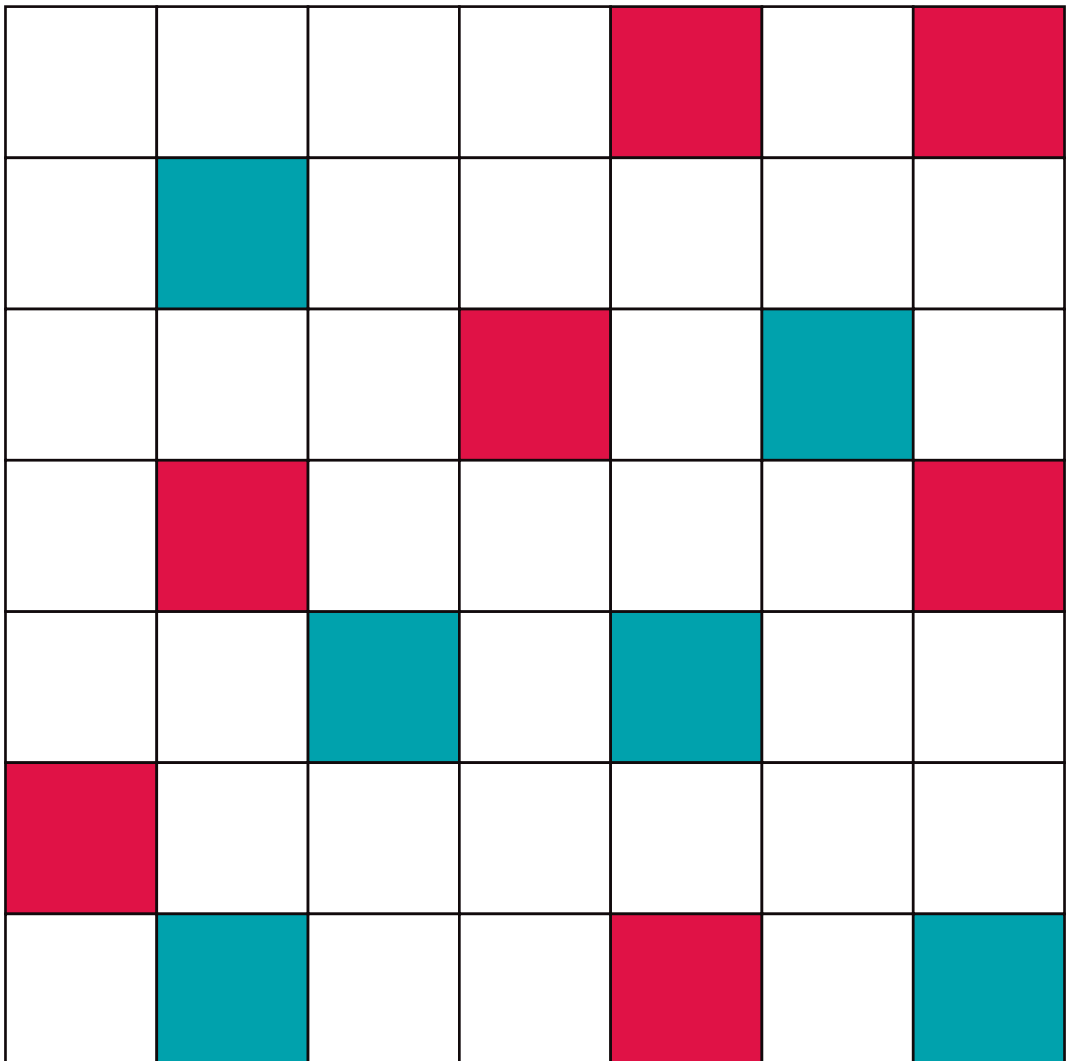


A CULTURAL WORKER'S GUIDE TO A GALAXY OF DATA

Burak Sayin



TRANS EUROPE HALLES, 2022

TRANS EUROPE HALLES
ISBN: 978-91-527-1601-4
Copyright: Trans Europe Halles

Author of the publication:

Burak Sayin

Design:

Darko Kujundžić

The European Commission support for the production of this publication does not constitute an endorsement of the contents. Neither the Commission or Trans Europe Halles can be held responsible for any use which may be made of the information contained therein, which reflects the views only of the authors.



