Federated data as a commons

a third way to subject-centric and collective-centric approaches to data epistemology and politics

Calzati, Stefano

DOI
10.1108/JICES-09-2021-0097

Publication date
2022

Document Version
Final published version

Published in
Journal of Information, Communication and Ethics in Society

Citation (APA)
Calzati, S. (2022). Federated data as a commons: a third way to subject-centric and collective-centric approaches to data epistemology and politics. *Journal of Information, Communication and Ethics in Society*, *21*(1), 16-29. https://doi.org/10.1108/JICES-09-2021-0097

Federated data as a commons: a third way to subject-centric and collective-centric approaches to data epistemology and politics

Stefano Calzati
Department of Urbanism, TU Delft, Delft, The Netherlands

Received 20 September 2021
Revised 12 January 2022
Accepted 7 March 2022

Abstract

Purpose – This study advances a reconceptualization of data and information which overcomes normative understandings often contained in data policies at national and international levels. This study aims to propose a conceptual framework that moves beyond subject- and collective-centric normative understandings.

Design/methodology/approach – To do so, this study discusses the European Union (EU) and China's approaches to data-driven technologies highlighting their similarities and differences when it comes to the vision underpinning how tech innovation is shaped.

Findings – Regardless of the different attention to the subject (the EU) and the collective (China), the normative understandings of technology by both actors remain trapped into a positivist approach that overlooks all that is not and cannot be turned into data, thus hindering the elaboration of a more holistic ecological thinking merging humans and technologies.

Originality/value – Revising the philosophical and political debate on data and data-driven technologies, a third way is elaborated, i.e. federated data as commons. This third way puts the subject as part by default of a collective at the centre of discussion. This framing can serve as the basis for elaborating sociotechnical alternatives when it comes to define and regulate the mash-up of humans and technology.

Keywords Data epistemology, EU, China, Federated data, Commons, Data policies, AI ethics

Paper type Conceptual paper

1. Introduction

This article explores the (dis)empowerment of subjects and collectives in relation to data-driven technologies [1] through the lenses of data epistemology and politics. This is so largely favoured by the current normative and isomorphic understanding of data vs information, which underpins both subject-centric (e.g. the European Union) and collective-centric (e.g. China) visions on tech innovation. The goal of the article is then three-folded. Firstly, to unpack such normative understanding through the lens of data epistemology. The ethos at the basis of data-driven technologies – i.e. "accountability" and "performativity" – is discussed: beyond being techno-social artefacts, data also have a performative side, i.e. they are agents (re)enacting a precise (positivist-rationalist) worldview. This opens up interesting epistemological issues on what we can actually know,

Journal of Information, Communication and Ethics in Society
Emerald Publishing Limited
1477-996X
DOI 10.1108/JICES-09-2021-0097

as humans, through data, as well as on the politics – privileging self-determination or sovereignty – in-forming data performances.

Secondly, the article reviews the visions on tech innovation unfolded by the EU and China in policy documents. A comparison between the European Ethics Guidelines for Trustworthy artificial intelligence (AI) [2] and China's New Generation Artificial Intelligence Development Plan [3] is advanced. The two documents are taken as examples of different (geo)political visions on tech governance and innovation. While these documents raise the same concerns over the power relation binding humans to AIs (and call for human oversight on AI), the former tends to foreground a subject-based vision, the latter a collective-based vision. And yet, both visions remain trapped within the normative understanding of data discussed in Section 2.

From here, in the fourth section, it is advanced that the human-tech relation needs to be rethought as an entangled whole. In this regard, the article suggests a possible third path – i.e. federated data as a commons – that moves beyond subject-centric and collective-centric visions. In so doing, the discussion can provide the basis for elaborating sociotechnical alternatives that posit the merging of humans and technology highlighting a new (and more sustainable) ecology in the making.

2. Cutting through data epistemology and (geo) politics

2.1 Data epistemology

Data are usually considered as the unrefined version of information. Davis and Olson (1985, p. 200), for instance, write that "information is data that has been processed into a form that is meaningful to the recipient"; similarly, Silver and Silver (1989, p. 6) define data as "the raw material that is processed and refined to generate information". This is a normative stance which run through – often undetected – academic articles and policy documents alike: Gomarasca (2010, p. 139) writes that "the substantial distinction between them [data and information] is comparable to the difference between an apparently disorganized set of letters and a word assembled with the same letters". Such stance considers information as the meaningful organization of data, assuming that the two are inherently isomorphic concepts, the latter being the bricks of the former. In this article, however, a difference standpoint is taken. Not only the relation between data and information is subverted – with information preceding data – but also it is claimed that the two are not isomorphic concepts. This allows to deconstruct the "naturalization" of data and to open space for a discussion on their pragmatics.

