Chapter to be included in New Challenges for Data Design published by Springer.
Draft version, do not diffuse

Chapter 5
Sketching with Data

Fabien Girardin[1].

Abstract The growing deployment of networked infrastructures has dramatically increased the amount of logs byproducts of people digital activities (i.e. digital footprints). That intangible material takes the form of cellular network activity, aggregated credit card transactions, real-time traffic information, user-generated content or social network feeds. The capacity to transform this type of big data into insights, products or services has called for new practices at the crossroad of design and 'data science'. This paper will discuss transversal incline of the innovation with digital footprints and describe how sketching with data offers useful interfaces to the many stakeholders of innovative projects.

1 Introduction

The explosion in the use of mobile devices and social networks has generated large datasets of digital footprints. For instance, visitors to a city have many ways of leaving voluntary or involuntary electronic trails: prior to their visits tourists generate server log entries when they consult digital maps or travel web sites; during their visit they leave traces on wireless networks whenever they use their mobile phones; and after their visit they may add online reviews and photos. Broadly speaking then, there are two types of footprint: active and passive. Passive tracks are left through interaction with infrastructures, such as mobile phone, that produces entries in locational logs, while active prints come from the users themselves when they georeference tweets or their workouts with Nike+ enabled shoes.

We have been active observers and contributors since the dawn of that data deluge working on "making data talk" [1] and materializing services from multiple sources of digital footprints: cellphones, cars, shared bikes, digital cameras, credit cards. For instance, our analysis on georeferenced photos suggested that exploiting this dataset to know who visits different parts of the city at different times can lead to the provision of customized services (or advertising), the rescheduling

[1] Fabien Girardin (✉)
Near Future Laboratory, Av. Général Guisan 19 CP 242, Sierre, Switzerland
e-mail: fabien@nearfuturelaboratory.com

of monuments opening times, the reallocation of existing service infrastructures or the evaluation of specific urban strategies.

1.1 A vision that became reality

The low cost and high availability of digital footprints now provide a new material to understand urban processes and design innovative services. Indeed, only a decade ago, the possibility of producing fully dynamic time-space diagrams from the fusion of human activities data and novel forms of analysis was only discussed in the conditional. For instance, Zook et al [2] envisioned in 2004: *When many individual diagrams are aggregated to the level of cities and regions, these visualizations may provide geographers, for the first time, with truly dynamic maps of dynamic human processes. One might imagine them as twenty-first century "weather maps" of social processes.*

It is only very recently that the presence of digital footprints provides new means to analyze a city in real-time and to replay its processes. These new potentials attracted the interest of large technological organizations, local authorities and urban service providers in 'big data'. For instance, in a follow-up of our analysis on georeferenced photos, we particularly showed the capacity to quantify the evolution of the attractiveness of urban space with a case study of the area of the New York City Waterfalls, a public art project of four man-made waterfalls rising from the New York Harbor [3]. Methods to study the impact of an event of this nature are traditionally based on the collection of static information such as surveys and ticket-based people counts, which allow to estimate visitors' presence in specific areas over time. In contrast, our contribution made use of the dynamic data that visitors generate, such as the density and distribution of digital footprints in different areas of interest and over time. Our analysis provided novel ways to quantify the impact of a public event on the distribution of visitors and on the evolution of the attractiveness of the points of interest in proximity. Local authorities used the results as part of their evaluation of the economic impact of the New York City Waterfalls.

In that specific project, we analyzed two types of digital footprints generated by phones or mobile devices that were in physical proximity to the Waterfalls: cellular network activity and photo activity. Cellular network activity was measured by analyzing aggregate statistical data about number of calls, text messages, and overall amount of network traffic generated at each AT&T antenna every hour. Photo activity was measured by adding up the number of photographers present in different areas of the city, and the number of photos they took in each location. We acquired this data by analyzing photo taken from the photo sharing web site, Flickr. A major implication was to apply prior to any analysis conscientious, principled, and evident measures to protect people's privacy.

1.2 Societal implications

Among the other pioneers in that type of investigations we had to ensure that the social advantages of these applications were not in conflict with important privacy requirements. Digital footprints are both immensely empowering (for the people and places able to construct and consume them) and potentially overpowering as institutional and state forces are able to better harness information with growing personal and spatial specificity. In consequence, there are ethical and privacy implications to grapple with. In conjunction with people's own representation of traceability, there is a legitimate concern about the drift of research on digital footprints. Particularly our work exemplifies the shift from large-scale top-down big brother thread on privacy issues to more local bottom-up little sister types of people monitoring, which makes the whole notion of opting out of technology adoption one of whether to opt out of society.

