Dismantling Digital Cages

Examining Design Practices for Public Algorithmic Systems

Nouws, Sem; Janssen, M.F.W.H.A.; Dobbe, Roel

DOI
10.1007/978-3-031-15086-9_20

Publication date
2022

Document Version
Final published version

Published in
Electronic Government - 21st IFIP WG 8.5 International Conference, EGOV 2022, Proceedings

Dismantling Digital Cages: Examining Design Practices for Public Algorithmic Systems

Sem Nouws(✉), Marijn Janssen, and Roel Dobbe

Delft University of Technology, Delft, The Netherlands
{s.j.j.nouws,m.f.w.h.a.janssen,r.i.j.dobbe}@tudelft.nl

Abstract. Algorithmic systems used in public administration can create or reinforce *digital cages*. A digital cage refers to algorithmic systems or information architectures that create their own reality through formalization, frequently resulting in incorrect automated decisions with severe impact on citizens. Although much research has identified how algorithmic artefacts can contribute to digital cages and their unintended consequences, the emergence of digital cages from human actions and institutions is poorly understood. Embracing a broader lens on how technology, human activity, and institutions shape each other, this paper explores what design practices in public organizations can result in the emergence of digital cages. Using Orlikowski's structurational model of technology, we found four design practices in observations and interviews conducted at a consortium of public organizations. This study shows that design processes of public algorithmic systems (1) are often narrowly focused on technical artefacts, (2) disregard the normative basis for these systems, (3) depend on involved actors' awareness of socio-technics in public algorithmic systems, (4) and are approached as linear rather than iterative. These four practices indicate that institutions and human actions in design processes can contribute to the emergence of digital cages, but also that institutional – opposed to technical – possibilities to address their unintended consequences are often ignored. Further research is needed to examine how design processes in public organizations can evolve into socio-technical processes, can become more democratic, and how power asymmetries in the design process can be mitigated.

Keywords: Public algorithmic system · Digital cage · Design process · Structuration

1 Introduction

Algorithmic systems used in public administration can have a detrimental impact on citizens. The predictive, structuring, and learning capacities of algorithmic applications are used to, for example, allocate social services, assess livability of neighborhoods, or for predictive policing [35]. However, practice shows that incorrect decisions are made by or based on these systems. For example, Ranchordas & Scarcella [24] discuss two cases in the USA and the Netherlands in which risk indication models were used to

Published by Springer Nature Switzerland AG 2022
M. Janssen et al. (Eds.): EGOV 2022, LNCS 13391, pp. 307–322, 2022.
https://doi.org/10.1007/978-3-031-15086-9_20

predict cases of fraud. In both cases, many citizens were falsely accused of fraud, had to pay back received benefits, and were not able to rectify the incorrect decisions. This caused high debts, and mental and physical health problems for the affected citizens. Following Peeters & Widlak [23], we consider these incorrect decisions and their impact as unintended consequences of digital cages. A *digital cage* refers to the rigidity of algorithms and information architectures that results in automation creating its own reality.

Research on the digital cage and its consequences has increased in recent years. Harms produced by public algorithmic systems have been extensively described. For example, studies show that algorithmic decision-making is often discriminatory (e.g., [26, 30]), that systems make incorrect decisions (e.g., [23]), that both citizens and users lose agency when these systems are used (e.g., [22, 31]), and that algorithmic systems are inscrutable (e.g., [1, 14, 29]). These findings urge public organizations to search for instruments to ensure safe, explainable and accountable algorithms.

Whereas most research is focused on the effects of algorithmic systems on citizens, the ways in which digital cages emerge are poorly understood. Digital cages and their unintended consequences result from the implementation of digital and analytical technologies in public administration (assuming that these cages do not intentionally emerge from political or strategic motives). Several authors have studied practices that may partly explain the emergence of digital cages. For example, the increase of chain decisions [34], the emphasis on technocratic governance [13], or the search for a complete view on citizens [4] can explain the materialization of algorithmic systems with severe impacts on citizens. Other scholars look at the development of AI systems in general, for example, the way in which vagueness is often wrongly addressed by developers and therewith results in citizen harms [5]. Finally, some authors specifically examine the dynamics between actors in public organizations. For example, Van der Voort et al. [25] analyzed the interactions between data-analysts and policy-makers. These authors all provide leads to approach digital cages from a socio-technical perspective. However, this perspective has not been used to study the specific case of emerging digital cages in public algorithmic systems. This impedes initiatives of public organizations to prevent, mitigate, or correct these cages.