Everything can be considered as information from any given entry point of enquiry. Physicist John Wheeler, initiator of the information-theoretic paradigm according to which information is fundamentally constitutive of reality, wittily said "it from bit", implying that reality springs out from information; he also added (quoted in Horgan, 1997, p. 84) that "at the heart of everything is a question, not an answer", pointing to the inherent speculative nature of all that is "given". From this perspective, data can be considered as (de)finite [4] bits, i.e. a *certain* fabrication of (certain) information. This helps elucidate the performative side of data, that is, "the best possible input/output equation" (Lyotard, 1986, p. 46) of any process. More to the point, as Gond *et al.* (2016, p. 447), "Lyotard's central message in problematizing performativity is [. . .] that we should be suspicious of the effects that the overriding importance attached to efficiency in the postmodern condition might have." In other words, data not only bear an agency (Dencik *et al.*, 2016; Kitchin, 2014; Metcalf and Crawford, 2016), but they are agents, and as such, they (re)enact a precise worldview, which finds its roots in rational-positivism or better, the instrumentalization of reason for mastering purposes, leading to what Hardcastle (1995) calls "executive theories". As bits of

information that can be collected, tested and shared, data in-form an understanding of reality as an already objectified dimension, while knowledge is equalled to an efficient process of objectification of that dimension (Locke, 2015). In fact, this process is based upon a *conditio sine qua non* – accountability – which draws originally upon the idea of describing by counting; thus a priori conceiving phenomena as quantities: to account, Werner Sombart (1987, p. 119) argues, means "to pursue the basic thought to grasp all phenomena only in quantities, the basic thought, thus, of quantification".

This brings to the surface a systemic flaw that has to do with the intrinsic partiality – being a parcelling of reality – of all data and it questions, consequentially, the epistemology that depends on them. "The fundamental problem", Wolfgang Drechsler (2019, p. 230) contends:

> Is that if they [data] necessarily show only partial aspects [. . .] then this means that one can always construct a set of indicators that proves any answer one wants to the question posed.

Data-driven technologies and data-led decisions are often *straightforwardly* perceived (if not regarded) as trustworthy, based on a self-validating conception of data as "reliable organized information", for the very fact of being accountable. This (auto)hypostatization obliterates all issues of partiality and interpretability of data, let alone the sociocultural fabrication of all data. It is not only that we are datafying (and automating) an increasing number of processes but also that this datafication is getting prominence over what is not (cannot be) datafied yet.

So, the question arises: what kind of information are we able to grasp beyond that being datafied? Do we even know the questions we *can* ask? As Lunch and Del Casino (2019) note:

> This requires an attentiveness to the various forms of information that may not be captured by technical devices and translated into digital data streams as well as the affective relationships among devices, bodies and spaces that escape digital capture.

At stake is not only where to look, but *what to look for* and *how*, i.e. epistemologically qualitative issues. Unfortunately, however, the answer often provided is framed in quantitative terms. A sufficient amount of data – it is claimed – will allow to get an ever more accurate picture of the phenomenon under scrutiny. "With enough data, the numbers speak for themselves", wrote Chris Anderson (2008) back in. And it was right – from a consequentialist point of view, at least. The problem is that the push towards datafication has obfuscated all alternatives.

In *Godel, Escher, Bach*, Douglas Hofstadter (1979, p. 395) acutely notes, with regard to Turing's work, that "Turing is suggesting that it is only a matter of complexity and that above a certain level of complexity a qualitative difference appears so that 'super-critical' machines will be quite unlike the simple ones hitherto envisaged." By comparing computer and brain, Hofstadter goes on to suggest that there seems to be a threshold above which mere computation cannot do the whole job alone, so to speak. The whole is irreducible to a breaking down into simpler (algorithmic) processes. Information and data are two partially incommensurable systems. Passing from one to the other is not a smooth process; it is a process of translation, with gains as well as losses. All translations, after all, are always an adaptation and betrayal of the "original": the root of the Latin word "traducere", which literally means to "bring across" and from which "traduzione", that is, "translation", comes from, is the same as "trader", which means "to deliver/to give", from which "tradimento", that is, "betrayal", derives. What data do is to trade the thickness and unboundedness of information – i.e. of phenomenological reality – in exchange for a formattable and generalizable grasping of that same reality.

Here it is argued that by accumulating and aggregating big data – i.e. by thickening the data-driven shaping of reality and by delegating to data an increasing power – we are thinning the epistemological pillars on which (Western) modernity is based, notably that same rationalist-positivist [5] worldview that has led to today's data-driven scenarios in the first place. Data are not bounded to meaningfulness (like language is); they respond to a different logic of signification – pure accountability – that escapes human discernment and understanding. And so, if that is the case, how can we say that data-led decision will be really good? For whom? And how can we say that data-driven technologies perform well? According to which parameters? It is the decoupling between actions and decisions based on (anthropological) meaningfulness, on the one hand, and actions and decision based on accountable (meaningless) performativity, on the other hand, that has produced a disorientation in the way we understand and make sense of our increasingly technologized reality.

