Data Literacy in the Digital Age: Experience Design for a Workplace Learning Solution

by

Oladipupo J. Doherty

B.S Mechanical Engineering, University of Virginia, 2013

Submitted to the Integrated Design & Management Program

in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Author .....
Integrated Design & Management Program
May 15, 2020

Certified by.....
Tony Hu
Senior Lecturer and Academic Director, Integrated Design & Management Program
Thesis Supervisor

Accepted by.....
Matthew Kressy
Founding Director, Integrated Design and Management Program

Data Literacy in the Digital Age: Service Design for a Workplace Learning Solution

by

Oladipupo J. Doherty

Submitted to the Integrated Design & Management Program on May 15, 2020, in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Management

Abstract

It is estimated that 2.5 Quintillion bytes of data is produced everyday[1]. Majority of this data originates from many business processes rapidly migrating to cloud solutions. Although this is an economic benefit to the organizations that are making this transition, many of the internal teams at the forefront of this change are ill-equipped to understand, analyze, and take advantage of these business insights; in other words there exists a deficiency in data-literacy.

Our motivation was to explore the impact of this digital phenomenon on human productivity across 3 major fronts: traditional customer service channels, brand/crowd-generated instructional content, and peer networks. A user research survey was conducted to better understand the problem space, and an inference was made to address the issues surrounding data-literacy in the workplace. This concluded with a service framework, which was conceptualized to provide a starting point for organizations to begin this cultural shift, and empower their employees with the skills necessary to take full advantage of its vast data resource.

Thesis Supervisor: Tony Hu
Title: Senior Lecturer and Academic Director, Integrated Design & Management Program

[1] "Every Day Big Data Statistics - 2.5 Quintillion Bytes of ... - Dihuni." 10 Apr. 2020, https://www.dihuni.com/2020/04/10/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/. Accessed 13 May. 2020.

Acknowledgements

This thesis would not have been possible without the immeasurable contributions from my family, and their steadfast assurance in my ability to change the world. To Mum and Dad, thank you for all the sacrifices you've made to make me the man I am today. To my future wife Dami, the inspiration for this work is as much mine as it's yours. And to my brothers and sister, who have been a sounding board to all my crazy ideas.

Learning rarely happens in silos, and I have to thank Tony for sharing his experience, guidance, and support on this project. I'd also like to thank Matt, Andy, and the rest of the IDM family for believing in me and giving me the opportunity to contribute, be part of the program, and contribute to the MIT community. You are all family to me.

Lastly, thank you to every member of the class of '18 for making my two years such a welcoming experience.

Table Of Contents

List Of Figures

List Of Tables

1. Introduction

In 2007 Steve jobs introduced the world to the iPhone. This was a polarizing experience, with many citing how it couldn't possibly change the established norms at the time. To put things into context then the market was headed by players like Nokia, Blackberry, and Motorola, who defined what a smartphone was back then.

Over the decade hereafter, the iphone and the android would consequently recreate what the idea of a smartphone would be. A flat touch screen interface, with voice and text communication features, an internet connection, and most importantly, an app store.

Beyond anything else, these app stores empowered developers to create apps. These are downloadable programs that achieve a core user job. The ability to create impact at scale gradually increased as adoption levels of the smartphone grew. And in 2020 out of 7.2 billion humans on the planet we have 4.4 billion smartphones in the world. This has reshaped the dynamics of business and opened up the market to more granular ways of approaching the end user.

Brands have also leveraged this new device type in many ways. Many hotels, airline companies, retail chains and various other industries have sought to find a unique means of reaching their

customer, deepening their engagement with their brand, and providing support. Thus, the relationship with the consumer has fundamentally changed.

Many established platforms have recognized this, and are already adjusting to the times. Facebook for example, has created various tools across their ecosystem of platforms (Instagram, Whatsapp, and Messenger) to leverage their reach in giving brands and businesses the ability to reach their customers. These can be through Whatsapp business storefronts, or through commerce capabilities on facebook pages. As of 2019 Facebook has 60 million pages, with 6 million of them paying advertisers. Similarly Instagram has 25 million business pages, with 50% of them not owning a website.

At the core of this paper is the hypothesis that given the limited bandwidth of the consumers in downloading multiple apps to cater to different brand issue resolution needs, can we device a centralized system for the customer, that can act as an interface between the various brands used by a consumer and the Issue Resolution (IR) teams of the respective brands themselves?

1.1 Hypothesis

Naturally, with the download rate of apps per user about 50 apps per year, it creates a unique opportunity for discoverability, but a narrow chance at building a strong relationship/ bond with the user. It is with this high bar in creating an engaging experience with the user that sets the expectations for the user very high. This has led other applications to provide tangential utility that is outside of their core product offerings.

There exists an opportunity in the value chain between a consumer and brand whereby information on brands could be bundled through data collection, and then simplified in a user friendly format.

1.2 Research Methodology

1.2.1 Quantitative Methods

1. Google Trends Analysis: We utilized Google's trend analysis platform to better understand keywords around the expression of some of the issues we encountered in our research, and their difference in popularity over time.

2. Research Publications: Research tools like Statista and Pitchbook allow us to get a customer and market centric understanding of how consumers perceive these opportunity spaces, how it fits into their current needs, the potential market opportunity, and the feasibility of this process.

1.2.2 Qualitative Methods

1. 1-1 Interviews: 7 interviews were conducted in person, and remotely to understand the 3 major domains to explore in understanding how users go about solving issues they have with brands they use.

2. Demographic Surveys: After the 1-1 interviews, subjects were given a survey to complete. This survey captured their general biological information, demographic data, language preferences, academic achievements to cross reference with the qualitative data to elicit any correlations. This method was constantly used throughout the product development process.

1.3 Subject Demographics

Many of our interview respondents were between the ages of 20 - 30 years, and are current graduate students of Massachusetts Institute of Technology.

