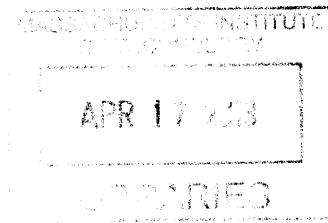


Data Driven Health System

ARCHIVES

by

Melissa Beth Rosen Ceruolo



B.S. Mechanical Engineering, 2001
Carnegie Mellon University

Submitted to the System Design and Management Program
In Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management

at the
Massachusetts Institute of Technology

February 2013

© 2013 Melissa Rosen Ceruolo
All rights reserved

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author _____

Melissa Rosen Ceruolo
System Design and Management Program
January 18, 2013

Certified by _____

Luis Perez-Breva, PhD
Thesis Supervisor
Research Scientist and Lecturer, School of Engineering

Accepted by _____

Patrick Hale
Director
System Design and Management Program

Data Driven Health System

by

Melissa Beth Rosen Ceruolo

B.S. Mechanical Engineering, 2001
Carnegie Mellon University

Submitted to the System Design and Management Program on
January 18, 2013

In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering and Management

ABSTRACT

Effective use of data is believed to be the key to address systemic inefficiencies in health innovation and delivery, and to significantly enhance value creation for patients and all stakeholders. However, there is no definition for health data. Rather, data in health is an assortment of observations and reports varying from science to clinical notes and reimbursement claims that emerge from practice rather than design. What is health data?

In this thesis we try to answer that question by looking at the system of health almost exclusively as a system that generates, transforms, and interprets data. We overview the different meanings data has throughout the health system, we analyze systematically the inefficiencies and trends as they emerge from data, and propose a new architecture for the system of health in which data is not present by accident. The result of this thesis is a new architecture for the system of health that is consistent with its present state but also consistent with a future learning system and a redefinition of value in health care that is patient and information centric.

Thesis Supervisor: Luis Perez-Breva, PhD
Title: Research Scientist and Lecturer, School of Engineering

ACKNOWLEDGEMENTS

With sincere gratitude, I thank Patrick Hale and MIT's System Design and Management program for giving me the opportunity to engage in research at MIT and for giving me the tools I need to be successful in the next phases of my career as an engineering leader.

This research is sponsored by MIT Center for Biomedical Innovation. I thank the Executive Director, Gigi Hirsch, for her support and the opportunity to be a part of the New Drug Development Paradigm initiative (NEWDIGS).

I gratefully thank my advisor, Luis Perez-Breva, for

- ❖ his commitment to our work and dedication to academia and innovation
- ❖ the freedom to explore with his patient guidance
- ❖ careful critique and strive toward perfection
- ❖ instilling in me ownership and precision
- ❖ his unique perspective and sense of humor

I would like to acknowledge my managers Michel Bruehwiler, Christian Péclat, and Walter Huber at Helbling Precision Engineering, Inc. for their support of my education and for nurturing my early knowledge of the health system in the medical device industry.

Lastly, I thank my husband, Christopher Ceruolo, for his strength, perseverance, and passion - constant inspirations leading me toward my true potential.

I dedicate this thesis to my parents who have instilled in me the value of hard work and laid the foundation for my success with their timeless support.

TABLE OF CONTENTS

ABSTRACT	3
ACKNOWLEDGEMENTS	4
TABLE OF CONTENTS.....	5
TABLE OF FIGURES	8
TABLE OF TABLES.....	10
TABLE OF ACRONYMS.....	11
CHAPTER 1: INTRODUCTION	12
1.1 Background on the Health System.....	14
1.1.1 Stakeholder Definition	14
1.1.2 Stakeholder Network Flow	16
1.1.3 System Problem Statement.....	17
1.2 Methodology.....	18
1.3 Thesis Outline	19
CHAPTER 2: HEALTH DATA LANDSCAPE	20
2.1 Landscape overview	21
2.2 Provider Data.....	22
2.2.1 Administrative Data	24
2.2.2 Clinical Data.....	25
2.3 Patient Data	29
2.3.1 Personal Health Record.....	30
2.3.2 Physiological Measurements	31
2.3.3 Genetic Data.....	34
2.4 Payor Data.....	35
2.5 Manufacturer Data	40
2.5.1 Drug Discovery.....	42
2.5.2 Drug Development.....	42
2.5.3 Clinical Trials	45
2.6 Regulator Data.....	47
2.6.1 New Drug Application (NDA).....	47
2.6.2 Label	49
2.6.3 Adverse Events.....	51

2.6.4	FDA Data Resources.....	51
2.7	Distributor Data	53
2.8	Data Landscape Summary	57
CHAPTER 3: ANALYSIS OF HEALTH DATA		58
3.1	Cross-Stakeholder Communication.....	59
3.2	Dependencies.....	61
3.3	Attributes	66
3.4	Value Assessment	68
CHAPTER 4: PATIENT INTERACTION DATA FLOW		72
4.1	Data Flow Diagram.....	72
4.1.1	Input Data.....	74
4.1.2	Output Data.....	74
4.1.3	Communication Between Provider and Payor	75
4.1.4	Communication Between Manufacturer and Public	76
4.1.5	Communication Between Patient and Provider	76
4.1.6	Communication Between Providers.....	77
CHAPTER 5: DATA DRIVEN TRENDS.....		79
5.1	System Input (A)	82
5.1.1	Wearable Sensors	82
5.1.2	Direct to Consumer Genetic Testing	84
5.1.3	Online Resources.....	84
5.2	Clinical Knowledge Warehouse (B)	85
5.3	Clinical Decision Support (C)	87
5.4	Claim Management Analytics (D)	88
5.5	Genetic Testing (E)	88
5.6	Open Innovation Initiatives (F)	90
5.7	Product Licensing (G)	91
5.8	Distribution Management and Analytics (H)	92
5.9	Localized Feedback (I)	93
5.10	Summary of Trends.....	94
CHAPTER 6: CONCEPT OF A LEARNING HEALTH SYSTEM		95
6.1	Functional Architecture	95
6.1.1	Data Categories in Figure 40	96
6.1.2	Process Categories in Figure 40.....	97
6.2	System Concept Emerging From Figure 40	98
6.2.1	A Learning Health System	99
6.3	System Problem Statement, Revisited.....	101
CHAPTER 7: CONCLUSION		102

7.1	Findings Supporting New Vision	103
7.2	Future Work	104
REFERENCES		106
APPENDIX A: REFERENCED WEBSITES		111
APPENDIX B: STAKEHOLDERS' NEEDS.....		112
APPENDIX C: PARTITIONED DSM, SECTION 3.2		113
APPENDIX D: TAMIFLU CASE STUDY		114

TABLE OF FIGURES

FIGURE 1: STAKEHOLDER NETWORK FLOW.....	16
FIGURE 2: WHAT IS DATA?.....	20
FIGURE 3: VISUALIZATION OF HEALTH DATA CATEGORIES.....	21
FIGURE 4: EHR SCREENSHOT FROM NEXGEN HEALTHCARE.....	23
FIGURE 5: VISUALIZATION OF PROVIDER DATA CATEGORIES.....	24
FIGURE 6: CATEGORIZATION OF PROVIDER ADMINISTRATIVE DATA.....	25
FIGURE 7: CATEGORIZATION OF PROVIDER CLINICAL DATA.....	26
FIGURE 8: VISUAL ILLUSTRATION OF PROVIDER CLINICAL SUBCATEGORY, PATIENT STATUS.....	27
FIGURE 9: ICD-10 SCREENSHOT OF LUNG CANCER DISEASE CLASSIFICATION	28
FIGURE 10: ZWEENAHEALTH PHR FEATURES	31
FIGURE 11: CORVENTIS' MOBILE PATIENT MANAGEMENT SYSTEM FOR HEART FAILURE PATIENTS.....	34
FIGURE 12: CMS-1500 SMARTFORM, HEALTH INSURANCE CLAIM FORM.....	36
FIGURE 13: CATEGORIZATION OF PAYOR DATA	38
FIGURE 14: CATEGORIES OF DATA IN THE DRUG DEVELOPMENT AND APPROVAL PROCESS.....	41
FIGURE 15: DATA INCLUDED IN IND	42
FIGURE 16: SECTIONS FROM CASE REPORT FORM	44
FIGURE 17: CATEGORIZATION OF CASE REPORT FORM (CRF) DATA	44
FIGURE 18: DATA INCLUDED IN CLINICAL STUDY REPORT.....	45
FIGURE 19: SIX SECTIONS OF THE NEW DRUG APPLICATION (NDA).....	48
FIGURE 20: ADDITIONAL DATA INCLUDED IN A NEW DRUG APPLICATION (NDA)	48
FIGURE 21: SCREENSHOT OF AVASTIN LABEL FROM DAILYMED.....	50
FIGURE 22: DATA CONTENTS IN A DRUG LABEL.....	50
FIGURE 23: DATA ELEMENTS FOR DRUG AVAILABLE FROM FDA	52
FIGURE 24: LIST OF RESOURCES AVAILABLE ON FDA WEBSITE.....	52
FIGURE 25: PBM NETWORK.....	53
FIGURE 26: CATEGORIZATION OF DISTRIBUTOR DATA	54
FIGURE 27: ADJACENCY MATRIX FOR CROSS-STAKEHOLDER COMMUNICATION	59
FIGURE 28: NETWORK DIAGRAM OF DATA FLOW BETWEEN STAKEHOLDERS	60
FIGURE 29: PARTITIONED DSM OF FIRST ORDER STAKEHOLDER COMMUNICATION DATA.....	62
FIGURE 30: 60 SELECTED SUBCATEGORIES OF DATA FROM CHAPTER 2.....	63
FIGURE 31: NON-PARTITIONED DSM OF 60 DATA SUBCATEGORIES.....	64
FIGURE 32: PARTITIONED DSM OF 60 DATA SUBCATEGORIES, DIRECT DEPENDENCIES	65
FIGURE 33: DATA ATTRIBUTE MAP	67
FIGURE 34: PATIENT INTERACTION DATA FLOW DIAGRAM	73

FIGURE 35: DATA DRIVEN TREND CATEGORIES A-I MAPPED ONTO INTERACTION DIAGRAM	80
FIGURE 36: VENDOR LANDSCAPE FOR WEARABLE TECHNOLOGY.....	83
FIGURE 37: STAGES OF MEANINGFUL USE.....	86
FIGURE 38: MYRIAD CANCER PRODUCTS OFFERINGS.....	89
FIGURE 39: TRANSPARENCY LIFE SCIENCES PROCESS LEVERAGE POINTS.....	91
FIGURE 40: FUNCTIONAL ARCHITECTURE OF THE INTERACTION DIAGRAM	96
FIGURE 41: SYNTHESIZED SYSTEM IN OBJECT-PROCESS NOTATION.....	98
FIGURE 42: LEARNING HEALTH SYSTEM CONCEPT	99

TABLE OF TABLES

TABLE 1: DESCRIPTION OF PHYSIOLOGICAL MEASUREMENTS.....	32
TABLE 2: PATIENT MONITORING DEVICES.....	33
TABLE 3: CLINICAL TRIAL PHASES	45
TABLE 4: ASSESSMENT OF METHODS AND TOOLS USED TO ANALYZE DATA.....	58
TABLE 5: SAMPLE ATTRIBUTE CATEGORIES	67
TABLE 6: DATA VALUE ANALYSIS.....	69
TABLE 7: RESULTS OF DATA VALUE ANALYSIS	70
TABLE 8: QUALITATIVE ASSESSMENT OF FUNCTION FOR SELECTED DATA CATEGORIES.....	70
TABLE 9: DESCRIPTION AND EXAMPLES OF DATA DRIVEN INDUSTRY TRENDS.....	82

TABLE OF ACRONYMS

ACO	Accountable Care Organization
AMA	American Medical Association
APCD	All-Payor Claims Database
BMJ	British Medical Journal
CDC	Centers for Disease Control and Prevention
CDDS	Clinical Decision Support System
CMS	Centers for Medicare and Medicaid Services
CRF	Case Report Form
CSR	Clinical Study Report
DFD	Data Flow Diagramming
DHHS	Department of Health and Human Services
DOD	Department of Defense
DSM	Design Structure Matrix
EHR	Electronic Health Record
EMA	European Medicines Agency
EMR	Electronic Medical Record
ETASU	Elements to Assure Safe Use
FAERS	FDA Adverse Event Reporting System
FDA	U.S. Food and Drug Administration
GCP	Good Clinical Practice
GSK	GlaxoSmithKline
HGP	Human Genome Project
HHS	Health and Human Services
HIPAA	Health Insurance Portability and Accountability Act
HITECH	Health Information Technology for Economic and Clinical Health
HRQoL	Health Related Quality of Life
ICD	International Classification of Diseases
IND	Investigational New Drug
NDA	New Drug Application
NDC	National Drug Code
NIH	National Institutes of Health
NLM	National Library of Medicine
PCAST	President's Council of Advisors on Science and Technology (US)
PCMH	Patient Centered Medical Home
PBM	Pharmacy Benefits Management
PHR	Personal Health Record
PRO	Patient Reported Outcomes
RDE	Remote Data Entry
REMS	Risk Evaluation and Mitigation Strategies
RSM	Remote Site Monitoring
SPS	System Problem Statement
VA	Department of Veteran's Affairs
VUE	Visual Understanding Environment
WHO	World Health Organization

CHAPTER 1: INTRODUCTION

The health industry has evolved from the scientific practice of medicine to a system that spans across many disciplines, public and private initiatives, and vastly different scales. This interconnectedness, in addition to the increase of the aging population and the multiplication of disease categories, has led to a system that is continuously forced to reconcile managing complexity with the opportunities presented by ad hoc and local advancement. The Institute of Medicine estimates that the United States spends \$2.6 trillion on health, with \$750 billion wasted on procedural inefficiencies, unnecessary services, and inflated costs; this amounts to a 30 percent loss, believed to be a symptom of a suboptimal system that fails to meet the needs of its stakeholders (Institute of Medicine, 2012).

Data in the health system is becoming more pervasive and its definition expands as new devices and technologies are created. However, the system has yet to evolve into an information system and data is managed within established stakeholder silos and is generally inaccessible from the outside, with no standard grid for data to flow or incentive to share it. This data is an assortment of observations and reports varying from science to clinical notes and reimbursement claims that emerge from practice rather than design. We believe this data holds more information than is currently used and has potential to significantly improve the state of the health system.

The objective of this work is to arrive at a definition of health data and use that definition to understand the system of health for its ability to generate, transform, and interpret data. We view this as a prerequisite to understand the system of health as a learning system. With that in mind, we overview the different meanings data has throughout the health system, we analyze the inefficiencies and trends as they emerge from data, and propose a new architecture for the system of health in which data can no longer be thought of as accidental.

The result of this thesis is a new architecture for the system of health that is consistent with its present state but also consonant with the aspirations of a learning health system. Our interpretation of learning, borrowing from the artificial intelligence literature, is the process by which models and prior beliefs are updated given new data. We try to reconcile this definition with the aspirations of learning in health.

We use the insight gained from analyzing the inefficiencies, in combination with the principles of system design, to envision a health system that is patient and information centric.

Through the methodical evaluation of data types, sources, and interfaces, we introduce the conundrums in the current system. The two data problems outlined below summarize the questions and impediments to learning that emerge from our overview of the system through the lens of data.

Data problem #1: Data input to the clinical decisions that drive the entire system is unnecessarily limited.

- 1.1 Data sourced outside the clinical environment, beyond the purview of the provider system, does not have a defined use in the current health system. This includes patient sourced data from physiological monitors that do not yet have a standard entry into the care system.
- 1.2 Clinical decisions are made based on the data available. The selection of data collected to make these decisions is largely driven by consensus expectation and provider specific intuition. This suggests the introduction of unintended bias and inconsistencies that easily propagate to every decision.
- 1.3 There is no standard mechanism for the patient to provide feedback to the product manufacturer or to the care system after an intervention.

Data problem #2: There is no mechanism for data to flow through the system. Rather the flow of data is easily interrupted, which results in suboptimal decisions and missed opportunities for knowledge sharing.

- 2.1 There is no standard way for specific patient information to transfer between providers; either through referrals or during care transitions.

- The result is a patient history at the point of care that is guaranteed to be incomplete.
- 2.2 Payors make decisions to reimburse for care based on transactional data from the provider. The content of this data has limited clinical information and is not a complete representation of the patient's health status.
 - 2.3 Medical product data generated during the development process is not released in full after the product has been approved for commercialization. The system relies on the summary data that is hard to interpret in the context of a single patient, such as labels, medication guides, and publications.

1.1 Background on the Health System

In this section we summarize the critical aspects of the health system that are needed to understand the analysis done in this thesis, namely its six main stakeholders – patients, providers, payors, manufacturers, regulators, and distributors – and the way in which value is currently delivered. We use Crawley's system architecture methodology to guide our definition of the system and to introduce the system problem statement that can then be used to analyze alignment between stakeholders.

1.1.1 Stakeholder Definition

In principle, the patient is the direct beneficiary of the health system. The current system considers individuals as patients upon their entry to the care system. The only individuals today that are always considered patients are those with chronic conditions. For our analysis of the future scenario we will eventually broaden this definition to consider an individual who aspires to manage his health – either from within or outside the care system.

Manufacturers of medical products, including drug therapies and devices, are solution providers to the system. Manufacturers benefit from the revenue generated from the sale of their products. The global pharmaceutical market is significant with revenue

reaching \$880 billion (Hirschler, 2012). The scope of the manufacturer stakeholder in this thesis is primarily pharmaceutical companies, or drug sponsors.

Manufacturers and patients have remarkably different objectives. Manufacturers benefit from volume and repeated use of their products, while patients seek personalized treatments and unique cures. These needs are indicative of the nature of tensions in the system.

The other stakeholders are providers, payors, distributors, and regulators. Providers are individuals or institutions that provide health care services and payors are the funding arm for these services. There are nearly 200 major private payors in the United States, the leaders including Aetna, Humana, Kaiser, Unitedhealth, and Wellpoint, with collective revenues reported at \$884 billion (HCAN, 2012). Payors coordinate directly with distributors and providers to reimburse for the cost of products and services.

Distributors close the loop back to the patient by managing the medical product stream from the manufacturers. The scope of the distributors used in this thesis is primarily drug distribution from pharmaceutical companies to patients. This type of distributor is also referred to as a pharmacy benefits manager, or PBM. The two largest PBMs, CVS Caremark and Express Scripts/Medco, process the prescriptions of an estimated 200 million people in the United States, with annual revenues of each of these companies exceeding \$15 billion (NCPA, 2011).

Regulators assume the role of watchdog of the system to ensure product safety and ultimately resolve the aforementioned tensions. Regulators coordinate closely with the manufacturers and oversee the commercialization of products. The scope of the regulator stakeholder used in this thesis is primarily the United States FDA's regulation of drug products.

Combined, these six stakeholders account for economic activity north of \$2 trillion.

1.1.2 Stakeholder Network Flow

Figure 1 summarizes the nature of the flows between the stakeholders in five ways: data, services, products, information, and money. We develop these flows from interviews with stakeholders and review of the literature. The network flow is a convenient way to synthesize the vastness of the system and we use the outcome to identify the needs of each stakeholder (Appendix B). Following Crawley's framework, we interpret these needs into goals to derive the system problem statement (section 1.1.3).

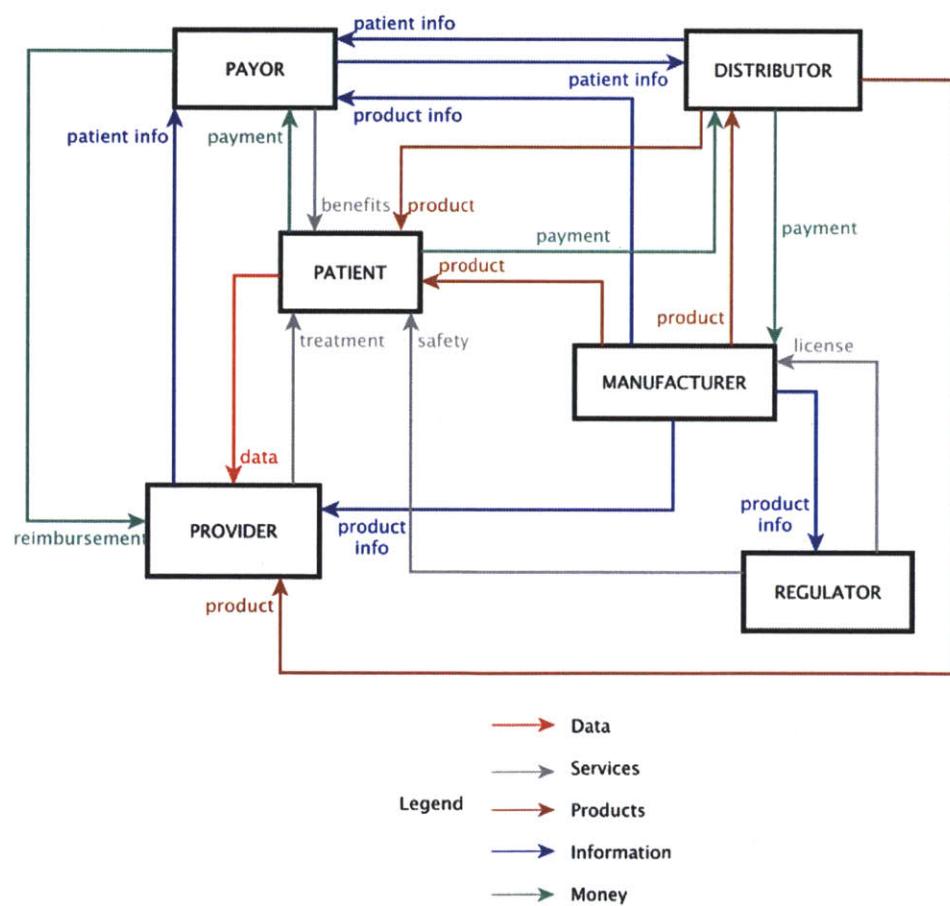


Figure 1: Stakeholder Network Flow

A significant takeaway from Figure 1 is that individual health data, that is, observables about a patient's health, are only exchanges at the patient-provider interface. All other exchanges between stakeholders are on products, payment, services, and summary information. The patient is the source of the data and is the receiver of care services, however, information does not flow directly to the patient. Rather, it is managed by the other stakeholders. As a matter of fact patients may find it cumbersome to gain access to their own complete health record. This contradicts the concept of a patient-centric system that has been previously defined by the Institute of Medicine (Institute of Medicine, 2001). The network flow diagram shows that the patient supplies the system with unlimited amounts of narrowly defined data and is not the primary receiver of information.

1.1.3 System Problem Statement

According to Crawley's framework, the system problem statement takes the form of a "to, by, using" sentence. Based on our stakeholder analysis, we derive the following system problem statement for the current health system:

To treat symptoms and diseases,
By providing care services,
Using available medical information and products.

A recurring observation, pervasive in the literature, is the divergence of the current health system with the intended system. The inefficiencies presented earlier in this chapter indicate that the runaway costs and stagnant number of new drugs approved are symptoms of this divergence. Our analysis of the current system aligns with this shift, as we have indicated with the limited role of the patient. We believe that the root of the problem may be found by examining how the system utilizes the information it generates. Hence the objective of this thesis to focus on the data each

stakeholder needs and how they capture, use, and report the information as they fulfill these needs.

1.2 Methodology

We collected information about data in the health system through stakeholder interviews, from industry conferences and academic seminars, and through review of the most recent literature. Our analysis of current stakeholders' perception of data emerges from the reports published by key opinion leaders in health, namely the United States Food and Drug Administration (FDA), European Medicines Agency, United States President's Council of Advisors on Science and Technology (PCAST) reports, and several reports by the Institute of Medicine. We complemented this information with interviews to founders of companies working on implementing new health innovations.

Throughout the thesis we use several system and visualization tools. We use Crawley's framework of system design to define the stakeholder network, identify stakeholder needs, and derive a system problem statement. We continue to use the stakeholder method to organize the data, as a first level analysis. After compiling the data, we apply adjacency and design structure matrices to assess the cross-stakeholder communication and to define data dependencies. We use process modeling methods, namely data flow diagramming, to represent the flow of data as the patient interactions with the health system. We analyze our data flow diagram from various perspectives, and we arrive at an architecture for a system design concept using topology-preserving transformations of the flow diagram. We compare this method to object-process modeling to verify the system characteristics, identifying form and function. We compare our system representation with examples of data driven trends to ground our process and provide relevant insight to the analysis.

1.3 Thesis Outline

The thesis is organized as follows: In chapter 2, we describe the kinds of data that are associated with each stakeholder, creating the landscape of data referenced throughout the thesis. We explore the sources of data, how it is used, and where it resides in the system. The data analysis continues in chapter 3 with visualizations of data flow, defining the communication paths and dependencies. Taking a perspective beyond the stakeholders, we further compare the data against a set of attributes and assess value. In chapter 4 we merge the data analysis methods to derive the patient interaction data flow diagram to holistically view and analyze data as a whole. We discuss the impact when stakeholders have partial information through structured review of the interaction diagram. We continue referencing the interaction diagram in chapter 5, while we discuss the data driven trends and current state of the industry. Chapter 6 we revisit the system structure and through an architectural transformation derive the system vision for an improved health system that is based on knowledge generation and learning.

CHAPTER 2: HEALTH DATA LANDSCAPE

There is no single interpretation of data in the current health system. Figure 2 shows a visualization of several kinds of data to allude to the complexity of the system and the data it manages. In this chapter we define the meaning of health data. Results of this research are a better understanding of what data means in the health system and more insight into the role of each stakeholder.

