 158–159, 160–165, 213
digitization of, 36
EHR data for research, 160–164
future partnerships for generating big data initiatives, 447–450
messiness of, 35
missing data, 177
need to interpret, 185
nursing, 223–225
in nursing, 187–188
ownership issues, 44–45
patient, 144–145, 147, 150, 259–260
as reusable resource, 35
sampling vs. using all data, 34–35
sharable and comparable data in nursing,
120–127
Index

sharing, between CDRs for research purposes, 339
standardizing patient data for CDR, 144–145
Data application, 40
Database
delirium and VA’s VINCI database, 287–298
large healthcare research databases, 158–159
Mini-Sentinel Distributed Database, 214
Data cleansing
in OptumLabs statin project, 177–178
Data clustering, 186
Data Compression Software to Make Biomedical Big Data Universally Available, 233
Data discovery index (DDI) prototype, 193
Data extraction
in OptumLabs statin project, 176–177
Datafication, 36
Data Index, 193
Data mash-ups, 73
Data mining
achievements and challenges of, 281
defined, 83
healthcare application, 84
integrated analysis protocols, 275–277
overview of process, 83–84
synopsis of importance advances in, 275–280
Data models
combining data form distinct systems, 123
traditional federated data compared to pSCANNER, 219–220
Data munging, 41
Data processing
speed of, big data trend, 5, 67
Data product, in data science process, 40, 42
Data repositories, purposes of, 142–143
Data science
books on, 50–51
characteristics of mature capabilities in, 44
compared to big data, 116
compared to Fordism, 188–189
conferences on, 50
defined, 40, 116
foundation of, 35
as fourth paradigm, 185, 186
growth of, 188–189
initiative supporting research and, 191–195
process of, 40–42
resources for, 50–51
skills needed for, 40
Data Science Association
Code of Conduct, 45
Data Science Initiative, 232–233
Data science process
steps of, 40–42
teams for, 43–45
Data scientists
demand for, 443–444
nurse data scientist, 66
projected need for, 189
role in data science process, 41–42
Data sources
administrative data, 81
for big data, 81–82
electronic health records, 81–82
genetic data, 82
research data, 82
routine data, 81–82
social media data, 82
VA consolidating, 19
VA expanding, 22
Data standards. See Standardized data
Data storage
Hadoop for, 38
increase in, big data trend, 67
names and amounts of memory, 37–38, 80
Data systems
interoperability of, used by nurse leaders, 117

Data visualization
books on, 52
components of, 42–43
conferences on, 52
curriculum opportunities and, 382
in data science process, 41, 42
most well-known, 43
Nightingale and, 184
overview of process, 43
resources for, 52–53
uses of, 280
Data wrangler, 35
Decision trees, 278
Deep Dive Program, 379
Deep learning, 279
Defense Advanced Research Projects Agency (DARPA), 194–195
Delaney, C. W., 189
Delirium
defined, 303
as difficult clinical problem, 288–289
highest risk patients, 304
identifying patients risk for, 291–292
impact of, 303
Delirium (cont.)
interdisciplinary approach to prevention, 307–310
Kaiser Permanente interprofessional
EHR data to prevent delirium
initiative, 304–310
risk score for, 308
VA’s VINCI database case study, 287–298
Democratization of production, 435–436
Department of Defense
initiative supporting data science, 194–195
Department of Energy
initiative supporting data science, 194
Diabetes, predictive model of, 164–165
Diabetes Collaborative Registry, 142
Diagnose.me, 437
Diagnosis Related Group (DRG), 99–100
Diagnostic analytics
inpatient nursing unit example, 125–126
Digital research, 183
Digitization, 36
Direct nursing costs
per patient episode in acute care case
study, 359–363
Distributed data network
characteristics of, 213
Mini-Sentinel Distributed Database, 214
used in national clinical research
network, 213
DNA repository, COMMONS, 233
Doctoral education in nursing
common core, 411–414
Doctorate in Nursing Practice (DNP),
378, 414–416
nurse data scientist, 419
overview of, 410–411
PhD program, 378–379, 417–419
practice-focused doctorate, 414–416
research-focused doctorate, 417–419
resources for implementation, 420–423
Doctorate in Nursing Practice (DNP),
378, 410–423
Documentation, nurses time spent
using EHR, 352–357
Domesday Book, 34
Dr. Foster, 434
Dumbill, E., 40

