

Visual Analytics in Precision Medicine: Using Mixed Methods to Support Stakeholder Data Needs

by

Aldo Ahkin Barrera-Machuca

Biotechnology Engineering, BSc, Tec de Monterrey, 2015

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the

School of Interactive Arts and Technology
Faculty of Communication, Art and Technology

© Aldo Barrera 2020

SIMON FRASER UNIVERSITY

Summer 2020

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Approval

Name: Aldo Ahkin Barrera Machuca

Degree: Master of Science

Title: Visual Analytics in Precision Medicine: Using Mixed Methods to Support Stakeholder Data Needs

Examining Committee:

- Chair:** Halil Erhan
Associate Professor
- Brian Fisher**
Senior Supervisor
Professor
- Sylvain Moreno**
Supervisor
Associate Professor
- Richard Arias-Hernandez**
External Supervisor
Associate Professor of Teaching, UBC iSchool
- Diane Gromala**
Internal Examiner
Professor

Date Defended/Approved: June 26th, 2020

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

- c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Precision medicine solutions require health consumers to increasingly interact with digital interfaces to report their medical history and conditions. This is challenging since the symptoms of an illness can often be located to an activity or a body part and untrained health consumers struggle to communicate them clearly. To address this problem, I designed a mixed methods study, where I first conducted short-term ethnography in a precision medicine company to understand the data requirements of a set of health data analysts. This exploration led to methodological and design guidelines that translated into an interactive data-capture system that visually mapped a controlled vocabulary of human disease phenotypes to a graphical depiction of the body. Results showed that describing the experience of illness in a somatic representation gave health consumers a more accurate and descriptive understanding of their illness, and by doing so captured more reliable data for analysts. The representation can support health care workers to provide more accurate analyses, aid caregivers in managing health risks, and empower health consumers to take action to better their health.

Keywords: Visual Analytics; Precision Medicine; Interactive Visualization; HCI; Design Thinking; Cognitive Mediators

Dedication

Para mis papás, que siempre me han estado ahí para mí, y que su apoyo siempre ha sido el suelo firme donde he plantado mis pies. Aunque vivamos en países diferentes, llevo conmigo el vuelo que me dieron con su amor y educación; siempre estarán conmigo.

Para Mayra y Fernando, que me recibieron en su casa y que me han apoyado durante la maestría, muchas gracias por todo, en especial su paciencia, sus consejos y su cariño. Para Kelli, que nos encontramos en este camino y que juntos hemos crecido como científicos y personas.

Para mis amigos que también empezaron de nuevo en Vancouver.

For those who I met at the beginning of this journey and that now I can proudly call them friends.

Acknowledgements

I would like to thank the Mitacs Accelerate program for supporting my ethnographic fieldwork in a great precision medicine Startup. Special thanks to Antonio, Gail, John, Wilma, and Katie for having me always feel welcome in the Data Team and for supporting my research. Also, special thanks to Nadya Calderon for helping me to navigate my way in the Integrated Science Lab and the Mitacs internship. Additionally, I would like to thank Aaron Johnson and Thara Vayali for helping me with their clinical expertise.

I would like to extend my gratitude to my supervisory committee, starting with my senior supervisor Brian Fisher for supporting me through funding, always giving me new books and papers to read, and for finding value in my research. Thanks to Sylvain Moreno for guiding me when I was losing track of my goal and for helping me to grow as a research scientist in the digital healthcare industry, and to Richard Arias-Hernandez for coaching me in how to conduct workplace ethnography and always being supportive when I faced methodological challenges.

I would also like to thank those professors who shaped my research and were instrumental at different parts of my MSc. Special thanks to Dara Culhane for teaching me how to become an ethnographer and guiding my journey into anthropology, to Carman Neustaedter for being my methodology and methods guru, and for helping me to find flaws in my research design. Thanks to Lyn Bartram for helping me to solidify my Visual Analytics and visualization knowledge, and special thanks to Peter Chow-White and the GeNA Lab for embracing me as part of their lab and helping me to maintain momentum when writing my thesis.

Finally, I would like to thank Tiffany Taylor and the SIAT administrative personnel for helping me to navigate the department's bureaucracy, and everyone else who I did not mention by name but who contributed to my thesis.

Table of Contents

Approval.....	ii
Ethics Statement.....	iii
Abstract.....	iv
Dedication.....	v
Acknowledgements.....	vi
Table of Contents.....	vii
List of Tables.....	xii
List of Figures.....	xiii
List of Acronyms.....	xiv
Glossary.....	xv
Preface.....	1
Becoming Data: Part 1.....	1
The Commute: Human as Data Chimeras.....	1
The Waiting Room: Human-Data multisensorial relations.....	3
Rite of passage.....	3
Chapter 1. Introduction.....	5
1.1. Challenges for Precision Medicine.....	6
1.1.1. The Disinformation Era of Big Data.....	6
The Data Context of the Omics.....	7
1.1.2. Catering to Multiple Stakeholders.....	7
The Health Data Analysts' Perspective.....	7
The Health Consumer's Perspective.....	8
1.1.3. Wicked Problems.....	9
1.2. Visual Analytics.....	10
1.2.1. Design of VA Systems.....	11
1.3. Definitions.....	15
1.3.1. Visualization.....	15
1.3.2. Biomolecular & Descriptive Data.....	17
1.3.3. Digital Records: EHRs, PHRs and MHRs.....	17
1.4. Research Overview.....	18
1.4.1. Research Setting.....	18
1.4.2. Research Objectives.....	19
1.4.3. Research Goal and Contributions.....	20
1.5. Thesis Outline.....	21
1.6. Chapter Summary.....	21
Chapter 2. Literature Review.....	23
2.1. Design Thinking.....	24
2.1.1. The Origins of Design Thinking.....	24
2.1.2. The Nielsen-Norman Design Thinking Framework.....	26

2.1.3.	Design Thinking Pioneers for HCI.....	27
	Herbert A. Simon	28
	Donald A. Norman.....	28
	Donald A. Schön	28
	W. Brian Arthur	29
2.1.4.	A Cognitive Framework for Design Thinking in HCI	30
	The Designer as an HCI Broker	31
2.2.	Ethnography and HCI	32
2.2.1.	Ethnography across disciplines	34
2.2.2.	Adapted Ethnography in Healthcare Research	36
2.2.3.	Cognitive Ethnography	36
2.2.4.	Short-Term Ethnography for Healthcare HCI	37
2.3.	Visual Analytics Design Considerations.....	39
2.3.1.	Cognitive artefacts.....	39
	Cognitive amplifiers and prostheses.....	40
	Cognitive mediators	41
2.3.2.	Target Audience	43
	Health consumers and Health Data Analysts	43
2.3.3.	System Considerations.....	44
	Visual Communication	44
	Usability	44
2.3.4.	Data Ontologies.....	45
	The Human Phenotype Ontology	46
	Medical Terminologies	47
	Data Ontology Visualization	48
2.4.	Chapter Summary	48
Chapter 3.	Methodology.....	50
3.1.	Research Overview	50
3.1.1.	Research Philosophy.....	50
3.1.2.	Research Design.....	51
3.1.3.	Research Questions	53
	Mixed Methods Question.....	53
	Qualitative Question – Short-term ethnography.....	54
	Quantitative Question – User Testing.....	55
3.2.	Design Thinking: Understand - Empathize.....	55
3.2.1.	Empathy	55
3.2.2.	Short-term Ethnography	56
3.2.3.	Study Participants.....	57
	The Researcher	57
	The Data Team	58
	The Physicians.....	59
3.2.4.	Interview Data Transcription	59
3.2.5.	Thematic Coding	60

3.3.	Design Thinking: Understand - Define	61
3.3.1.	Wicked Problems.....	61
3.3.2.	The Unit of Analysis.....	62
3.4.	Design Thinking: Explore - Ideate.....	63
3.4.1.	Mapping the Human Phenotype Ontology (HPO)	64
	Defining the Body Categories as Buttons	66
	Challenges & Limitations.....	67
3.5.	Design Thinking: Explore – Prototype.....	68
3.5.1.	Extracting the HPO.....	68
3.5.2.	Prototyping	69
	The Colour Palette	69
	Visual Saliency	70
	Interaction	71
3.6.	Design Thinking: Materialize - Test.....	73
3.6.1.	Research Setting and Data Collection	73
3.6.2.	Study Procedure.....	73
	1. Ethics and Consent Review	73
	2. Health Literacy Questionnaire	73
	3. Training: Clinical Vignette Review	73
	4. User testing.....	74
	5. System Usability Scale Questionnaire.....	75
	6. Semi Structured Interview	75
3.6.3.	Participants.....	75
3.6.4.	Data Analysis.....	75
	User Testing.....	75
	Thematic Coding.....	76
3.7.	Chapter Summary	76
Chapter 4. Results & Discussion		77
4.1.	Results	77
4.1.1.	Bridging Health Data Analysts and Caregivers	77
4.1.2.	Bridging Health Consumers and Health Data Analysts	79
	Participant’s Health Literacy.....	79
	User Testing Results.....	79
	Qualitative Coding.....	82
4.1.3.	Understanding Health Consumer and Caregiver Communication	84
4.2.	Discussion.....	85
4.3.	Understanding and Designing for Stakeholders.....	85
4.3.1.	Stakeholders Analysis	85
	Analysts	85
	Health consumers	90
	Caregivers	91
4.3.2.	Designing for Health Consumers	92
	Reporting on Medical History and Symptoms.....	92

Accounting for Vulnerability	95
4.4. Data Provenance	96
4.4.1. Minimum Data Needed	99
4.4.2. Technical Limitations in the Generation of Data	100
4.4.3. Data Mining Limitations	101
4.5. Qualitative Hypothesis	102
4.6. Supporting Health Literacy	104
4.6.1. Automation	104
4.6.2. Accuracy and the Issues of Self-assessment	106
4.7. Prioritizing Interactivity and Visual Feedback	111
4.7.1. Facilitating Health consumer Communication	111
4.7.2. Usability	115
Usability Recommendations	118
4.8. Quantitative Hypothesis	119
4.9. Research Limitations	120
4.9.1. Threats to Validity	120
The Caregivers' Perspective	121
4.9.2. Methodological Limitations	122
Thematic Coding	122
Experiment	122
4.9.3. Design Limitations	122
Mapping the HPO	122
Defining Diseases and Categories	124
4.10. Overarching Hypothesis	124
4.11. Chapter Summary	127
Chapter 5. Conclusion & Future Work	129
5.1. Summary	129
5.1.1. Chapter breakdown	129
5.2. Contributions	133
5.3. Future Work	134
5.4. Chapter Summary	135
Epilogue	136
Becoming Data: Part 2	136
Back to work: Human Empathy to Data	136
References	138
Appendix A. Ethnographic Interview: Guiding questions	153
Appendix B. Coding Scheme: Ethnographic Interviews	154
Appendix C. User Testing: Researcher's guide	157
Appendix D. User Testing: Clinical Vignettes	158

Appendix E.	User Testing: Clinical Vignette’s Glossary	160
Appendix F.	User Testing: Health consumer History Form	161
Appendix G.	User Testing: HLS-EU-Q16 questionnaire	163
Appendix H.	User Testing: System Usability Scale.....	165
Appendix I.	User Testing: Structured Interview Questions	167
Appendix J.	Coding Scheme: User Testing.....	168

List of Tables

Table 3.1 Study participants and data collection summary	58
Table 3.2 Wicked Problems in Precision Medicine	61
Table 3.3 Activity Theory in a Nutshell	63
Table 3.4 User Testing Participants Summary.....	75
Table 4.1 Codes to Design Guidelines	78
Table 4.2 Measures of Intercoder Agreement.....	79
Table 4.3 Participant's Health Literacy	79
Table 4.4 Study Summary	81
Table 4.5 Experiment's Descriptive Stats	81
Table 4.6 System Usability Scale Summary	82
Table 4.7 Confirming Participant Understanding	82
Table 4.8 Analyst's Digital Artefacts	88
Table 4.9 An Asynchronous Conversation between Health consumers and Analysts..	126

List of Figures

Figure 1.1 The Four Nested Levels of Vis Design (Munzner, 2014).....	12
Figure 1.2 Stanford's 5-Step DT model	14
Figure 1.3 Diagram of the Causes of Mortality in the Army in the East	15
Figure 1.4 Robert Baker's cholera map of Leeds (1833).....	16
Figure 2.1 Stanford's 5-Step DT model	25
Figure 2.2 Nielsen-Norman DT model	26
Figure 2.3 Schön's Reflective Learning Cycle	29
Figure 2.4. Brian Fisher's Design and HCI Framework.....	30
Figure 2.5 Bob Spence's The Broker.....	31
Figure 3.1 Mapping the HPO to the body graphic.....	65
Figure 3.2 Mapping the HPO to body categories (V1.0)	66
Figure 3.3 Mapping the HPO to the prototype categories (V2.0).....	67
Figure 3.4 The Prototype's Colour Palette	69
Figure 3.5 Cardio Button with Visual Feedback.....	71
Figure 3.6 Prototype's Introductory Scene.....	72
Figure 3.7. Interactive Visualization Prototype.....	72
Figure 4.1 Prototype vs Control Score.....	80
Figure 4.2 Three Words That Describe the Analyst's Work	86
Figure 4.3 Engeström's Activity Theory model applied through the lens of Kuutti & Arvonen for health consumers filling medical history forms.	94
Figure 4.4 Score & False Positives in the Prototype and Control.....	106
Figure 4.5 Confirming User Understanding	109
Figure 4.6 Health Literacy vs Prototype Score.....	112
Figure 4.7 Prototype Score vs False Positives Scores	113
Figure 4.8 SUS Score vs Time (Size Adjusted by Prototype Score)	115
Figure 4.9 SUS Score vs Time (Size by Prototype Score, Colour by HL)	116

List of Acronyms

CSCW	Computer-Supported Collaborative Work
DT	Design Thinking
EHR	Electronic Health Record
EMR	Electronic Medical Record
GWAS	Genome-Wide Association Studies
HCI	Human-computer Interaction
HL	Health Literacy
HPO	Human Phenotype Ontology
LAC	Library and Archives Canada
PHR	Personal Health Record
QA/QC	Quality Assurance & Quality Control
SFU	Simon Fraser University
SME	Subject Matter Expert
STS	Science and Technology Studies
SUS	System Usability Scale
VA	Visual Analytics
HIBAR	Highly Integrative Basic And Responsive

Glossary

Biomolecular Data	An aggregated data set conformed of omics data
Genome	An organism's complete set of DNA.
Genome Wide Association Studies	Large research projects that evaluate thousands of genomes looking for variants that are present in greater numbers in subjects with a given disorder than in subjects without a disorder.
Genotype	The part of the genetic makeup of a cell, and therefore of any individual, which determines one of its characteristics.
Health Consumer	Individual who purchases or consumes digital health products
Health Data Analyst	Data analyst who uses biomolecular data to find correlations between the literature, the health consumer's EHR and actionable recommendation
Human Phenotype Ontology	A standardized, controlled vocabulary [that] allows phenotypic information to be described in an unambiguous fashion in medical publications and databases
Omics	The sciences that study the molecular components of biology: proteomics, metabolomics, and genomics
Phenotype	The composite of the organism's observable characteristics or traits.
Precision Medicine	An emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person
Wicked Problems	A class of social system problems which are ill-formulated, where the information is confusing, where there are many clients and decision makers with conflicting values, and where the ramifications in the whole system are thoroughly confusing

Preface.

Becoming Data: Part 1

I hesitated about leaving the warmth of my bed sheets. I could already feel the cold winter day on the West Coast and while I was putting my clothes on; I tried not to think about breakfast. I needed an 8 hour-fast for the blood sampling later that day and my stomach never understood reasons for such an ordeal. I was excited yet uneasy about the test. I had just started a new job, and along with the new skills and experience, one of the challenges was a one-way 90-minute-long commute. A simple commute that embodied my transformation into a data chimera, and I set about preparing for this.

The Commute: Human as Data Chimeras

The journey to the office started with a ten-minute walk breathing fresh morning air thick with the river's humidity. Music filtering through my cabled earphones served as my companion on these walks. I almost never waited more than a couple of minutes once at the station until a train followed by a bus transported me to the other side of town. The long ride focused me and allowed for some planning. Previous experience showed me that trying to read research papers almost always ended in yawns and glassy eyes. So early in the morning, the "knotty and strange, remote and insular, technical and specialized, forbidding and clannish" (Rothman, 2014) nature of academic language was too much of an obstacle, especially without coffee. Nonetheless, theory shouldn't need to hide behind the fences of academy but reach outside of it through *undisciplining the mind*, as an array of narrative-based ethnographies have shown (Elliott, 2016). Further, storytelling in Indigenous traditions demonstrates that the subjective nature of the story is not a limitation, but one of oral history's primary strengths. It is relevant to note that history and anthropology stem from oral tradition (Cruikshank, 1994).

The countdown to a profound overview of myself started when I came to join a precision medicine company in the fall of 2018. During my time working there, I transited 64 km on average per day, resulting in roughly a 3-hour commute according to my GPS data. Google Maps' algorithm always mediated my decision whether to take public transit or not. Yet, it would be disingenuous for me not to acknowledge that as I was training the algorithm, I simultaneously provided variables about how fast I walked, my daily routine and the city's traffic patterns. This was willingly and freely provided data, and thus I was also creating a digitalized version of myself.

Additionally, the interface's layout had been normalized in my daily life to the point that the user data mining hid in plain sight (Pasquale, 2015). I experienced the epiphany of realizing I was a pawn, "Just as the Greek chimera was a mythological animal that was part lion, goat, snake and monster, the [Echo] user is simultaneously a consumer, a resource, a worker, and a product." (p. VI) (Crawford & Joler, 2018). At the same time, how could I refuse the convenience of such an intelligent travel companion? Why should I look for an alternative when Google Maps was the King? It was certainly easier to use it than to look at timetables and bus schedules.

Just by using my cellphone, I was constantly providing more numbers that were added to my quantified past (Elsden, Kirk, & Durrant, 2016). In that sense, the more that western society relies on search engines and social networks to find what is wanted and needed, the more fluency those systems will wield in our choices and our realities. How much *me* was the digital profile being created based on my web searches and patterns? How was my personalized access to the internet via Google affecting these search results? How did the digital 'me' affect the life circumstances of the 'physical' me?

Looking one last time at my cellphone to confirm my arrival at the correct location, I closed Google Maps and entered the building. The thing is, I had granted Google permission to access my GPS position and app usage a long time before that moment, so even without seeing or actively touching my cellphone, I was still on duty as a data chimera.

The Waiting Room: Human-Data multisensorial relations

Skimming through a *Royalty Magazine* in the waiting room, I kept thinking about the point of no return that having a biomolecular profile with genetic data represented. It was more than discovering health risks in a report, but rather coming face to face with the fact that my raw biomolecular data now existed. It had left my bodily domains and was now up for grabs, if someone really wanted it (Regalado, 2019; Russell, 2019).

While trying to read, I played nervously with the texture of my cellphone's skin, spinning it between my left hand's thumb and middle finger, and making the accelerometer of the device dizzy. The blood pressure sensor in my smartwatch diligently reading my pulse every few minutes; one could say that the more smart devices attached to the body, the more efficient the chimera. With all the "voluntary" data collection that surrounded my connection to the digital world, it all boiled down to one question: if the creation of my quantitative past was invisible, was it also intangible?

Rite of passage

I returned to my senses when I was called into the blood sampling room. By this time, I was rubbing my palms against each other in an effort to evaporate the sweat building on them. The nurse advised me to look away if I was uncomfortable. I declined. I wanted to see so I could take a deep inhalation as soon the syringe pierced my skin. A short spike of pain was quickly followed by the Vacutainer tubes being filled with my blood: my raw, unprocessed data that was just beginning its journey from a bright red, organic liquid to intangible, invisible data. Not surprisingly, this was quick. A small pinch was compressing the essence of my health to later extract that data into the container of my digitalized self: an electronic health record. After another long inhalation, I saw the needle recede from my skin and leave it with the sensation of a slight pinch. A band-aid was now in its place, leaving only a ghost of the sudden pain.

I felt the relief of being done with the blood collection and my stomach not-so-quietly reminded me about breakfast. I knew I had been wanting to get the test done for so many years, but while putting pressure on the round band-aid, I was not completely sure that I still wanted it. I left the laboratory and headed back to the office, but not without consulting Google Maps first. On my way to the bus station I grabbed some breakfast, and

after finding a seat in the bus, I relaxed and took a bite of my muffin. But, as someone carelessly sat besides me in the bus, I felt a sudden pinch in my arm as the glue of the bandage removed some hair from my arm... and then it clicked.

I came to realize that I had just experienced an unspoken rite of passage, one that involved turning a part of my human essence into an aggregated record in our company's database. When making a short mental checklist, I realized all the analysts had had their samples taken to expand the database, thus all the team had gone through the same process of blood collection. Furthermore, even though all samples were anonymized and given internal identifiers, we were all conscious that the data team would be analyzing the biomolecular data and building the reports. I realized that this process gave all of us a sense of vulnerability and nakedness; one beyond being stripped of clothing, and being stripped down to skin and muscles, only leaving the hollow blueprints of a human in a database.

How much was this being 'data-naked' in front of the data team an actual rite of passage in the company? Would I still be a valued part of the team had I refused to take the test? How could I resist otherwise unattainable employee benefits, especially when I desired them? How could I refuse such a test when resistance meant either distrusting the science behind the product, the data stewardship capability of the company or the western biomedical agenda?

Chapter 1. Introduction

The first step is to make an inventory of the molecules that you are made of. The sum of all this molecular information can be called the "molecular you". The second step is to store all this information in digital form. The digital version of you will be an impressive amount of data, but not that much use to you unless you can use it to answer your questions about yourself.

Pieter Cullis (2015:31)

Approaching the end of the second decade of the 21st century, mobile management of one's own health is becoming increasingly prevalent in western societies. A few clicks and keywords quickly transform electronic devices from black mirrors into colorful visual communicators of personalized health information. In this context, precision medicine has been defined as "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person" (NIH, 2020). This has been fostered by the declining costs of biomolecular testing (NIH, 2016; Shendure & Ji, 2008) and Big Data advances in analysis methods for the *omics*, a term that is used to generally define the sciences that study the molecular components of biology: proteomics, metabolomics, and genomics.

These technologies serve as foundations for a paradigm shift towards personalized and preventive medicine, increasing the need for health care consumers to directly access and interact with the digital version of their own records. This new commercial availability of biomolecular testing has provided the public with a vast amount of data, which along with the rise of computing power has led to the need for displaying complex analyses in a way that humans can understand. Most computational and algorithmic solutions are aimed at creating and refining visualizations of healthcare data for caregivers, omitting essential human factors that pertain to health consumers and account for the situated nature of cognition.

Furthermore, there remains a large gap between advances in visualization and dedicated tools for exploration (Mougin et al., 2018). In the field of precision medicine, visualizations preserve the same properties in terms of analytic value and expressiveness, yet they contain a critical addition, the ability to explore the system and how that system reacts to the user's actions. Nonetheless, even with such interaction capability of modern systems, molecular data remains vast, complex, and difficult to interpret.

In this sense, western society already has the technology to make DNA sequencing more accessible (Shendure & Ji, 2008), yet the new frontier of challenges shifts now to making both health specialists and consumers acquire insights from the data and act upon it. Shifting the focus towards the human experience has its own challenges, since understanding the initial complex raw data sets requires knowledge and understanding from bioinformatics and data visualization. Unfortunately, the average health consumer does not possess that knowledge, let alone understand it to a point that they can make decisions. Thus, the real challenge relies in being able to convey the message to a population with varying levels of health literacy. It is at this crossroad that I look to make an impact. I sought to immerse myself in the community of the Data Team through the theoretical and methodological rigour from ethnomethods in order to understand the nature of their work and their data requirements to design human-centred visualizations that can be evaluated through User Testing.

The rest of this introductory chapter is divided in three major sections. The first one presents the current challenges for precision medicine, such as the data context, the need to put consumers and analysts in communication, and framing the problem. The second section presents the field of Visual Analytics and provides relevant definitions for the present work. The third section speaks to the research setting, objectives, goals, and contributions. The chapter concludes by providing an outline of the thesis.

1.1. Challenges for Precision Medicine

1.1.1. The Disinformation Era of Big Data

In parallel with the challenge of representing complex data sets for understanding is the reduced trust in the good faith of big technology companies around what they do with their user data. This has been exemplified by the rise of data-related scandals which

derive from the misuse of algorithms which handle complex data sets. This becomes more worrying when those data sets arise from social media and other sources which ubiquitously monitor the user, such as wearables and cellphones (Elsden et al., 2016). One case by researchers from Microsoft Research and Cambridge University argued that their model could predict someone's personality based on their Facebook feed. The algorithm was advertised to be capable of discerning psychological traits such as Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (Bachrach, Kosinski, Graepel, Kohli, & Stillwell, 2012). The discussion of the algorithm treated it like a black box, with no mention of how it incorporated posts regarding health and wellbeing. With claims that would make the algorithm a quasi-diagnostical tool for mental health, it was even worse that the researchers didn't focus on the social ramifications if their tool when it could clearly be misused as a *Weapon of Math Destruction* (i.e. in correlating health conditions to risk of recidivism, access to credit or housing) (O'Neil, 2016). Later, this research served as the foundation for the Cambridge Analytica scandal (Isaak & Hanna, 2018).

The Data Context of the Omics

Precision medicine occurs in a context of a continuous iteration of biomolecular-data-related products, as consumers of AncestryDNA products were eloquently informed in September 2019: "Your DNA doesn't change, but we now have 13,000 additional reference samples and powerful, new science to give you better ethnicity results." (Charlsie & Denne, 2019). After the update, a set of clients who purchased the product in November 2018 found new ancestral ties to Russia, Greece, the Balkans, and the Baltics. As stated by Paul Maier, a population geneticist at FamilyTreeDNA: "Finding the boundaries is itself kind of a frontier in science, so I would say that makes it kind of a science and an art" Paul Maier, (Agro & Denne, 2019).

1.1.2. Catering to Multiple Stakeholders

The Health Data Analysts' Perspective

An emerging stakeholder in healthcare, the health data analyst, has the role of going through omics data and finding correlations between scholarly literature, the health consumer's , and actionable recommendations. Their role comprises the management, analysis, and interpretation of healthcare data. They are in charge of applying their subject

matter expertise to provide actionable insights (University of Wisconsin, 2020). Additionally, with clinicians being “...required to essentially-continually interact with multiple computer screens to enter/view/extract information (e.g. electronic health record [EHR], etc.)” (Mosaly, Mazur, & Marks, 2016), this reinforces the need to consider the health analyst as a relevant stakeholder among health consumers and caregivers.

The health data analyst requires omics data as raw material for the analysis, as well as extensive, and precise, health consumer-generated symptomatic information to accurately report health test results. This more complex information poses challenges for health data analysts when trying to understand and incorporate the health consumer’s experience. Additionally, the results of the analysis need to be communicated back in ways that are comprehensible by the health consumers, their caregivers, and other analysts. Therefore, the communication loop between health consumer and caregiver in precision medicine is not always explicit or presential, with digital healthcare also requiring an asynchronous communication through health records. It is through the need to turn biomolecular data into understandable reports that the loop is expanded by including health data analysts who craft such reports.

The Health Consumer’s Perspective

In terms of precision medicine, the individuals who come in contact with digital healthcare are not always suffering an injury, illness or undergoing treatment by a health professional, but might be seeking ways to maintain or better their health. In that regard, the term *patient* limits the scope of the studied population, and with the current availability of omics-related products, the health *consumer* relationship with healthcare professional does not seem to be a requisite in precision medicine and an existent illness is not necessarily a pre-requisite to use or access precision medicine (Calderon, 2020).

The capacity to understand and represent the body and its conditions is variable across the health consumer care spectrum: “For consumers with reasonable health literacy, self-monitoring offers greater autonomy, aiming to reduce pressure on health resources.” (Anderson, Burford, & Emmerton, 2016:2). Conventional medical forms for biomolecular analysis are designed for physicians knowledgeable in the format and content. The high barrier for understanding complex ontologies and terminology used by physicians can impede communication with health consumers and caregivers (Robinson & Mundlos, 2010). Confronted with a poorly understood representation, health consumers

may attempt to research their conditions independently which can add to an already confused state.

The sensitive nature of a health consumer's medical history creates privacy concerns surrounding its disclosure. This fact that is augmented when health consumers are dealing with commercial precision medicine products since the acquisition of such products involves blind faith in the company's data stewardship capabilities. While there is a drive by the health consumer to learn more about their health or to find out about potential risks, the process involves health consumers completing a medical history profile. This makes it increasingly important to empathize with the vulnerability that the health consumer is facing. Such an aspect has been widely recognized in design processes; for example, empathy acts as one of the first steps in some design thinking models. While the general public remains oblivious or lay towards the pervasiveness of data collection and profiling (Pasquale, 2015), it is important to recognize that biomolecular data is an extra layer of knowledge to their *quantified past* (Elsden et al., 2016) and that once generated it will face privacy and security concerns. Therefore, being aware of the health consumer's vulnerable state, along with a robust data steward capability are essential aspects to be considered by the health data analysts.

A typical user concern is privacy, more so because of the clinical nature of the data being entered in the system. In direct opposition is the need of the health consumer to share the data with their physician, which requires the EHR/PHR to be exported in a printable format. Such reports, once they leave the security of the server, become liabilities for health consumers (Haggstrom et al., 2011).

1.1.3. Wicked Problems

The design of precision medicine technologies is a complex task that involves multiple stakeholders who have different levels of expertise and different access to information, and the communication between stakeholders is not always direct. Furthermore, with the increase in databases housing clinical and biomolecular data, the algorithms and the science itself are in an ever-changing position (Fleck, 1979).

These kinds of complex spaces can be framed as Horst Rittel's *wicked problems*, defined as "a class of social system problems which are ill-formulated, where the

information is confusing, where there are many clients and decision makers with conflicting values, and where the ramifications in the whole system are thoroughly confusing.” (Rittel & Webber, 1973:15). Therefore, technology designers in a rapidly changing world face wicked problems since machine learning, AI and data visualization are starting to make sense of vast amounts of data which constantly reshape the way society looks at medical technoscientific developments.

1.2. Visual Analytics

The field of Visual Analytics (VA), defined as “the science of analytical reasoning facilitated by interactive visual interfaces” (Thomas & Cook, 2005:4) is a scientific discipline that grew out of Information Visualization (InfoVis) and Scientific Visualization (SciVis) to become an interdisciplinary discipline that is mainly informed by computer science and cognitive science, placing a focus on the human as the center of research. The definition was later expanded: “Visual analytics combines automated analysis techniques with interactive visualisations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets” (Kohlhammer, Keim, Pohl, Santucci, & Andrienko, 2011:118).

Further, under the umbrella of *analytics*, VA is “highly interdisciplinary and combines various related research areas such as visualization, data mining, data management, data fusion, statistics and cognition science (among others).” (Kerren, Stasko, Fekete, & North, 2008:156). Finally, VA as a multidisciplinary field has the following focus areas (Thomas & Cook, 2005):

- Analytical reasoning techniques that enable users to obtain deep insights that directly support assessment, planning, and decision making
- Visual representations and interaction techniques that take advantage of the human eye’s broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once
- Data representations and transformations that convert all types of conflicting and dynamic data in ways that support visualization and analysis

- Techniques to support production, presentation, and dissemination of the results of an analysis to communicate information in the appropriate context to a variety of audiences.

In that sense, VA systems can harness the calculation ability of machines and expertise from computer science along the understanding of how humans think and analyze from cognitive science. Further, with analytics covering the mining of data, its management, and fusion, VA systems are not solely restricted to representing data, but can also serve to analyze the data infrastructure and capture data by studying human activity through looking at situations where humans think through computers.

Methods from VA can support integrative analysis that consider both the health data analysts' and health consumers' needs in an information communication loop. By doing so, it is possible to focus on data infrastructures that contain controlled vocabularies to facilitate the communication between both parties. In precision medicine, such controlled vocabularies can be found in the form of clinical data ontologies (Glueck et al., 2018; Köhler et al., 2017). From the analyst perspective, VA can be a field which bridges the results from the data-intensive algorithms and tools that support understanding. It is only through visualizations that sense can be made from the data, so patterns can become insights. It is this cascade of "Eureka" moments where analysts can detect the expected and discover the unexpected (Thomas & Cook, 2005), which eventually allows scientists to claim their algorithms *can* predict behaviors. From the health consumer's perspective, VA can provide frameworks to account for their experience in the design of analytical-support systems for analysts.

1.2.1. Design of VA Systems

When focusing on methodologies and methods in VA, the initial VA literature missed considering sociological and anthropological social science methods that accounted for the situated nature of human cognition. In that sense, Lucy Suchman was one of the pioneers in bringing methodologies of the social sciences to the study of people interacting with technology. In her exploration of the use of technology at Xerox PARC, Suchman found that: "Standard procedures are formulated in the interest of what things should come to, and not necessarily how they should arrive there" (Suchman & Woolgar,

1989:326). In the same line, human factor-based design was also proposing the design of artifacts that were usable and useful for the intended audience (Tory & Möller, 2004).

Therefore, the expansion into the social sciences from VA is a key aspect, especially when shifting the focus to the design and evaluation of visualizations. While several authors such as Tamara Munzner (Munzner, 2009, 2014), Ben Schneiderman (Shneiderman & Plaisant, 2006) , and John Stasko (Stasko, Görg, & Liu, 2008) recognize the value and importance of social sciences, specifically ethnography, as a way to understand how professionals use computers and technology, there is a lack of methodological rigour in how to conduct those types of research. One of the methodological challenges is that social sciences are treated as a black box and reduced to observation (See Figure 1.1).

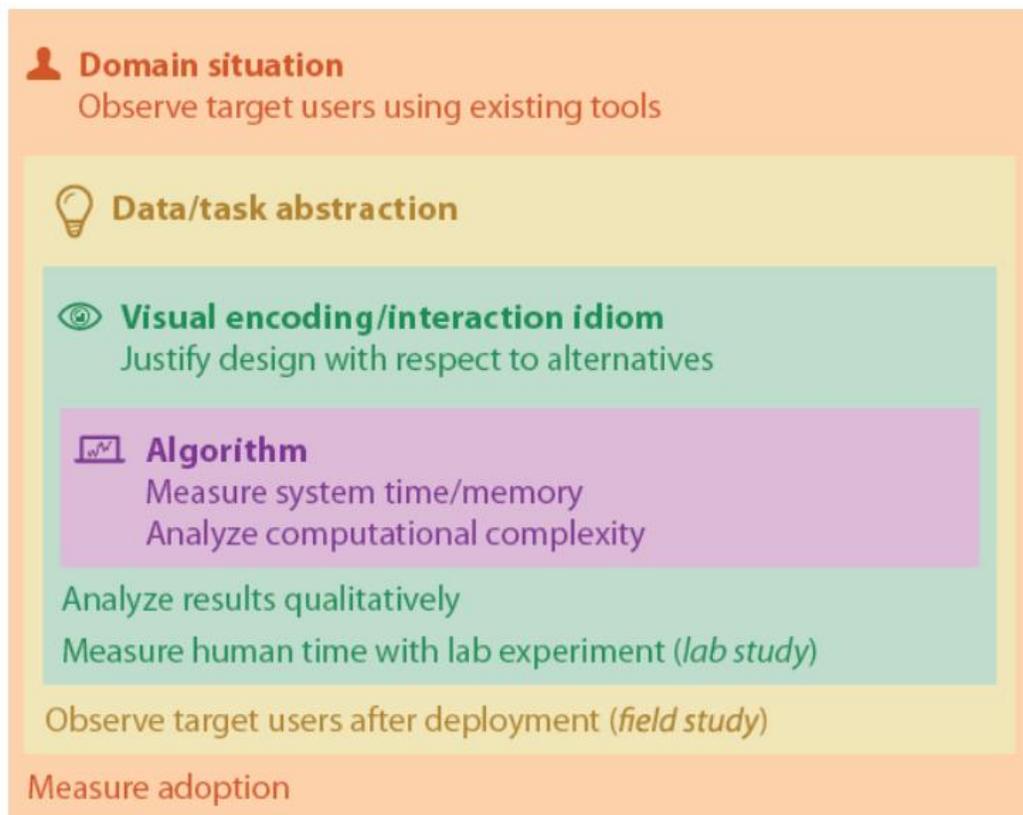


Figure 1.1 The Four Nested Levels of Vis Design (Munzner, 2014)

In a more caricaturesque way to put it, the incorporation of methodologies and methods from anthropology and sociology brings to mind Hunt-Lenox Globe's famous phrase where the social sciences are labeled as obscure places where "Here are dragons" (Stevens, DeCosta, Johnston, Lamb, & Pond, 1879:536). Furthermore, as a human-centred field, the validation of a visualization is a difficult task since there are a plurality of aspects that could be considered for the assessment, for example, how to define *better*, *effectively*, *engagement* or *insight*; more importantly how to define the *user* or what *quality* means (Munzner, 2015). It is in that problem space that Visual Analytics' focus on the human aspects of computation becomes more relevant for the design of interactive visualizations.

Design frameworks for VA systems should not be constrained by a single focus on the methodologies and methods from only one discipline (i.e. computer science) but include expertise from other disciplines, specifically design approaches that provide a more holistic perspective. In that sense, some of the critiques surrounding visualization (Lorensen, 2004) and the lack of application in visualization (Weber et al., 2017) can also be observed in Visual Analytics, especially the fact that the refinement of VA systems occurs from an algorithmic perspective (van der Maaten & Hinton, 2008) rather than by catering to the target audience's needs. In that regard, Cognitive Science's roots in anthropology and psychology provide frameworks to understand the human experience and the tools that surround it, more specifically, to analyse computers as cognitive tools that serve a specific purpose in a certain activity. In the same line, it is the social sciences (e.g. ethnography) that provide methods for centring the analysis on the study of human culture and the surrounding tools, allowing for better understanding the context of use to design systems that will match the final user's expectations and support their analytical needs through visual representations.

If trying to avoid the same critiques around visualization, then VA as a discipline needs to reflect on Lorensen's critique: if Visual Analytics wants to avoid death, now would be the time to go back to VA's roots as a highly interdisciplinary science. By doing so, it is possible to introduce new, interdisciplinary research that develops more robust applied science by building upon the current cognitive models and applying valid cognitive theories to produce engaging VA systems. Furthermore, the qualitative research of the human cognitive experience and skills of those individuals who will use the systems is important to address individual differences in cognition (T.M. Green & Fisher, 2011). Even more

important is developing interactive systems that help health consumers and specialists to make informed health decisions based on their biomolecular makeup.

It is in this sense that a pragmatic design framework, such as Design Thinking, can be a valuable addition to the methodological repertoire of Visual Analytics and can become essential for the design of human-centred technologies. The term *Design Thinking* (DT) was originally coined by Peter G. Rowe in his book *Design Thinking* about architecture, and its meaning has evolved to become widely used to address human concerns through design and being represented as a series of hexagonal steps (Jen, 2018) (See Figure 1.2). A Design Thinking framework can provide a better understanding and guidelines into how qualitative and quantitative research can be pragmatically used to design a Visual Analytics system centred on the human experience that allows health consumers to report and understand their own health.