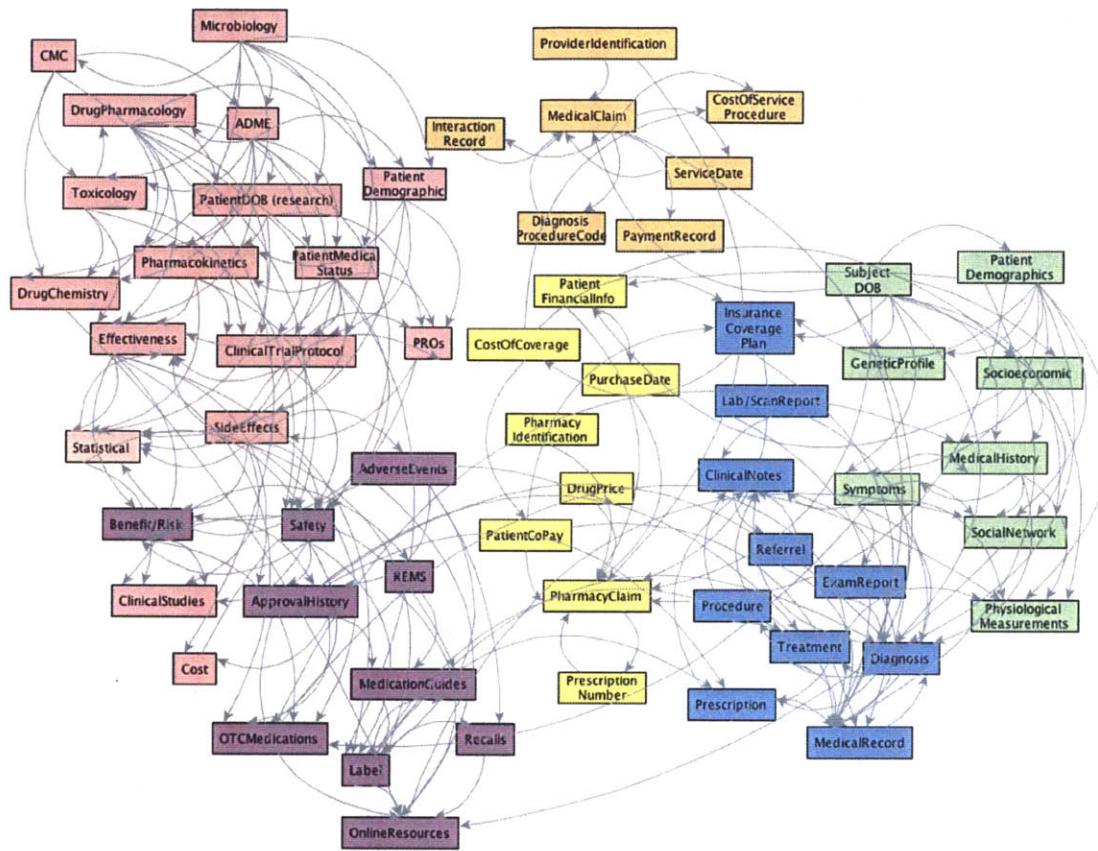


Figure 2: What is data?

This figure shows the landscape of kinds of data present in the health system. Colors reflect the stakeholder managing the data and arrows correspond with dependencies. Boxes indicate kinds of data of significant indexes used to sort data. This figure tries to capture the complexity of sources and meanings of data in the system of health absent a structured architecture for data in the system. The diagram is sorted by stakeholder; that sorting proves to be insufficient to understand the value of data, the inefficiencies that emerge from the lack of a data architecture, and the meaning one might want to ascribe to learning in health.

2.1 Landscape overview

We follow the stakeholder convention defined in chapter 1. Namely, there are six stakeholders - provider, patient, payor, manufacturer, regulator, and distributor. The data landscape assessment starts with a structured review of the kinds of data that each stakeholder manages. The data ranges from clinical measurements that inform patient health to administrative transaction records that reimburse care services. The breadth of kinds of data is as varied as the functions and needs of each stakeholder. We review various types of documents and enumerate the data housed within, resulting in nearly 500 distinct kinds of data. We sort the data into categories and subcategories to ease our analysis.

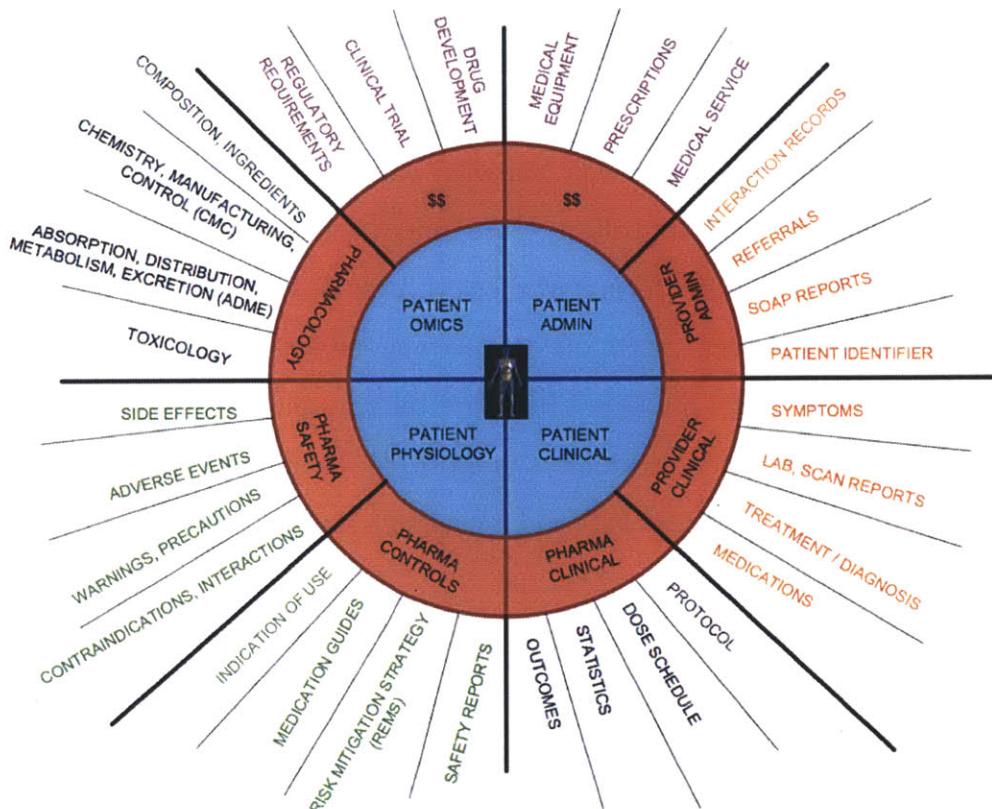


Figure 3: Visualization of Health Data Categories

In Figure 3 we outline the data categories with the patient at the center because our research suggests that no other element is shared across all the different kinds of data. We revisit this point when we review the flow of information creation in chapter 4.

Other data elements build off of the patient and are transformed, reformatted, and modified as they flow through the system among the stakeholders. The figure shows the extent of sources used to collect the data pieces and the organizational approach taken to envisage an understandable, yet comprehensive landscape.

This chapter looks at the landscape of data, without considering the level of adoption for each kind of data. The purpose is to give a holistic overview and we describe how the data is being used in subsequent chapter. Perceiving data from its stakeholder origin is convenient to start the assessment, but later we challenge the view and suggest alternate categories of data.

2.2 Provider Data

Providers are in direct contact with patients, with whom they often have multiple encounters a year. Providers collect, store, and maintain patients' detailed medical and interaction records. The format of these records is changing from paper-based to digital, with the worldwide adoption of electronic health records (EHRs). EHRs will provide electronic storage of patient data and the content is indicative of the type of data that is managed by the provider. Therefore, review of EHR data fields and format sufficiently defines the provider data and is used as a starting point for this analysis.

We compared ten major EHR vendors including AthenaHealth, Practice Fusion, Care360 Quest, Epic System, eClinicalworks, and NexGen Healthcare. Although the products themselves vary among vendors, the content of data within the EHR is comparable. Figure 4 shows an EHR screenshot from NexGen Healthcare, displaying the various fields of data.

Figure 4: EHR screenshot from NexGen Healthcare

("EHR Market," 2009)

There are approximately 60 kinds of data that can be inputted into this single EHR screen, collected during the patient interaction with the provider. Some are administrative data such as name of patient, time, date; and some are clinical data such as temperature, blood pressure, and allergies. Although the EHR has the data content we seek for our analysis, the format is intended for provider input and it is not useful for data analysis. We reorganize the data elements into the administrative and clinical categories and visualize the hierarchical representation.

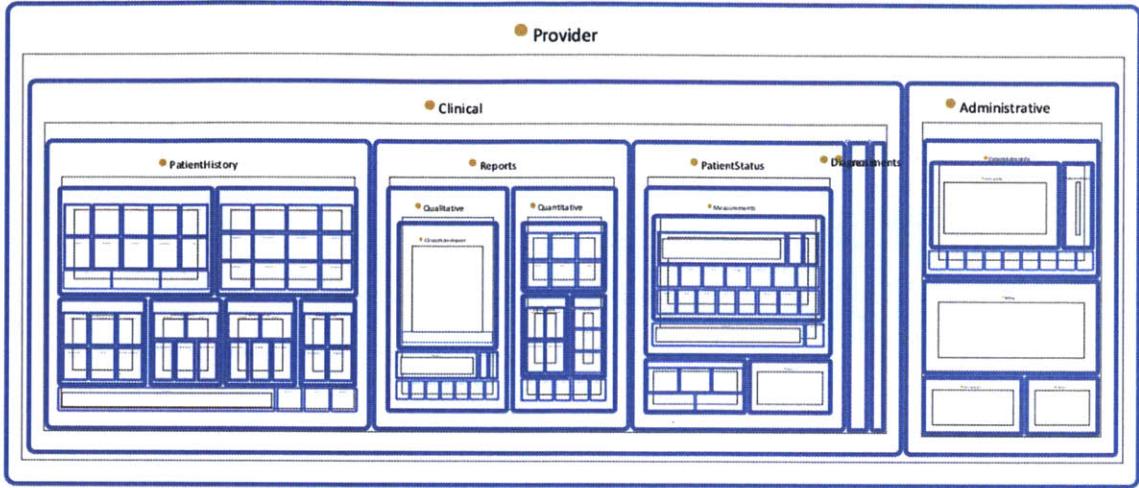


Figure 5: Visualization of Provider Data Categories

Created with Protégé software to compile and categorize the data contents
within the 10 EHR implementations

The next two sections describe the kinds of data within these two categories in more detail.

2.2.1 Administrative Data

Administrative data includes all of the information needed to identify a patient and track encounters between the patient and provider. Administrative data is organized into four categories: identifiable, encounter, billing, and follow-up; Figure 6 shows further details. Patient identifiable data include name, demographics, contact information, and status. Several more examples of data are enumerated in Figure 6.

Administrative	
PatientAdminInfo	Billing
Insurance	Codes
ContactInfo	Diagnosis
Name	Procedure
Address	Services
Telephone	Supplies
UniqueIdentifier	PaymentAmount
Demographics	PaymentSource
DOB	HMO
Gender	Medicaid
ZipCode	Medicare
Race	NoCharge
Ethnicity	PPO
Status	PrivateInsurance
Marital	SelfPay
Student	WorkerCompensation
Employment	
Occupation	
EmergencyContact	
Followup	EncounterRecord
Prescription	AdmissionRecord
Pharmacy	AppointmentRecord
Referral	DischargeRecord
	ReasonForVisit

Figure 6: Categorization of Provider Administrative Data

The encounter subcategory includes information about the patient interaction with the provider such as appointment records, admission/discharge records, and reason for visit. Referrals to other providers and corresponding contact information are included in the follow-up subcategory, in addition to prescription refill requests and pharmacy information. The billing subcategory contains the data used to generate a claim – diagnosis and procedure codes – and information about a patient’s health insurance carrier and plan (section 2.4).

2.2.2 Clinical Data

Figure 7 shows the five main subcategories within clinical data: history, status, reports, diagnosis, and treatment. Data in a patient’s clinical history – diseases, family health, past surgeries or operations, and administered vaccines – are inputs to provider diagnosis. During a patient encounter, the provider characterizes patient status by collecting subjective patient reported data like symptom descriptions, exercise and diet routines as well as provider-measured data such as height and weight.

Clinical		
PatientHistory	PatientStatus	Reports
AdverseReaction	Measurements	Qualitative
AllergyHistory	DeviceData	AutopsyReport
Allergy_SensitivityType	ElectricalMeasurements	ClinicalNotesReport
DateofLastOccurance	BCG	Appearance
ReactionSeverity	ECG	BodyBuild
ReactionTreatment	EEG	Deformities
ReactionType	SpO2	Demeanor
DiseaseHistory	VitalMeasurements	Gait
OTCTreatment	BloodGlucose	Habitus
SymptomDuration	BloodPressure	Hygiene
SymptomOnset	Diastolic	Mobility
SymptomType	Systolic	Speech
FamilyHistory	BodyMassIndex	VoiceQuality
AdultHealthHistory	BodyTemperature	PhysicalObservations
ChildHealthHistory	Crown-to-rumpLength	ConsultationReport
FatherAgeofDeath	FluidBalance	DischargeSummariesReport
FatherCauseofDeath	HeadCircumference	EmergencyReport
FatherHealthStatus	Hearing	HospitalizationReport
HereditaryDiseases	HeartRate	ImagingReport
MotherAgeofDeath	Height	CT
MotherCauseofDeath	Movement	MRI
MotherHealthStatus	Perspiration	ObservationsReport
SiblingAgeofDeath	Pulse	ProgressReport
SiblingCauseofDeath	PulseOximetry	RadiologyReport
SiblingHealthStatus	RespiratoryRate	SOAPReport
GeneticProfile	Vision	Quantitative
ImmunizationHistory	WaistToHipRatio	BiopsyReport
AgeAdministered	Weight	DiagnosticReport
DateAdministered	Social	DiagnosticTest
VaccineAdministeringPhysician	DailyRoutine	DiagnosticTestInterpretation
VaccineDose	DietPatterns	DiagnosticTestResult
VaccineLotNumber	ExercisePatterns	DiagnosticTestResultDate
VaccineType	SleepPatterns	LaboratoryReport
MedicalHistory	Symptoms	Blood
AgeofOnset	Duration	Histopathology
ConditionStatus	Location	Immunology
ConditionType	Onset	Lipid
DateDiagnosed	Severity	Microbiology
Treatment	Type	Urine
MedicationHistory	Diagnosis	OperativeReport
AllergicReaction	Disease	PathologyReport
Dosage	Condition	PhysicalExamReport
Drug	Disorder	ProceduralReport
Frequency	Overdose	ProcedureDate
PrescribedBy	Poisoning	ProcedureResult
PrescriptionDate	TraumaEvent	ProcedureType
PrescriptionNumber		SurgicalReport
QuantityForm		
QuantityNumber		
Route		
SourceofMedicationList		
StartDate		
StopDate		
SubstanceUseHistory	Treatments	
AlcoholConsumption	Procedures	
CaffeineConsumption	Devices	
DrugUse	Medication	
TobaccoUse	Therapy	
SurgicalHistory		

Figure 7: Categorization of Provider Clinical Data

Figure 8 represents graphically the patient status subcategory from Figure 7. This illustration is useful for the patient input data discussion.

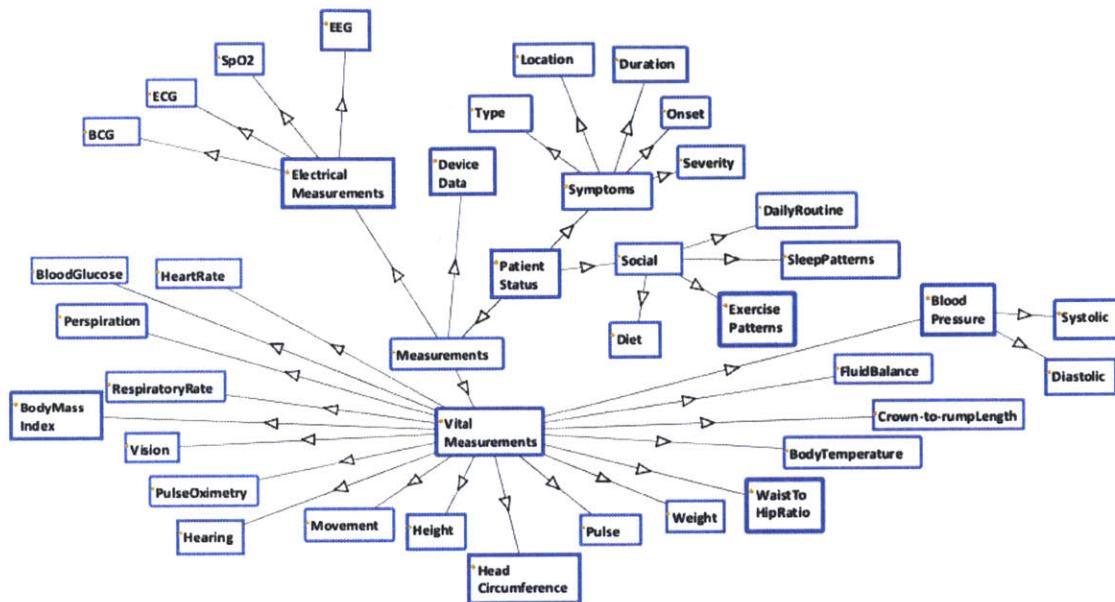


Figure 8: Visual Illustration of Provider Clinical Subcategory, Patient Status

Clinical data that is collected in reports contains both quantitative and qualitative data elements. Qualitative reports consist of consultation notes and progress reports. These are typically subjective measurements or observations made by clinicians. These reports, especially scanning reports require a degree of interpretation. Examples of quantitative reports are physical exams, biopsy reports, and most laboratory reports. This is actual measured and physical data that typically has a numerical value that is documented during the patient encounter or on a test report.

Diagnosis data are statements of provider's interpretations of a patient's state based on known, defined symptoms. This may be diseases, disorders, or conditions. There are also several forms of diagnoses that include naming of a disease or dysfunction, indication of degree of abnormality, or nosological coding. Each clinical diagnosis has a corresponding diagnosis code, classified by the international classification of diseases

(ICD). During the diagnosis process, physicians match their judgment on the state of the patient to a standard diagnosis. Selected a diagnosis code is a form of matching to a category system, with the ICD being the taxonomy of codes.

The ICD, updated and maintained by the World Health Organization (WHO), is used to classify diseases and other health problems and the codes are the basis for reimbursement. Figure 9 shows a screenshot from the ICD-10 online browser, which includes over 16,000 codes.

The screenshot shows the ICD-10 Version:2010 interface. The left sidebar lists categories under 'Malignant neoplasms of digestive organs'. The main panel displays the 'C34' chapter, specifically 'Malignant neoplasm of bronchus and lung'. It includes sub-categories for 'Main bronchus', 'Carina', 'Hilus (of lung)', 'Upper lobe, bronchus or lung', 'Middle lobe, bronchus or lung', 'Lower lobe, bronchus or lung', and 'Bronchus or lung, unspecified'. A note for 'Overlapping lesion of bronchus and lung' is present. The right sidebar includes links for 'Advanced Search', 'ICD-10', 'Versions - Languages', and 'Info'.

Figure 9: ICD-10 Screenshot of Lung Cancer Disease Classification

("ICD-10 Version:2010," 2010)

ICD-10 CM and ICD-10 PCS are the national modifications of ICD-10, which include more detail about clinical modifications of diseases (CM) and procedures (PCS). These databases are maintained by the United States' Centers for Disease Control and Prevention (CDC) and contain 68,000 and 76,000 codes respectively ("ICD - ICD-10-CM - International Classification of Diseases, Tenth Revision, Clinical Modification," 2012).

The provider treatment subcategory contains information about the therapy or medication that is prescribed as a result of a patient encounter or report analysis. There are many other subcategories of medication, including drug type, dosage, and frequency, as indicated in the medication history subset. The provider clinical record includes the drug prescribing history, but does not account for when the drug is actually distributed to the patient.

We make three observations based on the assessment of provider data. First, some of the data collected is subjective and the provider needs to interpret the data based on experience and knowledge. Second, the diagnosis is a result of a matching process to taxonomy of codes, indicating that the provider must select a pre-defined category that most represents the patient state. Third, the provider data collection is more focused on diagnosis and intervention selecting, rather than monitoring for outcomes.

The EHR and ICD standards suggest commonality, however there is still much operational and workflow variance among providers, resulting in a wide range of data quality. As well, patients visit several different providers during their lifetime, so the records kept by a single provider do not always reflect the patient's complete medical history. Next, we discuss data that is collected and owned by the patient that can be used to supplement the data managed by the provider.

2.3 Patient Data

From a data perspective, patients are the ultimate source, with data being collected through testing, sensing, monitoring and direct query. However, the way in which the health system incorporates this data varies by stakeholder.

New technology is enabling the collection of data outside of the traditional provider environment. The kinds of data discussed in this section are patient input data into

personal health records, dynamic physiological signals, and patient acquired genetic information.

2.3.1 Personal Health Record

Much of the data captured in a patient's medical record or EHR originates from the patient. During patient encounters with providers, patients share information about their symptoms and behaviors, and report on treatment feedback. In most cases, this is where data input stops. Patients typically do not document this kind of data, so the provider record is relied on for data collection and storage.

In some cases, patients may decide to maintain a health diary or personal health record (PHR), although currently not the norm as PHR adoption is low. PHRs are effectively like EHRs but are managed by the patient. A patient may use a PHR to collect all health information and maintaining a personal medical record for individual access. In addition to information from their EHR like allergies, lab results, and medications PHRs may include patient-reported outcome data and passively collected data from monitoring devices. Examples of PHR vendors for the patient consumer include Doclopedia, EmryStick, JuniperHealth, MiVIA, and ZweenaHealth.



Figure 10: ZweenaHealth PHR Features

("ZweenaHealth," 2012)

Services like Zweena claim to enable 24/7 access to complete medical records through an online secure platform. A benefit to the patient may be convenient management of the compilation of their health data from various sources. Patients own their PHR and have the option to permit external access or integration to a provider EHR system. The data in the PHR is from the perspective of the patient, while the EHR remains provider-focused. Both datasets may be used in tandem to yield a more holistic view of the patient's medical status, given the increase adoption of PHRs.

2.3.2 Physiological Measurements

Often the physiological signals measured during a patient-provider encounter, like blood pressure or heart rate, are not reflections of day-to-day average measurements. As well, these point-of-care collected data points do not give indication to conditions like arrhythmia, seizures, or other episodic driven disorders. Remote physiological measurements create a new category of data, which is in between encounter measurements, and may give a different view on of data. The results can provide additional data to enhance the clinical assessment of the patient. Table 1 lists several of the physiological signals that can be dynamically measured.

Measurement	Description	Method of Measurement
body temperature	temperature of mouth, skin, ear, under arm, or anus (oral normothermia 37°C)	thermometer, thermocouple, thermistor
pulse	arterial palpation of the heartbeat to determine cardiac performance	finger touch, LED/photodiode, stethoscope
heart rate	number of heart beats per unit time (typ 60-90 bpm), measure of exercise efficiency	photoplethysmography sensor, electrode monitor on chest or wrist
blood pressure	pressure exerted by circulating blood upon the walls of blood vessels (avg 110/65 – 140/90 mmHg), measure of cardiovascular health	sphygmomanometer, pulse wave velocity
respiratory rate	number of breaths per unit time (60 sec) typically 12-20 for adults, indicator of potential respiratory dysfunction	counting breaths, transducing sensor
fluid balance	human homeostasis, amount of fluid lost from the body is equal to the amount of fluid taken in	observation, body weight, urine output, blood chemistry
perspiration	thermoregulation, production of a fluid consisting primarily of water as well as various dissolved solids (chiefly chlorides), that is excreted by the sweat glands	thermal sensor, LED/pH sensor
movement	physical activity	accelerometer, pedometer, motion sensor
vision	visual acuity, acuteness or clearness of vision, which is dependent on the sharpness of the retinal focus within the eye and the sensitivity of the interpretative faculty of the brain	Snellen chart
hearing	perceive sound by detecting vibrations through an organ such as the ear	audiometer
weight	lbs / kgs, needed for medicine dosing	mass scale
height	inches / meters, needed for medicine dosing	linear scale
body mass index	BMI, or Quetelet index, heuristic proxy for human body fat based on an individual's weight and height, indication of obesity	calculated from weight and height measurements
waist to hip ratio	girth ratio - measure of regional fat distribution	calibrated tape
blood glucose	measurable amount of glucose (sugar) in the blood	enzyme electrode, glucose oxidase strip
EEG	Electroencephalography, electrical activity in brain to assess brain death, seizures	scalp electrodes
ECG, EKG	Electrocardiography, electrical activity in heart to assess regularity of heart beats	chest electrodes, sensor
BCG	ballistocardiograph, measure of ballistic forces on the heart	accelerometer, electrodes
SpO ₂	oxygenation, oxygen saturation of tissue, blood	near infrared spectroscopy, pulse oximeter

Table 1: Description of Physiological Measurements

("Epson Enters Healthcare Business with Wristwatch-Type Pulse Monitor," 2012)
 ("Wearable Blood Pressure Sensor Offers 24/7 Continuous Monitoring," 2009)
 ("Measuring and Managing Fluid Balance," 2011)

Automatic collection of physiological measurements is possible with the use of devices and monitors. There are several hundred ambulatory devices that attach to the body to enable dynamic monitoring of vital signals. Some of these devices are offered direct to consumer (patient), others are provider prescribed. The collected data can either be stored on the device, smartphone, or synced to a PHR/EHR platform. Many of these systems, approved as medical devices, alert the patient and/or provider of

abnormalities and recommend interventions. Table 2 highlights the various features of selected devices with a range of usages.