E
Eastern Maine Health System (EMHS),
352–357
Ebola, tracking, 17

Education, nursing
accreditation of graduate health
informatics programs, 391–396
AMIA informatic certification, 385–389
big data and academic nursing, 374–375
challenges ahead, 379–381
curriculum opportunities, 381–383
doctoral programs, 378–379, 407–423
faculty role changes, 446–447
future implications of big data in,
442–447
impact of data proliferation on academic
scholarship, 189–191
master’s education, 376
mobile apps for, 318–319
online education, 444–446
precision education for students,
444–446
quality outcomes and credentialing,
399–405
undergraduate education, 375–376
EHR. See Electronic health records (EHR)
Eko Core, 318
Electric stethoscope, features and uses of, 69
Electrocardiograms (ECG), computer
interpreted, 70
Electronic clinical quality
measures (eCQM), 337
Electronic health records (EHR)
challenges in accessing information for
clinicians, 305–306
characteristics/challenges of data from,
160, 161, 455
criteria for core functionality of EHR, 306
electronic integration of healthcare
systems, 256
eMERGE network, 204
extracting nursing data from, 97–98
federal mandates for use of, 13, 187
flowsheet data from, for clinical data
repositories, 140, 146–154
integrating individual patient care data in
HCS, 259–260
Kaiser Permanente delirium initiative and,
304–310
as medical data source, 81–82, 187,
302–303
mobile nursing apps for, 321, 322
nurses spending time using case study,
352–357
nursing studies using, with linked claims
data, 160–164
reimbursement reform and changes to,
15–16
secondary use of, 333
standard format for, 122
standardizing data with nursing
information models, 150–154
Trinity Health sepsis reduction
initiative, 258
undergraduate education and, 375
used in national clinical research network, 212–213
Veterans Administration’s system, 18–22
working toward sharable and
comparable data, 122
Electronic time capture, 353
Embedded variable selection, 271
eMERGE network, 204, 339
England. See National Health Services (NHS)
Environmental factors, contributing to
Type 2 diabetes, 199–200
Epidemiology, big data science and,
233–234
Epigenetics
modification and T2D, 201–202
overview of, 200–202
Estimating delivery date (EDD), mobile app
for, 320
Ethics, data ownership and privacy, 44–45
European Cross-Border Directive, 430–431
Evidence-based practice
in CPM framework, 245–247
to deal with variability in healthcare, 432
rapid dissemination of, by healthcare
systems, 257–258
Exploratory data analysis (EDA), 41
Explorys, 456
Exposome, 199
Extended clinical data (ECD), 336

Financial data
challenges of using, for nurse leaders,
117–119
electronic integration of healthcare
systems, 255
Fitness trackers, as medical data source, 82
Flow Health, 320
Flowsheet data
declared, 145
elements of, 145–146
importance of, 146, 150
as LOINC document, 146
organizational decisions on content of, 146
pressure ulcer study, 148–150
strengths and challenges of data from,
147–148
Food and Drug Administration (FDA)
Mini-Sentinel Distributed Database, 214
OpenFDA Initiative, 231–232
registry data for policy decisions, 231
Fordism, 188–189
Forrester Wave, 37
Fortini, Robert, 458
Fourth paradigm, 185, 373
Free-text comments
text mining and machine learning
application, 87–88
Fuller, Buckminster, 432
Function and Outcomes Research for
Comparative Effectiveness in Total
Joint Replacement (FORCE-TJR), 231

G
Gapminder, 43
Gartner Hype Cycle, 37
Gaugler, Joe, 75
Geisinger Health System, 159, 434
Genetic algorithms, 89–90, 279
Genetic data, 82
Genetics/genomics
clinical resources for, 203–205
contributing to Type 2 diabetes,
198–199
curriculum opportunities and, 382
epigentics, 200–201
genetic algorithms, 89–90, 279
scope of genomic nursing, 205
VA data source and, 22
Geographic information systems (GIS), 17–18
Geriatrics Cultural Navigator app, 324
GeriPsych Consultant app, 324
Global Comparators, 434
Global society, big data and, 427–439
Google, 316
MapReduce algorithm, 38
prediction of winter flu with big data, 4, 35
volume of data, 38
Google Play, 315
Governance, of PCORnet, 216
Graduate nursing education, 376–379.
   See also Doctoral education in nursing
   accreditation of health informatics
   programs, 391–396
Grady, Patricia, 7
Granger, Clive, 272
Graphical information system (GIS),
   450–451
Graphics processing units (GPU), 176
Gray, Jim, 185, 186
Gut microbiota, 201–202
   risk for T2D and, 200