In fact, these digital footprints have become inevitable in contemporary society and also necessary if we wish to enjoy many modern conveniences; we can no more be separated from it than we could be separated from the physical shadow cast by our body on a sunny day [2]. The growth of our data shadows is an ambiguous process, with varying levels of individual concern and the voluntarily trading of privacy for convenience in many cases.

In summary, at the same time as digital footprints give us new means to model human dynamics and develop new services, they also challenge current notions of privacy. The works described in this article attempt to appreciate and use the complexity and richness of digital footprints without crystallizing into authoritarian structures.

1.3 Methodological implications

The ability to replay the city shows that there are opportunities to propose novel ways to describe the urban environment and develop new solutions. However, there is a big assumption in seeing the world as consisting of bits of data that can be processed into information that then will naturally yield some value to people. Indeed, the understanding of a city and people goes beyond logging machine states and events. In consequence, let us not confuse the development of novel maps from previously uncollectable and inaccessible data with the possibility of producing *intelligent maps*. Our work precisely draws some critical considerations on the current state of the art. The first steps in our projects always aim to figure out: 1) What parts of reality the data reveal and 2) What we can do with them. For instance, not to confuse behaviors with endorsement, that can be considered as a limitation of our New York Waterfalls case study that used the density of digital footprints as indicators of urban attractiveness. In similar studies, calibrations with ground truth information are necessary. Alternatively, some questions or problems can be answered mixing both quantitative and qualitative methods:

- The qualitative analysis to inform the quantitative queries: This approach first focuses on people and their practices, without the assumption that something computational or data process is meant to fall out from that. This qualitative angle can then inform a quantitative analysis to generate more empirical evidences of a specific human behavior or pattern.
- The quantitative data mining to inform the qualitative enquiries: In that approach, the quantitative data help to reveal the emerging and abnormal behaviors, mainly raising questions. The qualitative angle can then help explaining phenomenon in situation.

With complementary perspective on people behaviors or the actual use of the space, it becomes for example possible to develop new types of "Post-Occupancy Evaluations" often overlooked in the practice of urban design and architecture or to design new services for and with local authorities, businesses or individuals. Those approaches almost exclusively involve multidisciplinary teams.

1.4 A multidisciplinary process

Both societal and methodological implications require the involvement of many stakeholders often from different practices and objectives, from engineering to statistics, design, strategy planning, product management and law. The process of innovating with digital footprints demands several steps, each with their own set of skills, knowledge, questions and answers [Figure 1]: From the data access and collection techniques that feed data to obfuscations algorithms and big data management systems that are interrogated by basic data mining operation or advanced statistical inquiries. Often in parallel, information visualization techniques are used to build evidences and indicators. It is the engagement of multiple stakeholders of the project that provoke own questions, ideas or scenarios and therefore new queries to the data.

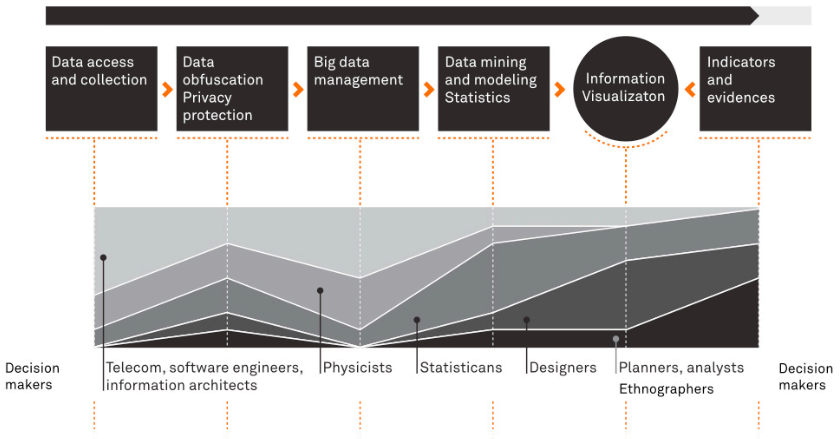

Fig. 1 Simplified overview of the multidisciplinary process of innovating with digital footprints. Each step demands specific skills, knowledge, questions and answers.