	Heart rate	Respiratory rate	Fluid status	Accelerometer / altimeter motion sensor	Perspiration / galvanic skin response	Blood Pressure	Skin Temperature	SPO ₂	ECG	Weight	Body fat	Body Mass Index (BMI)	Blood sugar
corventis	✓	✓	✓	✓					✓				
Sotera WIRELESS	✓	✓				✓	✓	✓	✓				
BODYMEDIA make your life a genuine life				✓	✓		✓						
fitbit				✓									
BASIS	✓		✓	✓		✓							
Withings						✓				✓	✓	✓	
VitaDock®						✓	✓			✓	✓	✓	✓
JAWBONE®				✓									
digifit Get Fit Stay Healthy	✓			✓		✓			✓				
VALENCELL	✓	✓		✓									
BAM BAM Labs	✓			✓									
CardioDefender	✓							✓					

Table 2: Patient Monitoring Devices

These monitoring devices improve signal capture from an individual data point to a data series collected over a period of time. Data trends can then be analyzed to develop a more complete assessment of patients' health. For instance, Corventis' AVIVO® system features a wearable device that attaches to the chest and measures

fluid status, heart rate, respiratory rate, posture, activity, and ECG. This data is transmitted to a handheld device that is connected to Corventis' clinical monitoring center. The physiological trends are reviewed by the patient's physician who uses the data to track heart failure and guide clinical decisions, enabling early intervention.(Corventis, 2012)

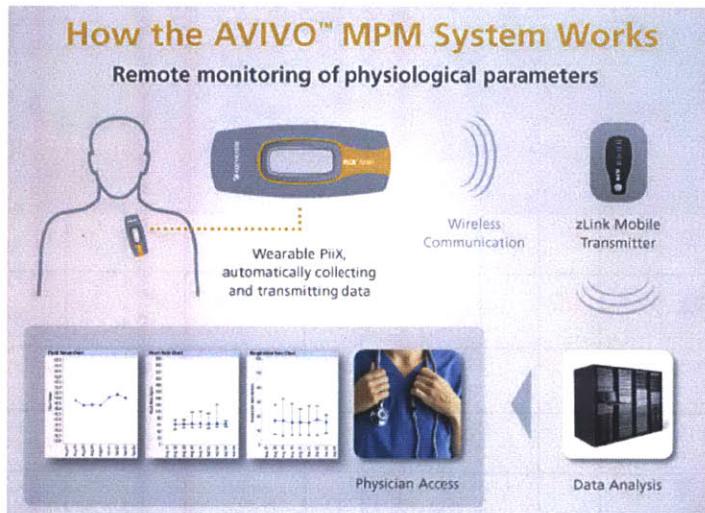


Figure 11: Corventis' Mobile Patient Management System for Heart Failure Patients

2.3.3 Genetic Data

Over 1,300 genetic tests have been developed since 1986 with about 1,000 currently available from testing laboratories ("Personalized Medicine: Trends and Prospects for the New Science of Genetic Testing and Molecular Diagnostics," 2012). Genetic data is determined by collecting a sample of DNA from blood or saliva that may be used to:

- diagnose disease
- confirm or rule out suspected genetic conditions
- determine risk of developing or passing on a genetic disorder
- inform the type or dose of a drug that will yield an effective response
- determine variety of predispositions, characteristics, and phenotypes

Testing service companies now offer DNA analysis direct to consumer. For approximately \$100, individuals can obtain genetic comparison of their DNA to 120 diseases, 57 traits, 48 carriers, and 21 drug responses (“Genetic Testing for Health, Disease & Ancestry; DNA Test - 23andMe,” 2012). This is a limited amount, compared to the 16,000 ICD disease codes. Direct to consumer genetic testing vendors include 23andMe, deCODE Genetics, easyDNA, and VuGene. It is intended and claimed by these vendors that genetic data may be used to individualize care and inform clinical decisions.

Advancements in science, technology, and information systems are enabling patients to take a more active role in their health and be at the forefront of their own data collection. Monitoring devices are infiltrating every day life and knowledge of the chemical makeup of the human body continues to advance, and so do the personal data collection opportunities. Services are beginning to cater to the well-informed patient, although adoption by patients remains low.

2.4 Payor Data

Most payors, public and private, use the data formats guided by the ICD standard, which enables comparison and we can analyze them together. Each interaction between the patient and provider is translated into a coded claim used for billing and reimbursement. Interactions include outpatient visits, inpatient stays, laboratory services, prescription filling, and medical device distribution. The claim includes the date and place of service and applicable codes, which are used to calculate the payment amount. The Centers for Medicare and Medicaid Services (CMS) provide numerous forms to assist payors in collecting the data required for provider reimbursement. Private health insurers, like Aetna and Tufts Health Plan, may use different claim forms than CMS-1500, but the content is nearly identical. A sample claim form is shown in Figure 12.

HEALTH INSURANCE CLAIM FORM APPROVED BY NATIONAL UNIFORM CLAIM COMMITTEE 08/05											
CARRIER											
PICA											
1. MEDICARE MEDICAID TRICARE CHAMPVA <input type="checkbox"/> (Medicare #) <input type="checkbox"/> (Medicaid #) <input type="checkbox"/> CHAMPUS <input type="checkbox"/> (Member ID#) <input type="checkbox"/> (Sponsor's SSN)				GROUP HEALTH PLAN FECA <input type="checkbox"/> (SSN or ID) <input type="checkbox"/> BLK LUNG <input checked="" type="checkbox"/> (SSN or ID) <input type="checkbox"/> OTHER				1a. INSURED'S I.D. NUMBER X987-1234A-032 (For Program in Item 1)			
2. PATIENT'S NAME (Last Name, First Name, Middle Initial) THOMPSON, ANNE MARIE, H				3. PATIENT'S BIRTH DATE SEX MM DD YY M F X				4. INSURED'S NAME (Last Name, First Name, Middle Initial) THOMPSON, ROBERT, H			
5. PATIENT'S ADDRESS (No., Street) 5813 CRADLE ROCK COURT				6. PATIENT RELATIONSHIP TO INSURED Self <input type="checkbox"/> Spouse <input checked="" type="checkbox"/> Child <input type="checkbox"/> Other <input type="checkbox"/>				7. INSURED'S ADDRESS (No., Street) 5813 CRADLE ROCK COURT			
CITY ST PAUL		STATE MN		CITY ST PAUL		STATE MN					
ZIP CODE 00123-0054		TELEPHONE (Include Area Code) (001) 5551212		CITY ST PAUL		STATE MN					
9. OTHER INSURED'S NAME (Last Name, First Name, Middle Initial) THOMPSON SR, ROBERT, G a. OTHER INSURED'S POLICY OR GROUP NUMBER X0987654321											
b. OTHER INSURED'S DATE OF BIRTH SEX MM DD YY M X F											
c. EMPLOYER'S NAME OR SCHOOL NAME US ARMY (RETIRED)											
d. INSURANCE PLAN NAME OR PROGRAM NAME MAMSI SENIOR PREFERRED PPO											
READ BACK OF FORM BEFORE COMPLETING & SIGNING THIS FORM.											
12. PATIENT'S OR AUTHORIZED PERSON'S SIGNATURE. I authorize the release of any medical or other information necessary to process this claim. I also request payment of government benefits either to myself or to the party who accepts assignment below. SOF											
SIGNED DATE 07/01/2007											
10. IS PATIENT'S CONDITION RELATED TO: a. EMPLOYMENT? (Current or Previous) <input type="checkbox"/> YES <input checked="" type="checkbox"/> NO b. AUTO ACCIDENT? <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO PLACE (State) MN c. OTHER ACCIDENT? <input type="checkbox"/> YES <input checked="" type="checkbox"/> NO											
11. INSURED'S POLICY GROUP OR FECA NUMBER BSBS54321											
a. INSURED'S DATE OF BIRTH MM DD YY 01 18 1956 SEX M X F											
b. EMPLOYER'S NAME OR SCHOOL NAME PRICE WATERHOUSE COOPERS											
c. INSURANCE PLAN NAME OR PROGRAM NAME BLUE CROSS BLUE SHIELD PPO											
d. IS THERE ANOTHER HEALTH BENEFIT PLAN? <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO If yes, return to and complete item 9-a-d.											
13. INSURED'S OR AUTHORIZED PERSON'S SIGNATURE I authorize payment of medical benefits to the undersigned physician or supplier for services described below. SOF											
SIGNED DATE 07/01/2007											
14. DATE OF CURRENT: ILLNESS (First symptom) OR 07/01/2007 INJURY (Accident) OR PREGNANCY(LMP)				15. IF PATIENT HAS HAD SAME OR SIMILAR ILLNESS, GIVE FIRST DATE MM DD YY 07 01 2007				16. DATES PATIENT UNABLE TO WORK IN CURRENT OCCUPATION MM DD YY MM DD YY FROM 07 01 2007 TO 07 20 2007			
17. NAME OF REFERRING PROVIDER OR OTHER SOURCE ROBERT SMITH MD				17a. 1B 987654321				16. HOSPITALIZATION DATES RELATED TO CURRENT SERVICES MM DD YY MM DD YY FROM 07 01 2007 TO 07 02 2007			
19. RESERVED FOR LOCAL USE 1234567890A								20. OUTSIDE LAB? \$ CHARGES <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO 150050			
21. DIAGNOSIS OR NATURE OF ILLNESS OR INJURY (Relate Items 1, 2, 3 or 4 to Item 24e by Line)											
1. E88 5 0				3. E90 1				22. MEDICAID RESUBMISSION CODE 12345678955 ORIGINAL REF. NO. ABC1234567890			
2. E23 4											
4. 903 5490											
24. a. DATE(S) OF SERVICE From MM DD YY To MM DD YY b. PLACE OF SERVICE EMO				c. D. PROCEDURES, SERVICES, OR SUPPLIES (Explain Unusual Circumstances) CPT/HCPCS MODIFIER				e. DIAGNOSIS POINTER			
1 7BEGIN 1245 END 1425 TIME 90 MINUTES 07 01 07 07 05 07 22 N 00770 25 26 LT RT 134				2 7BEGIN 1245 END 1415 07 01 07 07 01 07 22 Y 00770 P2 P2 P2 P2 134				3 ZZKAYE WALKER 07 01 07 07 01 07 12 N E1399 25 26 26 28 12			
4 N400026064871 IMMUNE GLOBULIN INTRAVENOUS UN2 07 01 07 07 01 07 11 N J1563 RT LT 28 29 13				5 VPA122BIC5D6E7G 07 01 07 07 02 07 11 N A6410 13				6 OZ00301134678906 07 01 08 07 07 07 11 Y A6410 25 26 RT LT 1 3			
7 FEDERAL TAX I.D. NUMBER 555666777888				8. PATIENT'S ACCOUNT NO. 20070613235249				9. ACCEPT ASSIGNMENT? <input checked="" type="checkbox"/> YES <input type="checkbox"/> NO			
								10. TOTAL CHARGE \$ 2930 50			
								11. AMOUNT PAID \$ 30 49			
								12. BALANCE DUE \$ 2900 01			
31. SIGNATURE OF PHYSICIAN OR SUPPLIER INCLUDING DEGREES OR CREDENTIALS If certify that the statements on the reverse apply to this bill and are made a part thereof.											
32. SERVICE FACILITY LOCATION INFORMATION LABCORP DIAGNOSTOC 123 HEALTHCARE LANE ST PAUL MN 00342-1111											
33. BILLING PROVIDER INFO & PH# (012) 5551212 THE PEDIATRICS GROUP 1234 MAIN STREET COLUMBIA MN 00123-0765											
APPROVED CMS-1500 FORM CMS-1500 (02/02) PLEASE PRINT OR TYPE											

Figure 12: CMS-1500 SmartForm, Health Insurance Claim Form

(“New CMS-1500 (08/05) Healthcare Claim PDF SmartForm With Built in 837P EDI Capabilities.” 2007)

As described in the provider section, clinical information about a patient is translated into a standard code that represents the patient diagnoses, procedures performed, services delivered, or supplies tendered. Example codes are shown in items 21, 24D, and 24E in Figure 12 and their meaning can be searched on various online databases, including the WHO and CMS. Most every disease and clinical service has a corresponding code, used nationally for billing. This format defines the communication structure between providers and payors.

Dates are recorded on a claim to capture the time sequence of clinical events. The codes captured in a claim represent the closest description of a diagnosis, procedure, or service for which a code exists. Due to the code renewal schedule, new procedures may be available that do not yet have a unique code. As well, due to the coding complexity and variations among providers, errors in claims are frequently observed. The recorded claim error rates for various payors have ranged from nine to sixty percent, according to the American Medical Association's (AMA) annual health insurer report card (American Medical Association, 2012).

Besides for the series of codes, the payors collect additional information about the patient and provider. Some of this information is also included on the claim form, while some is collected during enrollment. The five categories of data payors use to process reimbursements and manage patients are patient administration, product, medical claim, pharmacy claim, and provider information, as shown in Figure 13. These five categories represent five separate databases that payors manage.

PatientAdminInfo	MedicalClaim	PharmacyClaim
SocialSecurityNumber	RecordOfInteractions	DrugInfo
MemberIdentification	Outpatient	DrugCode (NDC)
AccountNumber	Inpatient	DrugName
ContactInfo	Emergency	PrescriptionInfo
Name	Dental	Number
Address	Dates	Date
Telephone	DateOfService	New/Refill
Demographics	DateOfFirstSymptom	DosePerDay
DOB	DateFirstConsulted	DaysSupply
Gender	DateOfAccident	Quantity
ZipCode	DateOfPregnancy	Strength
Race	DateOfPriorIllness	Generic
Ethnicity	DateOfDisability	FormularyCode
Status	DateReturnToWork	RouteOfAdministration
Marital	DateOfAdmission	Natureofillness
Student	DateOfHospitalization	NatureOfInjury
Employment	DateOfDischarge	PharmacyInfo
Authorization	DateOfPayments	PharmacyName
Signature	Codes	PharmacyIdentification
EnrollmentHistory	Diagnosis	PharmacyAddress
RelationshiptoInsurer	Primary (ICD)	PrescribingPhysicianInfo
InsurerInfo	Secondary (ICD)	PurchaseDate
SubscriberInfo	Procedures	Costs
BeneficiaryInfo	HCPCS	ChargeAmount
EmployerInfo	CPT	CoPay
SchoolInfo	CDT (dental)	Reimbursement
OtherHealthBenefitPlans	CodeModifier	
Product	ICD	ServiceProvider
ProductIdentification	Revenue	Name
ProductName	Services	Gender
PolicyNumber	Laboratory	DOB
PolicyGroup	Supplies	ContactInfo
ProgramName	MedicalDevices	ProviderDemographics
PlanType	Drug	ProviderIdentification
HMO	Resubmission	LicenseInfo
PPO	Costs	ProviderTaxID
POS	ChargeAmount	ProviderType
EPO	PaidAmount	Specialty
CoveragePlan	PrepaidAmount	PrimaryCarePhysician
BenefitLevel	CoPay	PrescribingPhysician
CoverageType	Coinsurance	PlaceOfService
Payments	Deductible	FacilityType
CoPays	Reimbursement	FederalTaxID
Deductibles	Days/Units	Address
Coinsurance	DenialInfo	Telephone
MedicalCoverage	PriorAuthorizationNumber	ReferralInfo
PrescriptionDrug		EMRVendor
DentalCoverage		AcceptingNewPatients
Vision		
BehavioralHealth		
Laboratory		
DiseaseManagement		
Disability		

Figure 13: Categorization of Payor Data

The data in the patient administrative and product categories may be used before a patient receives care to vet plan eligibility and determine coverage type. Once care has been received, claim processors assess patient eligibility and payment applicability of the services rendered. Most health care insurance plans have coverage for prescriptions and the details of drug distribution are recorded on a pharmacy claim. This includes the drug code, strength, dosage, and quantity as well as pharmacy and prescribing physician information. Pharmacy claims themselves do not include the diagnosis code relevant to the prescription, but may have a brief description of the patient's condition. Currently, there is no standard link between pharmacy claims and medical claims. Pharmacy related data is further discussed in section 2.7.

The last category is the provider, which includes data about the physician submitting the claim. The payor requests standard information (national identification number, license, specialty, and facility type) in order to properly identify the provider and verify credentials.

The volume and availability of claims have led to analytic efforts aimed at optimizing health care. Groups of payors combine claims about their patients/employers into commercial claims databases, such as PharMetrics, MarketScan, and Medco. Recently, several states in the United States have initiated the development of an all-payer claims database (APCD). This database combines claims of all members within a state for the purpose of facilitating a holistic view of cost and utilization (Patrick, Murray, Bigby, & Boros, 2012). These efforts to collate different kinds of data for the purposes of analysis and research stress further the need to ensure better quality assurance practices.

The transactions between the payor and provider rely on a series of codes, yet result in decisions that drive the health care system. With limited clinical details to support the claims, the payor's assessment of the value of patient care is based on claim

interpretation. As well, the payor manages five separate databases for each function it provides and experiences an abundance of information due to the nature of the data. The availability of data presents opportunities to relate and compare, regardless of the quality of the data.

2.5 Manufacturer Data

This section overviews the variety of data that is managed by the manufacturers of drug products. We describe the data that is collected during each development phase: pre-discovery, discovery, development, and clinical trials.

The time from drug discovery to adoption is approximately ten years and over that period manufacturers store and maintain data for compounds. The data originates from basic science, pre-discovery, discovery, and evolves into information about the safety, efficacy, and effectiveness of a compound in humans. Several kinds of data are introduced into this process and Figure 14 gives an overview from the perspective of data. We use the United States FDA's regulatory process as an example for this section.

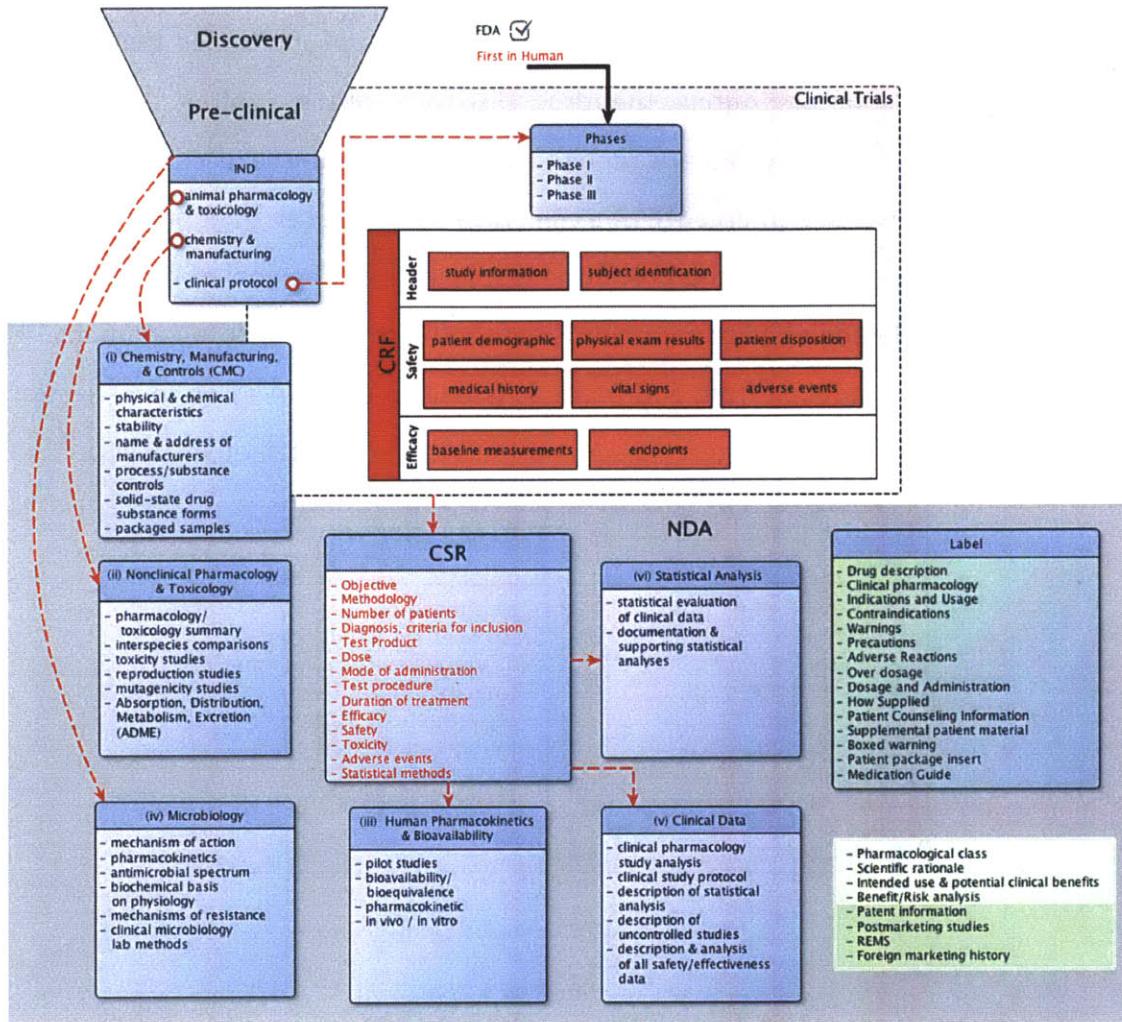


Figure 14: Categories of Data in the Drug Development and Approval Process

In order to license a drug, regulators must review the clinical trials and other documentation that supports the benefit claims of the drug made by the sponsor, or drug manufacturer. Manufacturers have incentive to adapt their processes and how they manage data to the rules and communications around the regulator.

The manufacturer and regulator sections that follow individually describe the data elements that are enumerated in Figure 14.

2.5.1 Drug Discovery

Scientific analysis of molecular targets and identification of molecules that interfere with such targets occurs during the drug discovery phase. After the target is determined, 3 to 7 more years are required to determine the lead compound. The result of the phase is the emergence of a candidate drug that can then be tested in the lab in vitro and in vivo, during the preclinical testing phase. The goal of this phase is to establish the safety and efficacy parameters before the compound enters humans. The lab and test data compiled from this phase is issued to the FDA as an investigational new drug (IND) submission. Figure 15 summarizes the data contained in an IND.

(i) Chemistry, Manufacturing, & Controls (CMC)	(ii) Nonclinical Pharmacology & Toxicology
<ul style="list-style-type: none">- physical & chemical characteristics- stability- name & address of manufacturers- process/substance controls- solid-state drug substance forms- packaged samples	<ul style="list-style-type: none">- pharmacology/ toxicology summary- interspecies comparisons- toxicity studies- reproduction studies- mutagenicity studies- Absorption, Distribution, Metabolism, Excretion (ADME)

Figure 15: Data Included in IND

(Investigational New Drug)

2.5.2 Drug Development

Once the regulator approves and reviews the IND, drug sponsors can then begin testing their product in humans. The clinical trial protocol is a formulary of the data that needs to be collected during a study to answer specific research questions. The test hypothesis is evaluated through measuring endpoints that are defined in the protocol. These endpoints are typically measurements of response rate or survival time. The protocol describes the type of people that may participate in the trial, schedule of tests, procedures, medications, dosages, and study length. Participants in a clinical trial are monitored regularly to assess health, disease state, and determine the safety and efficacy of the treatment.

The data is collected as per the protocol and reported on the case report form (CRF), which is the data-reporting document that is used during the trial. The CRF is designed to organize the data collection and allow for efficient analysis by biometrists and statisticians. CRFs are used to track subjects during clinical trials and the data contained is used to make decisions about the drug, benefits, risks, and marketability.

CRFs can either be in paper-based or digital formats. Electronic CRFs (eCRF) are referred to as remote site monitoring (RSM) or remote data entry (RDE). Procedurally, there is no difference between the CRF and the EHR, except the protocol provides a means of what to collect (prescriptive) and the CRF is designed for analysis and monitoring. Figure 17 shows an excerpt from a CRF and Figure 18 summarizes the nature of the CRF data.