Co-funded by the
Creative Europe Programme
of the European Union

A CULTURAL WORKER'S GUIDE TO A GALAXY OF DATA

4

**PREFACE: WHY
DID I WRITE THIS PUBLICATION?**

6

**INTRODUCTION: WHAT IS
IN THIS PUBLICATION?**

8

**AHMED'S STORY:
WHAT IS DATA?**

9

**WHAT DOES AHMED'S STORY TELL US
ABOUT DATA IN 13 POINTS**



teh.net

PREFACE: WHY DID I WRITE THIS PUBLICATION?

This publication’s journey started back in 2018, when we at the Trans Europe Halles Coordination Office attempted to build a Customer Relationship Management (CRM) system to streamline our data management efforts.

This publication’s journey started back in 2018, when we at the Trans Europe Halles Coordination Office attempted to build a Customer Relationship Management (CRM) system to streamline our data management efforts.

As a network organisation working in the cultural sector, we are connected to many people in our daily operations in a networked way – mirroring the internet. As the community manager, my work usually starts with an email address and a name on a laptop screen and evolves into video conversations via Skype, Zoom, Google Meets or Microsoft Teams. This has changed since the Covid-19 pandemic, but

that’s a different story for another publication (coming in March 2022).

I used to think that the only difference between our network and the internet is that we actually know the people in person on the other side of an email address or name on a laptop screen. Basically, we really can match the faces with email addresses in our databases. So, the CRM system was meant to organise that data to create a real-time digital map of our network. We would organisationally “know” who is doing what and where in terms of an individual’s position, an organisation’s location, types of organisation, turnover, types of business model and so on. It was sort of a “once the genie is out of the

bottle” type of situation. The idea was essentially to capitalise on our digital connections in a more professional manner that would benefit our members, our real-life connections with our members and our sector in general.

To cut a long story short, our CRM search evolved into a search for an Enterprise Resource Planning (ERP) system. We ended up subscribing to a solution provided by an Estonian start-up. It was actually a pretty agile system that could have combined a project management system with a database and even an invoicing system. That was part of the problem. We expected a lot from it and it failed. Fast forward one year, thanks to one of our intern's efforts, we managed to build a functional testbed for our database in a simple Google Sheet. Finally, we are now using Microsoft Project for our project management, we process our invoices manually and we have a testbed for our database.

As everyone in every universe knows, the grassroots and independent cultural sector is in a mess and this is totally OK (as in I am A Cyborg but that's OK [the 2006 Park Chan-wook Movie]). Please be advised there are more random references as the publication continues. This is a heavy topic and we only have 10 more pages to go so this is a heads up! In summary, Trans Europe Halles is not above the general mess that characterises the sector it represents. We tried to re-invent the wheel ourselves and partly succeeded. Along the way, we learned a lot and realised we didn't actually know

much about this topic. That's exactly why I am writing this publication. I'd like to write about the basics concerning data, so that you – dear reader – can save your precious time. I can tell you right away, this guide is not a set of recommendations. You will need to navigate your own way around it. There are basic things about data that everyone can do and learn and I will try to go over those in this guide. Bear with me. And most of all, DON'T PANIC!

INTRODUCTION: WHAT IS IN THIS PUBLICATION?

This guide is designed for cultural workers working in nano/micro cultural organisations, from cultural hubs to independent theatres. I am trying to simplify complex topics and maybe will even oversimplify things at times. That's why I will use stories, memes, graphics, popular references as my tools to unpack some complex terms.

Data is a massive topic to unpack in a short publication. Therefore, I intend to write a document that serves as a humble guide to the mysterious universe of data in the cultural sector. In a nutshell, when you open this publication, I hope you can find words that you are somehow familiar with but you couldn't really articulate before. I hope that you will also find my explanations useful in your daily work.

Whether you are aware of it or not, we all work with data on a daily basis. For example, do you remember this morning you opened your eyes

and turned on your bedside light? Boom! You are working with data (or actually working for "data"). You just increased the electricity consumption of your household (just a tiny bit) and this was measured by your electric counter, which is owned by some private energy distributor company (probably EDF, Engie, E-ON or ENEL, if you are living in Europe). They added your energy consumption to their statistics. So, now you are a part of their way of understanding the electric consumption of your household, your building, your neighbourhood, your city, your region and your country. If this is too

macro or too abstract (as in “abstract expressionism” [An art movement period covered from Wassily Kandinsky to Jackson Pollock]), let me give a micro and a possibly more familiar example:

