Politecnico di Milano
School of Design
Design & Engineering Master of Science

Design and Data Science Interface
in User and Context Research for Product Innovation

Academic Supervisor
Barbara Del Curto

Academic Co-Supervisor
Caterina Dastoli

Graduating Student
Audrey Saleh
Matricola 853484

Academic year: 2017-2018

Table of Contents

List of figures

List of tables

Abstract

Design methodologies for innovation have a user-centered background: design processes are starting with research on user and its context. Most of the tools used for this research have a qualitative approach, partly explained by the empathy required for this phase. The emergence of data science is an opportunity for designers to co-develop more quantitative methods for user research. The aim of the thesis is to present the diverse and new methods coming from the design-data science interface.

Designers are exploring new methods of data collection, with pre-designed channels (retail, sensors) or their own channels (data-product-data framework). Moreover, text mining offers great possibilities to scale up qualitative methods like netnography or lead user innovation. Big qual studies (the assemblage of previous qualitative studies and their secondary use) are also benefiting from text mining to ease the research in large data sets.

Implementation of these new methodologies are presented in three case studies: the beauty start-up Glossier and the leverage of an internet-based community for product development; the design agency IDEO and the combination of interviews and big data from sensors or surveys to define and target specific user segments; LEGO and mPath using skin conductance sensors to study kids' emotional profile when they are playing with difficult games, to develop a new way to engage parents. Finally, these new methodologies require an ethical use of data and an appropriate corporate culture to produce successful results.

Keywords: product development, design methodology, data science, user research, quantitative methodologies

Le metodologie di design per l'innovazione sono fondate su principi user-centered: i processi iniziano con una ricerca sull'utente e il suo contesto. La maggior parte degli strumenti utilizzati per questa ricerca sono approcci qualitativi, in parte a causa dell'empatia necessaria in questa fase. L'emergere della scienza dei dati diventa un'opportunità per i designer di co-sviluppare metodi più quantitativi per la ricerca sull'utente. L'obiettivo della tesi è di presentare i nuovi e diversi metodi derivanti dall'interfaccia design-scienza dei dati.

I designer stanno esplorando nuovi metodi di raccolta dei dati, con canali pre-progettati (settore del commercio al dettaglio, sensori, ecc) o propri canali (come ad esempio il Data-Product-Data design pattern). Metodi come il text mining, offrono poi grandi possibilità di potenziare le metodologie qualitative, come la netnografia o la Lead-User Innovation. Per citare un esempio, studi relativi ai "big qual" (l'assemblaggio di precedenti studi qualitativi e il loro uso secondario) sono notevolmente migliorati grazie al metodo del text mining, che ha permesso di semplificare la ricerca di grandi serie di dati.

L'implementazione di queste nuove metodologie è presentata attraverso tre casi studio: la start-up di cosmesi Glossier e l'utilizzo di una comunità basata su Internet per lo sviluppo di prodotti; l'agenzia di design IDEO e la combinazione di interviste e big data (da sensori o sondaggi) per individuare la segmentazione dell'utente; LEGO e mPath che utilizzano sensori di conduttanza cutanea per studiare il profilo emotivo dei bambini quando giocano con giochi difficili, per sviluppare un nuovo modo per coinvolgere i genitori. Infine, queste nuove metodologie richiedono un uso etico dei dati e un'adeguata cultura aziendale per ottenere risultati positivi.

Parole chiave: sviluppo prodotti, metodologia di design, scienza dei dati, ricerca sull'utente, metodologie quantitative

Introduction

Before writing this thesis, I had the opportunity to work as an intern in two diametrically opposed companies, at two clearly distinct positions. The first was a project manager role for innovative product in a cosmetics corporation. The mission was to pilot engineering, design, marketing and purchasing teams to foster and accelerate innovation development. I saw managers questioning the utility of at-home interviews with user, rejecting ethnographic research projects because of costs, without direct tangible profits, doubting about designers' proposals. The second work experience was a data analyst position at a Web giant firm with a strong data-driven culture. I discovered a new mantra there: if there is a problem, there are always data to justify and solve it.

The thesis subject came logically after these two experiences: can data science be a supportive tool for designers before concept generation? Does the introduction of data science methodology in designers' research phase reinforce their observations and give more weight to their future proposals? Which forms of design-data science interface exist when the two fields are paired for innovation development?

What is the landscape of methodologies combining design and data science to learn more about the user and its context, in the case of product innovation?

To answer these questions, Chapter 1 is dedicated to the place of design in innovation process and proposes an overview of usual methodologies implemented by designers to expand their knowledge on user and context. This chapter is the occasion to present two underlying themes of the thesis: the importance of empathy and the dichotomy between qualitative/quantitative methodologies.

Chapter 2 is first an introduction to the data science field, the basic principles and organization. Then, it presents three examples of successful designers – data scientists' collaboration (outside the specific scope of user and context research): data visualization, A/B/n testing and digital development.

After these two chapters setting the scenes of design and data science, Chapter 3 is a compilation of methodologies which designers can use by leveraging data science capabilities to get closer to the user: different channels of quantitative data collection, text mining for large scale ethnographic research… The chapter ends with a clear distinction between the marketing and the design purpose of these "data-science infused" methodologies.

To put into practice the methodologies described in the previous chapter, Chapter 4 is the occasion to present three cases studies of companies experimenting the integration of design-data science approaches to learn more about their user: Glossier, IDEO and the collaboration LEGO-mPath.

Finally, Chapter 5 concludes on a (non-exhaustive) overview of challenges that could restrain the implementation of design-data science methodologies.

Chapter 1. Innovation through design: process and tools

1. Development of design-based innovation approach

1.1. Design incursion in innovation field

Brief history of innovation: the tech roots

Innovation has been on everyone's lips since the beginning of the 21st century. If innovation is traditionally linked to the corporate world (to be more competitive), governments of Northern countries took over the concept and present it as a strategic lever for growth and development. Initiatives to promote commercial and social innovations are flourishing, notably in Europe with the world largest innovation program launched in 2014 (EC, 2011).

The very first manifestation of the concept of innovation comes from Ancient Greece as "introducing change into the established order" (Godin, 2015). However, the modern vision of innovation defined as a certain implementation of technology started with Adam Smith in 1760s', describing the productivity gained from the use of machinery in agriculture (Smith, 1776). But the major contributor is the Austrian economist Schumpeter, the "Godfather of Innovation" (The Economist, 1999), who developed a definition of healthy economy as an economy cyclically disrupted by technological innovations. Thus, each industrial revolution is associated to a series of technologic discoveries, seen as the driving force of all innovations waves, and as a result, economic prosperity (Schumpeter, 1939).

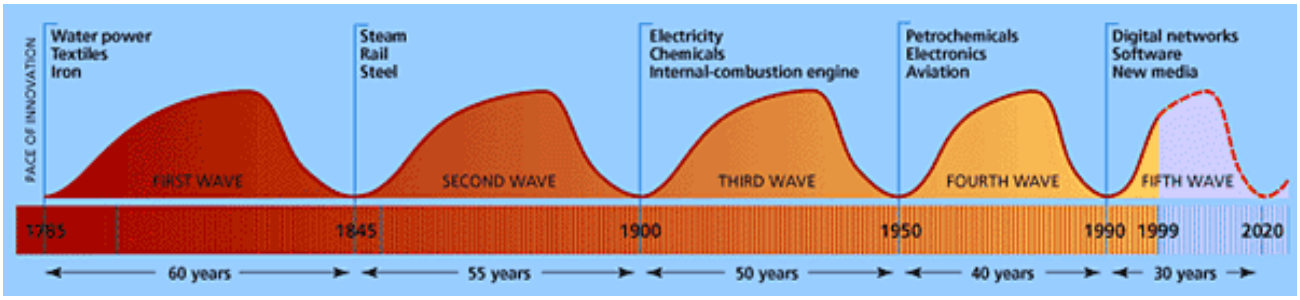
This perspective of innovation's technological roots is now challenged by historians showing that the fertile ground of innovation for the firsts industrial revolutions was a combination of numerous parameters, not just a result of technology (Bruland & Mowery, 2006):

- Property rights development
- Weakening of political powers (between 14th and 17th century in Europe, before industrial revolutions)
- Accumulation of knowledge and ease of knowledge circulation
- Creation of financial organization form like joint-stock organization

Defining innovation

The different attempts to define innovation these past 20 years seem to follow the broadening of innovation scope. The Oslo Manual, a reference edited by the Organization for Economic Cooperation and Development (OCDE) defines it as "the implementation of a new or significantly improved product (good or service) or process, a new marketing method, or a new organization in business practices, workplace, organization, or external relations." (OCDE, 2005, pp.46). As noticed by L. Cruickshank, it is

Figure 1.1: Schumpeter's innovation waves and technology associated (source: The Economist, https://www.economist.com/node/186628)



interesting to see that the word "technological" has been removed from the definition compared to the 1995 version (Cruickshank, 2010).

Moving out of a purely tech-driven definition of innovation is a first step, but several thinkers from 1980's, chose to go one step further and associate it with design. Tim Brown, a Design Thinking theorist, embodies this path:

> *"A purely technocentric view of innovation is less sustainable now than ever, and a management philosophy based only on selecting from existing strategies is likely to be overwhelmed by new developments at home or abroad. What we need are new choices—-new products that balance the needs of individuals and of society as a whole; new ideas that tackle the global challenges of health, poverty, and education; new strategies that result in differences that matter and a sense of purpose that engages everyone affected by them." (Brown, 2009)*

Design and Innovation: parallel development

This study does not try to define design. Nowadays, design is covering a multiplicity of fields, works, visions, and it is hard to find a definition satisfying all parties. The thesis will be focused on design as a creative process, aligned with the World Design Organization (ex-ICSID) definition: "Design is a strategic problem-solving process that drives innovation, builds business success, and leads to a better quality of life through innovative products, systems, services, and experiences." (WDO, 2018).

Thus, design definition is already containing the concept of innovation, which is not surprising given that the two concepts grew in parallel during the 20th century; both innovation and design deal with (Kahane, 2015):

- Production of newness and adoption of the novelty
- An equilibrium between rational and emotional

Factors of design innovation expansion

Several factors allowed design to take a place in the innovation landscape over the past 50 years:

- the Internet/App/Digital explosion gave birth to numerous start-ups having the possibility to develop a good digital product at low cost, compared to more mature markets where only strong and established companies have the financial means to invest in R&D to produce innovation. Consequently, pure technology was not enough for differentiation, and design became naturally the element dividing successful products and the others (Falguni, 2016).
- Moreover, alongside the digital craze, usability gained importance, first for digital products (e.g. the number of clicks necessary to perform an action being a staple in this field), and infused in the other product categories, putting designers at the center of these questions (Falguni, 2016)
- The development of a community of design thinkers, with some charismatic personalities like Don Norman, Harold G. Nelson and Tim Brown to advocate for the position of design as something more than "making things pretty" (Norman, 2010). Design found a new legitimacy, and designers were involved earlier and earlier in the innovation process, not anymore just at the final step like "the icing on the cake" but as a core element of the process. (Vial, 2010)

To conclude this brief historical recap, it cannot be denied that design concepts infused the innovation sphere, and conversely. If technology was a prerequisite for innovation the last centuries, this vision is challenged, and designers opened new innovation opportunities in terms of process, usage, meaning… (see following sections). Thus, this thesis is an attempt to show how design innovation[1] can use technology (here: data science) as a tool, and not only as an innovation vector. To see this phenomenon, it is necessary to have a certain understanding of design innovation processes commonly used, to be able to pinpoint later where data science can be relevant.

[1] Design innovation refers to "innovation led with a design methodology, design perspective" here. Design innovation has multiple definitions as seen by Mutlu & ER, 2011.

1.2. Overview of design innovation methodologies

Over the last 70 years, different approaches for innovation have been developed, with an upsurge observed from 2000's (simultaneously with the craze for both design and innovation). The main ones are described in the following table (Table 1.1). If some of them have strong roots in the tech field (Agile, Lean), most of them come from academicians with design background (Rolf Faste for Design Thinking, Roberto Verganti for Design-driven Innovation, Jake Knapp for Design Sprints).

Table 1.1: Main innovation processes overview (adapted from: Owyang, 2017)

Name	Founded	Differentiator	Type
Waterfall	1956 by H.Benington	Teams work independently on each stage	Traditional tech based process
User – centered Design	**1986 by D. Norman**	**End-user's needs, wants and limitations are the focus at all stages**	**Design-based process**
Design Thinking	**1980's by Stanford lab**	**Forces exploration of ideas beyond the familiar**	**Design-based process**
Agile	2001 by writers of Agile Manifesto	Can quickly and easily adapt to project changes	Tech-based process
Lean Startup	2008 by E.Ries	Low investment to test the market	Tech-based process
Design-driven Innovation	**2009 by R.Verganti**	**Innovation of meaning by leveraging "interpreters"**	**Design-based process**
Design Sprints	**2010 by J. Knapp**	**Produces a tested prototype in just one week**	**Design-based process**

User-Centered Design: the pioneer

It is interesting to note that User-centered Design (UCD), the first methodology taking inspiration from design, can be considered as the matrix of all the other design innovation methodologies e.g. Design Thinking would be a specific implementation or a "re-packaging" of UCD (WUC, 2014). All design approaches share the main principles of UCD:

- **user focus (via the user itself or mediators):** to early guide the development through activity goals, context of use, users' needs and tasks
- **early prototyping**: to embody and evaluate continuously design proposals, to confront them to end users' quickly
- **evaluation of use in context**: to ensure adequacy between usability and design
- **evolutionary system development**: to implement an iterative and incremental development process
- **professional and complementary team**: to lead the process thanks to effective multidisciplinary team
  (Gulliksen & Göransson, 2003)

However, even if the basic principles of the methodologies are relatively similar, they all have been developed in different directions, depending on the original industries having recourse to them, the founders' and leaders' background, and the practical tools promoted around the methods (Table 1.2). Thus, the approaches will be considered as different in the following paragraphs. The crucial diverging points were explored in several articles: UCD and Design Thinking (Knemeyer, 2015), UCD and Design-driven innovation (Norman & Verganti, 2014), Design Thinking and Design Sprints (Claes, 2017).

Table 1.2: Areas of specific development for design innovation methodologies

Methods	Emblematic fields[2]	Specificity
UCD	Goods, Interface, Hardware/Software	Contained the theoretical base of the other methods
Design Thinking	Goods, Services, Social concerns	Linked to the biggest practical toolbox, notably developed by IDEO (IDEO, 2015)
Design-driven Innovation	Service, Goods, Tech industry	Not related to users themselves, but on experts able to interpret users' needs and wants
Design Sprints	Start-ups, Digital, Tech industry	Highly related to a precise chronology (only five days)

[2] All methodologies are supposed to fit universally to every fields, but some methods are more widely spread in certain industries

1.3. A common starting point: knowing the user and the context

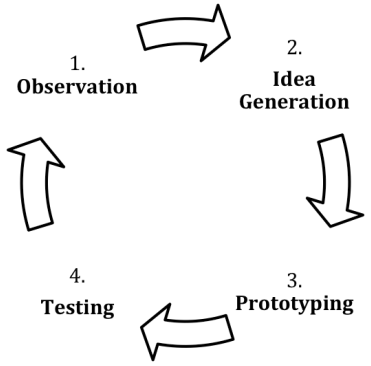
The previous design innovation methodologies panorama shows a development of different approaches. Each approach is characterized by a process divided in several steps. For the study, the focus is henceforth on the first step of the processes, and great similarities can be highlighted.

User-centered Design

As described by Don Norman himself, UCD is a 4-step process:

> *"Make observations on the intended target population, generate ideas, produce prototypes and test them. Repeat until satisfied" (Norman, 1998)*

Figure 1.2: Spiral UCD Process (source: Bell, 2016)[3]

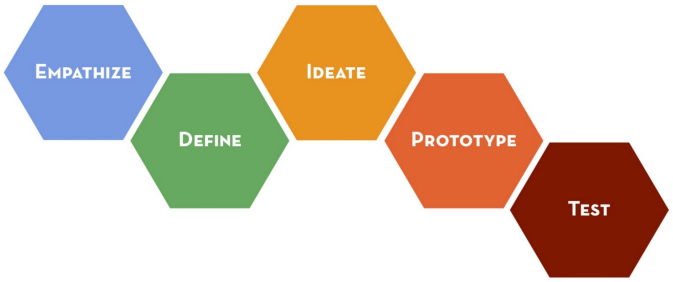


The first step, Observation, is made for: "the deep understanding of the goals the people are trying to accomplish and the impediments they experience" (Norman, 1998). This quest of users' knowledge is supported by the methods of applied ethnography. The emphasis is on direct observation and detailed analysis of targeted group (no early conclusion deriving from other groups studied). The contribution from marketing research (i.e. quantitative and generalist) is not denied, but the value of the Observation step comes from a small number of people deeply analyzed. The tools used are interviews, shadowing and self-recording - like ethnography (Sanders & Dandavate, 1999).

Design Thinking

The classical definition of Design Thinking process is in 5 steps, as defined by d.school (Stanford).

Figure 1.3: Design Thinking Process (source: d.school, 2010)



The goal of the first step, Empathize, is to gain an empathic understanding of problem at question. As for UCD, it involves a lot of observation, but not only. Additionally: active immersion is used to gain insight: by becoming the user, the designers perform the activity and have a privileged position to assess the problem. Interviews, shadowing and self-recording are again relevant, but more exotic tools flourished, led by the work of the creative agency IDEO and their Methods Cards (Collage, Peer observing peers, Card sort) (IDEO, 2015).

Design-driven Innovation

The innovation process for this method is focused on the preliminary work before concept generation: the design-driven research, which leads to a new meaning definition.

Figure 1.4: Design-driven Innovation Process (adapted from: Verganti, 2009)

In this method, the first step is slightly different: the aim is not to study users directly, but to leverage the deep knowledge of experts who are dedicating their work life to the study of the targeted users. This alternative path comes from Verganti's conclusions that users are only able to guide designers through incremental innovations, not radical ones (a basic user does not know that he will want something which has yet to be invented). Thus, designers

[3] As for the three other methods, the graphic illustrating the process are chosen to be consistent with the first graphic representation supported by the founders of the methodologies, the evolution of the style may be meaningful at some extent.

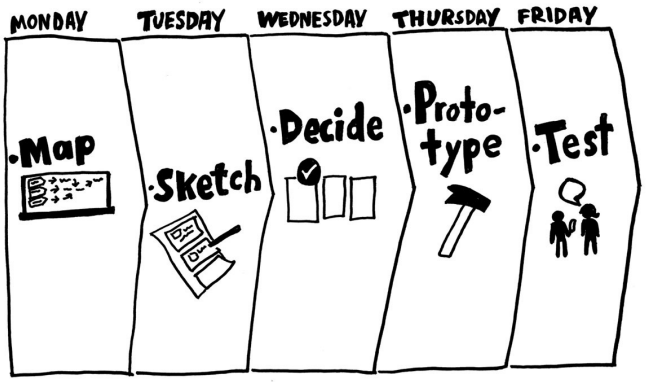
need the point of view of "professional thinkers": the interpreters (Verganti, 2009).

The challenge here is not in the methods of user and context's observations, but to find the interpreters that have already put in place users' observations and immersions tools. And the core goal of step one is to gather all the interpreters' knowledge and capitalize on them to define a new meaning.

Design Sprint

The Design Sprint process is similar to Design Thinking in terms of steps, but the particularity of design sprint is the chronology: the complete process should be done in 5 days (one step a day).

Figure 1.5: Design Sprint process (source: Knapp & Zeratsky, 2015)



The first step, Map (also called Understand cf. Google Design Sprint Tool Kit, 2017), aims at "exploring the business problem from all angles". As timing is brief (one day only), this first step is not described as a true user and context analysis, but it intends to gather experts point of view (similarly to Design-driven innovation), existing user researches and competitor audit. The result should be a synthesis of the different elements, to have a strong knowledge on user needs. Tools implemented are not distinct from the other methods: user interviews, user journey mapping, expert talks.

Synthesis: User research is the baseline of design innovation

In the four methods, the foundation stone of the process is user and context research. As pointed by Kujala (2003), user research does not have a precise definition, and can range from a general focus on user to a will of making users participate and actively take position during the design process. The user research has several forms, taking inspiration from ethnography, contextual research, usability testing, marketing.

There is no surprise that the design innovation processes rely on user and context analysis, because the positive effects of it has been proven the last 20 years, through qualitative and quantitative researches (Kujala, 2003). The benefits highlighted are:

- An increase in product quality thanks to more accurate user requirements
- A global reduction of product features thanks to a focus on what user truly needs/wants
- An improvement in product acceptance
- A more performant use of the product thanks to a better user's understanding of the product (Damodaran, 1996)

Even if user research and context has a price, once well-executed, it leads to higher sales and reduced costs in training and user supports. More generally, as summarized by Gould et al. (1987): "Extra effort in the early stages leads to much less effort later on and a good system at the end"[4].

Finally, the importance to start with user and context analysis is coherent from both design and business point of view. Thus, this first step of design innovation process is the focus of the following chapters of the thesis. There is already an extensive literature on user and context analysis, benefits and obstacles, how-to guide, methods and tools comparison: one remark emerging from this literature review is the natural predominance of qualitative research in user and context analysis phase. To which extent is this statement verified and what can we learn from it?

[4] It refers primarily to software development, but could be easily extended to goods.

2. Designers' tools for user and context exploration

2.1. Defining tools classification

Research methods spread in design community

User and context analysis is the foundation of design innovation process. That being said, the natural interrogation is the feasibility and the strategy implemented to perform user research. Leading research (in general, not only focused on users) is a common process for designers: researching for material, looking for historical development, search of design language… The four main types of researches used by designers are:

- Qualitative research: methods to capture people's thoughts, feelings and interpretations of what they are experimenting
- Quantitative research: methods to collect objective data, generally at larger scale than qualitative research, in order to describe, simplify or generalize a situation
- Visual research: methods to examine existing images and objects (medias and material) to find patterns and meanings
- Applied research: methods to investigate design practice in itself
  (Muratovski, 2016)

The framework of the study is user and context analysis, so the focus will be on qualitative and quantitative researches (considering that applied research embeds qualitative and quantitative methods). Thus, designers can leverage both qualitative and quantitative research for the first stage of the design innovation process. Moreover, methodological triangulation (combining different research methods) is recognized as a powerful instrument to enrich the user analysis (Thurmond, 2001). Sharing the same objective, "to understand and explain behavior and events, their components, antecedents, corollaries, and consequences" (Dzurec & Abraham, 1993), quantitative and qualitative studies have enough in common to be mixed, which allows a fuller understanding: qualitative input helps explaining exceptions or particularities standing out from global quantitative data for example. It reveals distorting points or reinforces aspects of the phenomena. Therefore, designers must have a range of qualitative and quantitative tools to nourish their primary user and context analysis.