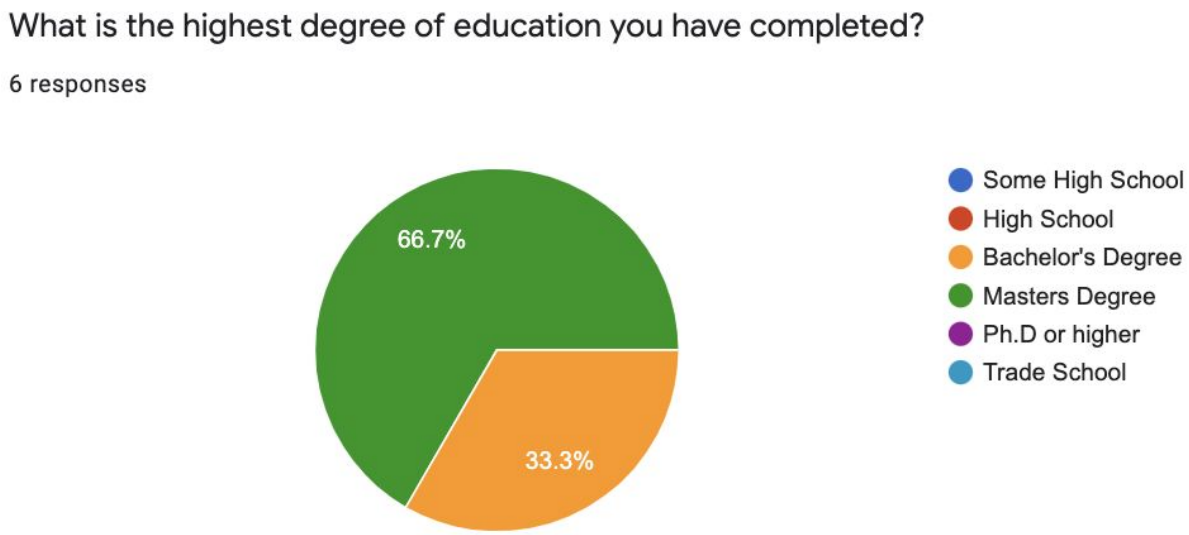

Figure 1.1: Share of respondents based of educational milestones

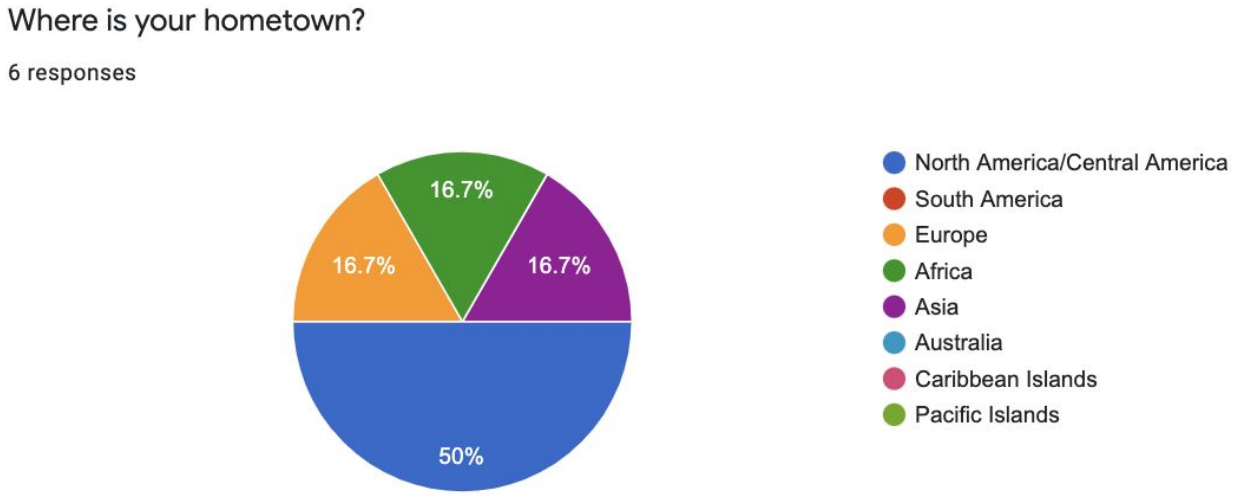

Figure 1.2: Share of respondents based on hometowns

Building on this and research, there are 3 major avenues people have explored is resolving issues they experience with products and services in their daily lives:

1. Customer Service
2. Instructional Content
3. Peer Network

These avenues were examined with our interview subjects to elicit qualitative information about assumptions they made regarding these channels, and current shortcomings, and the role their smartphones play into this process.

2. Analysis of Interviews

2. 1 Customer Support via Brand Resources

1. Calling as the last resort : When being asked what kind of channels they prefer, people cited digital methods to be superior to other more engaging channels. Email and Chat were the top methods of getting things done primarily because of the convenience of not waiting on the phone. There was a correlation between the urgency/ magnitude of the problem being experienced, and the preferred channel. For urgent issues most respondents felt it was better to call, and get a human intervention. Issues of this sort centered around returns for high value items, complex problems involving peculiar circumstances, and time pressing matters. Other less pressing matters, like minor technical problems with consumer electronics, and return of low value items, were usually approached via channels with longer perceived turn around times.

2. Google or Websites for information: When looking up customer support services, it was clear that people we either going to google to search for the customer information, or going directly to the brand's website to look up specific means to reach out to the CS department

3. Retail, Banking & Travel: There are 3 major industries that emerged out of this research that seemed to be the ones frequently engaged with by our interview subjects: Retail, Banking, and Travel. It is clear. When further asked about the reasons many cited instances like figuring out return policies, price comparisons, cancellation policies, or adjustment of a service or product purchased (changing ticket dates in the case of travel). It's also important to note that since the majority of our subjects were college students, some of our subjects that are still in the workforce had a radically different perspective, with one citing housing facility management as a major point of CS interaction. An assumption to make is that depending on the socio-economic situation, this layout of industry affinity might look different. More research would need to be done to validate this hypothesis, but it is safe to say the majority of young professionals will fall into this category.

2.2 Instructional Content

1) Automotive & Software Education: Many of our subjects cited most of their instructional content consumption centered around either learning highly specialized software skills for workplace productivity, or troubleshooting issues they have with their vehicles. These are very technical skills. Some software cited were Tableau, Excel, Photoshop, Sketch, etc. A very crucial thing to note here is how many of these tools have been slowly adopted into the workplace as technology is rapidly changing the landscape of business processes.

2) Immediacy of the situation: Many subjects also stated that when searching/ consuming these videos, speed is of the essence.

3) Standardization in format: During research, many subjects made comments around the quality of instructional content they came across.

    a. Convenience in annotation: Videos with text annotations, and time marked with respect to the beginning of the steps helped users to navigate through the content more conveniently, and allowed people who couldn't keep up to playback content, without estimating where the specific time of the step began.

    b. Short-form video content: Respondents commented on the length of instructional content they encountered playing a major role in how they perceive it’s quality. It was clear that many respondents preferred short content that was easy to digest

    c. Supplementary information: If additional troubleshooting might be required, which can be the case with highly technical products, having a reference guide was strongly preferred by our subjects.

4) Heuristic Learning System: The ideal solution subjects kept making reference to was that well designed software should have no manual to understand its use. When it comes to specialty software that's not usually the case. We find that many work environments have walkarounds, like powerpoints, PDFs, or human walkthroughs made by colleagues to ramp up new employees to existing business processes, but many of these are not personalizable, and would usually require constant human intervention if the form of updating the content and giving the user additional guidance.

2.3 On reaching out to network for assistance

1) Work Environments: The workplace seemed to be a popular environment for reaching out to peer networks. There was also an element of hierarchy in this process, with initial questions being asked to subordinates or same-level members of the company. These revolved around gaining clarity on certain internal processes to complete a task.

2) Social Stigma & Lack of Knowledge: A unique insight from our interview was how people perceived bringing up some issues they faced due to the concern of being labelled as "boring". People cited that in casual conversation they don't want to be seen as being uninteresting, and as such would be less likely to bring up these kinds of conversations.

3) Another interesting observation was the general lack of knowledge in peer-group expertise. Many people could not figure out which friends were potential knowledge experts in using products or brands they might have issues with. This might not be the case with workplace environments, as designations have been made with people assigned to departments, and

4) Power Friends: Out of our 7 respondents, 2 mentioned that they enjoy connecting people in their network with mutual friends that potentially share the same interests. These connections are thought about contextually, and were motivated either by professional career opportunities, dating opportunities, or through shared interests.