From a governance perspective Andreas Voßkuhle (2008, p. 18) notes that "if the citizen loses trust in the state's knowledge and thus in the rationality of its decisions, the readiness to follow sovereignty commands evaporates". According to Voßkuhle, the possibility for citizens to monitor the state's rational decisional processes – i.e. to keep the state accountable for its actions – is what makes the complying of citizens to the state's sovereignty possible. The more such accountability fades, the more citizens will question the state sovereignty. It is hardly surprising, then, that today we attest to a widespread erosion of the citizens' trust in public institutions: the implementation of data-driven technologies *within* these institutions, which further automate their already anonymous bureaucracy, might not be the only cause but is surely contributing to it by enlarging the gap between (automated) decisions and (human) discernment. The institutional transparency demanded to the modern state is no longer sufficient as soon as we delegate to data and algorithms an increasing portion of our governance. Data-driven decisional processes are based on a different form of legitimacy than rational transparency, insofar as their accountability resides in the performative functioning of algorithms themselves, i.e. the delivery of an efficient output. To be sure, however, it is not only that algorithms' functioning is obscure *per se* (the usual black box metaphor); to be opaque is especially the (geo)politics that accompanies their coming to being and implementation.

2.2 (Data) sovereignty and (data) self-determination

With the consolidation of global trade and information and communication technologies (ICT) infrastructures, it becomes necessary not only to extend discussions on technological and geopolitical power relations to "subalterns" – especially in the context of North-South relations (and increasingly West-East) – but also to rethink these same discussions, shifting towards a transnational perspective whereby the global mapping of power relations is accompanied by a critical questioning of how these relations rework data-based sovereignty and self-determination at "glocal" (global+local) level. As Wen (2021, p. 12) writes in his book devoted to *Huawei's Model* "the development of the global economy has been characterized by the transition towards transnationalized digital capitalism, within which information and communications technologies have increasingly played a pivotal role in restructuring the global capitalist system". This entails to undo, on the one hand, dichotomies such as global-local, especially when it comes to issues of "data sovereignty" (Purtova, 2017) and on the other hand, to rethink from within the matter of "data self-determination" (Buitelaar, 2017), to be tackled as an increasingly transfixed multiagential concern. In this respect, Wasserman (2018, p. 448) correctly notes that at stake is the remaking of power relations at all scales, which "ha[s] prompted different ways of thinking about categories such as the 'South', the

'global', the 'local' and the 'transnational'". As Calzati (2020) points out, what we witness is the emergence of federated forms of technological entanglements – contested internally as much as externally – in which the circulation of data, tech expertise, innovation and policies can be favoured as much as hindered by competing discourses, actors and technologies belonging to and traversing different networks *at once*.

Concerning the matter of data sovereignty, studies have shown the "misalignment" between the internet as a commons infrastructure and the legitimacy of sovereign powers (Mueller, 2019), as well as the shifting towards a multipolar scenario (Winseck, 2017) in internet governance. And yet, as seen, traditional categories such as "market" and "state", "national" and "international" might no longer be sufficient. For instance, as Yu and Goodnight (2020, p. 13) note with specific regard to China "cast in light of the cybersphere, China's so-called intranet also reveals entanglements with foreign capital, foreign technology, foreign markets and foreign labor". Hence, digital sovereignty must be reconsidered as a practice which, to begin with, contests and resists linear (agent-structure) readings. This is also why discursive and geopolitical approaches heralding competing visions on ICTs (e.g. multistakeholderism vs multilateralism, with the USA and the EU on the one hand and China, on the other hand, see Nonnecke, 2016) risk becoming heuristically unproductive. It is required a digging into the realpolitik of (supposedly) opposed champs to contextualize the different (competing and overlapping) visions on tech innovation of these actors, with specific regard to the (dis)empowerment of subjects and collectives.

This leads to the second issue, that of "data self-determination". Interestingly enough, such concept is accompanied and outnumbered in terms of frequency in publications by the concept of "informational self-determination". The point is that, while the use of the latter might signal a broader understanding of what "makes up" the subject compared to the former, the extent to which the two are in fact used quite interchangeably is symptomatic of a lack of critical understanding of the difference and incommensurability between information and data. At the intersection of the concepts of human dignity and autonomy, Buitelaar (2017, p. 136) defines informational self-determination as "the [subject] capacity to determine without coercion which information about him is and will be available and accessible". Here we retrace not only an isomorphic conceptualization of information and data, which leaves no room for a critical understanding of data as the fabrication of a certain portion of information but also a normative conception of personal information as something that can be rationally and unambiguously discerned. In fact, however, the emerging technological entanglements of which legal and natural subjects are part at various scales simultaneously makes the discernment of what is personal data increasingly problematic. As Purtova (2017, p. 189) writes "the distinction between the data that is personally identifiable and non-identifiable is collapsing", meaning that "data sometimes act as personal data and at other times as non-personal data". Likely, we will be moving towards forms of federated self-determination, that is, forms of self-determination that are not individual, nor universal, but collective – or better, "collectual" as the mixing of "individual" and "collective" (see below) – and extending up to where a certain technological apparatus (i.e. a mix of infrastructures, laws and policies) claims its own sovereignty. In this respect, data self-determination will increasingly be more a law-tech affair than a sociopolitical one.

To explore how the concepts of data sovereignty and data self-determination are given actualization, we look at the official documents that two major actors – the EU and China – have so far released. On this, we follow up on van Lente (2000, p. 43), according to whom, while "statements about future technological performance are not received as factual descriptions to be verified or falsified in due course [. . .] technological futures are forceful",

meaning that discourses on tech development contribute to realize those same scenarios they depict. Notably, the comparison between the European Ethical Guidelines on AI (2019) and the Chinese Development Plan on AI (2017) will allow highlighting differences and similarities in the way in which the EU and China think of and concretely give shape to the human-tech imbrications at individual and collective levels.