Qualitative vs Quantitative research: a questionable distinction?

The previous definitions of qualitative and quantitative research are voluntary vague. The distinction between both, and the relevance of the distinction has been discussed in literature for the last 30 years (Bryman, 1984, Hammersley, 1992, Allwood, 2012). Indeed, the frontier between qualitative and quantitative research is multifaceted, opposing pairs of concepts but moving the cursor of what is acceptable in one method or not depending on the context. A quick example is the acceptability of quantification in qualitative methods: some references banished it (Corbin & Strauss, 1998), whereas diary method, commonly classified as a qualitative method, can be focused on "quantitative question" (e.g. for a drug test, the user may document the hours of medication intake, the time required before noticing effects). This is why the criteria used later to distinguish qualitative and quantitative methods are adapted from Johnson & Christensen (2017) (and not fully aligned), to fit the framework of user and context analysis for first phase of design innovation process.

Criteria to qualify user research tools as qualitative or quantitative

A classical criterion to divide qualitative or quantitative methods is the scientific approach: top-down (the aim is to test hypotheses and verify them with data) for quantitative research and bottom-up (the goal is to build or generate knowledge and hypotheses from the data collected) (Johnson & Christensen, 2017). This opposition cannot be relevant for the framework of the study because the user and context analysis phase is intrinsically divergent (Brown, 2008): the purpose is to widen the opportunities, go further, discover insights. It would mean designers do not have the possibility to use quantitative tools for user and context analysis, whereas surveys, a quantitative method, is extensively spread for user exploration.

More generally, the criteria selected (Table 1.3) for the rest of the study are more focused on the nature of specific parts of the research process (e.g. data collection process, nature of data) because the global philosophy of the researches are identical: gaining relevant knowledge on the user, his context, his problematics, to feed the future concept generation. The case of mixed research is not explicitly described as it is considered as an in-between cursor for all criteria.

Table 1.3: Emphases of Qualitative and Quantitative research methods (adapted from: Johnson & Christensen, 2017)

Criterion	Quantitative research	Qualitative research
Ontology (i.e. nature of reality/truth)	Objective, material, structural, agreed-upon	Subjective, mental, personal and constructed
Common research objectives	Quantitative/numerical description, causal explanation	Qualitative/subjective description, empathetic understanding
Nature of observation	Isolate the causal effect of single variables	Attempt to understand insiders' views, meanings and perspective
Forms of data collected	Collect quantitative data based on precise measurement	Collect qualitative data such as in depth interviews, participant observations, field notes, and open-ended questions.
Nature of data collection instrument	Structured and validated data-collection instruments	The researcher is the primary data-collection instrument
Nature of data	Variables	Words, images, categories
Results	Generalizable findings providing representation of objective outsider viewpoint of populations	Particularistic findings, provision of insider viewpoints

2.2. Classifying the most common user research tools for designers

The corpus

To appreciate the range of tools available to designers to lead the user and context analysis, the study presented below examined the different methods listed in a corpus of famous "how-to" design innovation process guide. The selection of the corpus is based on Ricardo Martins (2017) article, analyzing the split of tools listed in these guides according to their relation to the process steps (user and context analysis, insights framing, concept generation, test and prototype, implementation). An interesting outcome in the perspective of the present study is the consequent share of user and context analysis tools identified in the corpus, 33% (145 methods on a total of 446 for the five process steps). The appetite for user and context analysis research method is noticeable.

The corpus is composed of:

- The Field Guide to Human-Centered Design, IDEO (2015), section "Inspiration"
- 101 Design Methods, Kumar (2012), sections "Know context" and "Know people"
- Design methods: 200 ways to apply design thinking, Curedale (2013), section "Know people and context"
- Service Design Practical access to an evolving field, Moritz (2009), section "Understanding"
- Research for designers, Muratovski (2016), sections "Qualitative research" and "Quantitative research"

The methodology

The 148 user and context analysis tools listed in the 5 guides have been assessed and classified through qualitative or quantitative typology in accordance with the criteria in Table 1.3. The category of mixed qualitative-quantitative method has not been considered because it could be possible to class the majority of the tools as mixed methods, depending on the final intentions of the designers: this choice forces a clear distinction. To waive any doubt, the two predominant criteria were: the will to look for an individual perception or a generalizable insight and the objectivity of data collected. Some tools appeared several times in the corpus (like surveys, focus group), the duplication is kept because the goal is to appreciate the occurrence of qualitative or quantitative methods in design literature, not the proper variety of tools.

An extract of the classification is presented on the next page (Table 1.4). The complete listing is in the Annex.

Table 1.4: Extract of listing and distinction of user and context analysis tools (complete listing Annex 1)

Corpus	Qualitative Research	Quantitative Research
The Field Guide to HCD, IDEO	Interviews (p.39)	
The Field Guide to HCD, IDEO	Group Interview (p.40)	
The Field Guide to HCD, IDEO	Conversation Starters (p.45)	
The Field Guide to HCD, IDEO	Extreme and Mainstreams (p.49)	
The Field Guide to HCD, IDEO	Immersion (p.52)	
101 Design Methods, Kumar	Buzz Reports (p.22)	
101 Design Methods, Kumar		Keyword Bibliometrics (p.32)
101 Design Methods, Kumar	Five Human Factors (p.102)	
101 Design Methods, Kumar	Ethnographic Interview (p.110)	
101 Design Methods, Kumar	User Pictures Interview (p.112)	
101 Design Methods, Kumar	Cultural Artifacts (p.114)	
101 Design Methods, Kumar	Image Sorting (p.116)	
Design Methods, Curedale	Behavorial Map (p.114)	
Design Methods, Curedale		Benchmarking (p.115)
Design Methods, Curedale	Bodystorming (p.119)	
Design Methods, Curedale	Boundary shifting (p.120)	
Design Methods, Curedale	Camera Journal (p.121)	
Design Methods, Curedale	Cultural probes (p.138)	
Design Methods, Curedale	Interview: conservative cards (p.193)	
Design Methods, Curedale	Shadowing (p.232)	
Design Methods, Curedale	Storyboards (p.236)	
Design Methods, Curedale		Eye tracking (p.238)
Design Methods, Curedale	Talk-out loud protocol (p.241)	
Service Design Practical, Moritz		Benchmarking (p.186)
Service Design Practical, Moritz		Client segmentation (p.186)
Service Design Practical, Moritz	Context analysis (p.187)	
Service Design Practical, Moritz	Contextual interviews (p.187)	
Service Design Practical, Moritz	Contextual inquiry (p.188)	
Service Design Practical, Moritz	Ethnography (p.189)	
Service Design Practical, Moritz	Experience test (p.190)	
Research for Designers, Muratovski	Ethnographic research (p.110) – Structured interviews	
Research for Designers, Muratovski	Ethnographic research (p.110) – Semi-structured interviews	
Research for Designers, Muratovski	Ethnographic research (p.110) – Focus group	
Research for Designers, Muratovski	Ethnographic research (p.110) – Oral History	
Research for Designers, Muratovski	Phenomenology (p.141) – In-depth interviews	
Research for Designers, Muratovski	Historical research (p.155)	
Research for Designers, Muratovski	Grounded theory (p.163)	
Research for Designers, Muratovski		Surveys (p.177)

2.3. Learnings

Qualitative research predominance

The first major result is the qualitative / quantitative split. There is a clear focus on the qualitative tools, in terms of occurrence (148 vs 36). Moreover, another hint shows the deeper study of qualitative methods, which are more granularly defined than quantitative ones. A good example is the distinction between interviews and surveys: unanimously classified as a qualitative tool for the first, a quantitative one for the second. Where survey method is presented as one tool (Curedale, 2013, Moritz 2009, Muratovski, 2016), interview is declined as several differentiated tools (IDEO, 2015, Curedale, 2013, Muratovski, 2016): structured, semi-structured, grouped, guided story-telling, contextual inquiry… up to 15 separated tools in Curedale (2013). However, a panel of survey methods exists: written survey, verbal survey, open-ended or closed-ended surveys, evaluative continua survey (with scale) (Glasgow, 2005), but they are not considered in the corpus as individual tools. If some authors mentioned particularities of surveys (the concepts of open-ended and closed-ended questions for Muratovski, 2016), the explanations are lighter than the ones for interview variety. There is a global tendency to highlight qualitative methods in design innovation process literature.

Typology of quantitative tools

The second finding is the nature of quantitative tools:

- A quarter of the quantitative methods presented are imported from marketing (SWOT analysis, market segmentation, benchmarking)
- A fifth of the methods are related to expert interviews, reading experts/academic paper[5]
- 14% of the methods are surveys or questionnaires
- 14% are the development of flow mapping: charts dissecting all the steps composing a phenomena, applying variables to "measure" parameters qualifying the step, (e.g. time allocated, resource allocated, satisfaction). This could be service blueprint, resource flow. The purpose is to have a clear and quantified picture of a phenomena to point out strengths and weaknesses.

The last category, flow mapping, is the only one that came directly from the design field, as a graphical interpretation of flow and process study adopted by designers. The other categories (60% of all quantitative tools) are borrowed from other disciplines. This is not problematic; this is also the case for qualitative studies, using ethnographic or phenomenological methods. But it can be interpreted as an opportunity space for quantitative methods developed by designers themselves.

The distortion between the number of occurrence of qualitative and quantitative methods implies less possibilities for designers to triangulate methods. Besides, a potential explanation of the qualitative methods over-representation may be correlated to a deeper designers' appetence for qualitative methods. How can qualitative method predominance be explained by the core principles of design?

Table 1.5: Results of the user and context analysis methods

Tools total	Qualitative tools total / Tools total	Quantitative tools total / Tools total	Typology of quantitative tools / Quantitative tools total			
			Marketing	Experts point of view	Survey/ Questionnaire	Flow mapping/ Step analysis
184	148	36	10	7	5	5
100 %	80,4 %	19,6 %	27,8 %	19,4 %	13,9 %	13,9 %

[5] These methodologies may be considered as qualitative in another context. They are labelled quantitative here because they come from a will of gaining a general picture of the phenomena, supported by people building a knowledge base for years, who led research. The interest is not thoughts and feelings of experts/researchers, but their rational point of view acquired.

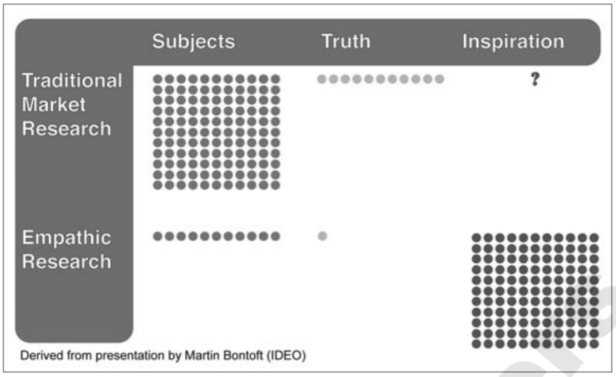
3. Empathy, designers' cornerstone

3.1. Empathy, essential corollary to design innovation process

With the transition from technocentric to design-driven innovation, a new parameter appeared at the core of the innovation process, besides R&D, feasibility, production, finance, distribution: the user. And with the tech-design translation, products are no longer only functional but also "supra-functional" (McDonagh & Joyce, 2010), including emotional, cultural, social and aspirational aspects[6]. To explore these new facets, designers leverage a new ability: empathy.

Empathy is defined as "the intuitive ability to identify with other people's thoughts and feelings – their motivations, emotional and mental models, values, priorities, preferences, and inner conflicts" (McDonagh, 2006). Empathy is the modality used by designers to collect users' data, opposed to traditional user data collection focused on large samples, search of generalizable truth, to finally deliver dehumanized data, missing all the supra-functional potentiality (McGinley & Donglong, 2011).

Figure 1.6: Empathic research for inspiration
(source: McGinley & Donglong, 2011)



Empathy implies a "direct and personal engagement and is dependent on the designer's willingness" (Battarbee, 2004). This is essential to surpass the classical "empathic horizon" of thirty-something Caucasian male (the archetype of the designers according to statistics, McGinley & Donglong, 2011) and to truly resonates with the vast range of user possibly targeted.

3.2. Empathy rhymes with qualitative researches?

Empathic design process is a four-step practice (Table 1.5)

Table 1.6: Four phases of empathy
(source: Kroupie & Sleeswijk Visser, 2009)

1. Discovery	Entering the user's world Achieve willingness
2. Immersion	Wandering around in the user's world Taking user's point of reference
3. Connection	Resonating with the user Achieve emotional resonance and find meanings
4. Detachment	Leaving the user's world Design with user perspectives

There is no concern about the relation between the four phases and qualitative methods (subjective bias, particularistic finding, researcher as the instrument of data collection). However, the immersion and connection phases can seem antithetical in the context of quantitative studies. "Immersion" and "Connection" are the opportunity to wander around and be surprised, detached from prejudices. But quantitative studies gain interest thanks to the strength of structured data collection, which needs to be in place at the very beginning of the process, as a core element: surprise is avoided. Moreover, quantitative studies often necessitate the intervention of data collection specialists (e.g. statisticians, ergonomists): designers are put aside from this step, with no immersion or connection chances (or only through the formatted data delivered from the specialists). This distortion creates an empathy gap between users and designers, explaining partly, the under-representation of quantitative tools. They are not intrinsically un-empathic, but they need an extra care to fit relevantly the design innovation process.

As stated through the following chapters, there are new opportunities to develop quantitative research methods for design, thanks to data science notably. Nevertheless, it would be easy to lose sight of the empathy challenge: these new methods should always "bring human information to life […] (present) user insights as fuller stories, (convey) liveliness through virtual material, and (give) scope for the design audience to complete the interpretations" (McGinley & Donglong).

[6] It does not mean that products were not vectors of some supra-functional aspects before design innovation process, but the conceptualization and systemic view of these components came with design.

References

Allwood, C. M. (2012). The distinction between qualitative and quantitative research methods is problematic. Quality {\&} Quantity, 46(5), 1417-1429.

Battarbee, K. (2004). Co-experience: Understanding user experiences in interaction. Aalto University; Aalto-yliopisto). Publication Series of the University of Art and Design Helsinki; A 51, 103 + app. 117. Retrieved from Aaltodoc database.

Bell, A. (2016). MaKey MaKey instrument invention part III: Constrain-to-create. Retrieved 06/01, 2018, from http://www.adampatrickbell.com/blog/makey-makey-instrument-invention-part-iii-constrain-to-create

Brown, T. (2008). What does design thinking feel like ? Retrieved 06/02, 2018, from https://designthinking.ideo.com/?p=51

Brown, T. (2009). Change by design (Harper Business ed.).

Bruland, K., & Mowery, D. C. (2006). Innovation through time . The Oxford Handbook of Innovation.

Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? The British Journal of Sociology, 35(1), 75-92.

Claes, G. (2017). When, which … design thinking, lean, design sprint, Agile?. Retrieved 06/01, 2018, from https://blog.usejournal.com/when-which-design-thinking-lean-design-sprint-agile-a4614fa778b9

Corbin, J., & Strauss, A. (2018). Basics of qualitative research (3rd ed.): Techniques and procedures for developing grounded theory. Thousand Oaks, California.

Cruickshank, L. (2010). The innovation dimension: Designing in a broader context. Design Issues, 26(2), 17-26.

Curedale, R.,. (2013). Design methods 1: 200 ways to apply design thinking. Topanga: Design Community College.

d.school. (2010). An introduction to design thinking. Retrieved 06/31, 2018, from https://dschool-old.stanford.edu/sandbox/groups/designresources/wiki/36873/attachments/74b3d/ModeGuideBOOTCAMP2010L.pdf

Daae, J., & Boks, C. (2015). A classification of user research methods for design for sustainable behaviour doi:https://doi.org/10.1016/j.jclepro.2014.04.056

Damodaran, L. (1996). User involvement in the systems design process-a practical guide for users. Behaviour & Information Technology, 15(6), 363-377.

Dzurec, L. C., & Abraham, I. L. (1993). The nature of inquiry: Linking quantitative and qualitative research. Advances in Nursing Science, 16(1,), 73-79.

European Commission. (2011). Horizon 2020 - the framework programme for research and innovation. Retrieved 04/25, 2018, from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011DC0808&from=EN

Falguni, D. (2016). How design became the darling of innovation. Retrieved 27/04, 2018, from https://www.forbes.com/sites/falgunidesai/2016/05/27/how-design-became-the-darling-of-innovation/#5362e4525b26

Glasgow, P. (2005). Fundamentals of survey research methodology. Unpublished manuscript.

Godin, B. (2015). Innovation: A conceptual history of an anonymous concept. Unpublished manuscript.

Google. (2017). Design sprint tool kit, understand. Retrieved 06/03, 2018, from https://designsprintkit.withgoogle.com/methods/understand/

Gould, J. D., Boies, S. J., Levy, S., Richards, J. T., & Schoonard, J. (1987). The 1984 olympic message system: A test of behavioral principles of system design. Commun.ACM, 30(9), 758-769.

Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., & Cajander, Å. (2003). Key principles for user-centred systems design. Behaviour & Information Technology, 22(6), 397-409.

Hammersley, M. (1992). Deconstructing the qualitative-quantitative divide. In Julia Brannen (Ed.), Mixing methods: Qualitative and quantitative research (). Aldershot: Avebury.

IDEO (Firm). (2015). The field guide to human-centered design : Design kit

Johnson, Burke, Christensen, Larry B. (2017). Educational research : Quantitative, qualitative, and mixed approaches.

Kahane, B. (2015). Design et innovation : deux faces d'une même médaille ? Sciences du Design, 1(1), 109-119.

Knapp, Jake,, Zeratsky, John,Kowitz, Braden,. (2016). Sprint : How to solve big problems and test new ideas in just five days.

Knemeyer, D. (2015). Design thinking and UX: Two sides of the same coin. Interactions, 22(5), 66-68.

Kouprie, M., & Sleeswijk Visser, F. (2009). A framework for empathy in design: Stepping into and out of the user's life.20, 437-448.

Kujala, S. (2003). User involvement: A review of the benefits and challenges. Behaviour & Information Technology, 22(1), 1-16.

Kumar, V.,. (2012). 101 design methods : A structured approach for driving innovation in your organization.

Martins, R. (2017). Analysis of available design implementation methods. A study about scarcity of implementation methods. The Design Journal, 20, S4730-S4733.

McDonagh, D. C. (2006). Empathic research approaches: The support the designer: A supra-qualitative research for designing model. Design Issues.

McDonagh, D., & Thomas, J. (2010). Disability + relevant design: Empathic design strategies supporting more effective new product design outcomes. The Design Journal, 13(2,), 180-196.

Mcginley, C., & Donglong, H. (2011). Designing with information and empathy: Delivering human information to designers.14, 187-206.

Moritz, Stefan. (2009). Service design: Practical access to an evolving field. [S.I.]: Lulu.com.

Muratovski, G. (2016). Research for designers: A guide to methods and practice.

Mutlu, B., & Er, A. (2011). Design innovation: Historical and theoretical perspectives on product innovation by design.

Norman, D. (2010). Design thinking: A useful myth. Retrieved 04/27, 2018, from http://www.core77.com/posts/16790/design-thinking-a-useful-myth-16790

Norman, D. (1998). The design of everyday things. London: MIT.

Norman, D., & Verganti, R. (2014). Incremental and radical innovation: Design research vs. technology and meaning change.30, 78-96.

OCDE/Eurostat. (20). Oslo manual: Guidelines for collecting and interpreting innovation data (3rd Edition ed.). Paris: Editions OCDE.

Owyang, J. (2017). Your guide to design and innovation methods. Retrieved 06/02, 2018, from https://medium.com/@jowyang/your-guide-to-design-and-innovation-methods-b28ca60d2ed6

Sanders, E., & Dandavate, U. (1999). Design for experiencing: New tools.

Schumpeter, J. A. (1939). BUSINESS CYCLES. A theoritical, historical and statistical analysis of the capitalist process (McGraw-Hill Book Company ed.).

Smith, A. (1776). In CreateSpace Independent Publishing Platform (Ed.), The wealth of nations.

The Economist. (1999). Catch the wave. Retrieved 04/25, 2018, from https://www.economist.com/node/186628

Thurmond, V. A. (2001). The point of triangulation. Journal of Nursing Scholarship, 33(3), 253-258.

Verganti, R,. (2009). Design-driven innovation : Changing the rules of competition by radically innovating what things mean.

Vial, S. (2010). Court traité du design. Paris: Presses Universitaires de France.

Vijay, K., & Patrick, W. Faster, cheaper, deeper user research. Design Management Journal (Former Series), 14(2), 50-57.

World Design Organisation. (2018). Definition of industrial design. Retrieved 04/26, 2018, from http://wdo.org/about/definition/

World Usability Congress. (2014). Design thinking or user centered design? same? Retrieved 06/01, 2018, from https://worldusabilitycongress.com/design-thinking-a-new-way-for-a-user-centred-design-process/

Chapter 2. Data science: an open window for abundant collaborations

1. What is data science?

1.1. Defining data science

Data science is a semantically new concept, with the first major use of this denomination by William Cleveland in 2001 "Data Science: An action plan to expand the field of statistics". Ever since, the craze around the new field has exponentially grown, becoming a buzz word that reached a pinnacle with the Harvard Business Review (2012) article "Data Scientist: The Sexiest Job of the 21st Century". However, defining data science remains a challenge. Data science refers to a "set of fundamental principles that support and guide the principled extraction of information and knowledge from data" (Provost & Fawcett, 2013).

This definition, although general and conservative, is questioned by statisticians because they consider that the data collection, mining and interpretation implicated in the definition is already included in statistics[7]. Data science would only be a reshuffle of statistics applied to tech industry to a part of the stat community (Schutt & O'Neil, 2013). This conflict is an opportunity to highlight a key ability for data science compared to statistics: tackling business problems from a data point of view. Thus, to implement a data-centric solving methodology, data science becomes a combination and re-interpretation of existing fields, and that is how data science is characterized as a full field (Fry, 2004):

- **Computer science:** to acquire and parse data
- **Mathematics, statistics and data mining**: to filter and mine data
- **Graphic design:** to represent and refine data
- **Info visualization, Human-Computer Interaction:** to interact with data

The use of "science" is another controversial point: if you need to so explicitly call something a science, there is automatically room for doubt about the real *scientificness* of the field. As mentioned by Schutt & O'Neil (2013), data science may be more related to a craft (an activity requiring particular skills) than a science. But it does not lower intrinsically the relevance of the field. And the dichotomy craft/science can be explained by the industrial roots of data science, opposed to academic ones for the other sciences.