Hence, the goal of this paper is to gain insight into what design practices of public algorithmic systems can contribute to the emergence of digital cages and their consequences. We study this emergence by using the structurational model of technology of Orlikowski [20] as an analytical lens. We use this model to empirically analyze design practices in public organizations based on observations and interviews. The analysis indicates four practices that can be related to the emergence of digital cages. We will close this paper by presenting the implications of these design practices.

2 Motivation and Background

To study unintended consequences of public algorithmic systems, Peeters & Widlak [23] demonstrated the usefulness of the digital cage as analytical concept. As the concept will be central in this paper, we will elaborate on it in this section. Thereafter, we show that the scope of digital cage research should be broadened by using the structurational model of technology and that practice can also use such a perspective.

2.1 Digital Cages

The use of algorithmic systems in public administration alters the bureaucratic organizations that they are situated in. The emergence of these systems has triggered refinements of Weber's [32] notion of bureaucracy (e.g., [15, 36]) as they change the nature of formalization in bureaucracy. For example, algorithmic decision-making based on code is less flexible compared to the deliberative practice of a legal system based on speech and written word [10]. Furthermore, algorithmic formalization results in what Janssen & Kuk [13] characterize as *technocratic governance*: "assuming that complex societal problems can be deconstructed into neatly defined, structured and well-scoped problems that can be solved algorithmically and in which political realities play no role" (p. 372).

Although Weber's notion of bureaucracy is not fully applicable to public administration in the current information society, bureaucracies using algorithmic and information systems do produce the digital equivalent of Weber's *iron cage:* the *digital cage.* The iron cage represents the continuous rationalization of society – through rules and procedures – over which individuals have no control. In the digital cage, these rules and procedures are (partly) replaced algorithmic and information systems. Peeters & Widlak [23] define the digital cage as "a highly disciplining infrastructure that rationalizes the execution of tasks through information architecture and algorithms instead of Weberian rules and procedures" (p. 182). Both civil servants and dependent technical or social systems are disciplined by the digital cage. Civil servants see their street-level discretion curtailed [34], their agency limited [22], and their behavior changed as their daily routines are governed by the cage [13]. Algorithmic systems can discipline other systems such as laws or other automated decision-making systems. For example, they can dictate the interpretation of a law and one small error in the algorithmic system can result in an accumulation of errors in systems that depend on the algorithm's outcome [34].

Adverse effects of digital cages are the exclusion of citizens and the obscuration of the decision-making process, which is another similarity to Weber's bureaucracy. His conception of the iron cage has often been compared to the excluding and disorienting implications of iron cages in Franz Kafka's work (e.g., [11]). Like iron cages, digital cages can create Kafkaesque situations in which citizens are caught up in a digital bureaucratic system without knowing how to solve their problems. For example, Peeters & Widlak [23] describe the case of a woman whose car was incorrectly registered. The registration ultimately led to large and erroneous tax debts at different public organizations. The woman was not informed about the incorrect registration. Moreover, she could not find out who made the mistake or which organization could correct the registration [33].

These adverse effects, which can be considered as unintended consequences of digital cages, are often studied from a deterministic perspective. The goal to make decision-making more fair, accurate, and efficient through algorithmic systems often backfires. This was the case for the two risk indication models mentioned in the introduction. Tax offices wanted to make detecting fraud more efficient and accurate but created situations in which the fraud cases detected by their algorithmic system did not resemble reality [24]. To the best of our knowledge, such cases are mostly studied from the deterministic perspective discussed above. In other words, research emphasizes the way in which the digital cages – its technology – disciplines human agents and other systems. However,

that ignores the way in which these digital cages – and therefore their unintended consequences – emerge. In order to prevent, mitigate, or correct the unintended consequences of digital cages, a broader perspective is needed that also includes human actions and the influence of institutions – i.e., social rules that structure the behavior of human agents [12] – on the emergence of these cages.

2.2 The Structuration Model of Technology and the Digital Cage

The structurational model of technology by Orlikowski [20] – based on Giddens' structuration theory [6–8] – is a framework that can show the shortcomings of a deterministic perspective on digital cages. The model is based on the idea that socio-technical systems comprise three components: a technical artefact, human agents, and institutions. Orlikowski describes four important interactions between these components. Firstly, technology mediates human action. Secondly, the use of technology can reinforce or transform institutions. Thirdly, technology is the product of human action. Finally, institutions structure human actions. This combination of interactions makes full control over the complete trajectory of outcomes impossible. Therefore, unintended consequences can emerge (see Orlikowski [20] for an elaboration on the model).