3. The individual and the collective: an irreconcilable gap?

"First law: 'A robot may not injure a human being or, through inaction, allow a human being to come to harm.'

Second law: 'A robot must obey the orders given to it by human beings except where such orders would conflict with the first law.'

Third law: 'A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.'

Zeroth law: 'A robot may not harm humanity, or, by inaction allow humanity to come to harm.'" Isaac Asimov (1942)

As Roberto Simanowski (2019, p. 436) acutely notes by portraying a near-future scenario ruled by a green-sustainable AI: "Once humans have finished discussing the reduction of carbon emissions, artificial intelligence could then ensure with the iron stubbornness of algorithmic if-then logic that the resolutions are implemented." In the words of Ben Goertzel (2011), this means that AI would become "a mildly superhuman supertechnology, whose job is to protect us from ourselves", even at the cost of our own behavioural nudging or physical annihilation. This would eventually subvert the long-lasting hegemony not only of humans over the environment but also of humans' control over themselves. From here, it emerges more clearly the partially irreducible gap between a conception of "humans" as individuals and humanity as a collective. Interestingly, such gap also appears in the two documents under analysis. Indeed, the main differences between the EU Ethics Guidelines on AI and China's development plan for AI concern the weight given, by each one, to the conception of "humans" as individuals and humanity as a collective, not dissimilarly from the articulation of the laws of robotics by Isaac Asimov (in excerpt).

3.1 The European Union vision on AI

The EU has become a leader in shaping the discourse on AI ethics. The Ethics Guidelines released in April 2019 by the High-Level Expert Group (HLEG) of the European Commission constitute the first pillar of the EU ethical framework on AI, to which a second document – the Policy and Investment Recommendations [6] (2019) – followed, after a couple of months. Central to both documents is the notion of "trustworthy AI". With this term, the HLEG took a firm stance concerning the conception of AI as a set of data-driven technologies that must be:

- lawful ("complying with all applicable laws and regulations");
- ethical ("ensuring adherence to ethical principles and values"); and
- robust ("developed in a responsible manner from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm").

While at the beginning of the Ethics Guidelines, the stress is put on humanity as a whole, stating that AI must be put "in the service of humanity and the common good, with the goal of improving human welfare and freedom", the subsequent disentanglement of the various principles that substantiate "trustworthy AI" makes it clear that the focus is first and foremost on the individual(s), in line with the rationalist enlightenment tradition.

The Ethics Guidelines document specifies that AI must be human-centric, i.e. designed by keeping human autonomy and integrity as default principles. Beyond that, the document introduces a double-sided concept of "fairness" as both the assurance that through AI "individuals and groups are free from unfair bias" and "the ability to contest and seek effective redress against decisions made by AI systems". This dual – substantive and procedural – understanding of fairness considers the recognition of one's own diversity and inclusion as a paradigmatic condition for the development of "trustworthy AI", also acknowledging that the individual *is* always part of a group. Conjointly, the notion of "explicability" is introduced in connection with human autonomy and decision-making. "Explicability" is characterized as "the ability to explain both the technical processes of an AI system and the related human decisions". Such characterization brings the rationale behind the human-technology interaction to the fore, with the anthropological factor that is (still) put in control. This, however, as we will see in Section 4, is increasingly problematic and difficult to apply in a concrete situation. More broadly, while it is true that since 2014 the European Commission (2021) has taken steps to facilitate the development of a citizen-centred, secure and competitive data landscape, it is nonetheless true that recent calls (European Commission, 2021) go in the direction of a market-driven conceptualization of AI innovation and data management, opening questions concerning the inclusiveness and sustainability of such calls.

3.2 The Chinese vision on AI

In 2017 China's state council released its Development Plan on AI. This civic-military plan includes a timeline that, if abided by, will make China the leader in AI research and development by 2030. Three are the major steps identified:

(1) by 2020, China aims to maintain competitiveness with other major powers and optimize its AI development environment while also establishing initial ethical norms, policies and regulations for vital areas of AI;

(2) by 2025, China aims to be world-leading in some AI applications, it targets an increase in the worth of its core AI industry to over ¥400bn ($58bn), and it plans to expand upon and codify in law, ethical standards for AI; and

(3) by 2030, China seeks to become the world's innovation centre for AI and articulate further upgrades in the laws and standards.

As it is typical with Chinese Governmental documents that provide general guidelines for various sectors, this too is characterized by some precise economic goals and more abstract conceptualizations of such goals. The framing of AI ethical principles falls under the latter of these two tendencies. As some researchers have noted (Roberts *et al.*, 2020), the document "outlines a clear intention to define ethical norms and standards, yet efforts to do so are at a fledgling stage, being broadly limited to high-level principles, lacking implementation". This implementation is what the EU has provided in the Policy and Investment Recommendations which are *based on* "trustworthy AI". China's state council, by contrast, has identified three broad areas – "international competition", "economic development" and "social governance" – to make the achievements of its goals possible. In this sense, China considers AI ethics as a *desideratum* to be sought along the way rather than a principle to be followed by default.