A functionalist approach

Indeed, data science was born in the Silicon Valley, with the initiative of several tech companies, Google ahead, to gather statisticians (like Rachel Schutt, who wrote *Doing Data Science* with Cathy O'Neil, quoted previously), computer scientists, engineers, physicists and social scientists in the same team (supported by data engineering teams) in order to develop insights for data-driven decisions, new approaches and explore causality. When other tech companies understood the potentiality of such teams, they started to post job offers entitled "Data scientists" in 2008 for the first time on LinkedIn; whereas Wikipedia page for data science came in 2012 and no academic institutions or degrees were mentioning data science at that time. Therefore, data scientists have been defined before data science itself, explaining the current functionalist approach to data science, and the semi blur around the definition.

Datafication

The fast rise of data science takes roots in the growth of the amount of data possibly collected since 2000. There is the "online" tip of the data collection iceberg: search engine historic, web shopping preferences, articles read, anything that people are mostly conscious of and see the

[7] Most of data science founders, in 2000's, had a statistician background, hence the interest of the divide

direct interest in their browsing habits. But the game changer is the ability to quantify offline behaviors. This is named "datafication" by Cukier & Mayer-Schoenberger (2013), and in this case, the awareness of the data collection is questionable:

> *"Datafication is a far broader activity [than digitalization]: taking all aspects of life and turning them into data. Google's augmented-reality glasses datafy the gaze. Twitter datafies stray thoughts. LinkedIn datafies professional networks" (Cukier & Mayer-Schoenberg, 2013)*

Datafication paved the way for new data developments by multiplying data sources. If data were already used to evaluate, benchmark, improve a product, a new category appeared: *data product*, built on data capacity itself, like friends' recommendations by Facebook, or personalized dynamic courses by education tech companies (Schutt & O'Neil, 2013).

Big data: Volume, Variety, Velocity

The thesis focuses on data science with design, not big data with design, because "the high-volume, high-velocity and high-variety information assets" (Gartner Inc.) is not necessary to gain relevant insights for designers: it can be done by applying data science process on limited data sets, as qualitative methods shown. However, big data gives opportunity to handle "heavy" data like videos, typically full of learnings, which were hard to handle before (Gandomi & Haider, 2015).

Thus, big data designates information assets characterized by the 3 V's (Ward & Barker, 2013). The first V, Volume refers to the increasing size of data sets. There are diverse attempts to size "how big" big data is (example in Schroeck et al., 2012, determining big data between terabytes and petabytes) but as Frank Buytendijk stated in a podcast (2014): "What is Big today, might be normal tomorrow". The elastic definition of big data volume from Schutt & O'Neil (2013) seems more reasonable: "Big is when you can't fit it on one machine" and so there is a need of different strategy to handle this size. The second V, Velocity, corresponds to speed at which data are generated and should be analyzed to take data-driven decisions or reinjected. The reaction capability is strategic and high rate is technically allowed thanks to big data. Finally, the last V is for Variety: "the structural heterogeneity in a dataset" (Gandomi & Haider, 2015). There is variety in the format of data:

- Structured data: data especially formatted word or numbers to be read and exploited via computers (database, spreadsheets). This is a minority (estimated to 5%) (The Economist, 2010)
- Unstructured data: data without organization to fit an analytical process via computers
- Semi-structured data: originally unstructured data that are tagged to make them machine-readable. This category is growing the fastest (Gandomi & Haider, 2015).

And there is also variety in the type of data: traditional (numerical, categorical, binary), texts, clickstream, geo-based location, sensor data, images, audio, video. Alongside the growth of types of data collected, techniques to analyze them improved (such as image or voice recognition) in order to make these valuable.

New paradigm from data explosion

The richness and possibility of data growing day after day, the former data approach were not sustainable anymore and a new paradigm emerges, with three pillars:

- **The abandon of sampling**: the use of samples is rooted in the statistician practice, however with current data science, there is a tendency to go for large data collection, to leverage the new opportunity of technology
- **The acceptance of inaccuracy**, correlated to the large data sets. The principle behind is to consider that it is more valuable to have a lot of data with a slightly lower quality than perfectly curated data. With the collection of more unstructured data, the cleaning phase is intrinsically harder, and messiness is inevitable (except huge costs); also, the new data science practice looks for patterns and trends, more than outliers and exceptions, with raise the tolerance towards data quality (Becker, Dunn King & McMullen, 2015).
- **The focus on correlation instead of causality**. (Cukier & Mayer-Schoenberg, 2013). The interest of huge data collection allows to identify trends and patterns, but not the origin of the pattern. Thus, data scientists focus on prediction according to variable items and repetition through times. In automotive repair, they are not looking to understand why the brake broke, but when the brake is likely to break next time.

1.2. Data science process and roles

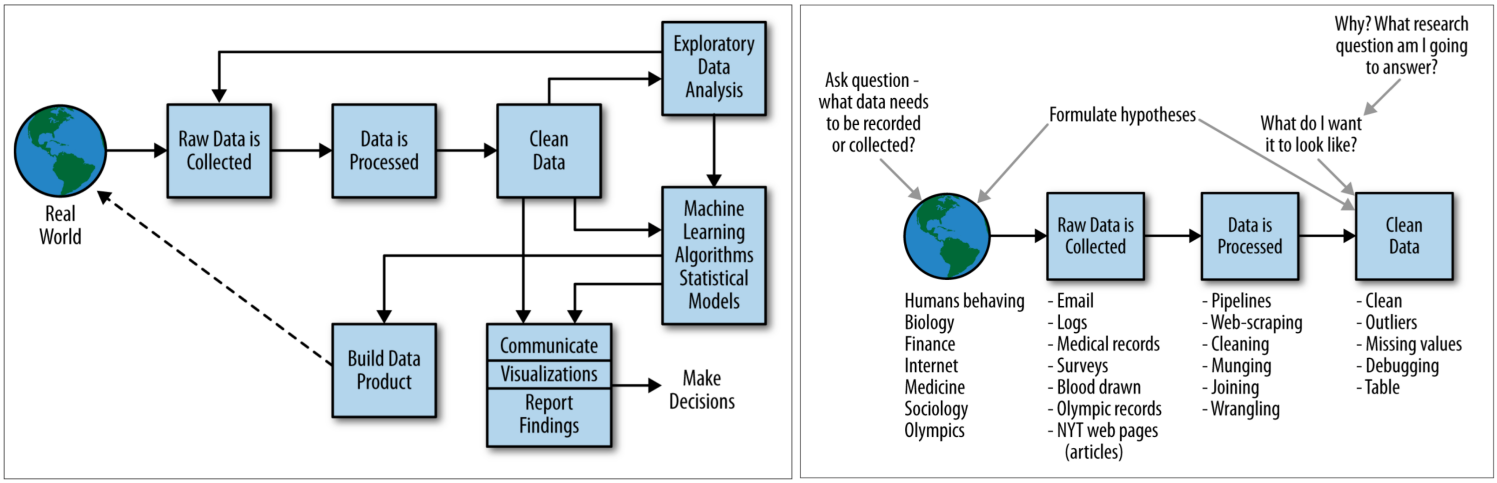
The process steps

Different canvas emerged since 2000's to codify the data science process. As data science arose from the professional field and not the academic one, everyone came with its own process, and lacks of systematic analysis and evaluation were frequent. Having a structured thinking helps to balance the necessity of human creativity an intuition and analytics tools capabilities at crucial steps (Provost & Fawcett, 2013). Frameworks can focus on data mining like KDD (Knowledge Discovery in Database), or adopt a more global approach, including business understanding, as SEMMA (Sample, Explore, Modify, Model, Assess) and CRISP-DM (CRoss-Industry Standard Process for Data Mining) (Azevedo & Filipe Santos, 2008). Alternatively, companies can create their own process like Microsoft with TDSP (Team Data Science Process) (Microsoft Azure, 2017). Nevertheless, as observed for design innovation process, there is a common base with similar steps:

- **Problem understanding**: definition of the objectives by identification of key variables and metrics measuring the success of the process. Then, refinement of the question expected to be answered (e.g. looking for a classification, an anomaly detection, a recommendation). Finally, definition of data sources (already existing or to be created)
- **Data acquisition**: production of data, first contact with it thanks to Exploratory Data Analysis, cleaning of the data, implementation of the data pipeline (the automatized periodical data refreshing)
- **Modeling**: application of diverse algorithms, machine learning or statistical models to the data sets, with calibrated optimal parameters. The results of every models tested are evaluated through the success metrics defined earlier
- **Deployment**: organization of the data output to deliver it to decision-making team, via reports, dashboards, spreadsheets, apps or creation of a data-based product. The knowledge gained through the data mining needs to be communicated.

It is interesting to note that when building a data product, an extra-step is required: the analysis of the feedback loop when data are reinjected in the real world (Schutt & O'Neil, 2013). Taking the example of votes for political election, the fact of having polls published every week (in a predictive perspective) influenced the electors' final behaviors (unwanted causal effect). Similar side effects happen with data product: the algorithm behind the Amazon recommendation tools ("You may also like…") is inducing a selection of recommended products that are more likely to sell, so more likely to appear in the recommendation, so more likely to sell: the data originally made to predict is acting on real behaviors and inducing the future data sets. The prediction/causation duality in data products needs to be considered and evaluated in product development and deployment, to be aware (and avoid?) ethical issues.

Figure 2.1: Data science process and Data scientist role in the process
(source: Schutt & O'Neil, 2013)



Data science: multi-levels, multi-functions

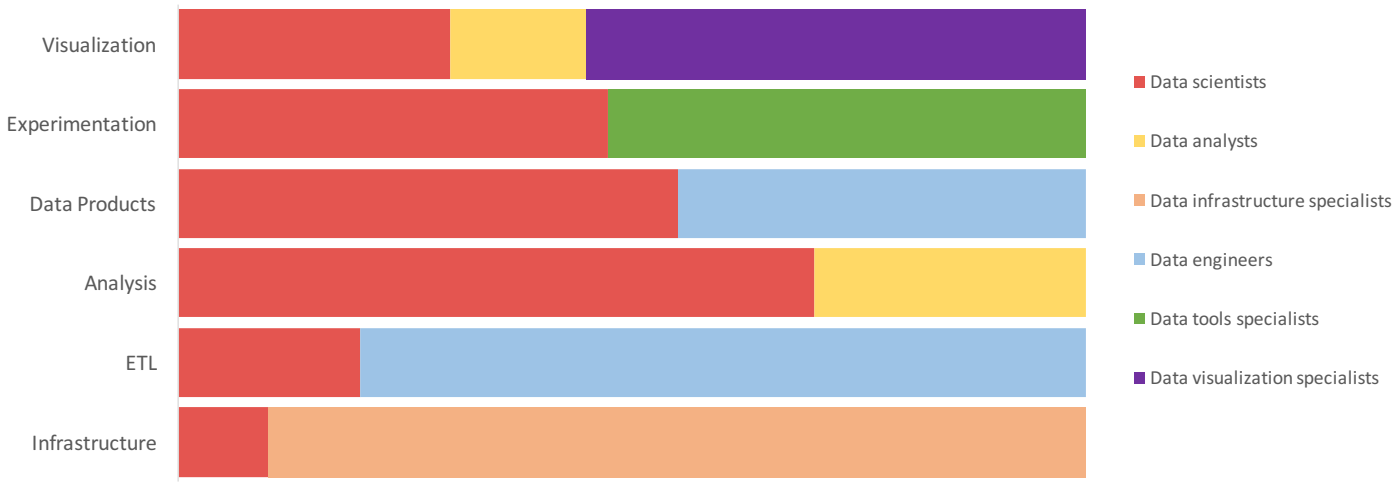
Like design, data science is not a unified field with workers sharing similar position, similar skills. Due to the increasing number of tools necessary to create and maintain the data process, roles are more and more separated and specialization emerged. In the case of companies with a limited data infrastructure, the data role is limited to data scientist position, the most generalist one. But for the others, data positions are specialized. It is important to raise awareness on these different categories to ensure the collaboration with other departments: designers may regularly need to work with data visualization specialists and data engineers, whereas they would have more restricted relations with data infrastructure specialists, hence the interest of a clear understanding of all the roles.

The framework presented is based on the organization in tech start-ups, especially Airbnb, but the general levels are common to most data science departments (Figure 2.2) (Johnson, 2017). The missions of all segments are summarized in Table 2.1. The six different jobs presented corresponds to different data skills: data infrastructure specialists have the knowledge on building and maintenance of data warehouses, a high point of the job is the insurance of data security; data engineers have an extensive knowledge on software engineering and programming tools; data analysts need a combination of programming tools skills (especially languages for database search like SQL) and standardized visualization tools (Tableau); data tools specialists are generally focused on the experimentation phase and are able to implement precise frameworks for tests; data visualization specialists have both the knowledge for visualization tools and UX/UI principles; finally, the data scientists is supposed to master all the skills (which explains the scarcity of data scientists) (Nelson, 2018).

Table 2.1: Data science departments

Levels	Functions
Visualization	Creation and maintenance of dashboards, synthesis of key data for data-driven decisions
Experimentation	Implementation of tests to measure the relevance of the products and decisions, detangling of causality
Data Products	Creation of algorithm, machine learning models for data product, analysis of the causality feedback loop
Analysis	Understanding user behaviors through key metrics, checking of previous and actual data reliance
ETL[8]	Curation of clean data, implementation of data pipelines to ensure data available in tables
Infrastructure	Sustainability of data warehouse and tools: the backbone of data work

Figure 2.2: Data stack, general departments organization and split of different data roles at each level (adapted from: Johnson, 2017)



[8] Extract, Transform, Load

2. Data science – Design: an existing bond

2.1. Data visualization

Data visualization definition

Data visualization is becoming a big trend since the middle of the 19th century, with a pinnacle in 2000, with the emergence of data visualization specialists as a job title (Friendly, 2008). Nevertheless, data visualization is part of history, and probably older than statistics and data science: the cartographers and surveyors work back to the 2nd century are first attempts of data visualization (Strecker & Cox, 2012). Famous pre-2000's examples include Napoleon's Russian campaign map, Mendeleev's periodic elements tables or London subway map.

Data visualization today is defined as the combination of three different visualization domains (Lindquist, 2011):

- Information visualization and data analytics: with the growth of data sets, computing and graph-making community joined their forces to develop process and tools to convert abstract data into visual-spatial forms, to nourish human intelligence since the 90's (Chen, 2006). Main stakes of the area are evaluation the efficiency of diverse graphical representations, automatized processes to display data, support for data exploration - in parallel with Exploratory Data Analysis development (Tukey, 1977)
- Visual facilitation for thinking and strategy: this refers to research on the power of diagramming in user engagement. The perspective is not data-driven but decision-making focused: the intention is leveraging data visualization to foster understanding of complex systems and ease discussions (Lindquist, 2011)
- Graphics and information display: this third part of data visualization is the one directly linked to design, exploring "aesthetics" and "beauty" (Strecker & Cox, 2012, Lindquist, 2011). Visual rendering is a pillar to make data impacting and embedding the story-telling underlying the data.

Data visualization specialists navigate around this three fields to develop their own practice, which explains the variety of functionalist approach in data visualization, like in data science more generally.

Design for data visualization: making data pretty?

The entry point of design in data visualization is often related to pure aesthetics in literature: in a chapter called "Think like a designer" (Knaflic, 2015), author delivers advice centered on layout, color, typography, contrast, alignment; publications like "Beautiful visualization" (Steele & Iliinsky, 2010) are entirely focused on design contribution in data "beauty". However, design contribution to the field cannot be reduced to the purely formal aspect: designers strive to give meaning, embodying a vision, a perspective in the product/service they are working on (Verganti, 2009). Thus, designers work on data's meaning through storytelling in collaboration with data scientists.

Data storytelling (or Narrative visualization, Segel & Heer, 2010) appeared with the need to make data memorable and make them stand out among the numerous data flourishing. Storytelling is a convenient and usual way to conserve and disseminate data (in the broad sense), and foster memory (Kosara & Mackinlay, 2013). Designers use their meaning and storytelling science to inject it in data visualization. Segel & Heer (2010) analyzed and classified the different techniques implemented by designers to communicate the storytelling: visual structure, visual highlights transition guidance, narrative ordering, narrative interactivity, messaging (the panel is broader than color and fonts).

On a different aspect, design (and especially interaction design) is highly involved in the haptic data visualization field. Haptic devices are "interfaces to computers or networks that exchange power (e.g., forces, vibrations, heat) through contact with some part of the user's body, following a programmed interactive algorithm" (MacLean, 2008). Research teams of developers, data visualization specialists and interaction designers work on possibility to make data model understood through haptic devices, but also make the user feels the (quantified) value of data (Paneels & Roberts, 2010). One of the added value of the design field in this case is the study of affordances to improve the usability of the devices.

Affordances is used in haptic design, but also in more classic data visualization tasks like dashboards. The following section presents a concrete example of benefits of the affordances design theory. There are more

spectacular examples, but this one proves the interest of design and data science collaboration on everyday basis.

An example: affordances in dashboard

Affordance is a design theory introduced by Don Norman (1988) based on Gibson's work (1977). Affordances are defined as "the actionable properties between the world and an actor" (Norman, 2013). This is what an object can do in user perception and it is crucial to play with affordances to induce the right use of a product. The four principles of affordances are:

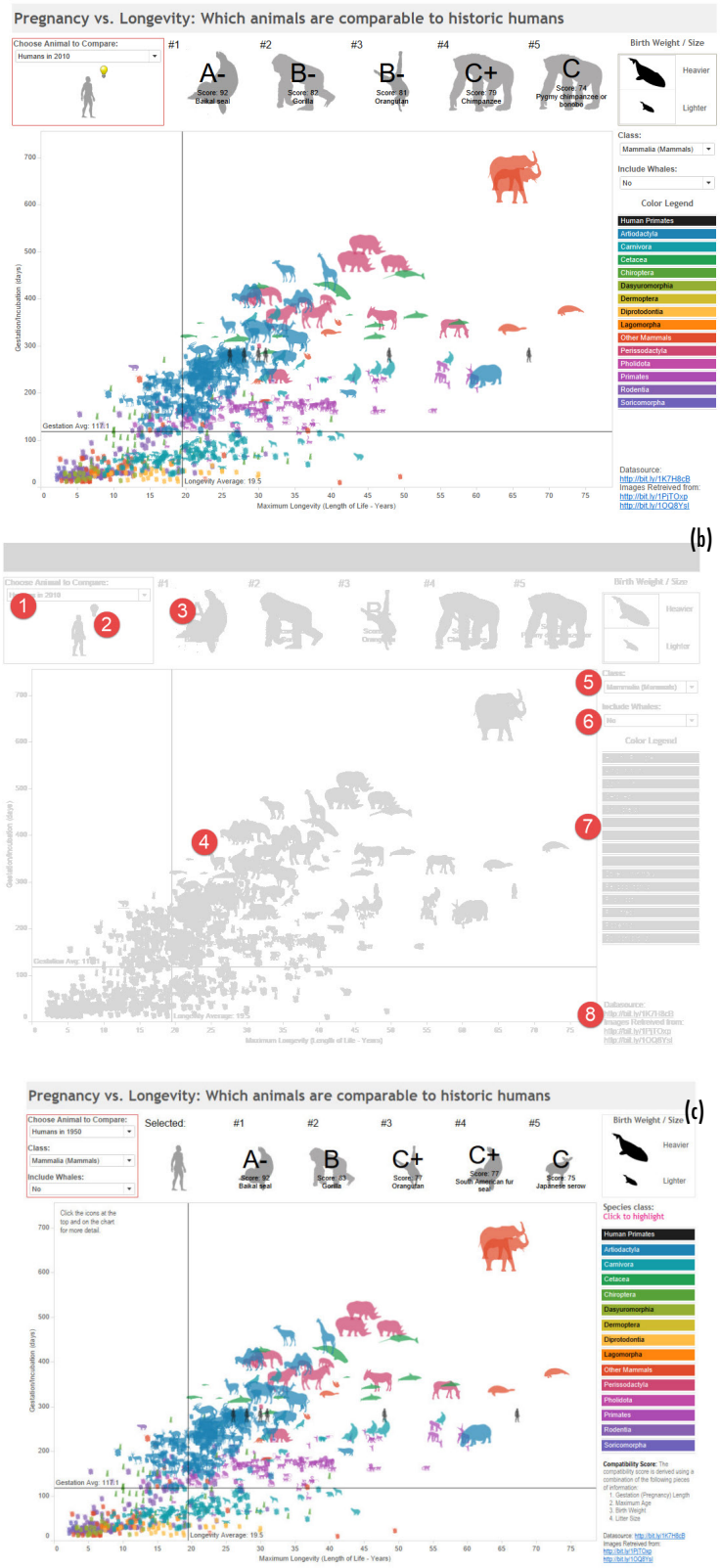
> *"1. Follow conventional usage, both in the choice of images and the allowable interactions.*
> *2. Use words to describe the desired action (e.g., "click here" or use labels in front of perceived objects).*
> *3. Use metaphor.*
> *4. Follow a coherent conceptual model so ◊ that once part of the interface is learned, the same principles apply to other parts."*

To show the utility of affordances theory in data visualization, Andy Cotgreave (2017) took a dashboard produced by a data visualization specialist and applied affordances principles to improve it (2017). The first step was to spot the affordances: (1), (5) and (6) are drop-down fillers; (2) is a lightbulb giving explanation on the grade when hovered; (3) and (4) are clickable animals to highlight them on scatterplot; (7) the color legend is clickable to highlight only a category of animals; (8) refers to data source.

Then, the affordance diagnosis: most of affordances are not visible instinctively because of lack of explanations (fail on affordance principle 2); similar affordances like drop down fillers are scattered (fail on affordance principle 4); the lightbulb does not carry an obvious interaction (fail on affordance principle 1).

Cotgreave rejoined the drop-down fillers together, text boxes were added for all the counterintuitive zones, lightbulb carrying explanations have been replaced by text in the corner. These minimal changes made the dashboard more readable and accessible. Users can benefit from the numerous actions more intuitively.
This example embodies the easy gain for data visualization by incorporating design theory, more than just color and fonts choice.