H
Hadoop
   for collecting and cleaning data step, 41
   exploring unstructured data, 39
   overview of, 38
   use of, 5
Handheld ultrasound, 70
Hardware, big-data analysis considerations,
   175–176
HBase, 38
HDFS (Hadoop Distributed File System), 38
Health, social determinants of, 17–18,
   236–238
Health and Social Care Information
   Centre (HSCIC)
   big data and, 26–27, 28
Healthcare, big data in
   benefits of, 228–229
   case studies of (see Case studies)
   data sources, 81–82
   evolving landscape, 158–159
   evolving roles for nurses, 64–68, 74–76
   extending reach of, with mHealth,
   326–327
   future implications of big data in,
   454–461
   global society and big data, 427–439
   healthcare research databases, 158–159
   history of mobile in, 316–317
   local aspect of, and variability of,
   428–434
   nursing applications, 86–90, 187–188
   patients as producers in, 436–437
   population health, 14–18
   research and, 186–187
   sources of, 11–12, 14
   tipping point for, 12–14
   transformation of healthcare systems by,
   253–262
HealthCare Cost Institute (HCCI), 158, 448
Healthcare Information and Management
   Systems Society (HIMSS), 374
   CNO-CNIO Vendor Roundtable Big Data
   Principles Workgroup, 235
mHealth defined, 314
Healthcare systems (HCS)
   big data characteristics emerging from,
   256–257
   electronic integration, 255–256
   growth in, and benefits of, 253–254
   integrating individual patient care data,
   259–260
   integration to manage patient populations,
   260–261
   past limitations of, 254–255
   rapid dissemination of evidence-based care
   by, 257–258
Health Care Systems Research Network, 339
Health coach, 66
Health Data Consortium, 229–230
Health Data Initiative
   goal of, 230
   Health Data Consortium, 229–230
   healthdata.gov website, 229
Health Data Initiative Forum (Health
   Datapalooza), 230
Health disparities
   big data to eliminate, 240–241
   defined, 238
   examples of, 240
   overview of, 238–240
Health equity
   big data to build equity using symptoms
   management, 240–241
   defined, 238
Health indicators
   health indicators warehouse, 237
   need for standardized, 432–433
Health Informatics Model, 245
Health information technology (health IT)
   benefits of, 332, 341
Health Information Technology for Economic
   and Clinical Health (HITECH) Act,
   121, 187, 247, 408
Health Information Technology Standards
   Panel (HITSP), 120–121
Healthsense, 75
Hearst Health Network, 432
Heart disease, interactive atlas of heart disease and stroke, 237
Heart failure, pSCAN data research, 221–223
Heart Pro III, 318
Hive, 38
HMO Research Network, 142
Home care, mobile apps for nurses, 322–323
Home Health and Hospice system, 322
Home Health Care Classification system (HHCC), 121
Home healthcare sector, 23–25
improving wound care, 24
powered by big data, 25
technology advances and shift toward, 68–69
telemonitoring to prevent hospital readmission, 23–24
HomeRN app, 322
Hospital Lead app, 321
Hospital readmission. See Readmission
Hospital RN app, 321
Hotel chains, use of big data, 13
HuGE Navigator, 234
Human resource data
 electronic integration of healthcare systems, 255–256
 used by nurse leaders, 117–118, 119

I
i2B2, 339
iGeriatrics app, 324
Informatics. See Nursing informatics (NI)
Information models. See Nursing information models
Institute of Medicine (IOM)
criteria for core functionality of EHR, 306
learning healthcare system, 447, 451
Interactive atlas of heart disease and stroke, 237
Interdisciplinary approach, to delirium prevention, 307–310
Interdisciplinary Plans of Care (IPOCs) case study, 132–134
International Classification for Nursing Practice (ICNP®), 121
International Classification of Diseases (ICD) codes
topic modeling and, 293–294
International Society of Nurses in Genetics, 205
Internet, data created daily, 67
Internet of Things
data collection as by product, 35
purpose to datafy everyday things, 36
Interoperability, of data systems used by nursing leaders, 117–118
Interprofessional collaborative practice and education (IPE)
NCDR creation and, 105–106

J
The Joint Commission
 National Patient Safety Goal, 345

K
Kahn, J. P., 228
Kahn Academy, 445
Kaiser Permanente, 122, 434
 interprofessional EHR data to prevent delirium initiative, 304–310
Kernel-based transformation, 269–270
Key-value stores, 39
K-fold cross validation, 163
Kitchin, R., 185
Kizer, K., 400–401
Klazinga, Nick, 432
K-mean cluster analysis, 86
Knewton, 445
Knowledge acquisition, 335–336
Knowledge discovery, 5
Knowledge dissemination, 336–337
Knowledge generation, 337–340
Knowledge processing, 340–341
Knowledge value benefits of, 62
 for clinical nurse practice, 68
Kranzberg, Melvin, 185