Throughout our projects we found the necessity to very quickly being able to visualize temporary results and share them with stakeholders of the project. We learned that this approach was useful to keep a proper momentum in projects that often need to fail, fork or win within a few weeks or months. As a consequence, more and more results of our investigations became interfaces or objects with a means of input and control rather than only static reports. As it became prominent in our creative and innovation process with digital footprints, we called that practice *sketching with data*. We will describe it in the rest of the paper with some examples of its virtues and the tool called Quadrigram developed from our experiences.

2 The practice of sketching

The practice of sketching is common in creative activities such as art or architecture often as a rough version of a work and preliminary attempts to complete something. For instance, in architecture [4] mentions "*Through visual artifacts, architects can transform, manipulate, and develop architectural concepts in anticipation of future construction. It may, in fact, be through this alteration that architectural ideas find form*".

Sketches are meant to be fast; they are often produced with great speed to capture rapid cycles of ideation. The look of those attempts is not as important as the role they play in the creative process. Indeed, sketches are not precise or visually compelling but their expression is vague enough to allow illusions and analogies. It is through iterative attempts that they accompany brainstorming for the author or dialogues among members of a project. However, unlike prototypes, they do not aim for fidelity nor desirability or to prove a concept. Rather, they are used to discover a concept at the beginning of a project. In fact, they can be employed in all stages of the design process to form or deform ideas, even as an observational recording long after the project is completed. For instance, Leonardo da Vinci sketches have become art objects in their own right, with many pages showing finished studies as well as sketches.

Sketches do not have to be captured on paper. They can take the form of 3D models or interactive visualizations, and often the intention takes precedent over the media [4]. When translated to the world of data, sketching has a lot more in common with exploration in a workshop rather than it does working with Photoshop or other graphic design application. The process consists in generating early results and insights from manipulating data queries, establishing a blended approach with multiple sources, sculpting with algorithms, extracting with filters and drawing results with libraries. Our use of sketches is meant as a medium for inspiration and transformation in multidisciplinary teams. In consequence, they need to be comprehended by a wide variety of professionals, from physicist and engineers to lawyers, decision makers and strategists. They become *boundary objects* to explore ideas and solutions [5]. Borrowed from sociology, *boundary objects are objects that are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites* [6]. Practically, they allow coordination without consensus as they can allow the specific understanding of each stakeholder to be reframed in the context of a common project. Instead of communicating across disciplines using vocabularies from different practices, a sketch reveals the data and its transformations in real running code and designs.

Sketches can be used to start a project with a multidisciplinary team that needs first to grasp the potentials of the data and their limitations. For instance, when an institution envisions the use of its own digital footprints to develop a real-time service, early sketches are needed to answer basic questions such as what is the expected amount of data, what are the expected signals in the data, or more precisely what time frequency can be associated to the notion of *real-time*. Practically, a time series visualization can help each team member understand whether their service can deliver new information every minute, every hour, every day or every week.

Besides the practice of prototyping often employed in the domains of engineering and experience design, there exist other creative forms to express ideas and concepts with programming and data. Under the umbrella term of *creative coding* there is a growing community of professionals who use the language of code and data as their medium. Their work, which often evolves through iterations of sketches, includes everything from digital art to elaborate interactive installations, all with the goal of expanding our sense of what is possible with data and software. In those approaches, the necessity to program in machine language cuts the direct relation between the physical action of the hand and the result. The tool becomes an abstract and is not more a direct extension of the body and the mind.

3 The tools

Based on our first experiences in sketching with data, we detected two increasing demands within innovative institutions. First, there is a necessity of multidisciplinary groups to think with liberty with data, outside of coding, scripting, wizard-based or blackbox solutions. Then, we perceived the demand to diffuse the power of information visualization within organizations to reach the hands of people with knowledge and ideas of what data mean. We were struggling to combine the tools that could allow people with the knowledge and data but not the technical skills to develop their own ideas and scenarios. We needed them to manipulate their data as artists or architects that explore ideas with freedom with an easy to erase pencil or chalk stick. The response came with the development in collaboration with the design studio Bestiario[2] of a visual programming environment called Quadrigram [Figure 2]. We wanted the tool to be particularly designed for iterative data exploration and explanation.