2.4 Affinity Diagram of Interview Voices

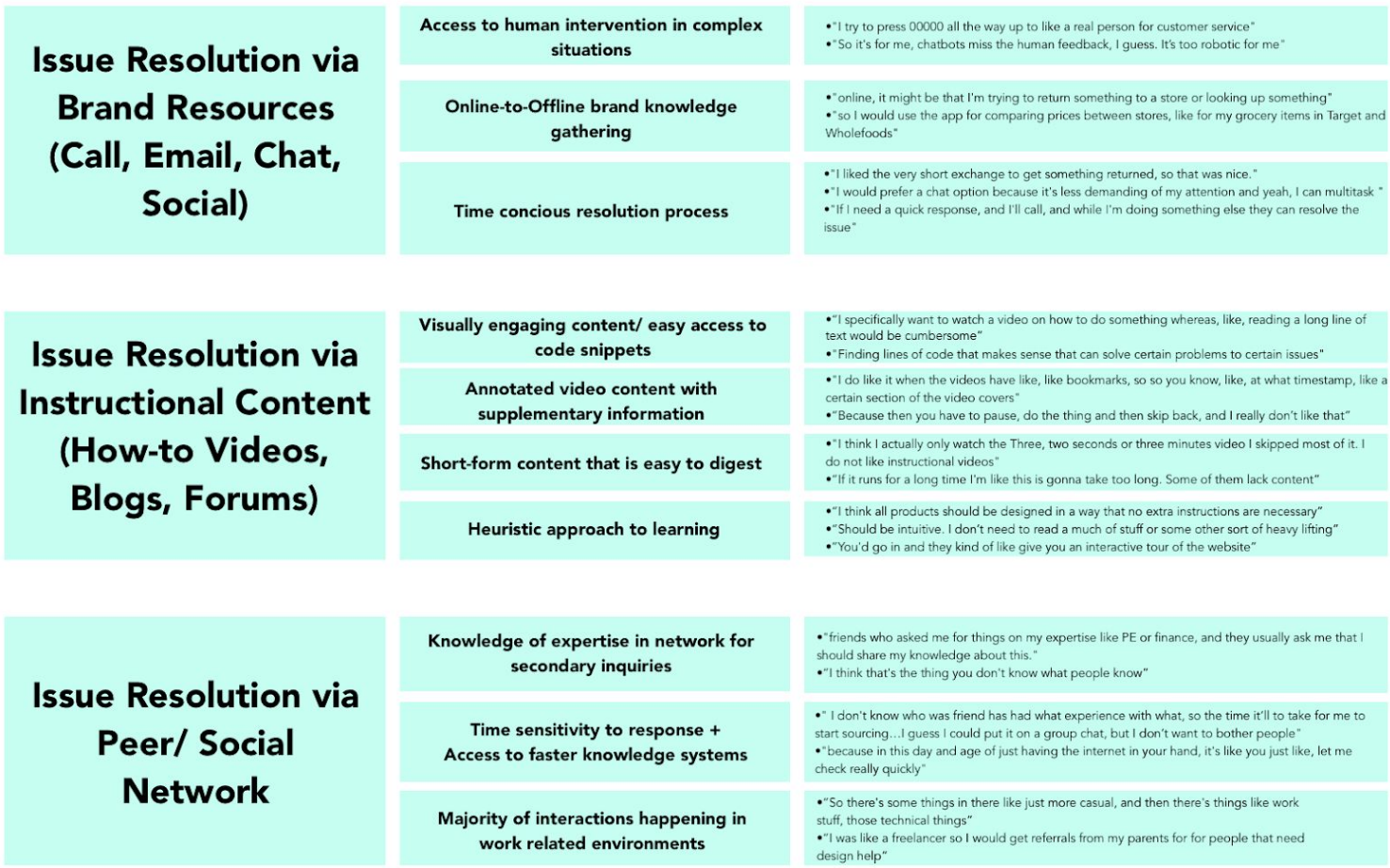

Figure 2.1: Affinity diagram of respondent's voices based on 1-1 interviews

The Affinity diagram above represents the findings from the research in a graphical manner. From left to right, the diagram breaks down the subject areas explored into summaries of the respondent voices heard during the interview, these break down into a series of actual statements made by the customer. By doing this we create a mapping of high and low points in the experience that customers remarked on. This provided us with insights we used to hone down on a problem area to focus on.

2.5 Down-selection Process

Based on the research done, we narrowed down our decision on the problem to focus on based off 3 criteria:

1) Frequency of User Encounters: This is an important measure of how eager the user is in finding a solution to the problem, and the potential for it being top of mind for them.

2) Market Survey of Customer Expenditure: This tests out the viability of the problem space in terms of how much customers are willing to spend in alleviating the problem.

3) Perceived Feasibility of Solution: This would be judged by ranking the complexity of implementation of the top 10 features in the respective category between 1 to 5, with 1

being the least difficult. The average score will be used as a metric to determine which category would be the most feasible.

2.5.1 Brand Channels

1) Frequency of User Encounters: From our interviews this varies between 1-5 times on a monthly basis.

2) Market Survey of Customer Expenditure: In a Forbes article in 2017, the CEO and Founder of Helpshift, Abinash Trip estimates that the customer service industry is approximately a $350 billion per year opportunity[2].

3) Perceived Feasibility of Solution: Capterra, a renowned software reviews website, currently lists over 200 software solutions[3] that cover the customer service industry.

[2] "Customer Service Is A $350 Billion Industry, And It's A Mess." 25 Sep. 2017, https://www.forbes.com/sites/blakemorgan/2017/09/25/customer-service-is-a-350b-industry-and-its-a-mess/. Accessed 13 Apr. 2020.
[3] "Best Help Desk Software | 2020 Reviews of the ... - Capterra." https://www.capterra.com/help-desk-software/. Accessed 13 Apr. 2020.

Brand Channels (Helpdesk)	Complexity Score (1-5)
Automated Routing	2
Customizable Branding	2
Document Storage	3
Email Integration	3
IT Asset Management	3
Interaction Tracking	3
Knowledge Base Management	3
Multi-Channel Communication	4
Network Monitoring	4
Real-time Chat	2
Average Complexity Score	2.9

Table 2.1: Complexity Score Chart for Brand Channel Software Features

2.5.2 Instructional Content

1) Frequency of User Encounters: Interviews conducted revealed that subjects experienced a need for instructional content 1-2 times on a weekly basis.

2) Market Survey of Customer Expenditure: Instructional content comes in the form of slideshows, videos, and walkthroughs. These are distributed across various platforms. The most popular that were made mention of are youtube videos. All these fall into the bucket of e-learning tools, outside brand-controlled customer service channels, used to advance user's skills and knowhow of a service. A recent report from Training Industry inc. mentioned workplace training is a $166 Billion industry[4].

3) Perceived Feasibility of Solution: The offerings in this space are vast, and based on Capterra many of the software tolls are categorized into the following segments, with the current number of offerings.

[4] "• Workplace training: market size North America 2018 | Statista." 12 Sep. 2019, https://www.statista.com/statistics/738412/size-of-the-workplace-training-market-north-america/. Accessed 13 Apr. 2020.

Instructional Content (via Learning Management Systems)	Complexity Score (1-5)
Asynchronous Learning	4
Blended Learning	5
Built-in Course Authoring	4
Gamification	4
Learner Portal	3
Mobile Learning	3
SCORM Compliance	3
Synchronous Learning	3
Video Conferencing	3
eCommerce	4
Average Complexity Score	3.6

Table 2.2: Complexity Score Chart for LMS Features

2.5.3 Peer Networks

4) Frequency of User Encounters: Majority of our respondents cited that they rarely have a need to ask their peers or social network for assistance. However, two of our respondents did cite that they do reach out to their network, albeit in a work environment. This happens on a monthly basis and involves questions around understanding business processes or seeking the intervention of an experienced employee regarding a specific task.

5) Market Survey of Customer Expenditure: Crowdsourced customer service has been around for a while. Services like Turbotax utilize the wisdom of their expert users on their platform to recommend tax strategies to their users.

6) Perceived Feasibility of Solution: Asides TurboTax, many startups are exploring the opportunity in building a peer network of power Directly, a Silicon-valley based startup, currently provides their clients a network of power users of their products and services, alongside machine learning capabilities to improve their customer service experience[5].

[5] "Directly raises $20 million for customer service powered by ...." 10 Apr. 2018, https://venturebeat.com/2018/04/10/directly-gets-20m-to-offer-customer-service-with-crowd-powered-ai/. Accessed 14 Apr. 2020.

Peer Network (Social Networking)	Complexity Score(1-5)
Activity/News Feed	3
Event Management	2
Group Management	3
Media Library	2
Real-time Chat	2
Social Media Integration	2
Social Media Tagging	2
Average Complexity Score	2.3

Table 2.3: Complexity Score Chart for Peer Networking Software Features

2.6 Time constraints, Software Learnability, and Data Abundance in the Workplace

An insight to highlight here is the prevalence of peer network and instructional content use in the workplace. Given the massive opportunity, frequency of issues encountered, and the various solutions that exist in the market, the problem space of focus would be providing specialized software reskilling and upskilling in the corporate sector. This appears in the form of frequent

questions regarding specialty software in the workplace. Examples of these include Microsoft Excel, Tableau, Structured Query Language (SQL), and Zendesk. A parallel that can be made here is the reliance of these software on Data.

The concern of time also played a major role across the board. Regardless of the nature of the situation, our subjects made a strong emphasis on the time to resolution (in the case of troubleshooting issues raised with the brand/ service) and time to acquisition (in the case of learning how to perform a new task on a new service).

3. The Impact of Data literacy in the Workplace

3.1 Importance of Data

From our interview findings, it is evident that a niche problem area exists around learnability of data-centric software systems within the workplace. Taking a look at the trend in rapid workplace digitisation, it is no surprise that we see this as a growing need for the global workforce in the near future.

In 2012, analysts estimated that 90% of the world's data was created within the previous two years (Vesset et al.,2014) . Majority of this data is unstructured, and comes in the form of documents, social media signals, and other sources not residing in conventional data storage systems.

With this phenomenon comes the growing need for analysis and inference of data. Professionals across every industry are encountering unique and complex data tools, and given the speed at which the digital industry is shifting, it is difficult to assume that these technical skills are understood by the majority of employees. What is more concerning about this trend are the numbers of leaders, managers, and decision makers in the workplace that do not feel comfortable using data-driven tools.

3.2 Data Literacy in Organisational Productivity

It is clear that the rapid growth rate of data usage has outpaced the skills and confidence of employees required to use it. This was evident in 2018 when Qlik, a data platform company, conducted a study on over 7,000 business decision-makers (junior managers and above) across the US, Europe, and Asia. The company found that only 24% of them were confident in their ability to read, analyze and work with data[6]. This skill is popularly called data literacy. Qlik further discovered that 32% of the C-suite professionals surveyed were data literate. These shortcomings are having great repercussions in holding decisions back from being data-driven; decisions that can give companies unique business insights and a competitive advantage.

