Hestia.ai
sustainable data solutions

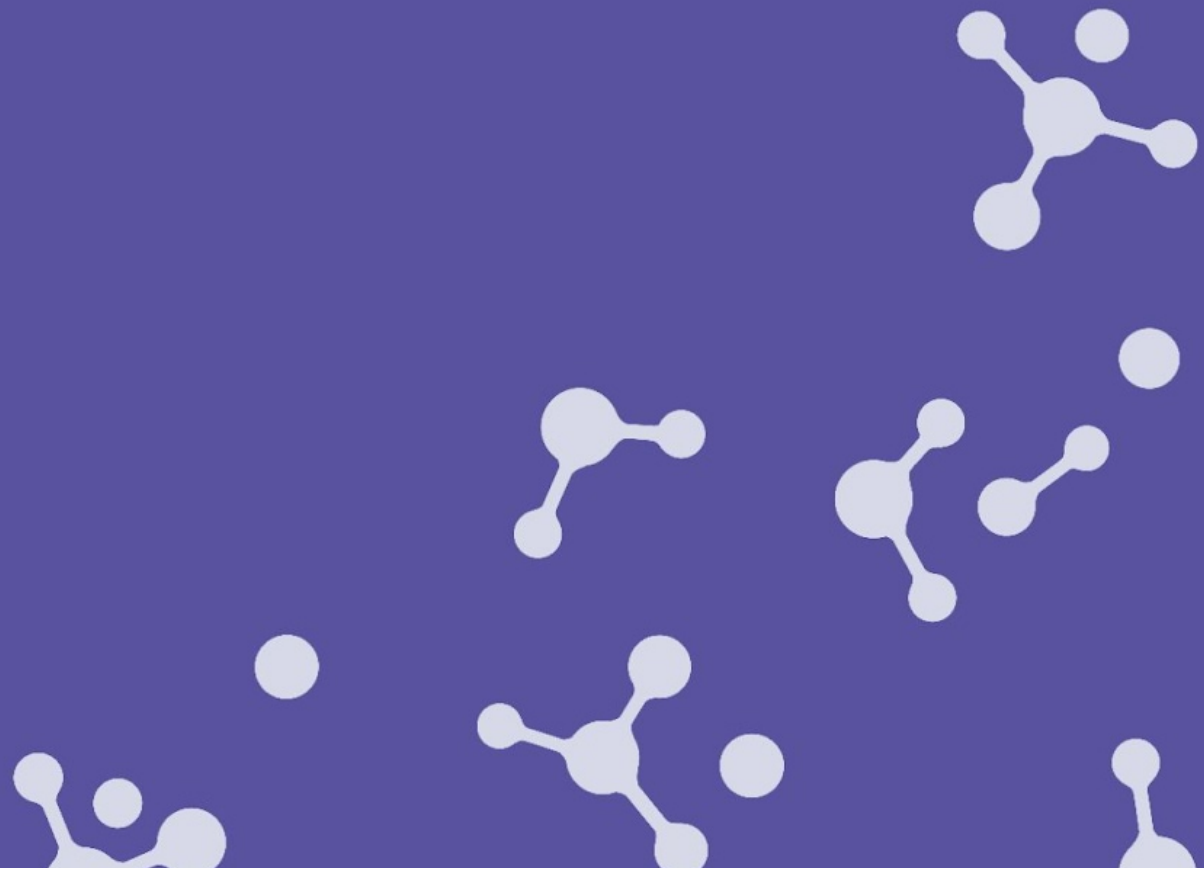
#digipower Investigation

Technical Reports

UNDERSTANDING INFLUENCE AND POWER IN THE DATA ECONOMY

A Narrative Report by Hestia.ai, May 2022

Authors: Jessica Pidoux, Jacob Gursky, Alex Bowyer, Paul-Olivier Dehaye



Hestia.ai
sustainable data solutions

Acknowledgements

The authors want to thank, for their trust, their time and their courage in engaging in something new, the study participants:

- Anders Adlercreutz,
- Leïla Chaibi,
- Filomena Chirico,
- Christian D'Cunha,
- Stephane Duguin,
- Atte Harjanne,
- Jyrki Katainen,
- Miapetra Kumpula-Natri,
- Dan Koivulaakso,
- Markus Lohi,
- Tom Packalén,
- Sirpa Pietkainen,
- Mark Scott,
- Niclas Storås, and
- Sari Tanus.

The authors also want to thank

- for his help, Konrad Kollnig (developer of Tracker Control),
- for their support, Tiina Härkönen, Kristo Lehtonen, Jukka Vahti, Riitta Vänskä and the SITRA staff,
- for their expertise, PersonalData.IO members and volunteers, and in particular Judith Herzog,
- for their dedication, members of data collectives *The Eyeballs*,
- the Migros Pioneer Fund and Robin Born, for their understanding of the relevance of #digipower to the ongoing *HestiaLabs* project, and their willingness to accommodate priority changes accordingly,

and of course

- Hestia.ai staff, who have all contributed differently to the investigation: Emmanuel Eckard, Charles Foucault-Dumas, Hugo Hueber, Andreas Kündig, Valentin Loftsson, François Quellec, Florian Singer, Luã Streit, Marie-Pierre Vidonne, Thomas Wilde

#digipower
Technical Reports:
Understanding Influence and Power
in the Data Economy

Hestia.ai
sustainable data solutions

Executive Summary

ABSTRACT

There is an imbalance of control over personal data between service providers and civil society. While service providers acquire knowledge and influence individuals' behaviour through data, individuals do not own their data, and the personal data ecosystem lacks the transparency necessary to be understood. The #digipower investigation tackles this imbalance by demonstrating the data economy's social consequences and providing solutions for a new economy in this technical report. This report is complemented by a case studies report illustrating flows and usages of data with practical examples from participants' data. The investigation maps out the ways that personal data is collected in both the physical and the digital realms. It reveals the commercial purposes driving the data economy and how these affect our private, public and social lives. In this investigation, fifteen participants, among them members of the European and Finnish parliaments, EU and Finnish civil servants, NGO directors, and journalists, were coached through a participative methodology that is highly replicable by anyone: a learning process situated in the participants' experiences as they recover their data through legal and technical means, to then make sense of it collectively. Using facts gathered by participants about dozens of companies like Twitter, the retailer Gigantti, Google, and newspapers like Aamulehti and Helsingin Sanomat, we introduce the concept of *Infrastructural Power:* the mechanisms by which providers and platforms exert their influences over today's data economy. We explain these mechanisms and their effects using industry jargon, so that civil society can take a role in challenging power dynamics around data. Four processes explain how infrastructural power is accumulated and exerted. These are broken into 10 specific mechanisms - such as inference, ranking, and data funnelling - that involve two feedback loops influencing users' actions and centralising data flows. Our reflective conclusions provide a manifesto and new scenarios for ensuring the data economy remains focused on the common good, including engineering design, proportionality of data collection, and social care.

GOAL

The #digipower investigation seeks to understand the distribution of power in the data economy, and in particular how that distribution of power changes as more of the economy becomes digitised, and ever larger amounts of personal data get collected.

METHODOLOGY

We assumed that it would be particularly interesting to consider our research question around test subjects who were already decision makers in a traditional form of power. We settled on fifteen participants that we selected amongst members of the Finnish and European Parliaments, civil servants, directors of NGOs, journalists, etc.

We then built a participative investigation, which required coaching the participants on how to retrieve their personal data, based on three lenses available to each of us:

- subject access requests, i.e., the right to view one's own data under the General Data Protection Regulation;
- data download portals, i.e., self-checkout transparency portals built by companies on a voluntary basis;
- a technical audit, through an Android app installed on a loan phone.

Each lens offered a different but complementary perspective on data flows surrounding our participants.

INDIVIDUAL FINDINGS

Through that process, the participants collected numerous facts about the data economy surrounding them. For instance, we found evidence of

- online retailer Gigantti sharing individuals' physical purchase information with Meta (Facebook),
- Boeing seeking to influence Finnish MPs to sell fighter planes (including the criterias Boeing used on Twitter),
- the UK Labour party and data broker Bisnode using Meta (Facebook)'s targeting tools,
- how Google blends data from searches with wifi signals to infer at which exact location one might be,
- systematic non-compliance with General Data Protection Regulation requirements, etc.

CASE STUDIES

Additionally, we weave many of the individual findings into four case studies as when analysed in a transversal way, they contribute together to bigger findings. The case studies are:

- *Who cares about my geolocation and why?*, which simply shows what can be collected and understood about position and movement in the physical space, and the consequences thereof;
- *When you view the web, the web views you*, which contrasts the previous case study with similar digital situations, and highlights the heightened potential for shaping the environment online;
- *Move fast and capture all signals, everywhere*, which is focused on the ability of Facebook to convince others to install their tracking tools, despite going against the long term own interest of those partners ;

- *Participants chasing their personal data*, which is focused on recounting the numerous difficulties that the participants faced.

All of the above is detailed in a separate methodology and case studies report, entitled *"Auditing the Data Economy through Personal Data Access"*.

BIG PICTURE

We also synthesised all of our heterogeneous, but complementary, findings into one coherent vision of the data economy, as part of a narrative report entitled "Understanding Influence and Power in the Data Economy".

We identified an overarching situation of power (*Infrastructural Power*) that functions through two distinct capabilities working in conjunction with two feedback loops:

- *Technical Capability* opens up the *Accumulating Information and Knowledge to Act* loop, for instance to influence an user of online services to make a purchase through extensive prior profiling.
- *Organisational Capability* enables the *Composing Complex Infrastructures for a Dominating Position* feedback loop, for instance to encourage website owners to transfer data to Facebook.

We also formulated much more precise decompositions of the mechanisms that can be used to acquire positions of power, and provided a taxonomy of the consequences of digital power.

RECOMMENDATIONS

As a guide towards a desirable future for the data economy, we formulate a *#digipower Manifesto*, rooted in values of *transparency through principled engineering design*, *proportionality of data collection*, and *social care*.

In order to reach that future, we make five recommendations:

- *Change the Narrative* of innovation around data, to encourage business practices that are more sustainable and in line with the General Data Protection Regulation;
- *Productivise #digipower*, i.e. replicate this approach to pedagogy situated in the personal data of the study participants, in order to seed multiple communities in line with the manifesto goals;
- *Increase Infrastructural Power of Civil Society*: build technical capability and organisational capability in order to have effective counterpowers in the digital economy;

- *Support Data Collectives* as a vehicle for reaching faster a more fair distribution of value in the data economy;
- *Enforce GDPR Properly* or risk affecting negatively innovative European businesses.

Hestia.ai
sustainable data solutions

1. Introduction

The #digipower investigation is carried out by Hestia.ai, commissioned by SITRA, on the digital power that controls today's data economy. In contrast to other traditional power forms, digital power is based on the deployment of techniques and the organisation of relationships around those techniques that harvest a vast amount of personal data. The data is collected in both physical and digital realms, and the commercial purposes driving the data economy affect at the same time our private and social lives, as well as public space.

The #digipower investigation presents in this narrative report the mechanisms and consequences on how digital power operates, and documents them through accessible facts that can be – in theory – obtained by anyone.

When trying to understand digital power, we end up always confronting the knowledge and resources that few privileged stakeholders have in the data economy, with the knowledge and resources that external stakeholders representing the interests of society have: these are our investigation participants supposedly placed in a favourable position, in politics and journalism, for ensuring data economy's integrity for the common good.

The results of the investigation are presented in two reports: *(i) a* case studies report entitled *"Auditing the Data Economy through Personal Data Access"* and *(ii)* a narrative report entitled *"Understanding Influence and Power in the Data Economy"*. Both reports are accompanied by a shared *Executive Summary*.

The executive summary presents the main findings of two perspectives on digital power: the bottom-up results of individual experiences when recovering personal data, the top-down results about the data economy which were produced from a collective analysis of individual experiences – a "big picture".

The methodology and cases studies report presents the bottom-up results, while the narrative report presented here discusses the top-down results of our investigation.

1.1. HISTORICAL CONTEXT FOR CONDUCTING THIS INVESTIGATION

In the early 70s, some of the biggest companies in the world were oil companies and car companies. Then the oil crisis hit, and local communities introduced various measures as a response, such as changing speed limits to reduce consumption. Since these events, awareness

of the looming climate crisis has increased, and we now have regulations limiting carbon emissions on individual cars. In 2014, an independent study led to the emissions cheating scandal and a market loss of tens of billions of dollars for Volkswagen, which was found to be evading certification tests for pollution standards[1]. The $50,000 study conducted at the University of West Virginia was as simple as sticking a sensor in the tailpipe of a car – crucially in real life road driving conditions.

In October 2003, a Harvard student in his dorm room managed, within the span of a week, to appropriate the idea of someone else, to breach security of Harvard's IT systems, to violate copyright, to violate privacy, to acknowledge all this and somehow, within two weeks after that, to get a disciplinary pass[2]. While he stated then "Issues about violating people's privacy [with his service] don't seem to be surmountable", his discourse has evolved now onto "I am sorry, we just need to collect more data, AI will fix it" (paraphrase). This business executive is now on top of one of the top ten[3] companies in the world by market capitalization, busy shaking hands with other executives at his level to redefine the next steps of the data economy (metaverse!) – while taking measures that increase Facebook's market dominance[4]. This would be – and remained – for a long time laughable, except that by now we have to face an escalating crisis whereby this data fuels business dynamics that in turn feed negative information dynamics. Indeed, it could be argued that elections are already won and wars already started thanks to slowly simmering disinformation fed through social networks.

Somehow, despite alarm bells ringing for years, we still have very little visibility on how those digital systems work or even good methodologies to address this question. Unlike with car emissions, we still cannot assess independently the behaviour of those systems in real life conditions.

It is in this context that SITRA approached Hestia.ai to conduct an investigation of the various forms of power leveraged through data by service providers within the data economy.

1.2. A NEW PARTICIPATORY INVESTIGATIVE PROCESS

Those two crises, with one catching up on the other, have similarities but they also have a crucial difference. One cannot simply adopt the approach of the West Virginia researchers with a test car: it is inescapable that the unit of inquiry in the data economy is an *individual* using a

[1] Volkswagen emissions scandal https://en.wikipedia.org/wiki/Volkswagen_emissions_scandal Wikipedia. Accessed March 2022.
[2] Katharine A. Kaplan, *Facemash Creator Survives Ad Board*, Harvard Crimson, November 19 2003, https://www.thecrimson.com/article/2003/11/19/facemash-creator-survives-ad-board-the/
[3] Just barely, at the time of writing, March 2022.
[4] *Mark Zuckerberg gave the order to kneecap Vine, emails show*, Mashable, December 2018, https://mashable.com/article/mark-zuckerberg-helped-thwart-vine

multitude of digital services. Addressing questions in the data economy thus introduces completely different requirements, which we resolved by building a participative investigative process leading to pedagogy situated around the participants' own data: we first assisted the test persons in recuperating copies of their personal data (through legal and technical means), and then helped them make sense of this data and the power dynamics present around it.

In addition, since the main questions were around power, we did not pick random participants: we selected our fifteen participants to be involved, in some way or another, in decisions affecting the data economy: members of the European and Finnish parliaments, EU and Finnish civil servants, NGO directors, journalists.

These fifteen participants ended up targeting dozens of companies, observed through multiple lenses (exercise of legal data rights, direct data downloads, and technical audits).

1.3. MAIN EVIDENCE ABOUT THE DATA ECONOMY

Through that process, the participants collected numerous facts about the data economy surrounding them. For instance, we found evidence of:

- online retailer Gigantti sharing individual purchase information with Meta (Facebook),
- Boeing seeking to influence Finnish MPs to sell fighter planes (including the criteria Boeing used on Twitter),
- the UK Labour party and data broker Bisnode using Meta (Facebook)'s targeting tools,
- how Google blends data from searches with wifi signals to infer at which exact location one might be,
- systematic non-compliance with General Data Protection Regulation requirements, etc

The evidence is presented as case studies drawn from the participants' experiences. They are detailed in a separate methodology and case studies report entitled "*Auditing the Data Economy through Personal Data Access*", which includes the following syntheses of the individual participant's findings:

- *Who cares about my geolocation and why?*
- *When you view the web, the web views you*
- *Move fast and capture all signals, everywhere*
- *Participants chasing their personal data*

The separate report also presents the general methodology information to increase study reproducibility. It is complementary to the analysis presented here.

1.4. SEEING THE BIG PICTURE BEYOND INDIVIDUAL STORIES

The present report is focused on communicating the data economy's big picture through a narrative that we hope will be helpful for structuring thoughts in society about systemic problems.

This narrative report contains six chapters. After the present introduction, we introduce our new narrative about the data economy in **Chapter 2,** where we describe the role of newly codified data rights for individuals in countering the aggressive assertions of the technology industry. We also establish the importance of using vocabulary from inside the technology industry – their jargon. We weave this jargon differently, to build a new vision of empowerment into the analysis of the data economy.

Chapter 3 presents the #digipower findings. It is the most important and extensive part of the report. It is structured into nine sections.

First, we start by introducing the importance of technical jargon concepts such as "context". The concept is then used for developing a new concept of *digiscape* that explains the process through which user actions are digitised in the data economy, in order to build a landscape that is influenceable. **Second**, we present a summary of case studies drawn for the individual stories of the #digipower participants that situate our findings. **Third**, we put in perspective what we consider the traditional power of the #digipower participants with the digital power of key stakeholders in the data economy. **Fourth**, we explain that the strongest form of power in the data economy can be defined as *Infrastructural Power*. **Fifth**, the infrastructural power is decomposed into four dimensions: (i) *Technical Capability* and (ii) *Organisational Capability,* which are combined for acquiring knowledge about (iii) *Individual Behaviour* and (iv) *Population Behaviour*. These are key pillars to obtain infrastructural power when they are accumulated by a stakeholder. **Sixth**, the combinations of those dimensions define four quadrants to explore the processes through which infrastructural power is exerted. We call these processes *levers* which are compelling forces to exert power. They are named as follows: (i) *Extracting Raw Data from Context*, (ii) *Converting Desire to Actions*, (iii) *Structuring Context and Content*, and (iv) *Orchestrating Behaviour*. **Seventh**, each of those levers are deconstructed in 10 specific mechanisms that enable the infrastructural power to be operational in the data economy. **Eight**, two feedback loops are developed for illustrating in a dynamic way how the infrastructural power is applied in practice and the effects that it produces over multiple stakeholders. The two feedback loops are: *Accumulating Information and Knowledge to Act*, and *Composing Complex Infrastructures for a Dominating Position*. **Finally**, in the consequences section we detail the consequences of all these mechanisms of infrastructural power acquisition, and align them to a taxonomy presented in the Appendix.