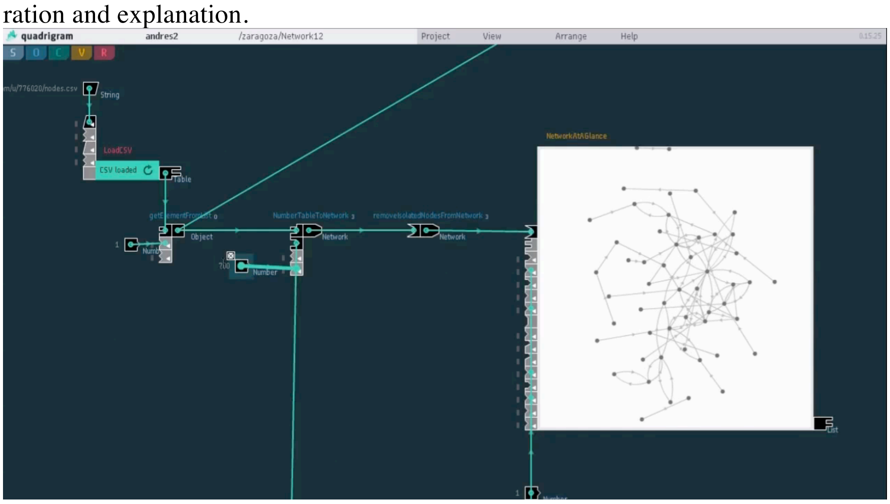

Fig. 2 Quadrigram is a Visual Programming Environment composed by pre-programmed modules that perform operations. The modules are linked between themselves to produce a visual workflow in a boundless canvas.

With Quadrigram, the abstract notions of programming became transparent as it offers the opportunity of manipulating data as a living material that can be shaped in real time. By living data we mean data that are constantly changing and accumulating. They can come from social network, sensor feeds, human activity, surveys, or any kind of operation that produces digital footprints. The tool was meant not only for *data scientists* but rather everybody with knowledge and ideas as an expansion of their mind. We designed it with the principles described in Table 1.

---

[2] Bestiario design agency http://www.bestiario.org/

Table 1. Summary of the key principles used to design Quadrigram.

Principle	Description
Non-linearity	The infinite nature of a boundless workspace provides a natural way of developing ideas without the need of keeping linearity. Therefore, users can sketch and arrange their ideas on demand, increasing naturalness.
Iteration	Quality is a consequence (among others) of understanding, the more you iterate the more you master, and one of the key techniques in Data Visualization is iteration.
Data as a living material	Data are constantly changing. The ability to work with real time data empowers to detect changes fast and act consequently, enhancing efficiency and control capacity.

To create Quadrigram as a tool to sketch with data we had to rethink the approach people take to work and create with data. We describe the four main challenges in the following sections.

3.1 Redefining work with data

The design of Quadrigram lays on this very idea of *feedback loop*. It is designed for iterative exploration and explanation. Each iteration or *sketch* is an opportunity to find new questions and provide answers with data. Data mutate, take different structure in order to unveil their multiple perspectives. We like to think that Quadrigram offers this unique ability to manipulate data as a living material that can be shaped in real time or as Mike Kuniavsky describes in Smart Things: Ubiquitous Computing User Experience Design: *Information is an agile material that needs a medium* [7].

With the diffusion of access to data (e.g. the open data movement), investigation with data has become utterly multidisciplinary. Projects embark teams with fast prototyped tools that promote the processing, interpretation, and reinterpretation of insights. For instance, our experience shows that the multiple perspectives extracted from the use of exploratory data visualizations is crucial to quickly answer some basic questions and provoke many better ones. With Quadrigram, we suggest a novel approach to work data in which analysis and visualizations are not the unique results, but rather the supporting elements of a co-creation process to extract value from data. In Quadrigram, the tools to sketch and prototype took the form of a Visual Programming Environment.

3.2 Reducing the barriers of data manipulation

Visual Programming Environments have flourished in the domain of information technologies, starting with LabVIEW[3] in the 80s and then spreading to the emerging fields mixing data with creativity such as architecture, motion graphic and music. In these domains, they have demonstrated virtues in reducing the barrier of entry for non-experts. In the Visual Programming Environment we developed, users manipulate in an interactive way pre-programmed modules represented as graphical elements. When connected, these modules form a 'data flow' (also called dataflow programming) that provide a constant *What You See Is What You Get* (WYSIWYG) view of the result of the program ideal for quick *trial and error* explorations. This way the tool allows for the evaluation of multiple pathways towards the correct solution or desired result. It inspires solution-finding for non-technical professionals by exposing the full flow of data.