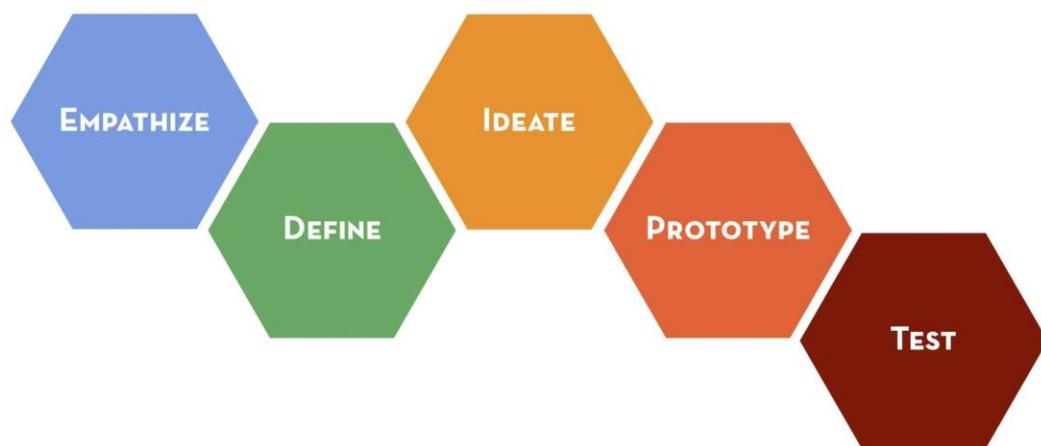


Figure 1.2 Stanford's 5-Step DT model

A more encompassing design thinking framework also begs the need to go back to VA's roots as a multidisciplinary discipline that bridges different knowledge domains to pragmatically solve problems that lie at the intersection of human cognition and computing. This would require embracing VA's nature as a "highly application oriented discipline driven by practical requirements" (Keim & Zhang, 2011:2). In that sense, VA can become a discipline to conduct HIBAR research that bridges basic and applied research

in service of society (The HIBAR Research Alliance, 2020). It is in that space that the design of digital healthcare technologies can be supported through VA systems.

1.3. Definitions

1.3.1. Visualization

From a historical perspective, symbolic and visual communication has given birth to graphic design, which has been prominent in books and magazines, scientific illustration, photography, cinema, television and, more recently, computer graphics (Buchanan, 1992). Data visualizations, in the way of static images, have been shown to have analytic and expressive values (Norman, 1993; Tufte, 1983).

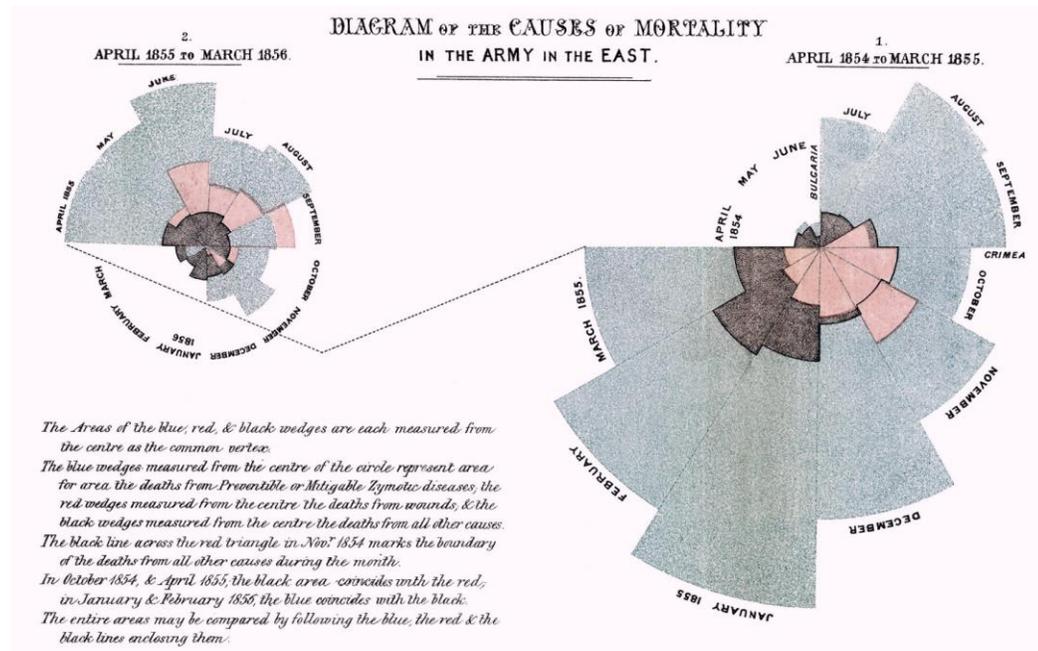


Figure 1.3 Diagram of the Causes of Mortality in the Army in the East

In the medical context, data visualizations have had a long-lasting presence and allowed for better understanding of phenomena, with some of the first recorded data visualizations coming from epidemiology, for example Florence Nightingale’s “Diagram of the Causes of Mortality in the Army in the East” (West, Borland, & Hammond, 2014) in

Figure 1.3 or Robert Baker's and John Snow's geovisualizations tracking cholera outbreaks in the UK from the 1830's to the 1850's (Friendly, 2008) in Figure 1.4.

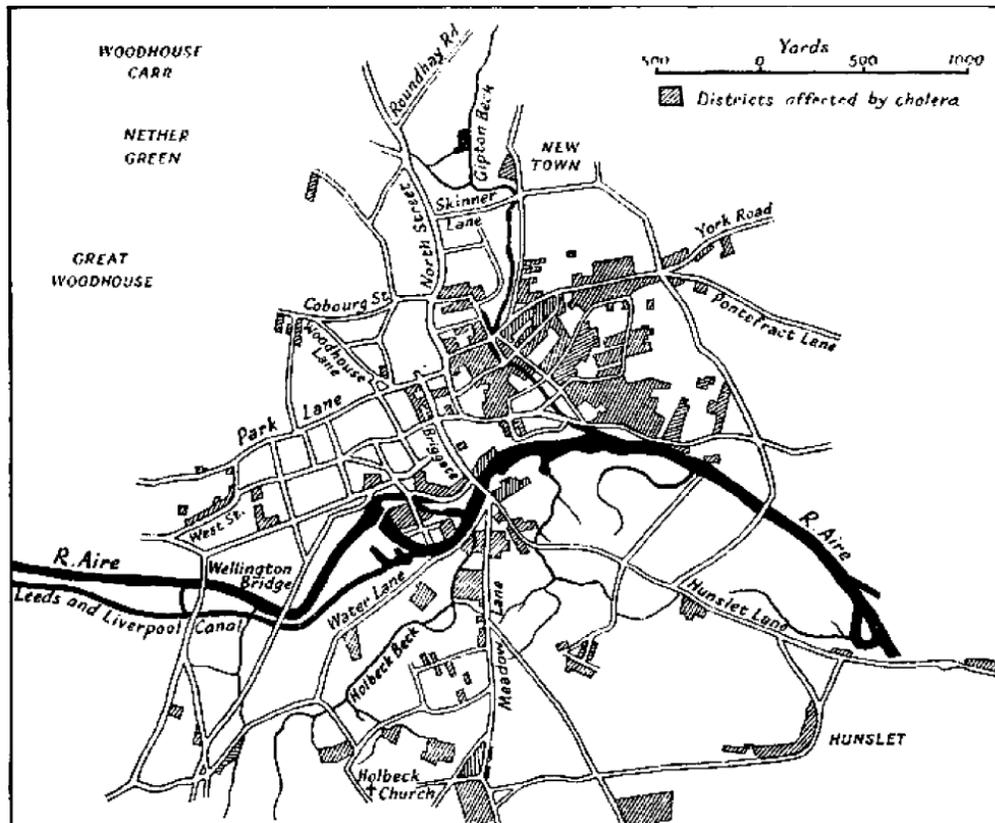


Figure 1.4 Robert Baker's cholera map of Leeds (1833)

As stated by Peters, language is “one of the key epistemological materials of which the world is made” (B. Peters, 2016); it is what shapes the way reality is constructed. While there is wide range of meanings and interpretations of what a visualization is, for the present work the term is understood in the context of HCI research, specifically from a psychological lens: A visualization is an external representation that harnesses the human innate visual intelligence to understand and generate insights from cases characterized by the complex nature of the data: “[Cognitive] mediators are external elements that facilitate actions or operations” (Arias-Hernandez, Green, & Fisher, 2012:13).

1.3.2. Biomolecular & Descriptive Data

The data generated by the omics is usually considered as big data, which creates challenges for its analysis and management in terms of size and complexity. For the present thesis, the aggregation of data sets stemming from the omics will be referred to as multi-profile molecular data, or more generally, biomolecular data. One aspect to take into consideration is the nature of data as being a proxy of “real” phenomena. In this case they represent concentrations of “biomarkers”, or organic molecules found in the human body. Additionally, it is understood as descriptive data or descriptive health consumer data as the information regarding the health consumer’s medical history, symptoms, and other conditions that affect their health.

1.3.3. Digital Records: EHRs, PHRs and MHRs

In terms of health records, there are three relevant types: Electronic Health Records (EHR), Electronic Medical Records (EMR) and Personal Health Records (PHR). While more recently the first two have become synonyms, there is a clinical distinction between them. EMRs are the digitalized versions of the paper charts, whereas the EHRs are comprehensive records that “contain information from all the clinicians involved in a health consumer’s care” (Office of the National Coordinator for Health Information Technology, 2019). PHRs relate to those that are designed, accessed, and managed by health consumers. It is important to mention that with the increase in availability of commercial molecular diagnostics products such as those sold by 23andMe, Counsyl, Natera in the United States and Microbiome Insights and Molecular You in Canada, EHRs are becoming a part of the health consumer experience and the line between EHRs and PHRs becomes less clear. This is even more so when the health data analyst becomes a relevant agent in the communication loop between health consumers and caregivers.

EHRs¹ are “the tools through which physicians, nurses, and other staff, enter, view, and share the information required to deliver high quality care” (Armijo, McDonnell, & Werner, 2009:8). They are designed to provide instant access across networks of caregivers and secure data housing for better care, yet they further perpetuate the

¹ For the present thesis, the interactive visualization is considered as an essential aspect of the design and evaluation of EHRs. I will make a distinction between EHR, EMR and PHR when needed.

depersonalization of health care (Duke, Frankel, & Reis, 2013). While serving as a valuable communication medium for health care workers, the EHR lacks the capacity for dialogue with the consumer. Nonetheless, they have the advantage of being part of the healthcare infrastructure in contrast to PHRs that operate mainly for the consumption and use of consumers, rather than for health professionals. Adding to the communication challenges of EHRs is the increasing complexity of medical tests and their interpretation (West et al., 2014).

1.4. Research Overview

1.4.1. Research Setting

The present thesis work was conducted in a Startup that focuses on personalized health. Their products involve a range of analyses that derive from biomolecular data, specifically genomics and proteomics, with the mission of using that data to provide actionable recommendations to their clients. The promise is to identify early signals of disease onset and wellness to provide the an individual personalized health plan.

The company was comprised of five major areas: 1) Marketing & Communication, 2) Customer service, 3) Software Development & Visualization, 4) Design and 5) Data Team. While most employees held full-time positions, there was freedom in working from home and accommodating one's own schedule, meaning that desks were mostly shared between people. Additionally, depending on the day and time, the amount of people at the office fluctuated, with Mondays and Tuesdays being the busiest due to meetings that involved inter-team work. I spent most of my time in a desk in three places, in the data team's director office, in the software development & visualization office, and near the design team's space.

Furthermore, the present research was conducted under the umbrella of a bigger project aiming to conduct translational research in visual analytics for personalized health. A thorough analysis of the relation between expert analysts and health consumers in direct-to-consumer precision medicine products was conducted by Calderon (2020), specifically in terms of the sociotechnical system where the interpretation of biomolecular data occurred. Through pair analytics and in-vivo studies in sense-making, Calderon analyzed the process of interpretation, documenting the different roles that health analysts

need to undertake when going through the data, what is considered evidence for the analysts, and the tools they employ for their analysis. Through the process, analysts encountered the issues of incomplete health consumer data and communicating uncertainty and complex evidence to a lay audience. The present thesis is guided by Calderon's recommendation for future work (Calderon, 2020:149):

“Future research will continue to investigate the digitalization of the experience of health, and its implications in the provision of care in the personalized health paradigm. [...] Phenotypes, or expressed health information in the form of symptoms, history of disease, lifestyle habits, or what is consider the biography of an individual, play a key role in the analysis of biomolecular data. More research in how to better capture a psychosocial picture of illness and health goals, will have an impact moving forward with more DTC services”

The focus of this research is to expand on the work by Calderon by focusing on the health consumer's perspective while accounting for the data needs of analysts. The aim is to design an interactive visualization that supports data and analytic needs of both audiences: health consumers and caregivers.

1.4.2. Research Objectives

- 1. Design and evaluate a health consumer-facing interactive system that supports the health data analysts' needs by mapping a medical data ontology as a visual controlled vocabulary**

I focused on gathering ideas that could address the data needs of analysts while considering the vulnerability of health consumers and their user experience with precision medicine technologies. I began with a literature review of data ontologies that were being used in the medical community, evaluating my findings based on the coverage of disease phenotypes and capability for scaffolding automatization and the implementation of algorithms (i.e. Genomic Wide Association Studies). The goal was to implement a data ontology as the database backbone for the health consumer-centred visualization. The ontology would serve two purposes, the first one as the data infrastructure for the health consumer-centred visualization, and the second to allow data scientists to visualize the health consumer's medical history and how the conditions and symptoms relate clinically.

For the design process, I focused on the creation of a prototype that was informed by a critical perspective of the data-driven healthcare system and the depersonalization of healthcare as a way to incorporate the industry requirements and a health consumer-centred philosophy in the design of visualizations for analytical purposes. The resulting prototype was evaluated through an experiment (User Testing) to assess whether the visualization helped health consumers to accurately report on their medical history.

2. Investigate the use of adapted ethnographic methods through a distributed cognition lens for the design and evaluation of visual analytics system for precision medicine in the context of interpreting biomolecular data.

I worked as a part of the company's data team. Being embedded in the organization, I took part in the team's meetings and activities to understand their daily activities, their challenges with creating the databases required for the analysis, and their experiences analyzing the biomolecular data. I utilized Nielsen-Norman's design thinking framework (See Section 2.1.2) to achieve empathy through ethnographic methods and put a focus on understanding the analysts' experience with analyzing biomolecular data. For this, I took field notes, conducted literature research to identify the constraints and critical aspects that needed to be present, and carried out ethnographic interviews with the analysts. The objective is to use an ethnographic approach to understand the health consumer's journey, along the data analysts' work, to design an interactive interface that is rooted in a data infrastructure that allows for catering to both audiences.

1.4.3. Research Goal and Contributions

The present thesis aims to provide an interactive data-capture system for health consumers that supports the data needs of health data analysts. By mapping a medical ontology to a graphical depiction of the body, health consumers can report on their conditions through a visual controlled vocabulary that supports automatization for precision medicine technologies.

Digital interfaces are becoming the standardized approach that western healthcare systems are using to deal with health consumer data (medical histories), yet they offer a filter-like barrier to modern medicine since only those few with medical literacy (Nolasco

et al., 2010) can harness the benefit of the health system. From a different perspective, the data captured through digital interfaces are the primary resource of data analysts to conduct biomedical profile analysis. If these systems are seen as a two-sided tool, then that lens allows for tracking the journey of it as both health consumers' tools and analysts' resources. In turn this allows for gaining a better understanding of what these systems represent to both populations and how data ontologies can be used as the infrastructure that supports this indirect interaction and communication between both of them.

The goal of the research work is to show how to incorporate ethnographic methods in a Design Thinking framework, specifically by departing from a superficial understanding of empathy in DT. This is achieved by incorporating short-term ethnography to gain rich descriptions of the experience of health consumers and analysts for the design of user-interfaces, and, more importantly, by designing interactive systems that allow health consumers to accurately report their data in a way that health data analysts can conduct analysis and create algorithms. Through the use of short-term ethnography to directly inform the design of VA systems, it is possible not to lose sight of the evolution that anthropology and ethnography have experienced, that means avoiding an oversimplification of the methodology and critical nature of the research.

1.5. Thesis Outline

This thesis is organized in 5 chapters. Following the introduction to and motivation of the research provided in chapter 1, chapter 2 contextualizes the research in relation to previous work in the area. Chapter 3 presents the research philosophy, methodology and methods, and the design of the interactive visualization. Chapter 4 deals with the results from the mixed methods inquiry, then reflects on how the design guidelines were created and discusses how the results from the qualitative and quantitative inquires triangulated the information; it finishes by addressing the research limitations. Chapter 5 focuses on conclusions and summarizes the thesis work and contributions.

1.6. Chapter Summary

Chapter 1 served as an introduction to this thesis, presenting the (dis)information context that health consumers face, the reduced trust in digital health products, and how health literacy becomes a barrier to access precision medicine products. This begs for

design to consider the perspectives from health data analysts and consumers, and to understand the *wicked* nature of HCI and precision medicine challenges. The field of Visual Analytics can provide frameworks to tackle such design problems by conducting highly interdisciplinary research. In the case of this thesis, by using a research-driven Design Thinking methodology through immersion in a precision medicine company to design an interactive data capture system that supports automation, and secondly, to curate methodological and design guidelines for Visual Analytics systems for precision medicine. Chapter 2 will continue by showing the supporting literature and main ideas that informed the research presented in this thesis; it focuses mainly on Design Thinking, Ethnography and HCI, and Visual Analytics Design considerations.

Chapter 2. Literature Review

Designers are exploring concrete integrations of knowledge that will combine theory with practice for new productive purposes, and this is the reason why we turn to design thinking for insight into the new liberal arts of technological culture².

Richard Buchanan (1992)

This chapter provides the research foundations and identifies the literature gap, synthesizing work from three major disciplines: design, ethnography, and visual analytics in the context of HCI. The first section presents an overview of HCI and Design Thinking, including its history and relevant Design Thinking frameworks and then focuses on key authors who pioneered an iterative approach to scientific thinking in design. The second section deals with ethnography and HCI; it shows how the methodology has been incorporated in the field, critical issues regarding ethnography across disciplines, and finishes by exploring adapted, short-term ethnographies and cognitive ethnography.

The final section focuses on an exploration of visual analytics design considerations for precision medicine. It starts with an exploration of cognitive theories in Visual Analytics and provides an overview of the underlying cognitive theories in the design of computational artefacts. It then moves to focus on system design frameworks, and the audiences, finishing by showing relevant data infrastructures, specifically data ontologies.

² Buchanan's understanding of art points towards science, and comes from Dewey who avoided treating science as primary and art as secondary: "The consideration that completed the ground for assimilating science to art is the fact that assignment of scientific status in any given case rests upon facts which are experimentally produced. Science is now the product of operations deliberately undertaken in conformity with a plan or project that has the properties of a working hypothesis." John Dewey, "By Nature and By Art," *Philosophy of Education (Problems of Men)* (1946; rpt. Totowa, New Jersey: Littlefield, Adams, 1958), 288.

2.1. Design Thinking

Design Thinking is a term that has shown a growth in its use, with applications being found in disciplines ranging from business to IT to healthcare. One of the main reasons for the rise in popularity is the seemingly universal applicability to any type of problem. Originally, the term was coined by Peter G. Rowe in his book *Design Thinking*³ about architecture design, and its meaning has evolved to become widely used to address human concerns through design (Jen, 2018).

2.1.1. The Origins of Design Thinking

Design Thinking (DT) as a design framework has found its origins both in academia and in business. Examples of an academic approach are Stanford's 5-Step model that defines DT as a methodology for creative problem solving (Hasso Plattner Institute of Design at Stanford University, 2020) (See Figure 2.1), and the Design Council's Double Diamond that defines it as "process of exploring an issue more widely or deeply (divergent thinking) and then taking focused action (convergent thinking)" (The Design Council, 2020).

In popular culture, it is Stanford's model that has become representative of the DT approaches, with the five steps being portrayed in different ways as any Google Images search would reveal. While this has led to the spread of the methodology and its implementations across different fields, it has also caused an oversimplification of the methodology. This has led to an overuse in the design industry to the point where the hexadiagonal diagrams have become a meme of design.

From a business perspective, a major player in the design industry is IDEO, which created its 3I's model (Ideation, Inspiration, and Implementation) that defines DT as "a human-centered approach to innovation that draws from the designer's toolkit to integrate

³ It is relevant to mention that there is considerable body of work in design thinking that predates Rowe's 1987 book. Nigel Cross points to two prior approaches, the first one in the 1920s and a second one in the 1960s (Cross, 2001). Examples for the former are *De Stijl* from Theo van Doesburg and *design studies* from Bauhaus, and for the latter can be found in section 2.1.3. Additional influential approaches stemmed from the industrial era in the 1920s in the USA (i.e. the work from Harley Earl Dreyfuss, Walter Dorwin Teague and Donald Desky), in the 1940's in Japan (i.e. Sony's post-war product design approaches) and in South America (i.e. Raul Prebín's *circumstantialist* ideology).

the needs of people, the possibilities of technology, and the requirements for business success” (IDEO, 2020). IDEO’s model follows a similar approach to the Design Council in terms of implementing convergent and divergent thinking as part of the process. In the design consulting industry, the 5-Step process has also been taken up by Pentagram, one of the world’s largest design consulting firms. Natasha Jen, a Pentagram partner, provides some critique regarding the spread of simple box diagrams as the solution to any and every problem, on her definition of DT (Jen, 2018):

[DT] packages a designer’s way of working for a non-designer audience by codifying their processes into a prescriptive, step-by-step approach to creative problem solving – claiming that it can be applied by anyone to any problem.

While it is expected that there will be overlaps in terms of the diagrams and models for DT, be it in the design industry or academia, it is interesting to note how the definitions only converge around creative thinking for problem-solving in an iterative cycle. There seems to be more of a lack of consensus in what Design Thinking really means. As stated by Buchanan, the term has evolved from being a “trade activity to a segmented profession to a field for technical research and to what now should be recognized as a new liberal art of technological culture” (Buchanan, 1992:5).

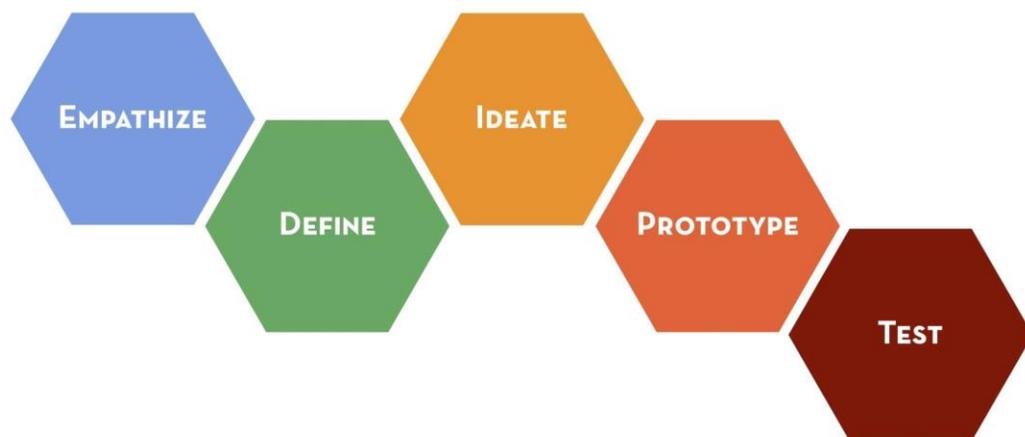


Figure 2.1 Stanford’s 5-Step DT model

2.1.2. The Nielsen-Norman Design Thinking Framework

In this context, a key design thinking framework comes from the intersection of the academic and business approaches in the form of the Nielsen-Norman Design Thinking framework (See Figure 2.2), which proposes deeper engagement of the DT model with conducting research.

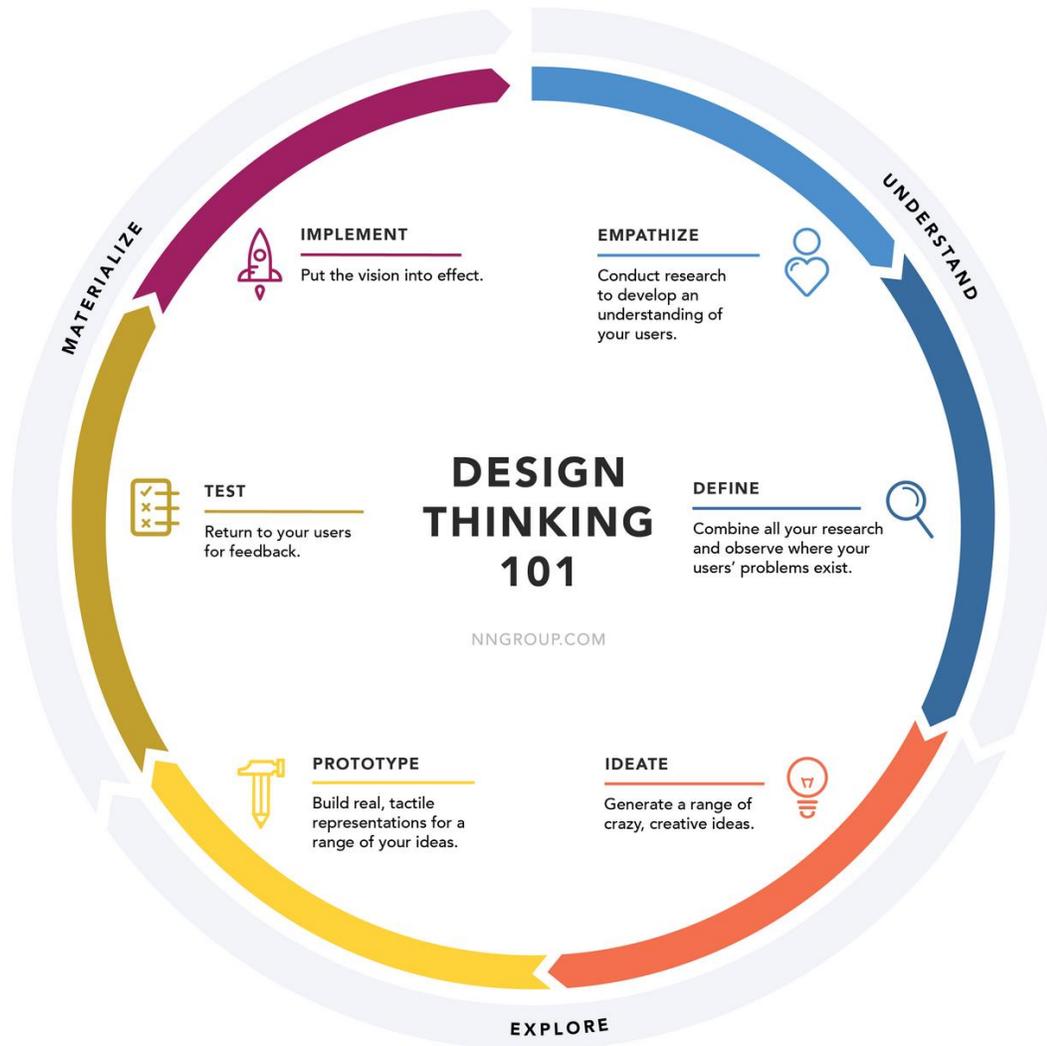


Figure 2.2 Nielsen-Norman DT model

When looking at their definition, the main contrast to other DT models is that it is quite broad: “an ideology supported by an accompanying process” (Nielsen-Norman Group, 2020). The ideology is rooted on user-centred design to problem-solving with the

goal of achieving a competitive advantage through innovation and differentiation, as for the process, it is divided in six phases that resemble Stanford's model with the addition of a last step: *Implement*. From another perspective, it is also similar to IDEO's in terms of including *Ideation* and *Implementation*.

In this evolution of DT and due to its rise in popularity, it is important to depart from oversimplifications and always maintain a critical attitude. It is up to the designer to always keep the critique, or as colloquially known "the crit", in mind. This does not mean that the box diagrams of the popular DT approaches are invalid or useless, but it means those representations are only a dimension of the methodology. Through the Nielsen-Norman approach, the HCI designer can benefit from the simple box diagrams (the process), but with a scientific approach to design that supports research methodologies.

In terms of HCI design, specifically for VA systems, the critique becomes even more important since it requires a critical approach for different perspectives or disciplines. When focusing on the design of digital healthcare technologies, incorporating the critique becomes even more challenging for precision medicine since computer science, medicine, kinesiology, and nutrition have different methods to gather and interpret evidence. This raises the need to look at similar frameworks that have been implemented across HCI to then be able to frame the task of designing and evaluating interactive visualizations for precision medicine.

2.1.3. Design Thinking Pioneers for HCI

The value of the Nielsen-Norman DT framework lies in the differentiation between ideology and process. As exemplified through different academic and business DT models, the process is similar among the different frameworks, but as pointed out in the previous section the critical mindset is what gives the ideology its value. In this sense, it is worthwhile to take a look at pioneers who thought of design and approached it from a robust academic perspective that involved design, technology, and humans.

In terms of pioneers in HCI and DT, the ideas of four key authors: i) Herbert A. Simon, ii) Donald A. Schön, iii) W. Brian Arthur, iv) Donald A. Norman, allow for creation of a holistic framework for HCI that has a scientific nature, is rooted in professional expertise, allows exploration beyond discipline boundaries and focuses on the design of

technology. Such a framework is critical for HCI since every design will have an impact on the target audience, and possibly society. In terms of healthcare, the impact will be on an already vulnerable population; it is in this sense that HCI design for healthcare is a practice or a craft, and it has an ethical responsibility and accountability to its users.

Herbert A. Simon

An iterative approach to design has been prevalent in scientific communities since Design Thinking pioneer Herbert Simon proposed that being a designer was not restricted to those with formal training and that in a way, everyone who “devise[s] courses of action aimed at changing existing situations into preferred ones”(Simon, 1996:111) is in fact a designer. Just as natural science uses the scientific method to build theories about the world and evaluate whether they are true, design could be conceptualized as a “Science of the Artificial” that uses design methods to improve a situation with the goal of ensuring that a design works in any likely set of external conditions and situations. Simon proposes the use of a Means-End analysis that compares a design relative to the goals for the thing designed, and “Satisficing” that happens when a decision is “good enough” to implement it in the design (Simon, 1996a).

Donald A. Norman

When focusing on HCI, Norman’s cognitive science approach allows for analyzing and evaluating the interaction between the user and technology, and the context of use. One of the key concepts explored by Norman was his notion of “affordances” as perceived and actual properties of objects that hint at their possible uses (Norman, 2002). As stated by Norman:

“The field I find myself engaged in (is) neither Cognitive Psychology, nor Cognitive Science, nor Human Factors. It is a type of applied Cognitive Science, trying to apply what is known from science to the design and construction of machines.”(Norman, 1986:31).

Donald A. Schön

From a reflective perspective, Schön (1990) wrote about what designers should do to improve. He proposed an iterative framework based on actively seeking out information, translating it into a form that can inform the designer’s practice, implementing it, and evaluating its effectiveness (See Figure 2.3). Schön’s understanding came from

Hans-Georg Gadamer: "Practice required knowledge.... which means that it is obliged to treat the knowledge available at the time as complete and certain" (Gadamer, 2004:4). Therefore, in Schön's view, designers were conceived as professional experts who learned the cognitive skill of design through guidance as a practice. This can be seen in his reflective cycle involving steps of analysis, integration and evaluation, where all steps require deep professional expertise.

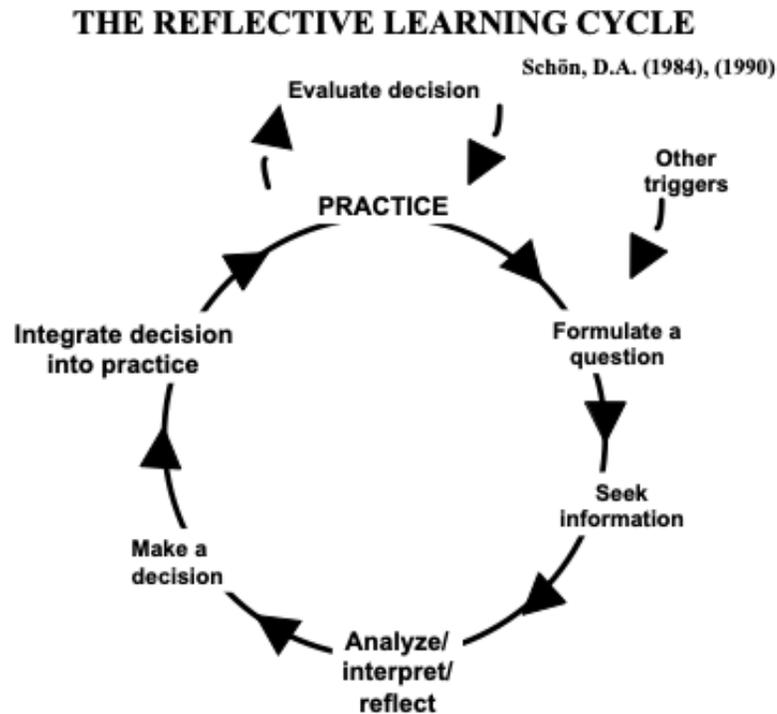


Figure 2.3 Schön's Reflective Learning Cycle

W. Brian Arthur

In terms of technology, Arthur (2009) examined how technology harnesses a phenomenon to achieve an objective. For example, when looking at a motor, internal combustion provides energy, a piston and crankshaft to turn energy into rotation and lubrication to reduce friction (Arthur, 2009). By focusing on a certain principle, technology provides new capabilities that lead to "method technologies" which can be combined to make more complex technologies. Nonetheless, such an understanding requires that designers learn the language of those method technologies. For HCI, it involves acquiring a level of language expertise that allows for understanding not only the technology but the

context of use. Since technology is embedded within human existence, it is critical to analyze how the interaction of technology and people can be designed.

2.1.4. A Cognitive Framework for Design Thinking in HCI

Simon and Schön recognized the need to improve design methods, be it through incorporation of new sources of knowledge into design (Schön, 2000) or through the creation of a design knowledge through a “science of the artificial” (Simon, 1996a). The work of Arthur serves as a guide between Simon and Schön, in that designers need to acquire a design language from the sciences. By doing so, “method technologies” from different fields, such as HCI, precision medicine, and ethnography can be combined for more complex technologies such as interactive visualization for precision medicine. Through Norman, it is possible to bridge the ideas of Simon, Schön, and Arthur through a user-centred philosophy for the design of technologies that will incorporate a human-computer interaction; this relation can be seen in Figure 2.4.

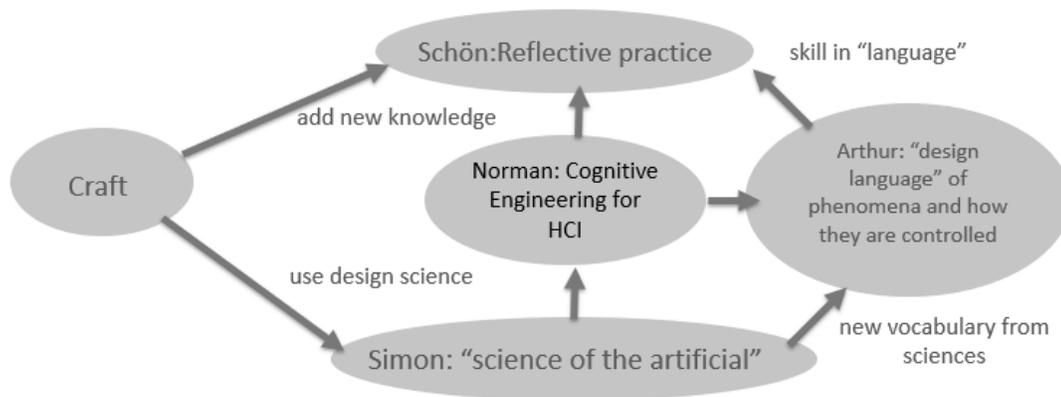


Figure 2.4. Brian Fisher's Design and HCI Framework⁴

⁴ This framework presents the relation between the main authors and ideas taught at IAT 201 “Intro to HCI and Cognition” by Brian Fisher. The course was for second-year design students, and I was involved as a TA and course designer for most of the duration of this degree.

Within the space of *wicked problems* and the use of DT approaches, achieving empathy or user understanding is a critical step. For this, designers need to depart from the oversimplification of the design thinking on polygon-based diagrams. They need to seek evidence-based approaches that support their design rationales and expand their toolkit of method technologies to better empathize with design stakeholders. This begs the need to expand the design framework by including methodologies that support naturalistic approaches for achieving user-understanding.

The Designer as an HCI Broker

Through this new framework, designers become knowledge brokers between science and design, agents who conduct meaningful reflection, and have an expanded design language that turns to method technologies to engineer more complex systems that support and distribute cognitive tasks. The HCI designer acquires the role of “The Broker” (Spence, 2011), a person who has expertise of different domains and can act as a bridge between cognitive science and design guidelines when focusing on a particular user action (See Figure 2.5). This role is important because discussions between designers and scientists often lead to confusion and breakdown in communication, without yielding productive integrations that lead to concrete results (Buchanan, 1992).

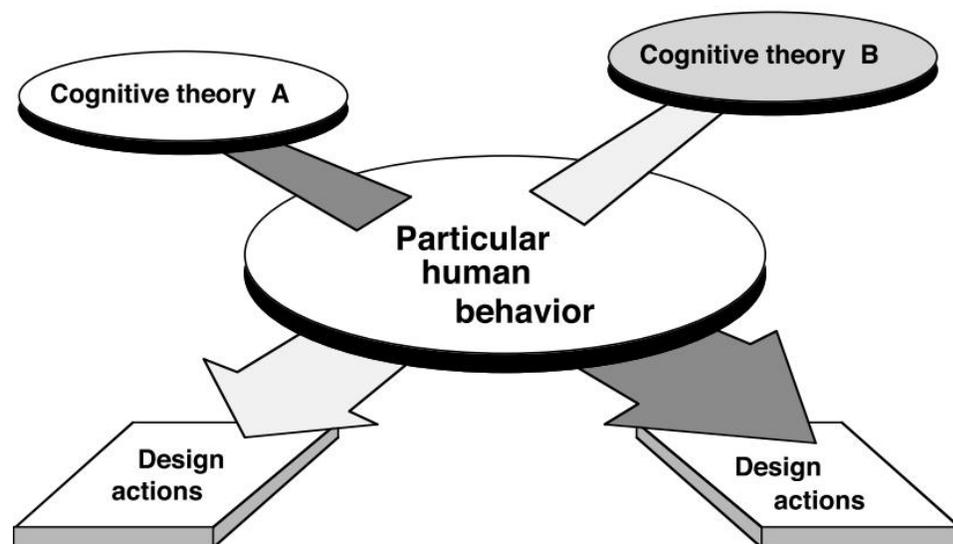


Figure 2.5 Bob Spence's The Broker

The Broker faces the need to have enough expertise from different fields to be able to interact with Subject Matter Experts, yet it cannot be expected that the designer is an expert in every field with which she gets in contact. In that sense, the HCI designer needs not to have *contributory expertise* (i.e. publishing research in that specific field) but *interactional expertise* (i.e. be able to critique the science) (H. M. Collins & Evans, 2002). Through interactional expertise the Broker can create a *trading zone*, which is an “interdisciplinary partnership in which two or more perspectives are combined and a new, shared language develops” (H. Collins, Evans, & Gorman, 2007:657). By doing so, it is possible to acquire multiple *method technologies* that can harness phenomena stemming from basic research in psychology and cognitive science into applied guidelines. This design approach should be rooted in scientific principles that support the chosen course of action. In that sense, the HCI designer can go beyond combining the ideas and frameworks from Schön, Norman, Simon and Arthur and continue the iterative reflective cycle itself.

Further, having interactional expertise can widen the scope of a trading zone beyond formal expertise to include uncertified experts that don't necessarily have an academic training but an empirical one (H. M. Collins & Evans, 2002). It is in that regard that the HCI designer needs a solid critical understanding of what comprises expertise in the studied field. This speaks again to the need to have naturalistic approaches that allow elucidating how expertise is reflected in the tasks and nature of the work, fully knowing that each trading zone between disciplines will have a certain degree of incommensurability (H. Collins et al., 2007).

2.2. Ethnography and HCI

The term *ethnography* can be used to refer to a research methodology, but also to its final product. As an outcome, an ethnography is a narrative that emerges from field immersion. It is collaborative and it has a critical nature to transform and engage with its audience. The field of Science and Technology Studies (STS) has shown that scientific inquiry and facts have been shown not to be completely *objective*, but negotiated in a social context (Fleck, 1979; Kuhn, 1963), happening along historical and linguistic processes. Colonialist ethnography has shown that a focus on text making and rhetoric highlights the constructed, artificial nature of cultural accounts (Clifford, 1986). As a research methodology or a process, ethnography is a qualitative research design that

focuses on describing and interpreting the shared and learned patterns of values, behaviours, beliefs, and language from a culture-sharing group. It involves extended periods of immersion with the culture that usually involve daily interactions, participant observation, and interviews with multiple members of the group (Creswell, 2018). As a methodology, it provides the advantage of getting rich knowledge of the meaning of the group's behaviour, language, and interactions; yet the process typically involves large amounts of data, the creation of personal relations with the studied group, and an acknowledgement of the highly subjective nature of the methodology. Derived from that, ethnography as an outcome, allows for providing a representation of reality that accounts for the subjective nature of its creation and that might be *more real* than the *true* representation. Nonetheless, the final product also requires for the researcher to know how to write in addition to having a solid critical framework of the field.