Figure 2.3: Eric Brow Longevity and Gestation in Human and Animals (Brown,2016)
(a) Original dashboard, (b) affordances spotted, (c) modified dashboard by Cotgreave



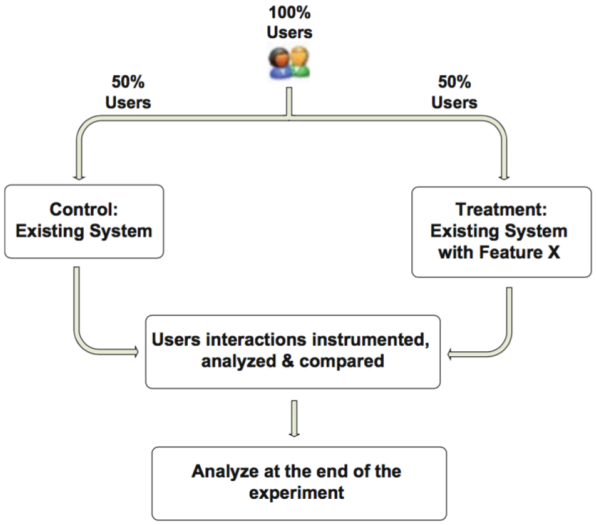
2.2. Experiments in digital design: A/B/n tests

Origins

User experience (UX) design is infused by the culture of experimentations, aligned with the will to stay closer to the user and to take data-driven decisions: testing is the opportunity to treat users as "co-developers" (O'Reilly, 2009). A/B/n tests is the most spread controlled experiments among UX design community: it consists in the random presentation to users of different variations of the design, with only one feature changing across the versions: version (A) – the control, version (B) – the variation, versions (n) – the potential other variations tested. Then, data scientists observe the "statistical significance" (p-value) of the results' difference between the versions. If p-value is under a preset level, the version (B) (or (n)) are considered as making an impact; if p-value if over the preset level, it is marked as coincidence (Greenland et al., 2016). To summarize, it quantifies benefits and drawbacks of design proposals compared to the actual version. Figure 2.4 shows the global structure of A/B tests.

A/B/n tests use in UX design (or interface design as called at the time) appears in the 60's at Bell systems, where engineers worked on the design of pushbutton telephone - "What are the desirable characteristics of pushbuttons for use in telephone sets?" (Deininger, 1960). Nowadays, A/B/n tests are heavily implemented by digital companies like Google and the famous A/B/n test of 40 shades of blue for the color font (Srinivasan, 2012). In digital development, prototypes cost less than in regular product industries, so designers and data scientist pairs easily to test versions (equals to prototypes), letting the user "unconsciously" making the choice amongst proposals. This point of view is coherent with "perpetual beta" state rooted in digital companies' DNA (W.H. Young, 2017): A/B tests are constantly run on websites like Facebook, Amazon, Airbnb, allowing perpetual user-revisions, continuing the iterative process of UX improvement.

Figure 2.4: A/B Testing - High-level structure of an online experiment (source: Kohavi & Longbotham, 2017)



In the case of data visualization, the relationship between design and data is: design supporting data science activity; whereas in A/B/n tests, this is the opposite: data science is supporting design activity (and design decisions) through the quantified experiment framework.

Designers' interest in A/B/n tests

However, at first glance, the scientific and empirical systematized approach driven by A/B/n tests can be pointed as a limitation for the creative process - which is comparable to Roberto Verganti's thoughts on how user-centered design kills radical creativity (2009). But 20 years of experience in A/B/n tests highlighted clear advantages:

- Confrontation to human bias: design process is necessarily impacted by natural bias. It is not problematic but A/B test offers a chance to filter what are the efficient design improvement and the design choice dictated by irrational motivations
- Deliverance from certain design guidelines: design culture is paved of design guidelines, stated before the digital era, and they are not always scientifically proved. With the use of A/B tests, designers gain freedom to test "unruly" design – the experiment will give an answer, no matter guidelines (Andrasec et al., 2011)
- Help in the hard process to assess the value of ideas: as Avinash Kaushik (2006) says "80 % of the time you/we are wrong about what a customer wants." This existing situation can be partly overridden thanks to A/B/n tests: at least, design proposals are ranked, it guides ideas selection through objective criteria, and assesses quantitatively the failure or success of proposals.
- Trustworthy results at controlled costs: A/B/n tests can be implemented at relatively low cost (numerous automatized tools exist) and the tests

can reach a large audience in the digital industry, getting solid number is easier than in other test methods. (Kohavi & Longbotham, 2017). And numbers are more positively perceived in the corporate world, due to habits (Muratovski, 2016)

There are drawbacks in A/B/n tests: it does not help in producing good design, if all design proposals tested are bad, it just ranks the bad proposals (Andrasec et al., 2011) and no daring decisions would be made if everything is A/B/n tests, whereas these "gut-feeling" decision are necessary to differentiate and challenge the status quo (Bowman, 2009). But wisely executed, A/B/n tests are a powerful tool, aligned with the iterative system promoted by design innovation process.

Example: Amazon shopping cart button

Amazon is well-known for its data-driven culture: any design proposals is A/B tested through "test cell", a variant with a small portion of users are exposed to the new design proposal. A/B tests gave Amazon designers the possibility to adapt through ecommerce explosion via incremental changes. To mention a few (Eisenberg, 2013):

1. The focus on reassuring clients (this is the debut of ecommerce) with 3 points-of-action assurances ("you can always remove it later", "Shopping with us is safe", "Guaranteed") in (a) disappearing progressively to nothing in (d): web shopping is no longer associated with worries
2. The lifting of "add to shopping cart" button: from a left-side rounded flat yellow and blue rectangle (a)-(c), to a 3D effect button (d)-(e), to a minimalist yellow and black rectangle with reduced text "Add to cart" (g): the aim is to clear the process
3. The color differences between "Add to shipping cart" and "Add to Wish List" buttons: similar color in (c)-(d), to more and more contrasting colors (light grey versus yellow for (g)). There is an intentional focus on a primary call-to-action, and secondary call-to-actions are muted.

More than an in-depth analysis, this example embodies "the perpetual beta" state: A/B/n tests acts as a catalyst of permanent reassessment. Designers constantly questions the shopping experience, and data science helps them to find a path: a win-win for the multidisciplinary team.

Figure 2.5: 15 years of Amazon shopping cart evolution (2003-2018) (adapted from: Einsenberg, 2013)

2.3. Global digital product developments

Continuity in the collaboration

As illustrated by the previous examples of designers and data scientists' collaboration (for data visualization and user experience experiments), links and contacts are highlighted between the two departments. This collaboration reaches a larger scale in digital companies, throughout the whole product development process. This can be explained by the integration of data scientists and UX designers at early stage in this field, so the chance of shared projects is higher.

There is no precise focus area of the product development process in this section: the aim is to give a glimpse of the global virtuous circle when a constant dialogue between data science and design exists. This is not surprising as both practices have clear similitudes (Table 2.2) (Bloom, 2017).

Table 2.2 Parallelism between design and data science process (adapted from: Bloom, 2017)

	Designers' practice	Data scientists' practice
1st step – Inspire	Identification of the latent needs and desires of users	Identification of latent mental model applied to view a particular problem
2nd step – Ideate	Generation of concepts, prototype and test with users	Generation of hypotheses, confirmed or rejected with data
3rd step – Implement	Development of the new product or integration of new features	Development of a computational model

User and context vision

The user and context perception is an axis of differentiation between design and data science. Designers and data scientists may have two opposite approaches: design process generally includes the generation of archetypal models to define the targeted user (age, gender, socio-professional category for the basic ones): this is the guide to develop concepts and ensuring they will fit to the model. On the contrary, data science models are not implementing this type of criteria: the very essence of data models is to give a proposal which should statistically fit with the behavior previously observed, whatever this behavior is made by whoever (Girardin & Lathia, 2017).

However, with the growing size of data sets, searching the right angle to detect relevant findings is harder, and as the final user is human after all, data scientists started to get interested in user research methods used by designers as a support before diving into data. It can take the form of practical tips on means to empathize with user like varying communication forms (speech, drawing, role-playing) or asking for specific stories (without data assumptions already in mind) (Bloom, 2012). But it can be a tighter data-design work as demonstrated by Heather Ford (2014) on the addition of ethnographic research in a data science project about Wikipedia and the sources habits: the data team had data for 67 million sources, but was not able to define why sources were more recurrent than others among the huge data sets. Ethnographic inputs - like the fact that Eastern Wikipedia writers were more likely to cite Western sources because they consider that readers will trust more Western sources (even if they are first documented through local sources)- helped to find a certain slicing of data, and the project succeeds. Re-injecting the human factor in data science can clear obscure and gigantic data sets.

Data product development

Data products (product based on data collection and generation such as shopping recommendation tools or search engines) are good examples of development process where designers and data scientists must work together from the beginning. The product UX needs a common brainstorming on the three phases of the experience:

1. The onboarding of user on a product with no data on the user, whereas the model is based on previous data collected
2. The evolution of the experience with the model changing with data addition. The model carries a portion of uncertainty and the product needs a frame to provide a satisfying, non-annoying, relatable experience
3. The offboarding of user. A choice needs to be made on storage of the data produced and potential re-use (Girardin & Lathia, 2017)

These three states have to be designed together to combine technical issues, relevance for user, business significance.

The example: Rise Science and IDEO

In 2016, IDEO, a design agency famous for the promotion of design thinking, helped the start-up Rise: they were initially called for a data visualization redesign support, but ultimately conducted a product development process to refine the complete experience.

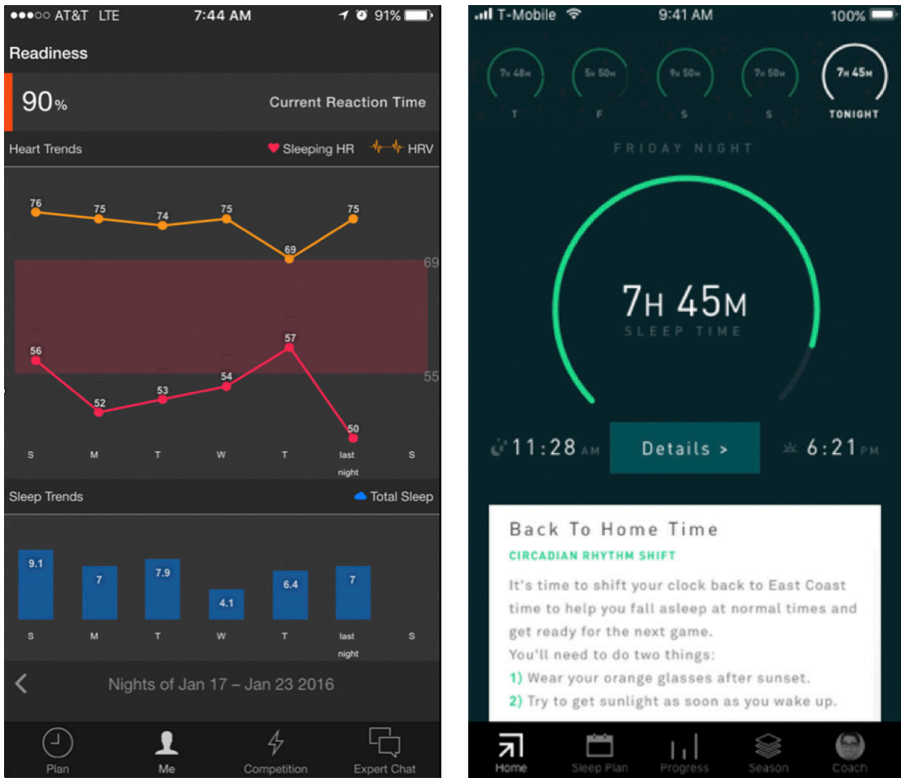
Rise is a " Sleep Improvement Program designed for elite athletes that have intense physical demands, time demands and constrained schedules because of travel or academics […] The goal of the program is to improve objective and subjective sleep outcomes." (Rise Science, 2017). Founded by three Tennessean engineering students researching in sleep science, they developped a first app, gathering data about athletes' health through biceps straps (heart rate, sleep length, restorative sleep) and built a data model to predict what could be an efficient sleeping plan to increase sleep length (Wetteren & Malmgrem, 2018). The data model was satisfying as athletes gained 1 hour of sleep by night in average (Rise Science, 2017). However, the engagement rate with the app was disappointing (15% after 20 weeks of test) (Schwab, 2017). The app was composed of charts of the previous week data (sleep time, sleep debt, time to fall asleep, sound environment) as shown in Figure 2.6 (a).

IDEO team implemented a proper user and context research plan using methods of athletes and coaches' interviews, immersion in college dorm rooms and sport facilities, thinking-out loud feedbacks. They noticed the amount of data athletes were exposed everyday between trainings, competition strategy, diet (and still college work). Adding a layer of data for sleep was burdensome. Thus, the app's homepage became a clock indicating the advised time to go to bed, simply, removing the confusing data (Figure 2.6 (b)). Moreover, the IDEO team remarked how athletes are formatted through the coach-trainee relationship: Rise will not be an app, but a sleeping coach. The app now includes a chatbot embodying the "coach" (mixed fed by algorithm and human answers).

The coaching is accompanied by bed sensors (to collect data), UV glasses and sleep mask to optimize the falling asleep phase, and athletes have clear indications on the time they should wear the glasses and mask, like the bed time. This became crucial with the long travels and change of time zones due to competition. The refinement of the concept went far beyond the pure date visualization, combining software and hardware, improving app engagement through coach analogy, focusing on data that matters for athletes. Rise is now adopted by several college teams through the USA, have a partnership with a star team of NBA League, the Chicago Bulls, and works on expanding the sleep coaching to companies' workers.

As outlined by Wetteren & Malmgrem from IDEO (2018), a project starting only with data perception (what data can do) will not automatically "reflect and respond to the functional, social, and emotional behavior of users". A balance between exploratory data analysis and design research. Onboarding data scientists and designers together is an efficient choice to improve the user experience, utility and understanding of data: this can be one of the differentiator for product success or failure.

Figure 2.6 Rise science app homepage, (a) before IDEO work (b) after IDEO work (source: Wetteren & Malmgrem, 2018)



References

Andrasec, D., Bloice, M., Lexer, G., & Zernig, J. (2011). *A/B testing literrature review.* Unpublished Graz University of Technology,

Azevedo, A., & Filipe Santos, M. (2008). KDD, Semma and CRISP-DM: A parallel overview., 182-185.

Becker, D., Dunn King, T., & McMullen, B. (2015). Big data, big data quality problem.

Bloom, E. (2017). *How adopting a design mindset made me a better data scientist.* Retrieved 06/20, 2018, from https://medium.com/capital-one-developers/how-adopting-a-design-mindset-made-me-a-better-data-scientist-863ab0fcd20b

Bowman, D. (2009). *Goodbye, google* Retrieved 06/20, 2018, from https://stopdesign.com/archive/2009/03/20/goodbye-google.html

Buytendijk, F. (2014). *Big data ethics: Privacy, risks and principles.* Unpublished manuscript.

Chen, C. (2006). *Information visualization: Beyond the horizon*

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171-209.

Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review, 69*(1), 21-26.

Cotgreave, A. (2017). *Affordances and signifiers: Applying design theory to your dashboards.* Retrieved 06/16, 2018, from http://gravyanecdote.com/bigbookofdashboards/affordances-and-signifiers/

Cukier, K. N., & Mayer-Schöenberger, V. (2013). The rise of big data: How it's changing the way we think about the world.*92*

Davenport, Thomas H., and D. J. Patil. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review 90, 10*, 70-76.

Deininger, R. (1960). Human factors engineering studies of the design and use of pushbutton telephone sets. *Bell System Technical Journal, 39*(4), 995-1012.

Eisenberg, B. (2013). *Hidden secrets of the amazon shopping cart 2.0.* Retrieved 06/20, 2018, from https://www.bryaneisenberg.com/hidden-secrets-of-amazon/

Ford, H. (2014). Big data and small: Collaborations between ethnographers and data scientists. *Big Data & Society, 1*(2), 2053951714544337.

Friendly, M. (2008). A brief history of data visualization. *Handbook of data visualization* (pp. 15-56). Berlin, Heidelberg: Springer Berlin Heidelberg.

Fry, B. J. (2004). Computational information design. Massachusetts Institute of Technology).

Gandomi, A., & Haider, M. (2015). *Beyond the hype: Big data concepts, methods, and analytics* doi:https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gartner Inc. (Unknown). *Gartner IT glossary.* Retrieved 06/10, 2018, from https://www.gartner.com/it-glossary/big-data/

Gibson, J.J. (1977). *The theory of affordances, in perceiving, acting, and knowing. towards an ecological psychology.* Hoboken, NJ: John Wiley \& Sons Inc; Hoboken, NJ: John Wiley \& Sons Inc.

Girardin, F., & Lathia, N. (2017). *When user experience designers partner with data scientists.* Association for the Advancement of Artificial Intelligence.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., et al. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*, 337-350.

IDEO. (2016). *A game-changing approach to sleep for athletes.* Retrieved 06/20, 2018, from https://www.ideo.com/case-study/a-game-changing-approach-to-sleep-for-athletes

Product School (Producer), & Johnson, T. (Director). (2017, 09/09). *Why analytics is key for product managers by airbnb data scientist.* [Video/DVD] Youtube:

Kaushik, A. (2006). *Experimentation and testing: A primer.* Retrieved 06/19, 2018, from https://www.kaushik.net/avinash/experimentation-and-testing-a-primer/

Knaflic, C. (2015). *Storytelling with data : A data visualization guide for business professionals.*

Kohavi, R., & Longbotham, R. (2017). Online controlled experiments and A/B testing., 922-929.

Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. *Computer, 46*(5), 44-50.

Lindquist, E. (2011). Surveying the world of visualization. *HC Coombs Policy Forum,*

MacLean, K. E. (2008). Haptic interaction design for everyday interfaces. *Reviews of Human Factors and Ergonomics, 4*(1), 149-194.

Microsoft Azure. (2017). *The team data science process lifecycle* Retrieved 06/07, 2018, from https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/lifecycle

Muratovski, G. (2016). *Research for designers: A guide to methods and practice*

Nelson, S. (2018). *4 types of data science jobs.* Retrieved 06/12, 2018, from https://blog.udacity.com/2018/01/4-types-data-science-jobs.html

Norman, D. (2013). *Affordances and design.* Retrieved 06/16, 2018, from https://www.jnd.org/dn.mss/affordances_and.html

Norman, D. A. (1988). *The design of everyday things*. New York: Basic Books.

O'Reilly, T. (2009). *What is web 2.0 - design patterns and business models for the next generation of software.* Retrieved 06/18, 2018, from https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=all

Paneels, S., & Roberts, J. C. (2010). Review of designs for haptic data visualization. *IEEE Transactions on Haptics, 3*(2), 119-137.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*(1), 51-59.

Rise Science. (2017). *Sleep improvement program for elite athletes.* Unpublished manuscript.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). *Analytics: The real-world use of big data, How innovative enterprises extract value from uncertain data* IBM.

Schutt, R., & O'Neil, C. (2013). *Doing data science: Straight talk from the frontline* O'Reilly Media, Inc.

Schwab, K. (2017). *Ideo designs A sleep system to help athletes get more shut-eye.* Retrieved 06/20, 2018, from https://www.fastcodesign.com/90129191/ideo-designs-a-sleep-system-to-help-athletes-get-more-shut-eye

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 1139-1148.

Srinivasan, R. (2012). *Testing 40 shades of blue – AB Testing.* Retrieved 06/19, 2018, from https://iterativepath.wordpress.com/2012/10/29/testing-40-shades-of-blue-ab-testing/

Steele, J., & Iliinsky, N. (2010). *Beautiful visualization: Looking at data through the eyes of experts* (1st ed.) O'Reilly Media, Inc. [download].

Strecker, J., & Cox, A. (2012). Data visualization in review: Part of the strategic evaluation on communicating research for influence. *Strategic Evaluation on Communicating Research for Influence: Part I,*

The Economist. (2010). Data, data everywhere: A special report on managing information. *The Economist,* (February 2010)

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, Mass.; Menlo Park, Calif.; London; Amsterdam: Addison-Wesley.

Verganti, R. (2009). *Design-driven innovation: Changing the rules of competition by radically innovating what things mean.*

W. H. Young, S. (2014). Improving library user experience with A/B testing: Principles and process.*1*

Ward, J. S., & Barker, A. (2013). *Undefined by data: A survey of big data definitions" arXiv: 1309*

Wetteren, J., & Malmgren, D. (2018). *What happens when data scientists and designers work together.* Retrieved 06/20, 2018, from https://hbr.org/2018/03/what-happens-when-data-scientists-and-designers-work-together

Transition

The first part of the thesis, Chapters 1 and 2, tackles numerous concepts between design and data science, necessary to understand the origin of both fields, the background of knowledge supporting the practice, variations in terms of processes and roles.

The first chapter focuses on the integration of design in innovation, originally dominated by the tech culture. This design integration passed through the implementation of different processes, always based on a common phase: an exploration of user and his context. The study of methods often exploited this first phase reveals a disparity between the use of qualitative and quantitative methods, which could be explained by the empathy needed to set up a strong contact with user.

The second chapter presents the big picture of data science and the core changes implied in data handling since the introduction of this new field. It also includes an overview of practical processes, departments and roles in data science, to have an overview of the daily practice. The end of the chapter is devoted to highlight the benefits designers and data scientists' collaboration through examples in data visualization, experimentations and digital product development.

These two chapters are not directly linked, but they are the foundations for the next sequence: designers have a restricted number of quantitative methodologies for the user and context research, whereas quantitative approaches have more and more impact in data-driven decision makings; can data scientists co-develop tools with designers to increase the sources of user and context data, explore new angles, highlight user insights at larger scale? This is the guiding question for the next chapters.

An overview of techniques leveraging data science to dig into user behaviors (and to point relevant insights) are presented in Chapter 3. Several case studies are analyzed in Chapter 4 to show the interlacing of these techniques to feed the user and context research in design process. Finally, Chapter 5 concludes with the challenges inherent to the use of such methodologies.