Visual programming language (VPL), as per its nature, facilitates the learning process, not only for those who are neophytes in programming, but also for those who know how to use non-visual programming languages. In comparison with standard programming languages and wizards, we wanted Quadrigram to lower the barrier of data manipulation, reducing time to produce sketches, inspire creativity and ensure flexibility [Table 2].

Table 2. Summary of the qualities of a tool that supports sketching with data.

Quality	Description
Power	Puts power in the hands of more people that have the knowledge, though not necessarily the technical ability to program solutions from scratch. Everyone from project managers to interns can now play a part.
Speed	Reduces development time and effectively the cost of implementation, by handling complex, programmatic patterns.
Creativity	Inspires solution-finding by exposing the full flow of data, whereas standard programming languages abstract and compartmentalize the process.
Flexibility	Ensures flexibility by allowing for multiple pathways towards a solution, working in the freedom of two-dimensions. By contrast, wizards and standard tools assume user-intent and impose fixed sequences that are rigid and one-dimensional, limiting the potential for discovery along the way.

[3] LabView : http://www.ni.com/labview

3.3 Creating a coherent language

A major challenge when grouping tools to work with data within a common Visual Programming Environment has been to define basic building blocks of a language. We started with an exploratory phase that led to the release of an experimental environment called Impure and its large sets (500) of diverse pre-programmed modules. This free solution generated a decent community of valorous 5000 users. We used Impure as testbed for our ideas and performed the necessary user studies to come up with a coherent basic language. We particularly focused on specific action verbs that enclose the most common operations on data (e.g. *sort, search, insert, merge, count, compare, replace, save*). These actions are performed on Data Structures (e.g. *create List, sort Table, replace String, cluster Network, compare Date, resize Rectangle, load Image*) within specific domains (e.g. Math, Geography, Statistics). The language is complemented with a set of Visualizers categorized according to their objectives to reveal aspects about the data (e.g. compare, contextualize, relate). Through this axiom of *actions – structure – domain,* user can find the appropriate module within a diverse toolset. However, that language did not eliminate all the barriers to manipulate data. For example, we found out that users needed some conceptual knowledge of data structures prior to sketching their own solutions.

3.4 Taking advantage of an ecosystem of great tools

Quadrigram was not meant to replace an existing tool or practice but rather to complement other more sophisticated tools of data analysis in order to embellish the last stretch of the run. Indeed, *big data* has been often understood in a vertical manner — terabytes of information. In Quadrigram we approach *big data* in a horizontal manner— through the multi-diversity of datasets. Like many other platforms, Quadrigram connects to various types of data sources (e.g. databases, APIs, files) to load data within a workspace. But we also wanted users with detailed needs to take advantage of R scripting[4] to perform advanced statistical method or Gephi[5] to layout large networks. The main challenge was to find and implement a protocol to communicate Quadrigram data structure back and forth with these great tools. In other words, we wanted users to perform analysis in R as part of their data flow. Similar to the architecture of distributed systems, the solution was to pass around serialized Quadrigram data structures, which offers a pretty unique mechanism to store and share results of data manipulations that we call

[4] The R Project for Statistical Computing: http://www.r-project.org/

[5] Gephi, an open source graph visualization and manipulation software https://gephi.org/

"memories". Consequently, Quadrigram is a *sponge* capable of absorbing information from many diverse sources and permits its users to visualize it as part of the activity of sketching.

When sketching with data and with Quadrigram, we aim at bringing data closer to existing investigation and innovation processes. In the following section, we highlight qualities of this practice in our experiences of answering questions and designing new services with a wide diversity of digital footprints.

4 The virtues of sketching with data

Originated within the posture of the investigators, our work has gradually evolved into helping institutions and companies in transforming digital footprints into insights, products or services. This practice requires the basic skills of *data science* (i.e. data analysis, information architecture, software engineering and creativity) with a capacity to play the interface with wide variety of professionals from physicists and engineers to lawyers, strategists and designers. This transversal incline of the investigations and innovation with digital footprints requires the knowledge of the different languages that shape technologies, report on the geography of their use, and describe people practices. The models of enquiries blend the qualitative field evidences with quantitative observations from logs with the use of sketches to share a common language.

4.1 Share a common language

As a first practical example, we would like to describe the capacity of sketches to bring diverse practices with their own specialisms and specific understanding of an idea to be reframed in the context of a common project. Indeed, many of our projects required the joint understanding of space (e.g. a territory, its rules, cultures, history), of the networks that compose the space (both physical infrastructures and digital activities) and the human behaviors manifested in that space. In a project called Footoscope we employed these prisms to explore new ways to analyze and experience football (i.e. soccer) with the increasing presence of data in the game. Sports have always kept a tight relationship with data to measure performances. It has been particularly the case to improve athletes' capabilities with motion analysis or objectify team sports that are easily fragmented into single events (e.g. Sabermetics). With new means of producing statistics via video and sensor technologies, other sports have started the search for objective knowledge.