In **Chapter 4**, we offer our own reflective conclusions on the entire investigation and its findings, as well as the methodology and its potential.

In **Chapter 5**, we critically discuss contemporary alternative futures, through our new infrastructural power lens. We propose that new stakeholders with different values and oriented towards the common good need to be empowered as to see a real change in society. For starting the change, we end this chapter with the #digipower manifesto.

Finally, in **Chapter 6** we present our recommendations. The first is really cultural, and hammers home the fact that we need to change the narrative of innovation around data. The second suggests productivising the #digipower investigation itself, in order to sustain it more continuously and diffuse its associated knowledge. The third recommends *Increasing Civil Society's Infrastructural Power*. More than an idea, we provide a full plan for achieving this, by building appropriate counterpower in the data economy, in a compositional way. We also discuss the way to achieve this through the organisation of groups, empowered with data and techniques, that set up their collective interests and tackle the sociopolitical issues that affect them. We call these groups: *Data Collectives*. The final recommendation provided relates to facilitating GDPR enforcement.

Hestia.ai
sustainable data solutions

2. Creating a New Narrative about the Data Economy

In the seminal 2019 work, *The Age of Surveillance Capitalism: The Fight For A Human Future At The New Frontier Of Power*[5], Shoshana Zuboff outlines six declarations that laid the foundation for Surveillance Capitalism.

Surveillance Capitalism is defined in part as *"a new economic order that claims human experience as free raw material for hidden commercial practices of extraction, prediction, and sales"* and *"an expropriation of critical human rights that is best understood as a coup from above: an overthrow of the people's sovereignty"*.

The self-declared rights of those who profit from data today, Zuboff claims, are these six,

- We claim human experience as raw material free for the taking. On the basis of this claim, we can ignore considerations of individuals' rights, interests, awareness, or comprehension.
- On the basis of our claim, we assert the right to take an individual's experience for translation into behavioural data.
- Our right to take, based on our claim of free raw material, confers the right to own the behavioural data derived from human experience.
- Our rights to take and to own confer the right to know what the data disclose.
- Our rights to take, to own, and to know confer the right to decide how we use our knowledge.
- Our rights to take, to own, to know, and to decide confer our rights to the conditions that preserve our rights to take, to own, to know, and to decide.

These declarations that reflect how the oligopoly controlling data flows works are now juxtaposed against the increasingly empowering nature of data rights (among them the right to access, the right to delete, and the right to opt out) enshrined in legal frameworks like the General Data Protection Regulation (GDPR)[6]; the European Union's data protection law[7] that imposes obligations to service providers anywhere in the world, so long as they target or collect data related to people in the EU.These data rights have the potential to give individuals the ability to reclaim not only their data, but the power that is derived from it within what has been termed "the data economy".

The *"data economy refers to the development of a digital economy where massive scale data is collected at fast speed"*, primarily by U.S tech giants, whose *"primary economic interest lies in opportunities to influence, on the basis of data analysis and probability computation, the small day-to-day choices made by millions of people."*[8].

---

[5] Zuboff, Shoshana. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Paperback edition. London: Profile Books, 2019.

[6] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 EUR-LEX. Accessed March 2022.

[7] a concept that is distinct from privacy, in EU law.

[8] Lammi, Minna, et Mika Pantzar. "The Data Economy: How Technological Change Has Altered the Role of the Citizen-Consumer". *Technology in Society*, vol. 59, novembre 2019, p. 101157. DOI.org (Crossref), https://doi.org/10.1016/j.techsoc.2019.101157.

In a 2014 report[9], the World Economic Forum (WEF) described the power imbalance of the data economy in direct terms: *"There is an imbalance in the amount of information about individuals held by, or that is accessible to, industry and governments, and the lack of knowledge and ability of the same individuals to control the use of that information."*.

Critically, *"a crisis of trust is developing, stemming from the use of personal data in ways that are inconsistent with individuals' preferences and expectations"*. The WEF called explicitly for the systematic empowerment of individuals to control how their data is used by others and to make it of value to themselves.

This report explores the results of an attempt to put this systematic empowerment into action, and develop a new narrative about the data economy and the power of data rights.

We are building a new narrative around the data economy that is grounded in the relationship between data and power. We propose a new taxonomy for understanding the mechanisms that regulate this economy and clarify the existent terminology that dominant players of the technological market use internally.

In their 2019 article *The Corporate Cultivation of Digital Resignation*[10], Draper and Turow demonstrate a framework for understanding how industry practices produce resignation to data collection among consumers, and how this creates *"uneven power relationships between companies and publics"*. Authors demonstrate that corporate practices actively encourage a sense of helplessness towards these surveillance-based power structures and identify several key strategies used by companies. Two of these, *jargon*, and *misnaming* involve using terminology as a form of manipulation and narrative building. First, they define jargon as *"terminology that is difficult for those outside a specific group to understand"*, stating that it *"not only generates confusion, but may frustrate efforts at comprehension"*. Secondly, misnaming describes *"efforts to obfuscate practices through the use of misleading labels"*.

This report creates a new narrative around the data economy by using industry terms to clarify and make them transparent to anyone. By using these terms in a project that empowers individuals to better control and understand their data, we counteract the industry strategies of jargon and misnaming.

Building on this, the individual data stories of our participants are couched in industry vocabulary to remove the obfuscating nature of the terms and reclaim them from stakeholders that have thus far benefited from their use. Furthermore, this report is an exploration of how effectively the rights afforded by the GDPR can empower individuals to counter digital resignation and better understand the uneven power structures of the data economy.

---

[9] Kearney, A. T. "Rethinking personal data: A new lens for strengthening trust." In World Economic Forum, vol. 1. 2014. https://www.weforum.org/reports/rethinking-personal-data

[10] Draper, N. A., and Turow, J. "The corporate cultivation of digital resignation". *New media & society*, *21*(8), 1824-1839, 2019, https://doi.org/10.1177/1461444819833331

The new narrative we propose is transparent and pedagogical. Throughout we provide the necessary scaffolds (i.e. concepts and theoretical tools for simplifying procedures and skills) in order to:

- acquire the ability to understand and claim rights to data;
- access data;
- acquire the ability to analyse the data retrieved;
- articulate strategies that straddle digital and physical realms in leveraging this data to address societal issues.

In this report we additionally create accessible taxonomies so anyone can understand power dynamics.

Hestia.ai
sustainable data solutions

3. #digipower Findings

3.1. SHAPING CONTEXT

Where we are, what surrounds us, the software we use, the environment in which we are living, here and now, is the **context** of our present. What we are doing, what we are reading, what we are buying, what we are looking at, what we are working on, here and now, is the **content** of our present.

In our digitised world, context and content are two essential notions. Transformed into data, the context and content of our lives can be modified by those who observe us discussing, working, entertaining, consuming, in order to serve their interests. This is worth developing further:

Context is the surroundings of a user: e.g., a shop, a neighbourhood, a group of friends, a social network app.

Content is what is contained within the context of a user: e.g., a purchase, a street name, a professional tie with a friend, a tweet on Twitter feeds.

Beware: the boundary between context and content is sometimes blurred, as the digitisation of a physical context eventually transforms it into content, as data. A typical example would be with a map that also displays the location of shops. This digitisation process develops a new kind of power in the hands of service providers who acquire knowledge capable of influencing human behaviour and the environment in which the behaviour takes place.

Context and content are also dependent on the observer's **viewpoint**[11]. They enable observers to capture user actions according to the purpose of who observes it. For instance, an observer is a service provider like Netflix that is looking to increase the number of views on music documentaries online by young people. Therefore, Netflix wants to observe the behaviour of young male and female individuals that attend concerts, watch online movies and go to cinema in specific locations (e.g., on YouTube from home, in a nearby cinema). The concerts, cinemas and locations are contexts. Within those contexts, Netflix wants to know at which time slots young people attend, the titles of the movies they watch, how they are going to the cinema, etc. This is all content to learn *about* young people's behaviour in a contextualised way, in order to later leverage this knowledge for the purpose of increasing streaming paid views.

In that sense, context's and content's definitions vary from one observer to another as they can be approached from multiple perspectives according to the observer's purposes. Another service provider like Deliveroo would be interested in observing the same users attending the same movie, but to know what type of food they eat when they go to the cinema and their willingness to eat before going by ordering online food to their place. Another content is then necessary: where these users live (another context), in order to deliver them food (another purpose for the observer).

---

[11] Linked to Michel Foucault's extension of the notion Panopticon from Latin *panoptes*: all seeing. The argument is centred on the relationship between power and the accumulation of knowledge where the observer controls others and the rules of society by having a widespread viewpoint. Foucault, Michel. *Discipline and Punish: The Birth of the Prison*. 2nd Vintage Books ed, Vintage Books, 1995.

Ultimately the value of the data arising from the observation of users who are constantly moving between intertwined online and offline situations is intrinsically determined by the observer and their business purposes.

Why are Context and Content Important?

In the current scenario where user actions overlap offline and online contexts (e.g., using Google Maps app to move physically from home to work), as well as digital and physical content (e.g., the level of an electric car battery, the position where a charging terminal is installed in a city), it is important to understand how the perspectives of context and content change according to different types of observers. It provides assertive insights about how data is collected, about what/whom, in which situations, and finally, how data is used by service providers for acquiring knowledge about users and shaping in a retroactive manner, the context and content of users.

The data economy relies on the digitisation of context and content to exploit individual's data, and at the same time, to shape them for its own benefit. The critical point is that *the more context and content are digitised, the more information a service provider obtains about individuals' lives (where one moves, navigates the internet, buys, thinks, and socialises), to increasingly influence our context, our behaviour and even the service provider's capacity to understand our behaviour.*

The Digiscape: Our New Hotel far from Home

We have named the hybrid landscape in which we now operate: the **digiscape**. The digiscape is a kind of Middle-earth in which an incredible amount of data is circulating, linking billions of individual contexts to billions of individual pieces of content, linking the circulation of people and goods (the cityscape), to the movements of ideas, opinions and desires (the infoscape).

This digiscape is pervasive in the way it is used today by service providers for changing users' state of mind and behaviour for commercial profit. Indeed, Apple's co-founder *"Wozniak said that he and his wife both recently deactivated their Facebook accounts over data privacy concerns. When he 'likes' a friend's post, the interaction isn't about connecting with someone he knows, the Apple co-founder said – it's about revealing his interests to advertisers."*[12]

For instance, a service provider who would know a person's behaviour in the city, could use that knowledge to influence the person towards buying a more expensive car. Another service provider could be interested in knowing a person's preference towards a particular political party, as well as the person's network and economic situation, in order to guide this preference towards a specific political orientation for voting. The Cambridge Analytica case demonstrates this.

Behaviour is easily manipulated in new ways in the digiscape in comparison to our non-digitised context.

---

[12] Apple co-founder Steve Wozniak: 'Of all Big Tech, Facebook is No. 1 that I don't like' https://www.cnbc.com/2022/03/23/why-apple-co-founder-steve-wozniak-deactivated-his-facebook-account.html CNBC, March 23, 2022.

For instance, consider this analysis of a Twitter exchange. A journalist gets asked *"Why post paywalled content on a free site?"* Her response is: *"I would say that if you walk in front of the showcase of a baker's shop, it's free, but if you want the bread inside, you have to pay"*. She explicitly equates the baker's showcase with her Twitter profile feed.

The Eyeballs, a collective formed around data issues related to the attention economy, reacted[13] to explain why this analogy breaks down[14]: *"Unfortunately it is not so simple, the showcase here is 'smart'[15] and it profiles both baker and pedestrians. It shapes the showcase to maximise its learnings, and, most importantly, it monetizes the window as an advertising spot."*

However, the difference between this digital shop window and the baker's shop window also tends to diminish. When we use our supermarket loyalty card in a physical store, our purchases become data that will be used by that supermarket, and perhaps by others, to better understand and target us – also on the ecommerce site.

In other words the digiscape is a continuum between two types of landscape we already illustrated: from the **infoscape** to the **cityscape**.[16]

When the little observed actions are physical, users find themselves in the *infoscape*. At the opposite end of the continuum, when more actions observed are physical, users are in the *cityscape* (whether you are in a city or not). We call it the cityscape to simply put an emphasis on its physicality: how even your path in the physical world becomes digital.

The more your context is transformed into content, the deeper you enter into the infoscape, where it is easier to influence you through data.

The digiscape can now be understood as incorporating two parts, the cityscape (dealing with data collection and power in the physical world) and the infoscape (dealing with data collection and power over information in both the digital and physical worlds).

Building a Digiscape

The process of building a digiscape is represented in figure 1. It consists of three steps: capturing raw data from devices to digitise the context, processing the information about the context for knowledge production, and using this knowledge to (re)shape the context. The process is explained in the following.

---

[13] The Eyeballs tweet in French https://twitter.com/TheEyeballsFr/status/1509471529674936325 Twitter, March 31, 2022.
[14] Why exactly this analogy breaks down is a teaching opportunity, see in particular the second item in the *Reflective Conclusions* chapter.
[15] in the sense of a smart device, a device that collects data and has some form of intelligence embedded
[16] A term coined by researchers to explain the way the experience of the geographical and physical space is intertwined with the 5G network, see Dr. Corinne Cath-Speth tweet https://twitter.com/C___CS/status/1507709910196363267 Twitter, March 26, 2022.

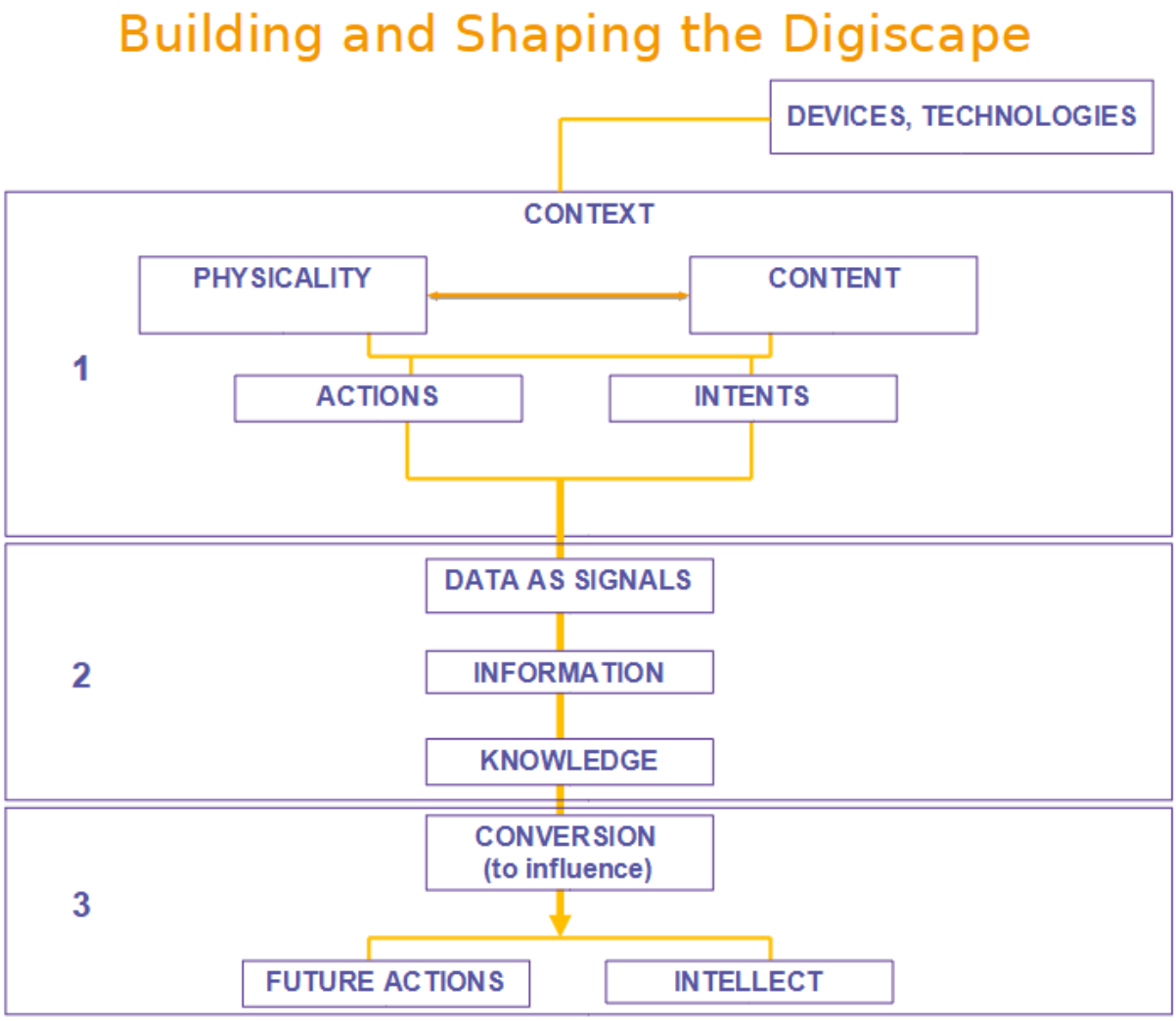


Figure 1. Building and shaping the digiscape in three steps

Step 1 (first figure box). The first step is building a digiscape through capturing the context in a digital format. Devices like smartphones, cameras, sensors, are used to digitally encode information about the context. Myriad digital technologies are also used to digitise context, for instance beacons, Bluetooth, and wifi.

These devices and technologies capture the physicality of the context, as well as its immaterial content: information. In a city, for instance, devices like satellites, cameras, and sensors, capture the city's geographical elements: the cars circulating there, the buildings surrounding the roads, as well as the names of the car brands, the images of a street, the number of traffic lights in that street, and events like the fact that some traffic lights are broken at a given time.