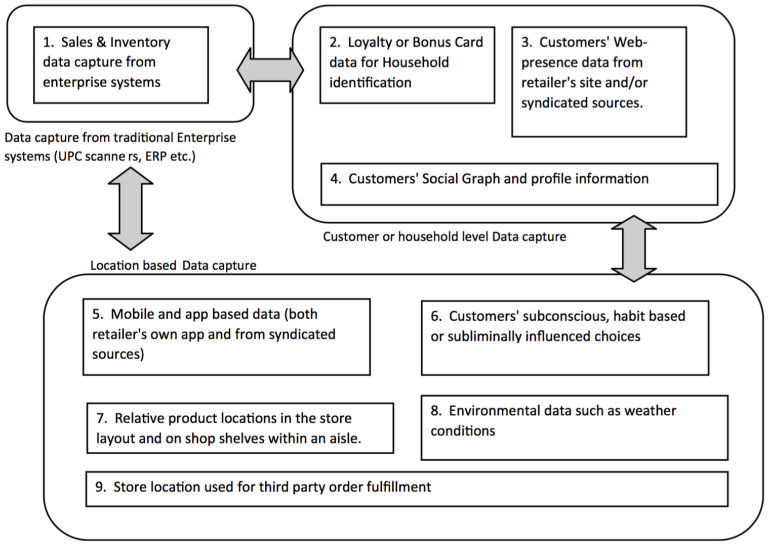
Chapter 3. Design – Data Science bond for user and context research

1. Gathering design insights through various channels

1.1. Data collection through retail channel

The Data science craze touched the retail sector quite early compared to other sectors: the supply chain management of any kind of store is a strategic concern. Assortment, pricing, store layout, multi-channel fulfillment are subjects that produce huge volume of data and have a high impact on sales: big data applications flourished early in the 2000's (Aktas & Meng, 2017). Besides supply chain, researchers started to explore possibilities to gather insights on customer behavior at large scale through the retail space. The methodologies here are mainly fitted to study the act of purchase, which is a more crucial point for marketers than product designers. However, they gave an understanding on user behavior in context, just like fly-on-the-wall ethnographic methodology, consisting of users' observation in context. New retail data sources enabled by data science are presented in Figure 3.1 and a focus on the sources directly linked to user behavior is developed below (#2, #3 and #6 on Figure 3.1)

Figure 3.1: New data sources in retail field (source: Bradlow et al., 2017)



Customers (and Users) profiling

Loyalty programs are a traditional way to get to know the user base of a product and its shopping habits, proposing a tacit pact with customers: in exchange for data (gender, age, email, status…) and purchase history, customers get promotions. The (already big) databases of customer profiles have been enriched by tracking IP address to cross brick-and-mortar and e-shop retails data, credit card recognition, and even individual user's social media activity if the user account is logged through a social media interface (Grewal, Roggerveen & Nordfält, 2017). Huge possibilities of machine learning exist: descriptive analysis to see the actual landscape of user base, predictive analysis to identify user segmentations and which segments is likely to grow/reduce (to adapt the product design for these segments tendency for example. This sounds like an idyllic path to learn from user demographics. However, these databases often reach extreme sizes and are sparse, making them hard to compute and data mine in reasonable business time. Data compression is the key, with the potential loss of information it induces, especially hiding or over interpreting outstanding behaviors, which are valuable for designers (Bradlow et al., 2017).

Path tracking

To understand shopping habits, path tracking reveals how customers moves in the store. Depending on the methodologies for data collection, the result can be a map of global people's flow (in case of person counting techniques) or the record of individual customer path. The technology used for tracking can be based on

devices – customers' smartphone directly or a device proposed by a store like an augmented shopping cart (Larson, Bradlow & Fader, 2005) – like Wi-Fi location, GPS location or Beacons (Bluetooth low energy radio waves). Alternatively, data tracking can pass through sensors – CCTV[9] via face recognition, thermal Imaging, infrared beams (Max, 2017). Marketing professionals interpreted the path data to study the correlation between shop space, path length and expense, the effect of store general layout and planned/unplanned expenses, most effective spread of product categories across retail space among others.

Understanding the customer habits through path tracking is an additional layer of user's understanding for designers. Moreover, path tracking techniques, that proved efficiency in stores, may be extended to other context to allow designers to collect objective data instead of what users say or what observers conclude (which always carry a certain degree of subjectivity).

1.2. Data collection through sensors

Grasping what users express as a need and wish is a thing. Grasping what users have as latent needs and wishes is another, even more difficult, because there is a useful medium missing: words. Majority of user and context studies are using dialogue. But to incite a fruitful dialogue on precisely what is unexpressed, observations mediums are needed. In general qualitative studies, the researcher is the instrument of observation, but it is limited in terms of bias and human sensorial capabilities. This is where designers can invoke sensors technology: by delivering quantified measurements, they reveal patterns and concerns[10]. However, the use of sensor in design research needs to be paired with qualitative studies because (Freach, 2014): 1) measurements from sensors, by their quantitative findings, may appear like "unarguable truth", and hide deeper reasoning, 2) experiments show that sensors results were particularly beneficial to validate/invalidate hypothesis from qualitative design studies, to provide legitimacy to intuitive ideas, 3) the addition of the sensor layer to qualitative data creates thicker data corpus, to be used for the actual product development and future ones.

Eye-tracking

Eye-tracking corresponds to the concept of eye motions measurements. From this measurement, it gives the pattern of observers' interest, their attention or inattention points (Duchowski, 2007). The basic measurements are location points of fixation, duration of fixation, number of fixations and ordered sequence of fixations; and the analysis is made through statistical methods to highlight: fixation clusters, attention patterns and saccade acceleration (Tannen, 2007). UX designers are intensively leveraging this technology to learn about interface interaction: what users look first, what they do not look at, what users are looking at before misusage (the cause of the issue) and how users learn the interface (Breeze, 2011). For now, devices for eye-tracking are more convenient for human-machine interaction (only the installation of a high speed camera on the computer or mobile, as the user has an almost fixed position) whereas in the case of product design, users needs to wear glasses. But the use of eye-tracking glasses is diffusing in design lab to learn about product-user interaction (Figure 3.2)

Figure 3.2: Example of eye-tracking via Tobii glasses - observation of an oven command board
Circles represents points of attention, diameters are proportional to fixation duration, numbers indicate the ordered sequence

The technology behind the camera is composed of face detection, user classification and three-dimensional eye-position extraction algorithms (Kang, 2016): data science made eye-tracking possible. Moreover, the systematic integration of eye-tracking modules on smartphones, laptops and desktop monitors is on his way, and the possible to eye-track

[9] Closed Circuit Television

[10] Additionally, sensor technologies are regularly used for product design evaluation to highlight friction points or success features (an example is give with eye-tracking in Guo et.al, 2016). This will not be detailed here as the focus is on user insights to generate new concepts, not directly assessment.

users at large scale is no longer science-fiction: descriptive and predictive machine learning analysis will have a key role to help designers interpreting this huge source of in-context data (no more restricted to labs because of the equipment) (Dickson, 2017).

Neuroimaging

Neuroimaging (also called brain imaging) is a research field on techniques' development to obtain representation of the nervous system. For the big picture, the goal is to see which parts of the brain are activated after a particular stimulus, and how to interpret these active areas. There are three main techniques in neuroimaging (Ariely & Berns, 2010):

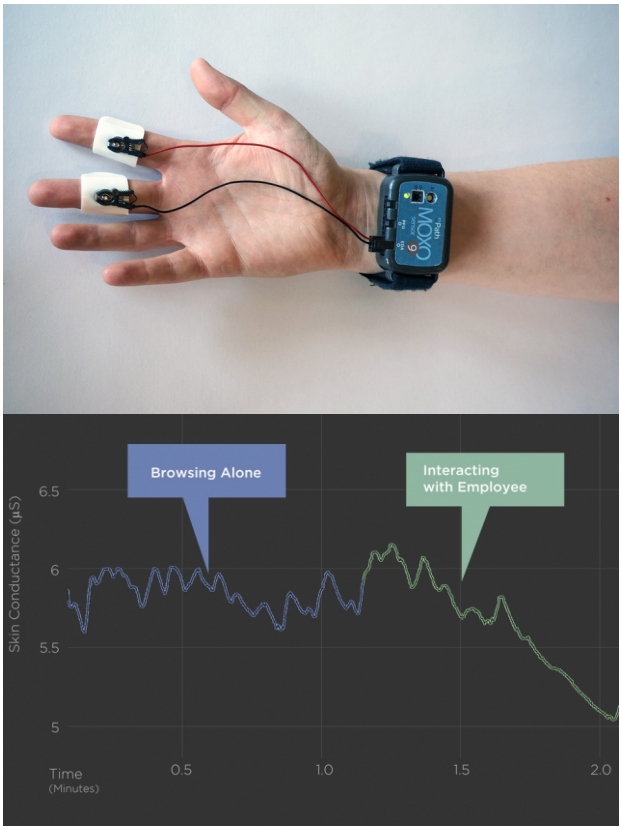
- Functional MRI: passing through MRI scanner to measure the blood oxygenation level-dependent signal
- Electroencephalography: electrodes applied to the scalp to measure the electric field variations of the nervous region underneath
- Magnetoencephalography: similar to the previous one, but measuring magnetic field change

Neuroimaging has been used in marketing and social research, to understand what was triggering certain emotions, and how to implement this triggers in future design. It has been applied in food industry (testing the effects of smells, tastes, assemblages, fat content on orbitofrontalcortex); in film editing to evaluate when emotional peaks are reached and how to condense them; in social science to pinpoint the emotions triggered when buying sustainable products (Goucher-Lambert, Moss & Cagan, 2016). The costs of neuroimaging set-ups (between 10,000$ and 2 million $) and the difficulty to precisely give a sense to brain parts activation are limitations to a widespread use of the technique, and it will not be a daily privileged source of data for designers in near future.

Stress sensors

In the same vein as neuroimaging, stress sensors aimed at discovering the emotional state of users when they are interacting with a product. Stress sensors developed by Elliot Hedman (founder of the start-up mPath), are called MOXO and are presented in Figure 3.3. It measures skin conductance variations: spikes in conductance may indicates stress and

Figure 3.3: (a) Moxo sensor
(b) Example of conductance measurement with Moxo
(source: Matheson, 2017)



frustration whereas dips are associated to disinterest or boredom. It allows a quantified and objective evaluation of interaction, with a precise timeline, for relatively low cost and convenient set-ups (Matheson, 2017). The MOXO will be studied more in details in Chapter 4.3 as one of the core element of the case studies concerning LEGO.

1.3. Designing for data collection

In this panorama of data collection options, designers are leveraging channels developed by others. But what would happen if designers could elaborate their own channel of data collections from A to Z? This is a possibility offered by including data-driven scenario in the product development pattern.

Data-driven scenario for product development

The launch of a product will result in data production from users: it can be directly expressed or induced feedback on the product with positive-negative feelings, usage or hijacked uses described, evaluation performance – the voice of customer (Griffin & Hauser, 1993). But it also includes nowadays indirect feedback, data collected during the use of the product e.g. demographics data if an account is created for the use, data collected from sensors integrated in the product (like Rise Science and the sleeping assistant). The idea of data-driven scenario for product development is taking this data production not as a corollary of the product development, but as the constitutive aim of the process.

In this perspective, Yu & Zhu (2016) propose a framework to supervise data-driven scenarios in product development: data-product-data design pattern. The model is first based on designers reflexions and decisions on the abstract data they wish to collect in a selected contextual space. Abstract data refers to a conceptual data, not already existing: numbers of visitors for a site by day is an abstract data, 5000 visitors a day is a (concrete) data. Next steps are to design and launch a first product, vectors of concrete data collection according to the abstract data targeted. Concrete data are extracted and mined for the product development of future products (Figure 3.4).

This framework can be applied to one of the most famous example of participatory design: the Nike iD platform. This commercial success story (Forbes, 2015) opened possibility for e-shoppers to personalize their sneakers (Ramaswamy, 2008). If the Nike iD platform has a clear functionality (offering mass customization), it is also a huge source for data collection, and valuable data for sneakers designers (fashion colors, geographical preferences, micro trends emerging) (Table 3.1).

Table 3.1: Application of data-product-data design pattern to Nike iD

Data-Product-Data pattern	Applied to Nike iD
Define contextual space	Nike's E-shopping space
Define valuable abstract data	Style trends
Put forward a solution for using abstract data	Platform of personalization asking for demographics data to create the account and allowing/recording design choices
Product design and development	Implementation of Nike iD
Obtain concrete data on the use of the product	Thousands of personalized sneakers created and recorded: Color chosen, soles preferences, straps/laces, embroidery preferences, gender, address, email
Innovate for new product development by data mining	Sneakers collection for next years adapted to the preferences emerging from the platform

By systemizing the implementation of data collection in functional product, designers have an opportunity to collect large data sets of insights they were especially looking for. Data mining processes support the data collection and helped designers to give sense to this huge volume of data.

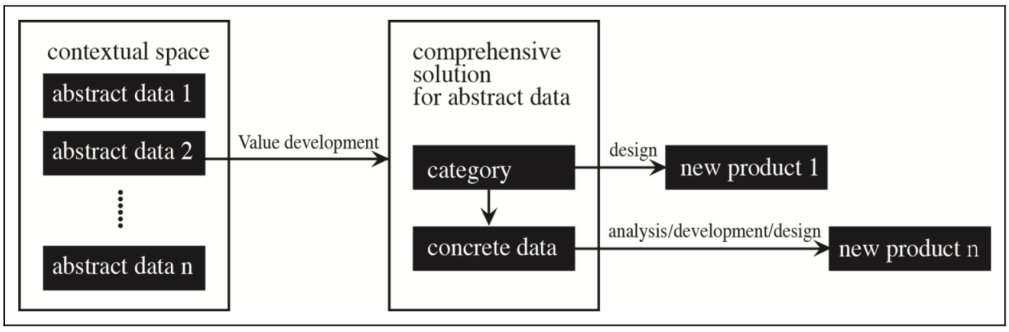


Figure 3.4: Data-Product-Data design pattern – collecting data spotted by designers to foster new product development (source: Yu & Zhu, 2016)

2. Qualitative methodologies reinterpretation through big data

2.1. Netnography: qualitative ethnography scaled up

The web is obviously an important data collection channel. From the last 20 years, it has become a place of exchange, a communities' aggregator, a communication vector for profit/non-profit organizations. The advent of social medias combined with the spread of connected devices creates an unforeseeable mass of data. Marketing (and business in general) has intensively explored possibilities and applications of big data mined in web (Lavalle et al.,2011). The results of these analytics studies can be studied and exploited by designers, but as they are not the primary recipients (so the results are not shaped to fit design approach), this facet is not detailed in the present thesis. However, there is a branch of another field tackling the mass of human interactions observed on the web to collect user and context insights: netnography, derived from ethnography (which is already widely used in the primary phase of design innovation process).

Netnography, a definition

Netnography is a range of analytic approaches and processes leveraged to gain access to online communities and finally produce insights on user behavior[11] (Kozinets, 2015). The interests of the method are: 1) the open and impartial voice taken by user when expressing online (partly thanks to a certain anonymity), the bias induced by an observation set up disappeared online; 2) a unique gathering of expert user exchanging together, which means that the panel of participants is not representative of the globality of real users, but they are more creative, engaged, original, and outliers always carry an interesting value in ethnography; 3) the users engaged in online communities are not only discussing about feedbacks on products, but also explaining their issues and their new concepts developments to face their problems (Pollock, Lüttgen & Piller, 2014).

Netnography's process is composed of 6 steps (Kozinets, 2002):

- Entrée: identification of online communities appropriate for the subject, and learning about the individuals interacting
- Data collection and analysis: recording of computer-mediated communications of the community and netnographer's fields notes.
- Analysis: study of the communications through "coding, content analysis, data linking, data display, and theory-building functions" (classification assisted by computer) and metaphoric-symbolic interpretation (by the netnographer)
- Providing trustworthy interpretation: evaluation of the limitations of the online medium and techniques
- Research ethics: framing the ethics of the study (disclosing the existence of the research to the community, permission for transcription…)
- Member checks: implementation of a feedback loop from the online community to comment netnography results.

Netnography leads to the highlight and discovery of consumer trends guiding future innovative product development process (example for electric vehicles in Pollock, Lüttgen & Piller, 2014).

Netnography supported by data science

Netnography burgeoned in the 90's, before big data craze. The analysis of data was helped by software for recording and classification, but each communication had to be added and tagged by the netnographer: the limitations of netnography size was the number of communications a netnographer could read/see and process in reasonable time. Netnography combines high-quality contextual results focused on a limited number of communities.

On the opposite, data scientists developed social media monitoring (SMM) tools allowing "the active monitoring of social media channels for information about a company or organization" (Lexicon FT, 2018). These commercial tools are used to record customer's conversations about brand, product, feedbacks and

[11] High level of nuances exists in defining netnography, versus online ethnography notably. For a complete panorama, see Costello, McDermott & Wallace (2017).

stories on a maximum of channels to spot hot topics, crisis, positive or negative feelings toward the brand (Reid & Duffy, 2018), through web crawlers[12], clustering, machine learning, creation of dashboards and word clouds. The results coming from these tools have a broad spectrum, high responsiveness, objective findings but they are lacking contexts, accuracy and motivations behind the communication, and textual analysis software is still in its infancy, itemizing words, missing irony, passion, tragedy, hope (Kozinets, Scaraboto & Parmentier, 2018).

Introducing SMM tools in netnography and vice-versa corrects the drawbacks of both methods. Social media monitoring tools:

- help in finding the online communities and the relevant communications for the netnographer (example of Twitter hashtag crawling by SSM to gather all the tweets with selected hashtags in Arvidsson &Caliandro, 2016)
- support the analysis phase by a categorization of communications (without predefined criteria) thanks to algorithm in easily handled dashboards, topic clouds based on frequency of words usage, categorization by direct mentions or retweets, to give a primary global mapping and investigation path for further "classical" qualitative studies (Reid & Duffy, 2018)
- include the different layers of online communication in the recording and analysis: text, actions (likes, shares), hyperlinks, images, search engines and location settings (Reid & Duffy, 2018)

The outcomes are rich insights, with a high granularity, in minimized time and costs, to understand user behavior through social media behavior.

Designers' interest

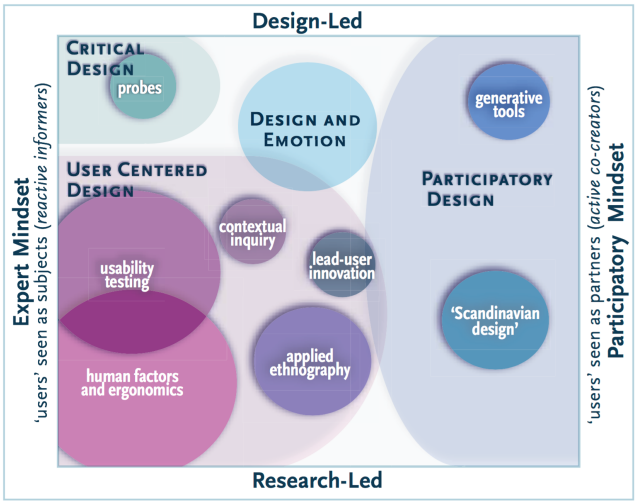
User and context research in design innovation process may already contained the usual netnography methodology. However, "data-augmented" netnography offers an original path for designers to collect qualitative insights at a quantitative scale, to reinforce observations from other qualitative methods (confirmation of trends and hot topics among the expert communities) and detect outliers' behaviors within large communities. Netnography paired with SMM has a moderate cost compared to regular ethnographic methods and measurable results, which made it particularly suitable for a first user and context exploration and orienting further research. Of course, designers can work in tandem with netnographer, but with the help of SMM to initiate data analysis, a trained designer alone may hold the key for an initial netnography.

2.2. Insights from participatory design at large scales

Participatory design

In correlation to netnography enabled by crawling, clustering and machine learning, rummaging the web automatically and systematically opens insights opportunities in the special case of participatory design innovation process, and co-creation in general. Participatory design is adjacent to the user-centered design innovation processes presented in Chapter 1: in user-centered design, the user is a subject; in participatory design, the user is a partner as a co-creator sharing creativity with the design team (*co-creation*) or as a co-designer when the creativity is shared at all stages of the design process (research, brainstorm, prototype, test, business model…) (*co-design*) (Sanders & Stappers, 2008) (Figure 3.5). And there is an interesting practice at the edge of the two approaches: lead-user innovation.

Figure 3.5: Design research topography (source: Sanders & Stappers, 2008)



[12] "A web crawler is a program or automated script which browses the World Wide Web in a methodical, automated manner. (ScienceDaily, 2018)

Lead-user innovation

Lead-user innovation is an innovation process capitalizing on lead users: users 1) facing problematics that will be common in next months for the rest of the user base and 2) having a proven interest in finding a solution for these problematics (Hippel, 1986). These lead users have already developed concepts of solution or more: built homemade solutions. Identifying these lead users gives a huge competitive advantage as they are unique sources of latent needs insights and even direct sources for concept generation. Different techniques have been combined to spot these lead-users: snowball sampling, screening, broadcasting (Tuomela, 2013), pyramid sampling (Poetz & Hippel, 2015), or classical netnography. They are time-consuming, because they require lots of interviews and qualitative (and manual) studies in general, have low chance of success (like broadcasting), and relies mostly on self-assessment. That is where data science comes.

As mentioned previously for netnography, the active online communities are composed of non-representative users, with high level of expertise and engagement: a perfect ground for lead users. Moreover, the characterization of lead users has been subjected to several studies, from Belz & Wenke notably (2010), stating the six criteria revealing a lead-user personality concerning a product:

- Ahead of trend (having new needs)
- Dissatisfaction (towards what is available on the market)
- Use experience (repeated and intense interaction with the product)
- Product knowledge (extensive knowledge on the product, the competitors, the environment)
- Involvement (commitment with the market)
- Opinion leadership

Thus, a large data set is available on the web (activities of communities on Youtube, Facebook, Twitter or specialized blogs, forums) and research criteria are enunciated: if researchers can detect the six criteria through different algorithms, they have the possibility to manage an automatized search of lead users leveraging data science capabilities, for a fraction of the time and cost of regular lead users research. This path has been semi-explored in Sanchez, Giacalone & Goduscheit (2018) through the study of Youtube and Facebook interactions of the EDM festival Tommorowland community: they gathered 9,149 communications, manually coded and tagged them according to the criteria, and recorded them in data management software. This is a "semi-use" of data science as the leveraged data volume handling and data visualization capabilities, but not algorithmic and machine learning side. A more holistic approach (from the data science point of view) comes from Tuarob & Tucker (2015), creating a complete algorithm detecting lead user criteria among 2.1×10^9 tweets on smartphone communities (with no manual coding). The interest is in the combinations of diverse machine learning capabilities: quantifying emotions script (grading positivity and negativity of comments on 1-5 scale), feature extraction algorithm, POS tagger (Part-of-Speech tagger), iterative learning algorithm identifying product functionalities-adjectives associated-frequency in the tweets, stemming and clustering algorithm to clean data, inverse document frequency techniques. The outcome is a ranking of all users contained in the data set depending on their compliance with lead-user criteria. This example opens vast opportunities for designers in their user and research context: 1) they will identify lead users faster than ever, and can invite them to join the design process (indirect insights), 2) they will detect latent needs expressed by lead users, and the solutions they are proposing, automatically (direct insights).

2.3. Big qual studies: combining best of both worlds

To end this overview of augmented qualitative user and context research methods through data science, the focus will be on a process developed by social researchers: big qual studies.