Since public algorithmic systems are socio-technical systems, they can be studied with the structuration model of technology. In these systems, the technical artefact is a machine-based application (cf. [19]) – either rule-based or case-based – situated in an information architecture. Human agents are involved in or affected by the system as the application automates, supports, or augments (parts of) decision-making in public administration. Finally, the public algorithmic system includes institutions that form the basis for an algorithmic system, institutions that constrain and structure the usage of algorithmic system, and institutions that organize the design process of such systems. Without the human agents and institutions, the information architecture does not work. Therefore, the disciplining nature of digital cages cannot only arise from the information architecture.

Research on digital cages has mostly focused on two types of interactions described by the structurational model. First, descriptions of how the digital cage disciplines users and dependent systems focus on how technology mediates human action. For example, the reduction of street-level discretion by the system. Second, research showing the changes of algorithm use to public administration emphasize the reinforcement or transformation of institutions by technology. This can be observed in the technocratic governance that arises because of using public algorithmic systems. The way in which human actions – structured by institutions or not – contribute to digital cages is underexplored. Therefore, this research focuses on influence of human action and institutions on the emergence of digital cages and their unintended consequences.

2.3 The Design Process of Public Algorithmic Systems

In practice, the design process of public algorithmic systems also shows ignorance of the third and fourth interactions in the structurational model. This is problematic as human actions and institutions are steering the design process. After all, the *design process* is a set of deliberative actions – performed by agents who are not directly affected by the

system – to shape or change the public AI systems in reaction to a question, a problem statement, or emergent behavior of the system.

The emergence of public algorithmic systems confronts public organizations with the shortcomings of their current design and policy-making processes. The democratic basis of design processes is often obscure, the political debate in design processes is underdeveloped, and the design process emphasizes technological fixes [17, 18]. Similarly, Van Zoonen [27] observes that the "transition to data-driven social policy almost completely takes place out of political and social view" (p. 3) and that the design process is in an institutional void.

Considering the flaws in current design processes, this research focuses on the influence of current design practices – i.e., a specific constellation of human action and institutions in the design process – on the emergence of digital cages and their unintended consequences. When considering human action, one can examine both the design and use mode of systems. These are analytical lenses to distinguish between two general classes of activities [20]. In the design mode, the system is intentionally created by designing actors. The system is also changed when users operate the system. This article focuses mostly on the design mode, as this is the part that public organizations mostly focus on. However, the two modes cannot fully be examined separately. Therefore, this article will sometimes also refer to the usage of public algorithmic systems.

3 Research Methods

We identified design practices that may result in digital cages and their unintended consequences by using an explorative case study. We analyzed how institutions and human actions in those organizations influence the structure of technical artefacts. Public organizations are considered as designing actors of public algorithmic systems. An explorative case study research was performed to study current design processes in public organizations through the structurational model of technology.

The explorative research was conducted at a consortium of Dutch public organizations that collaboratively realized policy instruments for public control on the development and use of algorithms. They developed an algorithm register, procurement conditions, a governance framework for development and use, and guidelines for objection procedures in government. The consortium consisted of several Dutch public organizations – five municipalities, three provinces, and three executive agencies – that were supported by the Ministry of Internal Affairs. The consortium was selected as a useful case, since the development of the instruments also asked for a reflection on the design process of algorithms. Furthermore, the participating organizations represented the majority of public organization types in the Netherlands; the consortium, therefore, also provided insight in differences between organizational design processes.

Since the consortium had just started with brainstorming about the policy instruments, the explorative research started with observations of the biweekly meetings of the core team members of the consortium. In these meetings, the project leaders shared the progress of the instruments and shared insights from within their own organizations. In total, 10 meetings, taking place from March to December 2020, were used for observations. The participants of the meetings differed, but the 5 core team members – representing two municipalities and two provinces – were consistently involved in the

meetings. The observations were documented in a logbook. Documents shared during the meetings were also included in the logbook.