L
Lasso methods, 163
Latent Dirichlet Allocation (LDA), 293
Learning healthcare system (LHS), 447–449, 451, 452
Leave One Out estimator, 276
Location data, 39
LOINC (Logical Observation Identifiers Names and Codes), 121
 nurse flowsheet as, 146
 Nursing Management Minimum Data Set (NMMDS), 7, 189
Loukides, M., 40
M
Machine learning, 277
bias and, 164
covariate balance using genetic algorithm, 89–90
for data cleaning and formatting, 41
free-text comment application, 87–88
K-fold cross validation, 163
nursing application, 87–89
in OptumLabs statin project, 178–179
overview of process, 85–86, 163
predictive model of diabetes, 164–165
skills for, 40
supervised, 86, 178
topic modeling, 293–294
treatment-effect estimation and, 163
unsupervised, 86, 178
Magic Quadrant, 36–37
Magnet Recognition Program, 401–402
Mahajan, Satish, 222
Management of Atrial Fibrillation app, 324
Map of Medicine, 432
MapReduce algorithm, 38, 68, 176
Markel Foundation
Connecting for Health, 16–17
network of networks approach, 16
Markov Boundaries, 274–275
Mayer-Schönberger, V., 4
Mayo Clinic, OptumLabs partnership, 159, 166, 381, 449, 453
McKeown, Thomas, 17
Meaningful Use Electronic Health Record (EHR) Incentive Program (HITSP), 121, 302
Meaningful Use mandates, 13
MedCalX, 321–322
Medicaid. See Centers for Medicare and Medicaid Services (CMS)
Medical tourism, 430–431
Medicare. See Centers for Medicare and Medicaid Services (CMS)
MedSnapID, 320
MedSnap Verify Services, 320
Memory storage, 37–38, 80
Messiness of data, 35, 39
mHealth (mobile health), 313–328
apps for nurses, 318–324
apps for patients, 324–326
defined, 314
mobile device and app history, 316–317
nurses’ roles and process, 317–318
value of, combined with analytics, 326–328
Microsoft, 4
Minard, Charles Joseph, 43
Mini-Sentinel, 143, 339
overview of, 159, 214, 449
Mio Global, 73
Missing data, 177
Mobile apps, for nursing, 313–328
for acute care, 321–322
for care coordination, 323–324
for home care, 322–323
market distribution by app categories, 315
mobile device and app history, 316–317
nurses’ roles and process, 317–318
for nursing education, 318–319
for patients, 13, 324–325
primary care apps, 319–321
Modified Early Warning (MEW) system, 116
Moore, Gordon, 67
Moore’s Law, 67
Mortality in the Crimean War, 42, 43
Multimodal analysis methods, 277
N
NANDA-Nursing Diagnoses, Definitions, and Classification, 121
Napoleon’s March on Moscow, 43
National Cardiovascular Data Registry, 142
National Center Data Repository (NCDR) characteristics of, 106–107
creation of, 103–106
ecosystem and infrastructure of, 110–113
National center for HIV/AIDS, viral hepatitis, STD, and TB prevention atlas, 237
National clinical research networks
distributed data networks and, 213–214
electronic data for, 212–213
Mini-Sentinel Distributed Database, 214
PCORnet as, 214–224
vision of, 212
National Database of Nursing Quality Indicators (NDNQI), 122
overview of, 143
National environmental public health tracking network, 238
National Health Services (NHS)
big data initiatives, 27–28
health information technology and, 26–27
overview of healthcare in England, 25–26
Spine, 26–27
National Human Genome Research Institute (NHGRI), 205
National Innovation Network (NIN) for the National Center for Interprofessional Practice and Education, 103, 104
<table>
<thead>
<tr>
<th>Term/Phrase</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Institute of Health and Care Excellence (NICE), 27</td>
<td></td>
</tr>
<tr>
<td>National Institute of Nursing Research (NINR)</td>
<td></td>
</tr>
<tr>
<td>Big Data in Symptoms Research Boot Camp, 240–241</td>
<td></td>
</tr>
<tr>
<td>big data research programs, 234</td>
<td></td>
</tr>
<tr>
<td>mission of, 142</td>
<td></td>
</tr>
<tr>
<td>Nurse Scientist Translational Research Interest Group, 142</td>
<td></td>
</tr>
<tr>
<td>nursing research defined, 185</td>
<td></td>
</tr>
<tr>
<td>promoting nursing research, 7</td>
<td></td>
</tr>
<tr>
<td>National Institutes of Health (NIH)</td>
<td></td>
</tr>
<tr>
<td>big data definition, 8</td>
<td></td>
</tr>
<tr>
<td>Big Data to Knowledge (BD2K) program, 188–189, 191, 192–193, 232–233</td>
<td></td>
</tr>
<tr>
<td>Clinical Translational Science Awards (CTSA), 140–142</td>
<td></td>
</tr>
<tr>
<td>COMMONS, 233</td>
<td></td>
</tr>
<tr>
<td>Data Science Initiative, 232–233</td>
<td></td>
</tr>
<tr>
<td>eMERGE network, 204</td>
<td></td>
</tr>
<tr>
<td>mHealth defined, 314</td>
<td></td>
</tr>
<tr>
<td>Roadmap Epigenetics Mapping Consortium, 205</td>