I have just bought a Kraftwerk ticket and it got me thinking about selling tickets at venues like yours. Back in the 1990s, people probably would have started queuing for tickets in front of your venue on the ticket release day, which was probably one month or one week before the concert. Maybe you would have even decided to sell the tickets on the day. Maybe you were a membership organisation and your members had early access. Today, you would start selling the tickets 5-6 months before the gig, and the Kraftwerk gig will be sold out in a month. It is very convenient. First, you know already that it is sold out, so you can plan accordingly in terms of logistics and staff.:

- **You know that 737 people will attend the concert with tickets and the average age of participants is 49.7.**

- **323 people attending are female, 300 people attending are male, 72 people attending are non-binary, 42 people attending prefer not to mention their gender preference.**

- **350 people have never heard of your venue before this event.**

- **530 people heard about the event via Facebook, 200 people heard about the event via your website, 7 people didn't answer.**

- **250 people subscribed to your newsletter, so you have direct access to 250 email inboxes.**

- **You even know that there are 50 Fredriks, 75 Annas etc. coming to the event. This part is a bit tricky, I will come to back to this later. But I have a feeling that you've got my point by now.**

You might have noticed that the topic of “data” in general became very popular in the last decade, especially on the other side of the Atlantic. The data trend finally reached European shores and our daily lives with platforms like Google and Facebook. In the last two years, our exposure to data has increased with the shock of the Covid-19 pandemic. We have been introduced to a number of cases and deaths on “World-o-meter”. Quite simply, the use of data has become omnipresent and data itself has become a popular narrative tool for so many people. And, what does this mean for the nano/player in the cultural sectors? How can we navigate the BIG world of BIG DATA or just data? How can data help you? Let's take a closer look!

MORE ABOUT DATA

AHMED'S STORY

Hello, dear reader. My name is Ahmed. I am going to try to explain to you what data is. I am a time-traveller who was born in the Nile Valley in 1623 BC. Like my father, I worked a piece of farmland for my alien overlords. Wheat was our thing. Every year after our harvest, we hit the road to take the year's crop to tall triangular buildings where our alien overlords lived. They were also our gods.

We were told by some strangely dressed folk that these overlords wanted development and progress. We really never saw or met them. These strangely dressed folk saw them all the time, and told us about their encounters. They were tall and mighty, we heard. They had superpowers. They lived for eternity. Unfortunately, they were stuck on this piece of land. Their ship's engine had broken down during their inter galactic travel. They were far from home and they wanted to go back so badly. They were homesick. Somehow, our wheat was helping them to go back home. The more wheat we brought, the more we helped them to progress, therefore helping them to get back home. In return, we were blessed. Maybe we could have joined them once they fixed their ship, but only after we passed on to the afterlife. What an opportunity!

Every year, we went there, we gave our wheat, we showed our papyrus with our family name. The strangely dressed people took our wheat and weighed it. Then, they marked some boxes on their papyri and noted our family name down. Life continued, we repeated the same ritual every year for some time.

Little did we know about the flood that would change everything. One day, we woke up and all our crops had gone. We went to the temple hoping to pray a little. Maybe our overlord would have mercy on us. They had some. When we arrived at the temple, it was the end of life as we knew it. People were climbing over each to be able to reach the gates. The word was that a ship was leaving the city, but there were only 12 seats on it. One could join the lottery, they said, and so I did. I was asked to give a drop of my blood, so I gave it to them. They took it and wrote 478,935 next to it. So, we started to wait. It was a long wait. Finally, these strangely dressed people started to write a number on a triangular plate. And, here it was 478,935. I entered the temple, I saw a light and

boom, I was in a party. Strangely dressed people were jumping. There was this mechanical sound almost deafening me. I asked the lady next to me: “What happened? Where am I?” She said: “Read the Rhind Mathematical Papyrus, Ahmed, you will understand.” And, we were in London.

WHAT DOES AHMED’S STORY TELL US ABOUT THE DATA IN 13 POINTS?

1. DATA AND INFORMATION ARE NOT THE SAME THING.

Whenever I need to write about something, I start the process by checking the dictionary meanings of the word I am writing about to see exactly what it means. Since the lingua franca of today is English, I checked Lexico Dictionary (powered by Oxford University Press) and the Miriam Webster’s Dictionary for the definition of data. Finally, I took a peek at the Organisation for Economic Co-operation and Development (OECD)’s Glossary of Statistical terms (a highly recommended resource for textbook definitions).