In fact, the Chinese authorities' guidance in the matter of ethical AI can be considered as a reflection of the "Chinese way" to ethics. Such a way echoes Confucianism and foregrounds the well-being of the collective over the individual. The Confucian collective is regarded as a dimension characterized by an ethos of distributed harmony to be achieved *even* to the detriment of the individual (Wong, 2013). This is also why "social governance" is

one of the key areas identified by the government: the individual is subsumed under the social collective. Such ethical AI vision may even lead to a form of AI as envisioned by Simanowski, i.e. one acting against the individual for the sake of the many. And in this scenario, the Chinese authorities would maintain the role as guarantors of collective well-being. Overall, China's state-led approach to tech innovation has been shown to present shortcomings, notably by being encumbered by bureaucracy (Sun, 2007) and by the lack of a robust link between research & development and industry (Huang and Wu, 2012).

Here, we are far from the EU vision that calls for a public discussion on AI ethics: "important ethical questions should be approached with the help of a wide consultation of civil society", it is stated in the Policy and Investment Recommendations. Rather, China adopts a top-down authoritarian approach which remains so far unquestioned. Of course, at stake is not an evaluation of which approach is "best", but the extent to which they concretely give shape to a precise techno-human ecosystem (and the extent to which they will be able to co-exist). As Luciano Floridi *et al.* (2018, p. 698) note:

> Humans should always retain the power to decide which decisions to take, exercising the freedom to choose where necessary and ceding it in cases where overriding reasons, such as efficacy, may outweigh the loss of control over decision-making.

Most importantly, such power and control of humans over automated data-driven technologies and decisions should be not only individual and ex post but *collective* and ex ante. After all, why should we consider autonomy only as an individual/personal matter? Here, it emerges more clearly the need to rethink from scratch the relation human-technology.

4. Towards an entangled understanding of the human-technology relation

The two examples discussed above show the normative discourses surrounding the implementation of data-driven solutions in different geopolitical and sociocultural settings. At stake is not only the correct identification of new forms of "distributed moral responsibility" (Floridi, 2016) that reconceptualizes the human-technology relation – in either individual or collective terms – but also the more radical idea that a clear-cut distinction between "human" and "technology" is increasingly problematic. Disentangling "entities" – be they human or technological – that are more and more imbricated becomes a perilous task not only conceptually but also pragmatically. Along these lines, a 2018 IEEE White Paper [7] talks of "symbiotic autonomous systems" to herald a new techno-human dimension:

> We are starting to see the emergence of a Digital Age in which the material to be manipulated is no longer (just) atoms but also bits. We are entering into this new age through a symbiotic relationship with our digital tools. These new tools have become complex entities that are probably better referred to as systems as they are starting to operate autonomously, due to a growing flexibility and awareness.

This means that AI decisions are and will increasingly be the result of entangled performances among bio-techno-social actors, in which the anthropological centrality gets inevitably contested, if not marginalized. Overall, we will be living (in) dynamic, interactive "milieux" in which the tech-human dualism will be resolved as a multi(f)actorial coalescence which will define its own horizon of existence and action at all times. In the words of Rose (2017, p. 789), it is necessary to "theorize (digital) posthuman agency by thinking it as always already (digitally) sociotechnical". As such, this emerging condition bypasses and reworks both the European vision on AI that foregrounds individual(s) and the Chinese one that foregrounds the social collective. To be envisioned is the shift from a networked (rationalist-positivist) approach to an entangled approach that requires discussing the consubstantial

human-technology relation. It is a new "collectual" dimension that is being fostered, where "individual" and "collective" are transfixed and constantly remoulded in practice.

In fact, the emerging entangled scenario is one where techno-human actors can be seen as co-subjective and ever-evolving rather than as entities making up the whole societal dimension. The basic idea is that "human" and "technology" cannot be taken as one-dimensional concepts: each of them also presupposes the other. Hence, the best way to address such entangled scenario is by de-essentializing it and acknowledging the lack of internal hierarchical subordination. It is necessary to think in terms of a whole new physis being unfolded rather than of "elements" interacting in a networked fashion. This idea foregrounds what could be called "peer inter-aggregation". Peer inter-aggregation can be understood as a form of shared practice, or a "commoning" (De Angelis, 2017), which is non-appropriative by default (knowledge, technology, assets and outputs are not owned, in the commercial sense of the term, but summoned up and recirculated); collaborative by design (it considers all nodes and edges of the system as integral and necessary to the system's flourishing) and collectively sustainable in its goals (indeed, commons for the community).

Moreover, talking of "peer inter-aggregation" allows moving beyond an anthropocentric vision of the techno-human entanglement towards the conception of a system of peers where nodes (aggregates) are not only humans but also they encompass organic and inorganic elements, as well as intangible assets. This is why the term "aggregate" fits particularly well: nodes are not fixed entities but multifaceted instantiations that mutually reconfigure each other in different scenarios and over time. Fundamentally, what technology can do must be evaluated in "humane" (Benjamin, 2003) terms and not only at the service of people. It is important, indeed, to recognize the ecology in the making of which both people and technology are part: "humane", then, does not stand for "anthropological", but "sustainable". Here, the concept of "homeotechnologies" by Peter Sloterdijk (2000, 91) comes in handy: differently from what he calls "allotechnologies", which "violate" natural resources, "homeotechnologies are developed based on "ecology" [and] entail a strategy of 'cooperation', of 'dialogue' with nature".