To test and complement the insights of the observations, semi-structured interviews of 90 min were conducted with the four project leaders after the observations. Four of the five core members – i.e., one lead developer and three policy makers involved in compliance; representing two municipalities and one province – were interviewed. The core team members were the project leaders of the instruments, and therefore had a full overview of the progress of the consortium. The interviewees were first asked to describe and reflect on the design process within their own organizations. Thereafter, they were asked to evaluate the policy instruments of the consortium. The topic guide of the interviews was based on the Institutional Analysis and Development (IAD) framework by Elinor Ostrom [21]. This analytical framework provides 11 important variables that need to be considered when looking at human action and the emergence of institutions.

We coded the observatory records and interviews transcripts thematically. The coding was focused on isolating the human actions and institutions in the design process. The IAD framework was used for the isolation of the factors. Thereafter, the relationships and interactions between the different human actions and institutions were derived from the records and transcripts by using the structurational model of technology [20].

4 Results: Observed Design Practices

This section discusses the design practices within public organizations following from institutions and human actions. The explorative study identifies four design practices: (1) a narrow focus on technology; (2) disregard of the normative basis for algorithmic systems; (3) designers are unaware of socio-technical components and interactions; and (4) caught in linear design processes. The four practices will be discussed here.

4.1 Narrow Focus on Technology

The silo structure of public organizations results in a narrow focus on technology. The observations and interviews showed that generally three organizational elements – each with its own jargon name – are involved in the design process: the *business*, *ICT*, and *compliance*. The domain-specific departments, which public organizations call "the business", commission a public algorithmic system which they will use in executing their tasks. Next, the algorithmic system is developed by an "ICT" function or department, or by an external party. Finally, "compliance" functions or departments ensure that the algorithmic systems are in line with rules and regulations. Three interactions arise in the horizontal relationship between the three organizational elements.

First, when algorithmic systems are developed internally, the system is developed by ICT or data-analytics departments – which are considered facilitative elements in the organization. In theory, the business should be in the lead in defining the goal and boundaries of a system in a policy. However, they largely depend on the knowledge of ICT to create the information architecture for their executive tasks. The business has to be told about what is technically possible and what is not. Due to this great dependence on technical knowledge, ICT naturally becomes the lead in designing public algorithmic

systems. Often, the business and ICT work together, however, the strong position of ICT creates an imbalance in this collaboration.

Second, the algorithmic system is usually developed by a private or external party, since most public organizations do not have the capacity to develop these systems. Concerning algorithm development an interviewee stated: "It is also the starting point of the government. People say we do not want to disrupt markets. We are the coordinator, we outsource what we can outsource [to external parties]… Well, that is also the direction that the government has taken in recent years. [trans.]" The external party controls the development of the technical artefact. The consortium developed procurement conditions to strengthen the position of public organizations in this dependent relationship. However, the conditions do not change the fact that the public domain is partly designed by private parties and that they make part of the political trade-offs. One interviewee stated: "…when a product is outsourced, there are conditions that a developer needs to adhere to. That also includes … policy goals or policy principles… But, the translation to an algorithm, the technical translation. They fully trust the developer to do the translation properly. [trans.]".

Third, compliance departments that ensure non-technical aspects of systems– such as safety and security, privacy, and ethics – are often viewed as burdensome in the design process. Compliance officers increasingly have a say in developing and using public algorithmic systems, especially privacy officers after the introduction of the GDPR. However, these actors often have few means and a relatively weak position to intervene in the design process. They tend to be informed late, making it harder to stop a project with high sunk costs. One interviewee stated about their position as compliance department: "The tricky thing is, we cannot control the money. The money is at 'the business', we cannot stop them. [trans.]" Another interviewee indicated how this leads to being considered as a burden: "… as compliance you are considered as hindering. Whereas if we would be involved early on, there are lots of possibilities to ensure compliance with frameworks and guidelines. At present, we are considered burdensome at the end of a project and, well, there is a continuous battle between innovation and formulating compliance frameworks and guidelines within the organization. [trans.]".

The three interactions result in a strong and reinforced position for techno-focused actors (e.g., engineers, data-analysts) throughout the design process of public algorithmic systems. Interviewees also provided examples of the leading role of ICT. For example, ICT provides products for several domain departments that are strictly separated. ICT is naturally put in a position in which they become aware of possible connections between different departments. One interviewee stated concerning the interactions between departments and their role as developer: "The product owners [i.e., domain department] mostly do not know each other, they do not actively exchange [about their projects and insights]. Encouraging and facilitating these exchanges, using insights more broadly within the organization, and educating product owners is really our role. [trans.]" Technology seems to become a vehicle for making public organizations work more comprehensive; however, not in an interdisciplinary way. ICT is the leading force – with its specific perspective. The assignment of this role is not a deliberate choice but a gradual and natural process.