Lexico’s definitions of data are as follows:

1. Facts and statistics collected together for reference or analysis.
 - 1.1 The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.
 - 1.2 [Philosophy] Things known or assumed as facts, making the basis of reasoning or calculation.

Miriam Webster’s definitions of data

1. factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation
2. information in digital form that can be transmitted or processed
3. information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful.

*** OECD’s definition (referring to The Oxford Dictionary of Statistical Terms as the source)**

Characteristics or information, usually numerical, that are collected through observation. Context: Data is the physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means.

** OECD provides “context” to have more detailed description of some terms*

I haven't forgotten what I wrote in the beginning. Data and information are not the same thing, BUT they are used interchangeably a lot. That's a language thing and there is not much we can do about it, but a more specialised textbook provides a simpler way out. According to the A-level Information Technology support guide of Cambridge International AS, data is information without meaning. The guide's definition is as follows:

The concept of data as it is used in the syllabus is commonly referred to as 'raw' data – a collection of text, numbers and symbols with no meaning. Data therefore has to be processed, or provided with a context, before it can have meaning.

2. MATHEMATICS AND COLLECTING INFORMATION IS NOTHING NEW (4000-5000 BC). MODERN STATISTICS AND DATA IS RELATIVELY NEW (17TH CENTURY).

Not to dive into the murky waters of history here, but we have always measured things. Apart from our survival being dependent on measuring things (i.e. food rationing), we travelled, built and traded – all of which are pretty impossible to accomplish without basic arithmetics and geometry. As a result, we prospered. There are 7 billion of us living on 7 continents. Not so bad, right? It is not all about measuring, of course, but technology has contributed a lot to our progress, hasn't it? And that is mostly about applying collective accumulated information (something we will later call "knowledge"). Anyhow, eventually, we got pretty good at measuring and pushed the boundaries in the form of statistics. But what for? Well, Cogito, ergo sum (see René Descartes). In a nutshell, some European minds thought that having a say in one's own will and destiny was a cool thing. Some thought that, if they compiled enough numbers together (i.e. number of annual births and number of annual deaths), they could calculate how many people would be living in a city in five years. This was witchcraft back then (17th century). If one knows how many people there are going to be in a city in the next five years, then one can plan many things accordingly. That's how God died, according to some (see Hegel and Nietzsche). To wrap up this point, how we looked at numbers has radically changed over the last 300 years. Numbers were always around, but rather how we thought about them changed and is still changing today.

3. THE INVENTION OF COMPUTERS (POST-WORLD WAR II) CHANGED THE COURSE OF DATA AND HOW WE PROCESS DATA.

This is an obvious one, but an important point. Before computers started to compute, our relationship with data was very different. Data existed in different shapes and forms and we already had the scientific methods to collect data (as

in Observe>Question>Hypothesise>Test>Conclude>Redo). However, our brains simply didn't have the "muscle power" to do what was needed to analyse big sums of data. In other words, we didn't have the sheer mental capacity to make the calculations with numbers, information and probabilities that our data sets consisted of.

At first, computers were merely very complex calculators that were simply programmed to work on one specific problem. Instead of pushing the numbers on calculators ourselves, we designed a machine to push the buttons of a calculator for us. And, we used the principles of probability theory to build it [The best known example is the Turing machine depicted in the movie *The Imitation Game*; lesser known examples are Z3 by Konrad Zuse and Harvard Mark I by John von Neumann]. Why did people think of this? They were looking to answer Entscheidungsproblem [see Gottfried Wilhelm Leibniz, Kurt Gödel, Wilhelm Ackerman and David Hilbert] – “the decision problem” – which is simply the possibility of boiling down every decision to a yes/no question.

We are still using very complex calculators in our homes, but what needs to be calculated is a bit different than during the 1950s and 60s at the height of the Cold War. The history of technological development is somehow reflected in the etymology of the word “data”, which is derived from the Latin word “given thing”. It is the plural form, for example, “given things”. The word data became popular in the 1600s and you can see how the definition has developed until the 1940s. You can also see how data-related words entered into daily use in the 1950s, 60s and 70s.

1640s > a fact given or granted [classical plural of datum, from Latin datum “(thing) given”, neuter past participle of dare “to give” (from PIE root *do- “to give”).