In this context, data can be seen as the ultimate commodification of technologization. The distinction between information and data returns here: information is surrounding us regardless of the means we have to catch it; data are a certain fabrication of that same information made possible under certain technological conditions. As Hess and Ostrom (2007, p. 10) explain:

> [. . .] [t]his ability to capture the previously uncapturable creates a fundamental change in the nature of the resource, with the resource being converted from a nonrivalrous, nonexclusionary public good into a common-pool resource that needs to be managed, monitored and protected, to ensure sustainability and preservation.

This also means that practices of data sovereignty will increasingly remodel the local-global scale towards forms of transpatial articulation whereby the affiliation to certain legal-ethical frameworks gets repeatedly reworked, customized and traversed by external geopolitical pushes as much as internal tensions which depend on tech innovation. This, among other things, decouples and expands the nexus territoriality-authority), questioning what it means to be a subject in legal, social, as well as technical terms.

Through the concept of peer inter-aggregation, the "investment" is made, from the outset, on the community as an ecology of people, ideas, technology, knowledge, resources and capital, rather than on the reification of relations and goals. Regardless of its form, such ecology demands the coupling of accountability and legitimacy not only as legal concepts but

also thick political practices in which the "public" – as a formalized sector and above all, as ex ante collective assessment – regains central stage. In this context, a discussion (and a reworking) of the idea of "federated data" comes in handy. Federated data are public-private consortia – regulated by precise rules and agreements – in which actors do not necessarily own the data they process, but rather borrow them from the "federation" of which they are part. As soon as such borrowing is redefined in view of a communal interest – for instance, by requiring that a certain value derived from the use of data is recirculated within the consortium – federated data become entangled systems with self-sustaining cycles. In other words, federated data can be thought of and shared as a commoning practice. In this regard, while single actors can still be identified, their constantly redefined co-operation makes the federation work as a whole. This, in turn, redefines issues of data sovereignty and self-determination as a matter that cuts across the local-global spectrum and demands to rethink legal and ethical frameworks towards sociotechnical ecologies that take a "collectual" perspective by default (note that this collectual perspective is respectful of privacy insofar as data would be circulated on a voluntary basis and regulated by legal agreements).

While federation is the context, so to speak, the core idea is to operationalize data as a commons. The term "commons" originally identified common-pool resources (CPRs) – such as fisheries or forests – characterized by non-excludability and rivalry. These terms point to the fact that:

- it is difficult to forbid access and use of CPRs to any potential beneficiary; and
- the use of CPRs depletes them and reduces further use by others.

The comprehensive fieldwork-based work by Elinor Ostrom (1990, 2010) showed that the self-management of CPRs by communities can be more efficient and sustainable than market-driven or state-led approaches, provided that formal and informal principles and roles are designed and abided to.

By now, the commons has spilled over onto other realms, such as digital infrastructures and services. Through ICTs, the digital commons has taken up a global connotation alongside its local nature. As part of the digital commons, data commons characterizes a regime in which actors – physical or juridical – join forces in the collection, pooling and use of data (and digital knowledge) subservient to the delivery of products or services. Data commons initiatives aim to counteract the centralized corporate and/or authoritarian ownership and use of data – either by tech companies or states – by giving these back to citizens, with the goal to foster sustainable collective data practices. In 2016, for instance, the Catalan municipality launched a "new social pact on data": various initiatives informed the new digital agenda, among which platform cooperatives based on data commons regimes, allowing citizens to own and keep control over their data (Morozov and Bria, 2018). This, in fact, is the key tenet: by acknowledging the sociotechnical agential nature of data and by recognizing the deep imbrication between technology and people, the conceptualization of federated data as a commons represents the preliminary step for a radical reshaping of humans' way of living, sidelining profit-oriented resources' exploitation and foregrounding sustainable well-being. Problems, however, remain: notably, preliminary studies on data commons initiatives (Calzada and Almirall, 2019, 2020) show:

- the need for similar experiments at different scales along the global-local axis; and
- a lack of theoretical guidance and applied systemic policy governing such initiatives.

5. Conclusion

In the first part of the article, data-driven technologies were framed within an epistemological and (geo)political perspective. On the one hand, it was highlighted that data – beyond embedding precise agencies – are also performative, i.e. agents that put forth a precise worldview based on accountability and performativity; on the other hand, the (geo) political implications that surround the coming into being and implementation of data-driven technologies were explored, with specific regard to issues of data sovereignty and data self-determination.

To take a pragmatic stance towards such issues, in the second part of the article, the European Ethics Guidelines on Trustworthy AI and China's New Generation Artificial Intelligence Development Plan were scrutinized and compared. In so doing, it was possible to assess the extent to which such documents, although sharing basic principles such as the need for humans to always maintain the upper hand over data-driven technologies, unfold a different perspective concerning how AI should be implemented in society. While the EU unfolds a subject-centric approach, China subordinates the individual to the collective, and in this scenario, state authorities play the role as guarantor of social ethics. And yet, both visions are still rooted in a normative positivist understanding of human and technology which keep them as separate entities.