Further, when looking at human-centred design for HCI, Simonsen and Kensing argue for the need to continue incorporating ethnography in design with a proactive role (Simonsen & Kensing, 1998). One in which the ethnographer is an active participant in the design process due to their in-depth knowledge of the culture and practices, since designers “use informants or data from informants to confirm their own understanding of the situation or practice in question” (Simonsen & Kensing, 1998:21). Such an implementation of ethnographic methods was also previously seen at Xerox PARC where researchers used ethnography for technology design. In addition to Suchman and Woolgar's work (Suchman, 1989), other research lines aimed to formulate the relation between ethnographic field methods and design, specifically for the creation of artefacts that support the activities of users in their context (Blomberg, Giacomi, Mosher, & Swenton-Wall, 1993). This shows how HCI has evolved from its stance during the 1990s of relegating most aspects of the design as engineering and technical problems (Hughes, King, Rodden, & Anderson, 1995) to gaining an understanding of the value of the social context for the design of interactive systems. Additionally, ethnography has been increasingly adopted as a valuable, naturalistic methodology to elicit design requirements. This has led to the rise of adapted approaches to ethnography that are shaped by the time constraints and scale in which HCI (Hughes et al., 1995) and healthcare workers are embedded (Pink & Morgan, 2013). In these fields, the contact period with the community is limited due to the nature of the fast-paced lives of healthcare workers (N. Scrimshaw &

Gleason, 1992), but also due to the increasingly short production realization cycles in HCI (Millen, 2000).

In this sense, Hughes et al. (1995) propose *quick and dirty ethnography*, which was later adapted by Millen (2000) as *rapid ethnography*, both being condensed methodological approaches for ethnography in HCI. *Quick and dirty ethnography* focuses on informing strategic decision making, while *rapid ethnography* focuses on condensed interviews, observation, and collaborative analysis. These adapted approaches also suggest video analysis, which allows capturing much information in a short period of time with the trade-off of long data analysis times. To reduce timelines, short and adapted ethnography reduces the lens and focuses on the narrower aspects that are critical for the research. Through this reduction, having multiple ethnographers or fostering co-analysis with the participants is key for rapid ethnography. This also means carefully choosing the informants, be it people who move freely about a group (*liminal*) or who are employees of the organization (*corporate*) (Millen, 2000).

2.2.1. Ethnography across disciplines

The travel of ethnography from anthropology to other disciplines, for example HCI and healthcare, was not unnoticed by the community when the terms *ethnography* and *ethnographic fieldwork* started to be more frequent outside of anthropology. This initial separation between anthropology and ethnography started in the 1970s when different disciplines incorporated the methodology, and the terminology was changed from *participant observation* to *ethnography* (Bryman, 2012). These concerns stemmed from the epistemological nature of ethnography, as Howell notes:

“Anthropology has nothing to offer the world without ethnography. Ethnography is just an empty practice without a concern for the disciplinary debates in anthropology departments” (Howell, 2017:15)

The separation of ethnography as a research methodology from its foundations in anthropology had epistemological consequences in terms of how to approach the methods and evolution of anthropology. Using the term *ethnography* became an umbrella to “humanize” studies in other disciplines (Howell, 2017). These actions put the field of anthropology/ethnography in a position where it needed to defend the validity of its methods against the “objective” omnipotence and verifiability of “facts” (Stoller, 2016).

Nonetheless, it is also some of those fields that evoke the critique that are in desperate need for re-centring their research on the human aspects (e.g.. sociology for society, medicine for health, geography for space, psychology for the mind or culture) (Kazubowski-Houston & Magnat, 2018). Kazubowski-Houston and Magnat pose questions surrounding the transdisciplinary travels of ethnography:

“What is lost and gained when ethnography “travels” across disciplines? How can ethnography’s transdisciplinary travels contribute to how we might conceptualize, reimagine, and practice ethnography today and in the years to come? What does it mean for ethnography to “travel” within a competitive and profit-driven neoliberal academia, where the pursuit of knowledge is no longer seen as a public good and an end in and of itself?” (Kazubowski-Houston & Magnat, 2018).

From the anthropology community, the main pushback against the appropriation of the term is that two essential aspects of ethnography are overlooked: 1) the immersion in the host culture for extended periods of time, and 2) the engagement with literature from anthropology. In terms of the first aspect, immersion is meant to be more than a couple interviews and observations; it is about becoming a part of the community. Such process might last months or years and involves the creation of meaningful bonds and relationships with the informers⁵ (O’Connell Davidson, 2008).

When transmuting the methodology to other fields, one of the risks is to reduce the dimensions and complexity of anthropology and ethnography. And by doing so, to lose sight of the partiality of the discovered truths, the need for reflection during the process, and a commitment to cultural critique. Importing ethnography to other disciplines requires the reinvention of the term, along with its epistemology and implications. This means that the ethnographer should never forget that she does not begin from a position of authority, or expertise, that channels empirical facts for its discovery and analysis (Carter, 2018). Nonetheless, the use of ethnographic methods should not be insulated from being employed in other disciplines, since ethnography is, in itself, an emergent interdisciplinary

⁵ The term *informer* is outdated; more inclusive terms are now used emphasize a shared working relationship rather than an outsider “expert” doing a study “on” a group of insiders.

phenomenon. "Its authority and rhetoric have spread to many fields where "culture" is a newly problematic object of description and critique." (Clifford, 1986:3) .

2.2.2. Adapted Ethnography in Healthcare Research

In healthcare, Rapid-Assessment Procedures (RAP) present a procedure and guidelines for health workers and social scientists to incorporate ethnographic methods in public health during short-periods of data collection (N. Scrimshaw & Gleason, 1992). A RAP approach begs considering epistemological and methodological slippery slopes that relate to the intrinsic limitations of rapid procedures: i) accuracy, ii) representativeness, iii) cultural inappropriateness, and iv) subjectivity; and extrinsic limitations such as giving an improper weight to the contextual aspects of the qualitative approach in the big scheme of the research strategy (N. Scrimshaw & Gleason, 1992):

"In practice, however, it is the investigator's judgement call as to what he can ignore and how much imprecision he can appropriately tolerate, without self-destroying penalties. How rapid isn't too rapid? The question to be asked may be: when is rapid not slow enough?"

Therefore, the limitations from RAP can be mitigated through participatory approaches or co-analysis, a robust background of the ethnographer in the methodology to avoid distortions, and the acknowledgement that, while RAP is meant to broaden the use of social investigation across disciplines, it is not a replacement for ethnography (N. Scrimshaw & Gleason, 1992). Furthermore, RAP ethnography has gained traction in medical anthropology, having been deployed for nutrition assessment for collecting data for designing guidelines for the prevention and treatment of AIDS (S. Scrimshaw, Carballo, Ramos, & Blair, 1991), as well as for sanitation (Cifuentes et al., 2006).

2.2.3. Cognitive Ethnography

The cognitive dimension of *cognitive ethnography* stems from studying problem-solving practices situated in real-world science environments (Nersessian, 2013). In this sense, Hutchins solidified the methodology of cognitive ethnography in his study of navigation, where he compares the cognitive processes and their differences between the US Navy's Palau and Micronesian navigators (Hutchins, 2000).

In his research, Hutchins modelled navigation as a computational system, one where many agents must coordinate actions without needing a complete understanding of the overall systemic process. In this manner, when analyzing the system as a whole, it became clear that tasks were still accomplished even when no one knew how the process as whole worked. The system was able to correct errors, support dynamic reconfiguration and be broken down for analysis at multiple levels and loci (Hutchins, 2000). Further, the cognitive system could be broken down into three categories:

- *Scripts*: Sequences of tasks in which individuals with specific roles are trained.
- *Artifacts*: Cognitive artifacts that support the task and enable cognition to be technologically distributed.
- *Protocols & Channels*: Communication protocols and channels for coordination of the agents.

Therefore, when utilizing a cognitive lens to ethnography, it is essential to be able to have a unit of analysis that allows to consider the artefacts and get an understanding of the system as a whole, one that grants the system its own cognitive properties. Such systems can distribute cognition among people, procedures, and artifacts. Therefore, to understand a community of health data analysts, it is important to consider that the knowledge is distributed among all the experts, their processes, and the tools they use. It is the system which possesses all the knowledge of how to analyze and interpret the biomolecular data, but also the knowledge of how to face the bioinformatical challenges of such a discipline. In this case, the intelligent system needs to be robust, resilient, and adaptable in problem solving. In the context of biomedicine and bioengineering, cognitive ethnography has helped to provide a naturalistic approach to the study of cognition in biomedical engineering research laboratories (Nersessian, 2013). By doing so, it has shown the value of focusing on such an interdisciplinary space where conceptual frameworks, methodologies and epistemic values are diverse.

2.2.4. Short-Term Ethnography for Healthcare HCI

Adapted, short methodological adaptations of ethnography have been useful for system design in HCI and healthcare while being adept to address the worries from the

anthropological community. As mentioned in previous sections, while methods such as interviews, observations and collaborative analysis are part of the quick methodologies, they seldom engage with critical discussions in anthropology and miss the proper incorporation of literature from anthropology. In this sense, *quick and dirty* or *rapid* ethnography are lacking that dimension, it is in that gap that *short-term ethnography* provides a relevant framework for HCI. Short-term ethnography can be understood as a *focused* approach to ethnography, rather than a short-range superficial methodology (Pink & Morgan, 2013). In addition to the considerations of shorter timespans, short-term ethnography emphasizes the qualities of ethnographic research, furthermore the methodology stems from the experience of the authors in ethnographic fieldwork in healthcare

For short-term ethnography, an ethnographic place is not understood as a fieldwork locality but rather as the aggregation and entanglement out of which ethnographic knowledge arises from (Pink & Morgan, 2013). Such a place extends to the audience considerations and how to represent the knowledge of a multidimensional world when reduced to letters and words in an academic publication. The qualities that comprise short-term ethnography are the *intensity of the research encounter*, *a focus on details*, *the ethnographic-theoretical dialogue*, and *the audio-visual and other traces of ethnographic encounters*. Below a more description of these qualities:

The *intensity of the research encounter* requires ethnographers to take an active role and find themselves at the centre of the action. By doing so, the intensity of the research encounter becomes part of the way of knowing. *A focus on details* aims to circumvent the short-range nature of the methodology, which in a healthcare context might even extend to ethical considerations, through making correspondences between the research participants experience and the ethnographer's own personal experiences (Pink & Morgan, 2013).

The ethnographic-theoretical dialog demands a critical engagement with contemporary debates and literature. It is essential to incorporate the evolution that the field has had, in particular the fact that the ethnographer is no longer an outside observer who relies on scholarly expertise to describe research subjects and their epistemologies. It is about working with communities, rather than studying communities. Ethnographies are no longer written for an academic audience that excludes the studied participants

(Erickson, 2011). When departing from the authoritative stance of ethnography, feminist, indigenous, and non-western ethnographers question the colonialist approach of ethnography (Kazubowski-Houston & Magnat, 2018). When reflecting on the representation of studied individuals, it is critical to understand the partiality in truths and subjectivity that derives from studying the human experience. Writing ethnography is not a matter of reporting the “facts”, but about making decisions about how the individuals will be represented.

Finally, the *Audiovisual and Other Traces of Ethnographic Encounters* refer to the production of multimedia materials that allow for a constant re-engagement with the ethnographic place and to provide focused analysis on the artefacts that are essential in the daily work of the participants.

Short-term ethnography provides a framework for HCI that can be deployed in short-range fieldwork, specifically for involvement in healthcare. The framework focuses on the contextual aspects of the nature of the research, engaging in a theoretical dialog and elucidating the essential aspects that define the culture and nature of work. In this sense, such a framework is very useful for HCI in precision medicine, where there are rapid product cycles that might eventually affect the lives of health consumers. That is where the value of short-term ethnography comes from, as a way to move away from quick and dirty, rapid ethnography approaches and their shortcomings.

2.3. Visual Analytics Design Considerations

2.3.1. Cognitive artefacts

From an historical perspective within HCI, cognitive artefacts were supposed to increase human capability to “approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems.” (Engelbart, 1962:1). Engelbart’s team was directly involved with developments in HCI that remain valid today, such as the mouse and the Graphical User Interface (GUI), and his views of computers as cognitive amplifiers permeated into contemporary research approaches. Engelbart thought of computers as artefacts that help overcome limits of the human mind, such as memory, thought and reasoning.

When considering the design of digital technologies for healthcare, specifically for precision medicine, it is a worthwhile exercise to take a step back and analyse current cognitive theories used in Visual Analytics to select a cognitive paradigm that can fit the development of novel interactive visualizations for precision medicine. For scientific inquiry, practitioners need to remain up to date and have an active scrutiny and understanding of assumptions behind the theoretical foundations of their craft. As stated by Ribarsky and Fisher: “our visual analytics researcher must understand how to map results from individual studies and reviews of multiple studies onto the more complex tasks and visual environments used in visual analysis.”(Ribarsky & Fisher, 2016:1447).

Previous analyses in VA have found that there are three main conceptualizations of visualizations in visual analytics: cognitive amplifiers, cognitive prostheses and cognitive mediators (Arias-Hernandez et al., 2012). In all cases, a visualization is a tool to harness the human innate visual intelligence to understand and generate insights from cases characterized by the complex nature of the data, yet the understandings of these artefacts are informed by different approaches to cognition. When differentiating cognitive artefacts into the three categories, the differences rely on assumptions such as: seeing humans as passive or active cognitive agents, whether cognition is a universal feature or if it adapts to the situation; and who (or what) is granted agency within the cognitive system (Arias-Hernandez et al., 2012).

Cognitive amplifiers and prostheses

The understanding of visualizations as cognitive amplifiers is rooted in different approaches to distributed cognition. They share the notion that cognition is not limited to the brain’s internal states, and that it is rather mediated between internal and external representations (Zhang & Norman, 2003). The first approach of “external cognition” is informed by cognitive sciences à la Norman (Norman, 1993; Zhang & Norman, 2003) or Larkin & Simon (Larkin & Simon, 1987); and the second one comes from cognitive anthropology à la Hutchins (2000) or Lave (1988). The main difference between both perspectives is whether “external cognition” focuses on the social interactions within the system and the cultural background of the actors or not. In a cognitive anthropological perspective, cognitive processes are distributed in a network of collaborative agents and artefacts, as hinted by Latour “In politics as in science, when someone is said to “master” a question or to “dominate” a subject, you should normally look for the flat surface that

enables mastery (a map, a list, a census, the wall of a gallery, a card-index, a repertory); and you will find it.” (Latour, 1986:21). In this perspective, the background and sociocultural factors of the system are relevant for the analysis and the unit of analysis is a sociotechnical system, with an emphasis on the interaction between individuals and artefacts.

In terms of agency, both views of distributed cognition consider humans and artefacts as elements of the system. By doing so, agency becomes equally distributed within the system. That is, humans and artefacts are considered at the same level. In retrospect, one of the biggest flaws of this approach is that it is rooted in the assumption that there is a universal human cognition.

The understanding of visualizations as cognitive prostheses is directly derived from that of cognitive amplifiers. As its name suggests, these artefacts aim to restore functionality of this *universal* cognition. A visualization becomes a means to restore lost (or never acquired) “normal” cognitive abilities, and more importantly, since humans and artefacts have the same agency within the system, then it is justified to focus research on the artefact instead of on the human. This has led to “an overemphasis in ascriptions of agency to external, visual representations and interactive techniques, and a consequent underplay of the active, human cognitive agent” (Arias-Hernandez et al., 2012:15).

Cognitive mediators

When being more critical about the reduction of cognition to the confinements of a mind-body dualism (internal and external representations), and socio-cultural aspects (folk psychology) (Nersessian, 2004), it becomes relevant to challenge the assumption of a universal cognition. It becomes clear that human cognition is situational, and expertise plays a major role. This understanding comes from Activity Theory (Kaptelinin & Nardi, 2006; Nardi, 1995) and allows for shifting our understanding of visualizations as cognitive amplifiers towards understanding them as cognitive mediators. This, since it has been shown in the fields of social sciences that human cognition is molded by a series of factors, as well as situations. For this perspective, human activity is seen as the locus of cognition (Kaptelinin & Nardi, 2006) instead of the interaction between internal and external representations (Arias-Hernandez et al., 2012). The shift in focus allows for considering the situatedness of cognition, the personal background and knowledge of the user, and more importantly how those factors influence cognition in a specific situation.

Understanding visualizations as cognitive mediators allows for the acknowledgement of the situated nature of cognition and knowledge, and to depart from cognitive universalisms. Such a framework is essential when dealing with varying levels of health literacy and agency (Trnka, 2016) from health consumers. When thinking about precision medicine in Visual Analytics, one noteworthy aspect that has been carried over is the notion of a universal cognition: “a particular element of importance to health IT is intuitiveness or “learn-ability without teaching”.”(Armijo et al., 2009:11). This reduces the importance of the focus on the human agent and over relies on the representation for problem-solving. As shown previously, such an assumption becomes problematic when deciding where to focus the design.

Furthermore, for cognitive mediators only the humans are ascribed agency, departing from the symmetry in agency of distributed cognition and allowing for centring the research on the human experience. While understandings of cognitive amplifiers treat visualizations as the graphical depiction of information, cognitive mediators consider a visualization as an artefact co-determined by the graphical depiction and the user’s knowledge. Therefore, conceptualizing visualizations within precision medicine as cognitive mediators allows for placing the focus on the activity of health consumers reporting on medical history for lay users, and more importantly for choosing a cognitive framework that is in line with research in that field.

Drawing from Arias-Hernández’s research (Arias-Hernandez et al., 2012; Fisher, Green, & Arias-Hernandez, 2011), the human cognitive experience is still to be addressed in order to create better cognitive mediators. Placing the focus back on qualitative research that focuses on the human experience allows for the creation of human-centred design and to harness the cognitive situatedness. When focusing on precision medicine, human error needs to be understood under the bigger scope of the underlying causes stemming from the system. Moreover properly acknowledging the semantic relations in which the analyst is situated is a critical step for the design of information technologies (Hazlehurst, Gorman, & McMullen, 2008).

2.3.2. Target Audience

Health consumers and Health Data Analysts

When considering both audiences, health consumers and health data analysts, visualization and analytical tools become intertwined and need to support expert and lay users who have very different analytic goals and needs. Expert users are defined as data analysts whose task is to interpret biomolecular data, and lay users as those without a formal training in medicine, the –omics sciences or related fields. This *barrier* between audiences is one of the main challenges for VA systems to overcome (Maglogiannis, Billiris, & Valavanis, 2016). VA systems need to allow health consumers to accurately report on their medical history in a way that analysts can analyse the data and automatize the process, while still being able to create a report that is understandable by the health consumers and their caregivers. It is in this sense that Mougín states (Mougín et al., 2018:13):

“As omics data are now widely used as relevant features in the context of clinical decision making and clinical research, a bridge is being built between these data. Put simply, the identification of genotype-phenotype relationships is an essential support for precision medicine.”

Such an integration is becoming an increasingly relevant challenge for precision medicine, especially with the deployment of AI algorithms for classifying or analyzing biomolecular data (Topol, 2019). In this sense, analyzing the analyst’s perspective allows to elucidate critical requirements for VA systems, while a focus on the health consumer enables the design of functional tools for them.

The problem of designing VA systems for health consumers is a problem of coordination between human agents and health information technologies. With EHRs becoming more frequent, the new frontier of discovery shifts now to making both health specialists and consumers, acquire insights from the data and act upon it. Shifting the focus towards the human experience has its own challenges, since understanding the initial complex raw data sets requires knowledge and understanding from bioinformatics and data visualization. Unfortunately, the average consumer does not possess that knowledge, let alone understands it to a point that they can make decisions. The real challenge lies in being able to convey the message without needing the technical

knowledge background. In turn, this raises the need for tools that can resemble a two-sided dialogue in which the interface provides the health consumer information as well as medical understanding.

2.3.3. System Considerations

Visual Communication

Visual communication through digital interfaces should also consider the cognitive load of the user and try to avoid imposing additional limitations. Thus, the design of visual systems should avoid having elements that impede the communication (Zachry & Thralls, 2004). Furthermore, certain aspects of the communication loop can be represented through the interface: Identity becomes the origin of the fact or hypothesis, illustrative language can be mirrored through effective graphics techniques, intonation can be achieved by highlighting, and verbal aesthetics through a good design (Mueller, Garg, Nam, Berg, & McDonnell, 2011). Through this lens, it is possible to create the essence of a dialogue. In the context of precision medicine, manipulating these dimensions of an interactive visualization can be used as the first steps towards a visual dialogue between the health consumer and the health provider.

In the current digital era, after the realization that computers were increasingly used for more purposes than mere calculation (Oberquelle, Kupka, & Maass, 1983), studying the communication properties of computer software became increasingly important. Even when considering that the computer-based artefact doesn't possess any cognitive agency, it can be designed in a way that its behaviour is rationally accountable (Suchman, 1989). In that same line, Mueller et al. have proposed the Human-Computer Communication Protocol as a way to create an *intelligent* visual communication loop through the interaction with a VA interface. The key idea being that the feedback from the visualization can be thought of as an *answer* from the system, specifically when considering visual signs as triggers for information retrieval and compression (Mueller et al., 2011).

Usability

It is acknowledged that successful design will be evidence-based and involve the need to include several disciplines such as medicine, information and data science,

usability engineering, cognitive science, psychology, etc. (Armijo et al., 2009). From a usability perspective, there needs to be a strong emphasis on the cognitive ramifications that using an EHR might present to the caregivers (V. L. Patel & Kushniruk, 1998; Stead & Lin, 2009). The focus on the cognitive workload produced by using health records has been carried over from paper-based records, since the presentation of a health consumer's clinical information will have a direct impact on the clinician's decision-making (Marchionini, Rimer, & Wildemuth, 2007). Further, when there is incomplete or inconsistent display of information, it becomes an additional factor that compounds the inherent variability in physician knowledge (Elson, Faughnan, & Connelly, 1997).

The above-mentioned design considerations are restricted to studying usability from the physician's perspective, since traditionally EHRs have been designed for and used by clinicians (Mosaly et al., 2016). Nonetheless, for precision medicine, health consumers are finding themselves needing to interact with EHRs and PHRs as their main entry point to digital healthcare. This raises the need to study usability from a health consumer's perspective. In this space is that traditional HCI surveys, like the System Usability Scale (SUS) (Bangor, Kortum, & Miller, 2008; U.S. Department of Health & Human Services, 2020) can be used to assess a VA's system usability from a health consumer's perspective.

2.3.4. Data Ontologies

In the context of data infrastructures, the term *ontology* is used as a controlled vocabulary about a specific area of knowledge (Prodanov, 2011). Ontologies contain the "objects, concepts, and other entities that are presumed to exist in some area of interest, and the relations that hold among them"(Katifori, Halatsis, Lepouras, Vassilakis, & Giannopoulou, 2007:10). In medical research, ontologies are widely used. They refer to a set of disease risks and phenotypes, and categories that show their properties and the relations between them. Additionally, they allow for establishing hierarchical relations between concepts, which can be visualized in a way to support the analytic tasks (Glueck et al., 2018; Holten, 2006; B. Johnson & Shneiderman, 1993). As stated by Hoehndorf et al.:

"Their success lies in the combination of four main features present in almost all ontologies: provision of standard identifiers for classes and

relations that represent the phenomena within a domain; provision of a vocabulary for a domain; provision of metadata that describes the intended meaning of the classes and relations in ontologies; and the provision of machine-readable axioms and definitions that enable computational access to some aspects of the meaning of classes and relations.” (Hoehndorf, Schofield, & Gkoutos, 2015:1069)

Data ontologies become robust data infrastructure assets since working with clinical data presents the challenge of structured but uncoded data elements (Mate et al., 2015). While there is a strong push from the medical community to standardize terminologies, a vast amount of existing literature lags behind these efforts (Köhler et al., 2017). For precision medicine, having a robust data ontology which allows for capturing diverse terminologies for a disease in a way that it maintains the hierarchical relations of the concepts is highly valuable. This allows for the implementation of algorithmic searches, and automatization and data mining (Jensen, Jensen, & Brunak, 2012).

Previous work in clinical data architecture for has been rooted in the integration of EHR data, biomolecular data, imaging data, and descriptive health consumer observations (Maglogiannis et al., 2016). This points to the need for integrative systems that provide sufficient clinical and biomolecular data to clinicians to make decisions, yet it also shows the need for a standardized vocabulary that allows for a successful integration of data. From a different perspective, the use of ontologies for EHR interoperability has proven to be challenging due to the lack of semantic heterogeneity of EHRs (González, Blobel, & López, 2012).

With the introduction of Deep Neural Networks and AI in healthcare, there is a need for accurate and descriptive health consumer data that can be fed into such algorithms. Previous studies have found that machine and deep learning algorithms that worked on data from health records were able to predict relevant clinical parameters, such as in-hospital mortality, sepsis, dementia, Alzheimer’s disease and suicide, among others (Topol, 2019).

The Human Phenotype Ontology

The Human Phenotype Ontology is a tool that provides a “standardized, controlled vocabulary [that] allows phenotypic information to be described in an unambiguous

fashion in medical publications and databases.” (Robinson & Mundlos, 2010:525). The HPO has two mayor strengths: the first one is that it has been adopted by several resources for genotype-phenotype and the second one is that it has lay synonyms. Examples of the former include GWAS Central and GWASdb, annotated health consumer record databases such as DECIPHER, and its integration into the UMLS starting on the 2015AB release (Köhler et al., 2017). While it is important to note that the HPO is only reaching maturity for a few themes such as non-malignant haematology and eye diseases, non-clinical uses such as capturing medical history information from health consumers can still be of benefit. As for the latter, the HPO contains “layperson” synonyms for some of the conditions, (e.g. Palmar pruritus vs Itchy palm), as well as short descriptions of them. While these present more graspable terms for the medical conditions and short explanations, most of the conditions remain as abstract concepts in relation to the health consumer’s embodied experience. The HPO’s strength in vocabulary, since it was curated by domain experts, has led to it being used for exploration and analytical support in building domain-specific knowledge and vocabulary (Demelo, Parsons, & Sedig, 2017).

Medical Terminologies

The International Statistical Classification of Diseases and Related Health Problems Revision 10 (ICD-10) is the WHO’s clinical, hierarchical controlled vocabulary which aims to standardize the terminologies across nations. The 10th revision is used by more than 100 countries and has more than 20,000 scientific citations (WHO, 2020), yet country-specific modifications exist. It aims to support: i) easy storage, retrieval and analysis of health information for evidence-based decision-making, ii) to be universal across hospitals, regions, settings and countries, iii) comparisons in the same location across time periods (WHO, 2020). Another relevant medical controlled vocabulary is the Systematized Nomenclature of Medicine (SNOMED), which is described as “the most comprehensive clinical terminology in use around the world” (SNOMED International, 2020) and it aims to become a key vocabulary for meaningful use, EHR certification and health information exchange. SNOMED complies with federal regulations from the USA, and it is recommended for documenting patient descriptive data, encounter diagnosis, procedures and smoking status (SNOMED International, 2020).

Data Ontology Visualization

The visualization of biomedical ontologies presents the challenges of every hierarchical tree visualization, such as multiple inheritance, which leads to either duplication of the concept or occlusion generated by using multiple edges. This leads to the need for larger screens with the limit being the amount of visual information that can be processed. As such, visualizing ontologies needs to represent the subsumption relationships between concepts or classes. For biomedical ontologies, the associations between concepts define the relational expressions between concepts and provide additional layers of information. Approaches to visualizing ontologies have been widely studied in six groups: indented lists, node-link and trees, zoomable, space-filling, focus + context or distortion, and 3D Information landscapes (Katifori et al., 2007). In terms of applications, Yang et al. have studied use-cases of ontology visualization for exploration of large hierarchical structures (Yang, Wybrow, Li, Czauderna, & He, 2019).

2.4. Chapter Summary

Chapter 2 served to explore the story of Design Thinking, how it started in the problem-ridden field of architecture and then spread to become a framework to tackle innovative, human-centred problems. In HCI, the work of Simon, Schön, Norman and Arthur allows to create a cognitive framework for evidence-based design. In turn, the HCI designer needs to acquire interactional expertise from different fields to fulfill the role of an interdisciplinary knowledge broker. For the present research that means a critical understanding of ethnography as a process and a product, and how it has been used across disciplines, more specifically HCI. This shows that even when the methodology needs to be adapted (i.e. cognitive ethnography centred on cognition and its mediation through artefacts, or short-term ethnography accounting for the short lifecycles of HCI products), it needs to maintain its epistemological integrity to anthropology.

Finally, the Visual Analytics systems designer needs to know how to translate the qualitative research into design guidelines. More importantly how to turn the guidelines into integrated interactive systems that serve as cognitive mediators for health consumers, and secondly to be familiar with controlled vocabularies that support the data and automation needs of health data analysts. Chapter 3 will focus on providing a thorough explanation of the methodology and methods, especially how short-term ethnography and

user testing were layered to the Norman-Nielsen Design Thinking framework for the present research. The chapter explores the research design in detail, and also shows the process of how the VA system was designed.

Chapter 3. Methodology

[Wicked problems] are a class of social system problems which are ill-formulated, where the information is confusing, where there are many clients and decision makers with conflicting values, and where the ramifications in the whole system are thoroughly confusing.

Horst Rittel (1972:141)

Chapter 3 aims to provide a deep engagement with the theoretical foundations, and the methodology and research design as part of the Norman-Nielsen DT framework. The chapter presents an initial overview of the research philosophy, research design and research questions. The rest of the chapter is subdivided in the three sections: *Understand*, *Explore* & *Materialize* that capture the steps of the Norman-Nielsen DT approach (See Figure 2.2). The *Understand* sections deals with the process of generating empathy, more specifically a rich understanding of the stakeholders, through short-term ethnography as a methodology and frames the design problems. The section *Explore* is about the design process of creating the interactive visualization through the mapping of the Human Phenotype Ontology to a body graphic. The final section *Materialize* deals with the methodology for the evaluation of the prototype.

3.1. Research Overview

3.1.1. Research Philosophy

VA can be understood as Highly-Integrated Basic and Responsive Research (HIBAR) (The HIBAR Research Alliance, 2020) since it integrates design & development of data analysis/data visualization systems with basic research in perceptually-rich situations. This research involves looking at human cognition, communication and action, in situations such as those created by highly interactive graphical environments. VA works best for situations where computation-based analysis (e.g. statistics, machine learning) must be reconciled with human understanding, values, and ethics. The HIBAR approach

provides a framework application-aware research. In the context of precision medicine, a HIBAR approach allows to focus on basic research in visually-enabled reasoning so that the methodological and scientific results can be transferable to other applications.

Another idea that informed the design was Simon's notion of *satisficing* (Simon, 1996b), which informed how to analyze the results. The present research was focused on how to incorporate short-term ethnography into a design framework, so having "good enough" results was the focus rather than looking for achieving perfect results. That means that when using standardized survey evaluation methods such as the SUS (U.S. Department of Health & Human Services, 2020), or comparing the intercoder reliability for the qualitative data (Geisler & Swarts, 2019), as long as the results are on the threshold or slightly below the literature standards, they were deemed proper to continue with the research.

3.1.2. Research Design

- **Research framing:** Pragmatic worldview
 - Exploratory Sequential Mixed Methods
- **Qualitative methodology:** Short-term Ethnography
 - **Methods**
 - Observation
 - Ethnographic Journal
 - Ethnographic Interviews
- **Quantitative methodology:** Survey, True experiment
 - **Methods**
 - Questionnaires (HLS-EU-Q16, System Usability Scale)
 - User testing
 - Short, structured interviews

The present research used an interdisciplinary, exploratory mixed methods approach to incorporate a cognitive lens to short-term ethnography to analyze the human experience of health consumers to complement VA systems creation strategies. Mixed Methods approaches draw from a *pragmatic worldview*, since they focus on the actions, situations and consequences rather than on the conditions that preceded them as in postpositivism (Creswell, 2007). There is an emphasis on the research problem, in this case the *wicked problem*, which allows the research to employ methodologies and methods from different types of inquiry, be it quantitative or qualitative, that centre on using pluralistic approaches to derive knowledge about the problem (Creswell, 2013). This requires a critical understanding of the work in order to accommodate the methods that best fitted the research approach, and also to provide an ethical grounding for conducting the ethnography (O'Connell Davidson, 2008). The research is guided according to TCPS 2 guidelines, following their core principles of concern for welfare, respect for persons and justice.

The research is divided in two phases, an initial one where I sought to immerse myself at the company to compile design requirements, and a subsequent one where I looked to evaluate the design in terms of usability and data reliability. For this, I employed an exploratory sequential mixed methods approach (Creswell, 2013), where I initially implemented a qualitative methodology, more specifically, short-term ethnography (Pink & Morgan, 2013) as informed by critical ethnography (Kazubowski-Houston & Magnat, 2018) and cognitive ethnography (Hollan, Hutchins, & Kirsh, 2000) to explore the views of the health data analysts. After the data analysis and the design of the prototype, the quantitative phase sought to evaluate the prototype through a user-testing experiment. Exploratory sequential mixed methods are typically used for the initial design of artefacts before an evaluation stage with the target population (Creswell, 2013).

For the initial exploratory stage, the goal was to compile a set of needs from health data analysts that would serve as guidelines for the design of a health consumer-facing interactive visualization. By doing so, the researcher sought to provide evidence for the need of departing from universalisms in cognition and a focus on the user's situated knowledge. Secondly, the experimental design provides data to assess whether the interactive visualization could accurately capture health consumer descriptive data.

3.1.3. Research Questions

Mixed Methods Question

- **How to design an interactive visualization for capturing descriptive health consumer data to support automatization in precision medicine?**
 - **Hypothesis 1:** Through structured field immersion for understanding different stakeholders, a VA design approach leads to a visual mapping of a clinical data ontology that facilitates a health consumers' reporting of their descriptive data, because data architectures and human factors can be converted into design guidelines

The main question provides an example of how the fields of VA and HCI, as human-centred fields, need to continue exploring interdisciplinary approaches that allow for the integration of qualitative and quantitative research. The expected outcome is that by mapping a data ontology to a body graphics, the VA system can facilitate the health consumers' report of their descriptive data through a visual controlled vocabulary for automatization

The mixed methods question is intended to become a starting point for investigating two different audiences: health consumers and data analysts. It bounds the study in terms of scope and timeframe, within the constraints of the tech biomedicine ecosystem. This question intends to bridge two different approaches for knowledge making through a Design Thinking Framework. The ethnographic methods relate to achieving empathy with the health consumers and to understand the data requirements from the health analysts to design the interactive visualization. The user-study relates to the prototyping and evaluation stages of an iterative cycle.

The importance of the encompassing question is that it allows to relate the patterns emerging from the observation and interviews from the field work with the data needs. It allows to form a holistic picture of the process, since it considers the two main audiences of the work. Furthermore, it allows to provide a critique of the western data-driven healthcare and how data is conceptualized and treated by the data analysts. Furthermore,

the mixed methods approach allows to reduce the biases inherent to pure qualitative or quantitative research, this will in turn allow for “triangulation” of the information. The interviews will be transcribed with a focus on the analysts experience and how they respond to the questions (Bird, 2005; O’Dell & Willim, 2013; Pink, 2011).

Qualitative Question – Short-term ethnography

- **How can a distributed cognition lens for short-term ethnography inform the design of visual analytics systems in the context of correlating health consumer descriptive data with biomolecular data?**
 - **Hypothesis 2.1:** The use of a cognitive lens for short-term ethnography helps elucidate the vulnerability of health consumers and the data needs of analysts for the design and use of VA systems in the context of correlating health consumer descriptive data with biomolecular data.

Data analysts are immersed in an environment rich with highly sensitive and confidential data, data which represents the molecular entirety of a person. This question serves to understand how the use of a cognitive lens for short-term ethnography helps to elucidate the vulnerability of health consumers, and the data needs of analysts for the design and use of VA systems in the context of correlating health consumer descriptive data with biomolecular data

It allows for the emergence of patterns related to experiences of the analysts, how they conceive the health consumers, and more importantly, what is the procedure that health consumers undergo when purchasing precision medicine products. For this question, it is essential to mention that the ethnographic methods will be used as a part of a design thinking framework, with the focus on the design of the interactive visualizations rather than on a complete and thorough description of the data team as a culture-sharing entity. This difference allows to focus the data analysis only in the relevant aspects that will translate into a part of the design of the visualization

Finally, it is important to clarify that the present research utilizes a cognitive lens to short-term ethnography rather than being a cognitive ethnography per se. The elucidation derived from the cognitive lens is intended to design VA systems, which in turn

could then be the target of a cognitive ethnography. A cognitive ethnography would expand across more departments of the company and in time, yet such an approach would go beyond the timespan of a master's degree.

Quantitative Question – User Testing

1. Can an interactive system for capturing descriptive data help health consumers to accurately report their conditions?

- **Hypothesis 2.2:** The interactive system captures a health consumer's descriptive data more accurately than a traditional medical history form.

Reporting on medical history and symptoms is a challenging task for health consumers when they must do it via a screen rather than by having a conversation with their caregiver. This question looks to translate the needs from the health data analysts as identified from the short-term ethnography, specifically through the mapping of the Human Phenotype Ontology to a graphical representation of the body. The goal is to evaluate whether an interactive visualization, that considers the vulnerable state of health consumers, allows to capture accurate and descriptive data that facilitates the analytical work conducted by health data analysts. The expectation is that the interactive visualization allows health consumers to accurately report on their conditions. The results from the surveys and prototype evaluation were analyzed using descriptive statistics in JMP 14 and visualized in Tableau.

3.2. Design Thinking: Understand - Empathize

3.2.1. Empathy

Empathy is one of the main tenets of Design Thinking, and is defined by Merriam-Webster as: “the action of understanding, being aware of, being sensitive to, and vicariously experiencing the feelings, thoughts, and experience of another of either the past or present without having the feelings, thoughts, and experience fully communicated in an objectively explicit manner”. In that sense, Norman-Nielsen's framework allows to scaffold a naturalistic methodology for getting a user understanding that goes beyond

sensitivity and is richer. By doing so it might be possible to create parallels between the researcher and the user's experiences.

3.2.2. Short-term Ethnography

In order to understand the nature of the analyst's works, I focused on the qualities of short-term ethnography (Pink & Morgan, 2013) to accommodate my reduced timespan. I experienced the *intensity of the research encounter* by taking a leadership role in the initiative for re-designing the company's data-capture form. I worked with the company through a Mitacs Accelerate⁶ internship between September 2018 and April 2019, spending three days a week at the facilities. I joined the Data Team which had five members and was comprised by Subject Matter Experts (SMEs) in experimental medicine, genetics, kinesiology, and nutrition. My role in the company involved working with the data team and re-designing the client-facing data capture form. Additionally, I sought to take my lunch and coffee breaks at the same time other member of the data team would, and looked to join and help plan social activities outside the work environment.

In terms of *A focus on details*, I underwent the same process that health consumers experienced when purchasing precision medicine products from the company. This involved giving a blood sample, adding my descriptive data to my profile and later on, reviewing my report and following the actionable recommendations. By doing so I was able to correspondences between my experience and those of health consumers (Pink & Morgan, 2013). *The ethnographic-theoretical dialog* required to look at contemporary debates and literature, specifically by looking at other ethnographies, starting from ethnomethods in HCI, to critiques stemming from anthropology. This also led to incorporating the issues of property and genomic data that indigenous nations have suffered, and how the present research relates to those critiques.