Another more complex and subtle example of context capture occurs by observing a person's behaviour in social media. Consider the situation where a user is viewing and reading a Twitter feed. Twitter's app is a digital context embedded in a smartphone (physical context) with

dedicated icons to navigate within. But Twitter's app also becomes a distinct context full of content based on its information architecture. The content is the feed that is presented to that person with specific tweets that are whether volunteered posts by other users followed, or sponsored posts paid by advertisers.

In a context, the most important resources for service providers to use are **actions** and **intents**. An action is something done in this context, and intents are interests of doing something in this context.

In the previous example about the city, an action is the movement of a car for turning around in one street. An intent is the interest of going from one address to another one. An address that is known by an address entered into the GPS or a search for a restaurant/store on a search engine. In the other example above about Twitter, an action would be e.g., scrolling down the feed. An intent is the user's interest in reading a type of content that can be found in that feed, which might be observed by them stopping their scrolling to read a specific tweet. These actions and intents are digitised in the form of "signals" and become the main source of information about the user's context for service providers.

Step 2 (second figure box). The second step for building a digiscape is processing the information about the context for knowledge production. It allows service providers to know more about the population's behaviour than the simple facts indicated by the signals, to serve their economic and political purposes. Knowledge production is mainly done with probability calculations. One of the methods used is statistical inference. Only the service providers capturing a vast amount of data can apply it.

In the example about the city, an important piece of knowledge for a service provider would be to know if the driver is heading to her-his home or workplace. In the example about Twitter's feed, an important piece of knowledge would be to know if the user will follow a specific user or retweet a specific piece of content. Detailed examples and additional illustrations are presented in the methodology and case studies report.

The aforementioned two steps (i.e., digitising the context and processing information for knowledge production) build a digiscape: a digital landscape of your daily life that finally shapes your physical and digital realms through a third step.

Step 3 (third figure box). The final step is a change of state of the context, that is changing a person's action, their thinking, their movement, or the direction of their self-driving car.

The actions and intents that were originally captured in the context are later converted into new actions that serve the service providers' purposes.

For instance, when leaving a bar, a person's desire to walk can be converted into an action of taking a taxi when a driver is passing nearby in the cityscape. Moreover, the individual's perception of the fare can be influenced so that they are willing to pay a high fare for a ride because their smartphone battery is low.

Changing the state of the context can also be illustrated with the Twitter feed example in the infoscape. The user action of scrolling down a feed can be later converted into the action of retweeting. To retweet a type of content in the feed (that was recommended to the user) is a way of broadcasting an idea to a person who may have initially wanted to read a different type of content.

The control over context is leveraged for the benefit of the platform at the expense of the user. For example, Meta (Facebook) has been documented to prioritise content in users' feeds that results in the expression of emotions through emojis, as this content leads to higher engagement and thus higher profits for Meta (Facebook). One of these emojis, the angry face, is considered by Meta (Facebook) CEO Mark Zuckerberg as the equivalent of a "dislike" button. However, as the Washington Post reports[17], Meta (Facebook) weighted the angry face five times more important than a "like", which means the angry face shapes greatly the context of a users' news feed. The Washington Post article also demonstrates that the angry face emoji was likely to result in a user being systematically exposed to "misinformation, toxicity and low-quality news".

A person can highlight some element of the physical landscape (the cityscape), e.g., a geolocated space, as well as the content within that landscape (the infoscape), e.g., an advertisement appearing in the user's feed for the purpose of facilitating discussions. However, the key point to establish power in the data economy is that the physical and the content are essentially always intermeshed, with the balance tilting more or less on one side (the physical or the content) depending on the situation.

For instance, an individual will read content from a particular geographic location, or will use an app to move from one place to another. However, building new technological systems enable service providers to capture some context from both physical and immaterial content at once, to base their decisions on the broad and global view they have over society. An example of how context and content are intermeshed is that a news site could make content recommendations based on inferred geographical positioning of their reader.

The digiscape shows why the digitization of context[18] is so powerful to act over society. The reason is that context is understood by its physicality and immaterial content, the latter being more easily engineered by service providers, for instance through recommender systems. Moreover, an essential dimension of the digiscape is time. All actions in the context are captured by service providers as signals in real-time, at high speed, which means that these stakeholders accumulate a fine-grained and dynamic view over populations through time.

Therefore, *people need new skills to be oriented in the digiscape* that is currently being shaped by service providers, in particular the technological industry. People need to understand how the data economy can influence our actions and intellect for profit: the way we move and think, the

[17] "Five Points for Anger, One for a 'like': How Facebook's Formula Fostered Rage and Misinformation." *Washington Post*, 26 Oct. 2021, https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/ .

[18] i.e. the encoding of context information as data

way our attention is captured, the way our opinions and interactions are shaped both online and offline.

3.2. SUMMARY OF CASE STUDIES

The individual findings within the #digipower's investigation are formulated as case studies in a dedicated report (the separate methodology and case studies report entitled *"Auditing the Data Economy through Personal Data Access"*). In this section of the narrative report, we introduce four #digipower case studies, which are detailed in that report.

The first case study, *Who collects my geolocation and why?*, introduces the concepts of physical **context**, **intent** and (purchase) **conversion**, particularly when tied to a physical place. Intrinsically, the events being tracked on physical context are primarily about geographically relevant events, such as mobility or purchase events taking place in a physical store, and thus closer to the cityscape. We start there because of the relative ease of outlining stories taking place in the physical environment, compared to the digital environment.

By contrast, our second case study, *When you view the web, the web views you*, introduces much of the same concepts but in the infoscape, when more of the context is tied to **content**. In particular we highlight examples of **interest** profiles, but also show how the context in the infoscape can be shaped more than in the cityscape. Indeed, context is shaped because Twitter is the master of your context when you use it. The way context is shaped leads to an additional feedback loop: *Accumulating Information and Knowledge to Act* (explained in section 3.8), accelerated and steered through directly commercially available offerings. A canonical example for this is the ability to target based on "Conversation Topics" on Twitter.

The third case study, *Move fast and capture all signals, everywhere*, focuses on *Facebook Custom Audiences*. We describe the data flows towards Meta (Facebook) that we observed, and the incentives leading to the wide adoption of that tool despite it seemingly contradicting some of Meta's (Facebook's) clients' self-interests. Some of our analysis is rooted in emails made public as part of ongoing lawsuits regarding antitrust and consumer protection (advertisers being the customers). Overall, this case study is a clear outline of the capability amassed by Meta (Facebook) over its clients (advertisers) and (potential) competitors.

The fourth case study, *Participants chasing their data*, describes the power companies have over individuals when they actively seek to recover their data but their personal data rights are not actionable. This power might manifest itself simply, through insufficient compliance or lack of care for respecting data rights, but sometimes can manifest itself by having complex procedures that can cause delay or extra work for the individuals seeking access, without ever providing what they are asking. It would be legitimate to ponder then whether this is intentional. In this case study, we also compare the outcomes of the three lenses pursued for accessing data: data download portals, subject access requests and the technical audit through TrackerControl.

3.3. TRADITIONAL POWER VS DIGITAL POWER

The motivation of the #digipower investigation is originally based upon the contrast of traditional power with digital power. While we consider that more traditional forms of power are held by the participants, by virtue of their occupation, digital power is held by service providers exploiting the value of the participants' personal data.

As previously presented in the methodology and case studies' report, the #digipower participants have a particular role in society. The participants are decision makers, who can be seen as embodying traditional forms of power. As holders of this traditional power, most of the #digipower participants have an everyday professional responsibility to ensure a democratic society where personal data rights are respected, and the value built around data is fairly distributed across civil society and the economic sector (or at least not capitalised by a few dominant service providers).

However, traditional power is threatened, or even confiscated, by digital power. This is evident when judges on more technological lawsuits are changed because of their lack of knowledge about digital techniques. Digital power is now held by key players in the technological industry such as Google, Meta (Facebook), Amazon, and more broadly by service providers like big groups of supermarkets, digitally savvy newspapers, as well as more traditional strategies like lobby. These service providers are powered by data and limited to the goal of increasing user engagement or retention for the economic profit and not for the common good.

Digital power is therefore exerted at the expense of people's power over their own personal data. That is why throughout the #digipower investigation the participants made the pedagogical exercise of confronting digital power as if they were regular users of digital services, which in fact they mostly were.

This exercise gave two main insights.

The first was that *the value exploited from data is today is in the service providers' possession and not in the participants' possession, despite their power and their personal data rights*.

The second – unrelated to their everyday position of traditional influence – is that the participants obtained facts, from their personal experiences, about how data value is actually exploited. This helped them understand how their traditional power is diminished by a complex system. These outputs can ultimately empower the participants' political role so that the value created via data is allocated more fairly between individuals, companies, civil society, regulatory institutions and society as a whole.

The #digipower investigation deconstructed the opacity and complexity of digital power. We have identified and illustrated the mechanisms of digital power in order to provide everyone with a better understanding of how the data economy functions and affects individuals and society.

Our main finding is that this digital power is exerted through a complex assemblage that can be documented, and that we named *Infrastructural Power*. This Infrastructural Power is established via *technical* and *organisational capabilities* that are used to produce knowledge about society in sophisticated ways, on a large scale. We define this Infrastructural Power in the following section.

3.4. INFRASTRUCTURAL POWER

Our transversal analysis of participants' experiences demonstrate the mechanisms through which an "infrastructural power", a concept we introduce here, is established in the data economy by service providers like Google, Apple and Meta (Facebook) but also by physical stores and brands that are part of our daily life interactions.

We define **infrastructural power**[19] as the combination of a leading expertise for the development and implementation of technologies, with the organisation of social and commercial relationships between stakeholders, i.e., service providers and individuals using those services. For instance, there are many relationships established between a retail store like Gigantti and third parties like Bisnode (a data broker) for data exchange, which gives these services (both retailer and third party) a lot of knowledge about their clients in the overall population.

When service providers accumulate technical and organisational capabilities, they are able to first acquire raw data about individuals and the population's behaviour; second, to acquire knowledge about their behaviour; and finally, to act in an attempt to influence individuals and the population.

It is important to note that when we refer to "individual" and "population" we refer more specifically, on the one hand, to one individual service provider or one individual person using a service; on the other hand, to a population, as a large group, of service providers or to a population of individuals using those services. The general distinction covers different stakeholders in the data economy: consumers/users, apps, platforms, institutions/companies/services digitally-mediated. This high level typology is relevant to highlight throughout the whole investigation that infrastructural power, when acquired by a service provider, dominates and affects both service providers within the market and individuals

[19] The definition is highly inspired from Bowker, Geoffrey C., and Susan Leigh Star, 2000, *Sorting Things Out: Classification and Its Consequences*. Inside technology. Cambridge, Mass: MIT Press.

using services in their lives. This is illustrated in the following two examples concerning Amazon.

On the one hand, Amazon Web Services (AWS) are a form of infrastructural power domination over other services. The reason is that many software companies are dependent on AWS for database storage and other functionalities, while also working with other services or providing similar services that compete with Amazon. According to a 2021 Bloomberg article, Amazon has the potential to use its market dominance to "punish the companies that work with other cloud providers and favour those that it works with exclusively"[20]. Amazon has already demonstrated its ability and willingness to use its infrastructural power and market dominance in the realm of physical products to undercut competitors through manipulative pricing, such as when they took a loss on diaper sales in order to facilitate the purchase of diapers.com[21].

On the other hand, one way that Amazon dominates people who use their services is through deceptive and coercive design choices in user interfaces. With an infrastructural power, Amazon can then design, rollout, and test user patterns to increase Amazon's economic profit while reducing a person's will. Indeed, a report from Business Insider shows that Amazon was able to make Prime cancellations drop by 14% by implementing a new user interface designed to divorce people from their ability to make the decision they wanted. More specifically, "[Amazon's] project created multiple layers of questions and new offers before a Prime member could cancel their subscription in hopes of reducing member churn"[22]. Consequently, it became harder for a user to cancel a paid subscription and more generally, to know how her-his data is used by Amazon for other purposes than the ones that were initially consented. These are called dark design patterns that have a negative influence on the subject's decision-making process regarding her-his privacy[23].

Similar to the AWS example above, a final example of infrastructural power is the relationship between Meta (Facebook) and the companies that it treats as both customers and competitors. When Meta (Facebook) Marketplace was being developed, Meta had to consider whether retailers that rely on Facebook advertising would stop paying when it became clear Meta's

---

[20] McLaughlin, M., Bass, D., and Nix, N. "Amazon Cloud Unit Draws Antitrust Scrutiny From Khan's FTC." *Bloomberg*, 22 Dec. 2021, https://www.bloomberg.com/news/articles/2021-12-22/amazon-cloud-unit-draws-fresh-antitrust-scrutiny-from-khan-s-ftc?sref=TBDibEcD

[21] Lecher, Colin. "How low prices could make for an antitrust case against Amazon." *The Verge*, 13 May 2019, https://www.theverge.com/2019/5/13/18563379/amazon-predatory-pricing-antitrust-law

[22] Towey, H., and Kim, E. "Amazon used a sneaky tactic to make it harder to quit Prime and cancellations dropped 14%, according to leaked data." *Business Insider*, 15 Mar. 2022, ,https://www.businessinsider.com/amazon-project-iliad-made-cancel-prime-membership-harer-leaked-data-2022-3

[23] Jarovsky, Luiza. "Dark Patterns in Personal Data Collection: Definition, Taxonomy and Lawfulness." 1 Mar. 2022, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4048582#

desire was to be a direct competitor. This is described in a August 19, 2015 conversation between former ebay/Paypal employee and then Meta (Facebook) product manager Mary Ku and CEO Mark Zuckerberg:

"Impact on ads business: While we are building Marketplace for the long term, if we are not careful, we can have a short term negative impact on the ads business before we build out sustainable value. Several large advertisers are marketplaces and multi-channel retailers who may find our launch threatening to the extent that they may decide to pull ad spend or investment in key strategic ad products (e.g., dynamic product ads). The Facebook marketplace is good for partners who themselves are not marketplaces but clear messaging and value exchange will be needed to help them understand our intentions and value proposition."[24]

3.5. FOUR INFRASTRUCTURAL POWER DIMENSIONS

We defined four #digipower dimensions of Infrastructural Power to orient ourselves in the data economy, as when using a compass.

The dimensions will assist anyone in understanding:

- how infrastructural power gets established so it shapes the data economy;
- the consequences of this power over individuals, social life, and the innovation market.

The dimensions provide a structuring tool for building a transparent view on the data economy, based upon evidence from the #digipower participants' data. Through these dimensions, individuals gain legibility[25] over the infrastructural power being exerted on them, but also the capacity to organise their learnings and to relay these learnings to others. Ultimately, it upskills any reader of this report to regain some form of control over their personal data flows, and to help those around them to do the same.

There are four dimensions (Figure 2) in the infrastructural power: individual behaviour, population behaviour, technical capability and organisational capability.

---

[24] Lawsuit. Court Filing: MAXIMILIAN KLEIN, et al. Plaintiffs, v. FACEBOOK, INC., Defendant., Court:United States District Court, Northern District of California, legal reference20-CV-08570-LHK (N.D. Cal. Jul. 20, 2021), Document 244-3, https://www.courtlistener.com/docket/18714274/244/3/klein-v-meta-platforms-inc/

[25] Mortier, Richard and Haddadi, Hamed and Henderson, Tristan and McAuley, Derek and Crowcroft, Jon, Human-Data Interaction: The Human Face of the Data-Driven Society, 2014, http://dx.doi.org/10.2139/ssrn.2508051

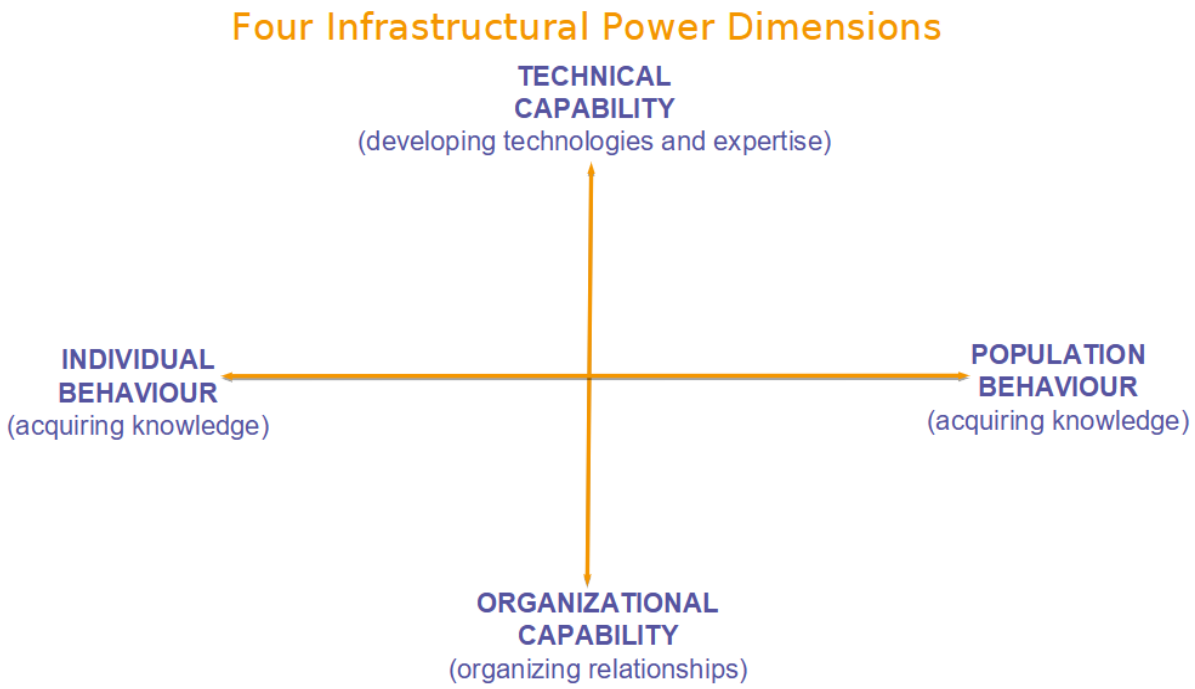


Figure 2. The four #digipower Dimensions for explaining Infrastructural Power

These dimensions present how infrastructural power is exerted in the data economy: When they are combined, infrastructural power is exerted by a service provider to acquire knowledge about individuals and the population, which also enables a service provider to act, based on this knowledge, over others and their context. As we explain the way the combination of dimensions become powerful, these dimensions could also help society in finding ways to obtain and then wield power for the common good.