Even in situations where C-suite competencies are present, the burden rests on the data science teams to assist other verticals within the organisation in fetching and analyzing the necessary data required to gain insights. This takes away from valuable time the data science teams can contribute to mission critical needs of the business, and constrains managers downstream that are unable to act on the insights in an agile manner. It is estimated that this data-induced procrastination results in organisations losing over 5 working days (43 hours) of productivity per

[6] "Lead with Data™ - The Data Literacy Project." https://thedataliteracyproject.org/files/downloads/Qlik%20-%20How%20to%20drive%20Data%20Literacy%20within%20the%20Enterprise_October%202018.pdf. Accessed 18 Apr. 2020.

employee per year (Accenture, 2020). The dollar cost equivalent has been charted by demographic below.

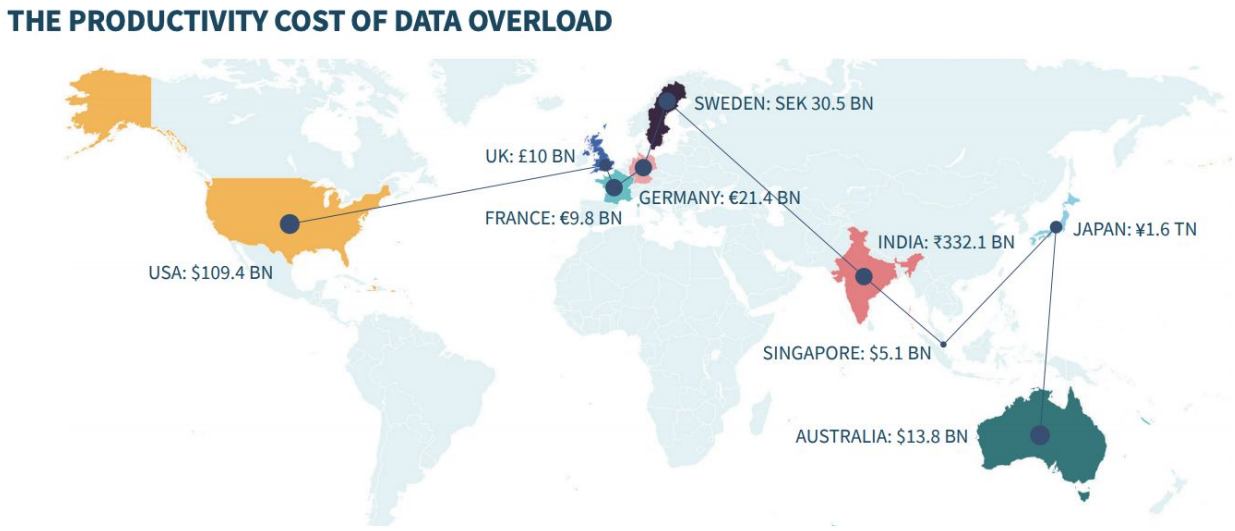

Figure 3.1: Global Productivity loss (annual) across demographics

3.3 Data Literacy and Human Capital

The problem of data literacy goes beyond just holding teams back from innovation and business impact; it also puts the workforce at risk from falling behind in career development. Rapid digitization of work has led to an increase in data generation, but also carries the potential for work to be comprehended and automated by machines. Many digitally manual tasks that existed

in the recent past are now being automated. These automation tools rely on vast amounts of data to intelligently complete these tasks; the same data that the workforce needs to embrace to gain a competitive edge in the workplace.

Deloitte Insights (2019) discusses this new skill demand trend eloquently in their Human capital trends report:

"The use of artificial intelligence (AI), cognitive technologies, and robotics to automate and augment work is on the rise, prompting the redesign of jobs in a growing number of domains. The jobs of today are more machine-powered and data-driven than in the past, and they also require more human skills in problem-solving, communication, interpretation, and design. As machines take over repeatable tasks and the work people do becomes less routine, many jobs will rapidly evolve into what we call "superjobs"—the newest job category that changes the landscape of how organisations think about work."

This surge of superjobs will cause undesirable shocks in several industries,especially from the perspective of the regular employee,who is currently experiencing a revival of automation anxiety (Akst, 2013). It is never easy for employees to change their habits, and adopt new tools in their workflow. In addition, the traditional method of asking colleagues in the workplace for advice in our findings is a current walkaround, but it is not scalable, especially given how distributed the workforce of the future will be. There are several losses in productivity from

time-zone differences, and employees working on resolving issues for several colleagues while missing the opportunity to progress on their tasks.

It is clear that the motivation to adopt digital skills is present in the workforce. The question arises around the current tools being used, and where they fall short of assisting employees while working in real-time.

4. Findings on Best Practices

4.1 The Current Solution Space

Firstly, it is important to acknowledge that outside the pain-points noted earlier employees experience in adopting new data-centric tools, many companies are making efforts to provide resources to alleviate this. These efforts can be bifurcated into two distinct paths - internal efforts, and external efforts. When Mimeo(2018), a print and digital content distribution company, created their annual report on the state Learning and Development (L&D) in the United States, they discovered that the majority (46%) of L&D teams are internally cultivated, while 54% of teams are outsourced through training consultancies.

These internal teams come in the form of digital academies that cater to all employees across the organisation's departments on building these skills. Companies like Google, Adobe, Guardian, and Bloomberg have nurtured data-literacy training resources for their employees, and this has even helped reduce employee turnover, and minimize the acquisition costs of new talent into the company by upskilling their current workforce. A good example of this initiative can be found at Airbnb's Data Science University.

4.2 Internal Organisation Efforts

Founded in 2016, the program offers a three tier curriculum that has been scaled out to their over 30 offices globally. These courses are completed through a series of in-person training sessions. The classes cover three important categories of data literacy - Data Access, Data Education, and Data Tools. The most appealing aspect of this program is in the technique used to measure success. Airbnb relies on Kirkpatrick's Model (1976). This model of training evaluation is divided into 4 parts - *Reaction, Learning, Behavior,* and *Results.*

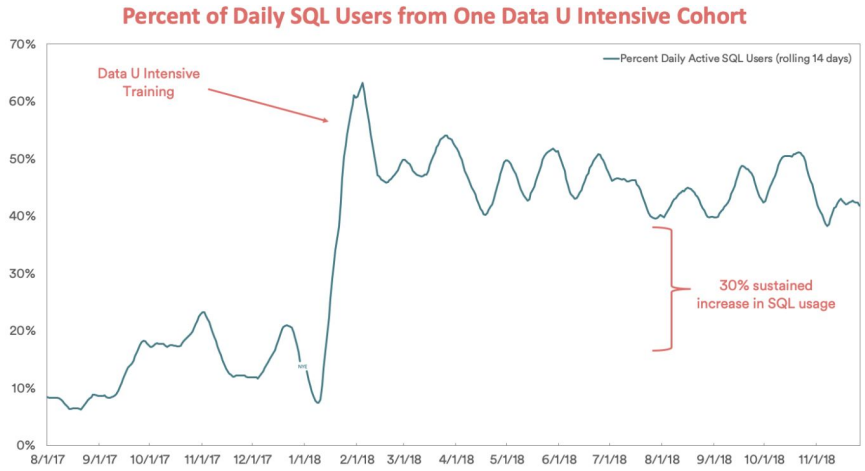

Figure 4.1: Percentage of active daily SQL users at Airbnb over time

One of the major metrics of success for the program is the volume of active SQL users in the company. Prior to the program's existence, about 25% of all the staff across Airbnb engaged

with data tools. This also included any access made to dashboards. In January of 2017, Airbnb's Experiences marketing management team had the opportunity to go through the Data university program, and after the program the volume of SQL users making requests to their data tools rose by a sustained 30% throughout the year.

4.3 External Organisation Efforts

Most of the entities covered in this category draw from a pedagogy that involves bringing in 3rd party L&D teams, mostly in the form of training agencies, to craft specialized training programs.

A big external entity that has played a role in shaping a data literacy culture for many organisations is General Assembly. In the past few years they have partnered with the likes of Adobe and Guardian to create tailor-made programs that cater to their workforce. Certain attributes of the program that stand out are:

1) Customisable format of the coursework, depending on the needs of the respective workforce, to cater to specific job roles
2) Enrolled staff are paired with subject matter experts and industry practitioners.
3) Problem sets and assignments are crafted in the context of the organisation's existing data and tools.