In terms of the *Audiovisual and Other Traces of Ethnographic Encounters*, for the ethnographic journal I used a password-protected Word file. For quick notes, I used the GoogleKeep app in my phone. The interviews were recorded using a Zoom H4n Pro

⁶ The Accelerate Fellowship is the flagship program of Mitacs, a Canadian nonprofit national research organization, and has supported over 10,000 research internships for graduate students and postdoctoral fellows since 2004 and has replaced Canada's Natural Sciences and Engineering Research Council's Industrial Postgraduate Scholarships Program. <https://www.budget.gc.ca/2015/docs/plan/budget2015-eng.pdf>

recorder (See Appendix A). During the fieldwork, the observational notes and ethnographic journal were written during lunchbreaks or after team meetings. For this, I took half an hour to one hour breaks every day I was at the office in order to record my observations and my feelings. For the interviews, I avoided having them at the office in order to create an *ethnographic place* (O'Dell & Willim, 2013). I sought to avoid issues of privacy since the kitchen was an open space, but more importantly, I looked for the interviewees to reflect and feel in a more comfortable setting. The Interviews were conducted in parks and cafés near the office.

3.2.3. Study Participants

For the ethnographic inquiry, there was a total of 5 participants for the interviews. The participant population consisted of the same group of health data experts that the senior PhD lab member worked with for her dissertation (Calderon, 2020). To maintain continuity, I will use the same pseudonyms for the present thesis⁷. For the user testing, there was a total of 8 undergraduate students from a local university, most of the students belonged to the design school.

The Researcher

For the purpose of clarifying the bias the researcher brings to the study (Creswell, 2013), I will present who I am in the context of the project by introducing myself and providing additional details about my person that I consider relevant. The goal of this process is to depart from a colonialist approach to ethnography where the researcher is presented as an objective pillar, and to represent myself as an active participant of the ethnographic inquiry.

My undergrad was in Biotechnology Engineering, but I also started formal education in Hispanic Letters and Literature, and later in Philosophy. As a graduate student, I am trained in ethnographic methods, information practices in contemporary society and bioaffective computing; with foundations in Science and Technology Studies (STS) and design approaches to computing. Throughout this degree I was involved for two years as a teaching and research assistant in an undergraduate introductory course

⁷ In Calderón's dissertation (2020), the role of the Nutrition Analyst was portrayed as Helena, whereas in the present thesis it will be broken down into two persons. The second nutrition analyst was a recent hire and still in training while this research was conducted.

to HCI. This provided me with a good theoretical and applied grasp of the design of user interfaces and user experience. Nearing the end of my degree, I was involved as a TA in a visualization and visual analytics graduate course. As a research assistant, I worked with Statistics Canada in assessing their design framework and operational constraints in their initiative for the design of data visualizations. Through this journey, I constantly found myself at the intersection between qualitative and quantitative methods.

For my audience, I am a rookie ethnographer, a visual analytics student who wants to understand the methodology and methods from ethnography to inform his own research practice. In a sense, I feel comfortable enough with constantly walking between computer science and the social sciences since I have always felt like belonging to two different worlds. For my ethnographic fieldwork, I focused on the analytical aspects of the data, and on gaining a better understanding of the health consumers and users of precision medicine products.

The Data Team

The Data Team worked in a pipeline completing the biomolecular analysis and, step-by-step, creating personalized and actionable recommendations. There were four major roles: 1) Health Data Review Analysts, 2) Nutrition Analysts, 3) Physical Activity Analysts and 4) Adjunct clinicians. The data team, comprised by the five participants of the interviews, oversaw the first three roles and completed them sequentially. They started with the health data review, followed by an assessment of physical activity, and finishing with nutrition. The summary of the study participants and settings can be found in Table 3.1.

Table 3.1 Study participants and data collection summary

Pseudonym	Role	Participant Background	Setting	Number of interviews	Time per interview
Antonio	Health Data Review Analysts	Biosciences, PhD	Taco Restaurant	1	61 min
Gail	Health Data Review Analysts	Genetics, PhD	Bubble Tea Café	1	65 min
Wilma	Nutrition Analysts	Bioinformatics, MSc	Ice-cream parlor	1	47 min
Katie	Nutrition Analyst	Canadian Certified Nutritionist	Local bakery	1	49 min

John	Physical Activity Analyst	Kinesiology and Biomolecular Sciences, PhD Candidate	Local park	2	63 min 59 min
			Total	6	344 min

The Physicians

It is important to mention that the analytical process is found between two processes conducted by a team of physicians: data quality assurance & control (QA/QC) and language review. For the QA/QC, the data team mined research literature databases to look for biomarkers that could be related to diseases or wellness. The role of the physician in this process was to provide a clinical perspective on the validity of the biomarkers in creating actionable recommendations. For the language review, the physician aimed to lower the language barrier for clients by rewording the recommendations into lay language. While the physician was not a member of the data team, the work overlapped in several areas during the creation of the personalized health plan. The physicians' feedback and experience complemented the biomolecular data analysis from the data team, creating a *trading zone* where they acted as interactional experts bridging the health consumers and data analysts' worlds. When designing the user testing, the team of physicians was very helpful in revising the language for the clinical vignettes. They helped to ensure the clinical accuracy of the cases, while also using a lay language.

3.2.4. Interview Data Transcription

Whereas the analysis begins during transcription, it is necessary to clearly know the methodology and research goals (Bird, 2005). In the same lines of Bird's proposal, I sought to have a theoretical framework before beginning the analysis since "Lacking a theoretical compass, there is no way to know where to begin." (Kaptelinin & Nardi, 2006:18). The transcription was not only meant to provide data, but also to create a character of each expert, especially since I knew more dimensions of the experts beyond their academic qualifications. This since I had had the chance of going out for food with them, playing board games, and casually chatting in the kitchen.

In a way, the process started before I went over my recordings, it started when I began to record my observation notes and interacting with the data team. For the interview transcription, the goal was to thematically code emerging patterns from the conversation, to present the data in codifiable units of meaning (Bird, 2005). Therefore, I focused only on those aspects that dealt with relevant information relating to the main research questions, staying away from the rigor in transcription demanded when performing discourse or conversation analysis. Additionally, I aimed to capture the narrative qualities and voice of the participant (Bird, 2005).

3.2.5. Thematic Coding

For the systematic coding of the key aspects from participant interviews, the methodology was taken from Geisler & Swarts (Geisler & Swarts, 2019). Aspects that related to the analyst's experiences with VA systems and the data interpretation process were identified from the interviews and captured in an Excel database. In relation to the qualitative research question, by centring the unit of analysis on human activity, and through the distributed cognition lens, it was possible to use the codes to navigate the values of the data team as a single unit. This allowed to better understand the work context where health data analysts were embedded. This would later lead to being able to translate the data team's needs into guidelines that could inform the design of the prototype. The coding scheme was created and verified against the first coding done by the researcher (See Appendix B). The coding scheme is ordered in descending order of frequency, i.e. "Data Challenges" is the first code, which means that in the first coding done by the researcher, it was the code with transcriptions related to it.

In order to validate the codes' reliability, two additional coders (N=2) were selected for calculating the interrater reliability. The first coder was a PhD candidate in linguistics and the second coder was a PhD in HCI. Both coders were given the database and the coding scheme. The simple agreement was calculated for each coder using a pivot table in Excel. Subsequently, the Krippendorff's alpha was calculated for the corrected agreement by creating an agreement matrix in Excel. According to the standards of Geisler & Swarts (2019), a simple reliability above 0.8 and a Krippendorff alpha above 0.7 represent an acceptable agreement between coders.

3.3. Design Thinking: Understand - Define

3.3.1. Wicked Problems

When analyzing the design of precision medicine technologies, framing them as wicked problems (See Table 3.2) allows to dissect relevant aspects to consider for the design. Furthermore, these kind of challenges that derived from a clinical context was explored by Fleck (1979) when tracing the history of syphilology as a way to analyze the genesis and development of scientific facts. This resonates with the last property of a wicked problem, technology designers have no right to be wrong (Rittel & Webber, 1973), especially when technology becomes intertwined with systems that are responsible for human life. Medicine can be considered as an ever-changing field that leaves no room for error to the clinician, and in precision medicine this extends to the technology designers themselves. In this sense, robust methodologies and frameworks that allow to design solutions to wicked problems are essential for precision medicine. Even more so because every problem is unique, solutions are not true or false and the Human Computer Interaction (HCI) adds a new layer of complexity.

Table 3.2 Wicked Problems in Precision Medicine

Properties of Wicked Problems	Design of precision medicine technologies
There is no definitive formulation of a wicked problem	With multiple stakeholders involved and a vast array of needs from everyone, the problem can be defined in numerous ways.
Wicked problems have no stopping rule	Precision Medicine will continue to evolve in terms of methods, robustness of databases and services. Healthcare will always accompany mankind.
Solutions to wicked problems are not true-or-false, but good-or-bad	Representing uncertainty, specifically health and disease risks, as well as predispositions can only be qualitatively analyzed
There is no immediate and no ultimate test of a solution to a wicked problem	The health consumer needs are vast and include counselling, access to information, health literacy, etc.
Every solution to a wicked problem is a "one-shot operation"; because there is no opportunity to learn by trial-and-error, every attempt counts significantly	Once the VA systems have been deployed and health consumers receive their personalized recommendations, there are ethical and legal ramifications.

Wicked problems do not have an enumerable (or an exhaustively describable) set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into the plan	There is variability in terms of VA systems, specifically when they are tailor-made for a company. Solutions can be addressed from a plurality of disciplines such as computer science, anthropology, psychology, etc.
Every wicked problem is essentially unique	Due to different legislation and different healthcare systems, the design of EHRs operate under different constraints across countries. Analysis and recommendations are dependent on the analysts' background and expertise
Every wicked problem can be considered to be a symptom of another problem	Precision Medicine is closely intertwined with the depersonalization and increased commercialization of healthcare, along with a disparity in access to modern healthcare technologies.
The existence of a discrepancy representing a wicked problem can be explained in numerous ways. The choice of explanation determines the nature of the problem's resolution	The solutions are bound by the design frameworks and scientific disciplines informing the design of precision medicine technologies
The planner has no right to be wrong	The health data analyst has an ethical responsibility to the health consumer. Not complying with ethical and legal bylaws might lead to the loss of license or closure of the company.

3.3.2. The Unit of Analysis

As previously mentioned (See Cognitive Mediators in Page 41), the unit of analysis for the present research is the activity of health consumers reporting on their medical history and symptoms. This decision to focus on the activity stems from Nardi and Kaptelinin's Activity Theory, as it allows for considering the user's goals within the study design for applied research (Kaptelinin & Nardi, 2006). While originally stemming from Soviet cultural-historical psychology, Activity Theory has had ripples in western approaches a lá Engeström's triangle diagram model (Engeström, 1990) and its application to Computer-Supported Collaborative Work (CSCW) by Kuutti & Arvonen in the field of medical care (Kuutti, 1992). Such frameworks provide further depth into how Activity Theory can be used for interaction analysis in precision medicine, specifically by allowing to scaffold distributed cognition models in a specific activity. As an analytical framework, Activity Theory provides advantages in terms of using empirical evidence, having a defined unit of analysis and considering agency for design (See Table 3.3).

Table 3.3. Activity Theory in a Nutshell

<p>Why implementing Activity Theory? Allows to include the user's goals (qualitative) in applied research (quantitative).</p>	<p><i>"the conceptual framework of activity theory can be used as a descriptive and orientational framework that facilitates the handling and interpretation of empirical evidence about complex phenomena of the technological mediation of everyday practices."</i> (p.97)(Kaptelinin & Nardi, 2006)</p>
<p>What is an Activity? The unit of analysis. Within the activity, the Interface has a certain identity.</p>	<p><i>"no properties of the subject and the object exist before and beyond activities (Leontiev, 1978)"</i>(p.31)(Kaptelinin & Nardi, 2006)</p>
<p>When to implement? For designing an interaction that conveys agency (Known audience).</p>	<p><i>"Transparency cannot be "built" into a system. Designers can only create preconditions for development of operations, but transparency itself is an emerging property of interaction as a whole."</i> (p.79)(Kaptelinin & Nardi, 2006)</p>

In terms of the quantitative research question, focusing on the activity of health consumers reporting on their symptoms and medical history allows to elucidate aspects that relate to the subjective nature of each individual's role and knowledge within the activity. This becomes relevant when analyzing the gap between the health consumer's actual symptoms and medical history, and the reported data. A focus on the activity allows to better understand the accuracy of the reported data, or lack thereof, and to understand the underlying causes that make data more or less reliable.

It is relevant to mention that during the main activity, there are subactivities which are out of the scope of the present research, such as health consumers consulting their questions around medical terminology with family and friends or via the web. This can also extend to health consumers checking their own PHRs or EHRs to consult their clinical history. While it was out of the scope of this research, each activity can be then further subdivided in time intervals to understand the evolution subject's object and use of artefacts (Engeström, 1990), and as previously mentioned, the possible addition of other actors.

3.4. Design Thinking: Explore - Ideate

One important aspect to account in the design process is the learning curve involved in the use of the system, as well as the familiarity that the user has with the gestures and general layout (J. Johnson, 2014). In this case, the body graphic used for

the visualization serves as visual feedback when the health consumer is interacting with the body categories. Additionally, it serves to show health consumers which conditions are related to which parts of body. The body graphic creates the preconditions for understanding the body graphic and what different parts represent in terms of diseases and health risks. Such an approach is in line with Nardi and Kaptelinin's notion that transparency is an emergent property of the system (Kaptelinin & Nardi, 2006).

Within the activity, the interface has a certain identity as represented through the visual dialogue, and it is the interface, which has a visual language and grammar (Wilkinson, 2005), that allows for health consumers to familiarize themselves with the system. The interactive visualization can fulfil the role of capturing descriptive health consumer data, but also serve to show the user colour schemes, connotations, and the meaning of the graphics that comprise the visualization system (J. Johnson, 2014). Another aspect that was kept in mind alongside the health consumer's experience, was to develop *good* interactive visualizations. The questions surrounding what constitutes a good visualization has been explored by constructing an economic model of visualization that comprises the interaction of three factors: i) the data, ii) the user, and iii) the visualization (Tera Marie Green, Ribarsky, & Fisher, 2009; Van Wijk, 2005). In this sense the researcher proposes the implementation of a data ontology as an essential aspect of the data since it allows to visualize the information according to the specific audience and goal of the artefact.

3.4.1. Mapping the Human Phenotype Ontology (HPO)

The body graphic was originally designed by the company's Vice President of Design and it was an integral part of the product's identity, the user interface and the health consumer's health data report. Creating a new vector image for a body graphic was out of the scope of this project and out of the expertise of the researcher. In order to operationalize the ontology, the first level of classes was identified using the Phenomizer tool, a web-based application for clinical diagnostics that is based on the Human Phenotype Ontology (HPO) (<https://hpo.jax.org/app/tools/phenomizer>).

The body graphic was an integration of different vector images that represented body organs and systems that could be individually mapped to certain categories. Mapping of the HPO's hierarchical structure traces the semantic relations of phenotypes

across body systems and organs and visualizes the symptoms in a body graphic. The parent node in the HPO is “organ abnormality”, out of which the other categories branch. This graphical re-interpretation of the ontology can better match the health consumer’s mental models of their body and experience. The assets were an instrumental part of the prototype due to their granularity, as it can be seen in Figure 3.1.

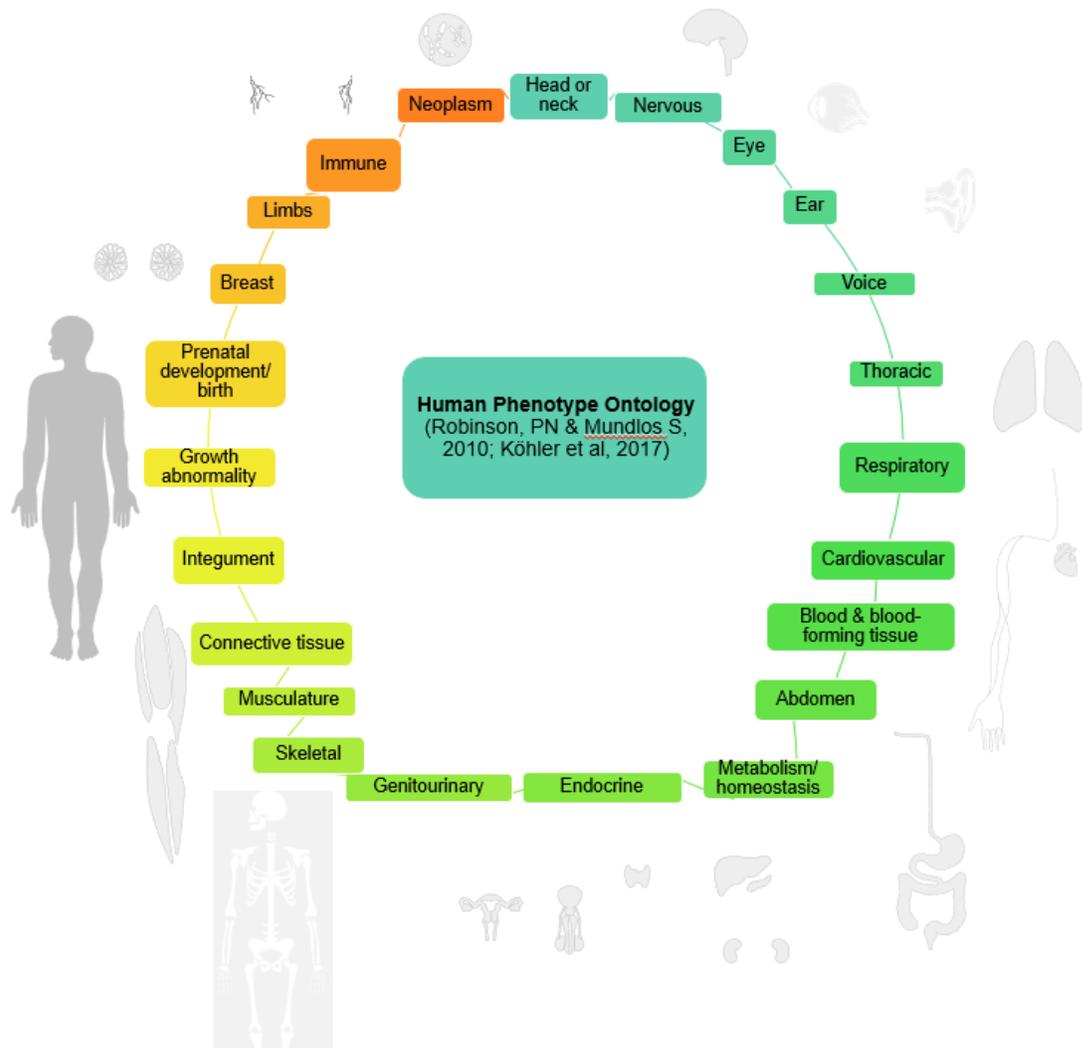


Figure 3.1 Mapping the HPO to the body graphic

Defining the Body Categories as Buttons

The subsequent step in mapping the HPO was to create body categories that would resonate with a lay audience. Even though the HPO has lay synonyms for different conditions, the body categories in the first level of the ontology remain in clinical language. In order to create the lay version of the categories, a quality assurance study was run (n=7) to check the participant's response to the initial categories that can be seen in Figure 3.2. The categories from the HPO in the red rectangle (right side on Figure 3.2) represent those categories to which there was no exact design asset to map. Based on the quality assurance study, the categories were refined to match colloquial terms or short forms such as "cardiovascular health" being reduced to "cardio" or "musculoskeletal health" to "Muscles and Bones", the new body categories can be seen in Figure 3.3. The buttons were rooted in the available design assets and the specific glyphs were chosen to be representational of the body category (i.e. a small heart for the Cardio category, a brain for Brain Health, etc.). One interesting aspect to mention is that during the quality assurance study, different conceptions of the dualism between mind and body emerged, emphasizing the fact that the mind vs body dichotomy is inherent to western culture and might not hold across cultural boundaries.

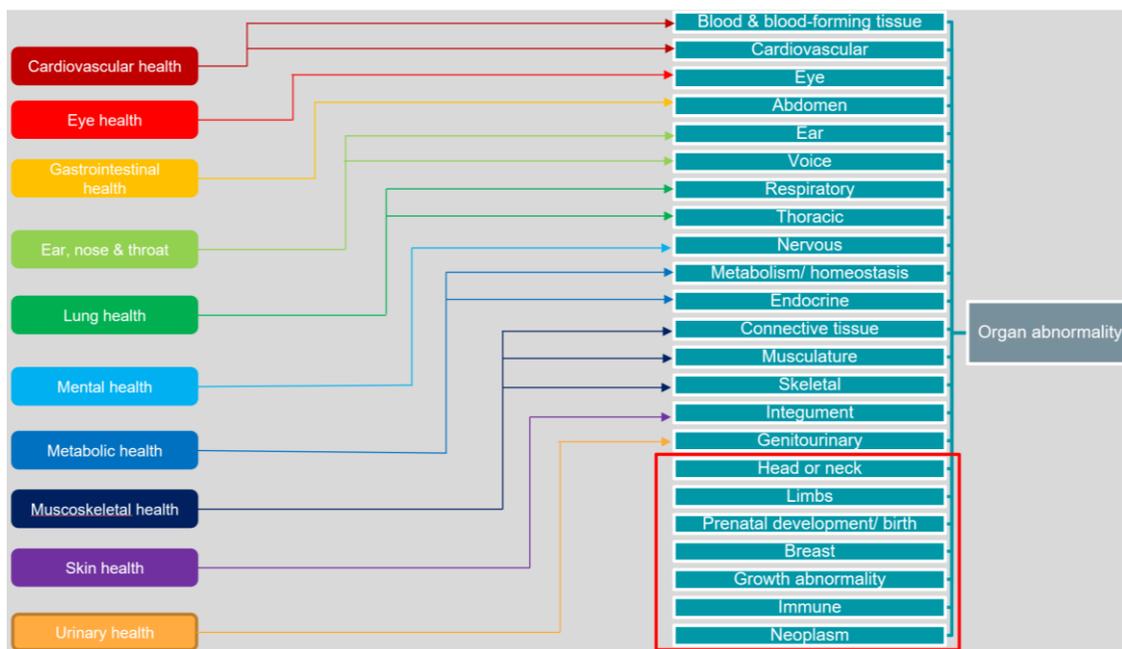


Figure 3.2 Mapping the HPO to body categories (V1.0)

Challenges & Limitations

While the design assets were a major advantage for the creation of the prototype, they also bound the visual representation. This can be seen in Figure 3.1 where there is not a full correspondence between the assets and the HPO's categories (i.e. limbs, head or neck, voice, abdomen). This was also an issue for those categories that are multifactorial such as neoplasms. For more abstract concepts that are hard to point to a specific organ or region, such as prenatal or birth abnormalities or immune system, concessions needed to be made with the available design assets (see Figure 3.3).

One of the main limitations is that the prototype's categories are meant to stay away from capturing mental health data. This is not only due to the lack of correspondence in the HPO, but due to ethical considerations of correlating mental health to specific biomarkers. Nonetheless, and as identified in the quality assurance study, other cultures might attribute body imbalances to mental health, which limits the capacity of the interactive visualization in capturing descriptive data cross cultures.

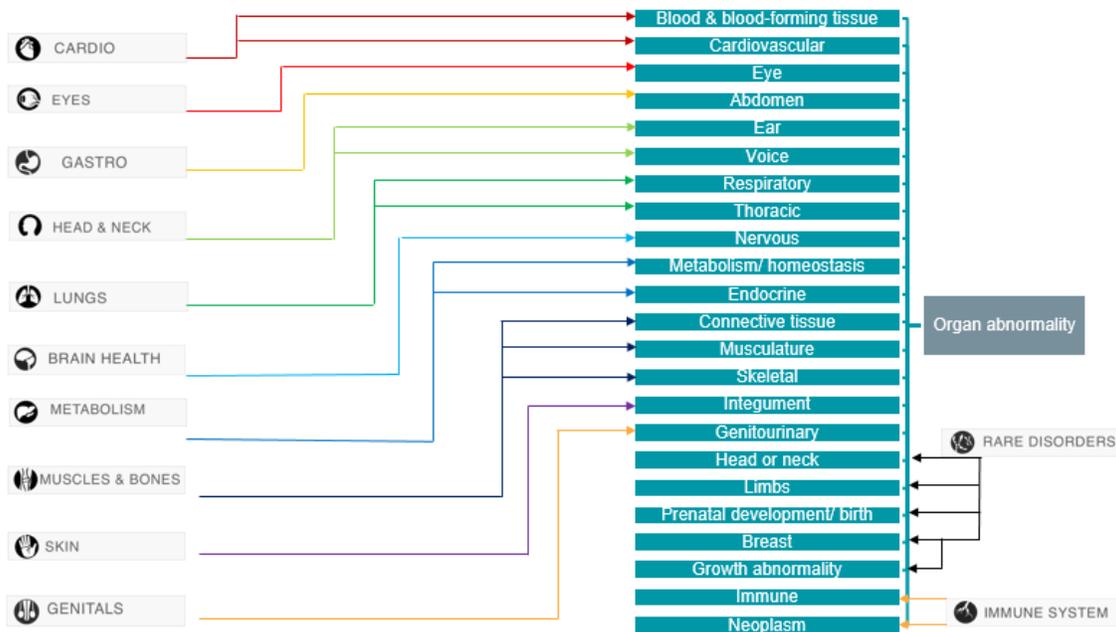


Figure 3.3 Mapping the HPO to the prototype categories (V2.0).

3.5. Design Thinking: Explore – Prototype

3.5.1. Extracting the HPO

The first step was to ask the company's Database gatekeeper to run a query on all the diseases on which the company reported results. This initial set of diseases derived from the curation of research literature and was based on the confidence level and available evidence of the reported correlation between biomarkers and clinical conditions. This decision to focus only on those conditions derived from two aspects: i) the company's medical history form was going to be used as a control in the user study, and ii) conversations with the data team on what type of claims can be done with the available literature and curated databases. More information on the process of determining evidence for analysis can be found on Calderón's dissertation (Calderon, 2020).

Data ontologies usually come either in .obo or .owl format, which remain plain text formats that capture the relations and hierarchies of the concepts and classes. In order to extract the relevant instances' IDs and their respective lay synonyms, a Python script was run that crosschecked the company's disease database's terms against the HPO's instances identifiers and lay synonyms and extracted those matches. The curated database was then inspected for accuracy. In this process, one of the main issues encountered was that some of the diseases' terms in the company's database were created in-house and represented the aggregation of several symptoms (i.e. inflammation treated as a health risk rather being dealt as a symptom). When running the script, all conditions that had *inflammation* were flagged, this involved manual screening of the company's disease database before running the script again. Later, for instances with multiple lay synonyms, the more accurate or fitting synonyms were chosen, for example Bladder cancer has three lay synonyms: i) Bladder neoplasm, ii) Bladder cancer, and iii) Bladder tumor. In that case, bladder cancer was chosen to avoid the terminology *neoplasm* and *tumour*, since there is the chance of the cancer being treated before metastasis or a major growth.

It is relevant to mention that most disease classes that act as nodes contain a definition, which could also be operationalized in VA systems. Further, some instances contain the UMLS and XREF identifiers which might provide a scaffold to link to other controlled vocabularies.

3.5.2. Prototyping

Having defined the categories as buttons and with a reduced database of diseases, the next step was the integration of all the assets. The prototype was built using Unity's 2019.1.f02 release. For the control, the disease categories and descriptive questions were digitalized and deployed using SFU's SurveyMonkey. The survey was designed with checkboxes for the conditions and text entry for descriptive data, the order of the categories was the same as the one used in the prototype. The decision for going with a digital version for the control rather than with a pen and paper one was made to avoid the introduction of bias that favoured the prototype just because of its digital nature.

The Colour Palette

One design constraint related to colour derived from the need to adhere to the company's colour palette. This restricted the available colours to choose from, since the palette was comprised of four main colours: red, yellow, blue and green. All the colours had a variety of two or three hues and there was an additional grayscale with ten levels of brightness (See Figure 3.4). This limited the use of colour for the main categories rather for the individual systems and organs. Nonetheless, such a limitation also allowed to avoid visual clutter, due to the overuse of colour.

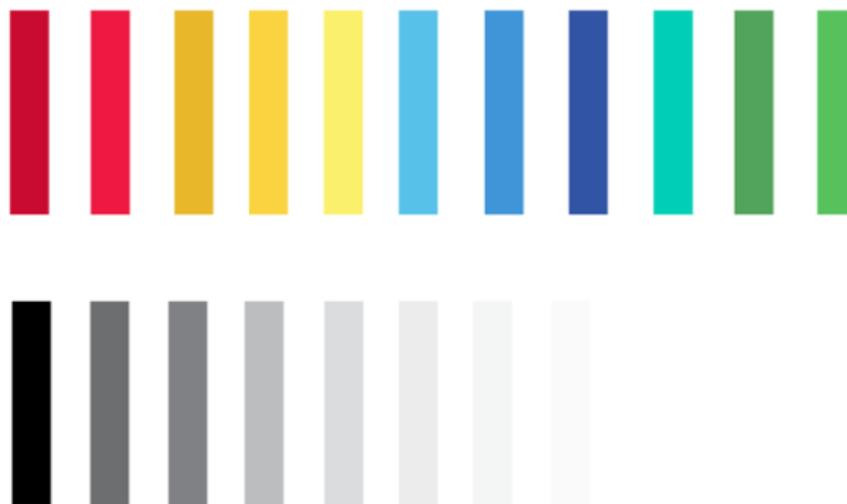


Figure 3.4 The Prototype's Colour Palette

Visual Saliency

As defined by William James, attention is “the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought” (James, 1890:403). In terms of Visual Analytics, designing for attention requires to harness preattentive processes, that is, automatic mechanisms that harness the parallel computing features of the Brain’s Visual Cortexes 1 and 2 (Ware, 2008b). This is to draw the user’s attention through visual saliency, especially when wanting to support bottom-up cognition to show the visual feedback. By doing so, it is possible to support Visual Thinking, defined as “Series of acts of attention, driving eye movements and tuning our pattern-finding circuits.” (Ware, 2008a:3). This is in line with Visual Analytics’ goal to support thinking through visual representations. As for preattentive features, the designer has at her hand variables such as colour, orientation, shapes and motion (Ware, 2012). For the present research, the chosen variables were colour for showing feedback and the shape and orientation of the somatic representation to allow the consumer to relate to the representation.

When designing for the user’s attention, visual saliency is a critical aspect to account in the process. Along orientation, shape and motion, colour serves as a preattentive feature for bottom-up attention. The way for the prototype to achieve this was through the use of a visual feedback mechanism that highlighted the selected organ, body system or region with the categories’ colour in order for the health consumer to relate the class with the related conditions. This can be seen in Figure 3.5 where driving attention also involves leaving the background in grayscale, and lower brightness and saturation than the highlighted areas. Since the interactive visualization would be one of their first contacts with the company’s assets, it served to familiarize the health consumers with the design assets. Ironically, the design assets were not only one of the more robust aspects of the prototype, but also presented limitations, specifically the vector images. When using the body graphic as a visual feedback, the line weight of the more granular assets made the highlighting barely noticeable. This serves as a subtle reminder of Weber’s Law when aiming to capture the user’s attention by achieving saliency using colour.

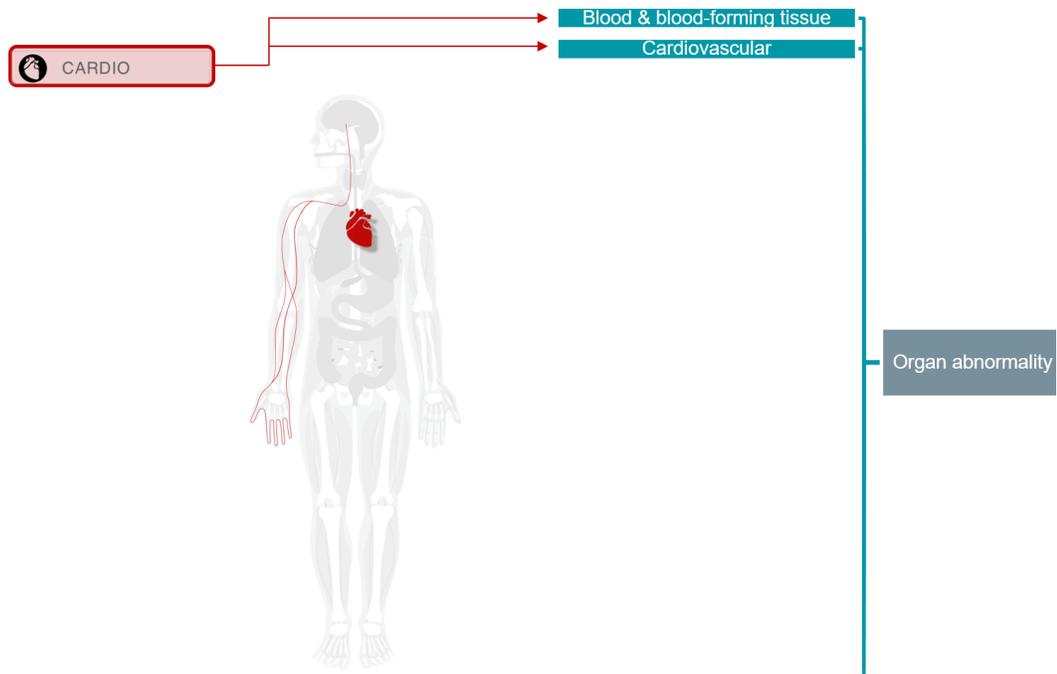


Figure 3.5 Cardio Button with Visual Feedback

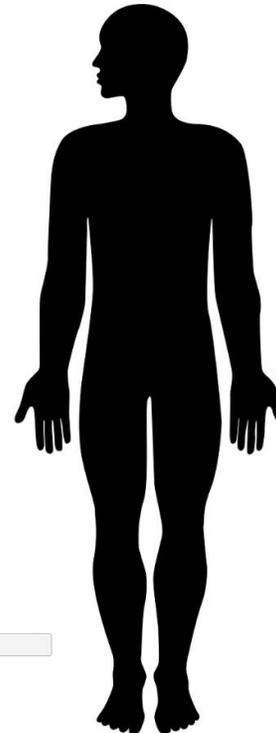
Interaction

Participants were first presented with an introductory screen that provided context and guidelines into how to deal with the body graphic. The initial screen also captured the most relevant symptoms that the health consumer would be suffering at the moment (See Figure 3.6). Once participants were done with the initial scene, they were led to the exploratory stage where they could report their medical history. Through interaction with the interactive visualization health consumers could navigate through functional systems such as the immune system, and body sections (Figure 3.7). This reduced the need to comprehend medical terminology. The lay synonyms of the HPO replaced medical terms for easier disease recognition during exploration. Finally, the data infrastructure was expandable, supporting the design of visualizations that match innovations in analytic methods.

Think of the body graphic on the right as a representation of you and your health.

Please check all the conditions that apply:

- Fatigue
- Weakness
- Fever
- Changes in appetite
- Recent weight gain, how much?
- Recent weight loss, how much?
- Poor concentration
- Increased Thirst



Next

Figure 3.6 Prototype's Introductory Scene

- HEAD & NECK
 - BRAIN HEALTH
 - EYES
- CHEST
 - CARDIO
 - LUNGS
- MUSCLES & BONES
 - BONES
 - JOINTS & MUSCLE
 - SKIN
- METABOLISM
 - KIDNEYS
 - METABOLISM
 - LIVER
- IMMUNE SYSTEM
 - RARE DISORDERS
 - IMMUNE SYSTEM
- GENITALS
 - GENITALS
- GUT
 - GASTRO

- High blood pressure
- Cardiovascular disease
- Disease of the heart muscle
- Heart attack
- Narrowing and hardening of arteries
- Elevated total cholesterol
- Decreased circulating cholesterol level
- Blood clot in vein
- Coronary artery disease
- Ischemia
- Blood clotting disorders
- Chest pain
- Palpitations
- Fainting
- Anemia
- Swollen legs or feet
- Stroke
- Angina

Next

Figure 3.7. Interactive Visualization Prototype.

3.6. Design Thinking: Materialize - Test

3.6.1. Research Setting and Data Collection

The user testing was conducted in a semi-private room in a graduate space located in Simon Fraser University. Recruitment happened in two ways, the first one via SFU's Research Management System (SONA), and the second one through the aid of social networks (i.e. Facebook's SFU "Graduate Students and postdocs" group). There was no monetary reimbursement and SFU students from participating courses got one course credit.

The health literacy and the system usability surveys were filled in paper and then digitalized using Excel. The prototype was coded in Unity and the user's clicks were recorded in a log. The interviews were recorded using a Zoom H4n Pro recorder.

3.6.2. Study Procedure

1. Ethics and Consent Review

Participants were initially explained about the goal of the project and the consent form to ensure participant agreement with the use of their data and procedure of the study. After the participants signed the consent form, their Participant ID was assigned and entered in the health literacy and usability questionnaires. The Participant ID was the only identifier used in the study from this point on.

2. Health Literacy Questionnaire

Participants were asked to fill the HLS-EU-Q16 questionnaire (Sørensen et al, 2013) (See Appendix G) and their scores were used to assign them to a group. There were three possible levels for Health Literacy (HL): i) inadequate HL, ii) problematic HL, and iii) sufficient HL. Participants with score resulting in an assessment of "Inadequate" were assigned to the layman group. Participants with results of "Sufficient" were assigned to the trained group.

3. Training: Clinical Vignette Review

Beyond the physician's involvement to help reduced the health literacy barrier for health consumers, the physicians also played an important role regarding the present

research, specifically in terms of the study design for the quantitative inquiry. Two clinical vignettes, one for males and one for females, were designed with the support of two physicians. The first physician helped to construct the clinical history and symptoms that resembled the target population: i) 19-35 years old university students with high work and stress loads, ii) seasonal allergies, iii) recent detection of Type 2 diabetes (See Appendixes D & E). The second physician provided a language review for the clinical vignette to *translate* clinical terms into lay synonyms (i.e. Allergic Rhinitis into Hay fever). By doing so, the clinical vignettes were designed to reduce the need for health literacy and to be easily understandable.

During the study, the participants were presented with the vignette and a glossary and were given up to 10 minutes to thoroughly read the scenario. Participants had the vignette at hand during the study and they were encouraged to make notes on it. The clinical vignette and glossaries were taken away before the interview in order to assess what symptoms and/or parts of the medical history the participants could remember.

4. User testing

The study's medical history vignettes were designed in lay language to allow participants to report on the conditions and symptoms from the scenario independently, and in as realistic a manner as possible—without guidance, coaching, praise, or critique from the researcher administering the study. It is important to mention that the participants never disclosed their own personal medical history but used the information in the vignette to simulate the scenario. The user test was separated into two critical tasks:

- Read the instructions for the data capture form
- Interact with the data capture form

During the interaction with the electronic health record the researcher took observational notes and recorded questions and comments made by the participant. The audio and screen were recorded with Camtasia.

5. System Usability Scale Questionnaire

After using both interfaces, the participants filled out the System Usability Scale (U.S. Department of Health & Human Services, 2020), an instrument for measuring the usability of a computational system. It is a 5-point Likert scale of 10 items.

6. Semi Structured Interview

An Open-ended, semi-structured interview was conducted after the participants finished filling the usability questionnaires. The questions intended to elucidate if having a direct feedback via the graphical representation had an impact on their behaviour while filling the form (See Appendix I).

3.6.3. Participants

The main population target was 18+ years old design undergraduates, but the recruiting encompassed the whole university community. Participants who were underage or were not fluent in verbal and written English were excluded. For the analysis, participants were given random names to avoid a further depersonalization of their roles by referring to them only by their ID numbers (See Table 3.4).

Table 3.4 User Testing Participants Summary

ID	RANDOM NAME	GENDER	AGE
ID01	Jessica	Female	24
ID02	Lauren	Female	24
ID03	Christopher	Male	31
ID04	Vincent	Male	22
ID05	Samantha	Female	21
ID06	Joshua	Male	28
ID07	Isabella	Female	23
ID08	Jack	Male	34

3.6.4. Data Analysis

User Testing

Descriptive statistics, such as mean and standard deviation, were run using JMP 14. Histograms for the Prototype and Control Scores. False positives were also charted

to assess the possibility of running inferential statistics. Due to the sample distribution and low population, it was decided to stick to descriptive statistics.