Service providers can capitalise the value they exploit from knowledge acquisition about individuals' and population's behaviour when they have the capability of developing techniques such as the Facebook Pixel and organising the relationships (e.g., who uses the Facebook Pixel) within the data economy. It allows them to collect a vast amount of data, as well to produce novel insights about populations and the market.

The left hand side of the horizontal axis in figure 2 refers to the first power dimension of acquiring knowledge about individual behaviour. The right hand side, refers to the second power dimension of acquiring knowledge about population behaviour.

The top of the vertical axis in figure 2, refers to the third power dimension of technical capability for developing technologies and expertise for acquiring knowledge. The bottom of the vertical axis refers to the fourth power dimension of organisational capability for organising relationships in order to acquire knowledge[26].

[26] These power dimensions and their corresponding definitions are inspired by the following work: Jon (Sean) Jasperson, Carte, T. A., Saunders, C. S., Butler, B. S., Henry J. P. Croes, & Weijun Zheng. (2002). "A

In the following we define the four dimensions.

Dimension 1: Individual Behaviour

In this dimension, power is held through knowing detailed information about individual's behaviour. For example, by collecting an individual's geolocation and device ID one can know where a person lives.

This is often the first type of power that people consider as a threat from data collection, leading to the common adage "If I have nothing to hide, I have nothing to fear". This form of power is demonstrated well through data brokers that act as credit reporting agencies, such as Experian: As described in Citron and Solove's taxonomy of privacy harms[27], individual data provides power over individuals' lives based on the integrity of the information and by removing from the control of the individual with whom the information can be shared – in other words someone would have power over you if you can't prevent them from sharing false information about you. In the case of Experian, an inaccurate profile with a false bankruptcy notation can keep an individual from receiving a loan, which is a measurable form of economic power over the individual that Experian controls.

Dimension 2: Population Behaviour

In this dimension, power is obtained by knowing a population's behaviour through data. For example, Google Search Trends will be indicative of the main topics different populations are concerned about around the world.

Another example would be finding in which locations diseases are spreading, as measured through the symptoms typed by the population into search engines, which show specific queries in certain locations[28].

A last example is the power of Google maps to measure traffic patterns based on population movements. The (often passive) collection of location information from devices offers vast insight into human movement. Its pervasiveness is often not fully considered until it produces consequences outside of Google's internal uses: in the physical realm and in vulnerable contexts. This has been powerfully demonstrated in Google maps usage by citizen investigators to track the Russian invasion of Ukraine[29]. Google made the decision to unilaterally close down Google maps' service in Ukraine as a result of this usage by investigators. This policy response

---

Metatriangulation Review", *Power and Information Technology Research*, MIS Quarterly, 26(4), 397–459. https://doi.org/10.2307/4132315

[27] Citron, D. K., and Solove, D. J. "Privacy harms." *GWU Legal Studies Research Paper No. 2021-11*, 9 Feb. 2021, https://dx.doi.org/10.2139/ssrn.3782222

[28] Experience has shown, however, that one should be careful in assessing such tools long term. See https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/

[29] Gordon, A., and Gault, M. "Google Maps Live Traffic Showed the Russian Invasion of Ukraine." MOTHERBOARD Tech by Vice, 24 Feb. 2022, https://www.vice.com/en/article/xgd7dd/google-maps-live-traffic-showed-the-russian-invasion-of-ukraine

demonstrates that the power derived from data for knowledge acquisition on population behaviour can have critical consequences to the lives of many[30].

Dimension 3: Technical Capability

In this dimension, power is obtained through the capability of developing state-of-the-art technologies and having a recognised know-how for acquiring knowledge. One example of a technical capability is Google and Apple's development of application programming interfaces (APIs), which serve as communications between platforms for capturing raw data. Meta's (Facebook's) very permissive API system was what allowed for the abuses of data in the Cambridge Analytica-Facebook scandal[31].

Third party libraries and software development kits (SDKs) within apps are pervasive and increasingly scrutinised sources of the technical capability to amass infrastructural power. SDKs are code incorporated into an app but built by a third party, such as Google. In this way, the developers of SDKs accomplish a broad reach and often collect data on the app users. Often they are necessary for using standard services, such as Google Maps[32]. As discussed by Feal et al. (2021) in their report *Don't Accept Candy From Strangers: An Analysis of Third-Party Mobile SDKs*[33], these SDKs are capable of undermining children's protections against data collection and provide social networks with ways to collect information outside of their internal platforms, to be later recombined. Paul-Olivier Dehaye (an author of this report) and Joel Reardon[34], an author of *Don't Accept Candy from Strangers*, have also outlined ways that vulnerabilities in SDKs, combined with their ubiquity, could be used to manipulate COVID-19 contact tracing efforts and execute attacks that simulate outbreaks of the disease.

Another example of this form of power is the use of machine learning methods for creating new knowledge by making inferences about the data extracted. Service providers have the technical and commercial architecture to collect large amounts of data and train probabilistic models. Therefore, they have the ability to generate new knowledge that can provide insights about individuals even if the raw data extracted is lost, intentionally deleted, anonymized or removed (whether to appear more privacy-friendly or to avoid the regulatory constraints of the GDPR

---

[30] Jamal, Urooba. "Google Maps disables live-traffic feature for Ukraine after reports it was being used to track ground activity during Russian invasion." yahoo!news, 28 Feb. 2022, https://news.yahoo.com/google-maps-disables-live-traffic-144053976.html

[31] Albright, Jonathan. "The Graph API: Key Points in the Facebook and Cambridge Analytica Debacle." *Tow Center*, 21 Mar. 2018, https://medium.com/tow-center/the-graph-api-key-points-in-the-facebook-and-cambridge-analytica-debacle-b69fe692d747

[32] Maps SDK for Android overview https://developers.google.com/maps/documentation/android-sdk/overview Google, Accessed March 2022

[33] Feal, Á., Gamba, J., Tapiador, J., Wijesekera, P., Reardon, J., Egelman, S., & Vallina-Rodriguez, N., 2021, "Don't Accept Candy from Strangers: An Analysis of Third-Party Mobile SDKs." *Data Protection and Privacy*, Volume 13: Data Protection and Artificial Intelligence, 13, 1.

[34] Dehaye, P. O., & Reardon, J., 2020, "Proximity tracing in an ecosystem of surveillance capitalism.", In *Proceedings of the 19th Workshop on Privacy in the Electronic Society,* (pp. 191-203).

and/or a local law). They also retain the capability of possibly linking this raw data with more precise or broader physical context in the future, as sensors improve.

Dimension 4: Organisational Capability

In this dimension, power is obtained through the capability of organising relationships for acquiring knowledge: relationships between service providers, and between service providers and service users.

A simple but effective demonstration of organisational capability is the automated phone line. The service provider puts the individual in a situation where they are coerced into a relationship created by the provider, and the individual likely has no alternative. The individual is forced to interact step by step through the phone line's options without much recourse, and the provider is able to pivot the conversations in desired directions to or away from solutions and human actors (for cost saving reasons). This form of organisational power is already growing into a system for developing the other three forms outlined above, via voice profiling[35].

Free Basics, also known as Internet.org, "is a partnership between [...] Facebook and six companies that plan to bring affordable access to selected Internet services to less developed countries."[36] The perfect example of a partnership service that takes advantage of places with weak internet infrastructure to create dependence on corporate ecosystems, i.e., a system formed by the interaction of stakeholders with their physical environment and linked to commercial interests, among vulnerable people. By controlling access to the internet and funnelling it through their ecosystem, Meta (Facebook) is able to control and coerce data collection.

3.6. HOW THE INFRASTRUCTURAL POWER OPERATES: FOUR QUADRANTS AND LEVERS

There are four quadrants that explain how the infrastructure becomes powerful, they are inhabited by four levers that offer an advantage to service providers. The four quadrants and respective levers are presented in figure 3.

[35] Turow, Joseph. "Opinion | Hear That? It's Your Voice Being Taken for Profit." *The New York Times*, 12 Sept. 2021, https://www.nytimes.com/2021/09/12/opinion/voice-surveillance-alexa.html.
[36] Internet.org, https://en.wikipedia.org/wiki/Internet.org Wikipedia, accessed March 2022.

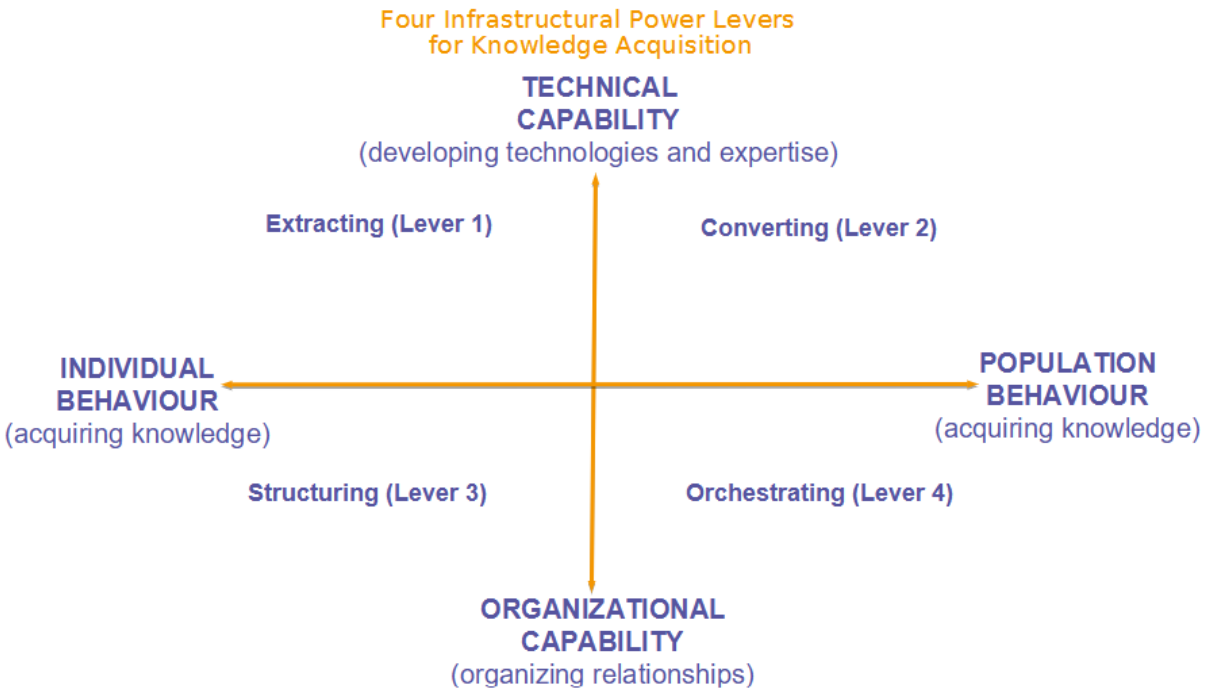


Figure 3. Four Infrastructural Power Levers

Lever 1 (Top-left Quadrant): Extracting Raw Data from Context

The first lever "extracting raw data from context" is performed by a service provider in the top-left quadrant. It combines two dimensions: technical capability and individual behaviour. It concerns any service provider that develops technologies and a leading expertise for processing data to acquire knowledge about individual behaviour, either directly when the individual uses a service provided by that service provider, or indirectly when third party services use technological building blocks it provides.

This first lever is deployed by service providers to know about intents in the digiscape, which are extracted from an individual's behaviour in context. When content and context are extracted as raw data, service providers seek to know about intents, in other words, to know an individual's interests and if the individual would be inclined to perform an action that was not previously planned.

More technically speaking, content about individual behaviour in the digiscape comes in the format of signals, e.g., a raw pixel when browsing a website built up into a data acquisition tool like Facebook Pixel.

Content is active or passive. Active content means that intents are are fulfilled by the individual's direct request online or offline. Passive content means that the system provokes intents through suggestions presented to the individual without her-his direct request.

Lever 2 (Top-right Quadrant): Converting Desires to Actions

The second lever "converting desires to actions" is performed by a service provider in the top-right quadrant. It presents the combination of two dimensions: technical capability and population behaviour. This means that service providers develop technologies and a leading expertise for processing data to acquire knowledge about population behaviour, that is a group of service providers and a group of individuals using those services, as well as the interactions between them.

This second lever is materialised by the conversion of pre-existing desires of populations into particular actions of interest to the service provider, thanks to the development of techniques. The ability to convert desires originates with the extraction of content, considered as intents that can be used to create new desires in the population.

More specifically, desires are predictions (sometimes relevant, sometimes not) made by service providers on the basis of past intents or actions, initially collected as raw data. The goal is to predict what an individual could desire in the future in order to guide this desire and convert it into a concrete action in the data ecosystem. For instance, to sell those desires to advertisers that pay for placing ads that seek to increase their sales[37].

These new actions are usually linked to commercial transactions but they may vary from one service provider to another, depending on its definition of conversion (rate). Indeed, a service provider like Meta (Facebook) could be interested in changing a person's opinion about a specific topic, whereas a retailer could be interested in ensuring that a person enters their physical store. These changes of states are the embodiment of "conversion" mechanisms.

Generally speaking, a conversion can be defined as an event (e.g., a click, entering into a physical store) that is digitally captured and considered as the moment at which an individual fulfils the service provider's will, whether its purpose is ensuring a commercial transaction (e.g., buying cigarettes, subscribing to exclusive content in a newspaper, virtually trying on a t-shirt[38]) or not.

---

[37] Apple co-founder Steve Wozniak: 'Of all Big Tech, Facebook is No. 1 that I don't like' https://www.cnbc.com/2022/03/23/why-apple-co-founder-steve-wozniak-deactivated-his-facebook-account.html CNBC, March 23, 2022.

[38] Amazon's new AI technique lets users virtually try on outfits https://venturebeat.com/2020/06/05/amazons-new-ai-technique-lets-users-virtually-try-on-outfits/ Venturebeat, May 6, 2020.

Lever 3 (Bottom-left Quadrant): Structuring Context and Content

The third lever “structuring context and content” is performed by a service provider in the bottom-left quadrant. It is located at the meeting point of two other infrastructural power dimensions: individual behaviour and organisational capability. This means that relationships are organised between a service provider and an individual using that service to acquire knowledge about the individual behaviour of different stakeholders.

It is the lever of deciding on the context structure and its associated content through the organisation of relationships in the data economy. A typical example would be Facebook forcing websites that want to offer a sharing functionality for their content to include a button that also sends data back to Facebook when the page is loaded, even if the user does not activate that button.

Lever 4 (Bottom-right Quadrant): Orchestrating Behaviour

The fourth lever “orchestrating behaviour” is performed by a service provider in the bottom-right quadrant. At the intersection of the two dimensions "population behaviour" and "organisational capability", it is used to organise relationships between service providers and people using these services to acquire knowledge about population behaviour.

An example would be an operating system designer, that decides which types of data apps can share between them, or which sensor data each app can access.

Activating this fourth lever allows one service provider to decide how others (groups of individuals and other service providers) should organise themselves and to give them guidance on how to organise themselves. This orchestration controls the acquisition of knowledge.

3.7. HOW INFRASTRUCTURAL POWER OPERATES: MECHANISMS

For each lever presented above we develop the mechanisms that make these levers operational. All the mechanisms related to its corresponding lever are presented in figure 4.

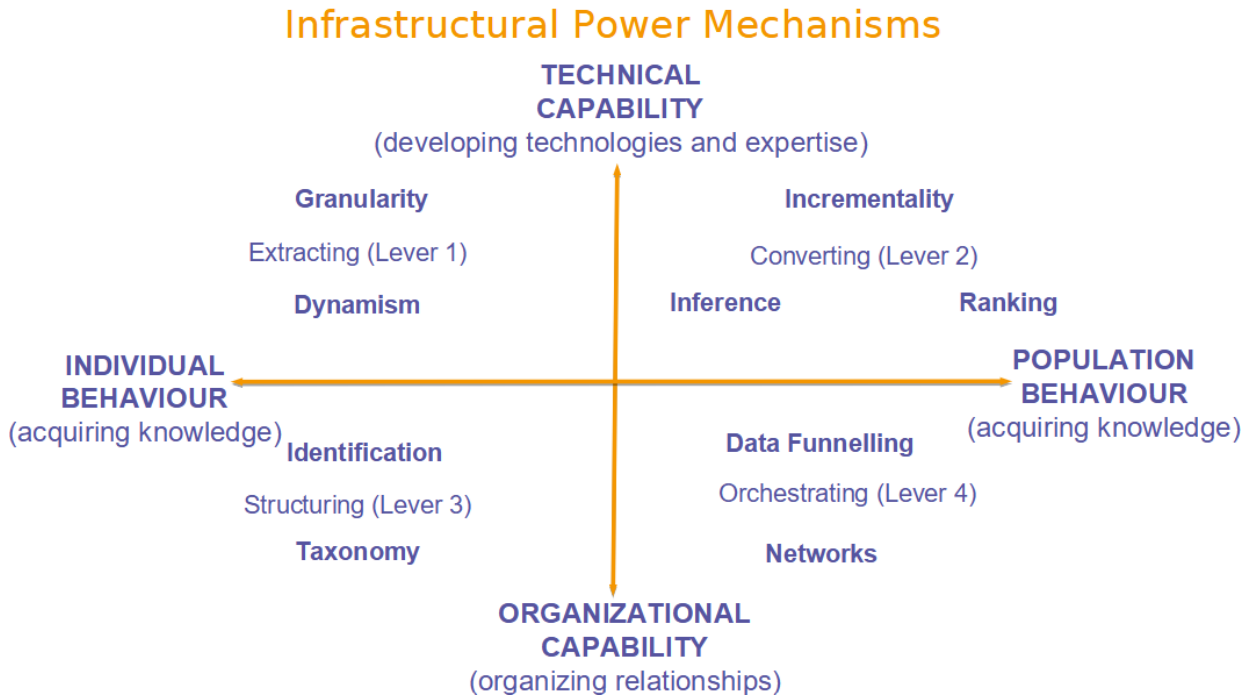


Figure 4. Infrastructural Power Mechanisms

Mechanisms for Deploying Lever 1: Granularity, Dynamism

Two mechanisms for lever one "extracting intent from context" are: (i) granularity, (ii) dynamism.