<td></td>
</tr>
<tr>
<td>Roadmap for Medical Research, 212, 451</td>
<td></td>
</tr>
<tr>
<td>National Partnership for Action to End Health Disparities (NPA), 239</td>
<td></td>
</tr>
<tr>
<td>National Patient-Centered Clinical Research Network (PCORnet). See PCORnet</td>
<td></td>
</tr>
<tr>
<td>National Quality Forum (NQF), 121–122, 403, 404</td>
<td></td>
</tr>
<tr>
<td>National Science Foundation (NSF)</td>
<td></td>
</tr>
<tr>
<td>Directorate for Computer &amp; Information Science &amp; Engineering (CISE), 193–194</td>
<td></td>
</tr>
<tr>
<td>information life cycle model, 190–191</td>
<td></td>
</tr>
<tr>
<td>Natural History of Disease (NHD), 166</td>
<td></td>
</tr>
<tr>
<td>Natural language processing tools, 41, 277</td>
<td></td>
</tr>
<tr>
<td>for analyzing patient reported outcome data, 162</td>
<td></td>
</tr>
<tr>
<td>improving delirium classification using, 292–295</td>
<td></td>
</tr>
<tr>
<td>NCH Healthcare System (NCH) alarm management case study, 345–351</td>
<td></td>
</tr>
<tr>
<td>Network Medicine, 279–280</td>
<td></td>
</tr>
<tr>
<td>Network of networks approach, 16</td>
<td></td>
</tr>
<tr>
<td>Network science, 279–280</td>
<td></td>
</tr>
<tr>
<td>Nightingale, Florence data visualization, 42, 43, 184</td>
<td></td>
</tr>
<tr>
<td>impact of research, 184</td>
<td></td>
</tr>
<tr>
<td>need for data, 96</td>
<td></td>
</tr>
<tr>
<td>polar-area diagrams, 332</td>
<td></td>
</tr>
<tr>
<td>Normalization, in OptumLab statin project, 178</td>
<td></td>
</tr>
<tr>
<td>NoSQL database, 4, 38, 39</td>
<td></td>
</tr>
<tr>
<td>Nurse data scientist, 66, 419</td>
<td></td>
</tr>
<tr>
<td>Nurse entrepreneur, 67</td>
<td></td>
</tr>
<tr>
<td>Nurse executives/administrators accountabilities and challenges of, 116–117</td>
<td></td>
</tr>
<tr>
<td>Chief Nurse Executive Big Data Checklist, 123–125</td>
<td></td>
</tr>
<tr>
<td>data systems interoperability and, 117</td>
<td></td>
</tr>
<tr>
<td>instilling data-driven culture through teams, 124–127</td>
<td></td>
</tr>
<tr>
<td>moving toward sharable and comparable data in nursing, 120–127</td>
<td></td>
</tr>
<tr>
<td>value of big data for, 119–120</td>
<td></td>
</tr>
<tr>
<td>Nurse practitioners, interdisciplinary approach to delirium prevention, 306–310</td>
<td></td>
</tr>
<tr>
<td>cost of care, 99–100</td>
<td></td>
</tr>
<tr>
<td>difficulty of measuring value of, 119</td>
<td></td>
</tr>
<tr>
<td>evolving roles and technology, 64–68, 74–76</td>
<td></td>
</tr>
<tr>
<td>interdisciplinary approach to delirium prevention, 306–310</td>
<td></td>
</tr>
<tr>
<td>mobile apps for, 313–328</td>
<td></td>
</tr>
<tr>
<td>quality patient outcomes and credentialing, 399–405</td>
<td></td>
</tr>
<tr>
<td>sharable and comparable data in nursing, 120–127</td>
<td></td>
</tr>
<tr>
<td>time spent using EHR, 352–357</td>
<td></td>
</tr>
<tr>
<td>value-based nursing care, 95–100</td>
<td></td>
</tr>
<tr>
<td>Nursing business intelligence and analytics (NBIA), 98</td>
<td></td>
</tr>
<tr>
<td>Nursing education. See Education, nursing</td>
<td></td>
</tr>
<tr>
<td>Nursing informatics (NI) accreditiation of graduate health informatics programs, 391–396</td>
<td></td>
</tr>
<tr>
<td>alarm management case study, 345–351</td>
<td></td>
</tr>
<tr>
<td>AMIA informatic certification, 385–389</td>
<td></td>
</tr>
<tr>
<td>benefits of, 334</td>
<td></td>
</tr>
<tr>
<td>big data and, 334–335</td>
<td></td>
</tr>
<tr>
<td>defined, 334</td>
<td></td>
</tr>
<tr>
<td>graduate specialty, 377–378</td>
<td></td>
</tr>
<tr>
<td>identifying direct nursing cost per patient episode in acute care case study, 359–363</td>
<td></td>
</tr>
<tr>
<td>knowledge acquisition, 335–336</td>
<td></td>
</tr>
<tr>
<td>knowledge dissemination and improvement, 336–337</td>
<td></td>
</tr>
<tr>
<td>knowledge generation, 337–340</td>
<td></td>
</tr>
<tr>
<td>knowledge processing, 340–341</td>
<td></td>
</tr>
<tr>
<td>nursing time in EHR case study, 352–357</td>
<td></td>
</tr>
</tbody>
</table>
Nursing informatics (NI) (cont.)
readmission prevention with
learning health system case study,
364–369
in undergraduate and graduate education,
374–379
Nursing information models
benefits of, 151
examples of, 151
need for, 151
Nursing Interventions Classification system
(NIC), 121
Nursing languages, standardized,
120–121
Nursing Management Minimum Data Set
(NMMDS), 7, 121, 189
Nursing Minimum Data Set (NMDS), 121
Nursing Outcomes Classification (NOC), 121
Nursing research. See Research, nursing