Definition

The first concepts of big qual studies appeared in 2002 with Jennifer Mason envisioning the future of qualitative research:

> *"Perhaps the most significant opportunity offered to qualitative data is the possibility of "scaling up" through data sharing, to produce cross-contextual under- standings and explanations" (quoted by Davidson et al., 2018)*

The idea to assemble small-scale archived qualitative studies to explore a new research question was born: big qual studies. Compared to the other methodologies presented above, the large data set does not come from web communication, but from archived qualitative studies database (that may concern Internet-based studies, but also regular social researches on site, in lab etc..). It became possible thanks to the acceleration of digital, easily accessible archives of social studies material (Moody, 2017), mainly populated by textual elements (transcripts, fields notes and diaries). Secondary analysis[13] of pooled qualitative studies is the occasion to add a quantitative facet to the social results, bringing objectivity and accuracy as the different studies pooled were not lead by the same researchers (minimizing bias), covering potentially different geographical and time area (Davidson et al., 2018).

Process

Davidson et al. (2018) proposed a process composed of qualitative studies and big data techniques back and forth to reveal the richness of such big qual studies, after the definition of the research question:

- **Constitution of the corpus from archived qualitative studies browsing**: for the first step the metadatas of each study must be carefully reviewed to evaluate their relevance towards the research question (type of data, socio-demographics characterizations, initial research question, spatiotemporal locations, research planning). The selected archives are collected, gathered via data management tools, but also reorganized for the secondary analysis (e.g. classification by age of the participants or level of studies across all the archives selected)
- **Recursive surface 'thematic' mapping**: a first mapping of the themes present in the corpus is made, assisted by data science. This can be done using different algorithms: simple calculations of words frequency in the corpus, searches of them through word proximity, matches of words and a preset lexicon, emotions grading algorithm. The methods used by Davidson et al. was keyness evaluation: comparison of keywords frequency between two sets of data.
- **Preliminary and in-depth interpretive analysis**: guided by the algorithm results, investigations of the most promising piece of data (by usual -and manual- qualitative techniques), to formulate a point of view for the research question

Opportunities for designers

The process presented here is applied to social science, not specifically design practice, but there are no strict constraints limiting the use of it for user and context research. The challenge is the availability of archives for design research: the concept is less developed in design, and as designers often work for companies, the firms would probably not accept to mutualize researches ordered and paid by them. However, at the scale of large corporations, a database compiling all the research made in the group could make sense. The cost of leading qualitative studies could be amortized by the perspective of secondary, tertiary analysis: these would have the image of a general investment for today and tomorrow, more than a cost. The thematic mapping would offer quantifiable observations, supporting the insights research.

2.4. Add-on: textual data preparation and analysis

This part on qualitative studies reinterpreted through data science is referring mostly to textual material - which is coherent as this is a privileged medium for qualitative studies, and textual analysis techniques are globally more developed than visual analysis ones in data science (a matter of complexity). There are several references in this part of textual data preparation and analysis operations. To give some clarity, major textual data techniques are gathered and briefly described in the table 3.2. The algorithms can interchangeably be used to support netnography, lead user or big qual studies.

[13] "Secondary analysis of qualitative data is the use of existing data to find answers to research questions that differ from the questions asked in the original research" (Long-Sutehall, Sque & Addington-Hall, 2011)

Table 3.2: Major textual analysis operations
(adapted from: Welbers, Van Atteveldt & Benoit, 2017)

Operations	Description	Input	Output
Data preparation			
Text importation	Importing and encoding set of files to make them usable by the computing environment selected	Flat text files (CVS, TXT), formatted (HTML, XML), complex (Excel, PDF)	Readable file encoded in UTF-8
String operations	Vectorization of the text in strings to ease later manipulations (remove, transform text)	Encoded files	Vectors of strings
Preprocessing *Tokenisation*	Split of the text in words	Vectors of strings	Vectors of tokens (= words)
Preprocessing *Normalization*	Standardization of the tokens to reduce the size of the vocabulary (removing lower/upper case distinction, singular/plural …)	Vectors of tokens	Vectors of uniformed tokens
Preprocessing *Removing stop words*	Removal of "poor valued" words like "the" or predefined set to reduce file size and improve accuracy	Vectors of tokens	Vectors of tokens (reduced)
Document-term matrix	Formatting of the text as a matrix, leading to a sparse matrix, an efficient format for further computations (DT matrix)	Vectors of token	DT Matrix with documents as rows, words as columns, words frequency as cells
Filtering	Removing terms exterior to a defined range of frequency to keep only informative words	DT Matrix	DT Matrix reduced
Weighting *Term frequency- Inverse Document Frequency (TD-IDF)*	Weighting the words frequency according to the distribution of the terms in the corpus	DT Matrix	DT Matrix, with weighted cells value
Data analysis			
Dictionary	Counting of the frequency of a set of select words or concepts (the dictionary)	DT Matrix	DT Matrix, with columns restricted to the ones matching the dictionary
Supervised machine learning *Positive-Negative detection algorithm*	Autonomous coding texts as positive, neutral, negative after being trained on texts manually coded as positive or negative	DT Matrix	Metadata variable assessing if the text is positive, neutral or negative
Unsupervised machine learning *Topic modelling*	Classification of documents in a defined n number of undefined topics through pattern recognition	DT Matrix	Matrix of n columns and cells containing most frequent words associated to the topic
Text statistics *Keyness*	Comparison of words frequency in two text sets (or two subsets of a text)	DT Matrix	Plot showing words frequency compared (generally)
Advanced analysis (suited to ethnographic/social research)			
Advanced Natural Language Processing *Part-of-Speech tagging*	Tagging of terms with morpho-syntactic categories (nouns, adjectives, verbs…)	DT Matrix	Matrix with a column for term categories
Word position	Preserving the words positions after tokenization (avoiding the bag-of-words effect)	Vectors of strings	DT Matrix + Data frame containing the number of the token and the position

3. Differentiating marketing and design insights from data science

3.1. Marketing - Data Science, a more common collaboration

Throughout the early research for this study, by presenting the subject (the idea of collecting user insights for designers, integrating more quantitative methodologies), a question came frequently: "Is it marketing analytics given to designers?". This interrogation opens a reflection on how design could leverage in a unique way data science through the intention more than the straight use of available tools, which has already been mastered by marketing professionals.

Origins of confusion

The marketing world has been quickly interested in data science, because marketers could directly benefit of this new science. Marketing is a customer-oriented field aiming at creating value for customers, and for this purpose, customer behavior was studied (Lindahl & Nordin, 2010). Before the 1990s', the data mainly collected was the act of purchase, as this was the only action including a recording phase. The creation of JavaScript tags in 1993 opened the door to new data collection channels: marketers could access to the complete customer action historic in e-commerce (browsing, wishlist, time spent on pages, clicks) (Bottégal, 2016). In 2000's, growing data flow led marketers to approach data science to be able to handle and make large data sets valuable: marketing analytics was born, with the emblematic Google Analytics tool in 2005, giving user-friendly access for non-tech professionals to advanced data science technology on traffic reporting. An everyday hint of marketing analytics is cookie, a data fragment added to browser when a website is visited, to track and make personalized action for the user (Clifton, 2008).

Thus, there is a stronger knowledge and understanding of marketing – data science collaboration than design – data science one in non-specialist population thanks to marketing analytics showcase. However, marketing professionals have already explored other areas for this partnership and it is nowadays completely integrated in the industry (both in terms of digital or physical aspect).

Data science application in marketing

Advertising and contents is now tightly linked with data science. With the multiplication of promotion channels, data science is used to track and monitor the performance of every promotions, an example of marketing analytics. This is descriptive data analysis (report of what happened), the most widely spread (Cook, 2018). But there is also a growing interest for predictive analysis (to predict the outcome for a promotion, updating the prediction with real-time data) and prescriptive analysis (including prediction and giving a proposal on what should be done in reaction of the prediction to optimize an objective) (Evans & Lindner, 2012). An experiment launched with IBM's artificial intelligence agent Watson and a marketing team introduced Watson to observe the success and failures of a set of ads on different medias and recommend which visuals should be replaced by another one depending on the channel (Johnston, 2017).

Pricing have been significantly improved thanks to data science by a more granular approach, especially in B2B[14] context, allowing companies to tie a precise price for every individual deal, by compiling "decision-escalation points, incentives, performance scoring, and more, based on a set of similar win/loss deals" (Baker, Kiewell & Winkler, 2014). On the B2C[15] side, personalized pricing focuses attention of marketers: it gives opportunity to adjust the price depending on key customers and key markets (through proper price variations or personalized deals and coupons), without being in the race of absolute minimal prices (Chen & Chen, 2017).

Spending optimization between social media, call-centers, traditional advertising, store animation (and more) is now monitored by real-time data, allowing a curated cost and optimized ROI[16] compared to what was formerly decided through "gut instinct" because of the lack of data (Court et al., 2015).

Finally, **trends forecasting** is emerging to detect hot topics from data, before experts view, is an

[14] B2B : Business to Business
[15] B2C: Business to Consumer
[16] ROI : Return On Investment

opportunity to gain competitive advantage. But the use is still limited (Import.io, 2015).

As marketing was at the forefront of data science usage, there is a well-known relationship between the two. Seeing the abundance of marketing insights produced thanks to data science, it may lead to think that the marketing data of customer segmentation, business trends or competitors' benchmark are enough to embody the quantitative addition that design could benefit from data science. This vision is reinforced by the trouble frontier between marketing and design.

3.2. Differentiating Design and Marketing

The confusing interplay of design and marketing is rooted in the unclear power game between the two fields. Marketing researchers tend to place design as an affiliated department who should be closely coordinated by customer wishes: the marketing research stimulate the creative process to increase the market success (Kotler & Rath, 1984). At the extreme, design is a vector of signs reconfiguration to enhance product desirability, to add perceived value in the marketing strategy (Heilbrunn & Barré, 2012), neglecting the research of new usages, paradigm shift, radical innovations. But design cannot be reduced to this marketing lineage as design is built on proper outcomes, processes and philosophy which are fundamentally distinct of marketing theoretical basis (Beverland, Gemser & Karpen, 2017).

Axis of differentiations

There are numerous dichotomy axis between design and marketing: the difference in professional identity, the vision of corporate identity, the threshold between commercial and artistic stakes etc. The study introduces four of the differentiations axis – the first three detailed below, the last one in the next section.

The **product approach** diverges between the two fields because marketing considers the product has a vector of their primary focus point, the brand, whereas design (even more product design) gives all the attention to the product. For marketing, the construction of the brand is perceived as the cornerstone of company success, as the competitive advantage to create emotional link with customers: the product is just one artefact transmitting the brand image, language and status (Salze-Mörling & Strannegård, 2004). Caterpillar is legitimate to develop robust boots for outdoor work because it is aligned with its core business of construction machine: same meaning, same language, same image (Holm & Ulla, 2005). On the opposite, designers express themselves through the product, not the brand. They are involved in the technical development, in contact with engineers and factories (Holm & Ulla, 2005). Designers' role is to challenge the impact and innovativeness of the product, even if the market is not ready (Bloch, 1995) or if it is not completely aligned with the brand image.

The question of **value creation** draws a line between marketing and design, mainly because of a lack of understanding of the outcome of design. If marketing is full of metrics, KPI[17], analytics dashboard, the framework to pinpoint and numerically measure design impact is more complex, and is not tracked daily. Criteria to assess design effectiveness in marketing literature includes aesthetics, symbolism functionality, usability, interaction, uniformity: the evaluation grid is evolving over and over, and confusion on design impact grows. No one would deny the existence of positive effect (the review in Beverland, Gemser & Karpen, 2017, gathered literature on the success observed through design), but the bias consisting on associating marketing as an investment and design as a cost still exists (Holm & Ulla, 2005).

The debate on **ethicality** in both marketing and design is growing in parallel with the rise of awareness on mass consumption effects, extrapolating to the effect of a marketing and design overflow[18]. There is a difference of accountability: marketing is accountable to the organization of which it belongs, to create value to serve their objectives; for design, there is an obvious accountability to the organization, but also accountability to the user (Vial, 2014)[19]. Designers redefine chocolate bar design to dissuade binge eating (with commercial success) (Wilner & Ghassan, 2017) and fighting against obesity; the design community advocates for more ethical UX design in social media (which leverages psychological and behavioral

[17]KPI: Key Performance Indicator

[19] The user accountability can be included in marketing through ethical marketing, but this is a different branch of the classical marketing concepts.

researches to retain longer the user on their apps) (Gray et al., 2018), design research focuses on how to encourage sustainable behavior (Casper, Debra & Pettersen, 2017). This is in line with the responsible vision advocated by Papanek (1971):

> *"In an age of mass production when everything must be planned and designed, design has become the most powerful tool with which man shapes his tools and environments (and, by extension, society and himself). This demands high social and moral responsibility from the designer. […] Design must be an innovative, highly creative, cross-disciplinary tool responsive to the needs of men."*

Thus, marketing and design practices are not driven by the same objectives. They might use similar channels to collect user and context insights, but the settings and the integration of the insights deserve different methodologies to best serve the goal of each ones.

3.3. Designers take on insights, compared to marketing

An axis has been left: the difference of approach in market and design research. The adjacent table proposes some contrasting criteria to distinguish the philosophy of both researches (Table 3.3).

The final concept giving a unique vision of insight to designers has been already presented previously: empathy. Marketing analytics are implemented to track to picture what is happening, not why. The user's exterior motivations do not appear in the numbers.
To illustrate the different approaches, a comparison between the interpretation of a web metrics, time spent on a site, for a transportation site is a good example (Kowal, 2017). The metric can show a high time on the website. In the marketing point of view, this is classically a positive metric: high time on a website equals to high traffic (so higher revenue in general), high level of interest in the offer. However, by adopting an empathic approach to the analytics, the result can be opposed. Having a look at the context of the days, and correlating the time on site data with environmental data (to recreate the user situation) like weather and transport traffic conditions, groups of user with completely different motivations are distinguishable. With the help of machine learning algorithm for clustering (gathering users according to group of similar criteria), three types of user motivation appears: 1) in the case of short time, and bad weather conditions, the user need a quick information to have an unplanned transport solution, 2) a high time on site and high time per page in normal weather and traffic conditions, the user is highly engaged in the experience, taking time to explore the information on purpose, 3) a low time spent on the website, bad transport traffic conditions and user arriving through referral sources, it embodies a panicking user, having a problem or discovering a problem of transport (because of strikes? because of breakdown ?). Trying to be at the place of the user, empathizing with him, immerging in his context: this is how designers have a unique interpretation of data science.

Table 3.3: Market and Design research comparison (adapted from: Mindflow, 2017)

Market Research	Design Research
Opinion Based: Aims to understand what people say about products and services.	**Observation Based:** Observes what and how people interact with products and services and not what they say.
Market focus: Aims to identify potential markets which includes understanding who will buy which products and opinions about a product.	**Context focus:** User research focuses on user, his needs and environment around, how and in what circumstances users interact with product, and observes what people actually do.
Historical data: Studying and analyzing what already exists to identify issues, opportunities and gaps.	**Behavioral data**: Studying and analyzing human behavior to identify insights, unmet needs, and unidentified trends
What: Identify what people feel	**Why**: Understand why people feel a certain way
Demographic study: Buying patterns change across demographics	**Cultural study:** Usage patterns change with change in sub-cultures.
Customer goal: Which segments of people are willing to buy a defined product or service?	**User goal:** Will the product or service satisfy the needs and expectations of user?

References

Aktas, E., & Meng, Y. (2017). An exploration of big data practices in retail sector. *Logistics, 1.*

Ariely, D., & Berns, G. S. (2010). Neuromarketing: The hope and hype of neuroimaging in business. *Nature Reviews.Neuroscience, 11*(4), 284-292.

Arvidsson, A., & Caliandro, A. (2016). Brand public. *Journal of Consumer Research, 42*(5), 727-748.

Baker, W., Kiewell, D. & Winkler, G. (2014). *Using big data to make better pricing decisions.* Retrieved 25/06, 2018, from https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions

Belz Frank-Martin, & Wenke, B. (2010). Netnography as a method of lead user identification. *Creativity and Innovation Management, 19*(3), 304-313.

Beverland, M. B., Gemser, G., & Karpen, I. O. (2017). Design, consumption and marketing: Outcomes, process, philosophy and future directions. *Journal of Marketing Management, 33*(3-4), 159-172.

Bloch, P. H. (1995). Seeking the ideal form: Product design and consumer response. *Journal of Marketing, 59*(3), 16-29.

Boks, C., Lilley, D., & Pettersen, I. (2017). The future of design for sustainable behaviour, revisited., 675-689.

Bottégal, B. (2016). *Brève histoire du digital analytics.* Unpublished manuscript.

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). *The role of big data and predictive analytics in retailing* doi:https://doi.org/10.1016/j.jretai.2016.12.004

Breeze, J. (2011). *Eye tracking: Best way to test rich app usability.* Retrieved 06/26, 2018, from https://uxmag.com/articles/eye-tracking-the-best-way-to-test-rich-app-usability

Chen, B., & Chen, J. (2017). *Compete in price or service?—A study of personalized pricing and money back guarantees* doi:https://doi.org/10.1016/j.jretai.2016.12.005

Clifton, B. (2008). Advanced web metrics with google analytics.

Cook, J. (2018). *How data science will change the marketing world as we know it.* Retrieved 06/25, 2018, from https://www.marketingprofs.com/articles/2018/33404/how-data-science-will-change-the-marketing-world-as-we-know-it

Costello, L., McDermott, M., & Wallace, R. (2017). Netnography: Range of practices, misperceptions, and missed opportunities. *International Journal of Qualitative Methods, 16*(1), 1609406917700647.

Court, D., Perrey, J., Gordon, J., McGuire, T., & Spillecke, D. (2015). *Marketing & sales big data, analytics, and the future of marketing & sales* McKinsey Company.

Davidson, E., Edwards, R., Jamieson, L., & Weller, S. (2018). Big data, qualitative style: A breadth-and-depth method for working with large amounts of secondary qualitative data. *Quality {\&} Quantity,*

Dickson, B. (2017). *Unlocking the potential of eye tracking technology.* Retrieved 06/28, 2018, from https://techcrunch.com/2017/02/19/unlocking-the-potential-of-eye-tracking-technology/

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*

Evans, J. R., & Lindner, C. H. (2012). Business analytics: the next frontier for decision sciences . *Decision Science Institute,*

Forbes. (2015). *How NIKEiD is helping nike's push for greater profits* Retrieved 06/30, 2018, from https://www.forbes.com/sites/greatspeculations/2015/07/09/how-nikeid-is-helping-nikes-push-for-greater-profits/#660ca0074e65

Freach, J. (2014). *Uncommon sense: The new role of sensing in design research.* Retrieved 06/30, 2018, from https://designmind.frogdesign.com/2014/04/uncommon-sense-the-new-role-of-sensing-in-design-research/

Goucher-Lambert, K., Moss, J., & Cagan, J. (2016). Using neuroimaging to understand moral product preference judgments involving sustainability., 007T06A023.

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* Montreal QC, Canada. pp. 534:1-534:14.

Grewal, D., Roggeveen, A., & Nordfält, J. (2017). The future of retailing.*93*

Griffin, A., & Hauser, J. R. (1993). The voice of the customer. *Marketing Science, 12*(1), 1-27.

Guo, F., Ding, Y., Liu, W., Liu, C., & Zhang, X. (2016). *Can eye-tracking data be measured to assess product design?: Visual attention mechanism should be considered* doi:https://doi.org/10.1016/j.ergon.2015.12.001

Heilbrunn, Benoît.,Barré, Bertrand.,. (2012). *Le packaging*. Paris: PUF.

Holm, L. S., & Ulla, J. (2005). Marketing and design: Rivals or partners? *Design Management Review, 16*(2), 36-41.

Import.io. (2015). *Why data science is the fastest growing industry in tech right now.* Retrieved 06/20, 2018, from https://www.import.io/post/why-data-science-is-the-fastest-growing-industry-in-tech-right-now/

Johnston, M. R. (2017). *Why marketers need to think like data scientists (and how to do it).* Retrieved 06/24, 2018, from https://contentmarketinginstitute.com/2017/11/marketers-data-scientists/

Kang, S. (2016). Multi-user identification-based eye-tracking algorithm using position estimation. *Sensors (Basel, Switzerland), 17*(1), 41.

Kotler, P., & Alexander Rath, G. (1984). Design: A powerful but neglected strategic tool. *Journal of Business Strategy, 5*(2), 16-21.

Kowal, J. (2017). *Machine empathy: How machine learning can bring real insight to UX design.* Retrieved 06/20, 2018, from https://www.deducive.com/blog/2017/6/20/machine-empathy-how-machine-learning-can-bring-real-insight-to-ux-design

Kozinets, R. V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research, 39*(1), 61-72.

Kozinets,Robert V.,,. (2015). *Netnography : Redefined*. Los Angeles: Sage Publications Ltd.

Kozinets, R. V., Scaraboto, D., & Parmentier, M. (2018). Evolving netnography: How brand auto-netnography, a netnographic sensibility, and more-than-human netnography can transform your research. *Journal of Marketing Management, 34*(3-4), 231-242.

Larson, J., Bradlow, E., & Fader, P. (2005). An exploratory look at supermarket shopping paths.*22*, 395-414.

Lavalle, S., Lesser, E., Shockley, R., S. Hopkins, M., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value.*52*, 21-32.

Lexicon FT. (2018). *Definition of social media monitoring* . Retrieved 06/26, 2018, from http://lexicon.ft.com/Term?term=social-media-monitoring

Lindahl, I., & Nordin, F. (2010). The interplay of design and marketing: A general model. *Irish Journal of Management,*

Long-Sutehall, T., Sque, M., & Addington-Hall, J. (2011). Secondary analysis of qualitative data: A valuable method for exploring sensitive issues with an elusive population? *Journal of Research in Nursing, 16*(4), 335-344.

Marion Poetz, & Hippel, E. v. (2015). To innovate better, find divergent thinkers. *Harvard Business Review, 93*(6,), 26-28.

Matheson, R. (2017). *Wearable device reveals consumer emotions* . Retrieved 06/30, 2018, from http://news.mit.edu/2017/wearable-device-reveals-consumer-emotions-0712

Max, R. (2017). *12 technologies to track people.* Retrieved 06/30, 2018, from http://www.behavioranalyticsretail.com/7-technologies-to-track-people/

Moody, V. (2017). *2016: A year of great progress in qualitative data archiving and exchange.* Retrieved 06/27, 2018, from http://blog.ukdataservice.ac.uk/2016-a-year-of-great-progress-in-qualitative-data-archiving-and-exchange/

Papanek, V. J.,. (2012). *Design for the real world : Human ecology and social change*. [Chicago]: Academy Chicago Publishers.

Pollok, P., Lüttgens, D., & Piller, F. (2014). Leading edge users and latent consumer needs in electromobility: Findings from a nethnographic study of user innovation in high-tech online communities.

Ramaswamy, V. (2008). Co-creating value through customers' experiences: The nike case. *Strategy & Leadership, 36*(5), 9-14.