Granularity is the mechanism to deploy a more fine-grained level of extraction: It allows service providers to know about an individual's most intimate context (heartbeat, for example) through signals about the human body, and on other hand, about the individual's external context (e.g. the use of electric car charging terminals) through signals about their environment.

Dynamism is the mechanism to extract content in real-time (i.e., in milliseconds) and at high speed throughout time, so content tracks the individual's digital biography. Dynamism refers to faster extraction and is complementary to granularity. It shows that service providers go beyond capturing static data.

Mechanisms for Deploying Lever 2: Incrementality, Inference, Ranking

There are three mechanisms for lever two "converting desires to actions": (i) incrementality, (ii) inference, (iii) ranking.

Incrementality is the mechanism to expand the quantity and multiplicity of data collected. It is to capture more data about the population's behaviour through constant increments of diverse

content in different contexts for being capable at a later time of knowing more. For example, gathering more data points of a user action (e.g. location, network, browser), or gathering more data points over longer time periods.

Inference is the mechanism for knowing new insights about the population's behaviour through computational analysis (statistical predictions) applied to existing data. For instance, a technique used by Google is semantic enrichment. When a user searches for an address on Google maps, Google is capable of, through inference, assigning a label "home" to this address. This means Google infers where the user lives even though the user did not make this explicit. Thus, Google captures new insights about the user's context.

Ranking is the mechanism for the attribution of a numerical value to the information extracted from users actions, e.g., qualities describing users, their behaviour within an app, and their relationship with others. Ranking, for a computing system, consists of placing information in a geometric space based on the value attributed. On the graphical user interface side, it is the technique of assigning a hierarchical position to information, e.g., a specific order that can be displayed in the list of the recommendation results of a recommendation system.

For instance, the order in which profiles are recommended to users in the dating app Tinder is based on their performance. Among multiple criteria considered, performance is defined by the number of likes that a profile receives, this is a mark of interest on a person's profile. Based on the number of likes received, a ranking is applied to assess the user's attractiveness in the application. This attractiveness positions a user in a specific order in the results' list. A user who receives many likes is presented in a better position (at the top of the list) than another user who rarely receives likes and has a worse position (at the bottom of the list). Thus, the former is more visible than the latter for finding a date.

A ranking is a lever of power as it enables service providers prompting actions in an organised way, sometimes optimised towards their own knowledge acquisition (e.g., who is more attractive according to the number of likes received).

Mechanisms for Deploying Lever 3: Identification, Taxonomy

The mechanisms for lever three "structuring the context and content" are two: (i) identification, (ii) taxonomy.

Identification is the mechanism to identify what is relevant to know to organise entities (e.g., an object, situation, company, person, transaction). It structures the relationship to be established between a service provider and an individual according to their description, as well as what can be known by whom in the relationship. One example is fingerprinting: "the process where a service gathers little bits of information about a user's machine, and puts those pieces together

to form a unique picture, or "fingerprint," of the user's device".[39]. This allows for corporate entities to structure the relationship between themselves and individuals through retargeted advertising aimed at increasing click-through rates. Another example relates to Facebook Pixel[40]: a technique that requires prior to its implementation to identify entities in order to measure the conversion of message exposure into action.

Taxonomy is the mechanism to form groups with techniques and organise the relationships of those groups with high-level classifications into which different entities previously identified can be placed. For instance, users are classified into advertising categories by their interests and their personal information by the Swedish digital marketing company Bisnode.

Mechanisms for Deploying Lever 4: Data Funnel, Networks

The mechanisms for lever four "orchestrating behaviour" are two: (i) data funnelling, (ii) networks.

Data funnelling is the technical mechanism of a service provider to direct the flows of data by organising data transactions and service providers' relationships. In other words, a service provider decides and directs who gets what for knowledge production. The transactions are made in two main forms: exchange, leakage.

First, data transactions are made in the form of exchange to gather more data from different sources or to obtain an additional expertise.

One particular example of this exchange is called "reciprocity", which is a form of power that one service can have over another by coercing or forcing a competitor to provide access to their data in exchange for access to the service's data or support in some way. This is a term and strategy explicitly used by Meta (Facebook). This strategy turns Meta (Facebook) into a broker among competitors and helps ensure Meta's (Facebook's) market dominance. In 2012 Osofsky, the then head of Facebook's Platform, described the rationale behind internal changes that limited API access and demanded reciprocity: "Policy changes: define competitive networks + require they have a deal with us, regardless of size. Maintain size-based thresholds for all other developers to force business deals. Require data reciprocity for user extended info to ensure we have the richest identity."[41]

---

[39] "What Is Fingerprinting?" Surveillance Self-Defense, *Electronic Frontier Foundation,*, 7 May 2020, https://ssd.eff.org/en/module/what-fingerprinting.
[40] "Facebook Pixel.", *Meta for Business*, Accessed 1 Apr. 2022, https://www.facebook.com/business/learn/facebook-ads-pixel
[41] Lawsuit. Court Filing: MAXIMILIAN KLEIN, et al. Plaintiffs, v. FACEBOOK, INC., Defendant., Court:United States District Court, Northern District of California, legal reference20-CV-08570-LHK (N.D. Cal. Jul. 20, 2021), Document 244-3, https://www.courtlistener.com/docket/18714274/244/3/klein-v-meta-platforms-inc/

A well known example of the dependence that other platforms have on Facebook APIs that still exist is the following. The Facebook Login SDK[42] provides multiple service providers a technical benefit: to authentify an individual's identity when registering into a service. While Meta (Facebook) enables those service providers that installed the SDK to access selected personal data of individuals like name, date of birth, friends, Meta (Facebook) asks in exchange receives the personal data extracted from the other service provider about the same individuals[43].

Another example is Deliveroo which provides the platform's data extracted from the population behaviour to a service like Paypal for obtaining a payment feature. The data can also be given to the state for understanding the public space and traffic, as Uber did by establishing data transactions in the Netherlands with local governments.

Finally, data transactions are made in the form of leakage. Malevolent actors can leak a service provider's data that concern the private life of individuals using that service. They can ask something in exchange for not making the database publicly accessible, or they can just use it for other purposes unknown without requesting anything. For instance, a hacker leaked the dating app Ashley Madison[44] and revealed the sexual preferences of the dating app's individuals.

Networks is the mechanism to organise links with the use of techniques between service providers and the population to know more about them. For instance, service providers use a "knowledge graph"[45] technique (also known as semantic network) to identify and make sense of interlinks and their content through labels and a graphical representation of the entities that are linked.

Service providers organise the association of individuals with other individuals and service providers to form social groups and to capture the interactions happening between them. Because of network effects, the more of these associations a service provider has under their control, the more useful the platform is for individuals and the more difficult it is for them to

---

[42] "Web - Facebook Login - Documentation." *Facebook for Developers*, Accessed 31 Mar. 2022, https://developers.facebook.com/docs/facebook-login/web/
[43] Guess What? Facebook Still Tracks You on Android Apps (Even If You Don't Have a Facebook Account), *Privacy International,* 27 Feb. 2020, https://web.archive.org/web/20200227161054/https://privacyinternational.org/blog/2758/guess-what-facebook-still-tracks-you-android-apps-even-if-you-dont-have-facebook-account
[44] "Ashley Madison data breach", Wikipedia, 2015 https://en.wikipedia.org/wiki/Ashley_Madison_data_breach
[45] What is a knowledge graph? https://www.ibm.com/cloud/learn/knowledge-graph IBM, Accessed March 2022.

leave or for competitors to arise[46]. Service providers use this control over social capital[47] to maximise benefits to themselves.

3.8. TWO INFRASTRUCTURAL POWER LOOPS

The infrastructural power tends to operate in the digiscape through two dynamics that are configured as loops (Figure 5).

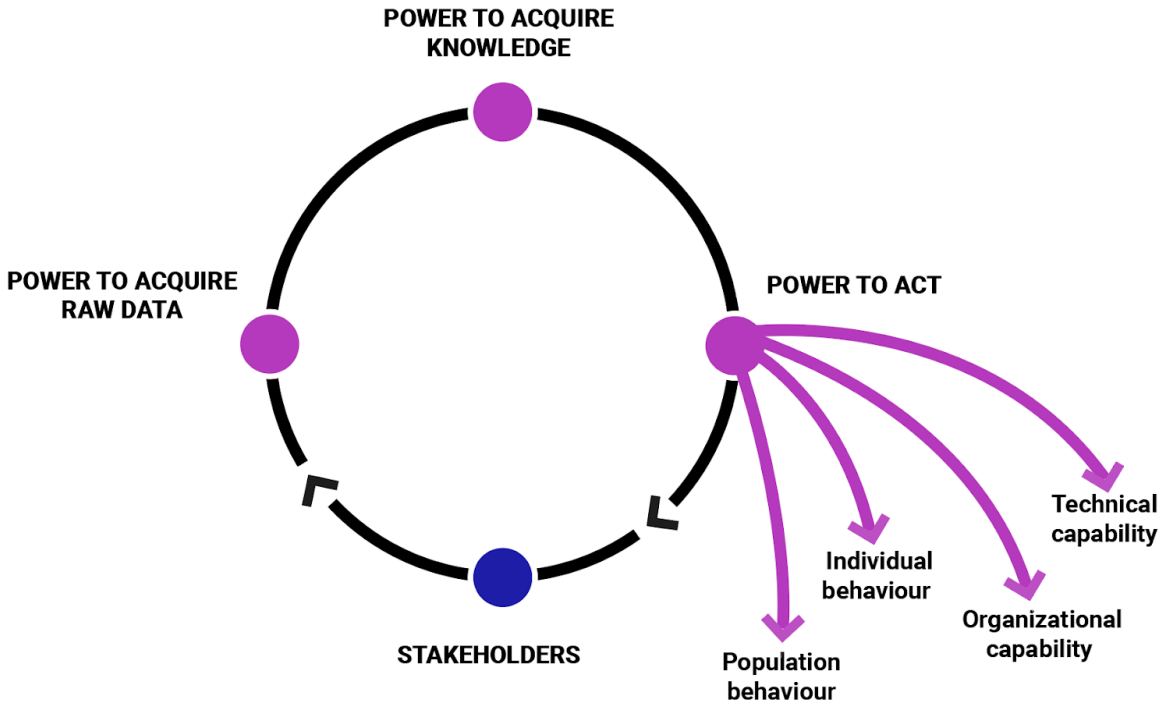


Figure 5. Infrastructural Power Loops

[46] Brien, B. and Cyphers, D.. "Facing Facebook: Data Portability and Interoperability Are Anti-Monopoly Medicine." Electronic Frontier Foundation, 24 July 2018, https://www.eff.org/el/deeplinks/2018/07/facing-facebook-data-portability-and-interoperability-are-anti-monopoly-medicine .

[47] Ellison, N. B., Steinfield, C., and Lampe, C. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. Journal of computer-mediated communication, 12(4), 1143-1168, 2007, https://doi.org/10.1111/j.1083-6101.2007.00367.x

Accumulating Information and Knowledge to Act

The first loop refers to the control of information and knowledge to dominate the population. It gives the power's holder the capacity of accumulating knowledge to act over stakeholders, i.e., a service provider, a user, a group of service providers, and users.

The loop operates as follows:

- acquiring raw data about the context
- acquiring knowledge about stakeholders in a contextualised way
- acting over context and stakeholders

The more context is defined through content (e.g., on a webpage, in an e-commerce site) by acquiring raw data, the easier it becomes to enrich the service provider's understanding of the context (e.g., through Natural Language Processing, A/B testing methods) by acquiring knowledge, but also the easier it becomes to shape that context (e.g., through matching or recommendation systems). The dynamic of acquiring raw data and knowledge creates an accelerating feedback loop whereby the context can be shaped to act over multiple stakeholders, at individual and population levels. For instance, influencing a person's state of mind (the intellect), creating unbalanced dependency relationships, diminishing an individual's critical capacity, and provoking the loss of singularity (i.e., the quality of feeling one of a kind).

The loop acts upon two types of stakeholders: service providers and users.

A first example refers to two dominating platforms in the technology industry that have developed sophisticated systems for knowledge acquisition. A platform like the iPhone operating system (iOS), or the Android operating system combined with the device manufacturer Samsung, is able to extract more knowledge about context from the raw data extracted through sensors than what a SDK or an app knows. However, the SDK knows more about the context than the app can know, not necessarily in terms of raw data but in terms of acquiring knowledge about population behaviour with technical capability. More specifically, through inference; a mechanism for deploying lever 2 (section 3).

The consequence of this loop is that the power's holder increases its knowledge by leading the technological state of the art and organising relationships for data capture, while the power holder decreases the knowledge of others by putting them into a dominated position with limited relationships and limited technologies to acquire knowledge. This produces a fundamentally unfair innovation dynamic in the data economy.

A second example refers to drivers of transportation systems like Uber, Lyft, Deliveroo and Free Now. Drivers find themselves in an asymmetric power relationship with respect to the service

providers. Indeed, drivers do not have the power to know about and to make decisions on matters that directly concern them, e.g., fixing prices, rating systems, assigning rides.
Drivers also have an obfuscated perception of their peers (other drivers) and riders (clients) using the app in comparison to the service provider. Consequently, by having the power of extracting raw data, service providers know more about the drivers' and riders' contexts than the workers themselves, e.g., where is the offer and demand, how many drivers are available, how many clients are waiting for a ride. While a driver only has access to the app interface and her-his surroundings offline, the extraction of raw data enables service providers to acquire knowledge from multiple drivers and riders with sophisticated techniques.

Service providers can profile individuals and develop a matching system for assigning specific drivers to riders. They can predict, through inference, risky events and the probability of a driver or a rider being an aggressor[48]. Service providers also fix the service fee of every trip to charge users according to every rider's profile (e.g., level of battery, location's social class), as well as the corresponding commissions for every driver (e.g., according to their rating). All these actions shape the working conditions of drivers and how much they have to earn according to fluctuant and arbitrary context conditions. It affects furthermore, the individuals' budget when predicting how much they are willing to pay for a ride according to their context as structured by the service provider.

Composing Complex Infrastructures for a Dominating Position

The second loop refers to the control of the innovation market through the term compositionality we introduce here from computer science vocabulary. This means that a service builds a complex technical architecture that is composed of multiple simpler architectures. It enables a power's holder to organise relationships between architectures to its benefit, this way knowing more about the population's behaviour across services. The most impactful decisions are about the protocols, the rules in which every composition is made, and the communication that is possible between services.

This loop accumulates the four dimensions (i.e., technical and organisational capabilities, individual and population behaviour) presented in section 3.5 for deploying each lever (i.e., extracting, structuring, orchestrating, converting). When service providers accumulate these dimensions and are able to deploy all levers, they can build a complex infrastructure for having an advantaged position, and placing others in a quasi permanently disadvantaged position.

The loop operates as follows:

---

[48] Belle Lin, "UBER PATENTS REVEAL EXPERIMENTS WITH PREDICTIVE ALGORITHMS TO IDENTIFY RISKY DRIVERS", TheIntercept_, October 30 2021, https://theintercept.com/2021/10/30/uber-patent-driver-risk-algorithms/

- Developing techniques for digitising individual and population behaviour
- Establishing relationships for digitising more individual and population behaviour than others
- Developing techniques for acquiring knowledge about individual and population behaviour
- Establishing relationships for acquiring more knowledge about individual and population behaviour than others

Because we focused in this investigation on retrieving personal data through an individual's perspective, we ended up seeing only a small part of this feedback loop, concentrated around the initial data collection and sometimes around the eventual outcome. This limitation is detailed in the methodology and case studies report. It is worth considering however that this mechanism is sustained through very long term actions by dominating players, often of the *embrace-then-choke* model. Some examples include:

- Google inked a deal with Mozilla for inclusion of Google Search in the Mozilla Firefox search bar, introducing a dependency that has been hard to wean off from. This has lead to 86% of Mozilla's 2021 revenue coming from Google[49].
- Google developed the browser Chrome. They eventually embraced interoperability standards for web extensions with Mozilla Firefox (WebExtensions API), but recently unilaterally upgraded the "standard" in a way that would make it impossible to develop effective ad blockers. This has led the developer of U-Block Origin, a major adblocker with credible claims of independence from Google (unlike say Adblock Plus, which receives a major part of its revenue from Google) to announce they might just have to throw in the towel. The current situation is that Firefox Mozilla Foundation is trying to figure out an independent road forward[50].
- Schema.org is a Google-initiated scheme that has obtained support from major other online service providers but is still dominated by Google[51]. It aims to facilitate data transfers between service providers. Its focus has been to facilitate indexing of web content, but the same mechanisms could also be used to facilitate transfers of data between providers, towards data portability (such as photos etc). However when the time came to implement GDPR data portability, Google, Meta (Facebook), Microsoft, Twitter etc went for a "clean room schema", which allowed them coupling the format of the data transferred with the transfer protocol itself[52]. The end result was dead on arrival.

---

[49] *Can Mozilla shake its Google addiction?*, AdExchanger, december 2021, https://www.adexchanger.com/ad-exchange-news/tuesday-14122021/
[50] *Manifest v3 update*, Mozilla Blog, May 2021 https://blog.mozilla.org/addons/2021/05/27/manifest-v3-update/
[51] *About*, schema.org, https://schema.org/docs/about.html
[52] *Integrate with schema.org*, Data Transfer Project Github, March 2018, https://github.com/google/data-transfer-project/issues/306

- Allegations in the *Six4Three* case, later substantiated by documents made public by a UK Parliamentary committee, show[53] that Meta (Facebook) engineered its permission system so that they could leverage access to extended functionalities in order to favour their own access to user data. Once they were sufficiently entrenched *and* could come up with a narrative disguising the move as a privacy friendly effort, they shut down access to APIs while differentiating even more between beneficiaries of long term exemptions. This state of affair, which played out until around 2015 (i.e. longer than it should have), directly opened the door to the very large scale transfer of data by Facebook to Cambridge Analytica.