O
Oak Ridge Graph Analytics for Medical
Innovation (ORiGAMI), 194
Observational Health Data Sciences and
Informatics (OHDSI), 339
Observational Medical Outcomes Partnership
(OMOP), 143, 220, 339
Observational research study designs,
339–340
Observational studies
bias in, 168–169
compared to randomized controlled trials,
168–169
Omaha System, 121
Omics, 203
Oncology Care Model, 162
O’Neil, C., 40
Online education, 444–446
OpenFDA Initiative, 231–232
Open Government Initiative, 230
Operational data, used by nurse leaders,
117, 119
OptumLabs
overview of, 159, 449
research collaborative, 166–168,
381, 453
OptumLabs Data Warehouse (OLDW), 166
developing model to predict adverse event
prescribing statins using, 172–179
Outliers, 177
outlier detection, 280
Overfitted structured data analysis, 266

P
P30 Center of Excellence for Biobehavioral
Approaches to Symptom
Management, 234
Pachman, D. R., 164
Pain scores, in patient reported outcome
data, 161–162
Pappas, S. H., 119
Patient(s)
education for, and readmission prevention
with learning health system, 364–369
mobile apps, 324–325
patient comments and text mining/machine
learning application, 87–88
patient engagement and pSCANNER,
219–223
patient portal use, 325–326
as producers in healthcare, 436–437
Patient care
future implications of big data and,
457–458, 459
personalization of, 460
quality patient outcomes and credentialing,
399–405
shift from traditional hospital to patient
centric, 67–68
Patient-centered care
future of, 460
mobile app for, 319–320
shift from traditional hospital to patient
centric, 67–68
Patient-Centered Medical Home model,
319–320, 458
Patient-Centered Outcomes Research
Institute (PCORI), 7, 158, 339, 452
launching PCORnet, 214–215
Patient-centered SCAlable National
Network for Effectiveness
Research (pSCANNER), 218–223
Patient data
integrating individual patient care
data in HCS, 259–260
standardizing, for clinical data repositories,
144–145, 147, 150
Patient-Powered Research Networks (PPRNs)
characteristics of, 215
in PCORnet, 215–216
Patient reported outcomes (PROs)
limitations of, 161
pain scores, 161–162
pilot programs using, 162
used in national clinical research
network, 213
Patient safety
alarm management case study, 345–351
National Patient Safety Goal, 345
Payment for population outcomes, 15
PCORnet, 142–143
Common Data Model (CDM), 216–217
creation and goals of, 214–215
current state of, 217–218
early demonstration studies, 217–218
future plans for, 218
goals of, 452
governance of, 216
nursing data’s role in, 223–225
overview of, 158–159, 448
partner networks, 215
pSCANNER, 218–223
Pearl, Judea, 272
Peer-reviewed research, 190
PeriOperative Nursing Data Set (PNDS), 121
Personal health record (PHR), 437
Pharmacists, interdisciplinary approach to delirium prevention, 306–310
PhD nursing doctorate programs, 378–379, 410–423
PhenX, 204–205
Physicians and physician assistants
interdisciplinary approach to delirium prevention, 306–310
Pig, 38
Point of care (POC) lab testing, 70–71
Policy translation, accelerating research finding to, 164–165
Polymodal analysis methods, 277
Popper, Karl, 85
Population health, 14–18
compared to public health, 15
curriculum opportunities and, 382
defined, 15, 260
gerographic information systems (GIS), 17–18
integrating data by HCS for, 260–261
mobile apps for care coordination, 323–324
payment for population outcomes, 15
reimbursement reforms and move towards, 15–17
social determinants and, 17–18
populations, sampling of, 34
Practice interoperability, 247
Precision education, 444–446
Precision (personalized) medicine, 73, 460
overview of, 205
Precision Medicine Initiative
Cohort Program, 205
Prediction, machine learning for, 163
Predictive analytics
curse of dimensionality, 266
inpatient nursing unit example, 126
overfitted models, 266
Support Vector Machines (SVMs), 267–269
Predictive modeling
advances in, 265–269
defined, 85
overview of, 85
Premier Health Alliance, 434
Prescriptive analytics, inpatient nursing unit example, 126–127
President’s Council of Advisors on Science and Technology (PCAST), 452
Pressure ulcer
CPF clinical practice guidelines risk for example, 249–250
flow sheet data from EHR study, 148–150
Trinity Health risk factor case study, 134
Pressure Ulcer Survey, 143
Prevent Group B Strep, 320
Prevention of Falls Guidelines app, 324
Primary care, mobile apps for nurses, 319–321
Privacy issues
data ownership and, 44–45
HCS data integration and importance of protecting, 262
pSCANNER distributed query mechanism and, 219–220
Proof of value, 44
Propeller app, 324
Propensity score matching, 89–90
pSCANNER, 218–223
heart failure readmission research, 221–223
Public health, compared to population health, 15
Public Health England (PHE), 27, 28
Python, to optimize model, 41
Q
Quality metrics for reimbursement
CMS reforms based on, 15–16
Quantified self-movement, 71–72
collecting data for, 259–260
R
R, to optimize model, 41
Random forests, 278
Randomized controlled trials (RCTs), 339–340
  compared to observational studies, 168–169
Readmission
  heart failure readmission research, 221–223
  prevention with learning health system, 364–369
  telemonitoring to prevent, 23–24
Redshift, 5
Reed, L., 119, 123
Regional Health Equity Councils (RHECs), 239
Regression, shrinkage regression methods, 271
Regularization, SVMs, 267, 269
Remote medical sensing, 260
Repeated nested n-fold cross validation (RNFCV), 275–277
Research, nursing
  big data in healthcare and, 185–187
  data proliferation’s impact on scholarship, 189–191
  data science as fourth paradigm of, 185, 186
  defined, 185, 190
  digital research, 183
  doctoral education in nursing and, 417–419
  future implications of big data in, 450–454
  impact of data proliferation on, 189–191
  importance of knowledge generation, 337–339
  information life cycle model, 190–191
  initiatives supporting data science and, 191–195
  multi-institutional research networks, 448–449
  national clinical research network, 212
  Nightingale’s contribution, 184
  observational research study designs, 339–340
  peer-reviewed research, 190
  randomized controlled trials (RCTs), 339–340
  research data as big data source, 82
  schema-on-read, 187
  schema-on-write, 186–187
  sharing data between CDRs for, 339
ResearchGate, 42
Resources
  for big data, 47–49
  for data science, 50–51
  for data visualization, 52–53
Retailers, using big data and analytics, 12–13
Risk model, identifying patients at risk for delirium, 291–292
Roadmap Epigenetics Mapping Consortium, 205
Roadmap for Medical Research, 212
Room and board charges, 96
Rose Diagram, 184
Rosling, H., 43
Routine data, as big data source, 81–82