1897 > a fact given as the basis for calculation in mathematical problems. Numerical facts collected for future reference.

1946 > transmittable and storable information by which computer operations are performed.

1954 > Data-processing.

1962 > Database; structured collection of data in a computer.

1970 > Data-entry.

4. DATA IN ACADEMIC LANGUAGE IS NOT THE SAME THING AS THE DATA YOU HEAR ABOUT IN THE MAIN(MULTI)-STREAM MEDIA.

Since the 1990s, we have been hearing about the rise of the internet. We hear about it on the news, on our Facebook and Instagram feeds. We hear about it in documentaries that tell us we are being observed 24/7 by some applications.

We watch those documentaries online. We also use data ourselves now. We use it a lot. We follow our cycles, our runs, our sleep patterns, our psyche, our chores through apps collecting data about ourselves. We are supposed to know ourselves better through data somehow. More recently, we followed infections rates, hospitalisation rates and death ratios across the globe during the Covid-19 pandemic. **Ourworldindata.org** has become our go-to page for data. Maybe due to positive attempts to communicate scientific knowledge to the general population, we are overexposed to the word data. However, the term data in academia has a different context and some prerequisites. In academia, data needs to be researched (not only collected). Research needs to follow scientific methods and it needs to be peer reviewed.

There are two main types of data:

Qualitative – describes things, i.e., it is descriptive information.

Quantitative – numerical information, i.e., numbers, statistics, measurements, etc.

Let me explain using our cat, Halüs, as an illustration.



Qualitative

- Halüs is black except for his paws, chest, chin and part of his tummy, which are white.
- His hair is short-to-medium in length.
- He spends most of the day asleep and tends to go out at night.

Quantitative

- His tail is 25cm long.
- He weighs 6kg.
- He has two sisters.
- He is 42cm long (excluding his tail).

5. DATA NEEDS OTHER THINGS FOR US TO MAKE SENSE OF IT (AS KNOWLEDGE).

In order for data to make sense, we need some things: Context, Information, Process, Standard and Knowledge. For example, I bought a ticket for the Kraftwerk concert in your venue. There are two other people who bought tickets for the concert. I will try to explain what that looks like in your database and give some context: Information, Process, Standard and Knowledge.

First Name	Last Name	City	Age	Gender	Phone	Email	Subs.
Burak	Sayin	Malmö	35	Male	74528059	burak@teh.net	Yes
Michelle	Saunders	Helsinborg	19	Female	75278346	freebird@fx.net	No
Nat	Williams	Lund	28	Non-binary	23485712	kip@ty.net	No

BASIC ID

PROFILE

MARKETING

The context here is the type of event, which is a Kraftwerk concert. Information is the text entered in every box that was given (remember the 17th century definition?) by the user. The process is the way I categorised the data (the Basic ID, profile and marketing). The standard is that we directly asked for the person's age (we could have asked for the date of birth) and another example of the standard would be that we asked about the gender and gave four options to the user (male, female, non-binary and prefer not to say). Finally, when those four things came together, we have knowledge about the user called Burak Sayin. And, we also have knowledge about two other users, Michelle and Nat. That means that we also have knowledge about the collective made up of three people in total. For example, two of them will come from another city. The average age of the group is 27.3. Two of them are not members of our club and so on and so forth. The rest of my points will be brief.

6. DATA IS ABOUT FIVE THINGS: VARIABLES, MEASURING, PROBABILITY, PREDICTABILITY AND PLANNING.

7. METADATA IS DATA THAT EXPLAINS OTHER DATA.

Metadata is very important, especially when it comes to making predictions about individuals by only looking at the collective data. For example, the metadata in the concert table above is the date and time when the information was given (Burak filled in the form at 14:32 on 1/1/2022).