In the domain of football, companies such as Prozone and Opta Sports have led the innovation in data collection. In parallel, some academics have been exploring this new terrain to apply their statistics-led methodology [8]. Similarly, designers have also started to transform these new measures (often in real-time) into sophisticated visualization to augment the spectator's experience.

In Footoscope, we explored new possibilities of sports *data-tainment*. Through multiple simple sketches we described visually the morphology and tactics of a football team according to raw data on its passing game (e.g. passes between players, positions of the players when receiving the ball, playing time) transformed into indicators and visualizations [Figure 3]. We sketched the results with Quadrigram with amateurs to help us decipher data of teams they know or want to explore. Sharing those attempts to understand the game differently allowed amateurs to extract and focus on key information that is otherwise hidden within football data. In a form of rapid visual business intelligence, this analysis and its visualization became the supporting elements of a co-creation process to extract value from data. For instance, in a team of amateurs we deciphered the World Cup 2010 statistics to reveal the network, spatial and behavioral elements of each team. That anecdotic work generated dialogues between statistics and people with knowledge of the terrain to produce a new apprehension of the game [ref]. When shared online, the sketches provided a common language for different practitioners to grasp the potential of data in football.

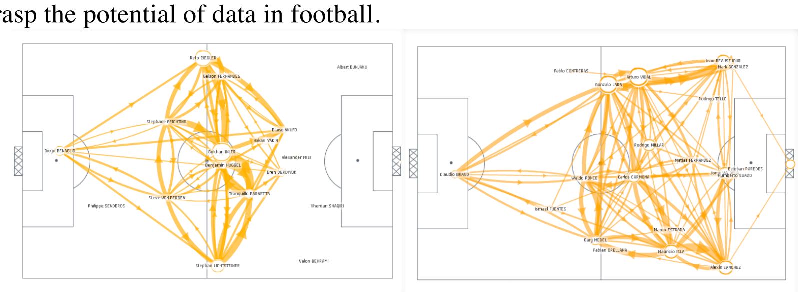

Fig. 3 Footoscope helped football amateurs decipher data of teams they know. Here two sketches produced during the investigation of the World Cup 2010 with the incapacity for Switzerland (on the left) to manage the distances between its lines, with its defense and strikers compacted at a short distance. This contrasts with a more balanced team that takes a greater advantage of spaces, such as Chile (on the right).

4.2 Qualify the results

In this second practical example on the virtues of sketching with data, we showcase our project at the Louvre Museum. Not only because we have kept fantastic memories of the breathtaking context, but also because we learned a lot from the analysis of digital footprints to provoke qualitative knowledge. The Louvre is, by far, the most visited museum in the world with 8.5 million visitors and more than 40,000 visitors at peak days. In Paris, it is one of the main drivers of "cultural enthusiasm" that is an inherent feature of the city. In consequence, the museum witnesses levels of congestion, which, beyond a certain threshold can be described as *hyper-congestion*. This phenomenon has some direct negative consequences on the quality of the visitor experience as well as on the organization and management of the Museum (e.g. increased stress level of the surveillance staff).

The Study, Evaluation and Foresight Department of the Museum performs extensive surveys, audience analysis and on-site observations to ensure a good visiting experience. However, the information they collect is punctual and only partially feeds the visitors flow models necessary to setup and evaluate some of the museum strategies. In an exploratory project, they wanted to investigate new solutions to help answer their concerns with *hyper-congestion*. In response, we first investigated the collection of new empirical data on the flows and occupancy levels of visitors in key areas of the Louvre and the developed diagnosis indicators that capture the changes of visitor behaviors according to the congestion in the museum.

In collaboration with a real-time traffic information provider, we specifically designed sensors that audited during short periods of two weeks the presence of Bluetooth-enabled mobile phones on a key trail that leads to the Venus de Milo. The analysis of the collected longitudinal measures of presence and flows of visitors quickly led to the development of an indicator that unveiled areas in which the congestion of a room changes the visitors presence times and flows. While unprecedented in the history of the Louvre, some results produced more questions than answers. We faced a new set of inquiries that quantitative evidences from sensors could not answer but field observations could. For instance, what event provoked the congestion, what aspects of the visiting experience were affected, or why some rooms do not show symptoms of *hyper-congestion*?