Study Code:	Randomization no:	<input type="text"/>	Subject Initials:	<input type="text"/>
CASE REPORT FORM				
STUDY TITLE				
Study reference number				
CLINICAL TRIAL SITE/UNIT:				
PRINCIPAL INVESTIGATOR:				
Subject Initials:				
Subject Randomization Number:				
<p><i>I am confident that the information supplied in this case record form is complete and accurate data. I confirm that the study was conducted in accordance with the protocol and any protocol amendments and that written informed consent was obtained prior to the study.</i></p> <p>Investigator's Signature: _____</p> <p>Date of signature: <input type="text"/> D <input type="text"/> d <input type="text"/> m <input type="text"/> m <input type="text"/> m <input type="text"/> y <input type="text"/> y <input type="text"/> y <input type="text"/> y</p>				
<p>VISIT 1 (SCREENING)</p> <p>Date: <input type="text"/> DD <input type="text"/> MM <input type="text"/> YYYY</p> <p>INFORMED CONSENT</p> <p>Please note: written informed consent must be given before any study specific procedures take place or any current therapy is discontinued for the purposes of participation in this study.</p> <p>Has the subject freely given written informed consent? Yes <input type="checkbox"/> No <input type="checkbox"/></p> <p>DEMOGRAPHIC DATA</p> <p>Age (yrs): <input type="text"/> Sex: Female <input type="checkbox"/> Male <input type="checkbox"/> Height (m): <input type="text"/> • <input type="text"/> Weight (kg): <input type="text"/> • <input type="text"/> Body Mass Index (BMI = Wt (kg)/H² (M)): <input type="text"/> • <input type="text"/></p> <p>SMOKING HABITS</p> <p>Does the subject smoke or use tobacco products? Yes <input type="checkbox"/> No <input type="checkbox"/> * How many cigarettes per day? <input type="text"/> Other, specify _____</p> <p>ALCOHOL CONSUMPTION</p> <p>Does the subject consume alcohol? Yes <input type="checkbox"/> No <input type="checkbox"/> If yes, how many units per week? <input type="text"/></p> <p>MEDICATIONS TAKEN</p> <p>Is the subject currently or previously taking any medication including OTC, vitamins and/or supplements? Yes <input type="checkbox"/> No <input type="checkbox"/> *Record all medication on Concomitant Medications page</p>				

Figure 16: Sections from Case Report Form

Adverse Events										
Has the patient experienced any Adverse Events since signing the Informed Consent? <input type="checkbox"/> Yes, specify below <input type="checkbox"/> No										
AE no.	Adverse Event (diagnosis if known) or symptoms/abnormalities	Start Date (date/time of symptom onset and Time (24 hour clock))	Stop Date (date/time of symptom resolution and Time (24 hour clock))	Outcome 1=Recovered 2=Recovering 3=Discontinued 4=Patient Died 5=Change in All Unknown	Severity 1=Mild 2=Moderate 3=Severe	Possible relationship to Study Drug 1=None 2=Probable 3=Definite 4=Discontinued Temporarily 5=Discontinued	Action taken with Study Drug 1=None 2=Temporary Reduction 3=Once Reduced 4=Discontinued Temporarily 5=Discontinued	Withdrawn due to AE?	Serious AE (SAE)?	If SAE Does It require immediate reporting? (see Protocol)?
		/ / : :	/ / : :			<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No
		/ / : :	/ / : :			<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No
		/ / : :	/ / : :			<input type="checkbox"/> Yes <input type="checkbox"/> No		<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No

Figure 17: Sections from Case Report Form (con't)

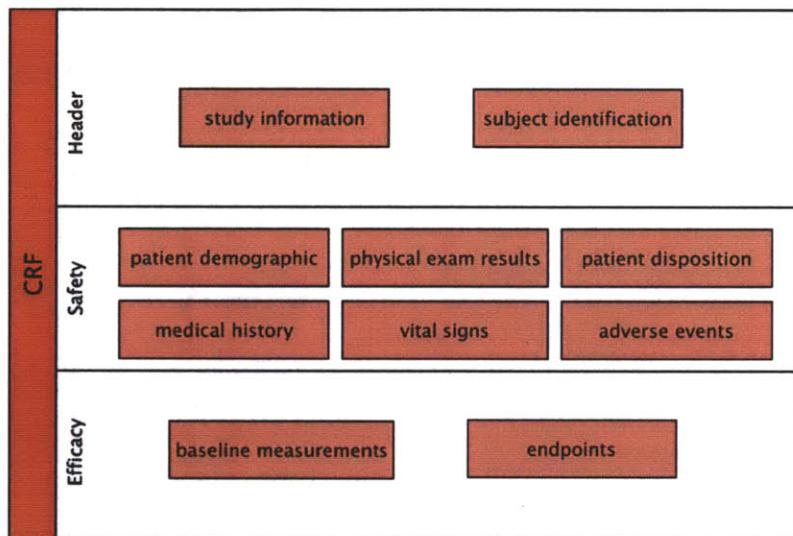


Figure 18: Categorization of Case Report Form (CRF) Data

There are three major parts of the CRF: header, safety, and efficacy. The header includes the study number, site/center number, subject identification number, and subject demographic information. The safety section includes data describing medical history, physical exam results, vital signs, patient disposition, concomitant medications, and adverse events. Adverse events include any untoward medical occurrence that results in death, is life threatening, requires hospitalization, or causes a significant incapacity. The final section of the CRF is efficacy, which includes information on baseline and endpoint measurements.

Once a study is complete, the CRF is collated into a clinical study report (CSR) and submitted to the FDA (Figure 19).

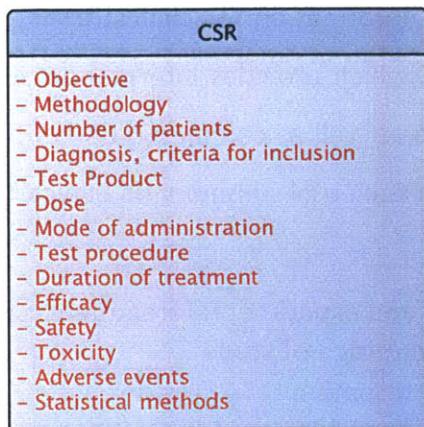


Figure 19: Data Included in Clinical Study Report

2.5.3 Clinical Trials

Clinical trials are divided into consecutive phases, with an increasing number of patients. The chart below describes the differences between each phase and the type of data that is collected at each stage.

	Number of people	Primary purpose	Secondary purpose	Study design
Phase I	20-80 healthy subjects	<ul style="list-style-type: none"> • safety • dosing • side effects 	<ul style="list-style-type: none"> • tolerability • drug reaction (ADME) • pharmacokinetics • pharmacodynamics 	
Phase II	100-500 patients with targeted condition, disease	<ul style="list-style-type: none"> • efficacy • safety • effectiveness 	<ul style="list-style-type: none"> • safe dose range • active dose range 	randomized, double-blinded
Phase III	1,000-5,000 patients (diverse)	<ul style="list-style-type: none"> • safety profile • expanded testing of effectiveness 		randomized blinded
Phase IIIb		additional safety data	test drug for additional conditions for which it may prove useful	
Phase IV		expand testing of proven drug to broader patient population	compare long term effectiveness and/or cost of drug to other marketed drugs	
Post Approval Studies	new age group, new patient types	focus on unknown side effects or risk factors		

Table 3: Clinical Trial Phases

(FDA Guidance for Industry: Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products, 2012)

There are many sources of information for clinical trials. The NIH manages the web database, ClinicalTrials.gov, which provides information about nearly 140,000 studies, both in the recruiting phase as well as completed trials. The kinds of information that are reported and listed with each trial summary on ClinicalTrials.gov are (NIH, 2007):

- disease or condition and experimental treatments studied
- title, description, and design of study
- requirements for participation
- locations where the study is available
- contact information
- links to relevant information at other health web sites, such MedlinePlus and PubMed

Some of the completed trials have reported results such as (NIH, 2007):

- participant flow
- baseline characteristics
- outcome measures and statistical analyses
- adverse events information
- administrative information

The same kinds of trial information are also shared by private sponsors through clinical registries.

The manufacturer makes decisions about whether to pursue a drug through review of the data created in the drug discovery and development phases. The data is collected into various standard forms including INDs, CRFs, CSRs, and populated in clinical trial databases. We believe the CRF contains the raw patient data that is most similar to the EHR. Manufacturers share INDs and CSRs with the regulator, but do not share the CRFs, which contain the most pertinent, patient level data.

2.6 Regulator Data

Regulators work closely with manufacturers during the approval process and provide standards for data transfer. The regulator is not permitted to share data, even if two manufacturers are developing a similar drug, due to confidentiality policies. The next sections describe the contents of the new drug application (NDA) submission, drug label, and other data that is collected and maintained by the FDA.

2.6.1 New Drug Application (NDA)

The NDA is a submission format to the FDA for drugs that have passed clinical trials. The NDA must be approved by the regulatory bodies prior to commercialization, for the sale and marketing of new drugs.

The goals of the NDA are to provide enough information to permit a FDA reviewer to reach the following key decisions (FDA, 2012a):

- Whether the drug is safe and effective in its proposed use(s), and whether the benefits of the drug outweigh the risks.
- Whether the drug's proposed labeling (package insert) is appropriate, and what it should contain.
- Whether the methods used in manufacturing the drug and the controls used to maintain the drug's quality are adequate to preserve the drug's identity, strength, quality, and purity.

The documentation required in a NDA should tell the drug's whole story: what happened during the clinical tests, what the ingredients of the drug are, the results of the animal studies, how the drug behaves in the body, and how it is manufactured, processed and packaged (FDA, 2012a). In order to answer these questions and provide a comprehensive documentation package, the FDA requires that the NDA be organized into the following six technical sections:

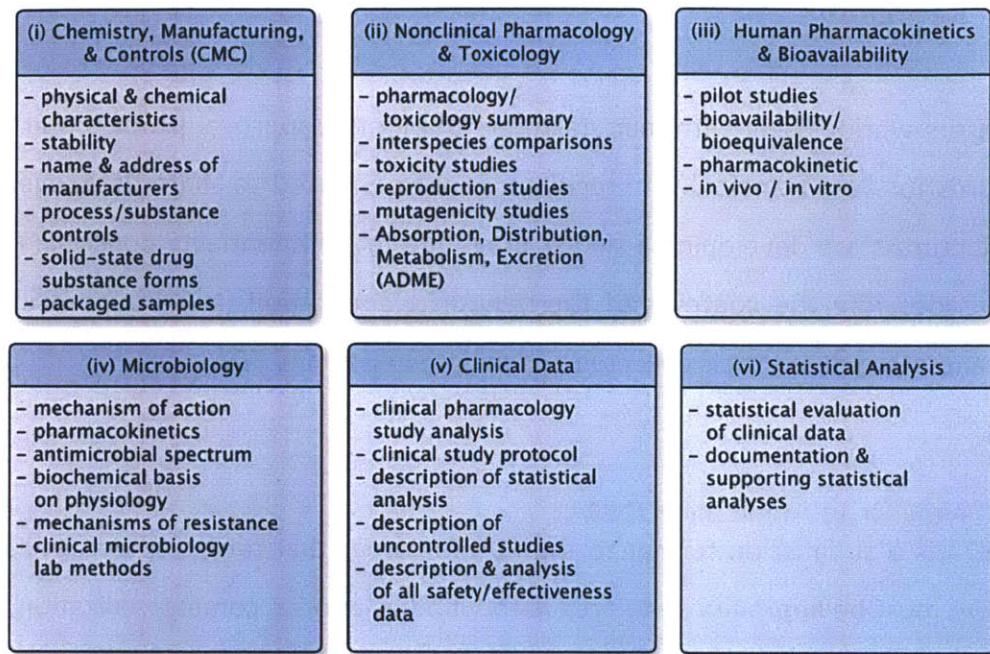


Figure 20: Six Sections of the New Drug Application (NDA)

Adapted from (FDA, 2012a)

Sections (i) chemistry and manufacturing and (ii) nonclinical pharmacology and toxicology are submitted to the FDA in the IND and are updated throughout the human trials. Section (iv) microbiology is informed by the data collected in the discovery phase and analysis of lab data from the clinic. The other three sections, (iii) human pharmacokinetics, (v) clinical data, and (vi) statistical analysis rely on input from the clinical trials and data reported on the CSR. The clinical data section must also contain the case reports for all patients who died during a study and for patients who did not complete a study because of any adverse event. Case report tabulations must be submitted for individual patients for the initial clinical pharmacology studies, controlled clinical studies, as well as all corresponding safety data.

Other information that is contained in the NDA submission is:

- Pharmacological class
- Scientific rationale
- Intended use & potential clinical benefits
- Benefit/Risk analysis
- Patent information
- Postmarketing studies
- REMS
- Foreign marketing history

Figure 21: Additional Data Included in a New Drug Application (NDA)

(FDA, 2003)

REMS are the FDA's risk evaluation and mitigation strategies that are enforced onto manufacturers to ensure that the benefits of a drug outweigh its risks (FDA, 2012b). Depending on the safety profile of the drug, REMS may be required for approval or if new safety information is learned about the product. REMS take various forms and are agreed upon between the manufacturer and FDA. The basic components of REMS include (FDA, 2012c):

- medication guide – pamphlet distributed with drug
- communication plan – informing key audiences about risks
- elements to assure safe use (ETASU) – interventions to reduce risk
- implementation system – monitoring and evaluating ETASU

The drug sponsor implements some or all of these measures to educate prescribers and patients and to monitor the safety of their product. The most rigorous REMS component is ETASU, which may restrict the usage of the drug to certain patient populations or require more stringent monitoring. The incentive for REMS adoption is to enable products with questionable safety implications to be available sooner for patients who could benefit from them.

2.6.2 Label

The drug label is a critical component of the NDA, as it contains specific information about how the drug is to be used, as well as any known adverse effects. All of the label information that gets submitted to the FDA for an approved drug is provided publically on National Library of Medicine's (NLM) DailyMed website. A screenshot for the cancer drug Avastin from DailyMed is shown in Figure 22.

AVASTIN (bevacizumab) injection, solution
[Genentech, Inc.]

RxNorm Names

► [Review RxNorm Normal Forms](#)

Permanent Link:<http://dailymed.nlm.nih.gov/dailymed/lookup.cfm?setid=939b5d1f9fb2-4499-80ef-0607aa6b114e>

Category	DEA Schedule	Marketing Status
HUMAN PRESCRIPTION DRUG LABEL		Biologic Licensing Application

Drug Label Sections

Description	Clinical Pharmacology	Indications & Usage	Contraindications	Warnings	Precautions	Adverse Reactions
Overdosage	Dosage & Administration	How Supplied	Patient Counseling Information	Supplemental Patient Material	Boxed Warning	
Patient Package Insert	Highlights	Full Table of Contents	Medication Guide			

HIGHLIGHTS OF PRESCRIBING INFORMATION

These highlights do not include all the information needed to use AVASTIN safely and effectively. See full prescribing information for AVASTIN.

AVASTIN® (bevacizumab)
Solution for intravenous infusion
Initial U.S. Approval: 2004

WARNING: GASTROINTESTINAL PERFORATIONS, SURGERY AND WOUND HEALING COMPLICATIONS, and HEMORRHAGE

See full prescribing information for complete boxed warning.

- Gastrointestinal Perforation: Occurs in up to 2.4% of Avastin-treated patients. Discontinue Avastin for gastrointestinal perforation. (5.1)
- Surgery and Wound Healing Complications: Discontinue in patients with wound dehiscence. Discontinue at least 28 days prior to elective surgery. Do not initiate Avastin for at least 28 days after surgery and until the surgical wound is fully healed. (5.2)
- Hemorrhage: Severe or fatal hemorrhage, hemoptysis, gastrointestinal bleeding, CNS hemorrhage, and vaginal bleeding are increased in Avastin- treated patients. Do not administer Avastin to patients with serious hemorrhage or recent hemoptysis. (5.3)

Figure 22: Screenshot of Avastin Label from DailyMed

(NLM, 2012)

The manufacturer, or drug sponsor, uses the data collected through the development process to create the label. The FDA must approve the label contents, as this is the most widely used dataset describing the drug that is available to the public. The contents of the label include:

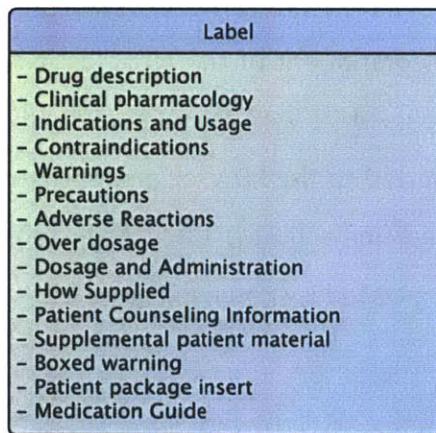


Figure 23: Data Contents in a Drug Label

(NLM, 2012)

During clinical trials and interactions with patients, manufacturers collect patient reported data that is used to prove the labeling claims. Patient reported outcomes (PROs) are patient psychometrics collected on questionnaires during clinical trials. The subjective data pertains to symptoms, health status, general health perceptions, and health related quality of life (HRQoL). These data points are direct patient reported measurements of physical, psychological, and social observations and are used to supplement the other data collected during a clinical study.

2.6.3 Adverse Events

During clinical trials, several pieces of data are collected when an adverse event is experienced including duration, severity, laboratory abnormalities, action taken, and outcome. In order to be able to record adverse events after a product is commercialized, the FDA has a passive adverse event reporting system, FAERS. Adverse event reporting is voluntary and open to anyone through the FDA's MedWatch website. Clinical reviewers at the FDA will regularly scan the database and assess for new safety concerns, which may result in regulatory action including updating the label, restricting drug use, or recalling the product. Adverse event reporting is an important feedback mechanism, although the quality and volume of the data does not make it a sufficient safety monitoring system (Heinrich, 2000).

2.6.4 FDA Data Resources

The FDA, and other regulatory bodies, maintain a lot of data about medical products and make several resources available to the public. A typical drug profile on the FDA website contains the following pieces of data.

DrugAdmin	DrugChemical
DrugName	ActiveIngredient
DrugCode	InactiveIngredient
FDAApplicationNumber	ChemicalType
Manufacturer	Strength
DrugShortages	DosageForm/Route
ApprovalHistory	Tablet
ApprovalStatus	Oral
ReviewClassification	Injection
Letters	TherapeuticEquivalent
Reviews	
Statements	
DrugSafety	DrugClinical
DrugRecalls	ClinicalTrialResult
MedicationErrors	IndicationOfUse
DrugSafety	MedicationGuides
DrugAlerts	MarketingStatus
	OverTheCounter
	Prescription

Figure 24: Data Elements for Drug Available from FDA

Additionally, the FDA maintains several databases on approved drugs and corresponding safety. These data sources are maintained and useful in capturing detailed information on drugs and other medical products.

Drug Approvals and Databases	Drug Safety and Availability
Adverse Event Reporting System (AERS)	Drug Alerts and Statements
Approved Drug Products with Therapeutic Equivalence Evaluations (Orange Book)	Importing Prescription Drugs
Bioresearch Monitoring Information System (BMIS)	Medication Guides
Clinical Investigator Inspection List (CLIIIL)	Drug Safety Communications
Dissolution Methods Database	Drug Shortages
Drug Establishments Current Registration Site	Postmarket Drug Safety Information for Patients and Providers
Drugs@FDA Database	Information by Drug Class
Inactive Ingredient Search for Approved Drug Products: Frequently Asked Questions	Medication Errors
National Drug Code Directory	FDA Drug Safety Newsletter
Postmarket Requirements and Commitments	Drug Safety Podcasts
Approved Drugs	Safe Use Initiative
	Drug Recalls
	Drug Integrity and Supply Chain Security
	Multistate outbreak of fungal meningitis and other infections

Figure 25: List of Resources Available on FDA Website

Regulators must coordinate closely with manufacturers to ensure that the data published accurately characterizes the medical product and is understandable to providers and patients. Regulators rely on the data supplied by the manufacturers to make crucial decisions about product licensing. The data flows from the patients, to the manufacturer, to the regulator, and finally to the public. Along the way, the data is summarized and reformatted, based on the manufacture selection process and guidelines enforced by the regulators.

2.7 Distributor Data

Distributors source data from the manufacturers and regulators, and provide data to the payors for product reimbursement. Before distributors, such as PBMs, can offer a medication to patients, pharmacies or hospitals, the product needs to be approved and added to their formulary. The PBM will then negotiate the price with the manufacturer and also coordinate fees with the different distribution channels. The PBM network in Figure 26 illustrates the product flow through various distribution channels and corresponding payment considerations.

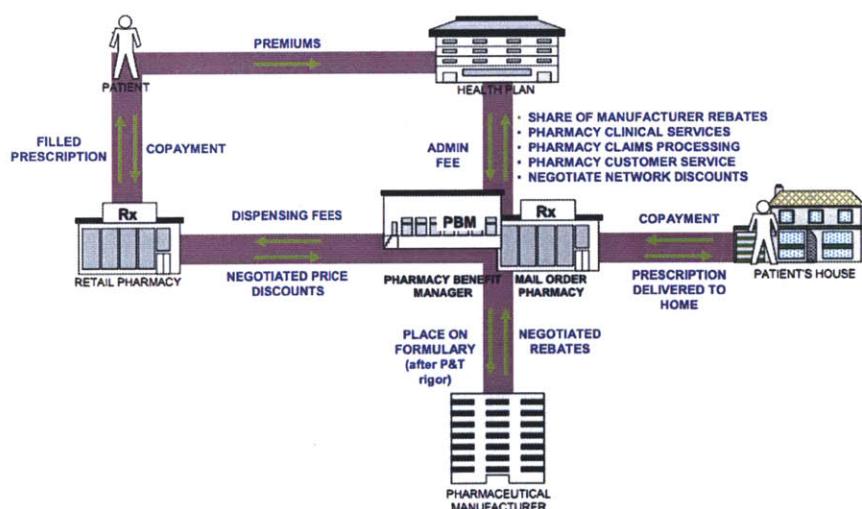


Figure 26: PBM Network

(Andrew, 2009)

The drug pricing data is a driver for the PBM network, as drugs will vary in price depending on the payor and type of distributor – mail order, retail pharmacy, etc. As well, drugs are typically categorized into three tiers, defining the patient copay range, with the most expensive drugs being tier three. These are typically non-formulary brands, while tier 1 is generics, and tier 2 is brand drugs on the formulary. Other pharmacy administrative data is listed in the second column of Figure 27.

PatientAdminInfo	PharmacyAdmin
SocialSecurityNumber	NationalDrugCode
DriverLicenceNumber	Formulary
MemberIdentification	Tier1
AccountNumber	Tier2
ContactInfo	Tier3
Name	Prices
Address	CoPayment
Telephone	AdministrationFee
Demographics	ManufacturerRebate
DOB	FlatDiscounts
Gender	PerformanceDiscounts
ZipCode	CombinationDiscounts
Race	DispensingFee
Ethnicity	AverageWholesalePrice
PaymentMethod	MaximumAllowableCost
BankInfo	UsualCustomaryPrice
CreditCardInfo	PharmacyInfo
Authorization	MailOrder
Signature	RetailPharmacy
PharmacyClaim	Specialty
DrugInfo	DrugInfoResources
DrugCode (NDC)	Primary
DrugName	PharmacyJournals
PrescriptionInfo	MedicationGuides
Number	DrugLabel
Date	Secondary
New/Refill	Medline
DosePerDay	PubMed
DaysSupply	Cochrane
Quantity	InternationalPharmaceuticalAbstracts
Strength	Tertiary
Generic	DrugFactsComparisons
FormularyCode	AmericanHospitalFormularyService
RouteOfAdministration	USPDispensingInformation
NatureofIllness	DrugInformationHandbook
NatureOfInjury	RedBook
PharmacyInfo	Micromedex
PharmacyName	ClinicalPharmacology
PharmacyIdentification	Lexi-Comp
PharmacyAddress	Medscape
PrescribingPhysicianInfo	
PurchaseDate	
Costs	
ChargeAmount	
CoPay	
Reimbursement	

Figure 27: Categorization of Distributor Data

(Ambizas, Ezzo, & Patel, 2009)

Every drug approved in the United States by the FDA has a unique ten-digit drug code, or national drug code (NDC). The makeup of the number identifies the labeler, product, and package information for a specific drug. The labeler is the manufacturer or distributor of the drug; the product code identifies the strength and dosage; and the package code signifies the package form and size (FDA, 2012d). NDCs can be queried on the FDA website, under the national drug code directory.

Payors contract with PBMs for prescription services. The distributor therefore initiates all the data within the pharmacy claim. As described in the payor section, this data includes the drug or product type, prescription number, dosage, and price, as well as patient identifiable information. The distributor also has the patient's contact information, account numbers, and credit card information used for purchases. PBMs track patients' entire prescription histories; recording prescriptions filled from various prescribers and picked up at different distributors (polypharmacy). This data is helpful in indicating drug interaction oversights or prescription drug abuse. The refill record also gives clues to medication adherence, and PBMs are now using this data to predict patient compliance and behavior.

The fourth data category of distributor data is the drug information resources. Distributors are often in direct contact with patients and need to supply medication guides and other information about the drug. Medication guides are included with several prescription drugs and must be distributed to the patient, when required by the FDA. These guides inform the patient of the instructions for use and describe the possible side effects and safety considerations. Distributors also maintain data about the drugs they distribute, and source this from released information from the manufacturer and journal published clinical trials. Additionally, there are many online resources for drug information that pharmacists use, including the United States National Library of Medicine's MEDLINE/PubMed, Redbook, and the FDA.

Similar to medical claims data, PBM data is also being repurposed, again due to the volume and availability. Several companies are using pharmacy data to gain insight into provider prescribing habits, as well as drug utilization and trends. These types of reports are sold to pharmaceutical companies for strategic marketing purposes. As well, insurers use this data during the underwriting process to verify a patient's application and eligibility for a plan based on their prescription history. Insurers can also identify risk using the types of medications recorded. Milliman and Optum Insight are two companies that buy PBM data that includes patient specific information, and sell the reports to payors for underwriting.

As previously discussed, pharmacy data is also being used to make predictions about patient behavior and clinical outcomes - identifying overuse, underuse, drug safety, and medication gaps. The intent for these analytic measures is to reduce cost and improve outcomes through clinical recommendations.

Similar to the ACPD, the prescription drug monitoring program (PDMP) is a statewide, government-administered electronic database that collects specific information about prescriptions filled for certain types of pharmaceutical drugs (Privacy Rights Clearinghouse, 2012). In this way, prescription drug usage for an entire population can be analyzed. The data collected by the PDMP is currently being used to monitor drug diversion and abuse for controlled substances. Depending on the state legislation, this data can be transferred to law enforcement for surveillance.

The data collected and managed by the distributor spans across several stakeholders. As manufacturers release drug specific information, distributors assess the clinical use of the drug and determine whether it should be added to their formulary. This data is then passed along to other distributors, payors, providers, and patients. The distributor is the intermediary between the product and patient, so the quality of the transferred information greatly impacts the patient experience. Finally, the wealth of

patient pharmacy data has led to a greater understanding of medication usage, which is currently being used for passive monitoring of prescription drugs.

2.8 Data Landscape Summary

The exploration of data from the lens of each stakeholder presents a view of the health system with over 500 different kinds of data. This data is scattered throughout the subsystems and the meaning varies depending on the beholder's perspective. For instance, data from the payor point of view is claims based, relying on standard codes to describe clinical status. Whereas data from the manufacturer point of view is product specific and needs to characterize disease therapy and report actual clinical effects. As a result, there is no single way to define data in health.

Through discussion of each stakeholder, we begin to examine how the data flows through the system. The transactions that occur between the stakeholders define the direction of flow. Some of the data transfers among subsystems, but is fragmented or represented in different forms. Other data remains within its silo and is used only by its originator. We also observed that when an abundant of data is available in a standard format, it is being repurposed. Both medical and pharmacy claims data are being used for different purposes than originally intended. Data in the health system is dynamic and the response of each stakeholder greatly impacts how it is utilized.

The data defined in this chapter is used for subsequent analyses in this thesis. Chapter 3 explores alternate methods to organize the data in order to further the evaluation and gain a deeper perspective of data in the health system.