Thematic Coding

The same methodology from Geisler & Swarts (Geisler & Swarts, 2019) was employed to create a coding scheme (See Appendix J), one PhD candidate in Engineering (N=1) was asked to do a second coding to calculate the simple agreement and Krippendorff's alpha for corrected agreement. The goal of this coding scheme was to find parallels between the health data analysts' codes to craft design guidelines. The recommendations are meant to account for the information needs from the expert users, and the experiences and challenges faced by health consumers.

3.7. Chapter Summary

Chapter 3 served to provide a thorough description of the research design, showing how a pragmatic worldview supported the use of mixed methods to tackle the problems of technology design for health consumers of precision medicine solutions. In that sense, Norman-Nielsen DT framework provides a scaffold to sequentially explore the problem, starting with short-term ethnography for design and then moving to user testing for evaluation. This required for the qualitative approach to support the interviews in intercoder reliability measures, and for the quantitative aspect to use standardized surveys such as the SUS for usability and the HLS-EU-Q16 for health literacy.

Finally, the prototype was designed to support visual salience and grab the user's attention through the visual feedback. It was built based on the Human Phenotype Ontology as a data architecture, as it contained lay synonyms and integration with standardized controlled vocabularies such as the UMLS. Chapter 4 will focus on showing how the interviews with the data analysts led to crafting design and methodological guidelines, and how to understand and discuss the results from the experiment in relation to the design of digital healthcare technologies.

Chapter 4. Results & Discussion

This chapter aims to present and discuss the results from the research. The Chapter is divided in two parts, the first part presents the results from the research and the second part discusses the findings. The results show two design and two methodological guidelines that were created based on the field immersion working with the health data analysts at the company and the interviews from the user testing. The discussion is divided according to the guidelines found in the results, which in turn refer back to the hypothesis of the research. The Methodological Guidelines 1 and 2 refer to hypothesis 2.1 and the Design Guidelines 2 and 3 relate to Hypothesis 2.2. Then, the overarching Hypothesis 1 is discussed in relation to the research, design, and methodological limitations of this work. The discussion is framed like a conversation between health data analysts and health consumers.

4.1. Results

4.1.1. Bridging Health Data Analysts and Caregivers

In terms of understanding the communication loop between health data analysts and caregivers, eight main codes were identified. These codes were then translated into two methodological guidelines for short-term ethnography in VA, and two design guidelines for interactive data-capture systems. The set of aggregated codes to design guidelines can be seen in Table 4.1. The codes allowed to understand the nature of the analyst's work in relation to their job role, teamwork, and the need to constantly accommodate for the health consumer's lack of health literacy. In that sense, the health data analysts faced several data-related challenges that derived from getting unstructured, descriptive data from the consumers. In turn, this contrasted with the amount of data that they actually needed to conduct their job. Further, through the codes it was possible to see the sense of accountability that the analysts had, especially when talking about their current automation efforts and how the product had evolved over time.

When reflecting about short-term ethnography and methodological guidelines, by exploring the analyst’s job role, experience working as a team, and how their work is informed, it was possible to understand the analyst’s imperative of accommodating the lack of health consumer’s health literacy in their analysis. Therefore, the first design guideline refers to the need to accommodate all the stakeholders in the design process, and more importantly, to always design from the perspective of the most vulnerable population (Methodological Guideline 1 in Table 4.1). Finally, by inquiring about the minimum amount of descriptive information required to perform the interpretation of biomolecular data, and the different data-related challenges that analysts faced, it was possible to trace and critique the production of biomolecular data (Methodological Guideline 2 in Table 4.1).

As for the design guidelines, when looking at the ongoing automation efforts, and understanding the sense of accountability of the Data Team when interpreting data and providing recommendations, it was possible to reflect on how to capture the data more accurately using a controlled vocabulary (Design Guideline 1 in Table 4.1). Additionally, by exploring the evolution of the product it became clearer that there was a need to reduce abstraction by making the capture of medical history and descriptive data visual and interactive (Design Guideline 2 in Table 4.1).

Table 4.1 Codes to Design Guidelines

Code	Guidelines
Job Role	<u>Methodological Guideline 1:</u> Identify and design from the perspective of the most vulnerable stakeholder and prioritize those needs when incorporating other stakeholder’s requirements.
Teamwork	
Accommodating Health Literacy	
Data Challenges	<u>Methodological Guideline 2:</u> Trace the production of data to understand the shortcomings of its creation and always maintain a critical attitude to what the data should represent.
Minimum Data	<u>Design Guideline 1:</u> Support automation by using a controlled vocabulary, such as a clinical data ontology, to streamline the data capture process.
Accountability	
Automation	
Product Evolution	<u>Design Guideline 2:</u> Accommodate the health consumer’s health literacy through visual mapping of the controlled vocabulary to facilitate interaction and visual feedback.

For the coding scheme's intercoder reliability, simple agreement and Krippendorff's alpha were calculated. The rule of thumb of a corrected agreement above 0.7 was used for deeming the coding scheme good enough to proceed with the user testing (See Table 4.2).

Table 4.2 Measures of Intercoder Agreement

	N=2	Average	SD
Simple Agreement		0.729508	0.040984
Krippendorff's alpha		0.710333	0.042223

4.1.2. Bridging Health Consumers and Health Data Analysts

Participant's Health Literacy

Participants were divided according to their health literacy in two groups, three participants had sufficient health literacy, whereas five had problematic health literacy. (See Table 4.3). On average, the health literacy was 11 (SD=2.712) which falls under the problematic health literacy.

Table 4.3 Participant's Health Literacy

ID	Name	Health Literacy Score	Health Literacy		
01	Jessica	14	Sufficient		
02	Lauren	9	Problematic		
03	Christopher	9	Problematic	N=8	
04	Vincent	9	Problematic	Median	11.5
05	Samantha	11	Problematic	Mean	11.75
06	Joshua	14	Sufficient	Std Dev	2.7124
07	Isabella	12	Problematic		
08	Jack	16	Sufficient		

User Testing Results

The experiment compared the interactive visualization (prototype) against a digital version of traditional medical history forms (control). There were eight participants, 4 males and 4 females, with a range of ages going from 21 to 34. Each participant was given

a clinical vignette with 10 symptoms/conditions that they had to report in the experiment, the summary of results can be seen in Table 4.4. Relevant variables for the study were the score (how many conditions they reported), the duration time and how many conditions were reported that were not originally in the clinical vignette (false positives). A descriptive statistical analysis showed that the score mean for the prototype was 5 (SD= 1.852) and for the control was 5.625 (SD= 1.923), since the distribution wasn't normal the score median for the prototype was 4.5 and 5.5 for the control. The Control was able to capture up to one more condition than the Prototype (See Figure 4.1).

Prototype vs Control Scores

(Total score was out of 10)

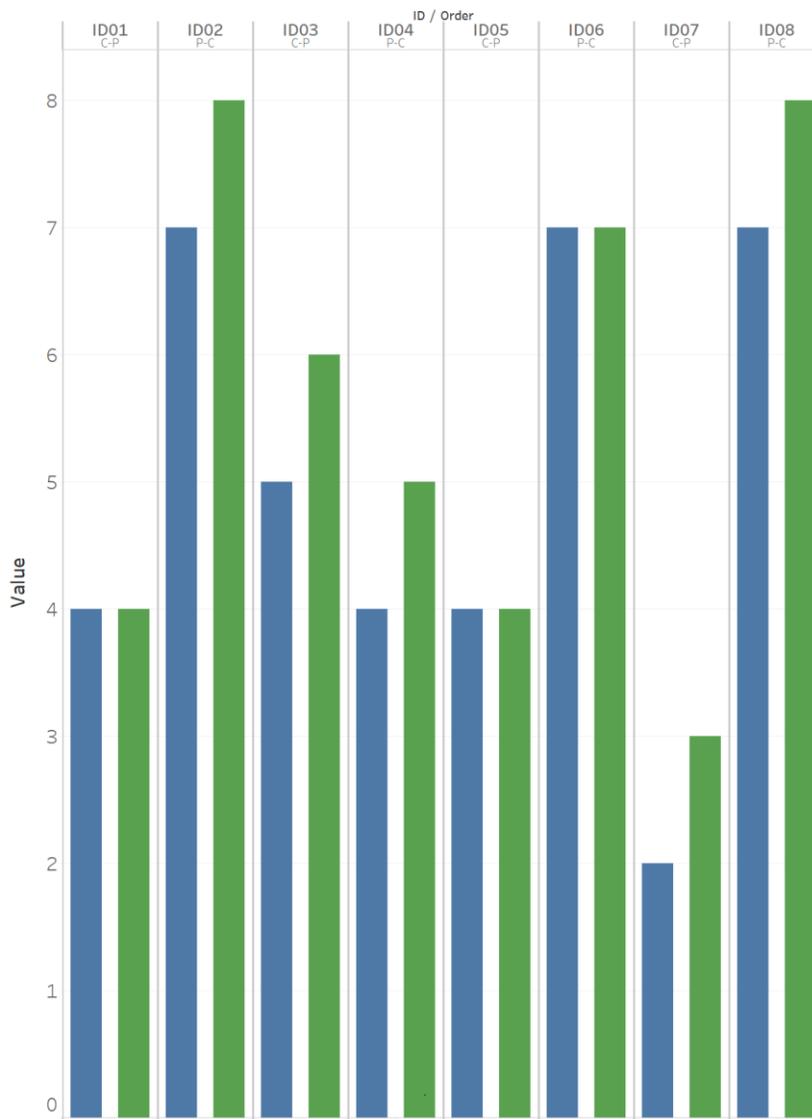


Figure 4.1 Prototype vs Control Score

In terms of false positives, the prototype captured on average 1.625 conditions (SD= 1.767) and the control 5 (SD=5.707) with a median for the prototype of 1.5 and 2.5 for the control. For the summary details see Table 4.5.

Table 4.4 Study Summary

<i>N=8</i>				Prototype			Control		
ID	Name	Gender	Age	Score	Duration [min]	False Positives	Score	Duration [min]	False Positives
01	Jessica	Female	24	4	5.78	3	4	15.5	16
02	Lauren	Female	24	7	1.52	2	8	3.67	2
03	Christopher	Male	31	5	1.58	0	6	3.72	1
04	Vincent	Male	22	4	6.12	2	5	7.93	3
05	Samantha	Female	21	4	2.83	5	4	9.68	6
06	Joshua	Male	28	7	7.22	0	7	3.3	0
07	Isabella	Female	23	2	1.93	1	3	9.72	11
08	Jack	Male	34	7	11.05	0	8	7.22	1

Table 4.5 Experiment's Descriptive Stats

<i>N=8 </i>	Prototype	Control
<i>Score is out of 10</i>		
Median	4.5	5.5
Mean	5	5.625
Std Dev	1.8516	1.9226
False Positives	1.5	2.5
Median		
False Positives	1.625	5
Mean		
False Positives	1.7678	5.7071
Std Dev		

For the system usability, the prototype had a mean of 68.75 and a median of 70. The rule of thumb for the SUS is that an interface is good with a score of 70 or more (See Table 4.6).

Table 4.6 System Usability Scale Summary

N=8	System Usability Scale (SUS)
Median	70
Mean	68.75
Std Dev	6.5465

In terms of confirmation of understanding, participants were asked all the details they could remember about the health consumer after they used both interfaces. On average, participants remembered 4 conditions (mean=4.375, SD=2.446). No descriptive statistics were calculated for the conditions that participants remembered incorrectly, that is those conditions which weren't present in the clinical vignette (See Table 4.7). The false positives were "weakness" two times, "brain cancer" one time and "fever" one time.

Table 4.7 Confirming Participant Understanding

ID	GENDER	AGE	CONFIRMA TION	FALSE POSITIVES		
Jessica	Female	24	3	0		
Lauren	Female	24	8	0	Median	4.5
Christopher	Male	31	6	0	Mean	4.375
Vincent	Male	22	2	1	Std Dev	2.4458
Samantha	Female	21	3	1		
Joshua	Male	28	6	0		
Isabella	Female	23	1	1		
Jack	Male	34	6	0		

Qualitative Coding

The participant interviews helped to better understand the background of health consumers, and later to dissect how they went about the interface and the activity. By looking at *Health Consumer Experiences* and *Health Literacy*, it was possible to see how they navigated their relationship with the healthcare system, and how their own health literacy constricted or supported that. The codes *Interaction* and *Value of Visualization* helped to elucidate how the participants went about using the interface, from simple

aspects as their initial wayfinding through the categories and visual feedback to their perceived advantages of the system. Finally, *Interface Limitations* and *Interface Recommendations* dealt with the issues that arose when they could not report on certain conditions, and further design considerations that would have better supported their activities. These codes helped to correlate the needs of both stakeholders to craft the design and methodological guidelines for the design of VA systems for precision medicine (quote samples are found below). For reliability, the simple agreement and Krippendorff's alpha were calculated. The agreement was below expectations with a simple agreement of 0.69 and Krippendorff's alpha of 0.66.

- Health consumer Experiences

Joshua (ID06): *When thinking about muscles and bones, I was also wondering if there was a "limbs" category, for example a friend of mine just disjointed the shoulder and I would look for limbs to report that*

- Health Literacy

Lauren (ID02): *Since I'm not a doctor or someone who studies biology or not. I don't know what category the disease belongs to.*

- Interaction

Christopher (ID03): *I thought about the problems, for instance hay fever, so hay fever should be either from the metabolism or the immune system, so I selected the immune system, going top to bottom, and then clicked on the disease.*

- Value of Visualization

Jessica (ID01): *The main purpose of the visualization was to help people to find and self-check the body in terms of health, because sometimes we go the doctor and cannot explain where I feel the illness. So, I could use this interface to check myself before I go to the doctor.*

- Interface Limitations

Samantha (ID05): *I'm not sure how the colours related to the organs or if they just served to differentiate parts of the body*

- Interface Recommendations

Isabella (ID07): *I couldn't understand the different diseases very much, so clicking the body graphic was good. It would be better to have more explanation of the words (diseases), like having an "I" that shows more information, because I couldn't understand them.*

4.1.3. Understanding Health Consumer and Caregiver Communication

The health consumer consultation continues to be the gold standard for providing clinical diagnoses to health consumers. In that sense, different healthcare systems across countries present different problems.

Joshua (ID06): *We have two issues in Canada related to medicine and this whole concept of depersonalization of healthcare: we don't have enough family doctors, and the people who do have family doctors, family practices are often overburdened. So, to get an appointment with your GP it takes like 2 weeks, maybe 1. That's basically it, we don't have enough family doctors, so you don't have these relationships. With a walk-in clinic you can go within a day or two and get checked out, a lot of people are going there but the issue is that the doctors don't really know you.*

When communicating this to their caregivers, new challenges arise for health consumers who are new to a country or the language:

Isabella (ID07): *I haven't been too long in Canada and one of my problems when talking to doctors it's hard for me describe where I have pain and what the actual location is.*

4.2. Discussion

The results show that there is a need to interlink health consumers and health data analysts when designing VA systems for precision medicine. By themselves, the quantitative results fail to provide a deeper insight into the situation of health consumers dealing with medical history forms. Through merging the qualitative results, and setting the voices of health consumers and the Data Team in a conversation, is that an ethnographic place can be created, one where there is a triangulation of both types to data to inform design guidelines that speak to and are catered to both audiences.

4.3. Understanding and Designing for Stakeholders

Methodological Guideline 1

Identify and design from the perspective of the most vulnerable stakeholder and prioritize those needs when incorporating other stakeholder's requirements

4.3.1. Stakeholders Analysis

Analysts

The Data Team was comprised of health data analysts whose role was to interpret and report biomolecular data in correlation to a health consumer's descriptive data to provide actionable health recommendations. Despite being a multidisciplinary team, comprised of members at different stages in their careers, the Data Team member had a consistent understanding of the definition of what a team was. For the health data analysts, a team was a group of individuals who have, work for, and contribute to a common goal. As stated by Antonio:

Antonio: *I want to say like, I like the machine analogy where you're talking about different moving parts and ultimately, they are all moving towards a certain goal. So, I think that really relates each individual cog in the system has their own dedicated role, but the team at the company really hasn't been perfect to that analogy. There have been*

multiple members of the team do fit certain roles, but those roles have evolved over time. And certain roles have either been removed, because the data showed they were not necessary vs. other roles... other tasks have been created. If we stay with the machine analogy, gears have been removed, certain grids have been adjusted, gears have been made larger.

Belonging to the Data Team was not only an opportunity to gain new knowledge and skills, but also a way to create interpersonal relation with the different members. In that sense, the Data Team also provided its members with a sense of belonging and pride in working for the organization. This can be seen in their choice of three words to describe working at the company (See Figure 4.2), where innovative was a common thread in 60% of the Data Team. Different dimensions related to the nature of the Data Team's work such as *ground-breaking, exciting, challenging, collaborative*. Others related to the spirit of the Data Team such as *diverse, fun, independent*. Finally, others related to the work involving sensitive data and the desire to provide understandable and actionable recommendations, like *emotional, fulfilling, and complex*.



Figure 4.2 Three Words That Describe the Analyst's Work

In this sense, the members of the Data Team conceived the team as a unit with a single goal. Considering Antonio's more holistic view of the Data Team, the team could be considered a cognitive unit in which, in alignment with rapid changes in the team's composition, all the cogs had a script, protocols and artefacts to carry out their jobs (Hollan et al., 2000):

- Protocols

Gail: *Generally, things have been quite smooth. Obviously in the beginning we had to work out the process to create the reports. We have three to four people involved in these reports, so figuring out that handover was really important. We pretty quickly started using Project management tools like Asana. We kind of did the handover outside the system that was built into the platform. In the platform there was a button to assign the report to a certain person, but all they would get is a notification in their email, it didn't tell them which one it was, because you can't send that information by email. We started putting codes for each of the reports in Asana and then we would move it to the next column. That's how someone would know if that person was assigned to them.*

- Scripts

Antonio: *Right now, we have an assembly line or batch process for creating the recommendations. When I'm done with something, these reports are given to the health consumer recommendations team and they are expected to go through, check the notes, and put a health consumer's recommendations together. If they have questions we go back and forth but it is almost like an assembly line*

Gail: *Once a report goes through QC, and is uploaded to the production DB, it's then assigned to either Antonio or myself. We do consult each other when we need the expertise of the other, for example Antonio knows more about the immune system than me.*

- Digital Artefacts

In addition to pen, paper, and other analogue artefacts for note keeping, most of the work of the Data Team revolved around the use of digital tools. Among them, the Google Suite was one of the main artefacts shared by the Data Team, followed by the MS Office Suite. The digital tools related to the different types of work that each individual member conducted. From artefacts that supported the communication protocols like Asana, Email client, and the internal platform to very specific tools such as Illustrator for creating biological diagrams, Tableau for data visualization, and Python & R for statistical analysis (See Table 4.8). The main databases that the Data Team used were PubMed, Google Scholar and PEN (Practice-based Evidence Nutrition).

Table 4.8 Analyst's Digital Artefacts

Antonio	John	Gail	Katie	Wilma
Email Client	Excel	Asana	Google Suite	Internal Platform
Excel	Google Suite	Google Suite	Monday	MySQL
Google Search Engine	Illustrator	GoogleScholar	PubMed	R
Google Suite	Internal Platform	Internal Platform	PEN	Tableau
PubMed	Python	PubMed		
RapidMiner	R	Slack		
Tableau				
Uberconference				
Word				

Through the data analysis and interpretation process, there was a notion of how the Data Team needed all its members to work together. Therefore, the Data Team⁸ can be considered a cognitive unit who was capable of:

1. Creating and curating databases for:
 - a. Genetics
 - b. Pharmacogenomics

⁸ From this point on, the Data Team will be dealt as a single unit, with the individual analyst quotes being attributed to the data team rather than to them. Exemptions will be stated clearly and made only if necessary.

- c. Proteome-related Health Risks
 - d. Nutrition Recommendations
 - e. Exercise Recommendations
2. Performing Quality Control and Quality Assurance (QA/QC) of biomolecular data
3. Selecting actionable and relevant biomarkers
4. Analyzing and interpreting the relevant biomarkers in correlation to descriptive health consumer data
5. Creating personalized reports:
 - a. The health consumer's health risks and highlights report
 - b. The actionable recommendations for lifestyle, nutrition, and exercise report.

Furthermore, when considering the perspective from anyone outside the company, especially health consumers, the individual members are not relevant or known to them. The individual expertise and skills of each data team member are only relevant in how they work together to produce a report that summarizes the health data and provides recommendations. Yet in order to produce those reports, health data analysts required consistent quality in the biomolecular data:

The Data Team: *My expectations are to get an omics panel to a point where we can get reliable values. It's been a little shaky because there has been quite a few profiles where we flagged errors from the companies to which we outsourced this stuff to, we've caught those errors. That's been the hardest thing to stomach, is the fact that what we are relying on to be robust and reliable may end up being actually not. If that's the case, then this whole thing is a complete waste of*

time. Because we are trying to make an association, an understanding of how to use multiple biomarkers to help assess someone's health.

We expected to have a certain level of consistency, but that would be with any scientific platform. It's the expectation of... if you say that you're using the technology, then you know which error rate/variance you're bringing with that technology and you can have a certain of consistency or reproducibility from one run to the other.

They also required good quality in descriptive health consumer data, specifically in order to be able to better tailor their recommendations to the health consumer's actual goals:

The Data Team: *We are getting a snapshot of information about them, so the information we can provide can be somewhat personalized in regards to their data that comes back. But it's not the same as meeting and doing an individual assessment. All of our answers are personalized to them, however it is automated.*

Health consumers

One of the main barriers of entry for health consumers to the healthcare system (Nolasco et al., 2010), specifically precision medicine, is the need to interact with digital or analogue artefacts to report on their conditions before seeing a caregiver. This leads to issues regarding the health consumers' health literacy and own experiences:

Vincent (ID04): *Since I'm not a medical student, I don't understand which situation of the body matches each illness.*

Jack (ID08): *It's confusing for diseases like hay fever because I thought it would be in the lungs and not in the immune system. I was thinking it was a respiratory problem*

Further, the consumer's challenges are not restricted to the embodied experience of illness, but also in not knowing how to abstract that feeling to a medical terminology they are not aware of:

Lauren (ID02): *It was hard for me to find where the type 2 diabetes was located, I had to click on each to find where to put it. The categories made sense but were hard to use because I'm not a doctor or someone who studies biology or not. I don't know what category the disease belongs to.*

This can lead to health consumers failing to fill the form under the assumption of having a good health, and failing to conduct a proper *division of labour* (Kaptelinin & Nardi, 2006), which would lead to the analysts having less descriptive health consumer data to correlate to the biomolecular data:

Vincent (ID04): *Usually when I go to the doctor, I don't have any medical history, so I just leave that part blank. So, I don't have much experience filling these medical history forms*

Therefore, health consumers require interfaces to report on their medical history and symptoms that support problematic health literacy, and that match their mental models of body and disease.

Caregivers

Caregivers rely on having accurate descriptive health consumer data along with a clinical history to provide a diagnosis for the health consumers. In the organization, physicians provided a pivotal role in facilitating health consumer understanding of the complex content:

The Data Team: *the Physician is like a magician with multiple roles, because she manages to break the science down into easy, tangible words that she has actually truncated 90% of it and given the health consumers a motivation and an action and they are gone, and done it and gotten healthy. In the end of the day, the person doesn't need to know about all the ins and outs about how the body works, they need to know what they need to in their day-to-day life to stay away from disease.*

For the present research, it is relevant to note that caregivers are considered as part of the Data Team. In reality, a proper analysis of caregiver's needs should focus on

those professionals who are providing active care and working directly with health consumers.

4.3.2. Designing for Health Consumers

Reporting on Medical History and Symptoms

During the analysis and interpretation process, health data analysts had an almost complete overview of the health consumer's biomolecular and descriptive data. Health consumers only interacted with the health data analysts through the company's medical history form. Therefore, in order to understand the nature of the health consumer's work, that is, the activity of health consumer's filling a medical history form, the activity was broken into its components according to Engeström's Activity Theory model (Engeström, 1990), as applied through the lens of Kuutti & Arvonen (Kuutti, 1992) (See Figure 4.3). The goal was to identify the areas where health consumers would need support:

- *Instruments/Tools*: Data capture forms, be it in the form of questionnaires or interactive visualizations. Web search engines can serve as the tools with which the health consumer will interact when reporting their descriptive clinical data. The main goal is that the artefacts serve to capture the descriptive health consumer data in a controlled way as to support the automatization of downstream elements of VA systems.
- *Rules*: Are defined by the data infrastructure, in this case by visually mapping the HPO to a body graphic. By doing so, the HPO defines what and how information will be captured (e.g. lay synonyms or capturing broader categories of the ontology). Further, by dividing the interactive visualization in stages, the health consumer gets to focus on single aspects on each interaction scene.
- *Division of labour*: While the present analysis focuses on health consumers using interactive systems for reporting on their descriptive data, the scope of Engeström's Activity Theory can be used to relate the health consumer's descriptive data as the main resource of the health analysts to conduct their work. In the present case, the quality and structure of the data is an essential variable for it to be actionable by health data analysts.

- *Subject:* Through the exploration of the interactive system, the visual feedback serves to create a dialogue through predetermined actions in response to a thought process (e.g. if the person wants to report on having diabetes, first they need to ponder about their medical history before exploring which button will lead to that condition).
- *Object:* The goal of the activity is directly related to the health consumer's health and wellness goals. Reporting on their medical history and symptoms constitutes the entry point for the dialogue with the medical provider or health data analyst. For precision medicine, it can be the health consumer's desire to lead a more balanced life, manage chronic conditions or increase their current performance.
- *Community:* When confronted with the high barrier of entry for precision medicine, the health consumer can ask support staff for advice or even contact their health provider when confusions arise. In the worst-case scenario, the health consumer would seek advice in open clinical websites that might mislead the health consumer or create further anxiety and confusion. Nonetheless, health-focused health consumers are often part of communities, be it digital or presential.

From the analyst's perspective, and due to varying health literacy among the "lay" people, it was challenging to accommodate the health consumer's health literacy. Nonetheless, while the technology and data are getting increasingly complex (Topol, 2019), there is a general understanding in western society that the blood has little particles that can be used to get an insight into one's health status:

The Data Team: *There is a population-level perception on data and scientific theorems and from there, the vast majority of us are trying to pull from some basis. Even your average individual that has no idea from an omics panel or a fancy blood test. Even they are familiar to an extent, because they have gone to the doctor and gotten a blood draw. They had some level of at least back and forth, they get a blood draw, they get a result, an action and then they get better.... or they didn't.*

So they have an understanding of the fact that it's possible that their blood has some information about them that can make me better.

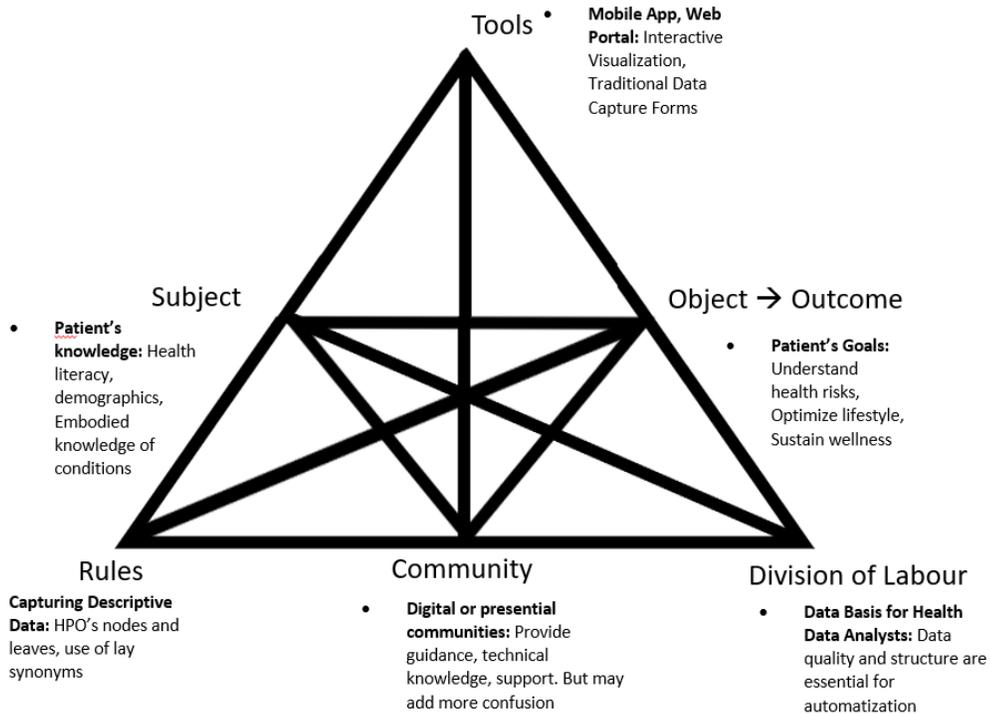


Figure 4.3 Engeström's Activity Theory model applied through the lens of Kuutti & Arvonen for health consumers filling medical history forms.

Furthermore, a bad result in a blood test would usually mean finding underlying conditions or health risks, whereas a good result meant that everything was working fine. It is under those eyes that health consumers see precision medicine products, as additional windows to their health, with the issue being that understanding the test results requires to deal with uncertainty in addition to the already vulnerable state in which they find themselves:

The Data Team: *There are some things that we know in humans more or less have to fall in a normal range. You kinda learn that by experience. For example, we once had a time where we were rolling*

out the exposome and we had a lot of markers that were out of range and we thought that it was weird, but we highlighted it for the analysis. An example was sodium, a lot of people have high sodium, people eat high-sodium foods all the time and then one of the customers went to their doctor and the doctor said that if that person's sodium was that high, she would be dead.

In that sense, accommodating a health consumer's health literacy becomes less about teaching the health consumers all the intricacies about the technology, but to make them be able to build on their existent basis of knowledge and mental models regarding blood tests to better assimilate the results and recommendations of a precision medicine test. Designing from a health consumer perspective also involved to understand the vast variability in how health consumers understand their own embodied experiences of disease, along other sociocultural limitations such as health consumers being new to a country or the language.

Accounting for Vulnerability

Digital health is becoming intertwined with AI solutions with an alarm light still going on in the background. The lack of understanding of the decisions and procedures that machine or deep learning algorithms make, creates an ever-increasing concern when the industry shifts their focus towards inscrutable, automated decision-making (Knight, 2017). Additionally, the free surrendering of such data puts the health consumer in a vulnerable position of disclosure, where the blood draw becomes a *test* in all sense of the word, since the report would show the health status of the consumer. Even with such a context, health consumers still retain the belief that improving communication may have a positive effect on their health outcomes, yet their perceptions on the medical information security were not positive (Small, Peddie, Ackerley, Hohl, & Balka, 2017). This context makes it even less clear for health consumers to understand why they should freely provide descriptive data in addition to highly sensible genetic data.

This vulnerability has been exemplified in multiple instances of data breaches that range from government-led initiatives (Regalado, 2019) to targeted hacks at vulnerable populations such as AIDS-positive patients. (Russell, 2019). Further, the privacy breaches become even scarier when there is documented account of algorithmic discrimination that limits access to healthcare, credit, and education among others (O'Neil, 2016). As stated

by Pasquale, there is a real impact on people’s lives, one that might even be invisible to those being oppressed: “Unlike the engineer, whose studies do nothing to the bridges she examines, a credit scoring system increases the chance of a consumer defaulting once it labels him as a risk and prices a loan accordingly.” (Pasquale, 2015:41).

From a biomolecular data perspective, there is a troubling past in technoscience misappropriating and commodifying genetic data from vulnerable populations, for example the trial between Arizona State University and the Havasupai tribe (Reardon & TallBear, 2012:238), when under the excuse of moving science forward, genetic samples were freely shared between labs for research. This simplification of genetic materials as a property, one that could be easily shared for science’s sake, undermined the property rights of the Havasupai tribe towards their own genetic makeup. While the genetic material from the Havasupai tribe was ultimately repatriated, the *ASU vs Havasupai* trial was settled out of court which left no legal precedent and exposed the limits of existing legal and regulatory frameworks (Reardon & TallBear, 2012:241).

Finally, this shows some of the challenges that health consumers will face when using precision medicine solutions, and that data privacy remains an area of significant concern (Small et al., 2017). In addition to the privacy concerns, the issues of property, and the risks of algorithmic discrimination, designing for the most vulnerable stakeholder requires having a critical understanding of where the data is coming from and what is done with it. More importantly, it places a burden on determining what data is essential to be collected and how that data might have collateral effects on the lives of the health consumers.

4.4. Data Provenance

Methodological Guideline 2

Trace the production of data to understand the shortcomings of its creation and always maintain a critical attitude to what the data should represent

After purchasing one of the packages from the company, the consumers of precision medicine products found themselves at the verge of transmuting a new dimension of themselves into data. Following the completion of the online registration and payment for the product, health consumers received via traditional mail a small box containing the kit for the blood draw and a flyer with instructions. This represented the first steps towards the health consumer having a profound overview of herself, a molecular-level understanding of whatever disorders they might have. In order to do so, the health consumer was instructed to:

- Choose a partner laboratory and schedule a blood draw
 - The list of laboratories was available in the company's webpage
- Bring the unopened kit & requisition forms to the laboratory
- Not eat/drink for at least 8 hours before the visit
- Not exercise for at least 24 hours before the visit

The health consumer's journey in becoming data was followed by the processing of the blood samples. This involved the clinical laboratory sending the samples to a different province, since the laboratories of the company were not located in the same city. After this journey, the blood samples were analyzed and broken down to single biomarkers as biomolecular data. These aggregated records travelled back electronically to the offices of the company where the data team would convert the numbers into actionable recommendations.

In order to better understand this process, an integral part of conducting short-term ethnography is for the ethnographer to put himself at the centre of the action (Pink & Morgan, 2013). It is in that regard that the researcher underwent the same process as a customer would during the fieldwork in order to explore empathy to consumers of precision medicine products. These results can be found in the Preface and Epilogue of this thesis as an essay exploration to understand the process that consumers of precision medicine products must undergo. Both parts of the essay exploration *Becoming Data* are meant to provide an additional perspective to the research, and to help create an *ethnographic place* (Howell, 2017; Pink & Morgan, 2013). While the present discussion has revolved

around the relation of the health consumers, caregivers and analysts, the researcher cannot be omitted from the process.

Through the process of deep immersion with the community, it was possible to trace the production of data and identify shortcomings of precision medicine that directly relate to the data. Furthermore, field immersion led to the researcher acting in different roles:

1. As a member of the Data Team:
 - a. Helping with data curation of DBs and visualizing data
 - b. Leading the project for the redesign of the company's onboarding data collection form
2. As a Researcher
 - a. Undergoing the process of *Becoming Data* to gain the perspective of a health consumer and understand their vulnerability
 - b. Conducting short-term ethnography with the Data Team
 - c. Designing and conducting user testing experiments

This multiple perspectives were essential lenses to be able to *unflatten* (Sousanis, 2015) the design of VA systems for healthcare, as it allowed for different vantage points to understand essential requirements. In that sense it was possible to make the invisible somewhat clear by providing a critical understanding of how health consumers, and for that matter everyone using a smartphone, is a constant data chimera. How the biomolecular data would now join the ranks of the digital version of health consumers, providing an additional layer of sensible and organic data to the database of a company and to the health consumer's *quantified past* (Elsden et al., 2016). As explored in both parts of the essay exploration *Becoming Data*, health consumers can be hesitant to provide their medical history, even if the outcome should be to their benefit due to the erosion of trust created by the industry itself (Charlsie & Denne, 2019; Russell, 2019). Further, that there must be an understanding from the researcher that each technology user's aggregated *quantified past* might have a real impact in the lives of the users, even

if the users might not even be aware of the existence of their digital self (O'Neil, 2016; Pasquale, 2015), and more importantly, that there is no going back.

4.4.1. Minimum Data Needed

Derived from the data context of precision medicine and the journey in understanding for consumers of precision medicine products, the sense of vulnerability in sharing one's medical history led to identifying the minimum data points that the analysts required to do their work. In that sense, the minimum aspects were:

- Data needed for adjusting reference ranges.

The Data Team: *Age is tremendously important for chronic diseases, if I see somebody with really high age but lot of... or if I see somebody with a low age and lot of chronic diseases, I know that not only is this report pretty important for the person to do something, because they have such a low age and they have all this chronic issues, it's even more vitally important for them to actually get ahead of this. Then age and sex are important.*

- Psychographic aspects regarding the health consumer's motivations

The Data Team: *The kind of data that the health consumer provides really is very similar to, I would say like the data they provide in a medical consultation. Maybe we ask a few more in-depth questions, but it's usually of physical nature. They are volunteering their own physical characteristics as well, age, sex, those kinds of things. We do ask them a little bit about what their motivations are, kind of like the psychographic aspects of things.*

- Relevant family history, medical history and medications background

The Data Team: *I think, family history, very basic family history. I don't want to know about your great aunt who is twice removed, so basic family history. I would maybe even limit it to certain conditions that are relevant to us. Sometimes we get things that we don't think about them but that when people have them they explain a lot. Medical history and*

medications. Yeah, medications would be important for the health data. The other things, I mean it is good to have a general idea about the person's lifestyle. I want enough information to build a visual of this person, kind of be able to guess who they are, what their life is like.

4.4.2. Technical Limitations in the Generation of Data

One of the assumptions of state-of-the-art technology is that it works perfectly and is fully validated for use, which might not always be the case. It is in that sense that the data interpretation can be hindered by the very same process of how the data is created:

The Data Team: *I know for a fact that a lot of the new technologies on the market have massive problems that no one wants to mention or address. I can speak to my masters in how we looked at brand new, high tech, next generation sequencing (next generation Illumina), which would give you all the transcript in a cell. We were having issues, for example we were working with muscle and all of a sudden, you see a pancreatic presence in there. And there is no freaking way that we all, contaminated and mixed the samples, we did it so many times. And then we noticed that the technology would have this phenomena called bar-code hopping, where there is actually no presence of that physical molecule in that sample, but the identification for that, from one sample to the other, hopped for some reason within the technology. Because this machine has to run for like 8 hours, in order for it to be able to take a look and go through the process of amplifying and identifying and quantifying all the RNA that are present in a cell. When we find this out, the people who got the shit for it was the grad students, "you guys fucked up!" you get shat on because you made a mistake. So, you go and do it again, you have three people in a tiny little fridge, I sat in a fridge to do this experiment with my colleagues. We went through an insane process; we had a full system set in place to avoid anyone from touching or contaminating the samples and we still got the problem. And then you go to the guy who is the tech, who is running the machine and he is in absolute complete denial that he messed up anything. And everyone thinks the other*

messed up, and then one guy says "what if the machine has a problem?" and we are all like "shut up". And a year and a half later, when you keep reporting this and sharing the data with the company (because the company is giving you some leeway and the reagents and the machine because you're testing it for them). They go and say "Oh BTW we find this phenomena called bar-code hopping" and they give you instructions to operate the machine to fix the problem.

Further, the risk of having so much variability deriving from aspects that are beyond the control of the Data Team reinforces the notion that designing VA systems for precision medicine can be framed as a wicked problem (Buchanan, 1992; Rittel & Webber, 1973). Due to the constantly evolving data context, there is no definitive formulation of a wicked problem and there can not be a true or false solution, there is rather a gradient between good and bad.

4.4.3. Data Mining Limitations

The technical limitations of the data extended to the different tools that were used to mine the health literature databases, specifically from problems derived from the acronyms, short forms of molecules, or language-related issues when parsing through large databases:

***The Data Team:** Back when the company was very new, they were using literature mining tools that the bioinformatics lab had made, but the problem with that is that these tools are not very good yet. They tend to dig up a lot of very strange things, I'll give you a more recent example, we were trying to calculate a confidence score for lead, the metal, and inflammation...so I think the way Antonio was doing this analysis, but basically I had given him a paper that said "high lead in the blood is related to inflammation", I mean no surprises there, but then he did this calculation were you go into PubMed, you put in the PMID and then the keywords you are looking for and it checks how many papers have been published with those keywords and how many citations they have. So you can do it for a specific paper or in general and that's the accurate score. And we got a stupidly high score, like*

10000 something like that... and that seemed surprising because you wouldn't think there would be that many papers about lead and inflammation. And it is because it's just looking for the word "LEAD", so it found everything that has like "lead to inflammation". So that's the problem with these data mining tools, they still don't have the intelligence to weed those things out automatically, that's where a lot of the tedious human work comes in and why the company is so fun to build.