In response to these interrogations, we returned to the sketches produced as part of our data analysis. Yet this time, we did not complete them for the decision makers but for the security staff. Indeed, on-side personnel offered untapped knowledge on visitor practices and flow management strategies. So we setup different meetings at the museum and used our sketches to have the staff qualify the results of the audits. Their evidences from the field explained some irregularities and completed the understanding of visitor behaviors. For instance, a simple decision to close a door provoked changes in the measures of visitor flows.

In that experience, we learned the types of questions the analysis of digital footprints can answer. For instance, "how many observations can we produce?", "what do the data tell about a population?", "what evolutions can we measure over time?", "can we categorize these evolutions?", "what are the trends and the outliers?" or "what are the flows that connect different places?" [9]. Yet, the understanding of an environment such as the Louvre goes beyond logging machine states and events. This project showed that the qualitative view from the staff reinforced the quantitative measures and consolidated the overall knowledge on *hyper-congestion*.

4.3 Innovate with data

In this final practical example with highlight the virtues of sketching when conceptualizing and building services based on digital footprints. With data, there is always the risk that teams jump to technical solutions before evaluating whether solutions will work. Our approach in that domain often focuses on building the simplest possible thing. Indeed, once we can prove something is working and we can prove that users want it, the next step is to improve the service. So we start considering if each task can be divided into small actions that end-users or experts can perform. That way the design and development of a data product or service starts with something simple that lets a team determine whether there is an interest to go further. Sometimes, the idea behind these sketches will survive into the finished version because they represent some fundamentally good concepts that the team might not have seen otherwise; sometimes, they'll be replaced by a different approach or technique.

We applied this approach in several data product projects including steering the development of mobile recommender systems or interactive dashboards of commercial activities for municipalities and retailers. Following the spirit of sketching, the results can be produced quickly (e.g., in a few days, if not a few hours), they are never complete or precise but they are vague and good enough to let you know whether it's worth going further. This approach has some similarity with prototyping that embraces the notion of the minimum viable product and the simplest thing that could possibly work. However, sketching with data does not aim for fidelity nor desirability or to prove a concept. Rather, we use them to form and deform ideas and concepts within a multidisciplinary team. They can be employed in all stages of the design process, for instance to check the sanity of the data or evaluate some aspects of the user experience.

4.3.1 Checking the sanity of the data

One of the biggest challenges of working with data is getting the data in a useful form. As teams want to jump to trying to reach a common language and design the product, this critical task of cleaning the data is often overlooked. In fact, in the data science community there is a rule that says that 80% of the work in any data project is cleaning the data. For instance, the geocoding process of transforming database of postal addresses into geographic coordinates might lead to erroneous information due to interpolation or ambiguous street names. Those common issues can be evaluated with mapping a sample of the dataset and verifying visually the accuracy ratings of each address [Figure 4]. This type of sketch is meant to quickly check the sanity of the raw data and whether the errors or imprecisions can be assumed over the course of the project.

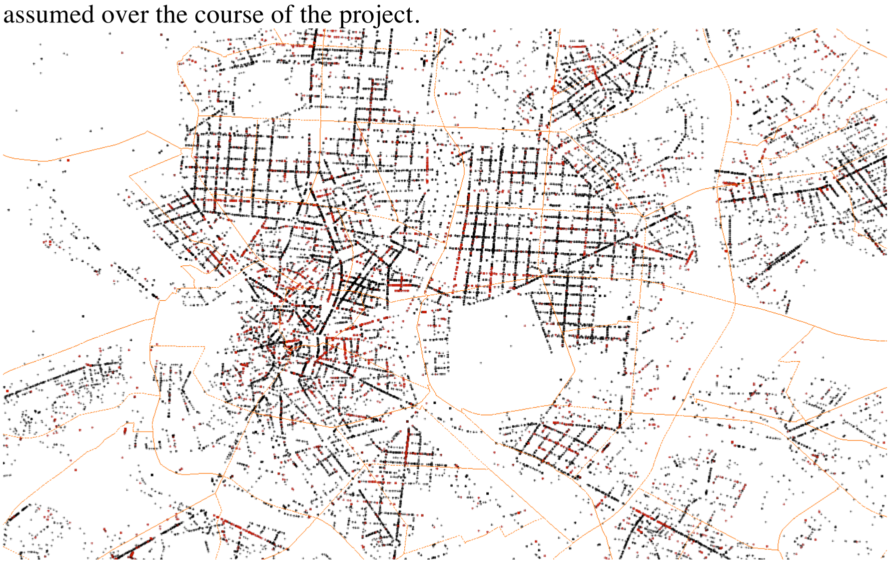

Fig. 4 Sketch of the result of transforming database of thousands of postal addresses in Madrid into geographic coordinates with the potentially erroneous results highlighted in red.