CHAPTER 3: ANALYSIS OF HEALTH DATA

In chapter 2 we present the data landscape through the lens the stakeholders. The stakeholders provide a convenient means to identify data, but fail to inform the system perspective for data. In this chapter, we view data beyond its stakeholder origin because we believe in order to understand the system we need to examine data as a whole. We take a three-tiered approach to deconstruct the system, looking at communication flows, defining data dependencies, and assessing system flexibility through data attributes.

Our analysis reveals that there is no single perspective to understand data. The limitations we observe suggest that the best approach to view the system is with a combination of methods – resulting in the patient interaction data flow diagram discussed in chapter 4.

Table 4 summarizes the main methods and tools we used, how we used them, and the observed limitations. We review some of these methods in the rest of the chapter.

Method / Tool	How we use it	Observed limitations
Adjacency matrix / Excel, Network diagram / yEd	illustrate kinds of data that transfer between stakeholders and direction of communication	does not show sequence of transactions, represents only high level communications, does not define data usage
Dependency analysis / Design structure matrix (DSM)	define upstream/downstream influences and dependency clusters	significance of indirect influences, complex relationships prevent precise dependency definition
Attribute analysis / Visual understanding environment (VUE)	link data to characteristics, visualize unconventional clusters	lack sense of data usage and flow
Value assessment / Excel	define purpose, value, and use of data to determine gaps	relationships across data not captured
Data flow diagramming (DFD) / yEd, ch5	map data flow during a patient's interaction with the health system	represents general data categories, individual data types are implicit
Object-process modeling (OPM) / yEd, ch7	verify system properties that result from DFD analysis	not intuitive to derive, need to translate from health terminology

Table 4: Assessment of Methods and Tools Used to Analyze Data

3.1 Cross-Stakeholder Communication

As a first level of analysis, we start by looking at the kinds of data that transfer between stakeholders. This defines the first order, or highest level, of communication. Figure 28 shows, as an adjacency matrix, the data flow from the stakeholders in the column to the stakeholders in the row. For instance, the manufacturer communicates product information to the regulator during the commercialization process. The kinds of communications are categorized by type, indicated by the different colored cells: raw, informative, standardized, and inferred.

	PATIENT	PROVIDER	PAYOR	MANUFACTURER	REGULATOR	DISTRIBUTOR
PATIENT		symptoms	enrollment	adverse events	adverse events	account, financial
PROVIDER	treatment		claims	adverse events	adverse events	prescribing
PAYOR	benefits	coverage				coverage
MANUFACTURER	label	label	price, label		product	price, label
REGULATOR	safety	safety	safety	licensing		safety
DISTRIBUTOR	product	product	claims	usage, payment		

LEGEND: Types of Communication

Raw
Informative
Standardized
Inferred

Figure 28: Adjacency Matrix for Cross-Stakeholder Communication

The types of communications are categorized in the legend. Raw data is communicated directly from its source; informative provides description, explanation, or education; standardized indicates policy or regulation; and inferred data is used for secondary communications.

Adverse events are differentiated, as the communication is event dependent, and not considered standard flow.

To ease interpretation of data flows, the adjacency matrix may also be represented as a network diagram (Figure 29).

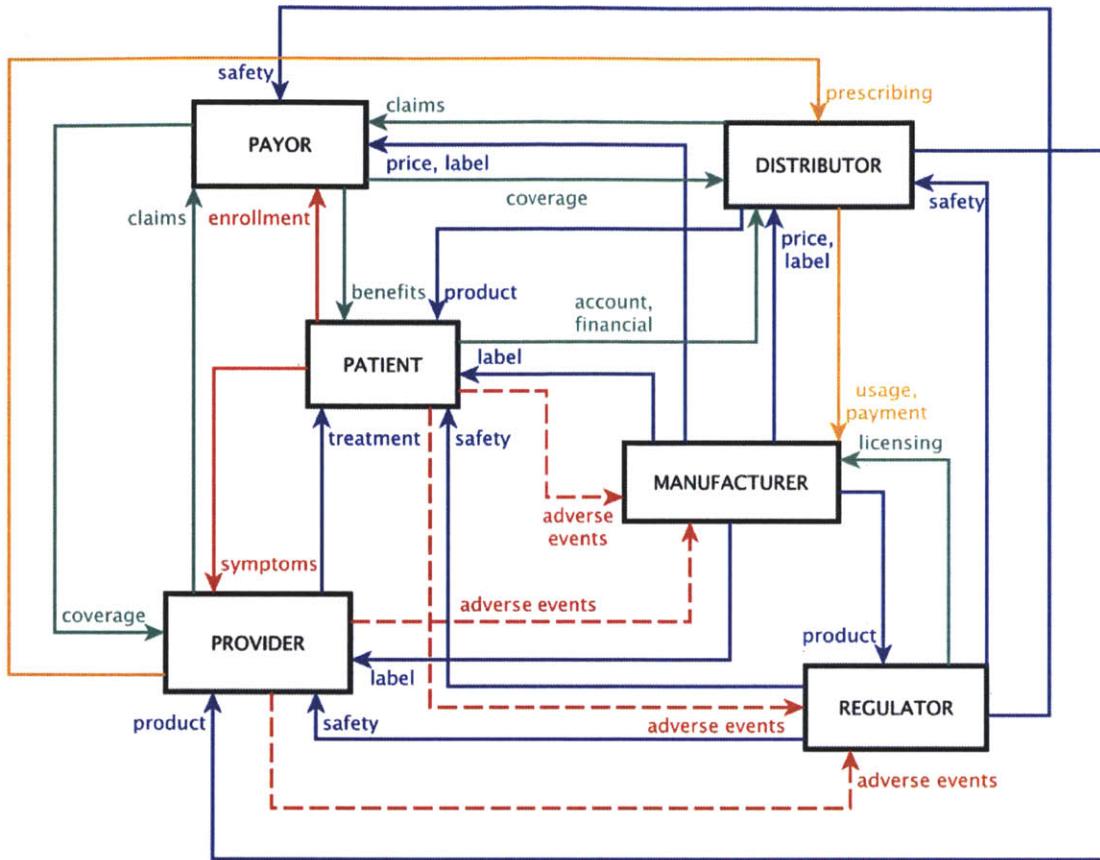


Figure 29: Network Diagram of Data Flow between Stakeholders

Reference legend in Figure 28. This diagram shows the kinds of communication and the direction of flow. As illustrated by the red lines, the raw data derives from the patient correspondence with the provider and payor. The dashed lines of raw data indicate intermittent transfer, and based on the reported adverse events rate, this occurrence is quite low. The blue lines indicate the informative data transfer given by the provider to the patient as treatment recommendations or through the product and label data transfer by the manufacturer. The green lines indicate the patient specific data that is standardized into a claim or coverage plan. These pieces of data have partial information about the patient but the transfer between stakeholders is automated, or follows an existing protocol. The yellow lines indicate inferred data, or data that flows through a secondary function. For example, the distributor gleans provider-prescribing data from the prescription record, and can analyze patterns for monitoring. This data may also get transferred to the manufacturers and used as marketing input, as previously discussed in section 2.7.

The adjacency matrix and corresponding network diagram help us to evaluate the stakeholder data defined in chapter 2. A takeaway from this analysis is that the patient

and manufacturer are the main sources of data and most of the other communications are data transfers downstream. Referring to Figure 29, the red and blue arrows indicate these two types of upstream communications, while the green and yellow arrows indicate the downstream communications. Another takeaway is the instances where limited or no communication transpires between stakeholders, as with payor to manufacturer and regulator.

These adjacency matrix and network diagram are useful in understanding the highest level of communication between the stakeholders. We define the kinds of data used for communication and the direction of flow, and further categorize communications by type. This representation does not yet show the sequence of transactions or data usage. We address these limitations in the following sections.

3.2 Dependencies

We use DSM to reveal the dependencies amongst the 17 first order communications, selected for the adjacency matrix in section 3.1. DSM is a structured method to evaluate and visualize dependencies. We plot the selected data onto the DSM to determine the dependency of each communication to the next. At this stage of the analysis, only the direct dependencies are considered and indicated along the vertical column for each kind of data – the data in the rows are inputs to the data in the columns. Partitioning the matrix sorts the rows and columns by dependency clusters (Figure 30).

PARTITIONED DSM		product	licensing	safety	label	treatment	claims	benefits	coverage	prescribing	account	product price	payment for product	product usage	adverse events	enrollment	symptoms	financial
		1	2	3	4	8	11	6	9	10	12	14	15	16	17	5	7	13
product	1	1	1	1	1	1						1			1			
licensing	2		2				1	1					1					
safety	3	1	1	3	1	1									1			
label	4	1	1		4	1									1		1	
treatment	8					8	1								1			1
claims	11						11	1	1	1				1				
benefits	6						1	6	1			1						
coverage	9						1	9		1								
prescribing	10								10			1						
account	12						1			12								
product price	14					1		1	1		14	1						
payment for product	15											15	1					
product usage	16										1	1	16					
adverse events	17		1	1	1	1								17				
enrollment	5							1			1				5			
symptoms	7						1	1						1		7		
financial	13							1	1			1					13	

Figure 30: Partitioned DSM of First Order Stakeholder Communication Data

The application of DSM to the kinds of communications between stakeholders reveals three dependency clusters. The first cluster groups data produced during product licensing that directly influences treatment decisions and adverse events. The other two clusters are administrative including claims, benefits, and product price data.

Treatment data is the most dependent data type with nine inputs, including product information, patient benefits, and reported symptoms. Adverse event data is next with five inputs. Both treatment and adverse events share dependencies on product, safety, and label, and system data.

The three empty columns on the right – enrollment, symptoms, and financial – are data sourced directly from the patient and therefore have no input from the other stakeholders. These components of patient data are used to inform six other data types (bottom three rows) indicating the wide usage of patient data.

Moving from communications across stakeholders, we go one level deeper to examine dependencies at the level of content, not just at the level of communication. We consider the main kinds of data managed by each stakeholder defined in chapter 2. Figure 31 shows the selection of contextual data that we use in our expanded DSM. This data matches the subcategories of data from the figures and tables in chapter 2.

PATIENT	PROVIDER	PAYOR	DISTRIBUTOR
PatientDOB	MedicalRecord	InsuranceCoveragePlan	PharmacyClaim
PatientDemographics	ExamReport	CostOfCoverage	DrugPrice
Socioeconomic	ClinicalNotes	PatientCoPay	PrescriptionNumber
MedicalHistory	Lab/ScanReport	ProviderIdentification	PharmacyIdentification
Symptoms	Diagnosis	InteractionRecord	PurchaseDate
PhysiologicalMeasurements	Treatment	ServiceDate	PatientFinancialInfo
OTCMedications	Prescription	Diagnosis/ProcedureCode	ClinicalStudies
GeneticProfile	Procedure	CostOfService/Procedure	MedicationGuides
OnlineResources	MedicalClaim	PaymentRecord	
SocialNetwork	Referral		
MANUFACTURER			REGULATOR
Microbiology	PatientMedicalStatus		Benefit/Risk
CMC	Pharmacokinetics		DrugChemistry
DrugPharmacology	PROs		Label
ADME	Statistical		ApprovalHistory
Toxicology	Safety		REMS
Cost	Effectiveness		AdverseEvents
ClinicalTrialProtocol	SideEffects		Recalls
PatientDemographic (research)			
PatientDOB (research)			

Figure 31: 60 Selected Subcategories of Data from Chapter 2

(Data sourced from Figure 6, Figure 7, Figure 13, Figure 27, and Table 1)

In assigning dependencies across the data subcategories, we rely on the description in chapter 2 to resolve any ambiguities about the nature of the data or meaning. Additionally, we only consider the dependencies that reflect mainstream usage. We also decided to differentiate between direct, indirect, and possible future dependencies, as shown in the DSM in Figure 32.

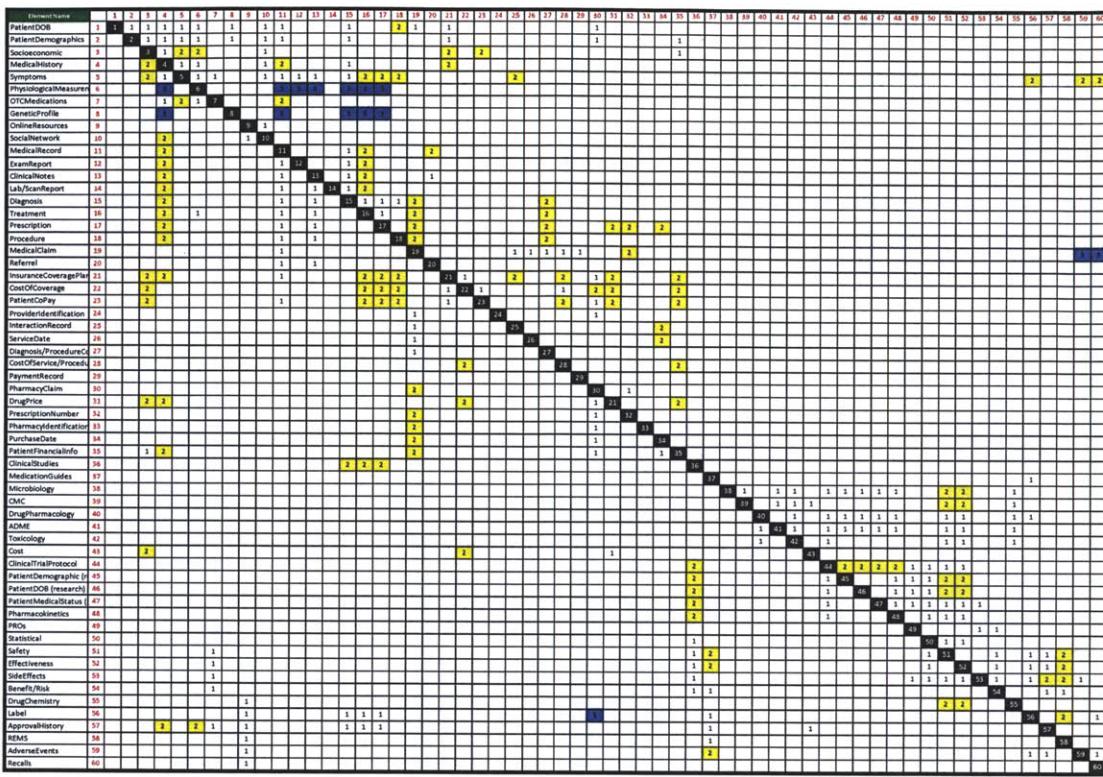


Figure 32: Non-partitioned DSM of 60 Data Subcategories

Legend:

1 = direct dependency

2 (yellow) = indirect/partial dependency

3 (blue) = possible future direct dependency

Figure 32 shows the adjacency matrix before application of the partitioning. We show this figure to highlight some of the non-direct types of dependencies – indirect, partial, and possible future dependencies that indicate subtleties in defining the dependencies. There are several indirect influences: cost of coverage is dependent on a patient's coverage plan and is indirectly influenced by the cost of services and drug price. As well, published studies and journal articles indirectly influence treatment and prescription selection. We considered several more relationships when assigning the dependencies.

Figure 33 shows excerpts from the partitioned DSM expressing only the direct correlations, after removing the level two and three dependencies. Refer to Appendix C for the full plot of the DSM.

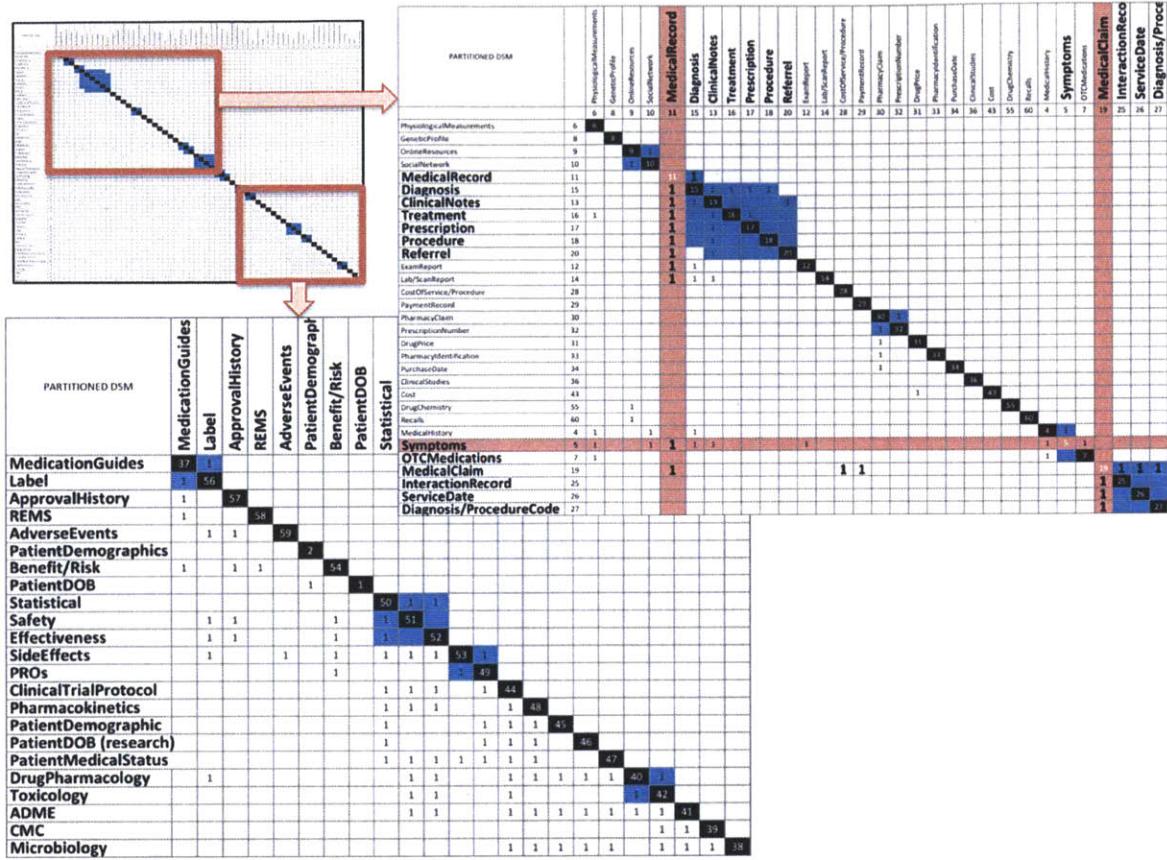


Figure 33: Partitioned DSM of 60 Data Subcategories, Direct Dependencies

The expanded DSM confirms our observations made with the dependency analysis of the communications between stakeholders (Figure 30) and adds further detail. For instance, although there are several data inputs into the medical record, its influence downstream is limited. The medical record has only one downstream direct influence (diagnosis), but has 14 dependencies. This suggests that when the patient symptoms and other 13 data types are recorded in the medical record, the flow stops – there is no other data category that requires input from the medical record. Consequently, the medical claim is dependent on the diagnosis and procedure codes inputted by the provider, but not dependent on the medical record. This suggests payor inference of clinical status using claims (section 2.4).

The cluster on the bottom right corner shows interdependencies with drug development and regulator data. Although the approval history, drug label, and adverse events all have downstream influences to diagnosis and treatment decisions, the bulk of the data does not correlate to the care system, even though similar data types (patient data collected during clinical trials) are shared between the two segments. We interpret this lack of dependency a result of two subsystems – one for care and one for product development – that are loosely connected. The impact is two disparate systems that have no standard flow for communication or feedback.

DSM is valuable for assessing dependencies on a subset of data and identifying clusters that suggest tight correlation, although it has shortcomings. The challenge of preciseness of the dependency definitions is a result of the faint relationships between discrete data types and abundance of indirect and partial correlations.

We now understand communication flow and data dependencies. In the following sections we look more closely at the data characteristics and usage in the system.

3.3 Attributes

We categorize data by its attributes to assess the nature of the data and relate kinds of data that are otherwise not relatable. Attributes tells us the dynamics of the data, availability, and constraints imposed on the data, all of which affect how the data is used – or can be used – in the health system. Categorizing data by its attributes delivers a descriptive assessment of the data as a whole and reveals unconventional clusters. In this section, we explore alternate uses of the data, outside of its current domain of use. We select a collection of non-related data from the various stakeholder categories in chapter 2. Based on the characteristics of the data, the following attributes are defined:

ATTRIBUTE	DESCRIPTION
static	not changing, always true
real time	immediately viewable dynamic data
controlled	hypothesis driven protocol
uncontrolled	not following standard protocol, random, real world
open	publically available, online
closed	private, protected under HIPAA
documented	standard within the health system
undocumented	outside of the health system
preventative	informative health information to protect and prevent
chronic management	repeated data for same condition
patient independent	general health and disease information
patient dependent	patient specific data

Table 5: Sample Attribute Categories

We link each kind of data to its corresponding attributes. The attributes are shown as ellipses on the map in Figure 34 and attribute pairs are shown in the same color. The rest of the nodes are the different kinds of data, drawn from the stakeholder sections in chapter 2.

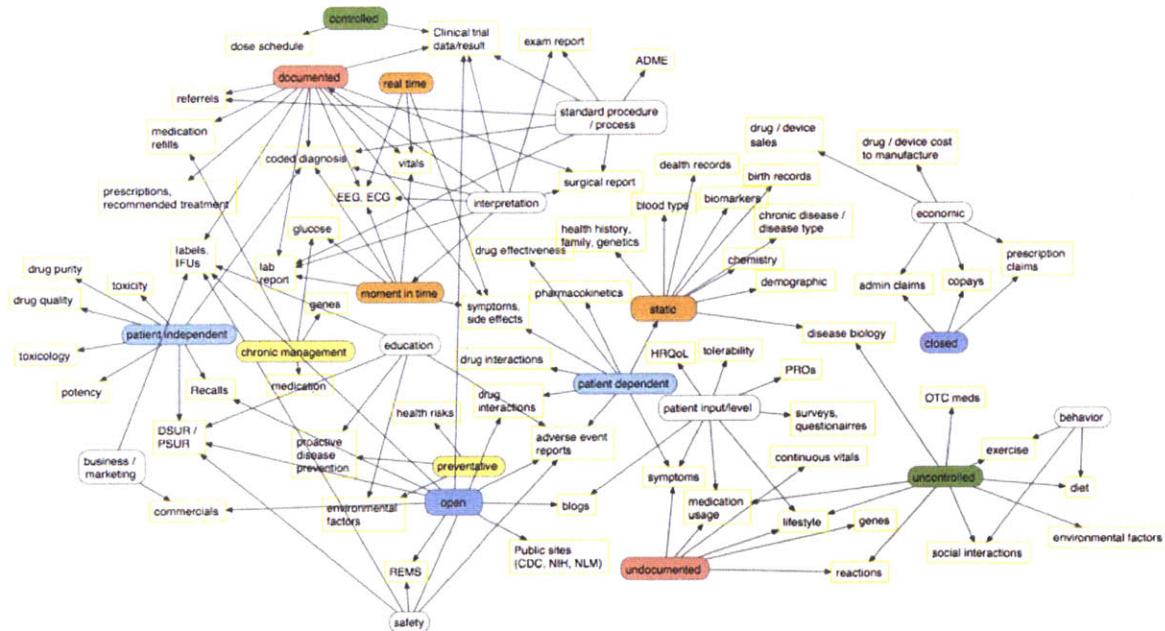


Figure 34: Data Attribute Map

This analysis is useful in identifying connections between otherwise unrelated kinds of data. For instance, adverse events caused by medical products and patient blogs are

public forms of data. Although it has not been conventional to relate these types of data, adverse event data could be gleaned off of patient blogs and other peer-to-peer health discussions, presenting an opportunity to use social media to supplement safety reporting systems (FAERS). AdverseEvents, Inc. is an online service that collates adverse event data and claims that actively monitoring social media and other online outlets can portray a more accurate view of drug safety (Overstreet, 2012). Even though this correlation does not originate from an existing stakeholder, it presents an opportunity to use input collected online to inform product safety.

There are several other examples of new correlations that we make from the attribute map that present opportunities for data usage beyond its original domain. For instance, some of the kinds of data that could inform care decisions are undocumented and uncontrolled, meaning use in the care system is not well defined. These include behavior, symptoms, vital trends, and medication usage. Collection of this data resides outside of the traditional clinical environment, and therefore are not typically documented or used for decision-making. Various trends in wearable monitors and device technologies make data collection at the patient feasible and are discussed in section 5.1.

The open, undocumented, and uncontrolled attributes indicate that there are clusters of data that are not formally tracked in the health system, resulting in fragmented data usage. Some of the kinds of data that share these attributes are represented in the lower half of the map in Figure 34. The system concept that is presented in chapter 6 considers the entry of these kinds of data into the health system.

3.4 Value Assessment

We determine the value of the data by assessing its purpose and how it is currently being used in the health system. We compare the value we believe the data to hold, based on stakeholder interviews and analysis of the literature, to how the data is

currently being used. The gaps observed indicate that the data has more potential than what it is currently being used for, presenting opportunities for improvement.

We evolve the stakeholder categories into seven, more functional, categories, shown in the first horizontal row in Table 6. For each category we list representative examples of data from chapter 2, describe the purpose of the data, value of the data, and how the data is currently being used.