As an alternative, it is advanced the need to rethink the human-technology relation in holistic terms, leading to the definition of new boundaries-in-practice that literally transfix the self-collective spectrum, as well as the local-global scale. A concretization of such conceptual idea is the (re)definition of federated data as a commons, whereby actors are led to cooperate in the sharing of data and recirculate the outcome of their (data) activity within the federation. This, in turn, can help ignite a reshaping of the way developed communities live, stressing sustainable well-being more than for-profit exploitation of natural, technological, or human resources. From this perspective, however, while data commons initiatives that attempt to repurpose data and technology for the community are emerging worldwide, limits remain concerning the scalability of the federated-data-as-a-commons approach.

Notes

1. The definition of AI given by the HLEG refers "to systems that display intelligent [rational] behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals". Here "data-driven technologies" and "AI" are synonyms to the extent to which AI, be it in its software or hardware form, needs data and algorithms to take autonomous decisions rationally. In Section 4, the term "technology" comes also to encompass digital devices and the ICTs at large.
2. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
3. http://fi.china-embassy.org/eng/kxjs/P020171025789108009001.pdf
4. The Latin "definire" means exactly to trace boundaries, to mark where something ends ("finire"), thus also deciding what something is.
5. "Positivism" comes from the Latin "positum", which means "grounded/established" and by extension, "what is posited/given".
6. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence
7. https://digitalreality.ieee.org/images/files/pdf/1SAS_WP3_Nov2019.pdf

References

Anderson, C. (2008), "The end of theory: the data deluge makes the scientific method obsolete", available at: www.wired.com/2008/06/pb-theory/

Benjamin, W. (2003), "On scheerbart", in Bullock, M. and Jennings, W. (Eds), *Walter Benjamin: Selected Writings*, Belknap Press, Cambridge, Vol. 4, pp. 386-388.

Buitelaar, J.C. (2017), "Post-mortem privacy and informational self-determination", *Ethics and Information Technology*, Vol. 19 No. 2, pp. 129-142.

Calzada, I. and Almirall, E. (2019), "Barcelona's grassroots-led urban experimentation: deciphering the 'data commons' policy scheme", SSNR, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3359268

Calzada, I. and Almirall, E. (2020), "Data ecosystems for protecting European citizens' digital rights", *Transforming Government: People, Process and Policy*, Vol. 14 No. 2, pp. 133-147.

Calzati, S. (2020), "Decolonising 'data colonialism': propositions for investigating the realpolitik of today's networked ecology", *Television and New Media*, Vol. 22 No. 8, pp. 914-929, doi: 10.1177/1527476420957267.

Davis, G.B. and Olson, M.H. (1985), *Management Information Systems: Conceptual Foundations, Structure and Development*, McGraw-Hill, New York, NY.

De Angelis, M. (2017), *Omnia Sunt Communia: On the Commons and the Transformation to Postcapitalism*, Zed Books, London.

Dencik, L., Hintz, A. and Cable, J. (2016), "Towards data justice? The ambiguity of anti-surveillance resistance in political activism", *Big Data and Society*, Vol. 3 No. 2, doi: 10.1177/2053951716679678.

Drechsler, W. (2019), "Kings and indicators: options for governing without numbers", in Prutsch, M.J. (Ed.), *Science, Numbers and Politics*, Palgrave Macmillan, London, pp. 227-262.

European Commission (2021), "Digital Europe program (DIGITAL)", available at: https://digital-strategy.ec.europa.eu/en/activities/digital-programme

Floridi, L. (2016), "Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions", *Philosophical Transactions*, Vol. 374 No. 2083, pp. 1-13.

Floridi, L., *et al.* (2018), "AI4People – an ethical framework for a good AI society: opportunities, risks, principles and recommendations", *Minds and Machines*, Vol. 28 No. 4, pp. 689-707.

Goertzel, B. (2011), "Does humanity need an AI nanny? h+magazine", available at: http://hplusmagazine.com/2011/08/17/does-humanity-need-an-ai-nanny

Gomarasca, M.A. (2010), "Basics of geomatics", *Appl Geomat 2*, pp. 137-146, doi: 10.1007/s12518-010-0029-6.

Gond, J.P., Cabantous, L., Harding, N. and Learmonth, M. (2016), "What do we mean by performativity in organizational and management theory? The uses and abuses of performativity", *International Journal of Management Reviews*, Vol. 18 No. 4, pp. 440-464.

Hardcastle, V.G. (1995), "A critique of information processing theories of consciousness", *Minds and Machines*, Vol. 5 No. 1, pp. 89-107.

Hess, C. and Ostrom, E. (2007), "Analyzing the knowledge commons", in Hess, C. and Ostrom, E. (Eds), *Understanding Knowledge as Commons*, MIT Press, Boston.

Hofstadter, D. (1979), *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, New York, NY.

Horgan, J. (1997), *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age*, Broadway, New York, NY.