8. SYSTEMICALLY COLLECTING DATA DOESN'T NECESSARILY MAKE THE COLLECTED DATA SCIENTIFIC.

9. YES, WE HAVE DATA OVERLORDS.

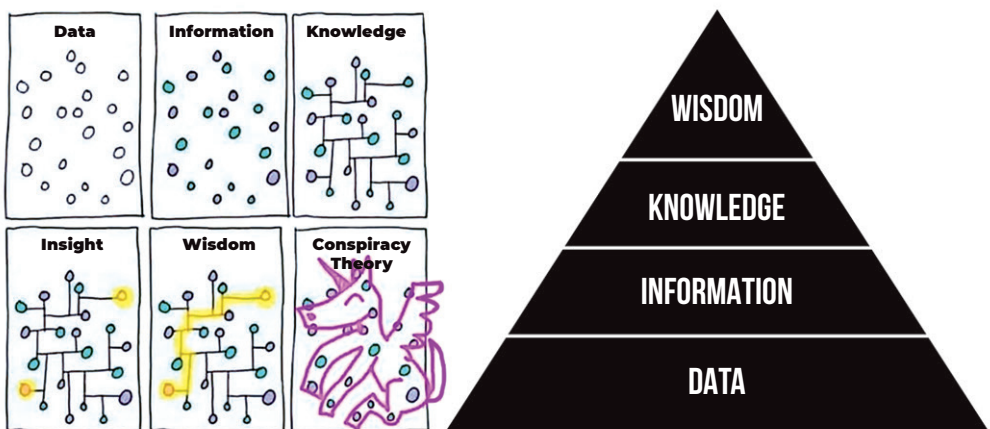
They are national statistical institutions, the International Labour Organization (ILO), the International Organization for Standardization (ISO), the Organisation for Economic Co-operation and Development (OECD). Without these organisations, it would be very difficult to have global trade, international travel, electricity and gas. They were mostly founded after World War II, so we don't fight over who (nation-states) means what (with numbers). We used to do that a lot. These organisations are heavily regulated by laws and regulations. They collect data officially and formally. They mostly don't need our consent to do so, but they can't share individualised versions of the data. They are there to inform others such as researchers, policymakers and citizens about societies and countries. There are also other organisations that collect data unofficially and informally. They are mostly private companies (Google, Facebook etc.). They need our consent to do so, but they are not regulated and oversight is very limited.

10. OUR RELATIONSHIP TO DATA HAS BECOME MORE ATEMPORAL (TIMELESS).

This is because how we learn has changed. We used to learn by reading books, but now we learn by extending our networks. The book has been replaced by the network. See writing by Bruce Sterling and William Gibson on this topic. Interestingly, the book *Neuromancer* provides a good narrative about what is going on regarding our realities today.

11. YOUR DATA MIGHT NOT HAVE MEANING BUT IT CERTAINLY HAS SOME VALUE (EVEN IF IT IS VERY VERY VERY LITTLE).

12. THIS MEME EXPLAINS IT ALL – CHECK OUT THE DIKW PYRAMID. ALSO, “CORRELATION IS NOT CAUSATION”.



13. WHY DOES IT MATTER FOR THE GRASSROOTS CULTURAL SECTOR? AND HOW CAN WE GET BY?

It matters because collecting data is the easiest way of adding value to your organisation. Just simply by asking five or six things about your audiences and users, you can make a difference. It will only take one week to set up basic data collection. One important thing to be aware of is that data collection projects never, ever end. They are always a work in progress because people change, ergo data changes.

For you, dear reader, maybe there could be two benefits. One is that you can get to know yourself (as an organisation) better, so that you can develop. Second is that you can get to know your audience better and increase your impact.

The topic of data is usually approached with caution, suspicion and scepticism in the cultural sector, simply because we are a social sector. We like human connections and we, as a sector, exist because of the human condition, which means always oscillating between isolation and connection. There are introverts out there who we can reach out to. We can take our positive energy to those in need, especially to help disperse the shadows of the last two isolated years. Data could be a powerful tool to do this. How much do we really make the most of this tool?

Have you Googled “How to make a Data Audit?”. It is very easy and it is free. You can use Typeform, Googleform or Jotform. Some of these are free or cost very little money. If you can't, then just use pen and paper. That's also a very reliable type of data collection. Just starting a Facebook event means that you are already selling your user data to Facebook. You are doing all the work for Facebook, while they are making money off your users and audience. It is very convenient, but at least add your own registration form so that you are gathering data too.

Once you start collecting data, try to systemise it. In other words, repeat the same method every time you collect data. For example, always separately ask for First name and Last name; always keep the date format the same; always ask about the users' city. Simple things become beautiful if they are done over and over again. Maybe you can try to develop your methods over time.

That's it for this publication. I hope you've enjoyed reading my musings on data and found something helpful. Please drop me a line at [**burak@teh.net**](mailto:burak@teh.net) for more information – and we may or may not add you to our database!