Finally, by tracing the production of data, it was possible to understand the technological context of precision medicine in relation to the efforts of achieving automation, data stewardship, and the nature of a health data analysts' job role. In that sense, tracing the data production was an integral step for the framing of the problem and in knowing what are the essential information aspects that are needed to conduct the analysis and interpretation of biomolecular data. Finally, visually mapping a robust data infrastructure to a health consumer interface can be seen as an integral part of the design of VA systems, since it allows to support the analytical needs of health data analysts by capturing data accurately.

4.5. Qualitative Hypothesis

The use of a cognitive lens for short-term ethnography helps elucidate the vulnerability of health consumers and the data needs of analysts for the design and use of VA systems in the context of correlating health consumer descriptive data with biomolecular data.

The hypothesis 2.1 was supported, since by using a cognitive lens (Hutchins, 2000) for short-term ethnography (Pink & Morgan, 2013) it was possible to capture the *intensity of the research encounter* through working with the Data Team (See Methodological Guideline 1 in p. 85) and having a *focus on details* by undergoing the same process a health consumer would (See Methodological Guideline 2 in p.96). Through this process it was possible to identify and capture design guidelines that related to the most relevant stakeholders, firstly the health consumers and secondly the data analysts for creating user interfaces (Millen, 2000). As presented in the introduction, the traditional

communication loop between health consumers and caregivers needs to be expanded for precision medicine to consider health data analysts, since the analyst's role is getting tacitly expanded to take human considerations when analyzing the data.

The use of a cognitive lens (Hutchins, 2000) brought to light the scripts, protocols, and artefacts that analysts required to conduct their work. The Data Team as a cognitive unit provided a robust framework for analysis beyond the individual shortcomings of the team members to understand how, through the integration of the expertise from Antonio, Gail, John, Wilma and Katie, the Data Team worked as a constantly evolving entity. Furthermore, by conceptualizing cognitive artefacts as mediators (Arias-Hernandez et al., 2012) it was possible to better provide design guidelines that considered the nature of the analysts' work. More importantly, to consider the activity of health consumers reporting on their conditions (Kaptelinin & Nardi, 2006) as embedded in the bigger context of a human-information discourse between health consumers and analysts. In that sense, the mediating activity of artefacts could be related to the stakeholders to better support their tasks and communication. This begged for the need to facilitate the communication between health consumers and analysts, which resulted in identifying eight themes that were translated into four design guidelines that informed the design of the prototype.

Further, through the research it was possible to relate the frameworks of Nardi & Kaptelinin's Activity theory (Kaptelinin & Nardi, 2006) and Hutchin's socially-distributed cognition (Hutchins, 2000), specifically around considerations of agency in relation to human agents (health consumers and health data analysts) and computers. Together, both frameworks can provide a scaffold to move from the analysis of the Data Team as a cognitive unit to the design of health consumer-facing VA systems that support the needs of both populations.

One of the main critiques that stemmed from the analysis of cognitive artefacts was that considering artefacts and humans as a system requires to attribute agency to both of them, with the main flaw of assuming a universal cognition (Arias-Hernandez et al., 2012). Nonetheless, this anthropological approach to cognition becomes very useful to get an overview of the system, specifically to dissect the protocols, scripts, and artifacts. By following Hutchin's socially-distributed approach, it was possible to understand the Data Team from the health consumers' perspective, especially since the report creation is a black box to them and they only see the final product. Even though the framework

proved to be robust when dissecting the system, as it allowed for considering the analyst's background and the interaction between actors, it fell short when aiming to design VA systems for precision medicine. This shortcoming derived from the assumption of a universal cognition, thus every health consumer should be able to use the system, yet in reality, health consumers differ vastly in terms of health and digital literacy.

Therefore, the assumption of a common expertise, excluding those populations who have had chronic conditions for years and are experts in their own embodied experience of the condition, fails to capture the heterogeneity of health consumer's health literacy. It is at this point when shifting from analysis to synthesis, that the adapted version of Activity Theory becomes useful. By using the activity as the unit of analysis, agency can be placed solely on the human agents, which in turn allows centring the design around the health consumers and follows HCI's mantra of human-centred design. By doing so, the design process can not only consider the background and communities of the health consumers, but also their level of health literacy. More importantly, it supports to accommodate the health consumer's vulnerability into the design of VA systems. Finally, depending on the researcher's task at hand, analysis or synthesis, both frameworks can complement themselves and proved useful when correlating research findings across communities or stakeholders who are in an information communication loop.

4.6. Supporting Health Literacy

Design Guideline 1

Support automation by using a controlled vocabulary, such as a clinical data ontology, to streamline the data capture process

4.6.1. Automation

One of the main internal projects was the automation effort being conducted by the Data Team. This derived from the need to scale up the *pipeline* to be able to augment the amount of reports being interpreted. Through the process, the Data Team looked to identify the bottlenecks in the interpretation process, in order to create rules that would save time and reduce human error.

The Data Team: We also realized that we were doing individual biomarkers and not giving people the big picture. For example, at the top it would say diabetes and then list some biomarkers, but there was no relation how the biomarkers related to the health risks. We started putting that in the description initially, but that was very time consuming and you had to modify it constantly. Now when you bookmark a health risk, automatically 10 biomarkers maximum (but you can add more), that are associated with that health risks are automatically put under that health risk. So now someone can see right away if they have diabetes and the ten, or less, related biomarkers that are out of range related to diabetes. Then they get a little paragraph that is not super wordy about what is diabetes, maybe some interpretation of the biomarkers, for instance: if diabetes comes up, but their glucose is low then we write something explaining why. Right now, that text is already prepopulated with different options, so you just delete the ones that aren't relevant. So that's the beginnings of the rules that eventually we hope the system can figure it out on their own and we don't have to spend time doing that.

The Data Team: For example, Katie knows which are the top ten foods which are high in X, so she'll list them. That's based on her expertise as a dietician. With all these rules, we won't recommend a supplement like vitamin A unless the biomarker is in the red cause you can overdose. We will always write "start with a multivitamin" instead of recommending them to take vitamin A straight up. If they are in the yellow maybe we just don't recommend it but tell them to eat more food, since they don't need the supplement. I think these kind of things can be written as rules, for example if Y biomarker is red then recommend W, this will do all the write ups in the background and hopefully that will speed up the process and there is less human error, cause I'm sure there will be human error if you're not careful when doing 100 reports.

Further, the focus on reducing human error in the automation efforts was not restricted to the Data Team's internal operations, but it was also another consideration

that extended to the health consumers and related the quality of the descriptive data being captured.

4.6.2. Accuracy and the Issues of Self-assessment

Participant Performance

Prototype Score in dark blue
 Prototype False Positives in light blue
 Control Score in dark green
 Control False Positives in light green

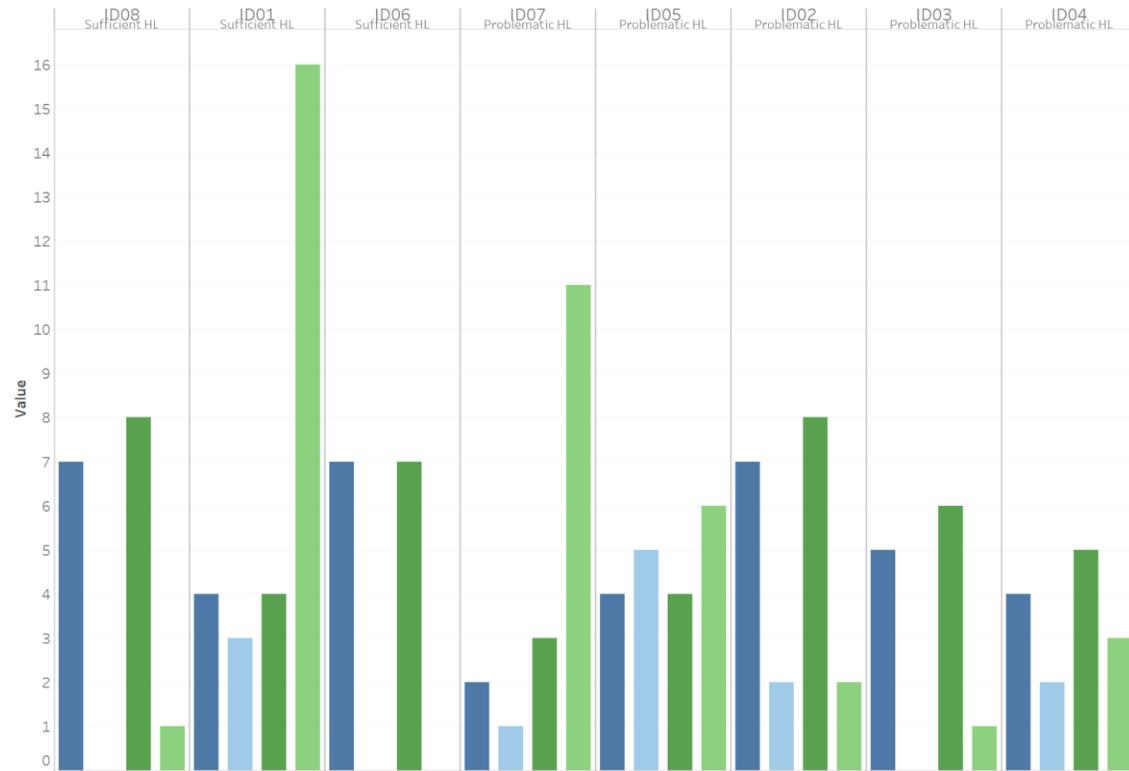


Figure 4.4 Score & False Positives in the Prototype and Control

While the results show a slightly better performance for the Control (median=5.5) than for the Prototype (median=4.5), the analysis should also focus beyond what gets captured accurately, specifically in avoiding the capture of descriptive data that will misrepresent the health consumer’s health status. In that sense, the number of false positives shows conditions that weren’t originally present in the clinical vignette that participants had to report. This can be seen in Figure 4.4, where the participants are arranged in descending order from left to right in terms of health literacy, with Jack (ID08) having the best self-reported health literacy and Vincent with the worst (ID04). The darker hues stand the Prototype (dark blue) and the Control (dark green) and the lighter hues show the false positives for the Prototype (light blue) and the Control (light green).

The Prototype's False Positives had a median of 1.5 conditions in the Prototype vs 2.5 conditions in the Control, additionally the Prototype consistently captured the same amount or less false positives than the control. When looking at Jessica (ID01) and Isabella (ID07), who have an average and below average Prototype performance respectively, this becomes especially interesting since the Control condition shows a very high number of false positives. On a deeper inspection of those captured conditions, the false positives showed a tendency of both participants leaning towards selecting different cancer conditions (i.e. blood cancer, stomach cancer, head cancer). This hints that consumers with low health literacy might attribute normal symptoms to more complex conditions such as cancers as part of their lack of understanding and confusion. When thinking about accuracy in terms of data capture, the focus should be kept in how the data will be used rather than on whether the captured data is an exact representation of someone's health, this becomes even more critical when it is the health consumers themselves who are reporting their medical history.

This notion of how people self-assess their capabilities and health was a topic of discussion in an informal conversation with the Data Team. People who did sports regularly would rank themselves as above average in a questionnaire about sports, but so would people who would walk occasionally. This was based on people's assumptions, since for the former the average person would regularly exercise, whereas for the latter the assumption was that the average person was sedentary. In that regard, both persons with very different lifestyles would self-report the same value. When looking back at the experiment's results, the issues of self-assessment were increased when participants had to rely on:

- Embodied Experiences of illness

***Vincent (ID04):** When I have a fever, it usually involves different sensations. Sometimes it's like burning or like a sore or throwing up.*

So, if I feel like throwing up, I'm not sure if it's from my stomach or what's causing it.

- Issues with the instruments/tools or artefacts

***Samantha (ID05):** The categories made sense to me, the left side was all the terms and the right side was all about the visualization. When I*

clicked the buttons, it was hard to remember which ones I had already clicked and which I didn't. When I double clicked, I wasn't sure at which stage (body category) I was at that time in the interface

Lauren (ID02): *When I was using it, at first I got confused because I didn't know what to put in it. I felt like it was not concise enough, it doesn't go too in depth. It's sort of broad, and if I don't know which area the sickness is it, then it's really hard to find it. And I can't find an option button where I can just input my sickness. Not all the sickness are listed on the prototype, so it could be hard to pinpoint the name. If I can't click the one that I have and then it's not shown and then something might not be accurate*

- False assumptions due to low health literacy

Christopher (ID03): *The categories required me to know to which disease the categories relate to. For type 2 diabetes I knew it was either the kidneys or the immune system⁹, but I didn't know where the problem was located*

Additionally, the issues of self-assessment could easily translate into missing data:

Jack (ID08): *I was trying to check the terms from the case with the ones in the interface, like tiredness and type 2 diabetes I know which systems are affected so I went to those directly. For the others since I didn't know where they belonged, I scanned each body category. I couldn't report the high blood glucose because I didn't see it.*

In order to better understand how the issues of self-assessment related to the Prototype in the three points mentioned above, the next step was to consider the participant's understanding of the clinical vignette. This can be seen in Figure 4.5, where the green dots represent the Prototype's Score, the Blue dots the Control's Score and the purple dots are the confirmation of understanding from the interviews. It is important to mention that this metric cannot be reliably used for understanding since it is known that working memory has a capacity of seven, plus or minus two (Miller, 1994) in addition to

⁹ For the prototype, Diabetes was considered a condition related to metabolism or the pancreas.

knowledge being in the lower stratum of learning according to Bloom's taxonomy (Bloom, 1956) .

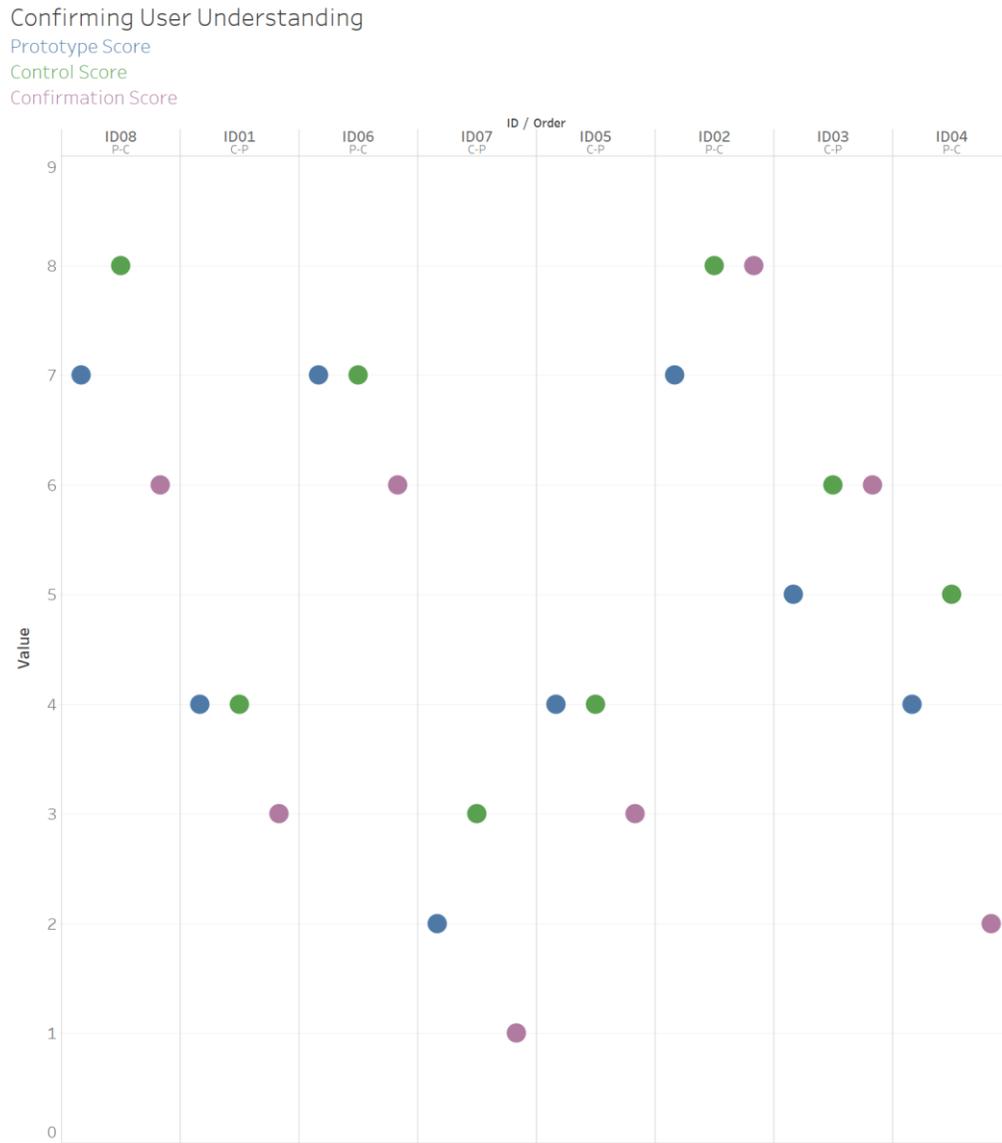


Figure 4.5 Confirming User Understanding

Nonetheless, the metric shows how there is a cluster of purple dots who remember three conditions or less: Samantha (ID05), Isabella (ID07), Vincent (ID04) and Jessica (ID01). This is consistent with Isabella (ID07) being new to Canada and Vincent (ID04) having close to no experience with medical history forms, and the three of them having a problematic health literacy. The only outlier was Jessica (ID01) who self-reported a sufficient health literacy with a score of 14. This reinforces that the issues of self-

assessment spill to the instruments that are used for measuring health literacy. The same effect can be seen on the opposite end for Lauren (ID02) who self-reported as having problematic health literacy but was the participant with best overall results. Being aware of that context, the data needs of analysts were more concerned with getting quality data that helped to create a more holistic view of the health consumer. More importantly, they accounted for such an issue in their interpretation process:

The Data Team: *The data that we use? the data from the biomarkers obviously, any of the biomarkers that are identified or the health risks that are identified. And then, in the phenome form, it's a little bit challenging because I've been torn between: do I make my recommendations based off the subjective information that they give me in the EHR or do I do it on the objective biomarkers that come back? Obviously, it has to be a little bit of both, but I definitely put more weight on the objective biomarkers, because I don't know how they interpreted that question, or if they are giving me all the information.*

Finally, when aiming to support automation and scaling up, and due to the ever-evolving nature of a data team, VA designers cannot rely on the individual analyst's expertise to accommodate the shortcomings of health consumer self-assessment. It is in that space where the HPO's "standardized, controlled vocabulary [that] allows phenotypic information to be described in an unambiguous fashion in medical publications and database" (Robinson & Mundlos, 2010: 525) could serve as a robust infrastructure, since the categories from the HPO can internally be turned in more relatable ones. In addition, the lay synonyms can be more understandable than medical terminology and allow to find relations with the UMLS and other data ontologies (Köhler et al., 2017).

This can lead to health consumers reporting fewer wrong conditions that in turn result in a better quality of data. Furthermore, due to the HPO containing the relations and hierarchy between conditions, this sets the road for automation through the creation of rules, and/or by facilitating the integration with literature data mining techniques. From a Visual Analytics perspective, this design guideline accommodates not only the situated nature of the health consumer's cognition (Arias-Hernandez et al., 2012), but it can support the design of ontology visualizations that help to explore and find new disease correlations (Demelo et al., 2017).

4.7. Prioritizing Interactivity and Visual Feedback

Design Guideline 2

Accommodate the health consumer's health literacy through visual mapping of the controlled vocabulary to facilitate interaction and visual feedback.

4.7.1. Facilitating Health consumer Communication

An essential barrier to successful communication between health consumers and caregivers is the health consumer's health literacy, in addition to their metacognition of their embodied disease experiences. When continuing with the analysis of the issues of self-assessment in relation to health literacy, if excluding Jessica (ID01) and Lauren (ID02), there seems to be a pattern of people with sufficient health literacy having a better Prototype score than the cluster of participants with problematic health literacy (See Figure 4.6) Further, when continuing with the same omission of Jessica (ID01) and Lauren (ID02), and looking at the prototype score versus the false positives (See Figure 4.7), participants who had more false positives tended to have a worse prototype score. This is in line with the expectations of the HLS-EU-Q16 survey, where better health literacy leads to consumers being able to talk about, and better understand their own health (Emiral G.O., Aygar H., Isiktekin B., Göktas S., Dagtekin G., 2018). This finding further hints that false positives might be related to a health consumer's own embodied experiences and their ability to properly self-asses. In that regard, the Prototype had a positive reception from the participants and the visual mapping of the HPO to the body graphic seemed to support the needs of different populations with varying health literacy:

- For people new to the country or the language:

Joshua (ID06): The only time I see the visualization would be good is if I didn't know the language. It would be useful for someone who speaks a totally different language then they can click on "eyes" and see where they light up.

Prototype Score vs SUS Score

Red and Orange: Problematic Health Literacy

Green: Sufficient Health Literacy

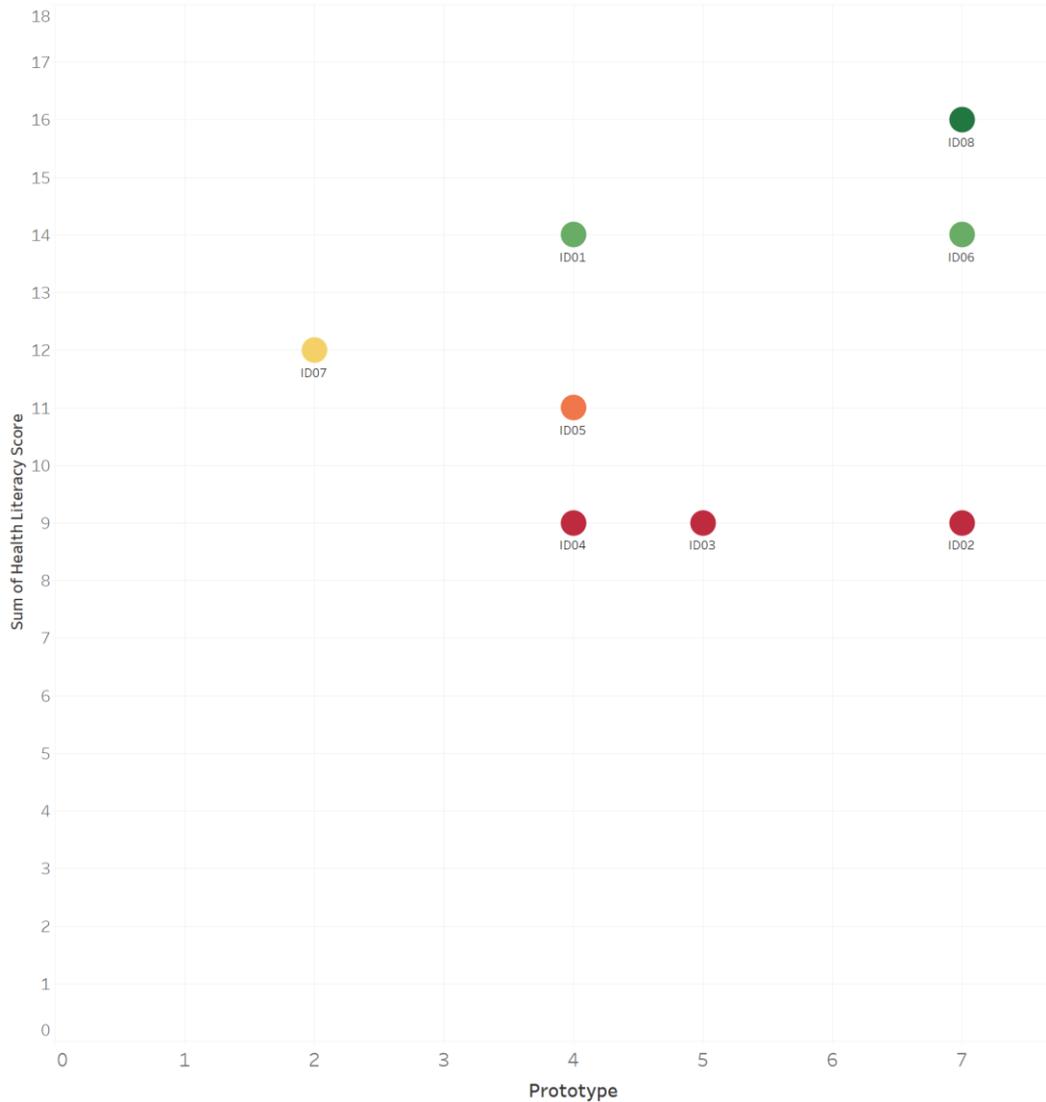


Figure 4.6 Health Literacy vs Prototype Score

Isabella (ID07): *The visualization helps the health consumer to fill the form as a health record and it makes it easier for the doctor to know more about the health consumer. After clicking, the highlighted body parts reminded me where the diseases were located. If I only had the diseases it would be hard to report since I can't understand them, and I don't know to which body part they relate to. The body graphic was really helpful to better communicate and self-record my medical history.*

Prototype Score vs False Positives

Red and Orange: Problematic Health Literacy

Green: Sufficient Health Literacy

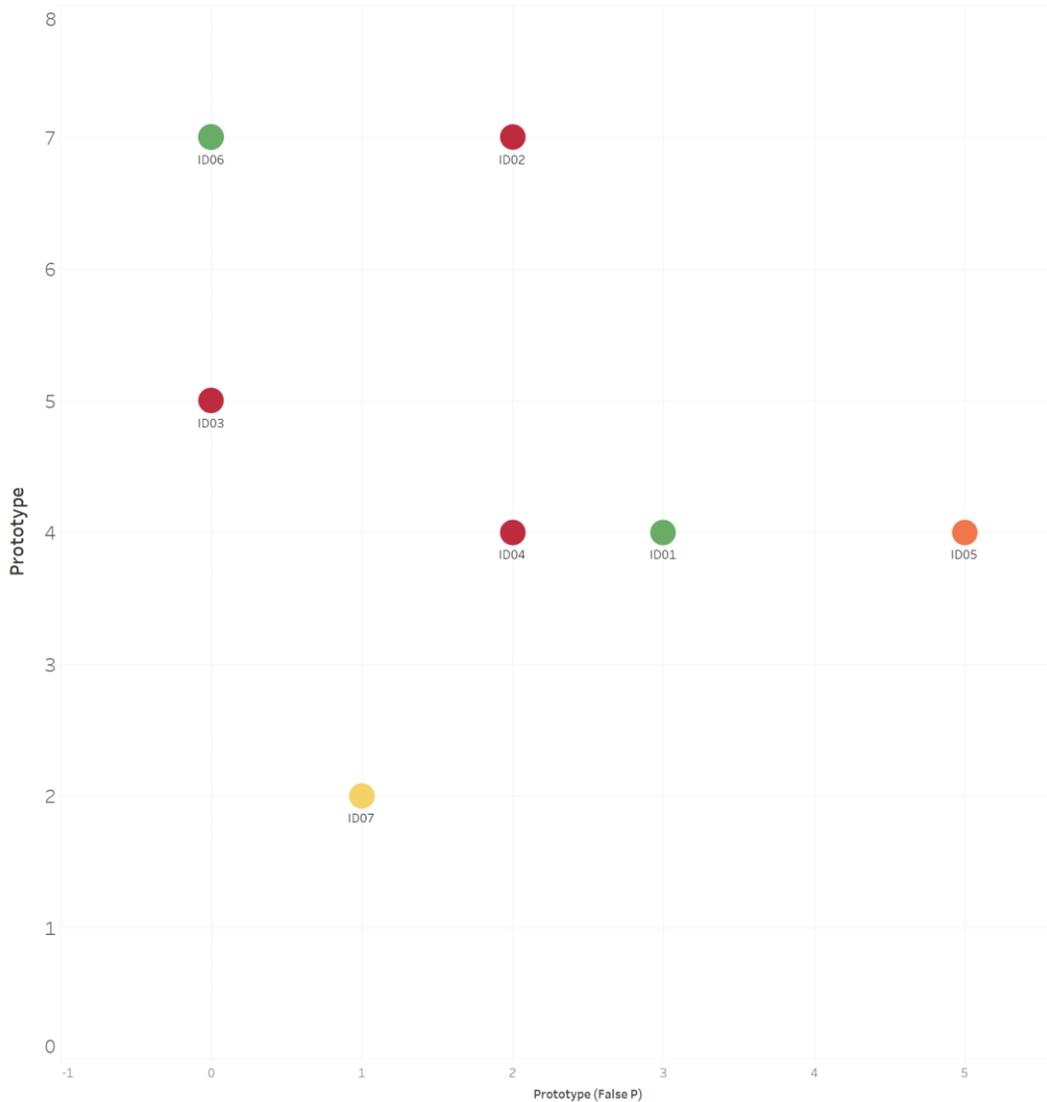


Figure 4.7 Prototype Score vs False Positives Scores

- For people with no knowledge of medical terminology and/or no experience dealing with medical history forms

Vincent (ID04): *If I'm looking to show which part of my body is hurting, the interface would help a lot. The interface can help a health consumer and a doctor to more easily understand which part of the body is having an issue.*

Samantha (ID05): *I can choose wherever I want and that can save a lot of time when communicating with my doctor, because if I just go and tell the Doctor where I have a discomfort, it might not be very clear. But by using the interface I can directly show the graphic and anywhere where I'm not comfortable*

- Additionally, the Prototype's added value is not restricted to health consumers with problematic health literacy

Jack (ID08): *The visualization was useful for things I didn't know like about the immune system. I think also it's useful to express where the discomfort comes from. If you're more knowledgeable about medical information, I think it will be easier for you. The highlighted body parts helped me to see where the health consumer was being affected.*

These perceived values and functionalities of the system by health consumers echo the data Team's consumer-driven considerations, specifically in terms of an existent concern regarding how the test results were reported back to the health consumer. More importantly, in terms of the results being an additional layer of information health consumers could bring to their caregivers:

The Data Team: *The health consumer recommendations process has been very iterative. At the beginning everything they got was biomarkers or health risks and the rest had to be written by hand. It was just a wall of text, but now it's in sections. We realized that we were writing the same thing every time for this marker, so let's start automating this. So now we have this DB of notes. Then we realized that we were giving people too much information, no one needs to read about 33 biomarkers in the highlight, so we made rules about how many biomarkers should be highlighted, what kind of biomarkers should be highlighted. So, we try and emphasize actionable biomarkers, if you took this to a doctor, would he care about that?*

4.7.2. Usability

The system usability score was another relevant metric for the analysis. While a system with a SUS score of 70 can be deemed good (Bangor et al., 2008), the results showed that the prototype was on that threshold (median=70; mean=68.75, SD=6.545). (See Figure 4.8).

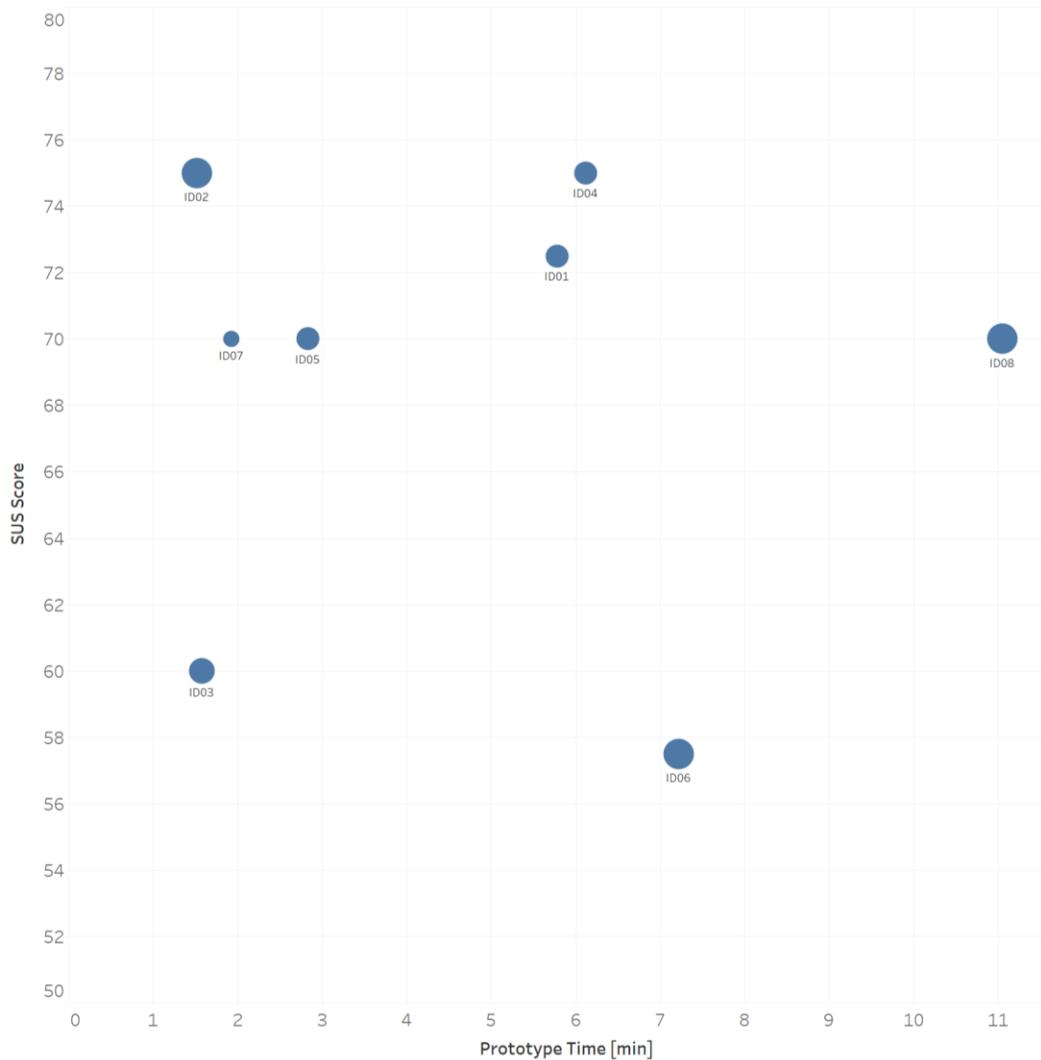


Figure 4.8 SUS Score vs Time (Size Adjusted by Prototype Score)

In order to further analyze the results, the Prototype usage time was a relevant variable, under the assumption that if the user stays in a system longer, then they will be interacting more with it. In Figure 4.8 it can be seen that the smaller the dots, the worse score when using the Prototype. This shows that participants with a better score did not

necessarily spend more time using the interface. In that same line, neither did spending more time in the interface lead to a better usability score. This lack of emergent visual patterns led to considering the impact on health literacy on the participant's use and interaction with the platform (See Figure 4.9), where the labels show participant ID and Prototype Score.



Figure 4.9 SUS Score vs Time (Size by Prototype Score, Colour by HL)

Through this additional layer, a pattern emerges that participants with problematic health literacy tended to spend less time in the interface, whereas participants with sufficient health literacy spent more time using the Prototype. Nonetheless, there did not seem to be a correlation between health literacy and usability score. When looking at the

interviews for participants who scored the system best, there was a trend in focusing on the positive aspects and benefits of the system:

Lauren (ID02): *The diagram gave me clarity, because of the different colours. At first, I thought it was just for decoration but now I'm seeing that the colours match the categories. After putting all the answers, when I look at it now, it is much clear to see which parts of my body are affected.*

Vincent (ID04): *When I used I just clicked on the buttons and I saw the visualization showing a match and I could see each part*

Participants in the middle range scored the system with a 70 to 72 usability score and focused on a more critical perspective of the interface and how it can be used:

Jessica (ID01): *The visualization helps me to understand in which part of the body the disease is located. Some diseases I don't understand. So at a first glance I don't know what they are. Sometimes the icon is so small I can't understand what it is, like the eye. At first, I need to click on the icon and then see which part of the body highlights, because if not it's hard to see what the icon is*

Samantha (ID05): *I interacted through clicks, when I clicked on the button categories, I saw the body light up with the illustrations and it was easier to use than the control. It didn't just only provide me with the words, but I could also directly see which part of the body I was considering.*

Jack (ID08): *The purpose of the system was to report everything that was wrong with the health consumer, it's a full report of the medical information of a health consumer*

Finally, the lower score tier came from participants with a more informed position towards design and interface design¹⁰, focusing on aspects regarding the user interface and user experience:

Joshua (ID06): *I didn't even notice the highlight during the interaction.*

I think that hierarchically, the conditions and the body graphic are given equal importance, so the eye doesn't need to go and the visualization wasn't essential to my task. After having clicked a category, there is no additional feedback once you choose a symptom, but now the user is used to having feedback after every click.

Christopher (ID03): *The interface allowed me to be faster, I was able to select on the conditions faster since I didn't need to read all the things. It really helped me. Also, in the control, there was no pictures and with the interface I was able to see all the questions at once. For example, with the control I didn't know which ones to choose, but when I see them on the same page at the same time, I could just click on any of the illnesses. I prefer using this interface than the control.*

Usability Recommendations

From the User Testing, key aspects were identified in terms of design aspects that need to be addressed for the next iteration of the prototype:

- Allow searching by text
- Allow interaction through the body graphic
- Provide a toggle to change the shown conditions between current symptoms and medical history
- Make clearer distinctions between categories and subcategories

Finally, in terms of usability and accommodating the health consumer's health literacy, the Prototype showed that the visual mapping of the HPO to the body graphic

¹⁰ While the participant's background wasn't part of the questions in the interview guide, Joshua shared his background as a Bachelor's in Psychology and his current work in User-Interface Design and Christopher talked about his own work as a Postdoc working on user interfaces.

was positively received by health consumers. They saw it as a useful tool for communicating with their caregivers. Furthermore, the Prototype could serve populations with different levels of health literacy. Future iterations should address the usability recommendations from participants.

4.8. Quantitative Hypothesis

The interactive visualization captures a health consumer's conditions more accurately than a traditional medical history form.

The results were not conclusive to support or reject hypothesis 2.2, but showed interesting new lines of inquiry. Relating the Data Team's with the participants' interviews allowed reconceptualizing accuracy in the context of precision medicine's Big Data challenges, specifically when considering the issues of self-assessment when health consumers have problematic to sufficient health literacy (Emiral G.O., Aygar H., Isiktekin B., Göktas S., Dagtekin G., 2018; Nolasco et al., 2010). This requires that interfaces not only capture the required data in a systematic way, but that the system itself reduces the possibility of capturing false positives.

The results showed that the Control captured more conditions than the Prototype, yet the Prototype allowed capturing descriptive health consumer data with less false positives than traditional medical history forms. This hints that the Prototype was able to capture more reliable data regardless of the participant's health literacy. While more evaluation is required, the results show that the visual mapping of ontologies for health consumers has the capability of becoming a key VA system for precision medicine.

In terms of reducing the impact of the issues with self-assessment, a controlled vocabulary, in this case the Human Phenotype Ontology (Köhler et al., 2017), provided the user with lay synonyms and categories that were more relatable than medical terminology. While there are issues that extend beyond the scope of the Prototype's capabilities (e.g. the varying state of clinical maturity of the HPO across different medical domains), it showed that the HPO was helpful for health consumers when reporting their data. The visual mapping of the HPO supported interaction and the results show that most

of the interface usability limitations can be easily integrated or fixed (i.e. toggle for medical history vs symptoms, allow text entry, etc.).

Further, this motivates future iterations to revolve around usability and accommodating the health consumer's health literacy by building upon shared mental models. Just as health consumers know that blood has information about one's health, the graphical representation of the body could be related to the western paradigm of diseases located to organs or systems. Additionally, the results are also in line with the notion that when there is incomplete or inconsistent display of information, it becomes an additional factor that compounds the inherent variability of the user's knowledge (Elson et al., 1997). This initial study supports the notion that healthcare design should be evidence-based and involve several disciplines (Armijo et al., 2009), and that HCI standardized surveys can be used for assessing usability (U.S. Department of Health & Human Services, 2020).