						CATEGORY of DATA							
							1	2	3	4	5	6	7
A EXAMPLES	Science Pre-Clinical	Clinical Research	Product Treatment, Drug, Therapy, Device	Clinical Reports/Records	Administrative	Real World Clinical	Patient Owned						
	chemistry	clinical trials	lab data	subjective notes	record of encounter	epidemiology studies	behavioral data						
	disease biology	protocol design	structured behavior	symptom	time sequence	patient registries	perceived symptoms						
	pharmacology	demographics	label	physiological metrics	procedure codes	surveys	physiological metrics						
	ADME	dose schedule	instructions	lab results	diagnosis codes	surveillance	genetic						
B PURPOSE of DATA	toxicology	statistics	interactions	scan results	treatment codes	safety reports	online resources						
		outcomes	adverse events	medications	pharmacy records		social communities						
		behavioral data	side effects	treatments	cost / payment	disease analysis							
		safety, efficacy,	manage/inform	medical record	financial transactions								
		effectiveness of	treatment post	keeping per patient			patient management of own health						
C VALUE of DATA													
	compound investigation	measure of benefit to risk	safety warnings & contraindications			population health management							
	human health	primary: patients with disease	guides treatment			view of disease population							
		secondary: manufacturer ROI	informs safety			real world' outcome, feedback							
	disease treatment	product licensing	decide treatment			statistical reports	patient self-knowledge support networks						
D HOW IS DATA													
	drug discovery	benefit-risk analysis	may bias diagnosis				patient knowledge						
E WHAT IS MISSING? C - D													
	chemistry/genomics of target patients	patient level analysis, feedback	patient level analysis	relationship to other patients									

Table 6: Data Value Analysis

The final row in Table 6 indicates the missing information when the value of data is compared to how data is being used today. We analyze the missing information and formulate our results in Table 7. This gap assessment is an important takeaway from the value analysis as it presents the specific losses of data and their impact on the current system.

Category of Data	Gap Assessment and Impact
Science	Limited individual patient (chemistry, genetic) data in the current drug discovery process may result in ineffective products for the target population.
Clinical Research	Outcomes from today's clinical trials only inform benefit-risk for the pre-selected patient group, limiting learning for other patient types.
Products	Available products may bias diagnosis, even with no knowledge of an individual's response to a selected therapy.
Clinical Record	Today's medical records are limited to individual patient management, as correlation among records does not exist.
Administrative	Administrative records (claims) are unidirectional with limited outcome data; therefore valid cost-benefit cannot be assessed.
Real World Clinical	Data collected from today's real world studies/events has limited use: assumptions made on general population, unknown denominator, and questionable reliability.
Patient	Data sourced from the patient is subjective, often unreliable, and has vast entry possibilities into the health system, making it difficult to compare among patients.

Table 7: Results of Data Value Analysis

Several of these hypotheses are referred to in other sections in this thesis, but here are specifically presented as the result of our value analysis.

Another way to analyze the value of data is to assess the comparability and usefulness in providing feedback to the system. Table 8 looks at the same categories of data from Table 6 and assesses whether the data is specific to the patient, can be compared, and whether it can be used to provide feedback.

Category of Data	Is the data patient specific?	Can different data sets be easily compared?	Can this data be used to introduce feedback?
Science	no	no	no
Clinical Research	no	yes	no
Products	no	yes	no
Clinical Record	yes	no	no
Administrative	no	yes	no
Real World Clinical	no	no	yes
Patient	yes	no	yes

Table 8: Qualitative Assessment of Function for Selected Data Categories

First, we observe that the data that is specific to the patient is difficult it is to compare. This implies that learning from one patient to another is limited and not a function of the current EHR system. Second, the data that already has standard transfer paths and protocol is most comparable, namely clinical research, product data, and administrative data. Claims data, for instance, has nationally approved formats using sets of coding standards, making it ideal for comparing (section 2.4). Thirdly, we observe that feedback is limited to data generated from patients and real world studies. Based on the results in Table 7, both of these data categories have limited value, although currently are the main source of feedback in the system.

The methods we use to assess the value of data helps to logically describe the inefficiencies in the current system. The result is a series of hypotheses that are driven by how data is being used today compared to its potential use. This analysis reveals opportunities for enhanced data use.

In this chapter we have uncovered several vantage points to understand the system and visualize data beyond its stakeholder origin. We conclude that there is not a single method that encompasses all of the insight needed to assess data as a whole. Rather, we combine elements of the different approaches to arrive at a representation that we can use to evolve our thinking of how data flows in the system. This representation is the patient interaction data flow diagram, discussed in chapter 4.

CHAPTER 4: PATIENT INTERACTION DATA FLOW

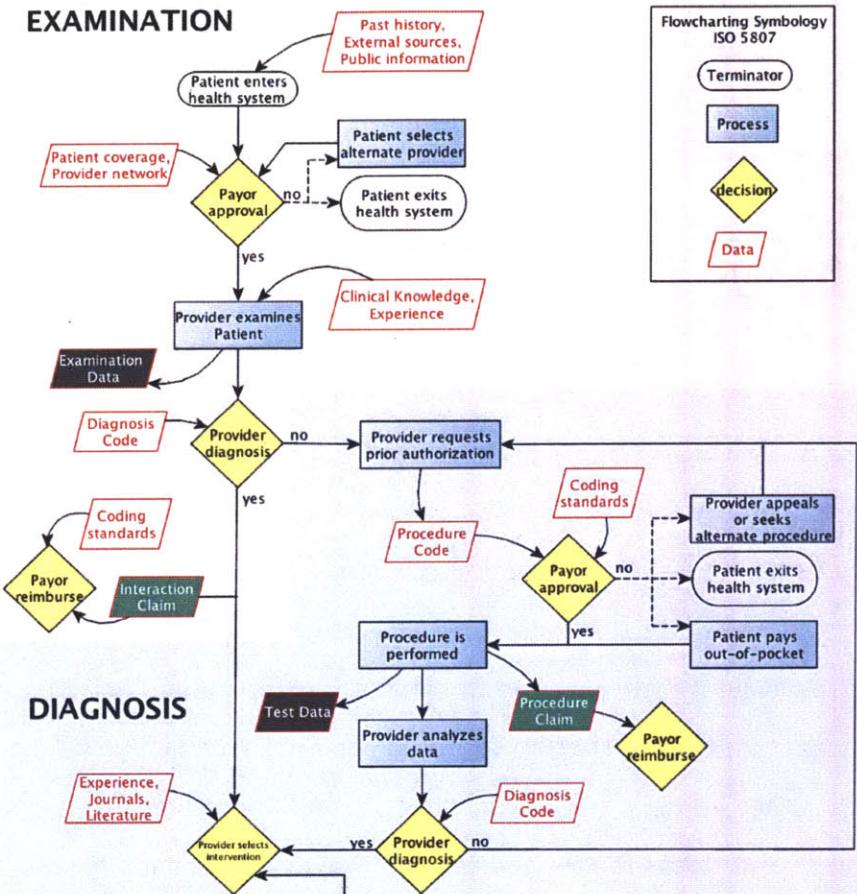
The methods used to analyze data in chapter 3 are effective in understanding and visualizing data usage from the perspective of the stakeholders that generated it; through communication, dependency, attributes, and value of their data. However, the stakeholder analysis does not inform how data is used to further the objectives of the system. In this chapter, we address that by assuming the perspective that the patient is the ultimate source of data and interpreting data flows to and from patients, as they interact with the health system.

4.1 Data Flow Diagram

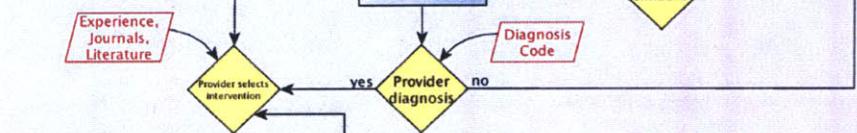
Data flow diagramming (DFD) is a method used to express the structure and flow of data. We combine our understanding of the data in the health system with the process modeling rules of DFD to create the patient interaction data flow diagram in Figure 35. The diagram defines the types of data, where the data interfaces with the system, and how the data impacts processes and decisions. Fragments of data combine to inform provider diagnoses and intervention selection, and fragments of data break away from the main flow during transactions among the system actors – during payor reimbursement and product licensing.

The patient interaction diagram provides the foundation for a systematic analysis, defining the logical data flow and also identifying instances where flow is interrupted. The data losses that occur when the flow is interrupted can be used to identify system inefficiencies and opportunities for improvement.

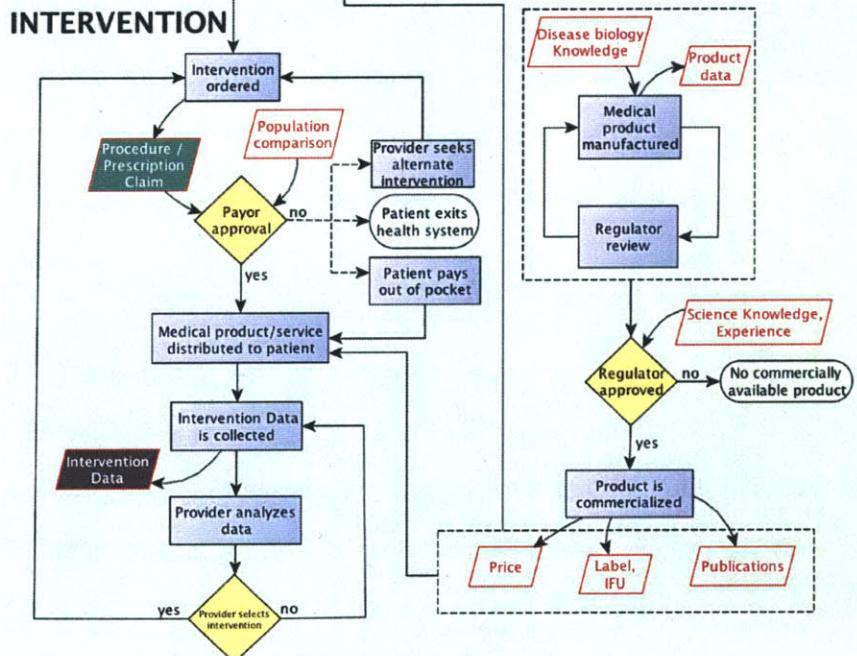
EXAMINATION



DIAGNOSIS



INTERVENTION



The diagram illustrates the three interaction phases: examination, diagnosis, and intervention. These phases align with previous observations of the health system and interviews with stakeholders. During the examination phase, the provider uses prior knowledge and experience to assess the state of the patient, along with subjective symptom descriptions and physiological measurements. When additional data points are required to make a diagnosis, a second data set is collected in the diagnosis phase. This includes data that results from diagnostic procedures, lab work, and imaging tests. The provider analyzes the data to select the intervention (procedures, surgeries, devices, drugs). The intervention phase includes a monitoring loop where the outcome of the intervention is observed. The payors, providers, and regulators drive the various processes and decisions. The patient is a passive actor, but a primary data source. The manufacturer is ancillary to the main flow, as an intervention supplier.

Figure 35: Patient Interaction Data Flow Diagram

The shaded regions in the diagram indicate the three interaction phases: examination, diagnosis, and intervention. Each phase reveals a scatter of input and output data, denoted by the red parallelograms. The following two sections describe the forms of data, and the subsequent sections discuss the impact of observed data loss, when partial data is used for decision-making.

4.1.1 Input Data

The input data feeds into the system at process and decision nodes and is used to execute processes and inform decisions. This input data is truly an assortment of data fragments; for instance, the patient's clinical history, patient provided symptom descriptions, administrative data regarding coverage policy, and the knowledge and experience of the provider all drive the examination phase. Other forms of input among the phases are non-patient specific data, such as publications, journal articles, organizational experience, and standards in science and medicine. There is no standard collection of input data for all patients and therefore the basis for decision-making varies widely from patient to patient. This poses a challenge to learning: variance across input data can be expected to translate into increased variance throughout the system all the way into the outcomes. However there are no mechanisms to capture lessons that would help reduce said variance.

4.1.2 Output Data

The data outputs are clinical or administrative. The gray-shaded data elements in the interaction diagram identify the clinical outputs at the point of care: examination data, test data, and intervention data. These are patient-specific and include physiological measurements, diagnostic results, and observations of patient state. The clinical output data are stored in a patient record and are used by providers to inform patient care and decision-making. During interactions with the same provider, the patient

record is enhanced with new forms of data, which may be further collated into an EHR (section 2.2).

Patient-specific administrative output data are derived from the contents of the clinical record. As described in section 2.2, reasons for visits, diagnoses, and interventions all have standard codes. The reimbursement claim, submitted to the payor, consists of the best available codes for each patient interaction. These claims are in each phase of the interaction diagram: interaction claim, procedure claim, and prescription claim (shaded in green).

Referring to interaction diagram, the right-hand section of the intervention phase describes the data flow during clinical research. Here, the clinical output data is driven by the trial protocol and is similar in content to the EHR, although more structured. Discussed in section 2.5, the clinical data for drug research is stored and managed in the trial CRF (case report form) that is used by the regulator during the approval process. Once a product is approved to market, the manufacturer releases product and safety information (denoted on the interaction diagram as label, and publications) that is considered an abbreviated summary of the clinical studies. This output data becomes input for provider intervention selection.

Communication between stakeholders is when flow is interrupted and data loss occurs. The following sections overview some examples of the impact of decisions based on partial information.

4.1.3 Communication Between Provider and Payor

The purpose of a claim is transactional. It does not contain clinical data or patient history, but rather is a record of an interaction. It is therefore arguable whether the claim is clinically meaningful and should be the mechanism used to determine care. For instance, payors review claims against insurance policies and standard procedures

and either approve or deny reimbursement. This is illustrated in all three phases of the interaction diagram and occurs either before (prior authorization) or after services is provided. The payor reimbursement decision has a strong impact that could result in a different medical course or early termination of care, having tremendous effect on the patient outcome.

4.1.4 Communication Between Manufacturer and Public

The product data in the interaction diagram represents the safety and efficacy data generated by medical product manufacturers during development and licensing. (Refer to section 2.5 for a complete list of product data). The purpose of this data is to satisfy regulatory requirements in order to grant approval to market, and once approved, only a subset of this data - expressed in the interaction diagram as price, label, instructions for use, and publications - is made available to the other stakeholders. Providers use this information to select interventions, payors to assess cost-benefit, and patients to research treatment options. The information only includes population statistics derived from the clinical studies which is hard to relate to the needs of a specific patient.

4.1.5 Communication Between Patient and Provider

Patients are data rich. However, the data collected during the provider interaction (examination data, test data, intervention data) is only a snapshot of the patient's clinical state, and may not always be sufficient to understand the entire clinical profile. Measurements taken at the point-of-care do not often reflect the normal patient state, nor are single data points effective to manage chronic conditions. Blood pressure, blood glucose, electroencephalography (EEG), electrocardiography (ECG), and like data can all be monitored outside the clinical environment, between patient interactions with providers, dramatically enhancing the data pool used for diagnosis and intervention.

Trends of remote patient monitoring (RPM) merge wearable sensors with wireless technology. Remotely collected data is transmitted to patient EHRs and decision support algorithms alert providers. Several examples of remote monitoring devices are listed in section 2.3.2. Other benefits to monitoring patients in their natural environment are reduced cost, reduced hospitalizations, continuous care, and patient engagement. See section 5.1.1 for a continuation of this discussion.

The entry of patient sourced data at the top of the interaction diagram indicates that the patient drives the entire flow of processes and decisions. The accuracy of diagnoses and effectiveness of interventions rely on the collection of input data. The provider's decision of what data to collect also makes a large impact on the care path. The selection of patient data that is used during patient-provider interaction – whether sourced within or outside the care system – greatly affects the clinical decisions made after the first encounter.

4.1.6 Communication Between Providers

Patient information maintained by one provider might not transfer to other providers, either through referrals or upon transitioning to different care systems. The result is a suboptimal starting point for care. Referrals occur between primary care physicians, secondary specialists, and tertiary care. (Gandhi et al., 2000) shows that communications between providers lack clarity about the reason for referral, have inadequate clinical notes, lack timeliness, and have unclear follow-up plans. This breakdown in data flow leads to poor continuity of care, delayed diagnoses, unnecessary testing, and risk of polypharmacy. In addition, complexity of health care plans and payor approval delays further strain the communication channel.

Data sharing at the interface of the phases among providers exemplifies the goals of the accountable care organization (ACO) model. ACOs claim to achieve better communication among providers by coordinating a care team network for each

patient. This collaboration is believed to enable data sharing among the providers who treat a patient.

Other instances of data flow between stakeholders are sparse, also observed in the adjacency matrix in chapter 3. A failsafe interaction that is not represented on the interaction diagram occurs during unexpected patient events. For instance, patients and providers may provide data to the manufacturer or regulator when an adverse event is experienced. Most regulatory systems have a mechanism to enable voluntary adverse event reporting (FAERS, MedWatch). This provides additional input data into the regulatory system, but does not alter the main flow. However, the value of these reporting systems are limited since less than 10 percent of adverse events get reported (Goldman, 1996).

The interaction diagram is used to show how data is being used at the point of care to treat symptoms. The decisions that are made at each node have great impact on patient outcome, but are only as good as the data that feeds into them. We observe interruptions of flow throughout all of the interaction phases and specifically among the interfaces between the stakeholders. Instances of data loss or missed opportunity for data usage are revealed as we consider the communication between stakeholders. In the next chapter, we discuss the trends that aim to improve data usage and assess their impact on data flow.

CHAPTER 5: DATA DRIVEN TRENDS

The interaction diagram we introduced in chapter 4 defines the flow of data that emerges from the patient's interaction with the health system. There we observed the natural tendency of the system design to fragment and silo data. In this chapter we focus on the solutions that are emerging to overcome difficulties associated with lack of data at different stages of the system. We inventory existing companies, initiatives and technologies that propose data empowering solutions throughout the system. When these solutions are overlaid onto the interaction diagram (Figure 36) they map to the inefficiencies identified in chapter 4. The ability for each solution to be represented by a category on the interaction diagram validates the diagram and shows a broader system problem – data driven solutions are local. That is, the space of solutions is as fragmented as the data itself. Local solutions address local inefficiencies but fall short of addressing the flow of data at the system level.

The emerging categories of data driven solutions are labeled A-I in Figure 36 and are further described in Table 9. The 9 categories arise from analysis of over 100 industry trends, of which 40 are shown in the example column of the table. Each of these examples addresses a local opportunity for improvement in data quality, clinical outcomes, health care costs, or process efficiency. Some of the trends are incremental improvements driven by the incumbents, while others are divergences propelled by new technologies.

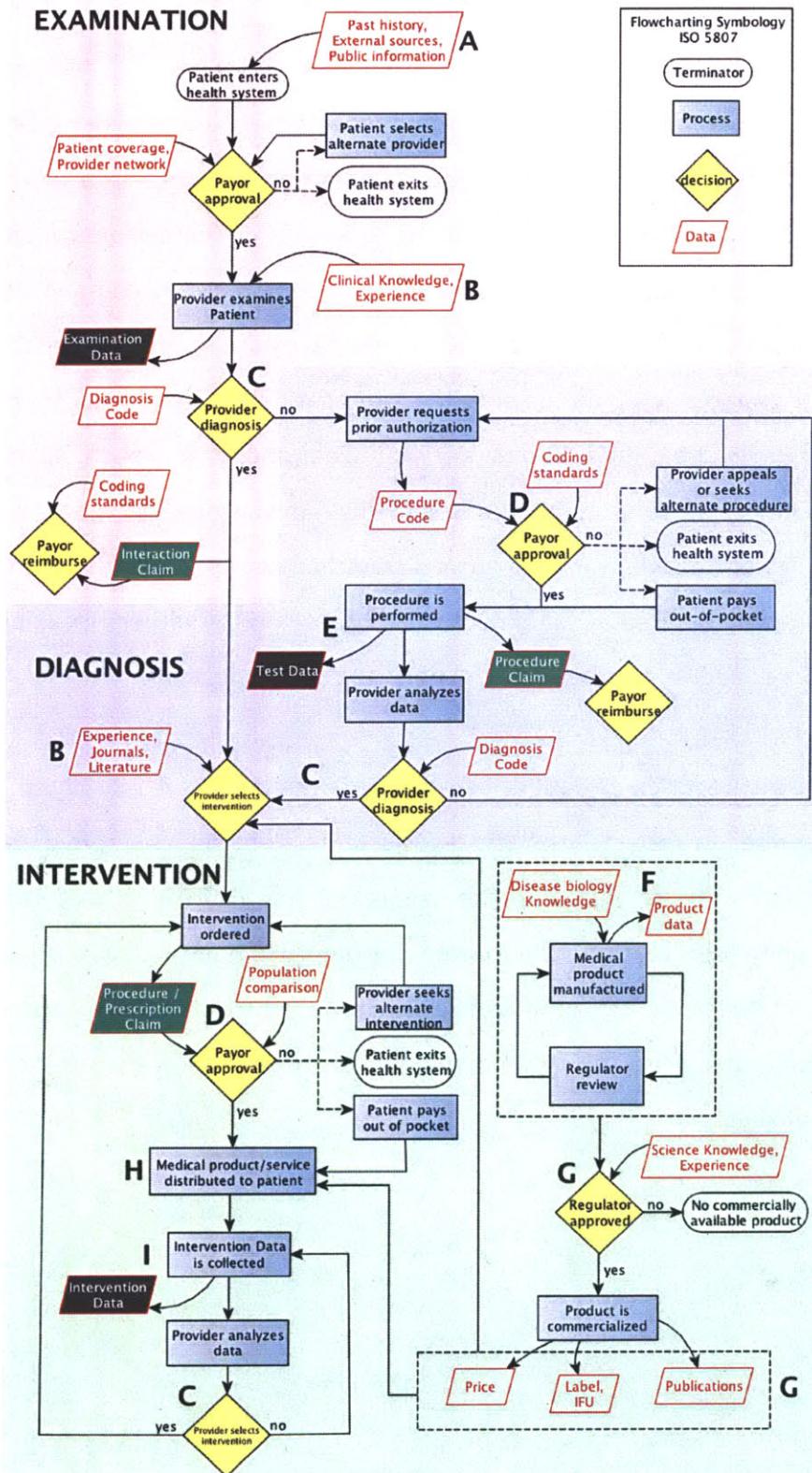
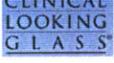
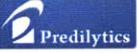


Figure 36: Data Driven Trend Categories A-I Mapped onto Interaction Diagram

	Category	Description	Examples
A	<u>System Input</u>	<ul style="list-style-type: none"> Personal monitoring devices Online health services DTC genetic testing & product advertisements Peer-to-peer healthcare 	    
B	<u>Clinical Knowledge Warehouse</u>	<ul style="list-style-type: none"> Point of care medical applications Electronic health records Clinical ontologies Provider network analytics 	   
C	<u>Clinical Decision Support</u>	<ul style="list-style-type: none"> Combine patient data with clinical knowledge to support diagnosis Evidence based risk identifier Increase provider productivity 	    
D	<u>Claim Management Analytics</u>	<ul style="list-style-type: none"> Applying analytics to claims data to reduce payor costs Identify population risks Prevent miscoding and fraud 	    
E	<u>Genetic Testing</u>	<ul style="list-style-type: none"> Use of DNA to screen for disorders and diseases Carrier screening, disease risk prediction Drug companion diagnostic, treatment effectiveness 	   
F	<u>Open Innovation Initiatives</u>	<ul style="list-style-type: none"> Drug discovery data sharing Open scientific collaboration platforms Transparency in clinical research 	   
G	<u>Product Licensing</u>	<ul style="list-style-type: none"> Innovating the drug approval process Staggered / adaptive licensing Pharmacovigilance considerations 	   
		•	

H	<u>Distribution Management & Analytics</u>	<ul style="list-style-type: none"> • Applying analytics to pharmacy claims to identify prescribing trends • Genetic Benefits Manager • Intelligent dispensing systems 	 EXPRESS SCRIPTS [®]  CAREMARK  
I	<u>Localized Feedback</u>	<ul style="list-style-type: none"> • Retrieving post-intervention data • Monitoring benefits, risk, and compliance • Reporting adverse events 	    

Table 9: Description and Examples of Data Driven Industry Trends

Discussion of Data Driven Trends

5.1 System Input (A)

Patients now have the ability to record their vital signs, discover their DNA, and research their symptoms before ever stepping into a physician's office (section 2.3). This information is indirectly entering the clinical environment through a narrow interface that relies on the discretion of each patient and provider. The result is a breadth of new data that does not yet have a standard entry path into the system or developed protocols for interpretation for clinical decision-making.

5.1.1 Wearable Sensors

The wearable sensing device market is forecasted to be a \$6 billion industry by 2016 (Ahadome, 2012), indicating the need to measure physiological signals outside of the care system. There are over 20 physiological measurements that can be recorded using a wearable device (Table 1) and the data is used to track health indicating metrics. Figure 37 shows examples of some of the companies that market these devices.



Figure 37: Vendor Landscape for Wearable Technology

(Ahadome, 2012)

The majority of these wearable sensors is available direct to consumer and has a broad range of applications, from weight loss to sleep management. Companies that intend for the data retrieved off of their devices to be used to inform clinical decisions must obtain medical device clearance (through the FDA in the United States). The regulatory requirements force the suppliers to narrow the domain of application and provide explicit protocols on device functionality and intended use for specific diseases or medical conditions.

For instance, Valencell, the producer of the biometric earbud PerformTek™, has announced plans to pursue FDA approval for clinical management of hypertension and diabetes ("Valencell - Health Management," 2013). Currently, PerformTek™ is positioned as a consumer device used for fitness training that feature an earbud technology that continuously measures heart rate, respiration rate, metabolic rate,

$\text{VO}_{2\text{max}}$, and ventilatory threshold. Since vital measurements in the ear are highly accurate, PerformTek™ is well positioned for medical device status if the data proves its use in managing targeted conditions – narrowing its application scope for regulatory purposes.

Once considered a medical device, the product may be prescribed and the data is accepted to use for clinical support. Since this is suggested after patient-provider interaction, this data is no longer characterized as system input and becomes feedback data, which is discussed later in section 5.9.

5.1.2 Direct to Consumer Genetic Testing

Direct to consumer genetic testing is forecasted to grow well over \$200 million over the next five years. 23andMe, deCODE genetics, DNA Direct, and GeneLink Biosciences are examples of companies that are leading this segment - providing test kits that can be purchased online for a relatively low cost. In some cases, the resulting data is helpful in discovering individual characteristics that relate to disease risk, drug sensitivity, and carrier status (“Genetic Testing for Health, Disease & Ancestry; DNA Test - 23andMe,” 2012). The various vendors are also using consumer data for research. This data is being collected outside the care system, and in order to be used for clinical decision-making, must be narrowed to specific applications.