4) Employees graduate these programs with a validated credential that holds value if/when the employee should choose to leave the firm.
5) The cost is between $4,000 to $15,000 per employee, which is 3 to 10 times more expensive than a plug and play Learning Management Systems (LMS)[7]

Another interesting take on this approach could be seen through a hands-on learning system designed by the Data Culture Project. DataBasics.io is a suite of simple data visualization tools, coupled with a curriculum, developed by Catherine D'Ignazio and Rahul Bhagava. This approach is unique for the following reasons:

1) The Data Culture project aims to be a free self service platform that introduces non technical individuals to foundational data tenets, in a creative and emotionally appealing manner.
2) The program strips away intimidating language surrounding the use of data, and through their tools, creates affordances that introduces the specialized vocabulary to beginners, thus becoming comfortable with its use
3) The program's suite of tools (WTFcsv, SameDiff, Word Counter, and Connect The Dots) offer anyone the ability to use their tools to conveniently explore summary statistics in a simple, visualised manner.

[7] "rethinking the build vs buy approach to talent - Josh Bersin." https://joshbersin.com/wp-content/uploads/2019/10/Build_vs_buy_Bersin_1.0.pdf. Accessed 30 Apr. 2020.

4) An end-to-end, staged process is taught on how to approach data for beginners. An example of this below shows the stages highlighted in the Data Culture program.

Figure 4.2: The Data Culture Project workshop process

The unique value proposition of the Data Culture Project is three-fold - First is its beginner-friendly approach to learning, second is its flexible curriculum that is supplemented with facilitation guides that empowers any data advocate in a firm to conduct a group exercise with ease. Lastly is the additional heuristics built into the tools in the form of tooltips, simplified vocabulary, and links to additional resources for independent and continued learning.

4.4 Learning in the Flow

Out of the many deficiencies in the existing learning services within organisations, the ability to enable the user to learn in the flow of completing a task presents the most painful part of the process. GP Strategies (2019) conducted a global survey of over 200 employees to understand at

what periods during their workday do they take time to learn. 80% of respondents noted learning only takes place when the necessity presents itself. This issue here is many of the training solutions existing today are structured like webinars, or physical tutorial classes, requiring the employee to finish hours of content, with no defined takeaway on how to apply these skills to the tasks they need to accomplish.

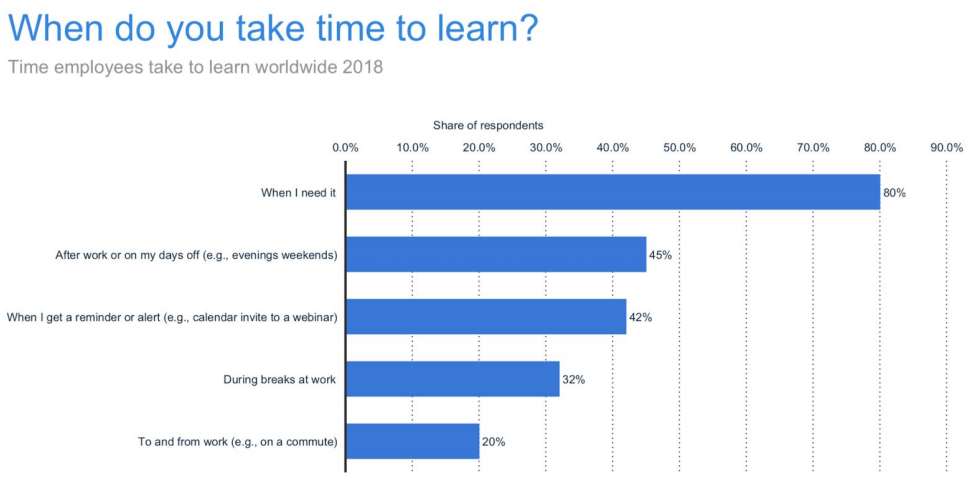

Figure 4.3: Feedback of respondents on time allocation towards learning

In an age of faced-paced change, where data literacy is a growing requirement, what is necessary is building both a culture, and a model of establishing data literacy on the go.

4.5 Summary

1) The Kirkpatrick model is a great tool to measure evaluation of training programs, but certain shortcomings exist. Research on these shortcomings have been done (Bates, 2004) around its lack of including individual or contextual interference in the quality of training; issues we have seen to play a primary role in the efficacy of the programs we have covered in this section.

2) Creating online resources does not equate to impact. A recontextualized, and experiential format has to make the content both engaging and relatable. The context is very important here, as members of the organisation would find it easier to see the usefulness of data literacy if the program is adapted to use the information they encounter during work. It also gives them the opportunity to build familiarity with the resources and tools the organisation offers.

3) People are on different paths of their data journey, and the best tools find flexible ways to accommodate individuals coming from various levels of data literacy. It is in this regard that not only resources in the forms of manuals, forums, and web pages should be created for data analysis tools (Excel, Tableau, SQL, etc), but affordances should be introduced in the interfaces of these tools to give novice users an un-intimidating learning experience.

4) Organisational data literacy is bifurcated between technical capacity, and qualitative understanding around data, as it flows through a lifecycle. We have identified that a balance of both of these capacities are necessary to achieve a healthy data culture. A reliance on technical skills alone introduces overtrust and over-confidence in the data, which reduces the individual/ team's ability to critically think about data throughout stages in it's cycle. A sole dependence on qualitative skills might provide critical thought, but would not be enough in enabling the individuals in teams and organisations alike to take agency of their data. This introduces feelings of intimidation of the domain of data, and encourages dependency on subject matter experts within the organisation (data scientists, engineers, etc), and thus reduces organisational productivity.

5. Proposed Concept & Recommendations

5.1 Problem Statement

Given our findings, the critical area in the data-literacy space is building adaptive tools into the workflow of non-technical members of an organisation, that can continuously present them with the means to make a decision on a given data related inquiry, and supplement both their qualitative and quantitative learning. Ideally this should be accomplished with little to no human intervention, but should provide the affordances for such an intervention to happen during the experience. This statement is illustrated by the empathy map of the non-technical employee below.

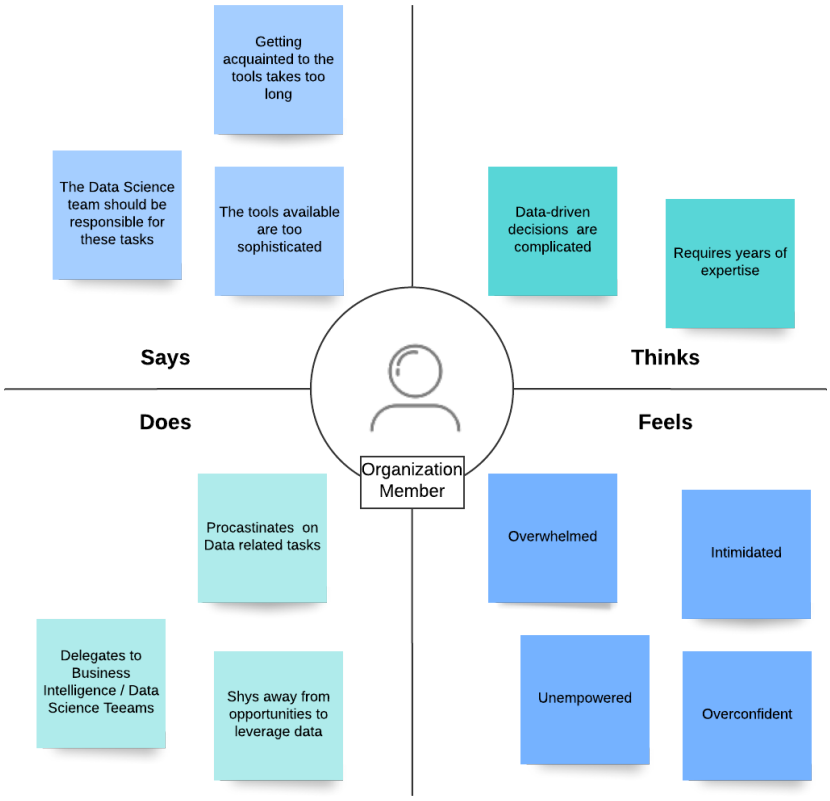

Figure 5.1: Empathy Map of the non-technical company employee

One of the interesting finds during our research was the polar attitude towards data in the workplace. This was bifurcated between individuals coming from a technical background, and non-technical background. Overconfidence was found in many subjects, with instances of using data to make decisions, without properly interrogating the context, which has led to frequent misinformed data decisions.