Reid, E., & Duffy, K. (2018). A netnographic sensibility: Developing the netnographic/social listening boundaries. *Journal of Marketing Management, 34*(3-4), 263-286.

Salzer-Mörling, M., & Strannegård, L. (2004). Silence of the brands. *European Journal of Marketing, 38*(1), 224-238.

Sanchez, V. V. S., Giacalone, D., & Goduscheit, R. C. (2018). Digital anthropology as method for lead user identification from unstructured big data. *Creativity and Innovation Management, 27*(1), 32-41.

Sanders, E. B. -., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Codesign, 4*(1), 5-18.

ScienceDaily. (2018). *Web crawler.* Retrieved 06/27, 2018, from https://www.sciencedaily.com/terms/web_crawler.htm

Tannen, R. (2007). *Eye tracking for product design research* Retrieved 06/28, 2018, from http://humanfactors.typepad.com/idsa/files/tannen_eye_tracking_for_product_design.pdf

Tuarob, S., & Tucker, C. S. (2015). Automated discovery of lead users and latent product features by mining large scale social media networks. *Journal of Mechanical Design, 137*(7), 071402-071402-11.

Tuomela, S. (2013). Lead-user identification, involvement and motivation in business-to-business context; edelläkäyttäjien tunnistaminen, osallistaminen ja motivointi yritystenvälisessä kontekstissa, 104 s. Retrieved from Aaltodoc database.

Vial, S. (2010). *Court traité du design*. Paris: Presses Universitaires de France.

Von Hippel, E. (1986). Lead users: A source of novel product concepts. *Management Science, 32*(7), 791-805.

Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). Text analysis in R. *Communication Methods and Measures, 11*(4), 245-265.

Wilner, S. J. S., & Ghassan, A. (2017). Tales of seduction and intrigue: Design as narrative agent of brand revitalisation. *Journal of Marketing Management, 33*(3-4), 173-202.

Yu, C., & Zhu, L. (2016). Product design pattern based on big data-driven scenario. *Advances in Mechanical Engineering, 8*.

Chapter 4. Case studies: application of design/data science methodologies for user and context research

1. Glossier: creating a beauty product line from blog comments

1.1. Brand presentation

Into the Gloss, a new treatment for beauty

In 2010, a Vogue assistant, Emily Weiss, opened a beauty blog: Into the Gloss. The trigger for this creation was a free space in beauty communication: between major brands (L'Oréal, Estée Lauder, Revlon) - having a highly curated and dehumanized communication purely aiming at ideally showcasing their brand's products - and personal blogging - where a single woman was presenting her beauty routine and tests, only adapted to her skin type and expectations. Into the Gloss was built to gather beauty vision and regimen of experts (models, makeup artist, editorialists) but also inspirational women in general (entrepreneurs, athletes, lawyers) (Entrepreneur, 2017). These interviews in women's bathroom, the "Top Shelf" (Figure 4.1), were giving them the opportunity to speak about beauty without suspicion of inadequacy, frivolity or vanity. More, it revealed a certain expertise by the assemblage of all the personal knowledge, and the community of readers shared and enlarged this expertise. And the statistics of community engagement are impressive:

- 1,5 million unique visitors each month (Business Insider, 2016)
- Around 100 comments under each blog posts
- 60% of readers returns almost everyday
- Top shelf interviews are long (around 15 min of reading) and the visitors are asking more; which is paradoxical in the era of tweets, GIF and fast and easy content (CNN, 2018)

Both the interviews and the reader-generated contents are data-points drawing a high-value and rich landscape of the millennial beauty expectations, and in

Figure 4.1: Extract of a Top Shelf interview (source: Into the Gloss, 2018)

Kirsty Godso, Master Trainer, Nike

SKINCARE

A huge part of my beauty routine is the fact that I'm moving a lot, and I sweat a lot. Being from the area of the world that I'm from, it's all about the glow. So, I love oils—I always use an oil as a base instead of moisturizer. At the moment I'm obsessed with mixing this Tata Harper Beautifying Face Oil with her Illuminating Moisturizer as my base.

I don't cleanse every time I work out, but I always have the Clinique Cleansing Wipes in my bag—those are really good. A lot of people react to cleansing wipes, but these smell really nice—I don't want to smell like I just poured alcohol on my face, and they don't make my face all tingly. Often I'm running from a class to a dinner or a meeting, so I can just wipe off with that. I also use this Panacea Skin Cleanser for when I'm washing my face. It's really buttery—I'm pretty lucky my skin's not reactive, because I love using that. I have a Dior Cleansing Milk that I like, and then I always moisturize with my oil. If I've been traveling a lot, I have some more intense hydrating products that I like. Moisture Surge by Clinique, and I'm obsessed with Giorgio Armani Prima Glow-On

2014, Emily Weiss used this resource to launch a product line: Glossier.

Glossier, co-creation at every stage

From 2014, Glossier products have been launched at a steady path (every 3 months in average) with a customer-centric obsession. The product development is highly based on community feedback, thoughts gathered for 8 years now, direct questions (which will be detailed in the next paragraphs). But the co-creation is also applied to the brand marketing strategy, as Emily Weiss says about her customer: "Involve her. Don't just make her *feel involved*" (Quartz, 2016). Thus, there is no ads with stars promoting the brand for juicy contract: the marketing is made by Into the Gloss readers and Glossier customers, posting photos and reviewing products on Instagram and YouTube. Capitalizing on peer-to-peer recommendations, which represents 70% of Glossier's sales and traffic, packaging and product design were created be "instagrammable" (Figure 4.2): pleasant to see with minimalist packaging and a hint of girly aesthetics, to foster the share on social media. Glossier teams believe in customers as micro-influencers who have the power to actively promote the brand through their family and friends circle, for free. Infusing co-creation in all aspects of Glossier seems a winning strategy: 600% sales growth in 2016, 1.3 million of followers on Instagram, raising of 34 million $ in 2017, international shipping on progress in 2018 (Digiday, 2017).

Figure 4.2: Glossier first set of products, launched in 2014 (source: Glossier, 2018)

1.2. Glossier own data collection channel: Into the Gloss

A media turning as a data channel

When Into the Gloss was created in 2010, it is hard to say that Emily Weiss had already built it as a data source for the brand that would be created 4 years later. The approach was probably more spontaneous than in the Nike iD platform case, but elements of the Data-Product-Data framework presented in Chapter 3.1.3 are clearly identifiable. Into the Gloss is the first "product"[20] launched to serve as a data gathering machine with two features:

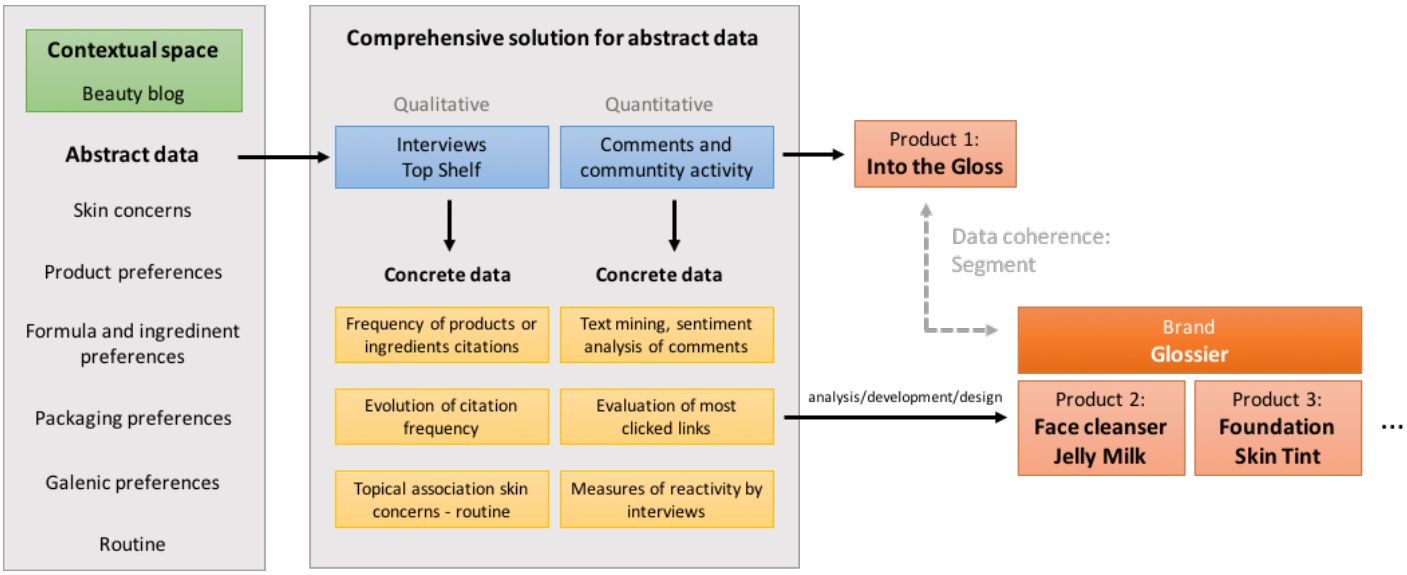
- Top Shelf interviews: by going into the bathroom, inspecting shelves of 521 women[21], taking photos and translating it in a detailed report published on the blog investigating the products, the integration of beauty in their lifestyle, their family story around it (almost an ethnographer work), this was a first strong data channel to understand women expectations on beauty products, in a more qualitative way
- Comments and community activity: the comments posted under the interviews are generally long, with a high rate of discussion between readers inside the comments - responding to a product recommended in the Tops Shelf, a reader can ask for a cheaper option to the other readers, if there is any contraindication in the use at the same type of two products, the reviews of others… With an average of 100 comments under every Top Shelf (so around 5000 comments in total), this is where the gold mine stands.

Data collection passes through these two funnels, which makes the essence of Into the Gloss. Then, there is a battery of analytics that could feed the brainstorming of designers for packaging concepts like preference of pumps instead of jar for bacterial concerns, solid perfume to adapt to millennials' mobile lifestyle, as well as chemical formulators on the galenic or ingredients people are starting to look for. The analytics includes: frequency of topics and evolution through months, sentiment analysis concerning

[20] Into the Gloss is not "a product" as tangible as personalized shoes could be in Nike ID case, but it is built with its own business model (through advertisement, like most media) and purpose.

[21] 521 between 2010 and 2017, the number surely grew during 2018

Figure 4.3: Data-Product-Data, applied to Into the Gloss and Glossier



product reviews, tracking of hyperlinks that are piquing the curiosity of readers.

With the detailed analysis of Into the Gloss, the specifications for a product appears naturally (or the team is directly asking the community, cf. next part). Glossier products always come from readers' insights: Into the Gloss is a permanent laboratory of user and context research to push new product development at Glossier. The scheme is summarized in Figure 4.3., taking the Data-Product-Data framework as a basis.

Segment: technology to reunite Into the Gloss and Glossier data

One of the challenge in data collection was the split of data between Into the Gloss and Glossier websites after the launch of the product line. If Into the Gloss remains the privileged channels for user interaction, Glossier's e-shop is gathering detailed review of shoppers on the product (example: 1206 reviews and ratings for Jelly Milk Cleanser). But tracking behaviors on two distinct websites is not possible because identification of users is different between two domains. Glossier teams paired with Segment, a data infrastructure start-up, to implement cross-domain tracking: the ability to merge two sessions of two related sites in one session through a single ID, thanks to second-party cookies (Google, 2018). With cross-domain tracking (Figure 4.4), Glossier is able "to "[compare] browsing behavior between the two groups [which] would allow Glossier to build Glossier products that fit the interests of Into the Gloss readers." (Segment, 2018). Besides, classical analytics and text mining, Glossier teams see how the readers are satisfied or disappointed from the products they partly co-created, and if it effectively triggers buy action. This point is not primary on user and context research, but it reinforces user knowledge in globality. It proves that involving users in product development is profitable: Into the Gloss readers are 40% more likely to buy Glossier than non-readers.

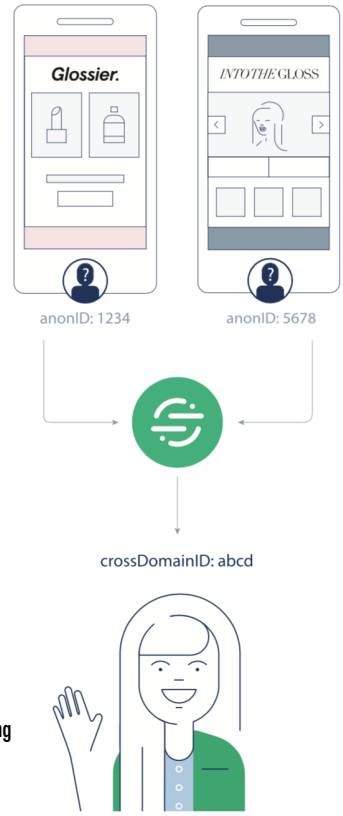


Figure 4.4: Cross-domain tracking applied to Glossier (source: Segment, 2018)

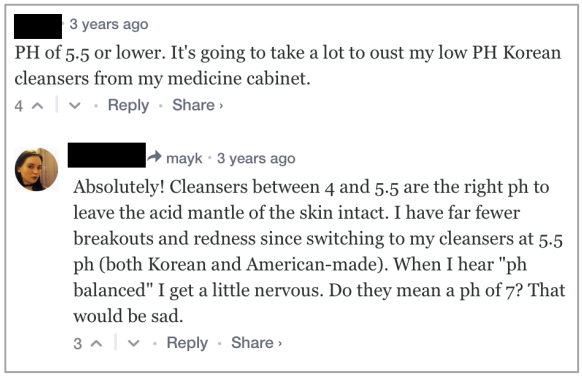
1.3. Glossier's netnography and lead user leverage

Beauty: a prolific category on the Internet

The beauty sphere is particularly active on the web: it is the second most popular category on YouTube, the famous Forbes' 30 Under 30 list highlighted five web beauty influencers in 2017 list (Forbes, 2017), e-commerce allows a panel of small brands out of beauty giants to emerge. By its high-shareability and peer-to-peer recommendations characteristics, beauty communities are bubbling with high-skilled users, and it is even possible to recognize several lead users quality in Into the Gloss community:

- **Ahead of trend and dissatisfaction:** readers often combine products together to obtain the results they are looking for and do not find
- **Use experience:** beauty products are used daily, and Into the Gloss readers demonstrates a rigorous application in their beauty routine
- **Product knowledge:** as it is easier to test multiple 12$ creams than multiple 2000$ laptops (as an example), readers can build an expertise from products comparison and develop a precise point of view (Figure 4.5)

Figure 4.5: Screenshot of comments for the Into the Gloss post asking for reader's face cleanser desire
Extremely precise expectations, supported by chemical arguments
(source: Into the Gloss, 2014a)



Thus, beauty communities like Glossier are a good choice for netnography and lead user analysis thanks to the rich content generated by community members and their high level of expertise. And it is exactly what Glossier did for its first product development.

Jelly Milk: the crowdsourced face cleanser

Glossier teams decided to leverage their community in a truly-direct way for the development of the first product, a face cleanser: launching a blog post with a single question: "What's Your Dream Face Wash?" (Into the Gloss, 2014b). 381 comments later, data scientists and designers of the team were ready to use netnography methods to dig them. With topics frequency analysis, a topic cloud can be created from the comments, to highlight the key points for user. The topic cloud shows the words the most frequently used in the comments; the bigger the word is, the more frequent it is presents in the comments (Figure 4.6).

Figure 4.6: Topic cloud based on the 381 comments answering "What's Your Dream Face Wash?"
Generated with TagCrowd

Accompanied with a qualitative analysis of comments, the most-wanted characteristics for a face cleanser were clear: balmy texture, scented-free, gentle formula, balanced pH, no oil and alcohol… This is exactly what Glossier Jelly Milk face cleanser is. Moreover, even if an "exfoliating action" were frequently requested, Glossier teams openly express on the blog why they would not include this feature (Into the Gloss, 2014a). The dialogue created for this co-created product was strong enough to precisely guide the designers and chemical formulator in the product development, ensuring a commercial success as readers had the feeling to finally found a Holy Grail thanks to a brand that truly listened to them. Data science was a vector to concretize the anecdotal impression emerging from the comment in balanced key points.

2. IDEO: data to reveal new user segmentations

The Glossier case study was focused on building a community of highly skilled users that would be able to express precisely their needs and wishes: designers would leverage data science to collect efficiently these needs. By contrast, the approach adopted by the design agency IDEO is the use of data science (combined to qualitative methods) to reveal or challenge what users are not able to express, through behaviors tracking. Two examples are presented below.

2.1. Drivers' behaviors in London

IDEO team supported the development of a data-driven service which helped city drivers to reduce their stress during their daily car commuting (Cooper Wright, 2015). In the first phase of user and context exploration, designers implemented a hybrid qualitative-quantitative methodology to define user needs. A pool of 15 London drivers had their driving behaviors tracked through different sensors installed in their car, matched with individual interviews before the driving tracking and after.

The driving data were tracked through steering wheel angle captors, break strength sensors, speed recorder, GPS locators. The data were visualized first with basic charts (bar charts showing acceleration or braking through the day), then more sophisticatedly with geographically-mapped routes combined with open data sources indicating weather and traffic conditions.

After the first round of interviews, a set of drivers were judging their driving behavior as very safe and good. But after the observation of tracked driving, this is exactly this set of over-confident drivers who had the worst measures in terms of driving: sudden peaks of acceleration/deceleration, dangerous braking moments… There was a clear dichotomy between what users say in qualitative methodology and what is observed via quantitative methods. This group of user, which was originally not a target for the new service to be designed as they apparently did not feel stress, became a priority in the product development: how fighting efficiently this excess of confidence?

In parallel, without extreme measures in driving behavior as observed for the first group, it can be difficult to detect "reasonable" nervousness by tracking mechanical features globally, and the original "wide-angled" interviews did not help. Thus, designers started to focus their analysis on driving moments in unordinary driving conditions (at night, or particularly bad weather conditions), and extracting these measures. From this point, different micro behaviors were observed showing a stress for a second group of drivers. By extrapolating, these micro behaviors may be the first signal of general nervousness in driving for these targeted users, and a second round of interviews confirmed it. The second target for the service was here.

75 hours of interviews and 20 million driving data points were gathered to identify potential users and what was their latent problematics.

2.2. The geek is not the stereotype you imagine

Similarly, IDEO team worked on the development new personal connected devices. They implemented the same type of process: interviews with 10 "tech devices" users for the qualitative side, and a classic method for quantitative analysis: a survey (Seeman, 2012). During the interviews, the team detected a surprising user: Rosaria, a 20-something paralegal juggling between her work, her family commitments notably the support she was giving for the schedule organization, and her own business of event-planning. To manage these all-day tasks, this user was equipped of numerous devices and apps to monitor her productivity, keep her schedule with minimal efforts, track the tasks. This description of an avid device buyer is far from what could be imagined of a "male tech geek" collecting smart devices in a Star Trek revival vibe. Was it an isolated case?

The high-scale survey showed that 14% of responders had a similar enthusiasm for smart devices as Rosaria. This group was mostly made of woman, seeking efficiency at all levels in their life, and having a high use of IPad (first generation). The qualitative observation meets the quantitative findings.

In these two examples, IDEO lead two distinct data collection channels, to contrast the results (in the case of driving service) or confirming them (for tech device project). Data science were used as a support of the data analysis in the case of Glossier; with IDEO, data science especially for the first example, is seen as a separate channel.

3. mPath-LEGO: emototyping to understand kids feeling

The Glossier case study shows how data science was used to text mine users' expressions; IDEO's case study demonstrates how data science helps to confront what people express/think and what is concretely observed and measured. But what can be done when the expression skills of the user are not complete, with kids for instance? Data science can also support this situation thanks to "emotional sensors tracking human feelings" and data process.

3.1. Emototyping: mixing sensors to build a deep picture from data

Elliot Hedman, a former MIT PhD student, founded the mPath start-up in 2014. His concept was to mix wearable stress sensors, analytics and eye-tracking glasses (or GoPro camera) to obtain a quantified landscape of human feeling when users are interacting with a product/a service: emototyping. The leading technology of mPath is the MOXO sensor, a bulky look-a-like smartwatch measuring skin conductance: peaks of skin conductance translate frustration, stress and dips disinterest and boredom (Chapter 3.2). Eye-tracking can catch where the user was looking at the exact moment of peaks/dips, and the algorithms developed correlate both data (Matheson, 2017).

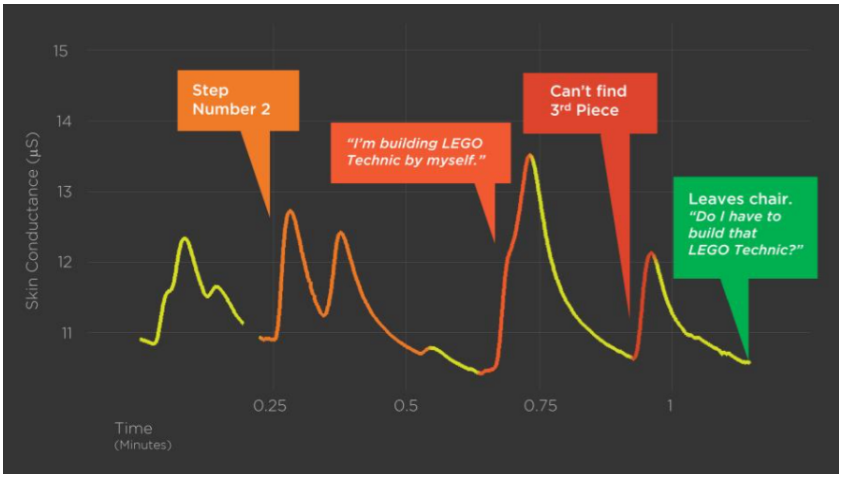
Emototyping was first develop in the context of research with children suffering from autism. MOXO sensors were created to evaluate children's stress level to alert when there was a risk of convulsing seizures due to stress. But during experiments at the hospital, Elliot Hedman observed that a kid calmly climbing a rock presented a peak of stress based on MOXO sensors, which looked paradoxical at first sight. Moreover, the kid explained that he was bored, but Hedman found an interpretation to this contradiction: to compensate his overwhelming feeling he became bored. "That was a really powerful moment for me. I knew there is so much we can do if we understand emotions better," says Hedman. He started to apply his technology to both research and commercial cases in user and context exploration, to act as a translator of emotions. An iconic implementation of emototyping was made with LEGO in 2014.

3.2. LEGO: understanding how difficult games challenge kids

In 2014, LEGO teams were working on the instructions for the LEGO Technic line: the most difficult one, with unstandardized gears, levers and even some dynamic pieces, targeting teens between 10 and 14 years old. To make these LEGOs more accessible, they wanted to include an instruction booklet, downloadable on tablet, with 3D models and animations to help the construction. However, there was no user evidence showing that this addition had an impact on users: they involved Elliot Hedman in the research to understand the emotions felt by kids during the LEGO construction (Hedman, 2014).