5.1.3 Online Resources

80% percent of internet users in the United States look online for health information (Fox, 2011). People seek answers to health questions, consult reviews of treatments and physicians, and connect with others with similar conditions. WebMD and MedicineNet are examples of online resources that provide detailed medical information and supportive communities to supplement professional care visits. Access to social media and peer networks influences the patient’s entry point into the health system, affecting when and where care interactions occur.

These three examples demonstrate the opportunities for collecting data outside of the clinical care setting. The data is being used at the patient level for individual assessment, which is enabling patient engagement. There is concern, however, about quality, opportunities for misusing, and confusion that are inherent with direct to consumer health products and online information.

Patient sourced data is not considered as typical system input, as it lacks a standard path into the system. As well, the broad applicability of many of the sensing devices and genetic testing limits the use in the clinical environment.

5.2 Clinical Knowledge Warehouse (B)

Several applications are now used at the point of care to enable physicians to quickly access clinical information online. Epocrates, for instance, makes guidelines, reference lists, tables, drug and disease information, dosing calculators, and other clinical resources readily available online. Epocrates has over 75 mobile applications for reference and education, suggesting information fragmentation that is overloading providers with many disconnected tools.

EHR technologies were conceived to help doctors manage this information overload, acting as an electronic warehouse system for patient data. The adoption of EHR technologies in the United States, driven by the HITECH Act, provides an electronic storage system for patient data. Several vendors are now certified EHR providers including AthenaHealth, eClinicalWorks, and CareCloud. Adoption, which is in process, intends to expand EHR functionality from data capture to cross system transmission and decision support tools. The three meaningful use stages outline the vision for growth (Figure 38).

Stage 1 2011-2012	Stage 2 2014	Stage 3 2016
Data capture and sharing	Advance clinical processes	Improved outcomes
Stage 1: Meaningful use criteria focus on:	Stage 2: Meaningful use criteria focus on:	Stage 3: Meaningful use criteria focus on:
Electronically capturing health information in a standardized format	More rigorous health information exchange (HIE)	Improving quality, safety, and efficiency, leading to improved health outcomes
Using that information to track key clinical conditions	Increased requirements for e-prescribing and incorporating lab results	Decision support for national high-priority conditions
Communicating that information for care coordination processes	Electronic transmission of patient care summaries across multiple settings	Patient access to self-management tools
Initiating the reporting of clinical quality measures and public health information	More patient-controlled data	Access to comprehensive patient data through patient-centered HIE
Using information to engage patients and their families in their care		Improving population health

Figure 38: Stages of Meaningful Use

(HealthIT.gov, 2012)

This multistage progression represents a substantial departure from the current functionality of medical records. Data collected today is used once, and is not forward accessible. Availability of this data for reuse and learning is the subject of nearly eight IOM reports is at the heart of clinical support systems that are discussed in the next section. The challenge is developing the science that will enable reuse of data for the purpose of learning.

5.3 Clinical Decision Support (C)

The availability of electronic patient data, in combination with clinical libraries and knowledge warehouses, lays the foundation for clinical decision support (CDS) tools. These tools are intended to assist providers in diagnosis and evaluation of patient data at the point of care. There are already several types of CDS systems with varied functionality and aspirations, including IndiGO, Clinical Looking Glass, UpToDate, and VisualDx.

Inputs into the CDS system are patient symptoms, physician observations, and test results, as well as medical journals, books, and online reference tools. Today's systems use decision support rules to analyze patient history, looking for patterns and gaps in care. The outputs may go from recommended tests, outlining possible diagnoses, interventions, and treatment plans. The expectation is physician productivity will increase as EHR platforms become more widely available and become integrated with CDS systems.

Archimedes IndiGO (Individualized Guidelines and Outcomes) is developed by Kaiser Permanente and is the commercialization of the Archimedes Model. IndiGO is a CDS system that takes patient-specific data into account to create individualized guidelines, calculate the statistical risk of common adverse events, and suggest interventions. Another CDS example is the Clinical Looking Glass (CLG), developed at the Montefiore Medical Center in New York. CLG provides decision support and allows the medical center to conduct retrospective studies.

As shown in the interaction diagram (Figure 36), CDS tools can be used throughout the three interaction phases and their value to providers relies on the quality of the input data and accuracy of the algorithms. Today we are limited to front end statistics.

5.4 Claim Management Analytics (D)

The extraordinary volume of claims data (nearly 400 terabytes for Medicare alone) presents an obvious source of analytics that has already been adopted in other industries. Several analytics companies are now providing services to payors and providers with the aspiration to increase quality of care through reducing waste, cost, and fraud. Companies like Predilytics and Verisk Health assert that claim databases can be used to predict outcomes, identify miscoding, and enhance revenue. These companies use machine learning on claims databases, demographic, and financial information to develop predictive models of disease risks, patient disenrollment, and readmission to hospital.

Claim analytics are used to extract information about interactions and not based on care metrics. The incentives are for cost management, and are not focused on improving outcomes. Additionally, as explained in section 2.4, the information in the claim is limited to codes and administrative data which does not have the medical value needed to make care decisions.

5.5 Genetic Testing (E)

We reviewed patient access to genetic testing in section 5.1.2. More laboratories are offering specialized genetic testing services directly to providers. Genetic tests are used in the clinical care setting to inform diagnosis, prognosis, risk prediction, prevention, and treatment selection for many disorders and diseases. More than 1,700 genetic tests are now available, but only a few are regulated by the FDA (“Genomics - HealthyPeople.gov,” 2012). Certified testing labs and molecular diagnostic companies send test kits to providers, clinicians collect patient’s blood sample or oral swab, and send it back to the lab for analysis. Several companies also provide educational services to the patient, once results are known. The availability of this individualized data is envisioned to yield a more objective diagnosis process.

There are several genetic tests available for cancer diagnosis and treatment. Myriad and Foundation Medicine provide tests that are intended to inform cancer risk and specify particular mutations or abnormalities in existing tumors. This genomic analysis, in some cases, helps to select care regiments. Myriad's product offerings are listed to the right.



Figure 39: Myriad Cancer Products Offerings

Companion diagnostic research is becoming a focus during drug development, with over eighty partnerships between pharmaceutical and diagnostic companies (Westenberg, 2012). Companion diagnostics can be used to stratify patients that are most likely to benefit from a drug. The intent is to prevent unnecessary costs due to ineffective medication usage and to help narrow the treatment landscape. Only a few products have been approved to market alongside therapeutics.

The data retrieved from genetic testing intends to help providers make better-informed clinical decisions, given that the tests are highly accurate. There are numerous concerns in adoption including false positives, testing overuse, genetic discrimination, and excessive distress that comes with positive disease prediction. The science is still advancing and we currently are only witnessing the very beginning of one avenue of personalized medicine.

5.6 Open Innovation Initiatives (F)

As shown in the interaction diagram and discussed in chapter 4, manufacturers produce exorbitant amounts of product data while developing new medical products. The breadth of data and span of development process introduces an opportunity for collaboration across companies and academia. We observe several innovation trends in the various phases of the commercialization process: discovery, clinical trials, and approval process.

Eli Lilly's Open Innovation Drug Discovery (OIDD) Platform and Sage Bionetworks are examples of collaboration during drug discovery. Both platforms are designed to expand the drug discovery and disease research networks through engaging the broader scientific community. Lilly's platform enables researchers to submit their molecular compounds for screening using Lilly's internal assays via a complimentary web-based application (Eli Lilly, 2011). Using Lilly's established evaluations tools, researchers can quickly learn about the potential of their compound and can further collaborate with Lilly on advancing development. The goal of open innovation during discovery is the use of collective efforts to reduce the time and cost of the phase.

Traditionally, only a subset of product data is made public, although there is increasing belief that the medical community could not only benefit from this data, but could add value to the development process. TLS is one such example. Transparency Life Sciences (TLS) has created an open platform to design and execute clinical trials using crowd sourcing. Involving all stakeholders, TLS welcomes input from medical experts, researchers, physicians, patients, families, and regulators. The web-based platform enables collaboration of all participants in protocol design and data analysis in real time. Figure 40 shows the steps involved in clinical trials and TLS's leverage point at each step.

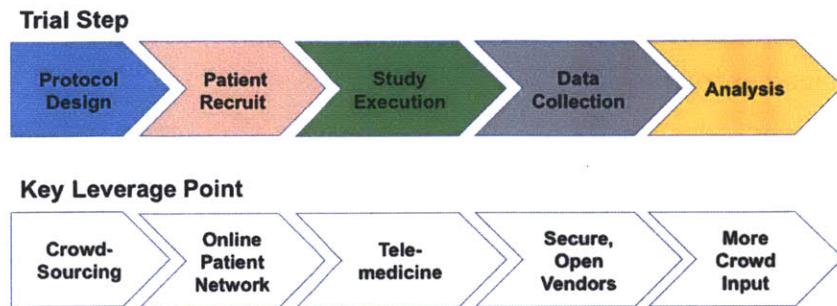


Figure 40: Transparency Life Sciences Process Leverage Points

(TLS, 2012)

Regulators are also proponents of open research data. The European Medicines Agency (EMA) published a memorandum in 2012 calling for the need of open clinical trial data, a response from a series of questionable trial data from Roche (Tamiflu) and GlaxoSmithKline (Avandia, Paxil). (Appendix D) The EMA proposes that clinical trial data should not be considered commercial, confidential information and patients who enroll in a trial do so with an assumption of contributing to medical knowledge (Eichler, Abadie, Breckenridge, Leufkens, & Rasi, 2012). Also in 2012, the British Medical Journal established the Open Data Campaign, which aims to achieve appropriate and necessary independent scrutiny of data from clinical trials (BMJ). GlaxoSmithKline announced support for BMJ's Open Data Campaign and is beginning the process to release individual-level data.

5.7 Product Licensing (G)

Several initiatives are now underway to enhance drug commercialization, led by the regulatory bodies of the United States and Europe. A current focus is on new drug development concepts to accelerate approval, which takes many names including:

- Staggered approval (EMA)
- Progressive reduction of uncertainty (FDA)
- Progressive authorization (Health Canada)

- Test bed for adaptive regulation (HSA Singapore)
- Managed entry (payors, HTAi)

Traditional drug licensing approaches, as described in section 2.6, are based on binary decisions, at the moment of licensing an experimental therapy is transformed into a safe treatment option (Eichler, Oye, et al., 2012). Regulatory decisions are based on incomplete data, but delaying approval even further would deny patient access to a potentially beneficial, life-saving drug. In a recent paper published in Nature by a consortium of researchers and stakeholders, led by MIT NEWDIGS, adaptive licensing is explained as a prospectively planned, adaptive approach to the regulation of new drugs, integrated with iterative phases of data gathering followed by regulatory evaluation and license adaption (Eichler, Oye, et al., 2012). Adaptive licensing, and other accelerated approval models, aspires to yield timely access to drugs through continuous monitoring once implemented.

5.8 Distribution Management and Analytics (H)

Prescription dispensing and expenditure in the United States is on an upward trend, reaching four billion scripts in 2011 (Lindsley, 2012). Distributors such as PBMs, as discussed in chapter 2, share the same incentives as payors to track medication alignment, assess compliance, and monitor safety risks. The volume of pharmacy claims give PBMs insight to prescribing patterns and many are leveraging analytics to provide added-value services for cost reduction. For instance, Express Scripts and CVS Caremark have recently begun pharmacogenomic programs, with the intention to intercept prescriptions and assess whether an appropriate genetic test is available to further inform treatment selection and prevent ineffective medication usage. Also dubbed, Genetics Benefit Management, these programs attempt to harness the benefits of genetic testing by educating both prescribers and patients of existing tests, and then providing the follow-up testing service. Generation Health, recently acquired

by CVS Caremark, analyzes prescription claims to identify patients that may benefit from genetic testing. This aims to ensure that patients receive the right therapy and right drug dosage, through optimizing the usage of genetic tests.

Drug licensing changes and other accelerated approval mechanisms, will require more monitoring and control of product distribution. One solution is the Medbox, a medication dispensing system that manages inventory and compliance. Medication is stored in a temperature-controlled unit and dispensing is controlled by biometric identification (fingerprint sample). The Medbox confirms that the user is a registered patient that has authorization to access the medication. Every transaction is documented and patient data is used to drive accurate dispensing.

5.9 Localized Feedback (I)

The final stage in the interaction diagram is the monitoring loop, where the intervention effect is assessed for benefits and risks. This is often the most difficult piece of data to obtain, as after intervention, the patient may not continue to engage with the health system. However, several new technologies are being developed to improve data collection post-intervention. Proteus Digital Health has created an ingestible sensor that is manufactured in a pill to detect ingestion time and physiologic data. High quality feedback related to the effects of the drug is recorded after the pill is swallowed. An external patch captures and transmits data to a mobile device. CardioMEMS is another technological breakthrough that measures blood pressure and heart rate in patients with chronic heart disease through a miniature sensor implanted into the pulmonary artery. The sensor wirelessly sends pressure readings to an external device. The data retrieved off of these devices is sent to the clinical care team, enhancing the available data for intervention assessment.

A slew of other remote monitoring devices that track vital signs, as previously discussed, can also be prescribed to patients to enhance the data pool. The Patient Centered Medical Home (PCMH) initiative provides an incentive for new companies in the monitoring device market. Several devices now have the capability to transfer real-time patient data, collected remotely, to EHR systems.

Another method of feedback is the reporting of adverse events to manufacturers and regulators. The FDA has previously relied only on the passive Adverse Event Reporting System (FAERS) for voluntary patient and provider reports of product safety issues. FAERS has proven to be insufficient with the low volume of reports and poor data quality. Mini-Sentinel is a FDA pilot project that is creating an active surveillance system, which uses claims and medical records to seek safety issues of marketed products. The system is driven by a safety hypothesis and queries data from dozens of partner sources for assessment. Although significant effort has been put into developing the system, its use is limited to research validation and is not designed to be used at the point of care.

5.10 Summary of Trends

The interaction diagram in Figure 36 serves as a comprehensive map to explore the current trends across the health system. Improvement initiatives are paired with the data inefficiencies described in chapter 4. Each initiative focuses on a segment of the system and aims to optimize the data for its own benefit. For instance, the basis of payor analytics is to reduce costs and it is arguable whether patient outcomes are a secondary or even tertiary priority. Another conundrum is the use of pharmacogenomics to identify patient subsets that will likely respond to a therapeutic. For a pharmaceutical or biotechnology company, this limits their market and greatly affects revenue potential. Even though companion diagnostics promise less waste and more efficiency, the limited number of patients is detrimental to the manufacturers' bottom line.

CHAPTER 6: CONCEPT OF A LEARNING HEALTH SYSTEM

In chapter 4 we introduced a flow diagram that explains data emerging from the patient interaction with the care system. And in chapter 5, we traversed the diagram to identify how data is used and identified emerging trends in data collection and use that point to opportunities for local enhancements to the system. Understanding how to translate these observations into system wide learning is the focus of this chapter.

In this chapter we propose a system architecture concept that favors learning. We start with a topological transformation of the interaction diagram (Figure 35) that moves us away from the sequential representation of patient interactions, in favor of groupings of form and function. The system architecture that results - shown in object-process notation in Figure 42 - has all of the elements needed to achieve optimal data usage that supports learning. The rest of the chapter connects this new architecture with the current system, the various local optimizations it is undergoing, and various interpretations of learning in the health system – such as the ones proposed in the IOM learning healthcare series of reports (Institute of Medicine, 2012).

6.1 Functional Architecture

In Figure 41, we reorganize the interaction diagram, focusing on form and function. Although they are visually very different, Figure 35 and Figure 41 are topologically identical; that is, they capture the same information about patient interaction but emphasize different aspects of the system. Figure 34 emphasizes the temporal sequence of events while Figure 41 emphasizes the nature and function of data and processes associated with that interaction. We identify five distinct clusters of activities: collected patient data, aggregated data, diagnosis/treatment selection, treatment approval, and product approval.

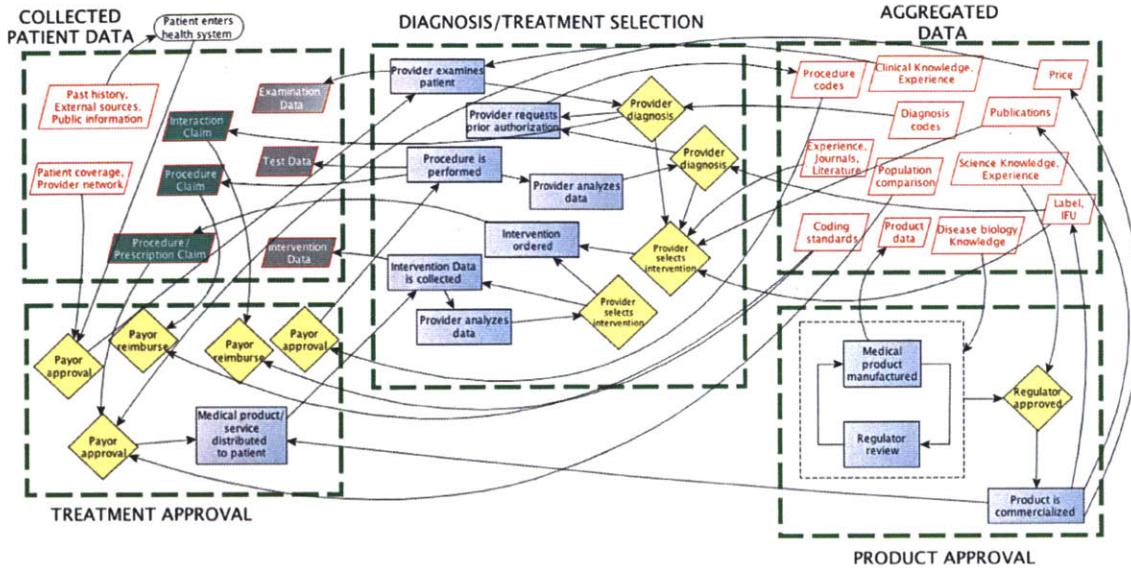


Figure 41: Functional Architecture of the Interaction Diagram

Figure 41 shows the clustering of activities as either data or process. Data comes in two flavors, either associated with an individual patient interaction with the system or aggregated and possibly anonymized. Process categories sort activities by the decisions and functions associated with treatment, its approval/distribution, and the development/approval of new products. The next two sections further review these activities.

6.1.1 Data Categories in Figure 41

The collected patient data category consists of all forms of data that are originally sourced from the patient, namely vital measurements, clinical test data, and medical claims. Since the patient is the common thread across this data category, we can envision a single data repository for each patient. Such a repository is technically feasible with advances in electronic medical records, personal health records, and real-time monitoring devices (sections 2.3, 5.1). A repository for patient data that is fully managed would allow patients to record their health record and state of health continuously – for health management or piecemeal sharing with care providers and other stakeholders.

The aggregated data category includes prior knowledge, experience, accrued population data, publications, policies, and product information. These are all the sources of information that result from aggregating data, be it cohort data, scientific or regulatory data, etc. These are the kinds of data that are currently being used to drive clinical decision support systems (CDSS), like IBM Watson and Archimedes IndiGO (section 5.3). Compiling this breadth of information into a single system could become the basis for a health information/knowledge exchange that supports learning.

The combination of these data categories results in knowledge creation driven by patient data. The data from the patient repository can be combined with the aggregated data to make care decisions on individual patients, inform public health, or used for research. This new data connection makes the system wide learning cycle explicit and, to the best of our knowledge, has yet to be implemented in the current system.

6.1.2 Process Categories in Figure 41

From the standpoint of data, there are three markedly different processes: diagnosis/treatment selection, treatment approval and distribution, and product approval. Diagnosis/treatment selection consists of various provider interactions with the patient to collect data for diagnosis. Each interaction between the patient and provider is reviewed for reimbursement by the payor - treatment approval process. The product approval process is embodied in the manufacturer and regulator loop and consists of the activities involved in manufacturing and licensing new medical products for distribution to the patient.

These processes have naturally different workflows and objectives; however, they all rely on the same set of input data from the two data categories described in section 6.1.1.

6.2 System Concept Emerging From Figure 41

All engineering systems have an object-process nature (Crawley, 2011). The division of the system into data and process categories represented in Figure 42 can be cast in terms of Crawley's architectural framework of systems. We do so by identifying collected patient data as operands and aggregated data as instruments. In Crawley's systems framework, instruments support the transformation of operands by processes. This interpretation, shown in Figure 42, confirms our expectation about the health system, namely that the three processes we identified act differently on data. It also helps identify why the current system struggles to incorporate learning. For learning to occur, we need a process that routinely reverses flow: the current operands need to become instruments and the current instruments need to become operands.

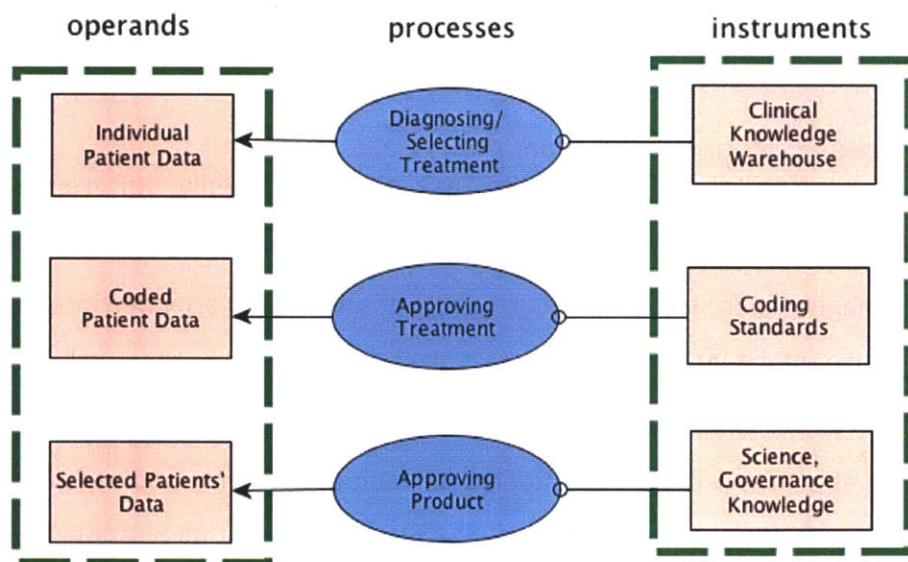


Figure 42: Synthesized System in Object-Process Notation

The simplified view of the functional architecture in object-process notation in Figure 42 allows us to use a different framework to validate our representation of the system.

6.2.1 A Learning Health System

Figure 43 shows a graphical representation of a new system that contains the topology of Figure 41, but in which we synthesized categories. We combine some of the arrows and, in doing so identify new functions that were not obvious in any of the previous diagrams that emphasized temporal sequence.

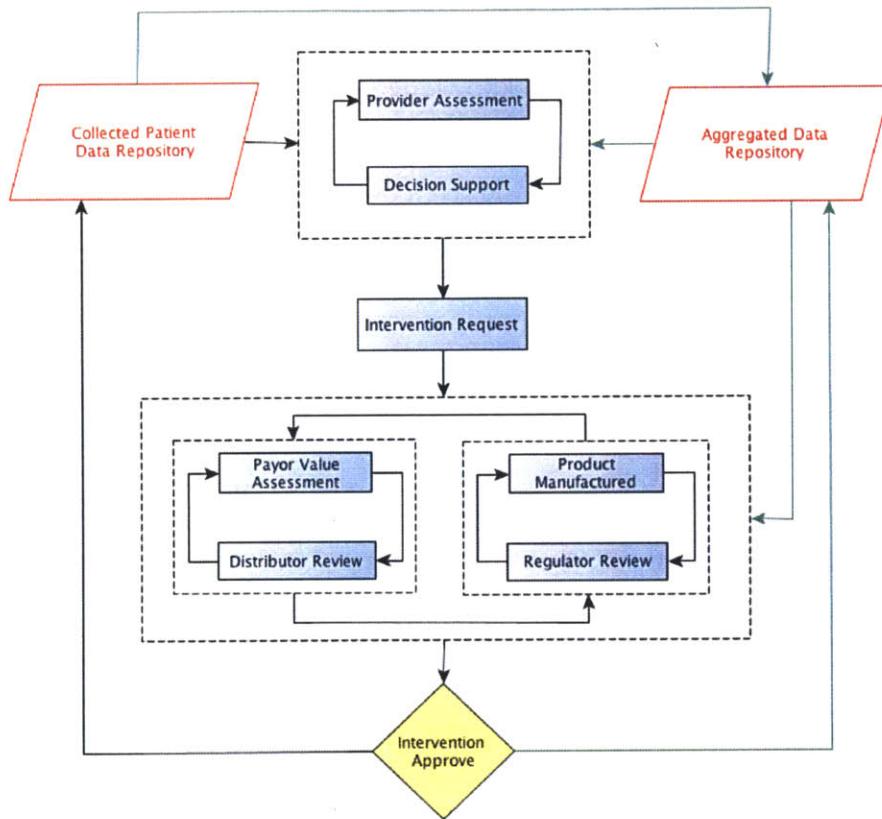


Figure 43: Learning Health System Concept

The current system and its three interaction phases – examination, diagnosis, and intervention – are still represented in this diagram, but they are represented now as different iterations around the single loop. For instance, during patient encounters, the provider uses newly collected data, information from past visits, and aggregated data for diagnosis and treatment selection. The next set of processes is triggered to approve the intervention selected by the provider. Once the intervention request is approved, patient data continues to feed back into both data repositories.

This system diagram allows us to identify elements that are implicit in the previous system that map to opportunities for learning that we identified in chapter 4. Collectively they represent the architecture vision of a health system designed for value assessment and knowledge management. The green arrows represent learning and are not standard in today's system. These arrows indicate the flow of data to and from the collected patient data repository and the aggregated data repository. The addition of the green arrows indicates that the system has two functions: improve individual patient care and contribute to global knowledge. The accrual of collected patient data into the aggregated data repository indicates that every patient contributes to overall learning. As well, every process performed and decision made also contributes to learning. The provider assessment/decision support function supports opportunities for direct learning from these two data repositories.