5.2 Key Assumptions

The needs list was constructed with each stakeholder taken into account. These stakeholders include:

1) The training officials belonging to the organisation
2) The novice in need of the data-literacy skills
3) The managerial entity of the employee that has an influence on the availability/ authorization of the employee into the program
4) The Data Tools required to accomplish the task (in this case the task is a business decision requiring some data insight in the form of a visualization)

There were also assumptions about the state of data literacy in an organisation. These assumptions were used in accordance to all the feedback received on the state of data-literacy in industry, and to simplify the problem space by creating boundary conditions:

1) The organisation has availability of tools present. These are for data fetching, data visualisation (dashboards).

2) The main goal of the concept is to supplement the data literacy journey of novice employees in the workplace quantitatively with respect to core technical skills, and qualitatively through improving their critical reasoning abilities.

5.3 Needs List

1. The concept must be accessible through a regularly used interface that the user encounters in his/her workflow.
2. The concept must provide a quick method in accomplishing the task
3. The concept must provide an efficient means of resolving data permission issues encountered.
4. The concept must not be intimidating. It must be inviting and introduce lexicon to the user with affordances.
5. The concept must be flexible to adjust to the skill-level of the user

6. The OZONE Program: Concept Overview

The concept we are proposing, called ‘Ozone’, is a service-based program that facilitates the stakeholder to delve into the data resources offered by the organisation in an inviting, un-intimidating manner. It also operates alongside a media campaign across the organisation that aims to sensitize, educate, and motivate the employees on engaging with both the tool and existing data amenities.

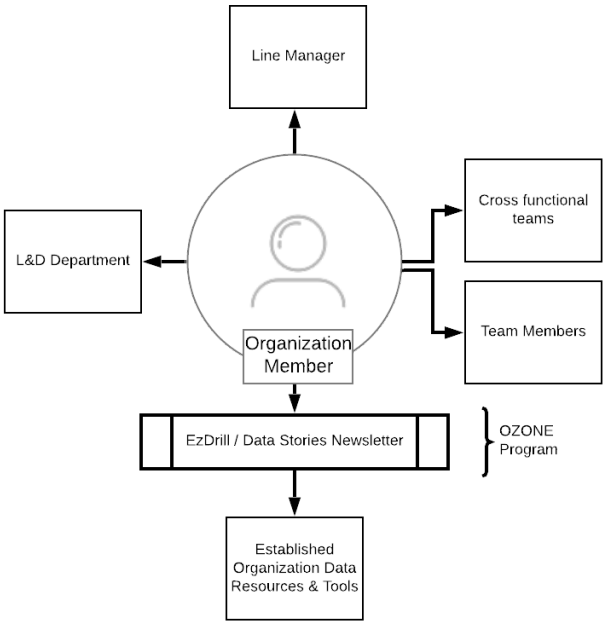

Figure 6.1: Stakeholder Map from the perspective of the non-technical employee

We envision the program to act as a supplement to the already existing learning and development initiatives within the firm. The program will also rely on identified stakeholders in the map above to ensure its impact.

For instance, the learning and development team have defined metrics that will be used to measure the effectiveness of the program and suggest iterations. Line management buy-in is also important as they have a unique influence on what teams can lay emphasis on, and by participating, either by sharing related content, verbally reaffirming its value, or even interacting with the program, will set a precedent that would not go unrecognized. Team members, both lateral and cross-functional, have an important part to play in the virality of the program, either via word of mouth, or through recognition from being featured in the campaign across the organisation.

An important metric of success for the program is the engagement rate of existing data resources. Ideally a sustained volume of queries through EzDrill, and other existing dashboards will indicate to us that the program is working in terms of improving the visibility and relevance of data across the organisation.

6.1 EzDrill

Our research indicated that the most common way employees in organisations encounter data is in a CSV format. But before the CSV format is obtained, a data pullout, in the form of an SQL request, has to be performed. This requires existing knowledge of the SQL language and an understanding of the data infrastructure - skills that will require long-term training.

EzDrill addresses the heart of this problem by providing a simple browser extension widget that allows non-technical employees in an organisation to ask questions about the data they encounter everyday in human readable language, and gain insights fast.

Feasibility of this concept has been well documented (Yin.P et al, 2016) whereby natural language is being parsed to interrogate a given knowledge base table. Fully fledged products have also been launched that solve the issue using a chatbot framework. A case in point is NLSQL[8] , a Natural Language to SQL translation platform.

In addition to this ability, affordances such as tooltips provide a quick way for people to understand specialized language around data. The autofill feature allows the user the opportunity

[8] "How It Works | NLSQL RPA BI Software." https://www.nlsql.com/how_it_works. Accessed 4 May. 2020.

to discover other datasets available for interrogation, inherently educating the user on the potential of the tool.

In addition to the autofill feature, an adaptive “popular questions asked” feature gives the user insights on other methods users have approached using the tool.

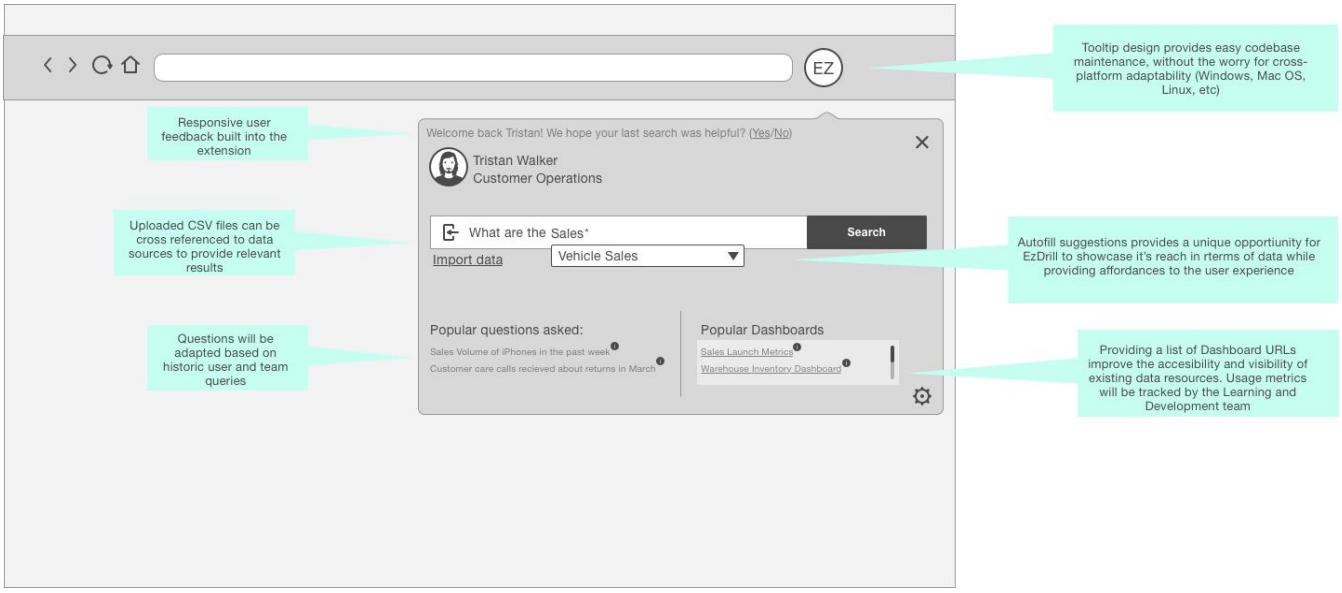

Figure 6.2: Annotated Wireframe of the EzDrill tool (Welcome View)