S
Safety Thermometer: Patient Harms and Harm Free Care, 28
Sampling data
  big data and shift away from, 34–35
  historical perspective on, 34–35
Schema-on-read, 187
Schema-on-write, 186–187
Scholarship. See also Education, nursing
  impact of data proliferation on, 189–191
Schutt, R., 40
Sensor technology
  activity monitors, 72–73
  quantified self-movement, 71
Sepsis/septic shock
  Trinity Health initiative to reduce, 257–258
Shared decision-making,
  characteristics of, 260
Shrinkage regression methods, 271
Sleep monitors, 72
SNOMED CT, 121
Social determinants of health, 17–18
  data sources for, 237–238
  overview of, 236
Social media data, as big data source, 82
Software
  big-data analysis considerations, 176
  NIH development awards, 193
  Sortable risk factors and health indicators, 238
  Spark, 38
  Sparse Candidate Algorithm, 273
  Spine, NHS, 26–27
  Stakeholder engagement, pSCANNER and, 219–221
Standardized clinical terminology
  coding to, 247
  WDL assessments mapped to, 248
Standardized data
  for clinical data repositories, 144–145, 147, 150
  national collaborative to standardize nursing data, 152–154
need for, to optimize data
within HCS, 261
nursing information models for, 151–154
Standards Coordinating Center (SCC), 188–189
Statins, developing predict model for adverse events from, 172–179
Stelland, Chris, 4
Stethoscopes
acoustic, 69
electric, 69
Stewardship programs
delirium patients and building, 295–297
overview, 290
Stonecrusys, 73
Stroke, interactive atlas of heart disease and, 237
Structured data, amount of, 80
Structured query language (SQL) tool using in OptumLabs statin project, 175
Super utilizer patients, 280
Supervised machine learning, 86, 178
Support Vector Machines (SVMs), 86
kernel-based transformation, 269–270
regularization, 267, 269
Symptom checkers, 73–74, 76
Symptom management, to build health equity, 240–241