3.9. CONSEQUENCES OF INFRASTRUCTURAL POWER MECHANISMS

The infrastructural power consequences on different stakeholders have been presented throughout the whole report: when defining power's dimensions (section 3.5), levers (section 3.6), mechanisms (section 3.7) and feedback loops (section 3.8). Those consequences are related to contemporary scenarios illustrating the way powerful service providers' operate in the data economy over society. The scenarios are supported with industry's vocabulary, lawsuit allegations and journalistic analyses. We present those scenarios in the aforementioned sections to show the magnitude of the infrastructural power in contexts of daily life and public affairs in addition to the #digipower participants' stories.

Our analysis of #digipower participants' data provide additional unrivalled facts for illustrating and typifying the consequences. Based on the #digipower stories and our investigation, the report presents a final taxonomy of consequences (see Appendix to this report) with their corresponding definitions so anyone in the data economy acknowledges them according to their experiences, in addition to what they have already gained as knowledge, based on previous sections, about the mechanisms producing those consequences.

The negative effects of infrastructural power are worse when they scale from one individual to a population, but are also hardest to comprehend. To better show when there is a scale jump on the effects, we present now a list of five sets of consequences isolating specific stakeholders: a person, a group of individuals, a commercial entity, and a group of commercial entities. We think anyone could root their own analysis of the infrastructural power impact in the same decomposition.

The five sets of consequences according to specific stakeholders and their relations are the following:

[53] *Internal Documents Show Facebook Has Never Deserved Our Trust or Our Data*, Vice, december 2018, https://www.vice.com/en/article/7xyenz/internal-documents-show-facebook-has-never-deserved-our-trust-or-our-data

1. A person that is a consumer of a service provider and is affected by the system design

An investigation by AlgorithmWatch[54] on the representation of the body in Instagram pictures shows the way the algorithm classifies a woman according to her images and the labels that are defined by the service provider to describe her body. A woman that shows more of her body skin is presented more often on Instagram's feed results than other women presenting less of their body skin.

In the Instagram investigation, the image labels used for analysing Instagram pictures are in part extracted by Google's Vision API "Detect Labels"[55] and later refined by the data scientists conducting the analysis. See the two indicators defined with their corresponding labels[56]: raciness, i.e., whether a picture contains sexually suggestive content like "skimpy or sheer clothing, strategically covered nudity, lewd or provocative poses, or close-ups of sensitive body areas"; and nudity, i.e., "describing body parts, underwear or swimwear" in the picture. In that sense, Google techniques influence how users are perceived by the app's algorithms and other users, and more broadly it affects knowledge production when scholars do not have full control and a view on how data structures are conceived and for what purposes.

The consequences from our taxonomy are (see Appendix for full definitions):

- **Alienation**, i.e., individuals become something else, dispossessed from what they are. They become data and they do not possess data's value produced by others.

---

[54] Undress or fail: Instagram's algorithm strong-arms users into showing skin (June 2020). Retrieved from: https://algorithmwatch.org/en/story/instagram-algorithm-nudity/

[55] Google Vision API Detect Labels (April 2021). Retrieved from: https://cloud.google.com/vision/docs/labels

[56] "Raciness: For each picture, the Vision API returns a safe search rating indicating whether or not a picture contained "racy" content. The feature is measured on an ordinal scale with the possible values VERY_LIKELY, LIKELY, POSSIBLE, UNLIKELY and VERY_UNLIKELY. Racy, in this context, refers to sexually suggestive content like "skimpy or sheer clothing, strategically covered nudity, lewd or provocative poses, or close-ups of sensitive body areas." For the purposes of our analysis, a picture labelled racy is one that received a raciness rating of either VERY_LIKELY or LIKELY. A non-racy picture is one rated either UNLIKELY or VERY_UNLIKELY racy. Images marked POSSIBLE are labelled as undecided. Nudity: The Vision API also returns a collection of labels that describe the content of each picture (e.g. Landscape, Vacation, Window, Thigh). To complement the safe search rating, I manually compiled a list of labels indicating nudity. I started by analysing which labels are most often associated with raciness to inform this list. These were mostly labels describing body parts, underwear or swimwear. For all labels tagged in more than 50 images, I then manually noted whether they indicated nudity, using samples of images to test our judgement. The relevant labels were: Abdomen, Bare Chested, Bikini, Bodybuilding, Brassiere, Chest, Lingerie, Muscle, Skin, Stomach, Swimwear, Thigh, Trunk, Undergarment, Waist. Since I was interested in exploring possible gender differences in the way Instagram's algorithms treat nudity, I filtered these labels for ones that are associated with one user gender in at least 90 % of pictures. I adjusted the resulting list so that only clearly gendered terms remained. The final list of labels indicating gendered nudity was: "Women: Brassiere, Lingerie, Undergarment, Bikini. Men: Bare Chested, Bodybuilder." Undress or fail: Does Instagram favour posts that show more skin? (2020). Retrieved from: https://docs.google.com/document/d/1L7A5hmskm3Y3huSXHNtIloiVijHD3dkDqubff4Yvkg8/edit#

- **Aggravation**, i.e., reinforcement of individual and societal problems like addiction
- **Amplification**, i.e., reproduction of discriminations, stereotypes and inequalities in society, which are ultimately amplified by computing and algorithmic capacities
- **Socioeconomic Discrimination** (mainly social in the example above), i.e., discrimination of social groups or invisibility of some individuals
- **Noncritical**, i.e., loss of the critical capacity to judge the individual's own opinion, sources of information and their reliability
- **Competitiveness**, i.e., rivalry is established between individuals by arbitrary rules
- **Digital Labour Exploitation**, i.e., loss of independent value while serving others to capitalise on your own data
- **Informational Blindness**, i.e., loss of visibility and comprehension on how structures and individuals influence oneself
- **Normalised Social Conformity**, i.e., loss of singularity, the self is defined by others based on the calculation of a normalised proxy
- **Risky Exposure**, i.e., being exposed to malevolent actors, which practices are facilitated by the infrastructure

2. Two or more consumers of a service provider affected as a population

Scholars[57] use personal data extracted from platforms for knowledge production about the population behaviour. They use it for predicting and explaining psychological causalities between sexual orientation and facial features[58], and for predicting attractiveness with profile pictures[59].

An extreme case is a pseudo-scientific analysis conducted with OkCupid data for neo-nazi racist assumptions[60]. Conducting this type of research violates the data-protection rights of users when there is no agreement with the company, nor a data management plan that requires a direct and clear user consent.

The related consequences from our taxonomy are (see Appendix for full definitions):

---

[57] Wang, Yilun, et Michal Kosinski, 2018, "Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images.", *Journal of Personality and Social Psychology* ,114, no 2: 246-57. https://doi.org/10.1037/pspa0000098

[58] Leuner, John, 2019, "A Replication Study: Machine Learning Models Are Capable of Predicting Sexual Orientation From Facial Images". ArXiv:1902.10739 [Cs], http://arxiv.org/abs/1902.10739 .

[59] Jekel, Charles F., and Raphael T. Haftka, 2018, "Classifying Online Dating Profiles on Tinder using FaceNet Facial Embeddings", no 6. http://arxiv.org/abs/1803.04347

[60] Kirkegaard, Emil O. W., and Julius D. Bjerrekær, 2016, "The OKCupid Dataset: A Very Large Public Dataset of Dating Site Users". *Open Differential Psychology 1*, no 1: 10. https://doi.org/10.26775/ODP.2016.11.03 .

- **Rights Obstruction or Violation,** i.e., personal data rights are fragilized and not guaranteed. Loss of data ownership and individuals are powerlessness
- **Privacy Threats,** i.e., loss of contextualised privacy control. The subject cannot decide what to share or not, to a restricted or a large public
- **Misidentification**, i.e., wrong assumptions made about one's identity and preferences
- **Socioeconomic and Political Class Polarisation** (mainly social in the example above), i.e., reinforcement of hierarchical socioeconomic and political classes, it reinforces the gap between them for polarising
- Socioeconomic Discrimination
- Informational Blindness
- Risky Exposure
- Predictive Harms
- Misidentification

3. A service provider controlling the service provider's ecosystem: deciding how much others pay/earn

Google AdWords Auction System is designed so every service provider is charged with different amounts for their ad visibility on Google search according to the most popular search terms. This is regulated by the "cost-per-click bidding [which] means that you pay for each click on your ads."[61]. Hal Varian, Chief economist at Google explains it as "pay just enough to beat the competition"[62]. The services that control the view over the data economy (like Google) do not have to discriminate against other corporate entities because Google has designed and decided on the bidding system that regulates the commercial visibility and position of others. As a consequence of this design, it rewards the already richest companies because they can set the maximum bid.

Similarly, Google has designed the measure of webpages' relevance which influences how visible they are according to Google's design of hyperlink networks: how pages are interlinked based on their navigation by users that go from one hyperlink to another one. Indeed, the PageRank algorithm used by Google Search that ranks website pages in their search engine results began as a tool to measure the structure of pages, but has become a dominant influence in how new socioeconomic and political structures on the Internet are formed. Dominique Cardon (2013) explains that "the design of PageRank has been lastingly associated with a particular representation of Internet which has a structuring effect on the ecosystem now formed by the web and its dominant search engine." [...] "PageRank [defines] the metrics to use to describe the relational forms of the social"[63]; that is how website pages are valued according

[61] Cost-per-click (CPC): Definition, https://support.google.com/google-ads/answer/116495?hl=en Google, Accessed March 2020.
[62] Pay Per Click Management - Insights on the Google AdWords Auction System, https://www.youtube.com/watch?v=tW3BRMld1c8 Google, Accessed March 2020.
[63] Cardon, D. (translation by Carey Libbrecht, L.) (2013). "Inside the Mind of PageRank: A study of Google's algorithm." *Réseaux*, 177, 63-95. https://doi.org/10.3917/res.177.0063

to the connections between them. But Google structures more critically, the market dynamic and the position of competitors within an ecosystem.

The consequences from our taxonomy are (see Appendix for full definitions):

- **Asymmetric Perception,** i.e., partial view and limited access to information, personal data, the infrastructure design
- **Market Exclusion,** i.e., exclusion from the technology market or being forced to have a permanent disadvantaged position in innovation
- **Servitude**, i.e., submission to a top-down establishment of "the good life": dictating what to do in the right way
- **Socioeconomic Instability,** i.e., unstable social and economic conditions when working, friending, etc., that are not guaranteed in the short or long-term
- Competitiveness
- Digital Labour Exploitation

4. A service provider in relation to other service providers: using power against competitors

Once a commercial entity has reached scale of collection and has organised relationships so that other commercial entities depend upon it, it has the infrastructural power to block out competitors as data collection and its own services increment in scale. The power to use data collection to systematically favour oneself was demonstrated in a 2017 conversation between then-Facebook Vice President Deborah Liu and Jon Eide, who at the time was Facebook's Director and Head of Monetization Applied Research and Strategy. In the conversation, Eide explicitly discusses how the organisational power of the Facebook Pixel, a relationship of trust between Facebook and those corporate entities who chose to use it, was becoming a tool to undermine competition because Facebook now desired to compete directly with eBay:

"Imagine eBay seeing this and realising that all of their pixel data is now being used to power our marketplace that enables others to compete directly with them with their data. How can we show Ebay that their disproportionately helping Ebay? Or any other advertiser determining that we are using their 1st party intent data (e.g. CA inclusion, etc.) in marketplace where they can't compete yet as a B2C?"[64]

In this scenario, Facebook's organisational control over a data collection system on which others are dependent, the Facebook Pixel, became power over another commercial entity when Facebook's own data collection and service offerings incremented to the point where they could compete directly with that entity.

---

[64] Lawsuit. Court Filing: MAXIMILIAN KLEIN, et al. Plaintiffs, v. FACEBOOK, INC., Defendant., Court:United States District Court, Northern District of California, legal reference20-CV-08570-LHK (N.D. Cal. Jul. 20, 2021), Document 244-3, https://www.courtlistener.com/docket/18714274/244/3/klein-v-meta-platforms-inc/

Another example concerns the social capital of Instagram. It shows how having a lot of users leads to social capital so it is hard for a competitor to compete with, even if the competitor has a better product. This combines both technical and organisational capabilities for gathering a lot of data. Zuckerberg stated: "...that there are network effects around social products and a finite number of different social mechanics to invent. Once someone wins at a specific mechanic, it's difficult for others to supplant them without doing something different. It's possible someone beats Instagram by building something that is better to the point that they get network migration, but this is harder as long as Instagram keeps running as a product."[46]

The consequences from our taxonomy are (see Appendix for full definitions):

- **Social Disconnectedness,** i.e, damage to social cohesion by influencing socialisation practices (in this case commercial practices that are also social)
- Competitiveness
- Market Exclusion
- Socioeconomic and Political Instability

This scenario is different from the previous one because in this scenation Meta (Facebook) or Google uses other corporations' dependence on them to undermine those corporations, while in the previous scenario Google bends other corporations to their norms, like the bidding system for advertisement.

5. Two or more commercial entities: when data exchange leads to personal data laws' infringement

Based on an organisational capability, data in online-dating platforms is mainly structured and collected for advertising purposes, as it is the main source of income for apps. This capability is linked to the technical one: through SDKs it is possible to exchange different types of data across services, which finally fixes the price to pay for ads according to the target audience that is using the app. This was recently confirmed by an investigation of the Norwegian Consumer Council[65]. The investigation resulted in a lawsuit, for a 10M Euro fine[66] against the dating app Grindr (10% of the app's annual revenue) for sharing the sexual orientations and precise geolocations of their users to third-party advertisers, without clear user consent. This case illustrates the concrete consequences of the way data structures and data processing are designed, on user's privacy, on companies' finances, and on reputations. Consequently, the commercial purposes of data structures, defined in advance, shape what is produced and observed as user behaviour.

The consequences from our taxonomy are (see Appendix for full definitions):

---

[65] Norwegian Consumer Council, technical report "Out Of Control"
https://fil.forbrukerradet.no/wp-content/uploads/2020/01/mnemonic-security-test-report-v1.0.pdf
[66] Historic victory for privacy as dating app receives gigantic fine (January 2021). Retrieved from:
https://www.forbrukerradet.no/news-in-english/historic-victory-for-privacy-as-dating-app-receives-gigantic-fine/

- **Opacity**, i.e., lack of comprehension or full ignorance about the systems composed that are processing personal data
- **Surveillance**, i.e., an ubiquitous and opaque observation over others to control that the observed individuals obey the rules established by the observer
- **Generalised Suspicion**, i.e, difficulty for trusting services acting on behalf of a person's autonomy
- Market Exclusion
- Asymmetric Perception
- Competitiveness
- Rights Obstruction or Violation
- Privacy Threats
- Informational Blindness

Hestia.ai
sustainable data solutions

4. Reflective Conclusions

We have already included some of our reflections on study logistics in the methodology and case studies report. Here we focus on five higher level reflections. Although some conclusions in this chapter could overlap with the ones in the methodology and case studies report, here we put forward our reflections with respect to the data economy, and not with respect to individual cases.

First, participants and investigators involved in this study found it very helpful to lay down examples (personal experiences and contemporary issues) in order to engage in sensemaking together. At the same time, there were difficulties. It is hard for one participant to see the big picture from just their own data, but it is also very hard to synthesise a coherent big picture from the plethora of facts provided by the multiplicity of channels pursued. It was a real challenge to achieve the promised goals of telling individual stories but also produce a high level "big picture" report.

The methods and new skills we developed together with participants and investigators enabled us to see the magnitude of the data economy. This was possible by combining multiple perspectives on the phenomenon according to every person's expertise: mathematics, sociology, engineering, data science, physics, journalism.

Second, participants seemed to benefit from the metaphoric parallels we exploited for pedagogical purposes between cityscape and infoscape. However our usage of those was not structured during the participant interviews, just as for the journalist we described in Section 3.1 *Shaping Context* of this report answering a reader question on Twitter with her baker showcase metaphor. By highlighting to her how exactly her metaphor broke down, we were able to get her to understand better the consequences of her own usage of Twitter. Within the #digipower context, by running multiple interviews, we progressively understood as coaches that the way the metaphors exactly break down has itself a structure. In other words, the engine of metaphors should itself be made explicit and structured in order to facilitate learning. This is what we did in the narration of Case Study 2 in the methodology and case studies report (infoscape: *When you view the web, the web views you*), systematically breaking down what carries over from Case Study 1 (cityscape: *Who collects my geolocation and why?*) and, more importantly, *what does not*. We discussed the pedagogical benefits more extensively in the methodology and case studies report.

While we explicitly told participants of the metaphors, we were not ourselves able to formalise the engine at the time of the interviews, only at the time of writing the methodology and case studies report. We believe the #digipower methodology and associated pedagogy would benefit from making the structure of that engine of metaphors explicit.

Third, we want to highlight the GDPR implementation (in)effectiveness. Indeed, as explained previously, we followed three different lenses for data access. Subject Access Requests (SAR)[67], Data Downloads portals[68], and technical audits[69] were all useful in achieving different outcomes, and beneficial in the aggregate. SAR were best at supporting inquisitive efforts into understanding, where the pedagogical approach was much more participative. They were also often disappointing, but provided a clear action point: fix the *enforcement* of GDPR. Data downloads afforded predictability, but they also came with the drawback of a biassed and company-sanitised view into the company's operations. Our data experience tools over the exported raw data provided a somewhat mitigating approach for the later point, but a Data Download approach needs to be supported with a critical comparison of the transparency provided by different platforms. The TrackerControl technical audit was helpful in providing a unified base for all the participants, but suffered some drawbacks in the actual visibility it had on the flows. All together, they provided multiple lenses on the same flows, affording the participant the possibility of confronting their views obtained through different channels. This was crucial in "unlocking" a more critical spirit towards the digital.