The payor assessment/distributor loop uses better information for value assessment and supports the notion of value for learning. The product manufactured/approval loop moves from outside of the system to within and by doing so, supports the models for accelerated licensing. These two subsystems act as a checks and balances system to define value as a combination of learning and outcomes. Therefore, as a design choice, we group these together as if they were a single subsystem.

The activities within the product manufactured/approval loop are concealed in the patient interaction diagram, but are analogous to the activities in the care system. The difference is that the manufacturer assumes the payor role by following the regulatory process. Integrating this loop into the care system, as proposed in Figure 43, enables continuous monitoring post intervention.

Collectively, the data repositories and emerging functions show value to the health system from lessons learned from broader data exposure. The system is designed to learn from each patient interaction, supplying better data to each subsystem function.

6.3 System Problem Statement, Revisited

In chapter 1 we defined the system problem statement based on our assessment of the stakeholder's needs and system goals. The future system vision deviates significantly from how we generally think about the health system; the future state for the health system we propose introduces learning as an integral component. This new vision emerges from our stakeholder interviews, analysis of the system inefficiencies, and concept vision of a learning health system. We express this shift by revising the system problem statement:

To maintain the health of individuals,

By learning at every interaction, activity, process, and function,

Using all the information acquired through data analysis, practice of medicine, and science.

The result of this new model is a system that benefits both the individual patient and at the same time supports knowledge generation and learning. Value is derived from every interaction and intervention, and lessons are drawn out of each process to better maintain the health of the next individual. Central to the future vision is the requirement for a custodial health "bank" for every individual and a health knowledge and information exchange. The evolution of individual information transforms into system wide learning.

CHAPTER 7: CONCLUSION

We envision a new architecture for the system of health where each patient interaction with the health system presents an opportunity to contribute to knowledge. We show that the system can be architected as a single loop, where data is generated, interpreted, and reused for analysis and decision-making. The loop considers value delivered to the patient directly and through learning. This data driven architecture for the health system is compatible with the current system but also points to a path to enable system wide learning. This architecture, introduced in chapter 6, is the main contribution of this thesis.

We took a principled stand that learning in health requires a deliberate approach to handling, collecting, analyzing, and interpreting data. That is, a learning health system is necessarily an information based system. Our analysis reveals that the current system does not effectively transfer data across interfaces; we showed that much in chapter 5 with our overview of data driven solutions that are emerging locally to compensate for partial access to information. Nevertheless, our data overview in chapter 2 reveals there are abundant opportunities to incorporate better data and better data practices into the health system. Chapter 3 tells us, though, that the traditional stakeholder view of the health system does not help us understand how to incorporate those into the existing system. The stakeholder agnostic view of health data we took in chapters 4 and 5 helped us uncover hidden functionality in the system, such as distributor review functions and the two markedly different categories of data that support our vision for a data driven system designed for value and learning. The discovery of said functions, the landscape overview of data, and the methodology we used to remove stakeholder and temporal biases from the traditional view of the system are also contributions of this thesis.

7.1 Findings Supporting New Vision

The following is a summary of the findings, spread over chapters 3-5, that support the vision laid out in chapter 6.

- The entry of data from outside the clinical environment into the care system is narrow and there is no standard method for the system to broadly accept these measurements. This data, in the form of physiological measurements for instance, presumes to offer a more comprehensive view of patient status enabling monitoring between encounters.
- Adoption of EHRs helps advance towards an improved management of patient records, but does not resolve uncertainties about the desired evolution of care delivery in a learning health system. Understanding how to compare different patient records and how to enhance interconnectivity across patient and provider systems will facilitate learning, but remains undefined.
- Automated generation of claims from EHRs may reduce translation errors but does not address systematic errors induced by the limitations of the disease code system, and the propagation of these errors to coverage decisions and outcomes.
- Data released by manufacturers has limited value because it is based on population statistics of a pre-defined patient group and is aggregated for public use, and it is hard to relate to the needs of a specific patient. The development process may benefit from contributing data for system wide learning.
- There is no standard mechanism to give feedback on approved products. However, the same processes are shared between the care system and the approval system, although they have different objectives. There is an opportunity to design monitoring and learning mechanisms in the care delivery process that feed from the lessons learned in the approval process.
- The drug approval process, by design, releases products that provide a benefit to only a small percentage of patients who receive them (Aspinall & Hamermesh, 2007). Patient specific data as input into the development process may improve the rates of efficacy for medical products, by, for instance, updating the attributes of the population likely to benefit from said medical product.

7.2 Future Work

We are at the cusp of a health system transformation. Our analysis suggests that the current stakeholder perspective is not yet fully compatible with a shift towards a health system that is personalized, predictive, preventive, and participatory, our conclusion matches the observations (Hood & Friend, 2011) make about the need to realign stakeholders to reduce waste and improve system inefficiency. The architecture of the system proposed in chapter 6 is a good starting point to engage in new research on the redefinition of the scope of stakeholders by the value associated with health data.

The patient's role needs to shift from a passive receiver of care to an engaged consumer. Objective data sourced from the patient is highly valuable, but needs to be complemented with systems capable of managing this voluminous amount of data to extract meaning. Patient specific information must also be comparable to enable broader usage across the process subsystems defined in chapter 6, such as in drug development – as a basis for personalized medicine. An avenue for future research is the development of new “big data” tools that enable the manageability and comparability of this data. The conclusions in chapter 6 may be used as an architecture to understand the areas in which such tools are needed, for instance, to support the transition from individual to collective data and back to aid decision-making, or to ensure that the delivery system derives lessons to be incorporated into the proposed health information exchange (aggregated data repository in Figure 43).

Our analysis highlights the weak connections between the product manufacturer subsystem with the care delivery system, distinct segments driving health that communicate only through a narrow interface – e.g. labels, medication guides, and adverse event self-reporting. New drug approval models are being proposed, such as adaptive licensing, that blur the boundaries between pre and post-approval. While the transition to a progressive means of licensing has received a lot of attention, an

uncertainty remains about how to link monitoring practices pre and post approval. The learning system we proposed suggests one approach to think about this problem.

The evolution of a health system from one that traces diseases to one that manages health implies a shift from an accidental accrual of health data to a deliberate design data streams. This data will be used to generate evidence to inform decisions and ultimately extract knowledge for continuous learning. This thesis lays the foundation and further research is proposed to continue development of the architecture of a data driven health system.

REFERENCES

- Ahadome, T. (2012). *Wearable Technology - A Global Market Overview*. Retrieved from http://www.wearable-technologies.com/c/document_library/get_file?uuid=65ab1125-6c83-4627-b8d2-bf56ac8e8fca&groupId=10192
- Ambizas, E., Ezzo, D., & Patel, P. (2009). Drug Information Resources for the Community Pharmacist. Retrieved from http://www.uspharmacist.com/continuing_education/ceviewtest/lessonid/106043/
- American Medical Association. (2012). *2012 National Health Insurer Report Care*. Retrieved from <http://www.ama-assn.org/resources/doc/psa/2012-nhirc-results.pdf>
- Andrew, K. (2009). *The Basics of Pharmacy Benefits Management*. Retrieved from http://www.anthem.com/shared/va/f5/s1/t0/pw_b135247.pdf
- Aspinall, M. G., & Hamermesh, R. G. (2007). Realizing the Promise of Personalized Medicine Realizing the Promise of Personalized Medicine. *Harvard Business Review*. Retrieved from <http://missionfacilitators.com/Articles/OrganizationalDevelopment/Articles/realizing the promise of personalized medicine.pdf>
- Corventis. (2012). Corventis™ - AVIVO. Retrieved January 6, 2013, from <http://www.corventis.com/us/avivo.asp>
- Crawley, E. (2011). Lecture Notes for MIT Course ESD.34 System Architecture.
- Doshi, P., Jefferson, T., & Del Mar, C. (2012). The imperative to share clinical study reports: recommendations from the Tamiflu experience. *PLoS medicine*, 9(4), e1001201. doi:10.1371/journal.pmed.1001201
- EHR Market. (2009). Retrieved January 1, 2013, from <http://www.ehrmarket.com/blog/2009/11/product-review-nextgen/>
- Eichler, H.-G., Abadie, E., Breckenridge, A., Leufkens, H., & Rasi, G. (2012). Open Clinical Trial Data for All? A View from Regulators. *PLoS medicine*, 9(4), e1001201. doi:10.1371/journal.pmed.1001201
- Eichler, H.-G., Oye, K., Baird, L., Abadie, E., Brown, J., Drum, C., & Hirsch, G. (2012). Adaptive Licensing: Taking the Next Step in the Evolution of Drug Approval. *Nature*, 91(3). Retrieved from <http://www.nature.com/clpt/journal/v91/n3/pdf/clpt2011345a.pdf>

Eli Lilly. (2011). Open Innovation Drug Discovery. Retrieved May 8, 2012, from
<https://openinnovation.lilly.com/dd/>

Epson Enters Healthcare Business with Wristwatch-Type Pulse Monitor. (2012).
Retrieved January 3, 2013, from
http://global.epson.com/innovation/technology_articles/201206_2.html

FDA. (2003). FDA's Review Process for New Drug Applications, (March).

FDA. (2012a). New Drug Application (NDA). Center for Drug Evaluation and Research.
Retrieved June 24, 2012, from
<http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/NewDrugApplicationNDA/default.htm>

FDA. (2012b). Postmarket Drug Safety Information for Patients and Providers - Approved Risk Evaluation and Mitigation Strategies (REMS). Center for Drug Evaluation and Research. Retrieved from
<http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111350.htm>

FDA. (2012c). *A Brief Overview of Risk Evaluation & Mitigation Strategies (REMS)*.
Retrieved from
<http://www.fda.gov/downloads/AboutFDA/Transparency/Basics/UCM328784.pdf>

FDA. (2012d). National Drug Code Query-Package Code. Retrieved from
<http://www.accessdata.fda.gov/scripts/cder/ndc/packagecode.cfm>

FDA Guidance for Industry: Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products. (2012). Rockville. Retrieved from
<http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm332181.pdf>

Fox, S. (2011). *Pew Internet: Peer-to-Peer Healthcare*. Retrieved from
http://pewinternet.org/~/media//Files/Reports/2011/Pew_P2PHealthcare_2011.pdf

Gandhi, T. K., Sittig, D. F., Franklin, M., Sussman, A. J., Fairchild, D. G., & Bates, D. W. (2000). Communication Breakdown in the Outpatient Referral Process. *Journal of general internal medicine*, 15(9), 626–31. Retrieved from
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1495590/>&rendertype=abstract

Genetic Testing for Health, Disease & Ancestry; DNA Test - 23andMe. (2012). Retrieved January 2, 2013, from <https://www.23andme.com/>

- Genomics - HealthyPeople.gov. (2012). Retrieved December 27, 2012, from
<http://www.healthypeople.gov/2020/topicsobjectives2020/overview.aspx?topicid=15>
- Goldman, S. (1996). *The Clinical Impact of Adverse Event Reporting*. Retrieved from
<http://www.fda.gov/downloads/safety/medwatch/ucm168505.pdf>
- HCAN. (2012). Medical Loss Ratio Provision Gives \$1.3 Billion Back to Consumers | Health Care for America Now. Retrieved from
<http://healthcareforamericanow.org/2012/04/26/medical-loss-ratio-provision-gives-1-3-billion-back-to-consumers/>
- HealthIT.gov. (2012). Meaningful Use | Policy Researchers & Implementers. Retrieved December 27, 2012, from <http://www.healthit.gov/policy-researchers-implementers/meaningful-use>
- Heinrich, J. (2000). *Adverse Drug Events: Substantial Problem but Magnitude Uncertain. Testimony Before the Committee on Health, Education, Labor, and Pensions, U.S. Senate*. Retrieved from <http://www.gao.gov/new.items/he00053t.pdf>
- Hirschler, B. (2012). Drug pipelines improving after years in doldrums | Reuters. London. Retrieved from <http://www.reuters.com/article/2012/06/26/us-drug-improving-idUSBRE85P06H20120626>
- Hood, L., & Friend, S. H. (2011). Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nature reviews Clinical oncology*, 8(3), 184–187. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21364692>
- ICD - ICD-10-CM - International Classification of Diseases, Tenth Revision, Clinical Modification. (2012). Retrieved from <http://www.cdc.gov/nchs/icd/icd10cm.htm>
- ICD-10 Version:2010. (2010). Retrieved January 2, 2013, from
<http://apps.who.int/classifications/icd10/browse/2010/en#/C34.1>
- Institute of Medicine. (2001). *Crossing the Quality Chasm*. Washington DC: National Academy Press.
- Institute of Medicine. (2012). Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. *The National Academies Press*.
- Lindsley, C. W. (2012). The Top Prescription Drugs of 2011 in the United States: Antipsychotics and Antidepressants Once Again Lead CNS Therapeutics. *ACS Chemical Neuroscience*, 3. Retrieved from
<http://pubs.acs.org/doi/pdf/10.1021/cn3000923>

- Measuring and Managing Fluid Balance. (2011). *Nursing Times*, 107(28), 12–16.
Retrieved from <http://www.nursingtimes.net/Journals/1/Files/2011/8/1/FluidbalanceCorr.pdf>
- NCPA. (2011). *NCPA Pharmacy Benefit Management Manual*. Retrieved from
http://www.ncpanet.org/pdf/leg/jan12/pbm_manual.pdf
- New CMS-1500 (08/05) Healthcare Claim PDF SmartForm With Built in 837P EDI Capabilities. (2007). Retrieved January 3, 2013, from <http://www.24-7pressrelease.com/press-release/pdf-smartforms-releases-the-new-cms1500-0805-healthcare-claimpdf-smartform-with-built-in-837p-edi-capabilities-31432.php>
- NIH. (2007). Understanding Clinical Trials. Retrieved May 8, 2012, from
<http://clinicaltrials.gov/ct2/info/understand#Q19>
- NLM. (2012). DailyMed. Retrieved June 24, 2012, from
<http://dailymed.nlm.nih.gov/dailymed/about.cfm?CFID=18030612&CFTOKEN=c5df27fcc5f620b7-82AAEF76-B196-9849-02E9B8DE88750060&jsessionid=84303dc2d1848a17596f7738a7d7e3f76384#content>
- Overstreet, B. (2012). A Closer Look at FDA's Adverse Event Reporting System. *Patient Safety and Quality Healthcare*. Retrieved from
<http://www.psqh.com/januaryfebruary-2012/1098-a-closer-look-at-fdas-adverse-event-reporting-system.html>
- Patrick, D., Murray, T., Bigby, J., & Boros, A. (2012). *The All-Payer Claims Database Health Care Claims Data Release - Documentation Guide* (pp. 1–294). Boston. Retrieved from <http://www.mass.gov/chia/docs/p/apcd/apcd-release-document-2012-10-31.pdf>
- Personalized Medicine: Trends and Prospects for the New Science of Genetic Testing and Molecular Diagnostics. (2012). Minnetonka. Retrieved from
http://www.unitedhealthgroup.com/hrm/UNH_WorkingPaper7.pdf
- Privacy Rights Clearinghouse. (2012). Your Prescriptions and Your Privacy. *California Medical Privacy Fact Sheet C4*. Retrieved January 5, 2013, from
<https://www.privacyrights.org/fs/fsC4/CA-medical-prescription-privacy>
- TLS. (2012). Transparency Life Sciences. Retrieved May 8, 2012, from
<http://transparencyls.com/>

Valencell - Health Management. (2013). Retrieved January 13, 2013, from
<http://valencelldev.clientmark.com/applications/other-applications>

Wearable Blood Pressure Sensor Offers 24/7 Continuous Monitoring. (2009). *MIT News Office*. Retrieved January 3, 2013, from
<http://web.mit.edu/newsoffice/2009/blood-pressure-tt0408.html>

Westenberg, D. (2012). Coming Wave Of Companion Diagnostics: Who's Making A Splash - Seeking Alpha. *Seeking Alpha*. Retrieved from
<http://seekingalpha.com/article/751671-coming-wave-of-companion-diagnostics-who-s-making-a-splash>

ZweenaHealth. (2012). Retrieved January 2, 2013, from
<http://www.zweenahealth.com/how-zweena-works>

APPENDIX A: REFERENCED WEBSITES

Chapter 2

<http://www.athenahealth.com/>
<http://www.practicefusion.com/>
<http://care360.questdiagnostics.com/>
<http://www.epic.com/>
<http://www.eclinicalworks.com/>
<http://www.nextgen.com/>
<http://www.doclopedia.com/>
<http://phr.emrystick.com/>
<https://juniperhealth.com/>
<https://www.mivia.org/>
<http://dnapolicy.org/>
<http://nih.gov/>
<http://www.cms.gov/>
<http://www.tuftshealthplan.com/>
<http://www.aetna.com/>
<http://www.fda.gov/default.htm>
<http://www.redbook.com/redbook/online/>
<http://www.nlm.nih.gov/bsd/pmresources>
<http://www.milliman.com/home/index.php>
<http://www.optuminsight.com/>
<http://www.clinicaltrials.gov/>
<http://www.fda.gov/>
<https://www.accessdata.fda.gov/scripts/medwatch>

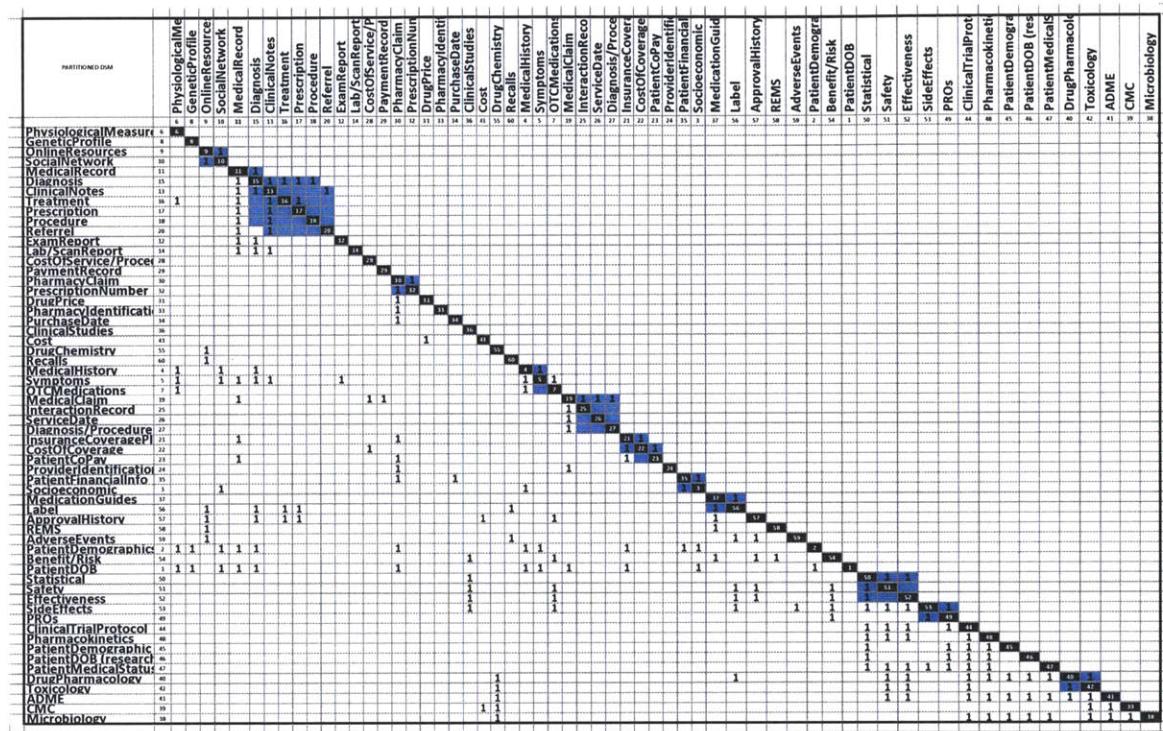
Chapter 5

<http://valencell.com/>
<http://www.electricfoxy.com/>
<https://www.23andme.com/>
<http://www.epocrates.com/>
<http://archimedesmodel.com/>
<http://www.predilytics.com/>
<http://www.veriskhealth.com/>
<http://www.foundationmedicine.com/index.php>
<http://www.myriad.com/>
<http://transparencyls.com/>
<http://www.express-scripts.com/corporate/>
<http://www.generationhealth.com/>
<http://www.thedispensingsolution.com/>
<http://proteusdigitalhealth.com/>
<http://www.cardiomems.com/>
<http://mini-sentinel.org/>

APPENDIX B: STAKEHOLDERS' NEEDS

Stakeholder	Needs
PATIENT	Good health, high quality of life
	Access to and affordable products, services, and care
	Education on health, wellness, and treatment to make informed decisions about prevention and care
	Tools to manage health and disease
	Access and engagement with network of others with similar conditions for support
	Ability to explain and communicate symptoms, feelings, state, etc.
PROVIDER	Privacy and security of personal data
	Education, knowledge, training on existing/new medical products, treatments, and tools
	Access to health facilities, equipment, and medical products (drugs, devices)
	Awareness by patient, provider, and payor networks
	Efficient, thorough, timely examination of patients to collect data
	Database for patient administration and clinical information
PAYOR	Communicate and transfer information to patients, provider networks, and payors
	Tools and applications to inform decision making and monitor patients to assess outcomes
	Payment from payors and patients
	Record of services performed by providers and record of medical products distributed to patients
	Knowledge of new/existing medical products, to evaluate payment coverage
MANUFACTURER	Payment from patients, patient employer, government
	Database to receive patient and provider information and to analyze and process claims
	Reduce long term costs by maintaining health of members minimizing services and products needed
	Knowledge of disease biology, molecular chemistry, medical science
	Materials to manufacture drugs
	Ability to test safety and effectiveness of drugs
	Sample human population to test drug
	Feedback on product outcome
	Obtain approval for license
	Reduced time to market
REGULATOR	Reduced development costs
	Revenue from customers (payors, patients)
	Funding for development, manufacturing costs
	Protect public safety
DISTRIBUTOR	Support/resources from government
	Complete information to evaluate drug/device products
	Monitoring product post-approval
	Access and storage of products
	Dispensing system to track incoming prescriptions, outgoing products, and transactions for reimbursement
	Product cost breakdown between payor and patient
	Patient information and prescription history to assess distribution safety

APPENDIX C: PARTITIONED DSM, SECTION 3.2



Direct Dependencies of 60 Kinds of Health Data

PATIENT	PROVIDER	PAYOR	DISTRIBUTOR
PatientDOB	MedicalRecord	InsuranceCoveragePlan	PharmacyClaim
PatientDemographics	ExamReport	CostOfCoverage	DrugPrice
Socioeconomic	ClinicalNotes	PatientCoPay	PrescriptionNumber
MedicalHistory	Lab/ScanReport	ProviderIdentification	PharmacyIdentification
Symptoms	Diagnosis	InteractionRecord	PurchaseDate
PhysiologicalMeasurements	Treatment	ServiceDate	PatientFinancialInfo
OTCMedications	Prescription	Diagnosis/ProcedureCode	ClinicalStudies
GeneticProfile	Procedure	CostOfService/Procedure	MedicationGuides
OnlineResources	MedicalClaim	PaymentRecord	
SocialNetwork	Referral		
MANUFACTURER			
Microbiology	PatientMedicalStatus		REGULATOR
CMC	Pharmacokinetics		Benefit/Risk
DrugPharmacology	PROs		DrugChemistry
ADME	Statistical		Label
Toxicology	Safety		ApprovalHistory
Cost	Effectiveness		REMS
ClinicalTrialProtocol	SideEffects		AdverseEvents
PatientDemographic (research)			Recalls
PatientDOB (research)			

APPENDIX D: TAMIFLU CASE STUDY

Tamiflu (oseltamivir) is Roche's influenza antiviral and during its 1999 approval by the FDA, the effectiveness of the drug was questionable and believed to be no better than aspirin or acetaminophen (Doshi, Jefferson, & Del Mar, 2012). However, there was tremendous pressure to commercialize Tamiflu, as governments around the world, following the guidance of the WHO, stockpiled the drug. With its assumed benefit, Tamiflu sales surged during the 2005 and 2009 influenza pandemics, raking in \$4.6 billion for Roche. Since then, there has been a growing controversy related to the effectiveness of Tamiflu and Roche's claims of reduced complications, hospitalizations, and prevention of transmissions have been challenged.

In 2009, the Cochrane Collaboration embarked on an independent analysis of the effectiveness of Tamiflu and compared published data with unpublished trial records. The Cochrane team, led by Peter Doshi and Tom Jefferson, claimed that the datasets made available by Roche provided incomplete and inconsistent evidence and there were several discrepancies between the published and unpublished data. As well, the published documents failed to acknowledge the serious adverse events (neurologic events such as hallucinations, convulsions, and encephalitis) which were identified during the clinical trials and have since been widely reported. The controversy over Tamiflu's clinical data continues, and Roche has still yet to provide the additional trial data that has been requested by the Cochrane team.

The Tamiflu example suggests that there is currently no mechanism that transcends the valuable data that is created during clinical trials beyond the development purview. This creates a fragile data connection between the manufacturer and other stakeholders, realizing a substantial data loss. Validation of this observation has recently emerged. Inspired by the Cochrane reviews of Tamiflu, the European Medicines Agency (EMA) has responded to the notion of open clinical trial data. The EMA agrees with the Cochrane claim that clinical trial data should not be considered commercial confidential information and patients who enroll in a trial do so with an assumption of contributing to medical knowledge (Eichler, Abadie, et al., 2012). However, independent review of trial data may risk quality, conflict of interest, and misuse. The EMA further proposes that availability of trial data may be realistic if personal data protection can be assured, quality standards of meta-analyses adopted, and rules for raw data sharing implemented.