4.9. Research Limitations

4.9.1. Threats to Validity

One relevant aspect to restate is the difference between patients and health consumers, specifically that participants of the study were not patients per se since they were not battling a disease. The participants were treated as health consumers, since in the shift of reactive to preventive medicine, healthcare consumers will not need to be sick to seek care, but rather try to maintain their health. It is in that regard that the participants were seen as precision medicine *health consumers*. Additionally, the Data Team worked in a for-profit organization and their main population of study was people who were interested in their health and had the economic means to pay for the test¹¹. Therefore, the main population who bought the product was either data savvy or had at least an understanding of precision medicine, hence having enough literacy and criteria to deem the product worth buying. Both aspects make the research results not a direct representation of reality, instead, they should be taken as suggestive of a future scenario where precision medicine is the norm and health data analysts are fully integrated in the

¹¹ At the time of the study (2018-2019), the most comprehensive package's retail price oscillated around \$1,000 CAD.

healthcare system. Nonetheless, concerns around data privacy and data stewardship will continue to accompany these kind of products ,and it is relevant to start identifying issues that will need to be addressed when precision medicine and clinical data ontologies are the norm and ripe for full clinical deployment.

The Caregivers' Perspective

One of the critical voices that wasn't thoroughly explored was the caregivers. While health consumers depend on EHR and PHRs to communicate their medical history, and data analysts use that information to provide analyses and reports, it is physicians who will ultimately effect on that information to provide a diagnosis. In that regard, caregivers in single-payer systems, specifically in British Columbia, Canada find themselves often overburdened, having less time to spend with individuals¹². Further, the introduction of new standards proves to be extremely challenging (i.e. new compassionate communication protocols, new electronic health systems, or new ontologies), as it requires time investment from such a time-sensitive community.

Along that same line, when thinking about the clinical deployment of the system and/or scaling up the research design for a larger user study, it would make sense to reconsider the chosen controlled vocabulary. While the HPO had a good integration with the UMLS and data-mining approaches such as GWAS, and the presence of lay synonyms, its target audience was mainly researchers. This can be seen in the fact the HPO is just achieving clinical maturity for certain assessments (Köhler et al., 2017). Further, it is well known that clinicians have overburdened schedules and introducing new controlled vocabularies or frameworks is very hard for them to assimilate into their practice. In that regard, the ICD-10 (or ICD-11 upon its release) or SNOMED are already embedded into their praxis and would better suit the audience. This does not mean that the HPO and the present research lacks validity, but it means that it still has a considerable way to go before clinical deployment.

Finally, while caregivers were involved in the research design and validation process, their contributions were on an advisory plane and their proper role as stakeholders wasn't explored in-depth.

¹² This was brought up in a casual conversation with a physician working through telemedicine in BC when talking about patient-physician communication challenges.

4.9.2. Methodological Limitations

Thematic Coding

Through the research, *Satisficing* (Simon, 1996b) was one of the main ideas informing the design approach, this means that a high intercoder reliability was not the goal, but to have a set of codes that was *good enough* to move forwards. In terms of the thematic coding, this translated into some text excerpts overlapping across codes and hence the low rates for both thematic codings. More careful consideration in choosing the interview extracts could have had a direct impact in the agreement rates.

Further, the agreement was calculated based on a single pass by the additional coders, rather than first sending the coders a stratified sample of the text database, adjusting the code and then sending the rest of the database (Geisler & Swarts, 2019) for calculating the agreement. This was done in the interest of time and the relatively small size of both qualitative interview databases.

Experiment

The results of the user testing were meant to provide descriptive statistics to better understand the impact of the design guidelines and how participants used the Prototype. One of the main limitations of the experiment is the low number of participants, along with no counterbalancing in terms of health literacy between groups. This reduced scope was chosen since part of the contributions of this work rely on the expanded design thinking methodology for design and evaluation, rather than on a thorough quantitative evaluation of the prototype. The experiment design drew from Usability Testing, but the main goal was not on the prototype's usability, but on finding trends and evidence that supported the overall methodological approach and might inform future iterations of the Prototype.

4.9.3. Design Limitations

Mapping the HPO

The HPO presents a robust standardized medical vocabulary with the addition of lay synonyms, yet it continues to evolve. This presents certain limitations that need to be considered, such as:

- The lay synonyms are not always helpful since some might only present a rewording of the condition. For example, conditions that arise from low levels “hypo” or from high levels “hyper” of a metabolite (i.e. “hyperaldosteronism” becoming “elevated plasma aldosterone” or “hypoaldosteronism” becoming “low blood aldosterone level”).
- Some synonyms contain abbreviations or acronyms. For example, “Hyperalphalipoproteinemia” becomes “High blood HDL level”, which might raise further confusion in health consumers since the acronyms might require additional knowledge.

Furthermore, mapping the HPO to a body graphic requires having a western view where diseases are linked to certain organs or systems, and cultural adaptations might be required when designing for other cultures. This was seen during the refinement of the prototype’s buttons in the quality assurance study, where some participants with Asian backgrounds saw cancer conditions as an outcome of mental health imbalances, rather than from issues in different organs or body systems. This relation derived from the understanding that if the mind was not well, the body would also become ill, which contrasts with the clear separation of mind and body in western medicine.

Further, in the case of psychiatric or other conditions that cannot be directly mapped to a specific organ, mapping them to a visual representation becomes an abstract challenge for the designer. This will require new representations to be created. For example, in the current prototype, such a challenge was faced with Skin conditions, Growth abnormalities and Rare Disorders.

- *Skin conditions*: Highlighting the border of the body graphic was an easy challenge from a coding perspective, yet it was a hard one from the perspective of the human eye. This challenge was caused by the thin border of the graphic that would be highlighted, more so in displays with restricted estate such as cellphones, since it was barely noticeable by the naked eye. As a result, I mapped the skin conditions to a representation of the hand, yet such a representation might be confusing.
- *Growth abnormalities & Rare Disorders*: The HPO supports the capture of limbs and growth abnormalities, yet displaying such conditions requires a

heavy input from the design perspective and it raises the question: with a vast number of growth abnormalities, is there a generalized way of showing them all? How can those representations be linked to the health consumer's notion of their own conditions? Therefore, I decided to map growth abnormalities along with Rare Disorders as a placeholder for those conditions.

Defining Diseases and Categories

Derived from the technical limitations of generating data, some diseases needed to be created internally to accommodate new research findings as the population grew in the database. While a high blood glucose level is a common indicator of diabetes, precision medicine's technology allows detecting biomarkers that might relate to early onset of the condition (Anwar et al., 2020). In that regard, precision medicine faces the challenge of possibly identifying new conditions, yet the slow process of clinical validation might take decades to catch up (Fleck, 1979). This puts the VA system designer in a mediating position between the fast-paced findings of the analysts with the slow process of knowledge solidification of caregivers. Either way, this is a bigger discussion that is out of the scope of the present research.

4.10. Overarching Hypothesis

Through structured field immersion for understanding different stakeholders, a VA design approach leads to a visual mapping of a clinical data ontology that facilitates a health consumers' reporting of their descriptive data, because data architectures and human factors can be converted into design guidelines

The overarching hypothesis 1 that encompassed the mixed methods approach was supported. Through the fieldwork it was possible to identify two methodological and two design guidelines for VA systems that support analysis in the context of correlating health consumer descriptive data with biomolecular data. Through the user testing, it was possible to evaluate its functionality and better understand the shortcomings of its usability for future iterations. Further, by the aggregation of the qualitative and quantitative data

sets, it was possible to create an asynchronous conversation between stakeholders to better understand and frame the results from both types of inquiry. Through the evaluation, the prototype showed promising results in terms of capturing more accurate and reliable descriptive health consumer data, but in contrast, it captured less conditions than the control. From the user perspective, the health literacy of the participants didn't show any effect on their ability to report the conditions, Nonetheless, the usability of the VA system did show a trend towards participants with a higher health literacy spending more time in the system, and therefore giving the system a better score. While the system's usability was on the lower boundary of being considered a good interface, most of the issues that the participants mentioned are easy to fix and integrate into a new iteration.

The design guidelines that derived from the analyst's interviews were validated by using simple and corrected intercoder reliability measures, which could then be linked to the health consumer's interviews (See Table 4.9). By doing so it was possible to triangulate the information (Creswell, 2007) and address the questions framed by Kazubowsky-Houston & Magnat:

What is lost and gained when ethnography “travels” across disciplines? How can ethnography’s transdisciplinary travels contribute to how we might conceptualize, reimagine, and practice ethnography today and in the years to come? What does it mean for ethnography to “travel” within a competitive and profit-driven neoliberal academia, where the pursuit of knowledge is no longer seen as a public good and an end in and of itself? (Kazubowski-Houston & Magnat, 2018:379)

What can be lost through the travel of ethnography to precision medicine is the critical inquiry of how data gets produced and analyzed, the notion that genomics has a negative connotation in vulnerable populations such as indigenous communities (Reardon & TallBear, 2012), and that discussions about property are becoming increasingly relevant in the western Anthropocene. This is even more so if through the use of precision medicine products some of the property rights to one's biomolecular data are not only freely surrendered, but there is an actual price for doing so. Further, precision medicine solutions do not necessarily address or fix the issues stemming from the depersonalization of healthcare, and they are also embedded in a conflicting space where data privacy is an ever-increasing concern. Furthermore, such products also increase the commercialization

of medicine by removing clinicians from self-diagnostic loops, in a similar way to what previous products have done in the past (i.e. pregnancy tests). Finally, in spaces where there are short lifecycles for products such as HCI, *quicker* versions of ethnography are required (Millen, 2000), yet that doesn't mean that the outcome itself cannot become an ethnographic place (Pink & Morgan, 2013). Further, it can maintain the epistemological integrity of ethnography as a methodology and an outcome.

Table 4.9 An Asynchronous Conversation between Health consumers and Analysts

Analysts' Code	Guidelines	Health Consumer's Code
Job Role	<u>Methodological Guideline 1:</u> Identify and design from the perspective of the most vulnerable stakeholder and prioritize those needs when incorporating other stakeholder's requirements.	Health Consumer Experiences
Teamwork		Health Literacy
Accommodating Health Literacy		Interaction
Data Challenges	<u>Methodological Guideline 2:</u> Trace the production of data to understand the shortcomings of its creation and always maintain a critical attitude to what the data should represent.	Becoming Data Essay (See Preface and Epilogue)
Minimum Data		
Accountability	<u>Design Guideline 1:</u> Support automation by using a controlled vocabulary, such as a clinical data ontology, to streamline the data capture process.	Interaction
Automation		Interface Limitations
Product Evolution	<u>Design Guideline 2:</u> Accommodate the health consumer's health literacy through visual mapping of the controlled vocabulary to facilitate interaction and visual feedback.	Value of Vis
		Interface Recommendations

When ethnography needs to embed itself in a profit-driven, neoliberal academia, then it can become a key asset in the methodological repertoire of researchers who are looking to critically analyze and move precision medicine forward. This is in line with a translational science approach such as HIBAR (The HIBAR Research Alliance, 2020) that requires a highly interdisciplinary integration of expertise across disciplines in order to translate basic into applied research. By doing so, scientific inquiries can be aimed to be in the interest of society. By tracing the data and putting stakeholders in the same space,

the pursuit of knowledge can support the public good, especially in spaces where wicked problems (Rittel & Webber, 1973) are the norm and achieving a shift from reactive to preventive medicine is to the benefit of the society. When considering the foundational literature for Visual Analytics, highly interdisciplinary approaches were one of the key aspects of the field (Thomas & Cook, 2005) and it was this mediation of the interdisciplinarity that made VA a discipline in itself (Kohlhammer et al., 2011). The incorporation of social science methods into the repertoire of VA is therefore in line with the foundational philosophy of VA and is in itself a contribution of this work.

Finally, by framing the research in a robust Design Thinking methodology, such as the Nielsen-Norman Design Thinking framework (Nielsen Norman Group, 2020), it is possible to design in a space where researchers and scientists can be considered designers (Norman, 2013) who use their professional expertise to incorporate new knowledge in their design practice (Schön, 2000) and harness social and natural phenomena to create new “method-technologies” (Arthur, 2009) that allow bridging omics with clinical data for decision-making (Topol, 2019). In that same regard, the research results show that short-term ethnography can become an asset in the *formative* stages of visualization design (Munzner, 2009) that lead to User Testing and/or Experiments in the *summative* stages. Furthermore, the integration of a distributed cognition framework (Hutchins, 2000) with activity theory (Kaptelinin & Nardi, 2006) proved to be useful for accommodating the analysis of a cognitive unit for the synthesis of a design that caters to not only the end population, but to all the involved stakeholders.

4.11. Chapter Summary

Chapter 4 presented and discussed the results from the inquiry. This made it possible to bridge the needs from health consumers and analysts by finding codes in the interviews that led to methodological guidelines for short-term ethnography: i) Understanding and designing for stakeholders and ii) Data provenance, and design guidelines that account for: i) the consumer’s vulnerability and ii) the consumer’s health literacy when creating VA systems for precision medicine. The user study showed that the issues of self-assessment affect the way consumers understand and report their own conditions and health, but that they also extend to the artefacts that determine their level of health literacy. This shows some of the intrinsic challenges of precision medicine, as

access to digital healthcare is mostly mediated by digital interfaces, removing the caregiver from the equation of reporting on descriptive data.

From a design perspective, it is important that the system fosters interactivity and that it visually maps abstract relations and hierarchies, such as controlled vocabularies, to support the usability of the system. In turn, the use of data ontologies supports the automation needs and the development of new algorithms for precision medicine. Additionally, it is important to choose an appropriate controlled vocabulary and be critical about its shortcomings when thinking about the limitations of the research. Finally, the research can be seen as a pilot study in conducting mixed methods for the design of interactive systems for precision medicine. In that light, Chapter 5 will continue by exploring possible avenues for future work and summarizing the contribution of this thesis.

Chapter 5. Conclusion & Future Work

5.1. Summary

The research shown in the present thesis centred around having a better understanding of how to design VA systems that facilitate health consumers to report their descriptive data. These systems are rooted on a data infrastructure that supports analysis, automation, and the development of algorithms (Hoehndorf et al., 2015; Köhler et al., 2017). In order to do so, the research introduces a distributed cognition lens (Hollan et al., 2000; Hutchins, 2000) for short-term ethnography as a valuable addition to the methodological repertoire of design thinking (Nielsen Norman Group, 2020). By doing so, it was possible to draft two new guidelines for short-term ethnography in Visual Analytics and HCI: i) *Understanding and designing for stakeholders*, and ii) *Data Provenance*. These guidelines expand on the qualities of short-term ethnography: *intensity of the research encounter, a focus on details, the ethnographic-theoretical dialogue, and the audio-visual and other traces of ethnographic encounters* (Pink & Morgan, 2013). Additionally, these recommendations were accompanied by two design guidelines: i) *Supporting Health Literacy*, and ii) *Prioritizing Interactivity and Visual Feedback*, that were later translated into an interactive visualization that provided an insight into capturing quality health consumer data.

Furthermore, this thesis showed the need to incorporate mixed methods when designing for vulnerable populations such as health consumers, and how by triangulating qualitative and quantitative inquiries (Creswell, 2013), a VA designer in digital healthcare can create interfaces that accommodate an anthropological view and critique of how data is dealt with in the precision medicine industry (Kazubowski-Houston & Magnat, 2018). This last point becomes increasingly relevant as the world starts to shift to a preventive approach in medicine and new careers in digital healthcare will require designers to be able to broker knowledge between stakeholders.

5.1.1. Chapter breakdown

The thesis was organized in 5 chapters and bookended with a short essay. The short essay framed the research from the author's perspective and was meant to show

the different vantage points in which the researcher/designer/author of this work stood during the research.

Chapter 1 presented the introduction and motivation of the research; the goal was to frame design in precision medicine as a wicked problem (Buchanan, 1992), where health consumers, data analysts and caregivers need to exchange reliable and actionable information. It showed the two main audiences involved in conducting the research, as well as providing the background of Visual Analytics as a highly interdisciplinary science that focuses on the design and evaluation of interactive systems that support analytical thinking (Fisher et al., 2011; Keim & Zhang, 2011; Kerren et al., 2008; Thomas & Cook, 2005). Chapter 1 concluded by presenting the research setting, objectives, goals and expected contributions of the work.

Chapter 2 was divided in four parts and contextualized the research in relation to previous work in the area. The first part provided an initial glimpse into design thinking and how the framework is understood from different perspectives: academia (Hasso Plattner Institute of Design at Stanford University, 2020; The Design Council, 2020), business (IDEO, 2020; Pentagram, 2018), and research (Nielsen Norman Group, 2020). This exploration focused on four relevant HCI Design Thinking pioneers: i) Herbert A. Simon (Simon, 1996a), ii) Donald A. Schön (Schön, 2000), iii) W. Brian Arthur (Arthur, 2009), and iv) Donald A. Norman (Norman, 1993). Further, it supported the choice of the Nielsen-Norman DT framework, which had its inception in a science-informed approach to DT that supports the consideration of cognition in the design process.

The second part provided an overview of ethnography, both as methodology (Creswell, 2018) and as outcome. This involved engaging with the critiques of anthropology from other disciplines who use ethnography to *humanize* studies or that appropriate the term but emancipate the methodology from the outcome (Kazubowski-Houston & Magnat, 2018). In that critique, the researcher explored how ethnographic methods have become part of the methodological toolset in healthcare (N. Scrimshaw & Gleason, 1992) and HCI (Blomberg et al., 1993; Millen, 2000; Suchman, 1989). The goal in this part was to propose a specific type of adapted ethnography, short-term ethnography (Pink & Morgan, 2013), as a methodology and an outcome, that considers the theoretical debates in anthropology. This was followed by an exploration of cognitive ethnography as a lens to understand the use of tools and how a group of people can distribute cognition

among each other to create a single cognitive unit. The third part dealt with the theoretical basis of understanding artefacts as cognitive mediators (Arias-Hernandez et al., 2012) and highlighted the need to understand the situated nature of cognition (Kaptelinin & Nardi, 2006). Additionally, it dealt with how, historically in HCI, digital interfaces have been considered means to *expanding human intellect* (Engelbart, 1962).

The fourth part focused on reviewing the literature related to the design and use of electronic health records (Vimla L. Patel & Kaufman, 1998), with a focus on creating a link to design frameworks in Visual Analytics (Munzner, 2009; Sedlmair, Meyer, & Munzner, 2012). By doing so, it was possible to identify key aspects in terms of usability (Marchionini et al., 2007; U.S. Department of Health & Human Services, 2020), and clinical data ontologies (Hoehndorf et al., 2015; Prodanov, 2011). This showed that most of the research is focused on understanding medical history forms from the context and burden to the caregiver rather than to the health consumer (Armijo, McDonnell, & Werner, 2005; Mosaly et al., 2016). This section also allowed for identification of data ontologies as a robust architecture for supporting automation, and ontology visualization as a catalyst of the discovery of new patterns and correlation between diseases (Robinson & Mundlos, 2010).

Chapter 3 had six sections and was framed according to the Nielsen-Norman DT framework (Nielsen Norman Group, 2020). The first section in this chapter was an overview of the HIBAR philosophy (The HIBAR Research Alliance, 2020) and how *satisficing* (Simon, 1996b) was an integral idea when making design decisions. The overview also contained the methodology and methods, research questions and hypothesis for this thesis. The second section showed how short-term ethnography was conducted at the organization to better understand stakeholders, and presented the design for the qualitative inquiry. This section touched on how participants were sampled and the methodology. The third section defined the unit of analysis as the activity of health consumers reporting on their medical history and symptoms, and framed the problems of designing VA systems in precision medicine as wicked problems.

The fourth and fifth sections showed the creative process of mapping the HPO to more relatable categories for a lay audience and presented how the visual mapping of the HPO to a body graphic was done. This involved challenges regarding the definition of categories, the visual allocation of abstract concepts to two-dimension designs, the colour

palette, designing for attention, and interaction. These sections also showed the technical steps of operationalizing the HPO as a data infrastructure and how that translated into the prototype used for the experiment. The final section provided the research design for the experiment and briefly described the survey instruments, as well as the how the descriptive statistics were calculated.

Chapter 4 dealt with the results from the mixed methods inquiry and was followed by the proposed guidelines and the discussion of the hypotheses. The chapter was written to present an asynchronous conversation between participants (health consumers) and the Data Team (health data analysts) as a way to create an *ethnographic place* (Howell, 2017; Pink & Morgan, 2013) and maintain the epistemological integrity of ethnography. The first part presented the results from the inquiry, starting with the codes derived from short-term ethnography and how they were translated into methodological and design guidelines that informed the prototype design. This part finished by presenting the results from the experiments through descriptive statistics and a set of data visualizations.

While the results were not enough to support the hypothesis of the prototype capturing more accurate descriptive health consumer data, they showed interesting and promising routes for future work in terms of mixed methods approaches for designing VA systems in precision medicine. The second part dealt with the first two guidelines, presented the relevant stakeholders for the analysis, specifically focusing on health consumers and caregivers. This allowed considering the Data Team as a cognitive unit which uses the descriptive health consumer data to correlate it to biomolecular data. Also, it served to reflect on the main challenges of health consumers when communicating with their health providers. This exercise required breaking down the health consumer activity into its components by using an adapted Activity Theory model from Engeström (Engeström, 1990) through the lens of Kuutti & Arvonen (Kuutti, 1992).

The third part focused on the results from the user study and analyzed Design Guidelines 1 and 2. This part discussed the issues of self-assessment and how the artefacts for reporting on medical history and symptoms compounded to the issues of health literacy to reduce the quality of captured data. Further, by focusing on the participant's health literacy and their usability rating of the prototype, it was possible to identify recommendations for the next prototype iteration, and to reflect on how to better cater health consumers with varying levels of health literacy. Furthermore, the discussion

elucidated data issues regarding genetic sequencing technology and data mining limitations. As part of the critical nature of ethnography, analysis in this section provided a critical perspective on precision medicine and its required data. The last part focused on the research limitations, dealing with the issues of validity, methodology, and design limitations.

5.2. Contributions

There are two contributions, the first one in terms of a data-capture system that visually maps a data ontology to a somatic representation. This contribution shows how to scaffold a controlled vocabulary to a lay-user interface that support interactivity and accommodates a lacking health literacy. The prototype demonstrated diverse functionalities, as identified by the user-testing participants, ranging from being a visual aid to reduce false positives to becoming a supportive artefact in the communication between health consumers and caregivers. In terms of precision medicine, the contribution continues the line of patient-centred research and bringing the voices of health consumers into the design cycle. Additionally, the system showed several promising areas for data-capture systems for precision medicine, such as the need to account for the issues of self-assessment, to have a more reliable transition from reactive to preventive medicine.

The second contribution lies in the two methodological guidelines for short-term ethnography, and two design guidelines for data-capture systems. These recommendations aim to help future researchers conduct naturalistic research for Visual Analytics in a precision medicine setting, and to bring back critical aspects of ethnography as a methodology and a research outcome. This contribution aims to become an integral part of patient-centered design that needs to accommodate the informational needs of other stakeholders, specifically data analysts. The guidelines are as follow:

Methodological Guideline 1: *Understanding and Designing for Stakeholders*

- Identify and design from the perspective of the most vulnerable stakeholder and prioritize those needs when incorporating other stakeholder's requirements.

Methodological Guideline 2: Data Provenance

- Trace the production of data to understand the shortcomings of its creation and always maintain a critical attitude to what the data should represent

Design Guideline 1: Supporting Health Literacy

- Support automation by using a controlled vocabulary, such as a clinical data ontology, to streamline the data capture process.

Design Guideline 2: Prioritizing Interactivity and Visual Feedback

- Accommodate the health consumer's health literacy through visual mapping of the controlled vocabulary to facilitate interaction and visual feedback.

These contributions provide some guidance for deploying mixed methods, specifically the use of short-term ethnography to inform an experiment in the context of designing VA systems in precision medicine. This contribution helps to expand on the social science methods in Visual Analytics by focusing on the situated nature of cognition and understanding the human agents involved in the activity. This allows incorporation of qualitative methods along with traditional VA methods for advancing the field and better supporting analytics, especially when working in a Big Data Space where there is sensitive data and visual exploratory analysis might better support the discovery of patterns and correlations.

5.3. Future Work

Future work should first look into fixing the current system limitations; that means that the next version of the prototype should have more refined categories, support interactivity through the body graphic, a text entry box, and a toggle to switch between medical history and symptoms. Additionally, another challenge for the next iteration is the design of assets that help to map more abstract conditions.

In the medium term, future work could shift the focus back to the analysts and assess whether the captured data from the participants could support the exploration of

medical history and diseases in relation to the analysis of biomolecular data. Due to the inherent complexity of visualizing ontologies, future work could scaffold existing visualization techniques rather than require the development of new tools. In that sense, new studies could trace the whole process, from health consumers providing data, to analysts correlating descriptive data with biomolecular data when creating the reports. In that regard, it would be possible to see how interactive visualizations could impact the work of analysts. Along that same line, the unit of analysis could focus on the activity as a whole rather than only on health consumers reporting their conditions

Another possible avenue would be to work directly with physicians to cater to their data needs to create a more streamlined process between the three stakeholders. This possible path would be the more relevant for the mass implementation of precision medicine technologies that provide complex data in a report that gives insights to health consumers and facilitates clinical diagnostics by caregivers. In that same line, incorporating the physician's voices in the product design process could help to address their challenges when communicating with, and understanding the patient's experience.

Long term future work could go back to the setting and then conduct a cognitive ethnography to see how the artefacts became embedded in the setting and how they are being used. This could close the loop between design, iteration, deployment of technology, and full integration by the analysts. By doing so, it would be possible to test the hypothesis of whether these kinds of approaches have a significant impact on the workflow of analysts.

5.4. Chapter Summary

Chapter 5 summarized the thesis, showing an overview of how the research was designed, the findings from the fieldwork, and the exploratory sequential data analysis. This led to designing a data-capture system, two methodological and two design guidelines for visual analytics in precision medicine. Short-term future work involves incorporating the feedback and recommendations from the participants of the user-testing. In the medium-term, it is to create data ontology visualizations that help to visualize the captured data to find new disease correlations, as well as the creation of new algorithms and supporting automation. Finally, in the long term, cognitive ethnography can serve to assess the deployment and use of these types of technologies in their own setting.

Epilogue.

Becoming Data: Part 2

Back to work: Human Empathy to Data

Sitting back at my cubicle, I removed the small, round band-aid from my arm and rubbed the tiny red spot on my skin. I reflected about my tangible relation to data and how I transmuted into it. My interaction to it happened through a terminal, be it my smartphone, tablet, or computer; and only then did data become tangible, through my fingers and wrists. The delicacy of biomolecular data was difficult to grasp through the cold, metallic and/or plastic materials of the terminals, the importance becoming lost in the commonality of the materials. Now that I had consciously become data, I realized a digital version of myself had existed and had been evolving along with me for years. Furthermore, I realized that my relation to data also hurt, that actual fluids of my body were extracted to create a digitized version of me. A version that could be aggregated to contain almost all information that quantified who I was.

Becoming data allowed me to better understand how, just like the physics behind Velcro's biomimicry, current technology companies have come to realize the value of many single data sources of one's life to know more about us, the data chimeras who consume, work for, and are used and sold as product. Most importantly, becoming data provided me with a true exercise in vulnerability as an ethnographer, one that allowed me to empathize with aggregated records sitting in a database. While there are ethical implications concerning an employee benefit such as a biomolecular test, it is an exercise in vulnerability that reminds me how data was both precious and intangible, delicate and invisible.

When I worked in my cubicle, data was odorless, its smell quickly sterilized by the air conditioner, yet I know for a fact that on the morning when I went to give a blood sample, my data smelt like my cologne, a lingering smell accompanying my sampling tubes until

they were put in a freezer and analyzed. Now all of my smell's essence has been stripped down, joining the rest of the sterilized data. Becoming data made me feel like a window shopper, seeing and touching the store's window, but with the actual smells and warmth of the goods behind a glass screen. Even though I feel a physical separation from it, I cannot forget anymore the importance of what that data represents, the molecular information of a human being, one who could be me.

References

- Agro, C., & Denne, L. (2019). Twins get some “mystifying” results when they put 5 DNA ancestry kits to the test. *CBC: Technology & Science - MARKETPLACE*, 1–9.
- Anderson, K., Burford, O., & Emmerton, L. (2016). Mobile health apps to facilitate self-care: A qualitative study of user experiences. *PLoS ONE*, *11*(5), 1–21. <https://doi.org/10.1371/journal.pone.0156164>
- Anwar, M. A., Barrera-Machuca, A. A., Calderon, N., Wang, W., Tausan, D., Vayali, T., ... Fraser, R. (2020). Value-based healthcare delivery through metabolomics-based personalized health platform. *Healthcare Management Forum*. <https://doi.org/10.1177/0840470420904733>
- Arias-Hernandez, R., Green, T., & Fisher, B. (2012). From Cognitive Amplifiers to Cognitive Prosthesis: Understandings of the Material Basis of Cognition in Visual Analytics. *Interdisciplinary Science Reviews*, *37*(1), 4–18. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Armijo, D., McDonnell, C., & Werner, K. (2005). *Selecting an EHR for Your Practice: Evaluating Usability*. *Journal of Biomedical Informatics* (Vol. 38). <https://doi.org/10.1136/amiajnl-2012-001458>
- Armijo, D., McDonnell, C., & Werner, K. (2009). *Electronic Health Record Usability Interface Design Considerations* (Vol. 09). Retrieved from <http://healthit.ahrq.gov/sites/default/files/docs/citation/09-10-0091-2-EF.pdf>
- Arthur, B. (2009). Phenomena. In *The Nature of Technology: How it is and how it evolves*. New York: Free Press.
- Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). Personality and patterns of Facebook usage. *Proceedings of the 3rd Annual ACM Web Science Conference on - WebSci '12*, 24–32. <https://doi.org/10.1145/2380718.2380722>
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Bird, C. M. (2005). How I stopped dreading and learned to love transcription. *Qualitative Inquiry*, *11*(2), 226–248. <https://doi.org/10.1177/1077800404273413>

- Blomberg, J., Giacomi, J., Mosher, A., & Swenton-Wall, P. (1993). Ethnographic field methods and their relation to design. In D. Schuler & A. Namioka (Eds.), *Participatory Design: Principles and Practices* (pp. 123–155). New Jersey: Lawrence Erlbaum Associates. <https://doi.org/10.1201/9780203744338>
- Bloom, B. (1956). *Taxonomy of educational objectives : the classification of educational goals*. (D. McKay, Ed.). New York.
- Bryman, A. (2012). *Social research methods*. Oxford University Press. New York.
- Buchanan, R. (1992). Wicked Problems in Design Thinking. *Design Issues*, 8(2), 5–21. <https://doi.org/10.2307/1511637>
- Calderon, N. A. (2020). *Visual Analytics in Personalized Health: A Study of the Expert Analyst – Health Consumer Relationship in a Direct-to-Consumer Data Model*. Simon Fraser University.
- Carter, T. F. (2018). Disciplinary (Per)Mutations of Ethnography. *Cultural Studies - Critical Methodologies*, 18(6), 392–399. <https://doi.org/10.1177/1532708617746423>
- Charlsie, A., & Denne, L. (2019). Twins get some “mystifying” results when they put 5 DNA ancestry kits to the test.
- Cifuentes, E., Alamo, U., Kendall, T., Brunkard, J., Kendall, T., Cifuentes, E., & Brunkard, P. J. (2006). Rapid Assessment Procedures in Environmental Sanitation Research: A Case Study from the Northern Border of Mexico. *Canadian Journal of Public Health / Revue Canadienne de Santé Publique*, 97(1), 24–28.
- Clifford, J. (1986). Introduction: Partial Truths Interdisciplinary. In J. Clifford & G. E. Marcus (Eds.), *Writing Culture: The Poetics and Politics of Ethnography* (pp. 1–26). Berkeley: University of California Press.
- Collins, H., Evans, R., & Gorman, M. (2007). Trading zones and interactional expertise. *Studies in History and Philosophy of Science Part A*, 38(4), 657–666. <https://doi.org/10.1016/j.shpsa.2007.09.003>
- Collins, H. M., & Evans, R. (2002). The third wave of science studies: Studies of expertise and experience. *Social Studies of Science*, 32(2), 235–296. <https://doi.org/10.1177/0306312702032002003>
- Crawford, K., & Joler, V. (2018). Anatomy of an AI System: The Amazon Echo As An

Anatomical Map of Human Labor.

- Creswell, J. W. (2007). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. Book (Vol. 2nd ed). <https://doi.org/10.1016/j.aenj.2008.02.005>
- Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. *Research design Qualitative quantitative and mixed methods approaches*. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Creswell, J. W. (2018). "Qualitative Inquiry and Research Design: Choosing Among Five Approaches."
- Cross, N. (2001). Designerly Ways of Knowing: Design Discipline Versus Design Science. *Design Issues*, 17(3), 49–55. <https://doi.org/10.1002/0471684228.egp06293>
- Cruikshank, J. (1994). Oral Tradition and Oral History: Reviewing Some Issues. *The Canadian Historical Review*, 75(3), 403–418. Retrieved from <https://muse.jhu.edu/article/574633/summary>
- Demelo, J., Parsons, P., & Sedig, K. (2017). Ontology-Driven Search and Triage: Design of a Web-Based Visual Interface for MEDLINE. *JMIR Medical Informatics*, 5(1), e4. <https://doi.org/10.2196/medinform.6918>
- Duke, P., Frankel, R. M., & Reis, S. (2013). How to Integrate the Electronic Health Record and Patient-Centered Communication Into the Medical Visit: A Skills-Based Approach. *Teaching and Learning in Medicine*, 25(4), 358–365. <https://doi.org/10.1080/10401334.2013.827981>
- Elliott, D. (2016). WRITING. In D. Elliott & D. Culhane (Eds.), *A Different Kind of Ethnography: Imaginative Practices and Creative Methodologies* (pp. 23–44). Toronto: University of Toronto Press.
- Elsden, C., Kirk, D. S., & Durrant, A. C. (2016). A Quantified Past: Toward Design for Remembering With Personal Informatics. *Human-Computer Interaction*, 31(6), 518–557. <https://doi.org/10.1080/07370024.2015.1093422>
- Elson, R. B., Faughnan, J. G., & Connelly, D. P. (1997). An Industrial Process View of Information Delivery to Support Clinical Decision Making: Implications for Systems Design and Process Measures. *Journal of the American Medical Informatics Association*, 4(4), 266–278. <https://doi.org/10.1136/jamia.1997.0040266>

- Emiral G.O., Aygar H., Isiktekin B., Göktas S., Dagtekin G., U. A. (2018). Health literacy scale-European union-Q16: a validity and reliability study in Turkey. *International Research Journal of Medical Sciences*, 6(1), 1–7. Retrieved from http://www.isca.in/MEDI_SCI/Archive/v6/i1/1.ISCA-IRJMeds-2017-016.php
- Engelbart, D. (1962). *Augmenting Human Intellect: A Conceptual Framework*. California.
- Engeström, Y. (1990). *Learning, Working and Imagining: Twelve Studies in Activity Theory*. Helsinki: Orienta-Konsultit Oy.
- Erickson, F. (2011). A History of Qualitative Inquiry in Social and Educational Research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research*.
- Fisher, B., Green, T. M., & Arias-Hernandez, R. (2011). Visual analytics as a translational cognitive science. *Topics in Cognitive Science*, 3(3), 609–625. <https://doi.org/10.1111/j.1756-8765.2011.01148.x>
- Fleck, L. (1979). *Genesis and development of a scientific fact*. (T. J. Trenn & R. K. Merton, Eds.). Chicago: University of Chicago Press.
- Friendly, M. (2008). A Brief History of Data Visualization. In C. Chen, W. Haerdle, & A. Unwin (Eds.), *Handbook of Data Visualization* (pp. 15–56). Berlin Heidelberg.
- Gadamer, H.-G. (2004). Theory, Technology, Praxis. In J. Gaiger & N. Walker (Eds.), *The Enigma of Health: The Art of Healing in a Scientific Age* (Reprint). Cambridge: Polity Press.
- Geisler, C., & Swarts, J. (2019). *Coding Streams of Language. Techniques for the Systematic Coding of Text, Talk, and Other Verbal Data*. Boulder: University Press of Colorado. <https://doi.org/10.1017/CBO9781107415324.004>
- Glueck, M., Naeini, M. P., Doshi-Velez, F., Chevalier, F., Khan, A., Wigdor, D., & Brudno, M. (2018). PhenoLines: Phenotype Comparison Visualizations for Disease Subtyping via Topic Models. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 371–381. <https://doi.org/10.1109/TVCG.2017.2745118>
- González, C., Blobel, B. G. M. E., & López, D. M. (2012). Formal specification of an ontology-based service for EHR interoperability. *Studies in Health Technology and Informatics*, 180(4), 78–82. <https://doi.org/10.3233/978-1-61499-101-4-78>

- Green, T.M., & Fisher, B. (2011). The Personal Equation of Complex Individual, 38–57.
- Green, Tera Marie, Ribarsky, W., & Fisher, B. D. (2009). Building and applying a human cognition model for visual analytics. *Information Visualization*, 8(1), 1–13. <https://doi.org/10.1057/palgrave.ivs.2008.28>
- Haggstrom, D. A., Saleem, J. J., Russ, A. L., Jones, J., Russell, S. A., & Chumbler, N. R. (2011). Lessons learned from usability testing of the VA's personal health record. *Journal of the American Medical Informatics Association*, 18(SUPPL. 1), 13–17. <https://doi.org/10.1136/amiajnl-2010-000082>
- Hasso Plattner Institute of Design at Stanford University. (2020). Get Started with Design Thinking. Retrieved January 31, 2020, from <https://dschool.stanford.edu/resources/getting-started-with-design-thinking>
- Hazlehurst, B., Gorman, P. N., & McMullen, C. K. (2008). Distributed cognition: An alternative model of cognition for medical informatics. *International Journal of Medical Informatics*, 77(4), 226–234. <https://doi.org/10.1016/j.ijmedinf.2007.04.008>
- Hoehndorf, R., Schofield, P. N., & Gkoutos, G. V. (2015). The role of ontologies in biological and biomedical research: A functional perspective. *Briefings in Bioinformatics*, 16(6), 1069–1080. <https://doi.org/10.1093/bib/bbv011>
- Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. *ACM Transactions on Computer-Human Interaction*, 7(2), 174–196. <https://doi.org/10.1145/353485.353487>
- Holten, D. (2006). Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 741–748. <https://doi.org/10.1109/TVCG.2006.147>
- Howell, S. (2017). Two or three things I love about ethnography. *HAU: Journal of Ethnographic Theory*, 7(1), 15–20. <https://doi.org/10.14318/hau7.1.004>
- Hughes, J., King, V., Rodden, T., & Anderson, H. (1995). The Role of Ethnography in Interactive Systems Design. *Interactions*, 56–65.
- Hutchins, E. (2000). *Cognition in the Wild* (Fourth). MIT Press.
- IDEO. (2020). Design Thinking. Retrieved January 31, 2020, from

<https://www.ideo.com/pages/design-thinking>

Isaak, J., & Hanna, M. J. (2018). User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, 51(8), 56–59.
<https://doi.org/10.1109/MC.2018.3191268>

James, W. (1890). *The principles of psychology*. New York: Henry Holt and company.

Jen, N. (2018). Design Thinking is Bullsh*t. United States: 99U [Video].

Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: Towards better research applications and clinical care. *Nature Reviews Genetics*, 13(6), 395–405. <https://doi.org/10.1038/nrg3208>

Johnson, B., & Shneiderman, B. (1993). Tree-maps: a space-filling approach to the visualization of hierarchical information structures. *Proceeding Visualization '91*, 284–291. <https://doi.org/10.1109/VISUAL.1991.175815>

Johnson, J. (2014). *Designing with the mind in mind simple: Simple guide to understanding user interface design guidelines*. Amsterdam: Elsevier Ltd.

Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: Activity theory and interaction design*. MIT Press.

Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., & Giannopoulou, E. (2007). Ontology visualization methods - A survey. *ACM Computing Surveys*, 39(4), 1–43.
<https://doi.org/10.1145/1287620.1287621>

Kazubowski-Houston, M., & Magnat, V. (2018). Introduction to Special Issue: The Transdisciplinary Travels of Ethnography. *Cultural Studies - Critical Methodologies*, 18(6), 379–391. <https://doi.org/10.1177/1532708617737100>

Keim, D., & Zhang, L. (2011). Solving problems with visual analytics: Challenges and applications. *ACM International Conference Proceeding Series*.
<https://doi.org/10.1145/2024288.2024290>

Kerren, A., Stasko, J., Fekete, J.-D., & North, C. (2008). Visual Analytics: Definition, Process, and Challenges. In A. Kerren, J. Stasko, J.-D. Fekete, & C. North (Eds.), *Information Visualization: Human-Centered Issues and Perspectives*. Berlin, Heidelberg: Springer. Retrieved from
<https://link.springer.com/content/pdf/10.1007%2F978-3-540-70956-5.pdf>

- Knight, W. (2017, April). The Dark Secret at the Heart of AI. *MIT Technology Review*, 1–10. Retrieved from <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>
- Köhler, S., Vasilevsky, N. A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., ... Robinson, P. N. (2017). The human phenotype ontology in 2017. *Nucleic Acids Research*, 45(D1), D865–D876. <https://doi.org/10.1093/nar/gkw1039>
- Kohlhammer, J., Keim, D., Pohl, M., Santucci, G., & Andrienko, G. (2011). Solving problems with visual analytics. *Procedia Computer Science*, 7, 117–120. <https://doi.org/10.1016/j.procs.2011.12.035>
- Kuhn, T. (1963). *The structure of scientific revolutions*. University of Chicago Press.
- Kuutti, K. (1992). Identifying potential CSCW applications by means of activity theory concepts: a case example. *Computer Supported Cooperative Work*, 233–240. Retrieved from <http://dl.acm.org/citation.cfm?id=150955>
- Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, 11(1), 65–100. [https://doi.org/10.1016/S0364-0213\(87\)80026-5](https://doi.org/10.1016/S0364-0213(87)80026-5)
- Latour, B. (1986). Visualization and Cognition. *Knowledge and Society*. <https://doi.org/10.1002/9780470979587.ch9>
- Lave, J. (1988). *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge, Massachusetts: Cambridge University Press.
- Lorensen, B. (2004). On the Death of Visualization. *Position Papers NIH/NSF Proc. Fall 2004 Workshop Visualization Research Challenges.*, 1(2), 5.
- Maglogiannis, I., Billiris, A., & Valavanis, I. (2016). Redesigning EHRs and Clinical Decision Support Systems for the Precision Medicine Era. *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 14, 313–316.
- Marchionini, G., Rimer, B. K., & Wildemuth, B. (2007). *Evidence Base for Personal Health Record Usability: Final Report to the National Cancer Institute* (Vol. 1). Chapel Hill.
- Mate, S., Köpcke, F., Toddenroth, D., Martin, M., Prokosch, H. U., Bürkle, T., & Ganslandt, T. (2015). Ontology-based data integration between clinical and

research systems. *PLoS ONE*, 10(1), 1–20.
<https://doi.org/10.1371/journal.pone.0116656>

- Millen, D. R. (2000). Rapid ethnography: Time deepening strategies for HCI field research. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS*, 280–286.
- Miller, G. A. (1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 101(2), 343–352.
<https://doi.org/10.1037/0033-295x.101.2.343>
- Mosaly, P. R., Mazur, L., & Marks, L. B. (2016). Usability evaluation of electronic health record system (EHRs) using subjective and objective measures. *CHIIR 2016 - Proceedings of the 2016 ACM Conference on Human Information Interaction and Retrieval*, 313–316. <https://doi.org/10.1145/2854946.2854985>
- Mougin, F., Auber, D., Bourqui, R., Diallo, G., Dutour, I., Jouhet, V., ... Thébault, P. (2018). Visualizing omics and clinical data: Which challenges for dealing with their variety? *Methods*, 132, 3–18. <https://doi.org/10.1016/j.ymeth.2017.08.012>
- Mueller, K., Garg, S., Nam, J. E., Berg, T., & McDonnell, K. T. (2011). Can computers master the art of communication?: A focus on visual analytics. *IEEE Computer Graphics and Applications*, 31(3), 14–21. <https://doi.org/10.1109/MCG.2011.39>
- Munzner, T. (2009). A Nested Process Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 921–928.
<https://doi.org/10.1109/TVCG.2009.111>
- Munzner, T. (2014). Analysis : Four Levels for Validation. In A. K. Peters (Ed.), *Visualization Analysis and Design*. Boca Raton, FL: CRC Press.
- Munzner, T. (2015). What's Vis, and Why Do It? In *Visualization Analysis and Design*. CRC Press.
- Nardi, B. A. (1995). *Studying Context: A Comparison of Activity Theory, Situated Action Models, and Distributed Cognition*. Flor and Hutchins.
- Nersessian, N. J. (2004). Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. *Scientific and Technological Thinking*, 17–56. <https://doi.org/10.4324/9781410611314>

- Nersessian, N. J. (2013). Interdisciplinarity in Action: Cognitive Ethnography of Bioengineering Sciences Research Laboratories. *Perspectives on Science*, 27(4), 553–581. <https://doi.org/10.1162/POSC>
- Nielsen Norman Group. (2020). Design thinking 101. Retrieved March 16, 2020, from <https://www.nngroup.com/articles/design-thinking/>
- NIH. (2016). The Cost of Sequencing a Human Genome. Retrieved from <https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/>
- NIH. (2020). What is precision medicine? Retrieved from <https://ghr.nlm.nih.gov/primer/precisionmedicine/definition>
- Nolasco, A., Barona, C., Tamayo-Fonseca, N., Irlles, M. A., Mas, R., Tuells, J., & Pereyra-Zamora, P. (2010). Alfabetización en salud: propiedades psicométricas del cuestionario HLS-EU-Q16, 23(xx), 727–735. <https://doi.org/10.1016/j.engappai.2010.01.015>
- Norman, D. (1986). Cognitive Engineering. In *User Centered System: Design New Perspectives on Human-computer Interaction* (p. 526). Taylor & Francis. <https://doi.org/10.1016/b978-0-444-70304-0.50008-x>
- Norman, D. (1993). *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. New York: Addison-Wesley.
- Norman, D. (2002). The Psychopathology of Everyday Things. In *The design of everyday things* (p. 257). New York: Basic Books.
- Norman, D. (2013). *The Design of Everyday Things* (Revised Ed). New York: Basic Books.
- O'Connell Davidson, J. (2008). If no means no, does yes mean yes? Consenting to research intimacies. *History of the Human Sciences*. <https://doi.org/10.1177/0952695108095511>
- O'Dell, T., & Willim, R. (2013). Transcription and the senses: Cultural analysis when it entails more than words. *Senses and Society*, 8(3), 314–334. <https://doi.org/10.2752/174589313X13712175020550>
- O'Neil, C. (2016). Weapons of Math Destruction. Talks at Google [Video]. Retrieved from <https://www.youtube.com/watch?v=TQHs8SA1qpk>

Oberquelle, H., Kupka, I., & Maass, S. (1983). A view of human-machine communication and co-operation. *International Journal of Man-Machine Studies*, 19(4), 309–333.

Office of the National Coordinator for Health Information Technology. (2019). What are the differences between electronic medical records, electronic health records, and personal health records? Retrieved from <https://www.healthit.gov/faq/what-are-differences-between-electronic-medical-records-electronic-health-records-and-personal>

Pasquale, F. (2015). *The Black Box Society* (First). Cambridge, Massachusetts: Harvard University Press.

Patel, V. L., & Kushniruk, A. W. (1998). Interface design for health care environments: the role of cognitive science. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, 29–37.

Patel, Vimla L., & Kaufman, D. R. (1998). Medical informatics and the science of cognition. *Journal of the American Medical Informatics Association*, 5(6), 493–502. <https://doi.org/10.1136/jamia.1998.0050493>

Pentagram. (2018). Design Thinking, Examined. Retrieved January 31, 2020, from <https://www.pentagram.com/news/design-thinking-examined>

Peters, B. (2016). Introduction. *Digital Keywords A Vocabulary of Information Society and Culture*.

Pink, S. (2011). Multimodality, multisensoriality and ethnographic knowing: Social semiotics and the phenomenology of perception. *Qualitative Research*, 11(3), 261–276. <https://doi.org/10.1177/14687941111399835>

Pink, S., & Morgan, J. (2013). Short-term ethnography: Intense routes to knowing. *Symbolic Interaction*, 36(3), 351–361. <https://doi.org/10.1002/symb.66>

Prodanov, D. (2011). Data Ontology and an Information System Realization for Web-Based Management of Image Measurements. *Frontiers in Neuroinformatics*, 5(November), 1–18. <https://doi.org/10.3389/fninf.2011.00025>

Reardon, J., & TallBear, K. (2012). “Your DNA Is Our History.” *Current Anthropology*, 53(S5), S233–S245. <https://doi.org/10.1086/662629>

Regalado, A. (2019). A consumer DNA testing company has given the FBI access to its

two million profiles. Retrieved from <https://www.technologyreview.com/the-download/612875/a-consumer-dna-testing-company-has-given-the-fbi-access-to-its-two-million/>

Ribarsky, W., & Fisher, B. (2016). The Human-Computer System: Towards an Operational Model for Problem Solving. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 1446–1455. <https://doi.org/10.1109/HICSS.2016.183>

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences*, 4, 155–169. <https://doi.org/10.1080/01636609209550084>

Robinson, P. N., & Mundlos, S. (2010). The Human Phenotype Ontology. *Clinical Genetics*, 77(6), 525–534. <https://doi.org/10.1111/j.1399-0004.2010.01436.x>

Rothman, J. (2014). Why Is Academic Writing So Academic? *The New Yorker*. <https://doi.org/10.377/0033-2909.l26.1.78>

Russell, J. (2019). Singapore says personal details of 14200 HIV patients were posted online. Retrieved from https://techcrunch.com/2019/01/28/singapore-14200-hiv-patient-details-online/?utm_source=tcfbpage&sr_share=facebook

Schön, D. A. (2000). *Reflective practitioner: how professionals think in action*. New York: BasicBooks.

Scrimshaw, N., & Gleason, G. (1992). *Rapid Assessment Procedures - Qualitative Methodologies for Planning and Evaluation of Health Related Programmes*. Boston: International Nutrition Foundation for Developing Countries (INFDC). Retrieved from <http://archive.unu.edu/unupress/food2/UIN08E/UIN08E00.HTM#Contents>

Scrimshaw, S., Carballo, M., Ramos, L., & Blair, B. (1991). The AIDS Rapid Anthropological Assessment Procedures: A Tool for Health Education Planning and Evaluation. *Health Education Quarterly*, 18(1), 111–123. https://doi.org/10.5005/jp/books/11413_20

Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design Study Methodolgy: Reflections from the Trenches and the Stacks, 1(12), 2431–2440.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135–1145. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18846087>

Schneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information

visualization tools, 1. <https://doi.org/10.1145/1168149.1168158>

Simon, H. A. (1996a). The Science of Design: Creating the Artificial. In *The Sciences of the Artificial* (pp. 111–138). MIT Press.
<https://doi.org/10.5040/9781474282932.0013>

Simon, H. A. (1996b). *The Sciences of the Artificial* (Third). Cambridge, Massachusetts: MIT Press. <https://doi.org/10.2307/3102825>

Simonsen, J., & Kensing, F. (1998). Make Room for Ethnography in Design! Overlooked collaborative and educational prospects. *Journal of Computer Documentation*, 22(1), 20–30. <https://doi.org/http://dx.doi.org/10.1145/571773.571781>

Small, S. S., Peddie, D., Ackerley, C., Hohl, C. M., & Balka, E. (2017). Patient Perceptions About Data Sharing & Privacy: Insights from ActionADE. *Studies in Health Technology and Informatics*, 241, 109–114. <https://doi.org/10.3233/978-1-61499-794-8-109>

SNOMED International. (2020). Why SNOMED CT? Retrieved from <https://www.snomed.org/snomed-ct/why-snomed-ct>

Sousanis, N. (2015). the importance of seeing double and and then some. In *Unflattering* (pp. 29–48). Cambridge, Massachusetts: Harvard University Press.

Spence, B. (2011). The Broker. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6431 LNCS, 10–22. https://doi.org/10.1007/978-3-642-19641-6_2

Stasko, J., Görg, C., & Liu, Z. (2008). Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2), 118–132.
<https://doi.org/10.1057/palgrave.ivs.9500180>

Stead, W. W., & Lin, H. S. (2009). Computational technology for effective health care: Immediate steps and strategic directions. *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions*, 1–120.
<https://doi.org/10.17226/12572>

Stevens, J. A., DeCosta, B. F., Johnston, H. P., Lamb, M. J., & Pond, N. G. (1879). *The magazine of american history, with notes and queries* (Bands 3-4). A. S. Barnes.

Stoller, P. (2016). In Defense of Ethnography. *Huffington Post*, 1–13. Retrieved from

https://www.huffpost.com/entry/in-defense-of-ethnography_b_8028542

Suchman, L. A. (1989). Plans and Situated Actions: The Problem of Human Machine Communication. *Contemporary Sociology*, 18(3), 414.
<https://doi.org/10.2307/2073874>

The Design Council. (2020). The Double Diamond: A universally accepted depiction of the design process. Retrieved January 31, 2020, from <https://www.designcouncil.org.uk/news-opinion/double-diamond-universally-accepted-depiction-design-process>

The HIBAR Research Alliance. (2020). The HIBAR Research Alliance: Enhancing Fundamental Research Excellence in Service to Society. Retrieved January 31, 2020, from <https://hibar-research.org/>

Thomas, J. J., & Cook, K. A. (2005). Illuminating the path: The research and development agenda for visual analytics. *IEEE Computer Society*, 184.
<https://doi.org/10.3389/fmicb.2011.00006>

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>

Tory, M., & Möller, T. (2004). Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics*, 10(1), 72–84.
<https://doi.org/10.1109/TVCG.2004.1260759>

Trnka, S. H. (2016). Digital Care: Agency and Temporality in Young People's Use of Health Apps. *Engaging Science, Technology, and Society*, 2, 248.
<https://doi.org/10.17351/ests2016.119>

Tufte, E. R. (1983). *The visual display of quantitative information* (Second Edi). Cheshire: Graphics Press.

U.S. Department of Health & Human Services. (2020). System Usability Scale (SUS). Retrieved February 4, 2020, from <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

University of Wisconsin. (2020). What Is a Healthcare Data Analyst? How Do You Become One? Retrieved January 31, 2020, from <https://himt.wisconsin.edu/experience-uw-himt/healthcare-data-analyst/>

- van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605. <https://doi.org/10.1007/s10479-011-0841-3>
- Van Wijk, J. J. (2005). The value of visualization. *Proceedings of the IEEE Visualization Conference*, 11. <https://doi.org/10.1109/VIS.2005.102>
- Ware, C. (2008a). Visual Queries. In *Visual Thinking : For Design* (pp. 1–22). Elsevier Science & Technology. <https://doi.org/10.1016/b978-012168030-5/50012-7>
- Ware, C. (2008b). What We Can Easily See. *Visual Thinking*, 23–42. <https://doi.org/10.1016/b978-0-12-370896-0.00002-0>
- Ware, C. (2012). *Information visualization : perception for design* (3rd ed.). Boston: Morgan Kaufmann.
- Weber, G. H., Carpendale, S., Ebert, D., Fisher, B., Hagen, H., Shneiderman, B., & Ynnerman, A. (2017). Apply or Die: On the Role and Assessment of Application Papers in Visualization. *IEEE Computer Graphics and Applications*, 38(3), 96–104. <https://doi.org/10.1109/MCG.2017.51>
- West, V. L., Borland, D., & Hammond, W. E. (2014). Innovative information visualization of electronic health record data: a systematic review. *Journal of the American Medical Informatics Association*, 330–339. <https://doi.org/10.1136/amiajnl-2014-002955>
- WHO. (2020). Classification of Diseases (ICD). Retrieved from <https://www.who.int/classifications/icd/en/>
- Wilkinson, L. (2005). *The Grammar of Graphics*. (J. Chambers, D. Hand, & W. Haerdle, Eds.) (2nd ed.).
- Yang, Y., Wybrow, M., Li, Y.-F., Czauderna, T., & He, Y. (2019). OntoPlot: A Novel Visualisation for Non-hierarchical Associations in Large Ontologies. *IEEE Transactions on Visualization and Computer Graphics*, 1–1. <https://doi.org/10.1109/tvcg.2019.2934557>
- Zachry, M., & Thralls, C. (2004). Cross-Disciplinary Exchanges: An Interview with Edward R. Tufte. *Technical Communications Quarterly*, 13(4), 447–462. https://doi.org/10.1207/s15427625tcq1304_5

Zhang, J., & Norman, D. (2003). Representations in distributed cognitive tasks.
Cognitive Science, 18(1), 87–122. [https://doi.org/10.1016/0364-0213\(94\)90021-3](https://doi.org/10.1016/0364-0213(94)90021-3)

Appendix A.

Ethnographic Interview: Guiding questions

Research question	Interview question
<p>How are team identity and membership, and trust created in the data team?</p>	<ul style="list-style-type: none"> - How would you explain your job role? - How would you define the word “team”? <ul style="list-style-type: none"> o What do you see is your role as a part of the Data Team? o What is your experience working with the team? o How do you work with other members of the Data team? o How do you identify as a part of the team? - Have you had a previous job/ internship? <ul style="list-style-type: none"> o Compare the job environment to your previous one? - Which 3 words would you pick to describe working at X
<p>How do the data analysts conceive their own work and its importance? How has that notion evolved since they started working at the company?</p>	<ul style="list-style-type: none"> - What do you do at X? - What did you think you would be doing when you started?
<p>What is the baseline data that the analysts require to be able to represent the health consumer case and make inferences about their conditions? What is the data analyst’s relation to the data and how does that impact their work philosophy?</p>	<ul style="list-style-type: none"> - What do you need to do your work? (tools, data, software, etc.) - What data does the health consumer usually give you? <ul style="list-style-type: none"> o Is there a minimum amount of data to do a health consumer report? (Who establishes it?) - If you were in charge for making things mandatory in the Phenome Form, what sections/aspects would they be?

Appendix B.

Coding Scheme: Ethnographic Interviews

Data Challenges

Definition: Code as *Data Challenges* any transcription excerpt in which the analyst refers to their experiences regarding the quality of the data and/or the interpretation of biomolecular data.

Data Challenges may include instances where the analyst...

- (a) Utilizes the examples of molecules or proteins to convey instances where they faced issues with the quality of the data
- (b) Talks about the challenge of providing personalized recommendations based on the data they have available
- (c) Explains the process of cleaning the data and/or data conversions in relation to the analysis and/or interpretation of the data

Minimum Health consumer Data

Definition: Code as *Minimum Health consumer Data* any transcription excerpt in which the analyst refers to the minimum amount of data they require to provide recommendations and a health report to the health consumers/consumers.

Minimum Health consumer Data may include instances where the analyst...

- (a) Talks about the type of data that health consumers/consumers usually provide in their medical history form.
- (b) Explains how different data variables (i.e. age, sex, gender, weight, height) relate to the data analysis.
- (c) Show which aspects of the health consumer's family & clinical history, nutrition, and diet journals are essential to carrying out the data analysis.

Automatization

Definition: Code as *Automatization* any transcription excerpt in which the analyst refers to prospects and/or steps they have made to automatize the analysis process.

Automatization may include instances where the analyst...

- (a) Talks about the work that has been done so far in terms of automatization.
- (b) Explains how the rules for automatization are constructed and/or which aspects are considered for the process
- (c) Talks about the creation/curation of databases as part of the automatization efforts
- (d) Talks about issues or challenges they have had when automatizing the analysis

Accommodating Health Literacy

Definition: Code as *Accommodating Health Literacy* any transcription excerpt in which the analyst refers to measures they have taken to make the scientific aspect of the product accessible to lay people.

Accommodating Health Literacy may include instances where the analyst...

- (a) Talks about the steps they have made to reduce the scientific/medical jargon
- (b) Explains how the need to accommodate the literacy of health consumers has contributed to the design of the product
- (c) Talks about the medical knowledge of the general population and how to relate the results to them.

Product Evolution

Definition: Code as *Product Evolution* any transcription excerpt in which the analyst refers to the product (Direct-to-Consumer Personalized Medicine) or how the product has evolved over time.

Product Evolution may include instances where the analyst...

- (a) Talks about their experience with the product and/or how they have contributed to it
- (b) Explains the process for creating the product

Job Role

Definition: Code as *Job Role* any transcription excerpt in which the analyst refers to the tasks they perform as part of their job description.

Job Role may include instances where the analyst...

- (a) Talks about the role they were assigned when they joined the company
- (b) Explains their daily/regular tasks at the company

Teamwork

Definition: Code as *Teamwork* any transcription excerpt in which the analyst refers to their experiences working with other members of the Data Team.

Teamwork may include instances where the analyst...

- (a) Talks about how they coordinate with other teammates (i.e. communication, consulting each other)
- (b) Show how the data team works together and/or explains where their job role is located within the production pipeline.

Accountability

Definition: Code as *Accountability* any transcription excerpt in which the analyst refers to how they understand their responsibility in providing accurate recommendations to the health consumers and consumers.

Accountability may include instances where the analyst...

- (a) Talks about their own struggles when providing recommendations and/or how responsible they feel for it
- (b) Shows their responsibility and caution when working with the biomolecular data

Appendix C.

User Testing: Researcher's guide

Title: Interactive Electronic Health Records: A Usability Study

Quick explanation

- Precision Medicine
- Challenges

Consent

- Data and privacy
- Study instructions: **You can ask any questions. Avoid consulting your cellphone.**
 - (i) We will give you a form with questions to assess your level of health knowledge
 - (ii) You will use two different medical history forms to report on the medical history of a fictive health consumer. We will give you the information about the health consumer.
 - (iii) We will give a form with questions to answer to understand the usability of the medical history forms.
 - (iv) We will conduct a short interview to ask you about your experience using the medical history forms.
- Signature

HLS Q-16

- Give and wait for questions

Clinical vignette

- You can make any marking on the provided case vignette.

Interfaces

- For both interfaces, focus your answers on the provided clinical vignette
- Tell them not to click next on the interface!
- **Take a print screen before they click on the interface (leave snipping tool open)**

SUS

- See glossary

Interview

Appendix D.

User Testing: Clinical Vignettes

FEMALE MEDICAL HISTORY VIGNETTE

You are a 25-year-old woman. For the last 2 weeks, you've been extremely thirsty, peeing frequently and you've lost 10lbs/5kg. While people have said you look nice, you've also been feeling extremely tired, your vision is blurry and you've had trouble focusing when you are studying for your exams. You know something is not right. You go to your GP and they said to come to the ER to get checked out. You are worried.

You get hay fever every spring, and take a daily puffer for asthma, and you have an Epipen because you had a severe allergic reaction to lobster once, 7 years ago - but you've never needed to use it. You've been on the pill for 9 years.

At the hospital you get some blood tests. Your blood sugar was really high, and the doctor said it was called Type 2 Diabetes. No one in your family has diabetes. Your mom's dad recently died from a lung condition, a heart failure and he was a chronic alcoholic.

They admit you to the ICU for 1 night and then you got transferred for 2 days to the regular ward. You were treated with some medicines and they sent you home to do Insulin shots. You have to do one type of insulin (Lantus) twice a day, 25 units at breakfast and 25 units at bedtime. You also have to take another type of Insulin (Novolog) before meals. You were also told to measure your blood sugar 4 times a day.

You went back to check in with a hospital nurse 3 days later. She scheduled you in with an endocrinologist to see if you need an insulin pump. Your appointment is in 2 weeks.

MALE MEDICAL HISTORY VIGNETTE

You are a 25-year-old man. For the last 2 weeks, you've been extremely thirsty, peeing frequently and you've lost 10lbs/5kg. While people have said you look nice, you've also been feeling extremely tired, your vision is blurry, and you've had trouble focusing when you are studying for your exams. You know something is not right. You go to your GP and they said to come to the ER to get checked out. You are worried.

You get hay fever every spring, and take a daily puffer for asthma, and you have an EpiPen because you had a severe allergic reaction to lobster once, 7 years ago - but you've never needed to use it.

At the hospital you get some blood tests. Your blood sugar was really high, and the doctor said it was called Type 2 Diabetes. No one in your family has diabetes. Your mom's dad recently died from a lung condition, a heart failure and he was a chronic alcoholic.

They admit you to the ICU for 1 night and then you got transferred for 2 days to the regular ward. You were treated with some medicines and they sent you home to do Insulin shots. You have to do one type of insulin (Lantus) twice a day, 25 units at breakfast and 25 units at bedtime. You also have to take another type of Insulin (Novolog) before meals. You were also told to measure your blood sugar 4 times a day.

You went back to check in with a hospital nurse 3 days later. She scheduled you in with an endocrinologist to see if you need an insulin pump. Your appointment is in 2 weeks.

Appendix E.

User Testing: Clinical Vignette's Glossary

GLOSSARY

GP:

General Practitioner, a medical doctor.

ER:

An **Emergency Room** is a medical treatment facility specializing in emergency medicine.

The pill:

Birth control pills are a medicine with hormones that you take every day to prevent pregnancy.

ICU:

An **Intensive Care Unit** is a special department of a hospital or health care facility that provides intensive treatment medicine.

Endocrinologist:

An endocrinologist specializes in all things relating to our hormones.

Cumbersome:

large or heavy and therefore difficult to carry or use; unwieldy ||
Slow or complicated and therefore inefficient.

NOTES

Appendix F.

User Testing: Health consumer History Form

HEALTH CONSUMER HISTORY FORM

Age: _____ Sex: <input type="checkbox"/> F <input type="checkbox"/> M	
---	--

BRAIN HEALTH	CARDIO	GASTRO
<input type="checkbox"/> Depression <input type="checkbox"/> Seizures <input type="checkbox"/> Parkinson's disease <input type="checkbox"/> Ischemic stroke <input type="checkbox"/> Alzheimer's disease <input type="checkbox"/> Meningioma <input type="checkbox"/> Attention deficit disorder <input type="checkbox"/> Schizophrenia <input type="checkbox"/> Brain cancer <input type="checkbox"/> Malignant astrocytoma <input type="checkbox"/> Multiple sclerosis <input type="checkbox"/> Head cancer <input type="checkbox"/> Traumatic brain injury <input type="checkbox"/> Acoustic neuroma <input type="checkbox"/> Headaches <input type="checkbox"/> Dizziness <input type="checkbox"/> Fainting or loss of consciousness <input type="checkbox"/> Numbness or tingling	<input type="checkbox"/> High blood pressure <input type="checkbox"/> Cardiovascular disease <input type="checkbox"/> Disease of the heart muscle <input type="checkbox"/> Heart attack <input type="checkbox"/> Narrowing and hardening of arteries <input type="checkbox"/> Elevated total cholesterol <input type="checkbox"/> Decreased circulating cholesterol level <input type="checkbox"/> Blood clot in vein <input type="checkbox"/> Coronary artery disease <input type="checkbox"/> Ischemia <input type="checkbox"/> Blood clotting disorders <input type="checkbox"/> Chest pain <input type="checkbox"/> Palpitations <input type="checkbox"/> Fainting <input type="checkbox"/> Anemia <input type="checkbox"/> Swollen legs or feet <input type="checkbox"/> Stroke <input type="checkbox"/> Angina	<input type="checkbox"/> Inflammation of the large intestine <input type="checkbox"/> Cancer of the pancreas <input type="checkbox"/> Gastric cancer <input type="checkbox"/> Ulcerative colitis <input type="checkbox"/> Crohn's disease <input type="checkbox"/> Gallbladder cancer <input type="checkbox"/> Colorectal cancer <input type="checkbox"/> Esophageal cancer <input type="checkbox"/> Nasopharyngeal cancer <input type="checkbox"/> Nausea <input type="checkbox"/> Heartburn <input type="checkbox"/> Stomach pain <input type="checkbox"/> Vomiting <input type="checkbox"/> Increasing constipation <input type="checkbox"/> Persistent diarrhea <input type="checkbox"/> Blood in stools <input type="checkbox"/> Stomach or peptic ulcer <input type="checkbox"/> Black stools
METABOLISM	KIDNEYS	RARE DISORDER
<input type="checkbox"/> Insulin resistance <input type="checkbox"/> Obesity <input type="checkbox"/> Iron deficiency <input type="checkbox"/> Low blood sugar <input type="checkbox"/> Liver cancer <input type="checkbox"/> High blood sugar <input type="checkbox"/> Hyperlipidemia <input type="checkbox"/> Vitamin A deficiency <input type="checkbox"/> Type 2 diabetes <input type="checkbox"/> Liver inflammation <input type="checkbox"/> Thyroid cancer <input type="checkbox"/> Vitamin B Deficiency <input type="checkbox"/> Metabolic syndrome <input type="checkbox"/> Pre-diabetes <input type="checkbox"/> Vitamin C Deficiency <input type="checkbox"/> Type XV glycogen storage disease	<input type="checkbox"/> Elevated plasma aldosterone <input type="checkbox"/> Pheochromocytoma <input type="checkbox"/> Low blood sodium levels <input type="checkbox"/> Neuroblastoma <input type="checkbox"/> Low blood albumin <input type="checkbox"/> High blood sodium levels <input type="checkbox"/> Low blood aldosterone level <input type="checkbox"/> Congenital adrenal hyperplasia <input type="checkbox"/> Bladder cancer <input type="checkbox"/> Kidney cancer <input type="checkbox"/> High blood albumin levels <input type="checkbox"/> Chronic kidney disease <input type="checkbox"/> Frequent or painful urination <input type="checkbox"/> Blood in urine <input type="checkbox"/> Kidney stones	<input type="checkbox"/> Spastic paraplegia <input type="checkbox"/> Methylmalonic acidemia <input type="checkbox"/> Cardiovascular disease <input type="checkbox"/> Ewing's sarcoma <input type="checkbox"/> Phenylketonuria <input type="checkbox"/> Niemann-Pick Disease <input type="checkbox"/> Cancer cachexia <input type="checkbox"/> Hyperlipoproteinemia Type V <input type="checkbox"/> Gaucher's disease <input type="checkbox"/> Pendred Syndrome <input type="checkbox"/> Familial mediterranean fever <input type="checkbox"/> Thrombocytopenia-absent radius syndrome <input type="checkbox"/> Li-Fraumeni syndrome <input type="checkbox"/> Cancer syndrome <input type="checkbox"/> Leiomyomatosis
GENERAL	IMMUNE	MUSCLES & BONES
<input type="checkbox"/> Recent weight gain; how much _____ <input type="checkbox"/> Recent weight loss: how much _____ <input type="checkbox"/> Fatigue <input type="checkbox"/> Weakness <input type="checkbox"/> Fever	<input type="checkbox"/> Blood cancer <input type="checkbox"/> Cancer of lymphatic system <input type="checkbox"/> Complement deficiency <input type="checkbox"/> Chronic lymphocytic leukemia <input type="checkbox"/> Myeloma <input type="checkbox"/> T Cell Lymphoma <input type="checkbox"/> B Cell Lymphoma	<input type="checkbox"/> Rheumatoid arthritis <input type="checkbox"/> Low blood phosphate level <input type="checkbox"/> Low blood calcium levels <input type="checkbox"/> High blood phosphate levels <input type="checkbox"/> High blood calcium levels <input type="checkbox"/> Vitamin D deficiency <input type="checkbox"/> Bone cancer

<input type="checkbox"/> Night sweats <input type="checkbox"/> Difficulty falling asleep <input type="checkbox"/> Difficulty staying asleep <input type="checkbox"/> Poor appetite (Changes in appetite) <input type="checkbox"/> Food cravings <input type="checkbox"/> Hallucinations <input type="checkbox"/> Paranoia Poor concentration <input type="checkbox"/> Anxiety Increased Thirst	<input type="checkbox"/> Non-Hodgkin's Lymphoma <input type="checkbox"/> Hodgkin's Lymphoma <input type="checkbox"/> Chronic fatigue syndrome <input type="checkbox"/> Lupus erythematosus <input type="checkbox"/> HIV/AIDS	<input type="checkbox"/> Secondary bone cancer <input type="checkbox"/> Neck cancer <input type="checkbox"/> Numbness <input type="checkbox"/> Joint pain <input type="checkbox"/> Muscle weakness <input type="checkbox"/> Joint swelling
SKIN	LUNGS	WOMEN ONLY
<input type="checkbox"/> Skin cancer (melanoma) <input type="checkbox"/> Xanthomas <input type="checkbox"/> Carcinoma <input type="checkbox"/> Oral Cancer <input type="checkbox"/> Redness <input type="checkbox"/> Rash <input type="checkbox"/> Nodules/bumps <input type="checkbox"/> Hair loss <input type="checkbox"/> Color changes of hands or feet	<input type="checkbox"/> Asthma <input type="checkbox"/> Chronic obstructive pulmonary disease <input type="checkbox"/> Lung cancer <input type="checkbox"/> Surfactant metabolism dysfunction <input type="checkbox"/> Shortness of breath <input type="checkbox"/> Cough <input type="checkbox"/> Pneumonia <input type="checkbox"/> Pulmonary embolism <input type="checkbox"/> Emphysema <input type="checkbox"/> Tuberculosis	<input type="checkbox"/> Breast cancer <input type="checkbox"/> Hyperemesis gravidarum <input type="checkbox"/> Pre-eclampsia <input type="checkbox"/> Ovarian tumor <input type="checkbox"/> Choriocarcinoma <input type="checkbox"/> Endometrial cancer <input type="checkbox"/> Cervical cancer <input type="checkbox"/> Vulvar cancer <input type="checkbox"/> Pediatric germ cell cancer
LIVER	EYES	MEN ONLY
<input type="checkbox"/> Alpha-1 Antitrypsin Deficiency <input type="checkbox"/> Cholangiocarcinomas <input type="checkbox"/> HBV-Related Hepatocellular Carcinoma <input type="checkbox"/> Non-alcoholic fatty liver disease <input type="checkbox"/> Liver Cirrhosis <input type="checkbox"/> Hepatitis <input type="checkbox"/> Hemangiomas	<input type="checkbox"/> Glaucoma <input type="checkbox"/> Redness <input type="checkbox"/> Loss of vision <input type="checkbox"/> Double or blurred vision <input type="checkbox"/> Dryness	<input type="checkbox"/> Testicular cancer <input type="checkbox"/> Prostate cancer <input type="checkbox"/> 5-alpha reductase deficiency <input type="checkbox"/> Pediatric germ cell cancer

Appendix G.

User Testing: HLS-EU-Q16 questionnaire

PARTICIPANT ID: _____

DATE: _____

HLS-EU-Q16 Questionnaire

Instructions: For each of the following questions, mark one answer that best describes how you feel about your health literacy.

Health Literacy: capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions

On a scale from very easy to very difficult, how easy would you say it is to:

		Very easy	Fairly easy	Fairly difficult	Very difficult	Don't know
1	find information on treatments of illnesses that concern you?					
2	find out where to get professional help when you are ill?					
3	understand what your doctor says to you?					
4	understand your doctor's or pharmacist's instruction on how to take a prescribed medicine?					

5	judge when you may need to get a second opinion from another doctor?					
6	use information the doctor gives you to make decisions about your illness?					
7	follow instructions from your doctor or pharmacist?					
8	find information on how to manage mental health problems like stress or depression?					
9	understand health warnings about behaviour such as smoking, low physical activity and drinking too much?					
10	understand why you need health screenings?					
11	judge if the information on health risks in the media is reliable?					
12	decide how you can protect yourself from illness based on information in the media?					
13	find out about activities that are good for your mental well-being?					
14	understand advice on health from family members or friends?					
15	understand information in the media on how to get healthier?					
16	judge which everyday behaviour is related to your health?					

Appendix H.

User Testing: System Usability Scale

PARTICIPANT ID: _____

DATE: _____

System Usability Scale

Instructions: For each of the following statements, mark one answer that best describes your reactions to the **interactive medical form** (“system”) *today*.

On a scale from very easy to very difficult, how easy would you say it is to:

		Strongly Agree					Strongly disagree				
1	I think that I would like to use this system frequently										
2	I found the system unnecessarily complex										
3	I thought the system was easy to use										
4	I think that I would need the support of a technical person to be able to use this system										
5	I found the various functions in this system were well integrated										
6	I thought there was too much inconsistency in this system										

7	I would imagine that most people would learn to use this system very quickly					
8	I found the system very cumbersome to use					
9	I felt very confident using the system					
10	I needed to learn a lot of things before I could get going with this system					

Appendix I.

User Testing: Structured Interview Questions

Research question	Interview question
<p>How useful is the interactive medical history form?</p> <ul style="list-style-type: none"> - In reporting medical history - In creating a dialogue with the health provider - As a learning aid 	<ul style="list-style-type: none"> - Think about the interface you just used to report on the medical history forms. <ul style="list-style-type: none"> o Tell me how you used the interface? o Tell me about any information that you couldn't report?? - How did you find the body categories for reporting on the medical history? <ul style="list-style-type: none"> o Did the categories make sense? o Was there something confusing?
<p>How useful is the interactive medical history form?</p> <ul style="list-style-type: none"> - In reporting medical history - In creating a dialogue with the health provider - As a learning aid 	<ul style="list-style-type: none"> - What was the main purpose of the interface during the exploration? <ul style="list-style-type: none"> o How did the visualization help you in reporting the medical history? o How did the feedback make you feel? - Think about the interface with the body graphic. If this was a part of the conversation with your caregiver, did you feel heard? <ul style="list-style-type: none"> o How useful was the interface in helping to have a better communication? o Tell me about the aspects that you remember the most?
<p>How useful is the interactive medical history form?</p> <ul style="list-style-type: none"> - In reporting medical history - In creating a dialogue with the health provider - As a learning aid 	<ul style="list-style-type: none"> - Tell me about something you remember about the medical condition you had to report. - Tell me about something that you did not know before the session? - Try to remember about your last bio class. Do you remember which topics you covered? <ul style="list-style-type: none"> o Was something related to the interface you just explored?

Appendix J.

Coding Scheme: User Testing

Value of Visualization

Definition: Code as *Value of Visualization* any transcription excerpt in which the health consumer refers to the purpose of the visualization or the advantages it provides against traditional medical history forms.

Value of Visualization may include instances where the health consumer...

- (d) Shows advantages of the interface and/or how it was useful or helpful
- (e) Talks about aspects of the interface/visualization/prototype that helped them to better understand the conditions/illnesses
- (f) Explains the purpose of the interface/visualization/prototype in reporting their medical history

Interface Limitations

Definition: Code as *Interface Limitations* any transcription excerpt in which the health consumer talks about aspects of the interface/visualization that made it hard or confusing to use.

Interface Limitations may include instances where the health consumer...

- (d) Talks about aspects of the medical history which they couldn't report
- (e) Explains instances in which the interface/visualization caused them confusion and/or didn't make sense
- (f) Shows to what extent they did not use the whole system (i.e. not using the visualization, the categories or the buttons).

Interaction

Definition: Code as *Interaction* any transcription excerpt in which the health consumer refers how they used the interface/visualization/prototype

Interaction may include instances where the health consumer...

- (a) Explains how they interacted with the interface/visualization/ prototype to report on the medical history
- (b) Talks about their exploration process and/or how they used the interface

Health Literacy

Definition: Code as *Health Literacy* any transcription excerpt in which the health consumer refers to their own knowledge about the health and body.

Health Literacy may include instances where the health consumer...

- (d) Talks about the shortcomings of their own medical literacy and/or if they have any formal training in medicine/biology/etc.
- (e) Explains how they understand their own body and disease in relation to the reported medical history

Health consumer Experiences

Definition: Code as *Health consumer Experiences* any transcription excerpt in which the health consumer refers to their own experiences with doctors and/or healthcare

Health consumer Experiences may include instances where the health consumer...

- (c) Talks their own experiences reporting their health conditions to doctors/caregivers
- (d) Talks about the experiences of friends or family when dealing with healthcare

Interface Recommendations

Definition: Code as *Interface Recommendations* any transcription excerpt in which the health consumer provides concrete recommendations for the interface/visualization.

Interface Recommendations may include instances where the health consumer...

- (c) Talks about aspects that would be good to have in the interface/visualization
- (d) Suggests features to include in the interface/visualization