4.3.2 Defining a clear focus

The practice of building products with digital footprints and its integration of user experience design processes is still relatively young. In consequence, many solutions have a tendency to engage users with the many, often irrelevant, insights that can emerge from data analysis and visualization. For instance, a wide variety of indicators can be developed when designing a dashboard for municipalities to measure the commercial impacts of events. Indicators that data scientists might fail to understand because local authorities and city planners simply do not grasp or communicate their potential. The quick production of interactive sketches with real data is often a good approach to evaluate the potential of different indicators with their views (e.g. distances to purchase of local citizens, commercial routes among the main areas of commercial influence, balance of trade with neighboring cities). These early results are material to collect feedback and iteratively engage users in considering the evolution of their work using new metrics and indicators. Consequently, each insight is designed according to the specific action the user can take leading to a product with a clear focus.

5 Conclusions

The growing quantity of digital footprints gives us new means to model human dynamics and develop new services. However, that evolution presents both societal and methodological implications that require the involvement of many stakeholders often from different practices and objectives from engineering to statistics, design, strategy planning, product management and law. Throughout our projects we found the necessity to very quickly being able to visualize temporary results and share them among different specialisms. As it became prominent in our creative and innovation process with digital footprints, we called that practice sketching with data. We participated in the development of a tool that treats data as an agile material. In Quadrigram, a sketch is an opportunity to find new questions and provide answers with data. A sketch becomes *boundary objects* for inspiration and dialogue in multidisciplinary teams.

Our experience shows that digital footprints are not sufficient to give full answers and solutions about people, their behaviors and usage of technology. Yet the world of 'data science' and computer science still lack in sensitiveness to the limitations of quantitative evidences and the models we can build on them. We have often been confronted to these limitations. Several of our projects with digital footprints taught us that there are insights that only the articulation of digital footprints and contextual observations can provide.

On a more general picture, sketching with data is an approach of the investigation and design with digital footprints that is close to academic research. Sketching implies taking time to think, open and contrast ideas, and staying humble, not being afraid of expressing doubts. When it comes to mixing practices, this practice of the researcher driven by doubt but confident in its methods is what we believe drives to relevant insights and solutions.

6. References

1. Girardin, F. (2009). Aspects of implicit and explicit human interactions with ubiquitous geographic information. PhD thesis, Universitat Pompeu Fabra.
2. Zook, M., Dodge, M., Aoyama, Y., and Townsend, A. (2004). New digital geographies: Information, communication, and place. *Geography and Technology*, pages 155–176.
3. Girardin, F., Vaccari, A., Gerber, A., Biderman, A., and Ratti, C. (2009). Quantifying urban attractiveness from the distribution and density of digital footprints. International Journal of Spatial Data Infrastructure Research, 4:175-200.
4. Schank-Smith, K. (2005). *Architect's Drawings: A Selection of Sketches by Worlds Famous Architects Through History*. Routledge.
5. Biddulph, M (2012). Prototyping apps using real data. Presentation at the EAN's World of Data event in London. http://www.slideshare.net/mattb/eans-world-of-data-prototyping-apps-using-real-data

6. Star, Susan; Griesemer, James (1989). "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39". Social Studies of Science 19 (3): 387–420
7. Kuniavsky, M. (2010). *Smart Things: Ubiquitous Computing User Experience Design: Ubiquitous Computing User Experience Design*.
8. López Peña, J., & Touchette, H. (2012). A network theory analysis of football strategies.
9. Yoshimura, Y., Girardin, F., Carrascal, J.P., Ratti, C., and Blat, J. (2012). New Tools for Studying Visitor Behaviours in Museums: A Case Study at the Louvre. In Fuchs, M., Ricci, Francesco., Cantoni, L. (Eds.), Information and Communication Technologies in Tourism 2012. Proceedings of the International conference in Helsingborg (ENTER 2012). Mörlenback: Springer Wien NewYork.