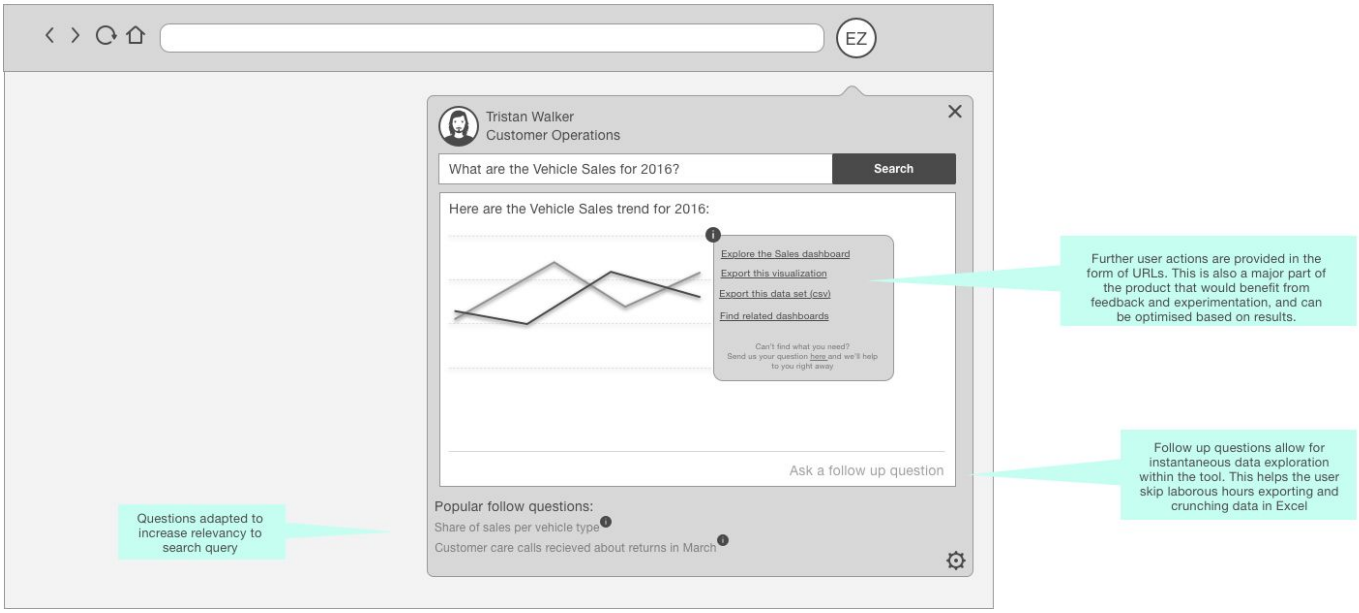

Figure 6.3: Annotated Wireframe of the EzDrill tool (Results View)

6.2 The “Data Stories” program

The motivation behind this concept was to develop a system of delivering organisation-wide data-literacy case studies and critical incidents that can be delivered via a communication channel to sensitize, educate, and motivate the members of the organisation on the value, and ease of implementation of a data-driven culture.

The program is divided into 2 parts: the monthly email newsletter, and a physical banner campaign across high traffic spots in the organisation’s physical space. It is structured in its messaging to be digested very quickly, and has clear calls to action (CTA) that enable the user to act on the messaging with ease.

The Data Stories Campaign was structured as a tool to spotlight individuals and teams (ideally non-technical) that have leveraged EzDrill, and other existing tools, to result in business impact. The poster achieves 3 major objectives - Educating the viewer on how data is leverage in the firm, inspiring the viewer by boldly stating the business impact, and giving the viewer the ability to act on this inspiration by registering for a future lunch & learn series, or interacting with the EzDrill tool directly.

The Data Stories monthly newsletter is a low touch, digital format of the poster campaign. In addition to all the objectives the poster program serves, this gives members of the firm the opportunity to share their testimonials and thoughts on the program with the L&D team. The

messaging within the newsletter is strictly limited to three, and a maximum of three sentences, with 1 piece of multimedia content, and 1 clear CTA is permitted. Rudimentary feedback from interview subjects noted that whilst email is a great channel of communication, information overload remains a problem as they feel overwhelmed by the volume of emails they read on a daily basis. This is one of the reasons we created clear content constraints, and limited the frequency of the newsletter to once a month.

Figure 6.4: Annotated Prototype of the Data Stories Poster

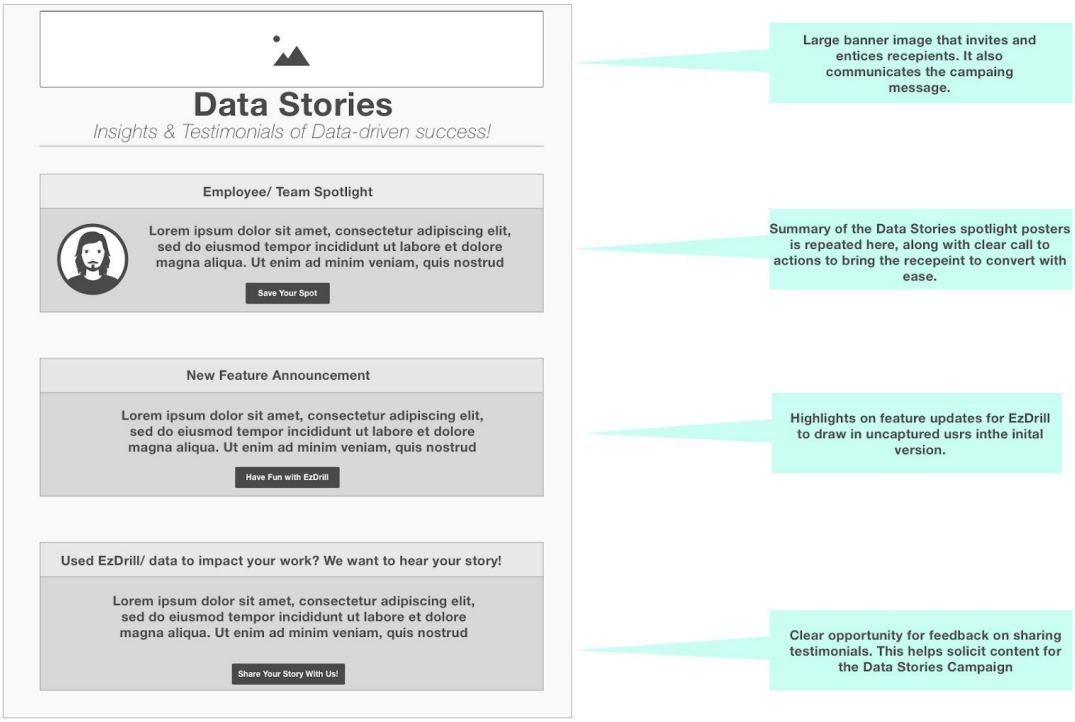

Figure 6.5: Annotated Prototype of the Data Stories Newsletter

6.3 Service Blueprint & Walkthrough

The service blueprint below represents the layout of the experience across time, stakeholders, and levels of interaction. The holistic experience between EzDrill and the Data Stories program is divided into 3 distinct parts.

The first part involved everything from the planning to initiation phase of the two programs. This part ideally should take 1 to 3 weeks for completion, and would involve primarily the Learning and Development (L&D) team to align with senior leadership in promoting visibility. It is crucial

that beta tests would have been run with teams that have been identified as gaining a high level of impact from this program. Such a beta test would provide the initial testimonial that would be published to kickstart the Data Stories program.

The second part, called “Analysis”, involves collecting and aggregating as much feedback on the program as possible. Feedback would come in the two forms. The initial qualitative feedback would be collected from the usage and engagement of both the EzDrill tool, and the Data Stories campaign. Examples of such feedback will be install rates, volume of queries per user passed over time, and engagement in the form of views and queries with existing dashboard and data resources. The qualitative feedback collected would come in the form of survey responses from the EZDrill tool, and pre-post surveys from the lunch and learn series from the Data Stories campaign. Examples of such feedback are the textual information supplemented into the Net Promoter Score (NPS) survey, and interviews the L&D team would have with teams that have been identified to have very low or high engagement rates with the program.

Once this data is collected, by the L&D team, in conjunction with other cross-functional teams such as IT support, would deliberate on adjustments, such as new feature additions and deprecations to be made to the EzDrill tool, and changes to the Data Stories campaign, to improve the efficacy for the entire organisation, or identified sub-teams that need specialized intervention. New content in the form of testimonials from employees or teams that have benefited heavily from the program can be collated, and scheduled. These would be prioritized

based on time and budgetary constraints, and fed into the cross-functional teams in charge of deploying these changes.

The Iteration stage gives adds between a week to a month of a buffer window to allow implementation and roll-out of the new changes that will be made to the program.