The set-up of the experiments includes 10 at-home interviews and game session recorded with children between 7 and 12 years old. Kids were given a LEGO Technic box; with or without tablet instructions; they would play alone first, but they had the possibility to call their parents to help them (which always happened). Hedman voluntarily chose games usually slightly too difficult for the children to see if the tablet instruction had an added-value to help them to overcome the challenge. They were equipped of MOXO sensors, of course, to track skin conductance (Figure 4.7).

Figure 4.7: Example of skin conductance graph over time, augmented with the different stage of LEGO construction reached by the kid (first and third inserts) and quotes of kid's reaction (second and fourth inserts) (source: Hedman, 2014)



Thanks to skin conductance measures, some expected behaviors could be confirmed:

- Peak of conductance, related to a peak of stress when looking for a piece or when finding out that what was built is not functioning as it should
- Conductance dips related to boredom when kid is facing several times misunderstandings, asking their parents if they can change game

But the measurements were also revealing unexpected behaviors, related to the instruction tablets and parents:

- The use of tablet instructions provokes jerky peaks of conductance when kids are looking at the 3D-rotating model: confusion and overwhelmed feeling are dominating the sequence. After 2 or 3 steps following the table instruction, kids naturally come back to paper instruction
- When parents come to help, there is an immediate decrease of conductance, even if they are facing issues: parents' presence ensures a linear conductance, traducing a serene approach toward the construction. Parents give kids the confidence to temper their emotions.

The final results of the study were not precisely about the tablet instructions, which were rapidly dropped out by kids, but the importance of parents' involvement in kids playing session with a challenging game. From this new perspective, LEGO deployed a dedicated communication intended for parents to reassess their key involvement all along the LEGO session: YouTube videos (LEGO, 2015), LEGO Technic tab on the official website to gather posts to help parents in their guide role (LEGO, 2018a), even the only picture showing humans on LEGO Technic home page is showing a father and a son together (LEGO, 2018b).

This last case study is probably the one demonstrating the most the importance of empathy when a designer deals with quantified data. A graph of skin conductance alone retains the same emotional impact as a graph of stock exchange fluctuations. But mirroring the graph with kids' gestures, quotes and stares humanized the peaks amplitude, slopes, acceleration/deceleration. Skin conductance is enrolled in a bigger story that reveals kids learning curves, their attitudes towards challenges, their trust in what they already experiment, the parental relationships. Developing designers' empathy ability is necessary to grasp user insights from any quantified, technologically-measured data.

Figure 4.8: Only picture presenting LEGO constructions with humans in situation on LEGO Technic homepage - a father and a son (source: LEGO, 2018b)

References

Business Insider. (2016). *This beauty startup has become so popular that it has 10,000 people on a waitlist for lipstick.* Retrieved 07/01, 2018, from http://www.businessinsider.fr/us/how-glossier-became-so-popular-2016-5

CNN. (2018). *Glossier CEO: Brands of the future will be 'co-created'.* Retrieved 07/01, 2018, from http://money.cnn.com/2018/03/12/news/companies/glossier-makeup-ceo/index.html

Cooper Wright, M. (2015). *Are you a good driver? how designers use data to get to the Truth.* Retrieved 07/02, 2018, from https://medium.com/design-x-data/are-you-a-good-driver-how-designers-use-data-to-get-to-the-truth-3c534fcaf9d2

Digiday. (2017). *How glossier uses data to make content and commerce work.* Retrieved 07/01, 2018, from https://digiday.com/marketing/glossier-uses-data-make-content-commerce-work/

Entrepreneur. (2017). *How glossier hacked social media to build A cult-like following.* Retrieved 07/01, 2018, from https://www.entrepreneur.com/article/298014

Forbes. (2017). *Meet the 2017 class of 30 under 30* Retrieved 07/01, 2018, from https://www.forbes.com/sites/carolinehoward/2017/01/03/meet-the-2017-class-of-30-under-30/#beef168178a8

Glossier. (2018). *Glossier, eshop homepage.* Retrieved 07/01, 2018, from https://www.glossier.com

Google. (2018). *About cross-domain tracking.* Retrieved 07/01, 2018, from https://support.google.com/analytics/answer/1033876?hl=en&ref_topic=2772342

Hedman, E. B. (2014). Thick psychophysiology for empathic design. Massachusetts Institute of Technology. Department of Architecture. Program in Media Arts and Sciences.).

Into the Gloss. (2014a). *The glossier cleanser: Where we are now.* Retrieved 07/01, 2018, from https://intothegloss.com/2015/10/glossier-cleanser-update/

Into the Gloss. (2014b). *The glossier cleanser: What's your dream face wash?.* Retrieved 07/01, 2018, from https://intothegloss.com/2015/01/emily-weiss-glossier-cleanser/

Into the Gloss. (2018). *Kirsty godo, master trainer, nike, top shelf.* Retrieved 07/01, 2018, from https://intothegloss.com/2018/06/kirsty-godso-workout/

LEGO. (2015). *Introduction for parents - LEGO MINDSTORMS EV3.* LEGO (Director). (2015, 11/17).[Video/DVD] Youtube:

LEGO. (2018a). *LEGO parents homepage.* Retrieved 07/02, 2018, from https://www.lego.com/en-us/family/make

LEGO. (2018b). *LEGO technic homepage.* Retrieved 07/03, 2018, from https://www.lego.com/en-us/themes/technic

Matheson, R. (2017). *Wearable device reveals consumer emotions* . Retrieved 06/30, 2018, from http://news.mit.edu/2017/wearable-device-reveals-consumer-emotions-0712

Quartz. (2016). *Glossier is building a multimillion-dollar millennial makeup empire with slack, instagram, and selfies.* Retrieved 07/01, 2018, from https://qz.com/847460/glossier-girls-emily-weiss-on-how-glossiers-customers-became-its-most-powerful-sales-force/

Seemann, J. (2012). *Hybrid insights: Where the quantitative meets the qualitative.* Retrieved 07/02, 2018, from https://hbr.org/product/hybrid-insights-where-the-quantitative-meets-the-qualitative/ROT182-PDF-ENG

Chapter 5. Challenges and conclusion

1. A panel of opportunities, a panel of challenges

1.1. Privacy and regulations

The thesis presented a spectrum of design-data science collaborations that could be highly beneficial for both designers and data scientists. However, there is a subject that should never be avoided when it comes to data science: privacy. Nowadays, data is an economical power and a strategic element: the purpose of the thesis is to show how designers can leverage data science to innovate better, so to give an economical advantage to the company they are working for. But behind any data points, there is a person, and the traces that he left through web browsing, through smart objects dispersed at home or at work, through sensors, should be treated as property rights (The Economist, 2010).

How can a designer have an impact in data protection when he is pairing its work with data science in user and context research? By applying human-centered design principles, keeping the person at the center of data collection. Designers must adopt the clearest communication to observed users on the means of data collection, the shareability of these data and the storage of them through time (Schwab, 2018).

Nevertheless, there is a process that cannot be directly controlled by designers: the potential exchange/resell of data to others aggregators/providers. These types of processes are generally regulated at the company-scale and data trading politics of firms are rarely built in collaboration with designers. Moreover, data warehouses and data clouds are usually operated by third-parties' companies, and resells or leakages are possible. Big qual studies presented in Chapter 3.2.3 (based on the re-use of former data collection) may be questioning: how to ensure that people (that accepted the first data collection) will accept and trust the secondary use?

Progresses are made to give back to people their property on data like GDPR legislation launched in May, 2018 (Sphere Identity, 2018). But the vigilance of designers is key to protect the primary "asset" of the job: people, future users.

1.2. Acceptance of design-data science bond in corporate culture

Another trouble spot, in the case of design for commercial purpose, is the integration of both design and data inside corporate culture. As stated in Chapter 1.1, innovation is originally associated to technology improvement. If the interest of design and data has been proved in innovation, it does not mean that large traditional companies are completely ready to onboard from the very beginning of the innovation process (the user and context research) both designers and data scientists' teams, alongside the usual engineering and marketing groups. The shift of culture is operating little by little, influenced by the start-ups ecosystem pushing and empowering their design and data teams, regularly recruiting UX designers and data analysts.

Besides, the reality of data management inside companies can be a supplementary hurdle for designers in data gathering. Logically, large firms collect larger data sets, which could be favorably leveraged by designers. But the access and sharing of data are often strictly regulated in between the different departments of a same company: for example, data sets originating from different geographical zones cannot be assembled for strategical reasons (in order to keep the competition between the zones in particular). Sometimes, for pure technical questions, like the use of two different cloud clients or two different encoding systems between two departments of a same firm, data assemblage turns as a technical exploit.

This quick overview of some design-data science challenges is not exhaustive, but the positive results of the collaboration may help to clear the path for an established association.

2. Conclusion

Over the year, the design field has taken an important role in product innovation development, proposing an original vision compared to the techno-centric focus formerly dominating innovation. The different methodologies developed by the design community for innovation are based on a common principle: the importance of understanding the user and his context. The panel of research methods to reach this knowledge is mainly composed of qualitative approaches, which could be explained by the fact that it is intrinsically easier to develop empathy (a cornerstone of designers' work) through qualitative methods than quantitative.

Widening the spectrum of approaches available to designers is important to add a new layer of insights and expertise: tools and processes deriving from data science open possibilities in this direction, especially towards quantitative methods. Apart from the user and context research field, the collaboration between designers and data scientists has already been explored: in data visualization, with the example of the design concept of affordances applied to dashboards; in design testing, with the implementation of A/B/n tests by data scientists to iteratively orient design decisions; finally, in digital development, the paired work of designers and data scientists foster idea generation, and add a "human dimension" to data like shown in Rise app development.

The combination of design and data science to develop new methods for user and research context takes several forms: the use of different channels for quantified data collection, via retail, via sensors, or via product/system especially designed for it; the reinterpretation and scaling-up of qualitative methods thanks to data science give a broader impact to netnography, lead-user innovation and big qual studies with the progress of text mining. If some of these methods are also used in marketing research, there is a distinction between the reasons motivating marketers or designers to leverage data science: marketing professionals want to know *what* is the situation, whereas designers seek for *why* the situation is happening in a certain way.

Implementation of these methodologies are demonstrated in practical case. The beauty startup Glossier created a unique channel of data collection through the blog Into the Gloss, and by applying netnography methodology, the product development team gained a detailed knowledge of expert beauty users expectations, to design the perfectly fit cosmetics. The agency IDEO uses a mixed of classical ethnographic studies and big data sets to detect certain customer segmentations. Lastly, the collaboration between LEGO and mPath shows the use ok skin conductance sensors and analytics algorithms to profile the emotions of kids, and point out the true causes of stress or serenity in game sessions, orienting the design development through a new relationship with parents, opposed to the initial plan which was a focus on instructions redesign.

To conclude, this landscape of possibilities offered from design-data science collaboration, and their concrete applications, proves that user and context research can be enriched with approaches more related to quantitative methods. For further researches, the development of image/video mining and the growing integration of sensors in daily products (notably through Internet of Things network) will surely facilitate the possibility of data collection; and designers will have the ability to pick in these new data to develop an uncharted understanding of user and context. But whatever will be the next generation of data collection, it will be essential to act ethically and keep in mind that there is an individual behind any data point.

References

Schwab, K. (2018). *What is GDPR, and why should designers care?.* Retrieved 07/05, 2018, from https://www.fastcompany.com/90171699/what-is-gdpr-and-why-should-designers-care

Sphere Identity. (2018). *GDPR and privacy by design: What developers need to know.* Retrieved 07/04, 2018, from https://medium.com/@sphereidentity/gdpr-and-privacy-by-design-what-developers-need-to-know-fa5a936da65a

The Economist. (2010). Data, data everywhere: A special report on managing information. *The Economist,* (February 2010)

Annex

Listing and classification of user and context research methods mentioned in the corpus (Chapter 1.2)

Corpus	*Qualitative methods*	*Quantitative methods*
The Field Guide to HCD, IDEO	Interviews (p.39)	
The Field Guide to HCD, IDEO	Group Interview (p.40)	
The Field Guide to HCD, IDEO		Expert interviews (p.43)
The Field Guide to HCD, IDEO	Conversation Starters (p.45)	
The Field Guide to HCD, IDEO	Extreme and Mainstreams (p.49)	
The Field Guide to HCD, IDEO	Immersion (p.52)	
The Field Guide to HCD, IDEO	Analogous Inspiration (p.53)	
The Field Guide to HCD, IDEO	Card Sort (p.57)	
The Field Guide to HCD, IDEO	Peers Observing Peers (p.60)	
The Field Guide to HCD, IDEO	Collage (p.61)	
The Field Guide to HCD, IDEO	Guided tour (p.64)	
The Field Guide to HCD, IDEO	Draw it (p.65)	
The Field Guide to HCD, IDEO		Resource Flow (p.67)
101 Design Methods, Kumar	Buzz Reports (p.22)	
101 Design Methods, Kumar	Popular Media (p.24)	
101 Design Methods, Kumar	Innovation Sourcebook (p.28)	
101 Design Methods, Kumar		Trends Experts (p.30)
101 Design Methods, Kumar		Keyword Bibliometrics (p.32)
101 Design Methods, Kumar		Ten Types of Innovation (p.34)
101 Design Methods, Kumar		Innovation Landscapes (p.36)
101 Design Methods, Kumar	Trends Matrix (p.38)	
101 Design Methods, Kumar	Convergence Map (p.40)	
101 Design Methods, Kumar	From..To Exploration (p.42)	
101 Design Methods, Kumar		Initial Opportunity Map (p.44)
101 Design Methods, Kumar		Offering-Activity-Culture Map (p.46)
101 Design Methods, Kumar		Popular Media Search (p.62)
101 Design Methods, Kumar		Publications Research (p.64)
101 Design Methods, Kumar	Eras Map (p.66)	
101 Design Methods, Kumar		Innovation Evolution Map (p.68)
101 Design Methods, Kumar	Analogous Modesl (p.70)	
101 Design Methods, Kumar		Competitors-Complementors Map (p.74)
101 Design Methods, Kumar		Industry Diagnostics (p.76)
101 Design Methods, Kumar		SWOT Analysis (p.80)
101 Design Methods, Kumar	Interest Groups discussion (p.84)	
101 Design Methods, Kumar	Research Participant Map (p.96)	
101 Design Methods, Kumar	Five Human Factors (p.102)	
101 Design Methods, Kumar	Poems (p.104)	
101 Design Methods, Kumar	Fields Visit (p.106)	
101 Design Methods, Kumar	Video Ethnography (p.108)	
101 Design Methods, Kumar	Ethnographic Interview (p.110)	
101 Design Methods, Kumar	User Pictures Interview (p.112)	
101 Design Methods, Kumar	Cultural Artifacts (p.114)	
101 Design Methods, Kumar	Image Sorting (p.116)	
101 Design Methods, Kumar	Experience Simulation (p.120)	
101 Design Methods, Kumar	Remote Research (p.124)	
101 Design Methods, Kumar	Field Activity (p.122)	
Design Methods, Curedale	Affinity diagram (p.112)	
Design Methods, Curedale	Behavorial Map (p.114)	
Design Methods, Curedale		Benchmarking (p.115)
Design Methods, Curedale		Blueprint (p.117)
Design Methods, Curedale	Bodystorming (p.119)	
Design Methods, Curedale	Boundary shifting (p.120)	
Design Methods, Curedale	Camera Journal (p.121)	
Design Methods, Curedale	Closed card sorting (p.122)	
Design Methods, Curedale	Coaching methods (p.124)	
Design Methods, Curedale	Cognitive maps (p.126)	
Design Methods, Curedale		Cognitive task analysis (p.128)

Corpus	*Qualitative methods*	*Quantitative methods*
Design Methods, Curedale	Collage (p.130)	
Design Methods, Curedale	Communication map (p.132)	
Design Methods, Curedale	Cradle to cradle (p.134)	
Design Methods, Curedale	Cultural Immersion (p.136)	
Design Methods, Curedale	Cultural inventory (p.137)	
Design Methods, Curedale	Cultural probes (p.138)	
Design Methods, Curedale	Customer experience audit (p.140)	
Design Methods, Curedale	Customer experience map (p.140)	
Design Methods, Curedale	Day in the life (p.144)	
Design Methods, Curedale	Diary study (p.145)	
Design Methods, Curedale	Day experience method (p.146)	
Design Methods, Curedale	Design workshop (p.148)	
Design Methods, Curedale	Workshop: creative toolkits (p.150)	
Design Methods, Curedale	Design charette (p.152)	
Design Methods, Curedale	1.5 day mini charette (p.153)	
Design Methods, Curedale	2.0 day design charette (p.154)	
Design Methods, Curedale	4.0 day architectural charette (p.155)	
Design Methods, Curedale	635 method design charette (p.156)	
Design Methods, Curedale	Historical methods (p.186)	
Design Methods, Curedale	Interview methods (p.188)	
Design Methods, Curedale	Interview: contextual inquiry (p.190)	
Design Methods, Curedale	Interview: contextual interviews (p.191)	
Design Methods, Curedale	Interview: contextual laddering (p.192)	
Design Methods, Curedale	Interview : conservative cards (p.193)	
Design Methods, Curedale	Interview emotion cards (p.194)	
Design Methods, Curedale	Interview: email (p.196)	
Design Methods, Curedale	Interview: extreme user (p.198)	
Design Methods, Curedale	Interview: group (p.199)	
Design Methods, Curedale	Interview: guided storytelling (p.200)	
Design Methods, Curedale	Interview: man in the street (p.202)	
Design Methods, Curedale	Interview one-on-one (p.204)	
Design Methods, Curedale	Interview: structured (p.205)	
Design Methods, Curedale	Interview: unstructured (p.206)	
Design Methods, Curedale	Interview: telephone (p.208)	
Design Methods, Curedale	Longitudinal analysis (p.209)	
Design Methods, Curedale	Magic thing (p.210)	
Design Methods, Curedale		Market segmentation (p.211)
Design Methods, Curedale	Mobile Ethnography (p.212)	
Design Methods, Curedale	Mobile diary study (p.214)	
Design Methods, Curedale	Mystery shopper (p.216)	
Design Methods, Curedale	Network map (p.217)	
Design Methods, Curedale	Observation (p.219)	
Design Methods, Curedale	Open card sort (p.220)	
Design Methods, Curedale	Personal inventory (p.222)	
Design Methods, Curedale	Personas (p.224)	
Design Methods, Curedale	Picture cards (p.226)	
Design Methods, Curedale	Pugh's matrix	
Design Methods, Curedale		Questionnaires (p.229)
Design Methods, Curedale		Remote evaluation (p.230)
Design Methods, Curedale	Shadowing (p.232)	
Design Methods, Curedale	Stakeholder map (p.234)	
Design Methods, Curedale	Storyboards (p.236)	
Design Methods, Curedale		Surveys (p.238)
Design Methods, Curedale		Swimlanes (p.240)
Design Methods, Curedale	Talk-out loud protocol (p.241)	
Design Methods, Curedale	Taxonomies (p.242)	
Design Methods, Curedale	Through other eyes (p.243)	
Design Methods, Curedale	Unfocus group (p.244)	
Design Methods, Curedale	Teachback (p.246)	
Design Methods, Curedale	Think out loud protocol (p.247)	
Design Methods, Curedale	Triangulation (p.248)	
Design Methods, Curedale	Wizard of oz (p.250)	
Service Design Practical, Moritz		Benchmarking (p.186)
Service Design Practical, Moritz		Client segmentation (p.186)
Service Design Practical, Moritz	Context analysis (p.187)	
Service Design Practical, Moritz	Contextual interviews (p.187)	
Service Design Practical, Moritz	Contextual inquiry (p.188)	
Service Design Practical, Moritz	Critical Incident Technique (p.188)	
Service Design Practical, Moritz		Ecology map (p.189)
Service Design Practical, Moritz	Ethnography (p.189)	

Corpus	*Qualitative methods*	*Quantitative methods*
Service Design Practical, Moritz	Experience test (p.190)	
Service Design Practical, Moritz		Expert interviews (p.190)
Service Design Practical, Moritz	Focus group (p.191)	
Service Design Practical, Moritz	Gap analysis (p.191)	
Service Design Practical, Moritz	Historical analysis (p.192)	
Service Design Practical, Moritz	Inconvenience analysis (p.192)	
Service Design Practical, Moritz	Interview (p.193)	
Service Design Practical, Moritz		Market segmentation (p.193)
Service Design Practical, Moritz	Mystery shopper (p.194)	
Service Design Practical, Moritz	Net scouting (p.194)	
Service Design Practical, Moritz	Observation (p.195)	
Service Design Practical, Moritz	Probes (p.195)	
Service Design Practical, Moritz		Reading (p.196)
Service Design Practical, Moritz	Service status (p.196)	
Service Design Practical, Moritz	Shadowing (p.197)	
Service Design Practical, Moritz	Thinking aloud (p.197)	
Service Design Practical, Moritz	Trend scouting (p.198)	
Service Design Practical, Moritz		User surveys (p.198)
Service Design Practical, Moritz	5 why's (p.199)	
Service Design Practical, Moritz		Tested and tried components (p.200)
Service Design Practical, Moritz	Inspirational specialists (p.201)	
Research for Designers, Muratovski	Case studies (p.106)	
Research for Designers, Muratovski	Ethnographic research (p.110) - Structured interviews	
Research for Designers, Muratovski	Ethnographic research (p.110) - Semi-structured interviews	
Research for Designers, Muratovski	Ethnographic research (p.110) - In-depth interviews	
Research for Designers, Muratovski	Ethnographic research (p.110) - Focus group	
Research for Designers, Muratovski	Ethnographic research (p.110) - Oral History	
Research for Designers, Muratovski	Ethnographic research (p.110) - Observing People	
Research for Designers, Muratovski	Ethnographic research (p.110) - Being with People	
Research for Designers, Muratovski	Ethnographic research (p.110) - Cutural Probe	
Research for Designers, Muratovski	Phenomenolgy (p.141) - Oral History	
Research for Designers, Muratovski	Phenomenolgy (p.141) - In-depth interviews	
Research for Designers, Muratovski	Historical research (p.155)	
Research for Designers, Muratovski	Grounded theory (p.163)	
Research for Designers, Muratovski		Surveys (p.177)
Research for Designers, Muratovski		User-Centered Design Research (p.201) - Usability Test Lab
Research for Designers, Muratovski		User-Centered Design Research (p.201) - Electronic Group Lab
Research for Designers, Muratovski		User-Centered Design Research (p.201) - Prototype Design Lab