Fourth, the investigation put forward the contrast of participants' traditional power with the digital power of dominating service providers. Participants have a political role that somehow influences the data economy, although they do not necessarily have deep technical capabilities. However, their traditional power includes, in particular, organisational capabilities that are of great benefit for mitigating the effects of the infrastructural power over society. This was anticipated by us from the get go, and we developed our tools accordingly: beyond a one-off we hoped to replicate a version of this study, possibly within the circles of some of our participants. This remains a possibility at the time of writing the report: we hope for instance to replicate these efforts through schools or community centres, which political leaders might have an easier time motivating than us.

Fifth, we saw the centrality of certains services like Google or Meta (Facebook). This network centrality should be qualified and measured in the data economy to better understand the power of dominating service providers. Indeed, the infrastructure becomes opaque for services and subjects because of its complex composition and centrality. Based on the compositionality mechanism, services have a facilitated anchoring in the data economy to centralise benefits via SDKs and APIs mainly distributed by Google, Amazon, Apple, and Meta (Facebook), for having access to data, to orchestrated social capital, and to a technical architecture that provides partial benefits in development costs and efficiency. This 2012 exchange between Mark

[67] active requests for personal data, as mandated by data protection laws such as the General Data Protection Regulation
[68] self-service portals for personal data, for which platforms tend to keep plausible deniability of (in)completeness in restituting all of the personal data
[69] as conducted during the study through an app called TrackerControl, looking at third-party trackers in Android apps

Zuckerberg and the Facebook executive team demonstrates how maintaining centrality dominates the strategies of these companies:

"We're trying to enable people to share everything they want, and to do it on Facebook. Sometimes the best way to enable people to share something is to have a developer build a special purpose app or network for that type of content and to make that app social by having Facebook plug into it. However, that may be good for the world but it's not good for us unless people also share back to Facebook and that content increases the value of our network. So ultimately, I think the purpose of platform—even the read side—is to increase sharing back into Facebook."[70]

Connected services like Uber or Deliveroo only gain a position in the data economy as "complementors" or facilitators of Google's power controlling all data processes, from data collection to data distribution. In other words, dependent services become the free workforce multiplying the sources of data collection while Google decides what to collect, about whom, when and where.

Favoured by an infrastructural power, services establish two main asymmetry relations identified in our investigation. A first asymmetry relation is established between a service and subjects: while a service gains power by capturing data, subjects are diminished from their data sovereignty and agency to act. The second asymmetry relation is established between services and a main complex service like Google: while the latter centralises power and the resulting benefits in the long-term, other services become dependent and dominated with the possibility of capturing only partial and short-term benefits in the data economy.

Finally, the data economy is vast and complex. While acknowledging many imperfections to our typologies of power dimensions (section 3.5), levers (section 3.6), mechanisms (section 3.7) and feedback loops (section 3.8), we hope that, beyond these reports, we have created through #digipower an entire methodology, patching individual understanding into collective sense-making.

[70] Lawsuit. Court Filing: In re: Facebook, Inc. Consumer Privacy User Profile Litig., Court: United States District Court, Northern District of California, legal reference 18-md-02843-VCJSC) (N.D. Cal. Dec. 30, 2021), Document 491, https://cand.uscourts.gov/wp-content/uploads/cases-of-interest/in-re-Facebook-consumer-privacy-VC/Second-Amended-Consolidated-Complaint-Dkt-491.pdf

Hestia.ai
sustainable data solutions

5. Alternative Futures

Engineer Stephen Diehl has described[71] Web3 as a "vapid marketing campaign that attempts to reframe the public's negative associations of crypto assets into a false narrative about disruption of legacy tech company hegemony."

Because we decompose in this report the mechanics behind this legacy tech company hegemony, there is simultaneously a risk of appropriation of our findings for furtherance of this "vapid marketing campaign" and an opportunity for us to tear down that very campaign. We would like to take that opportunity. In addition, it will be helpful to do so ahead of our upcoming recommendations in the next chapter, to provide stronger contrast with suggested alternatives.

We first discuss blockchain and then other crypto assets and in particular Web3. Finally, we formulate a contrasting manifesto for #digipower efforts.

5.1. BLOCKCHAIN AND THE ATOMIC BOMB

At the time of writing this report, the atomic bomb has been used twice in war. Though imperfect, the threat of mutual assured destruction[72] has prevented repeat incidents. However, even though nuclear weapons have not been used in recent decades, the *threat* of using them has been immensely impactful on global politics – if only through the United Nations and its Security Council embodying the need to find alternative diplomatic solutions.

Blockchain technology should be seen in a similar way: repurposing slightly the language of our findings from Chapter 3, blockchain is extreme *Technical Capability* geared towards nuking existing *Organisational Capability* through decentralisation. Except that, just like with the atomic bomb, after the actual functioning of the technology has been demonstrated once (which has happened through bitcoin[73]), the mere *threat* of mutual assured destruction is sufficient for incumbents to be forced to build alternative arrangements. Just as issues do not get resolved at the Security Council by dropping nukes within the Council chambers, there is however no obligation for incumbents to resolve their problem through blockchain.

In fact, it is fortunate that no such obligation exists: because the holders of technical capability do not want to think about the social process of organising value production, they coat what they do in under conceptualised consensus mechanisms that amount to "trust no one". This has a real environmental cost[74] while the technology is not ready with respect to scale, speed, or

[71] Stephen Diehl, *Web3 is Bullshit*, https://www.stephendiehl.com/blog/web3-bullshit.html
[72] https://en.wikipedia.org/wiki/Mutual_assured_destruction
[73] Note that the demonstration requirement is for technological functioning, rather than utility.
[74] See more about this topic at *What are the environmental impacts of cryptocurrencies?* https://www.businessinsider.com/personal-finance/cryptocurrency-environmental-impact?op=1 BusinessInsider, Mar 17, 2022.

privacy. All those technical drawbacks mean that incumbents will *always* have time to spot the threat of decentralisation, and preempt it. In fact such dynamics are fully anticipated, since most blockchains offer *permissionless* and *permissioned* modes of operation. The former can be difficult to run, expensive, privacy invasive, etc, while the permissioned mode covers these drawbacks but puts power still in the hands of the incumbents (and newcomers who play the political game well). Another consideration is that permissioned blockchains then become useless – there are always superior technical alternatives that amount to a shared database system[75].

In other words, blockchain is a technology that is catalysing change through the threat of its deployment, but which is unlikely to be effectively deployed beyond its very first technical validation (bitcoin)[76].

5.2. THIS SECTION IS A WASTE OF EVERYONE'S TIME: WHEN CAPTIVATING EVERYONE'S ATTENTION IS THE ONLY STRATEGY

According to one of the biggest socio-technical collaborative enterprises in the world, Wikipedia, web3 is defined as "an idea for a new iteration of the World Wide Web based on blockchain technology, which incorporates concepts such as decentralisation and token-based economics."[77]

This idea comes with opposing expectations: While "some experts argue that web3 will provide increased data security, scalability, and privacy for users and combat the influence of large technology companies. Others have raised concerns about a decentralised web, citing the potential for low moderation and the proliferation of harmful content, the centralization of wealth to a small group of investors and individuals, or a loss of privacy due to more expansive data collection." (Ibid.)

The key message of web3 is decentralised value production, that is why it is relevant to consider Wikipedia's design to discuss the promises of web3. However, one fundamental difference lies in the fact that Wikipedia's consensus mechanism for knowledge production is merely mediated through technology but is intrinsically social (it has a high organisational capability). Indeed, it

[75] See for instance Wüst, K., & Gervais, A. *Do you need a blockchain?*. In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)* (pp. 45-54). IEEE, https://doi.org/10.1109/CVCBT.2018.00011
[76] The only possible exception would be where very pressing global needs manifest themselves so fast and so urgently that they bypass any traditional methods of diplomacy at supranational level – the ultimate incumbents – and at the same time reach such a scale that they become indispensable.
[77] Wikipedia, https://en.wikipedia.org/wiki/Web3 Wikipedia, accessed on March 2022.

encourages dialogue between peers on the platform to figure out as a community the best articulation for moving forward. Unlike, ironically, web3, which at best offers vaporware[78] promises of consensus amongst dominating corporate players. Indeed, the main innovation between the concept of blockchain (pure technical capability) and web3 lies precisely in the optimization of "vaporwaring" itself. Indeed, web3 follows a long string of successive innovations in the financing of those systems: Initial Coin Offerings, Decentralised Autonomous Organisations, Non-Fungible Tokens, and now web3 are ways to bind the technical innovation to some incremental mechanisms to finance it, rooted in more and more speculation.

ICOs were about small pairings of developers with full control of the code together with financiers who had full control of the developers launching protocols to structure entire economic communities (thereby *increasing* centralisation). Money was raised at the time of ICOs based on "whitepapers" that outlined a dual strategy of a vague technical idea and a very precise marketing plan, often already leveraging a fear of missing out on fantastic gains - both from the coin buyers and from the investors who were offered the opportunity to jump ahead of the queue.

Decentralised Autonomous Organisations were basically toying with the idea of decentralising some aspects of ICOs, building in their highest profile incarnation decentralised venture funds[79].

Non-Fungible Tokens (NFTs) are an even more decentralised version of this speculative dynamic, where the entire speculation is built around an ambiguous object. Currently that object is said to be of artistic value. The art market is a smart choice because of the very subjective value of art, and the deep links of the art market with tax evasion schemes.

It is hard to say what web3 is about, except maybe a network of objects of ambiguous value, a network which is said itself to be of value thanks to the interoperability it creates between those ambiguous objects. However there are many lingering questions about the technical details, the governance, and the control of such a network.

Many of those questions can actually be at least partly answered by looking at *Move*, the programming language behind the Libra/Diem coin, Facebook/Libra's failed attempt at cryptocurrencies[80]. It does make sense to look at the programming language, as this is what is

[78] a product that has been promised for a long time, keeps on being promised, but will realistically never be delivered
[79] See for instance the spectacular technical failure associated to the initiative imaginatively named *The DAO*, which then turned into an organisational crisis
https://en.wikipedia.org/wiki/The_DAO_(organization)
[80] See *Move: A Language with Programmable Resources*, Diem Association
https://developers.diem.com/docs/technical-papers/move-paper/

immutable within the project, and has to be decided from the very beginning - and it is somewhat coercive on all the other participants.

We find the following notable in Move:

- a deep decomposition of the language into two complementary strands:
    - one strand allows description of just about any event in the digiscape, such as e.g. the need for a particular electric scooter located in a particular location to be charged, and therefore its transformation into an object that can be manipulated (a process called *reification*[81]);
    - the other strand allows complete *financialisation*[82] of such events, for instance one could buy futures in a scooter charging contract based on the weather.
- lack of precision on the process through which this language actually gets compiled or interpreted into running and interoperable code.

The latter is very significant: the need for interoperable code creates deep coordination problems between different deployments that can quickly tip power dynamics. Yet that second point is very important as it decides not who or how can the two complementary strands be coupled, but how the two strands can be *effectively* and by whom.

In other words, as an example, Move allows building *financial* marketplaces for electric scooter charging, but does not determine who can do high-frequency trading there, what are the rules of high-frequency trading, and who gets to decide the rules and the participants. One again, we see Meta/Facebook playing the *Composing Complex Infrastructure for a Dominating Position* feedback loop: leveraging some key technical capability withheld centrally in order to orchestrate dominance.

5.3. A #DIGIPOWER MANIFESTO

The values behind the #digipower investigation are very different and worth contrasting with the previously described futures (blockchain, web3). These values encourage developing for individuals (seen as situated within their communities) new forms of empowerment in the digiscape, through a systematically constructed theory of change.

First, the value of transparency through mathematical modelling for engineering design. Complex systems are only used if they are necessary for the collectivity's goals and their

[81] According to wikipedia, "Reification (in knowledge representation) is the process of turning a predicate or statement into an addressable object".
[82] a process whereby financial markets, financial institutions, and financial elites gain greater influence over economic policy and economic outcomes

composition remains possible to deconstruct[83]. This modelling guarantees that our systems remain explainable and accessible to give the possibility of reusing them and conducting audits. An example of this systematic and constant modelling lies in the second point of *Reflective Conclusions* in Chapter 4: our "engine for metaphors" itself has structure, which we immediately try to make explicit as part of our pedagogy.

Second, the value of proportionality for personal data protection. We create data pipelines for securing and sharing personal data while acknowledging that there *might* be a tension between protecting the community and individuals' interests, privacy and rights, all the while guaranteeing the technical efficiency of the system. This is our technical and ethical responsibility.

Finally, the third value is social care. We care about the individual and social concerns of collectivities and we are interested in supporting actors without a vision of the common good for building one with them. We support society through literacy programs where we learn together from/with individuals that have a common interest and concern around data. We give them a space and a voice, and we build participatory methodologies for upskilling and building together new data governance models along with trusted infrastructures.

Our values do not focus on privacy-by-design, nor on a paternalistic view over service providers and data subjects where only few stakeholders have the expertise and benefit disproportionately from the data. Instead, we are a collective pioneering new forms of social relationships, aiming towards a more locally oriented and federated proposition of mutual learnings.

[83] See Breiner, Sriram and Subrahmanian *Compositional Models for Complex Systems* in *Artificial Intelligence for the Internet of Everything* https://doi.org/10.1016/B978-0-12-817636-8.00013-2

Hestia.ai
sustainable data solutions

6. Recommendations

There are five recommendations resulting from the #digipower investigation – all clearly aimed to the #digipower Manifesto described in the previous section.

6.1. CHANGE THE NARRATIVE: INNOVATING DIFFERENTLY

The current narrative around the data economy centres on a myth of innovation through technology. However, the technology that has been leveraged on personal data is not that innovative. What is at stake is the dominating power that service providers hold from their centralised technological business.

For instance, Facebook's real origin story should not be one of a lone innovator creating complex services. Instead, as outlined in the introduction, it should be the story of an impetuous business executive who has been able to learn a crucial lesson from his first brush with organic product traction: always go faster than proper technical accountability. Or as sociologist Dominique Boullier expresses it, the technology industry dynamic is based on "running code and apologising". This is not a story of recruiting and training excellent engineers. It is a matter of training other business executives to constantly operate at best at the margins of legality – all in the pursuit of organisational capability as a precondition for investing into technical development, often simply copying features of competitors as Jessica Pidoux has demonstrated[84] with the dynamic of imitation and counter-imitation between worldwide dating platforms.

Not all companies operate like Meta (Facebook)[85], and this is precisely the point: the right of access acts as a leveller to the data economy providing a fairer playing field for entrepreneurs who want to innovate differently – and it is not just about ensuring freedom to innovate. It is about suppressing the impact neglectful innovators have over the entire market if they are left to amass organisational capability. However, the investors supporting this type of infrastructural power still need to buy into this mindshift (or at least some of them need to).

A potential accelerator in the immediate term is to document ongoing lawsuits extensively, and facilitate deconstruction of what amounts to corporate public relation efforts.

---

[84] Pidoux, Jessica, Pascale Kuntz, et Daniel Gatica-Perez, 2021, "Declarative Variables in Online Dating: A Mixed-Method Analysis of a Mimetic-Distinctive Mechanism" 5, n° CSCW1: 100-132. https://doi.org/10.1145/3449174.

[85] or Google, really: the Jedi Blue deal for instance is a shocking scandal of collusion to the detriment of advertisers and publishers, but unfortunately not yet sufficiently documented anywhere.

6.2. PRODUCTIVISE #DIGIPOWER-LIKE EFFORTS

As we hope we have demonstrated in our conclusions, #digipower should be understood as a methodology, not a one-off investigation.

We think this methodology should be deployed in a wide variety of settings and age ranges, in order to develop critical skills towards the digiscape within entire communities.

As soon as possible, kids should be encouraged to develop towards the digiscape similar skills as what they are encouraged to develop towards the physical or biological world[86]. Being as old as the first smartphone does not mean being armed against the insidious power mechanisms of the easy-to-use tools we have in our hands. Far from it.

But it really should not just be kids. How many (expensive) digitisation projects are done without knowing the ins and outs[87] of the data economy?

- Why would a business executive get to invest large sums in tech development, without fully understanding the dependencies they introduce in its business model towards tech giants?
- How can a public authority make difficult decisions on cloud hosting infrastructure for health, education, etc. without having the means to measure the risk it poses to its population (increasingly aware of the commercial use of this data)?
- How can journalists complain about the growing mistrust of their readership when their publishers stuff their websites and apps with profiling trackers?

The core strength of the #digipower methodology is in the pedagogical approach of rooting an individual's understanding of the data economy within their own data. If we were to run such an effort again, we would keep the approach but to make it more accessible, to give it a collective dimension beyond the individual experience, and for all the reasons described in the methodology and case studies report, **we would offer workshops instead of coaching sessions**.

We anticipate that there might be interest in such workshops from schools, ecosystem facilitators (innovation networks, incubators,...), executives wishing to approach the digitisation of their business systematically, human resources departments, journalist organisations, city planners, etc.

---

[86] An interesting question: would kids nowadays have first experienced exponential growth through real-life virality of a virus, virality of content on social media, or a biology textbook talking about reproducing rabbits?
[87] One example: https://twitter.com/TheEyeballsFr/status/1508838683486609418?s=20

While Hestia.ai has obvious conflicts of interest in recommending all this, all of the code generated for this investigation is available publicly at http://github.com/hestiaai and Hestia.ai welcomes collaboration.

It should also be acknowledged that this investigation benefitted from a high level of involvement of SITRA, Hestia.ai and commitment from the participants. All were aware they were trying something that had never been done before, and hoping the results would be directly useful at a *national and even supranational level*.

6.3. INCREASE CIVIL SOCIETY'S INFRASTRUCTURAL POWER

Problems in the digital domain are complex. Pick the example of an association specialised in defending women's rights. One instant they might be concerned with providing shelter to a woman in a difficult situation, and the next they might be trying to assist someone in preventing the viral spread of sexual content shared without consent by an ex-partner. They might be trying to be doing both at the same time for the same person. Both situations call for vastly different responses. In one case it is a matter of coordinating locally a complex network of support, and in the other to get heard by global platforms. The latter is unmanageable by a very large number of civil society actors, despite their perimeter of relevance expanding more and more in the digital space. A first blocker is to even understand the relevance of data protection law to a wide variety of matters. Meanwhile, a number of NGOs specialised in digital policy making consistently say they have no capacity to address such individual cases[88], but at the same time are on the lookout for "exemplary cases" to illustrate their advocacy. In a society of intense personalisation, it is critical to fix this problem at the intersection of individuality and countering mind boggling infrastructural power.