T
Teams
for data science process, 43–45, 165–166
emerging roles, 167, 168
instilling data-driven culture in nursing through, 124–127
in OptumLabs statin project, 173–174
Technology
for big data evaluation, 36–37
enabling technologies, 69–74
evolving nursing roles and, 64–68, 74–76
shift from traditional hospital to outpatient/home care, 68–69
Telehealth facilitator, 66
Telemonitoring, to prevent hospital readmission, 23–24
Text mining
defined, 84
free-text comment application, 87–88
healthcare application, 85
nursing application, 87–88
overview of process, 84
Time, nurses time spent using EHR, 352–357
Topel, Eric, 70
Topic modeling
defined, 293
used in classifying delirium, 293–294
Total Joint Replacement registry, 231
Training set, in machine learning, 85, 86, 163
Treatment-effect estimation, machine learning and, 163
Trinity Health
data warehouse, 131–132
general to specific and failure to diagnose case study, 136
Interdisciplinary Plans of Care (IPOCs) case study, 132–134
overview of, 130–132
pressure ulcer case study, 134
rapid dissemination of evidence-based care by, 257–258
venous thromboembolus (VTE) advisory case study, 135
Triple Aim
EHR data and, 303
generating data for analyzing interprofessional teams to improve outcomes, 103–113
goals of, 333
mobile apps and, 323
Trinity Health strategy, 131, 133, 136
Twitter, volume of data, 38
Type 2 diabetes (T2D)
environmental factors contributing to, 199–200
epigenetic modification, 200–202
genetic factors contributing to, 198–199
as significant health problem, 197–198

U
Ultrasound, handheld, 70
Undergraduate nursing education, 375–376
United Kingdom. See National Health Services (NHS)
United Network for Organ Sharing/Organ Procurement and Transplant Network (UNOS/OPTN), 142
University of Minnesota
CDR of nursing information, 144–145
Nursing Knowledge: Big Data Science initiative, 152, 235
OptumLabs partnership, 166
pressure ulcer flowsheet study, 148–150
Unstructured data
advances in, 277
amount of, 39, 80
Unsupervised machine learning, 86, 178
U.S Department of Health and Human Services (HHS)
  Health Data Initiative, 229–231
US Department of Veterans Affairs, 122

V
VA. See Veterans Health Administration (VHA)

Validation set, in machine learning, 85, 163

Value, of big data, 80
Value-based nursing care model development case study, 95–100
Value-based purchasing, moving toward, 15–16

Variability, of big data, 80
Varian, Hal, 45

Variety, in big data
  overview of, 39, 80
  produced in healthcare systems, 256

Velocity, of big data
  overview of, 39, 80
  produced in healthcare systems, 256

Venous thromboembolus (VTE) advisory case study, 135

Veracity, of big data
  produced in healthcare systems, 257

Veterans Health Administration (VHA), 18–23
  challenges in advanced analytics, 21
  Clinical Assessment, Reporting, and Tracking (CART) system, 22
  consolidating data sources, 19
  Corporate Data Warehouse (CDW), 19–21
  data governance, access and quality, 19–20
  delirium and VA's VINCI database, 287–298
  expanding data sources, 22
  genomics and, 22
  historical perspective on EHR platform, 18
  transition to advanced analytics, 19–21

Veterans Information Systems Technology Architecture (VistA), 18–22

Videos
  on big data, 48–49
  on data science, 51
  on data visualization, 53
  video files as data source, 82

VINCI (Veterans Informatics and Computing Infrastructure)
  delirium case study, 287–298
  overview of, 290–291

ViPAR network, 339

Visual cues, to encode data, 42

Visualization methods, uses of, 280

Volume, in big data, 332–333
  examples of, 38
  overview of, 37–38, 80
  produced in healthcare systems, 256
  sources of, 33–34

Vulnerable populations footprint tool, 238

W
Watson computer, 76
Watson Health, 456

Wearable fitness devices, 324–325

Websites
  on big data, 49
  on data science, 51
  on data visualization, 53

Westra, B. L., 189

Within defined limits (WDL) assessments
  mapped to standardized terminology, 248

World Health Organization (WHO)
  mHealth defined, 314
  social determinants of health, 236

Wound care
  improving outcome for home healthcare patients, 24

Wrappers, 270–271

X
XML parsers, 41

Y
Yahoo!, 68
YARN (Yet Another Resource Negotiator), 38

Youtube, volume of data, 38

Z
Zang, Qing, 293
Zerhouni, Elias, 141, 451–452
Zika, tracking, 17
ZocDoc, 437
Zynx Health, 432