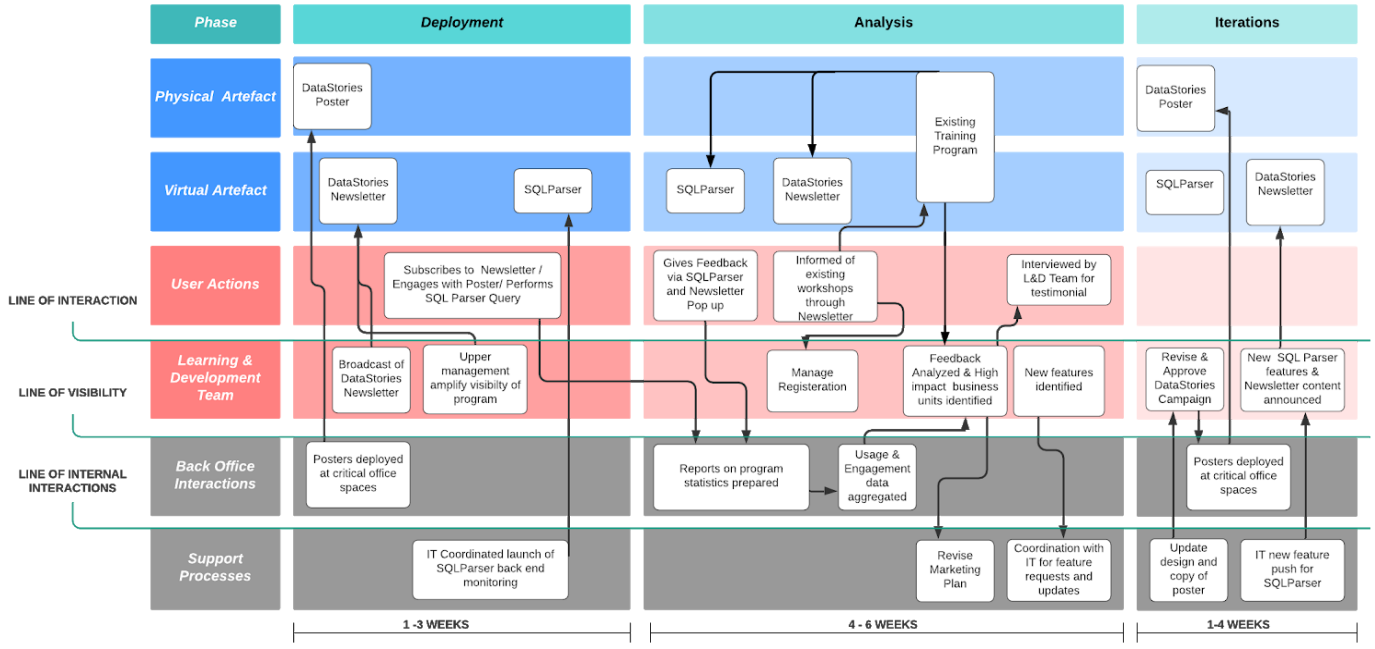

Figure 6.6: Service blueprint for the OZONE Concept

In Figure 14clear lines of visibility and interactions have been elaborated. From the top up is the first line of interaction. This line articulates the boundary between the user (non-technical

employee) and the organisation's operations that power the experience. The line of visibility separates the activities the employee might have privy to from the activities that will be happening without the employee's knowledge. Lastly, the line of internal interaction divides the L&D team members from other staff that do not directly support the interactions between the end user and employee.

6.4 Metrics For Success

Each part of the OZONE service experience has measurable Key Performance Indices (KPIs) attributed to it to ensure proper monitoring, guidance, and alignment on the impact of the program across stakeholders.

Figure 15 represents the list of metrics that can be monitored across the 2 programs under OZONE. The primary metric for the EzDrill is the percentage of daily active users (DAUs) that access the tool. This metric gives a sense of the adoption rate of the product, and could also indicate to us external factors that might drive this number up. An example could be the introduction of a EzDrill tutorial to all existing physical workshops.

EzDrill			
KPI	Units	Inference	Potential Data Slices/ Deep Dives
% of daily active users (DAUs)	# users/day	Number of people opening Ez Drill daily	Across teams
Daily volume of queries	# queries/ day	Adoption of EzDrill tool	Across teams
Daily volume of dashboard views	# views/ day	Adoption of existing data products	Across dashboards; Across teams
Volume of queries per user	# queries/ user	Power users of the EzDrill tool	Across teams
NPS Score	-100/+100	User loyalty and advocacy	Across teams; Interviews
Data Stories (Poster)			
QR Scan Volume	# scans/week	Engagement with poster	Across office spaces
Worskhop sign ups	# signups/week	Interest in tool/ topic	Across teams
Data Stories (Newsletter)			
Open Rate	# opens/ month	Engagement	Across teams
Sign ups	# signups/ month	Interest/ Engagement	Across teams
Feedback	# submissions	Interest	Across teams

Table 6.1: Key Performance Indicators (KPIs) for the OZONE Concept

The primary metric for the Data Stories campaign will be the number of sign ups for the workshops, and open rate of the email. These indices provide us with insights on the perceived interest of the program, and areas within the organisation that resonate with the messaging. As we note later in the conclusion, these metrics should have a means of being monitored in the event that this program be tested.

7. Conclusion

Data in today's workplace is generated by several business units, and many of them have never been tasked with analyzing such quantitative information before. This increases the demand for employers to constantly think of methods to not only upskill their workers, but inculcate a culture that encourages critical thinking of data. As much as the OZONE concept targets the underlying issue around catalyzing data literacy across non-technical staff in an organisation, we would need further testing of this concept in a company setting to figure out the points of weakness.

OZONE offers stakeholders within organisations, training agencies, and L&D teams facing rapid digitization, a starting point on how to address the human resource repercussions. Individually, Data Stories offers an economical opportunity to imagine how a cultural shift can be integrated into familiar business routines, while EzDrill offers an empowering experience for non-technical staff to utilize the power of data in their everyday decision making.

Unlike Data Stories, the concept of EzDrill will require upfront capital investments in setup and maintenance costs, and as such, would not be a trivial concept to implement. In addition, every company has a unique data architecture, and it might prove infeasible to develop this solution

from a ‘one size fits all’ approach. It is in this regard that additional field research would prove valuable to gain a holistic view of the data infrastructure of addressable companies. Understanding the popular data warehouse and pipeline formats can inform our engineering design decisions, and increase the adaptability of EzDrill.

Bibliography

[1] Forbes; “Watch out Corporate Learning, here comes disruption”: https://www.forbes.com/sites/joshbersin/2017/03/28/watch-out-corporate-learning-here-comes-disruption/#48d2665ddc59

[2] D’Ignazio, C., & Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *Journal of Community Informatics*, *12*(3), 83–107. http://academy.datawrapper.de/

[3] Bullhorn. (July 10, 2019). What is the top area in which your peers need the most training and development? [Graph]. In Statista. Retrieved April 20, 2020, from https://www.statista.com/statistics/1026870/opinions-of-staffing-firms-worldwide-peer-training-areas/

[4] Vesset, D., Olofson, C. W., Schubmehl, D., McDonough, B., Woodward, A., Stires, C., Fleming, M., Nad- karni, A., Zaidi, A., and Dialani, M. (2014). IDC FutureScape: Worldwide Big Data and Analytics 2015 Predictions [report]. International Data Corporation.

[5] Akst, Daniel. 2013. “What Can We Learn from Past Anxiety over Automation?” The Wilson Quarterly, Summer.

[6] GP Strategies. (June 14, 2019). When do you take time to learn? [Graph]. In Statista. Retrieved April 20, 2020, from https://www.statista.com/statistics/885973/time-employees-take-to-learn-worldwide/

[7] Deloitte Insights. (2019). Global Human Capital Trends 2019. Deloitte University Press, 1–112. https://www2.deloitte.com/content/dam/insights/us/articles/5136_HC-Trends-2019/DI_HC-Trends-2019.pdf

[8] Accenture. (2020). The Human Impact of Data Literacy.

[9] Mimeo. (2018). *State of Learning and Development 2018*.

[10] Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), Training and development handbook: A guide to human resource development. New York: McGraw Hill.

[11] Bates, R. (2004). *A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence*. https://doi.org/10.1016/j.evalprogplan.2004.04.011

[12] Yin, P., Lu, Z., Li, H., & Kao, B. (2016). *Neural Enquirer: Learning to Query Tables in Natural Language*.

[13] IDC. (April 4, 2019). Revenue from big data and business analytics worldwide from 2015 to 2022 (in billion U.S. dollars) [Graph]. In *Statista*. Retrieved May 14, 2020, from https://www.statista.com/statistics/551501/worldwide-big-data-business-analytics-revenue/