Huang, C. and Wu, Y. (2012), "State-led technological development: a case of china's nanotechnology development", *World Development*, Vol. 40 No. 5, pp. 970-982.

Kitchin, R. (2014), "Big data, new epistemologies and paradigm shifts", *Big Data and Society*, Vol. 1 No. 1, doi: 10.1177/2053951714528481.

Locke, K. (2015), "Performativity, performance and education", *Educational Philosophy and Theory*, Vol. 47 No. 3, pp. 247-259.

Lynch, C.R. and Del Casino, V.J. Jr. (2019), "Smart spaces, information processing, and the question of intelligence", *Annals of the American Association of Geographers*, doi: 10.1080/24694452.2019.1617103.

Lyotard, J.-F. (1986), *The Postmodern Condition: A Report on Knowledge*, University of MN Press, Minneapolis.

Metcalf, J. and Crawford, K. (2016), "Where are human subjects in big data research? The emerging ethics divide", *Big Data and Society*, Vol. 3 No. 1, doi: 10.1177/2053951716650211.

Morozov, E. and Bria, F. (2018), "Rethinking the smart city. Democratizing urban technology", Rosa Luxemburg Foundation, New York, NY, available at: https://onlineopen.org/media/article/583/open_essay_2018_morozov_rethinking.pdf

Mueller, M. (2019), "Sovereignty and cyberspace: Institutions and internet governance", available at: https://dlc.dlib.indiana.edu/dlc/bitstream/handle/10535/10410/5th-Ostrom-lecture-DLC.pdf?sequence=1&isAllowed=y (accessed 10 June 2021).

Nonnecke, B. (2016), "The transformative effects of multistakeholderism in internet governance: a case study of the East Africa internet governance forum", *Telecommunications Policy*, Vol. 40 No. 4, pp. 343-352.

Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, Cambridge.

Ostrom, E. (2010), "Beyond markets and states: polycentric governance of complex economic systems", *American Economic Review*, Vol. 100 No. 3, pp. 641-672.

Purtova, N. (2017), "Health data for common good: defining the boundaries and social dilemmas of data commons", in Adams, S., Purtova, N. and Leenes, R. (Eds), *Under Observation: The Interplay between Ehealth and Surveillance*, Springer, Berlin, pp. 177-210.

Roberts, H., Cowls, J., Morley, J., *et al.* (2020), "The Chinese approach to artificial intelligence: an analysis of policy, ethics and regulation", *AI and Society*, Vol. 36, pp. 59-77, doi: 10.1007/s00146-020-00992-2.

Rose, G. (2017), "Posthuman agency in the digitally mediated city", *Annals of the American Association of Geographers*, Vol. 107 No. 4, pp. 779-793.

Silver, G.A. and Silver, M.L. (1989), *Systems Analysis and Design*, Addison Wensley, Reading.

Simanowski, R. (2019), "On the ethics of algorithmic intelligence", *Social Research*, Vol. 86 No. 2, pp. 423-447.

Sloterdijk, P. (2000), *La Domestication de L'être*, Mille et une nuits, Paris.

Sombart, W. (1987), "Der moderne Kapitalismus: Historischsystematische Darstellung des gesamteuropäischen Wirtschaftslebens von seinen Anfängen bis zur Gegenwart", vol. 2, pt. 1: Das europäische Wirtschaftsleben im Zeitalter der Frühkapitalismus, vornehmlich im 16., 17. und 18. Jahrhundert. München: Deutscher Taschenbuch Verlag.

Sun, P. (2007), "Is the state-led industrial restructuring effective in transition China? Evidence from the steel sector", *Cambridge Journal of Economics*, Vol. 31 No. 4, pp. 601-624.

van Lente, H. (2000), "Forceful futures: from promise to requirement", in Brown, N. and Rappert, B. (Eds), *Contested Futures: A Sociology of Prospective Techno-Science*, Ashgate, New York, NY, pp. 43-64.

Voßkuhle, A. (2008), "Das konzept des rationalen staates", in Schuppert, G.F. and Voßkuhle, A. (Eds). *Governance Von Und Durch Wissen*, Nomos, Baden-Baden, pp. 11-33.

Wasserman, H. (2018), "Power, meaning and geopolitics: ethics as an entry point for global communication studies", *Journal of Communication*, Vol. 68 No. 2, pp. 441-451.

Wen, Y. (2021), *The Huawei Model: The Rise of China's Technology Giant*, University of IL Press, Urbana.

Winseck, D. (2017), "The geopolitical economy of the global internet infrastructure", *Journal of Information Policy*, Vol. 7 No. 1, pp. 228-267.

Wong, P.H. (2013), "Confucian social media: an oxymoron", *Dao*, Vol. 12 No. 3, pp. 283-296.

Yu, H. and Goodnight, T. (2020), "How to think about cyber sovereignty: the case of China", *Chinese Journal of Communication*, Vol. 13 No. 1, pp. 8-26.

Further reading

Milan, S. and Treré, E. (2019), "Big data from the South(s): beyond data universalism", *Television and New Media*, Vol. 20 No. 4, pp. 319-335.

Corresponding author

Stefano Calzati can be contacted at: s.calzati@tudelft.nl