The problem extends beyond NGOs though. Similarly, academics lack access to data for researching emerging problems, and platforms are playing one research group against another for favourable access to datasets – often under NDAs and certainly with biassing effects on scientific production. Yet the problems involve extremely intrusive data, since they often concern causality determinations between individual behaviours online, population-level inference and collective dynamics.

It seems natural that infrastructural power of dominant actors would call for matching infrastructural counterpower. Of course at the same time no one wants a landscape dominated by a handful of NGOs and universities dominating all the others. Plurality is important in media, academia or advocacy. At the same time, the mechanisms through which dominant players are

[88] Often, primarily, due to the complexity of managing the "customer service" side of the interface with the general public.

acquiring infrastructural power are still helpful towards understanding what should exist on the civil society side.

We think these organisational problems between digital civil society, thematic NGOs and researchers could be addressed through the mechanism of compositionality. As explained in section 3, compositionality is key to infrastructural power. It thus seems natural that it would beget a specific response from civil society. Instead, we think the two dimensions of technical and organisational capabilities building up around compositionality should be taken at face value, and that civil society projects should be funded accordingly: through coalitions involving multiple actors, each with different and *orthogonal* perspectives on building counter power.

Our starting example of the association defending women's rights is an illustration of how organisational power in civil society could be increased, by favouring collaboration between thematic civil society actors (NGOs, researchers, journalists) and digital NGOs around assisting individuals (or documenting their predicaments).

As for technical power, such civil society projects will always need assistance to manage the complexity of workflows between multiple civil society actors (while preserving privacy as much as possible), but also of the topic at hand (reverse engineering algorithms, data analysis, etc). It seems crucial to highlight that in this case *data processing flows are meant to increase the value of raw data towards creating social change*, i.e. that while the general objective is not commercial, the intermediary mean is essentially a service that has commercial value: increasing the value of raw data. In a certain sense, the situation is similar to that of civil society actors with respect to newsletter tools, for instance: no one thinks civil society should use a *nonprofit* newsletter service, but they should at the same time be conscious they cannot credibly use just any newsletter service (with a differentiated pressure within civil society: a nonprofit focused on digital matters will be more vigilant, and will in turn be mimicked in its tools choices by other nonprofits a bit more distant to digital matters). Therefore that side of the equation should be treated through a strategic procurement angle: rather than adopting commercial services a posteriori, civil society would seed funds (and ideas) through its own needs, and cross subsidy the emergence of more virtuous players that will bring new commercial offerings, which will in turn help realign commercial players not specialised on digital issues. This scenario is deeply conscious of the very limited funds of civil society, but *also* very aware of the accelerating value the credibility their use of some tools brings. It is also deeply conscious of slightly different needs and funding mechanisms of journalists, academics, and NGOs, all the while leading to a hopefully coherent effort centred on helping first and foremost individuals make sense of the data flows around them.

6.4. SUPPORT DATA COLLECTIVES

A data collective is a group of individuals who get together to manage their data collaboratively, and extract value from it that they would not be able to achieve otherwise.

It has a participative dimension that is not necessarily present within the concept of "data intermediaries" recently introduced in the recent Data Governance Act[89].

Increasing infrastructural power of civil society is not the panacea either. Journalists compete for scoops and can be instrumentalized, for instance by negotiating favourable press coverage. Academics compete for papers and can also be instrumentalized, through exclusive access to datasets or acquiring spin-offs. In a world of intense personalisation this might make it harder to get some voices heard.

In addition, in a world of data, issues of representativity and biases are key, and are more likely than not to span beyond advocacy groups. For instance queers who would allege a disparate impact of automated decision making on their community would need to use a control group, which cannot be limited to their allies.

The same data can also be relevant to multiple causes at once. All this points to an additional layer of organisation distinct from the civil society organisations themselves, with access to data (or the results of statistical computations) negotiated with intermediaries.

One could ask if the participatory dimension of collectives is really necessary, and indeed some of the ideas described here are operationalised as data *donations* to academic institutions[90] or advocacy groups[91].

In contrast, we think that research and advocacy are means toward systemic change, but that this change can be accelerated through the participative dimension of data collectives. Indeed, knowledge should be produced bottom up within the data collective, through a form of extreme citizen science. This should contribute to upskilling broadly the community, with new business models emerging more organically, as detailed knowledge of the data value creation process is produced. This will contribute to accelerating the investment and development of alternatives that have a chance of matching the speed of growth of current systems, particularly as organisational power starts to kick in between collectives.

The urgent needs for the development of data collectives are: education, legal (governance but also access to data), business model development, scale, and credibility.

---

[89]World Economic Forum, *Advancing Digital Agency: The Power of Data Intermediaries*, February 2022
[90] Data Donation Day, Utrecht University, January 2022
https://hds.sites.uu.nl/2022/01/15/data-donation-day/
[91] DataSkop, AlgorithmWatch https://algorithmwatch.org/en/dataskop/

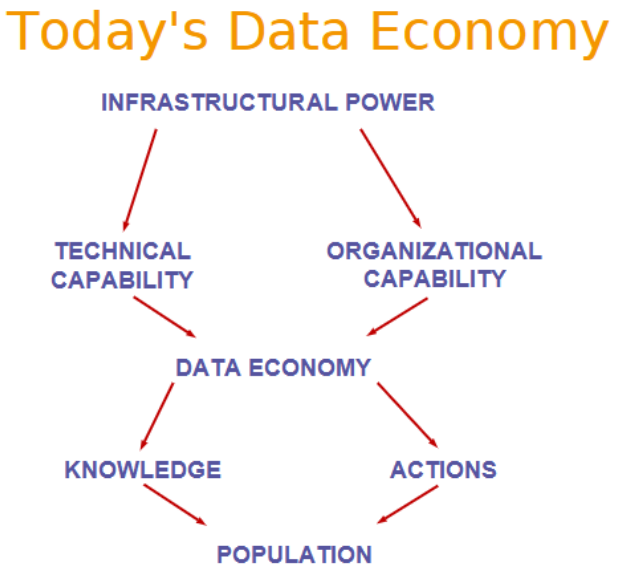


Figure 6. Today's Data Economy

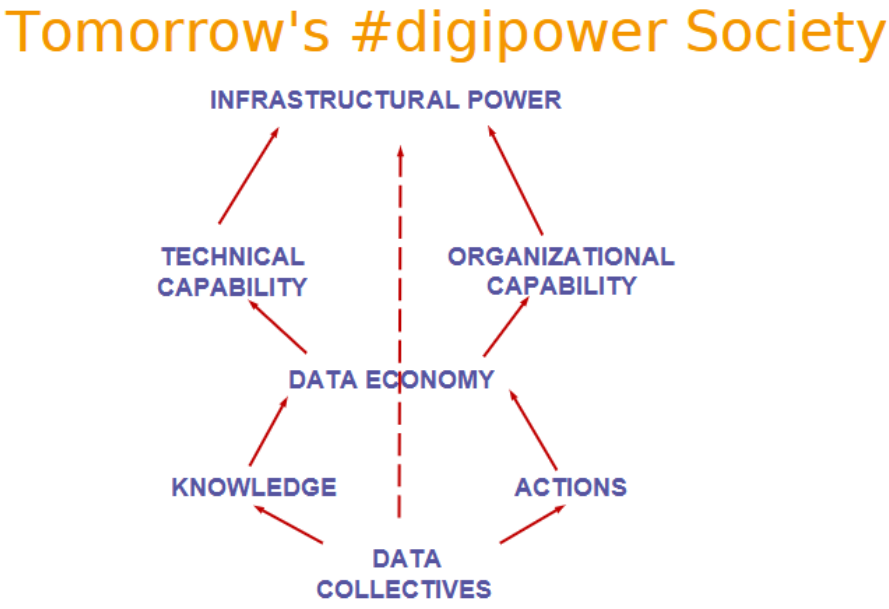


Figure 7. Tomorrow's #digipower Society

6.5. ENFORCE GDPR PROPERLY

To no one's surprise, the authors of this report think GDPR enforcement should be improved. It is essential for all the previous points that individuals are provided with effective means of access to their own data.

We believe a lot can be achieved already with the existing channels of access to data (such as Data Download portals), but are fearful of the asymmetry that uneven data access capabilities create. Companies do not tend to have equivalent responses depending on their country of operation, and their scale. In some countries, GDPR requests are just ignored or very shoddily complied with. In other countries, like Ireland with global companies, it is a much more subtle process of appearing to be compliant while not providing the most crucial information. The end result could be that some parts of the world learn faster than others what are the next generation of products worth investing in, and that European incumbents miss the train because they successfully lobby that GDPR is unfavourable to European companies.

Not all the data has been captured yet (see an illustration of Google's data partial access in figure 8), and enforcement should be done according to the end vision of a data economy based on self determination, rather than management of current business sensibilities.

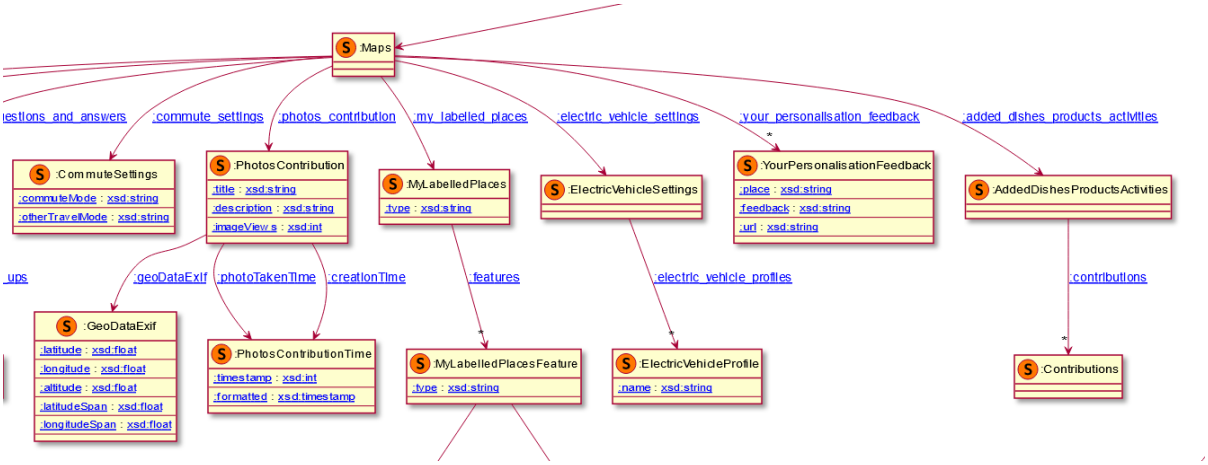


Figure 8: Not all data has been captured yet: Google's data download portal reveals data structures that stand ready to crowdsource data collection of electric vehicle recharging data. Meanwhile, mobility entrepreneurs, including some Uber drivers, are facing plenty of obstacles to gain access to their geolocation data as collected by their smart car manufacturers.

A common leitmotif is that data protection authorities do not have enough funding to act properly when facing the giants, introducing structural inequalities in the results of their enforcement. Data protection authorities should realise that a more endogenous enforcement

of rules in the data economy through externalising of sensemaking is highly desirable to them. They should therefore focus on strategic actions facilitating externalisation of this sensemaking, such as:

- making sure access is properly enforced in joint controllership situations, since this most directly touches on organisational capability of the dominant platforms[92],
- differentiated enforcement of the right of access for journalists, thematic NGOs or collaborative projects. While this approach might fly against a vision of equal access for all to their data, it might be substantiated by other goals (public interest in the transparency) and would echo somewhat differentiated enforcement of Freedom of Information laws,
- preserving the ability for scholars to research the data economy outside platforms' structures and rules.

In addition, if the idea of data collectives is to be seriously pursued, it introduces a vast array of new complications at the cutting edge of data protection law jurisprudence. Consultations by data protection authorities on the topic with the aim of providing actionable guidance would be highly desirable[93].

[92] For very structured guidance to do so, see Dehaye, Hahn and Jargalsaikhan, *Platforms and Personal Data Processing: The Potential for Achieving Systemic Transparency*, April 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3552930

[93] For the hanging questions at the intersection of the right of access and data collectives, see for instance PersonalData.IO's recent submission to the right of access consultation by the European Data Protection Board https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-012022-data-subject-rights-right_en

Hestia.ai
sustainable data solutions

Appendix: Consequences Taxonomy

Based on the results of the #digipower investigation we built a consequences taxonomy presented below. The taxonomy covers the consequences from the infrastructural power that is deployed by service providers which are affecting both individual and social groups as the cases we presented in chapter 3. This work-in-progress requires further development but should already help in guiding a reflective analysis on how society is affected in the data economy. Authors on which the definitions are based or inspired can be found at the end of these.

Consequence	Definition
Aggravation	Reinforcement of individual and societal problems like addiction[94]
Alienation	Individuals become something else, dispossessed from what they are. They become data and they do not possess data's value produced by others[95]. Loss of freedom to do, desire, choose, move, buy, give an opinion based on the individual's own will
Amplification	Reproduction of discriminations, stereotypes and inequalities in society, which are ultimately amplified by computing and algorithmic capacities[96]
Asymmetric Perception	Partial view and limited access to information, personal data, the infrastructure design
Captivity	Damage to individual and collective emancipation by blocking autonomous organisation and establishing dependency
Competitiveness	Rivalry is established between individuals by arbitrary rules. It favours the social positioning of some by disadvantaging the social positioning of others, e.g., popularity or attractiveness ranking in social networks
Digital Labour Exploitation	Loss of independent value while serving others to capitalise on your own data[97]
Dispossession	Being dispossessed of an individual's authorship concerning any related information about the self, and value belonging[98]
Generalised Suspicion	Difficulty for trusting services acting on behalf of a person's autonomy
Illiterate	Lack of skills to modify the infrastructure's composition, modify it or produce personal and collective benefits. A lambda user is in a worst position in comparison to developers that have technical skills to act on the infrastructure to a certain extent
Informational Blindness	Loss of visibility and comprehension on how structures and individuals influence oneself. It includes the loss of the ability to know about the infrastructure's existence and how it affects

[94] Marmet, S., Studer, J., Wicki, M., Khazaal, Y., & Gmel, G.. "Online Gambling's Associations With Gambling Disorder and Related Problems in a Representative Sample of Young Swiss Men". *Frontiers in psychiatry*, 12, 703118, 2021, https://doi.org/10.3389/fpsyt.2021.703118
[95] Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Paperback edition. London: Profile Books, 2019.
[96] Huszár, Ferenc, et al. "Algorithmic Amplification of Politics on Twitter". Proceedings of the National Academy of Sciences, vol. 119, nº 1, 2022, https://doi.org/10.1073/pnas.2025334119.
[97] Tubaro, P., Coville, M., Le Ludec, C. and Casilli, A. A.. "Hidden inequalities: the gendered labour of women on micro-tasking platforms." *Internet Policy Review*, 11(1), 2022. https://doi.org/10.14763/2022.1.1623
[98] See Zuboff

Market Exclusion	Exclusion from the technology market or being forced to have a permanent disadvantaged position in innovation
Misidentification	Wrong assumptions made about one's identity and preferences without the possibility of verifying or correcting them. Individuals can be identified by normalised categories for ads targeting and do not have the possibility of contesting them
Misinformation	Inability to find reliable information and to know when it's a reliable piece of information
Noncritical	Loss of the critical capacity to judge the individual's own opinion, sources of information and their reliability
Normalised Social Conformity	Loss of singularity, the self is defined by others based on the calculation of a normalised proxy, e.g., predicting rapists among Uber drivers
Opacity	Lack of comprehension or full ignorance about the systems composed that are processing personal data
Predictive Harms	Harms produced by statistical predictions made about the self which affect privacy[99]
Privacy Threats	Loss of contextualised privacy control. The subject cannot decide what to share or not, to a restricted or a large public
Reactivity for User Retention	A loss of making reflexive and informed decisions in contrast to acting in reactivity (i.e., within a state of alertness) to keep the accelerated pace that is dictated by service providers and their technological design[100]
Rights Obstruction or Violation	Personal data rights are fragilized and not guaranteed. Loss of data ownership and individuals are powerlessness
Risky Exposure	Being exposed to malevolent actors, which practices are facilitated by the infrastructure, e.g., scams, disinformation, bullying
Servitude	Submission to a top-down establishment of "the good life": dictating what to do in the right way
Social Disconnectedness	Damage to social cohesion by influencing socialisation practices
Socioeconomic and Political Class Polarisation	Reinforcement of hierarchical socioeconomic and political classes, it reinforces the gap between them for polarising, e.g., a service provider vs a user status, high class vs low class assignation according to volunteered data, business models distinguishing the benefits for users with paid vs free services, traditional models favoured like patriarchy
Socioeconomic and Political Discrimination	Discrimination of social groups or invisibility of some individuals, e.g., not inclusive categories, exclusive digital practices, dependence on having access to computational devices
Socioeconomic Instability	Unstable social and economic conditions when working, friending, etc., that are not guaranteed in the short or long-term
Surveillance	An ubiquitous and opaque observation over others to control that the observed individuals obey the rules established by the observer. The latter is at a distance so the observed individuals are unable to enter in contact[101]

[99] Citron, Danielle Keats and Solove, Daniel J., "Privacy Harms", GWU Legal Studies Research Paper No. 2021-11, GWU Law School Public Law Research Paper No. 2021-11, *Boston University Law Review*, Vol. 102, 2022, http://dx.doi.org/10.2139/ssrn.3782222
[100] Boullier, Dominique. *Comment sortir de l'emprise des réseaux sociaux.* le Passeur éditeur, 2020.
[101] Foucault, Michel. *Discipline and Punish: The Birth of the Prison.* 2nd Vintage Books ed, Vintage Books, 1995.