The central position of ICT results in an emphasis on the technical artefact in public algorithmic systems. Other departments consider ICT as a facilitative element to which the development of public algorithmic systems is 'thrown over the fence'. In this way, the development attains a sense of neutrality that also disguises the central role of ICT in the design process. Naturally, the technical artefact attains most attention, neglecting institutional and agential components of public algorithmic systems. Compliance departments cannot correct this due to their weak position.

4.2 Disregarding the Normative Basis for Public Algorithmic Systems

Vertical disconnections in public organization's structure and culture impedes addressing normative aspects in the design process. Most of the time, public algorithmic systems must aim to promote public values and not infringe them, but such high-level statements leave room for interpretation. Civil servants seem to specify these norms in an ad hoc manner.

Missing channels of vertical communication create distance between the political decision-making level and the operational levels. Various interviewees stated that public administration is traditionally well-equipped to formulate policies and attune them to the political debate; this is a familiar practice within public organizations. However, the translation of the work done on public algorithmic systems – on the operational level –to the political debate on the decision-making level often falls short. The awareness of political trade-offs in public algorithmic systems and the feeling of agency to make these trade-offs is often low at the operational level. For example, the consortium often discussed the need for using values in the design process; however, they did not acknowledge that those values can be in conflict.

The lack of vertical communication also follows from a lack of clarity, concreteness, and direction on normativity given by the decision-making level. This seems to emerge from the (political) discourse associated to public algorithmic systems, which is still based on the premise that technology is neutral (see also [13, 25]). AI is still seen as a technology with great benefits for public organizations – although it is unclear what the actual affordances are. The limitations of the technology are increasingly recognized, but tend to be considered as something that can be resolved by developers. Moreover, the political debate is mostly reactive. Decision-makers react to incidents such as the introduction of the GDPR (explaining the emphasis on privacy) and recently the child benefit scandal *(Toeslagenaffaire)* in the Netherlands (resulting in a push for more transparency). Because of this reactive practice, politicians or decision-makers have not provided univocal and holistic normative or evaluative frameworks for public algorithmic systems.

Furthermore, the consortium considered the lack of awareness of algorithmic system use among citizens as one of the flaws in the current public debate. Citizens usually do not know whether and how systems were involved in making decisions that do affect them. To overcome this information asymmetry, the consortium developed an algorithm register. However, when writing this article, it is unclear whether this register will provide meaningful insight into the workings of the systems – including their interaction with policy, laws and regulations – and will support a public debate on algorithms used. Discussions within the consortium showed that the register can be used for a wide variety

of goals and publics, for example, as internal archive for algorithms or as repository for citizens that want to appeal decisions. Furthermore, the effectiveness of the register depends on the status it is given by the public organization. The final form of the register will determine its support to the public debate.

The normative flaw in the design process makes it unclear for designers what values are applicable for specific public algorithmic systems and how trade-offs between these values have to be made. The lack of a public debate also hinders the formulation of such a normative framework. Currently, the design process is insufficiently embedded in political processes and deliberation about design choices is missing as these choices are not acknowledge as political choices. As engineers make these design choices, the narrow focus on technology is strengthened even more.

4.3 Designers are Unaware of Socio-technical Components and Interactions

All interviewees pointed out the importance of involved actors' approaches to the design process. Interviewees argued that, in the current situation, a project is approached more comprehensive or interdisciplinary when designers are involved that are aware of the socio-technics of the systems. One interviewee explained the dynamics within a project in which the awareness was high: "The specific people involved in this project were already thinking about these [socio-technical] themes. So, not just developing an application, but to think, okay, why do I program these choices in this model? [trans.]".

However, most employees in public organizations have little awareness of the different socio-technical components (and their interactions) of public algorithmic systems. They consider algorithmic systems as mere automation: a simple and objective tool for executive tasks. They are unaware of the fact that a public algorithmic system is an integral part of policies and regulations. In other words, they are unfamiliar with the fact that an algorithm can also influence the interpretation of those policies and regulations. Furthermore, the lack of awareness was observed in the difficulty of demarcating system boundaries. Interviewees stated that product owners (i.e., the business) are not always aware what is affected by an application and that they often do not recognize the products they use as algorithmic systems.

Whereas the first two design practices show the influence of institutions on the emergence of a focus on the technical artefact, this third design practice shows the importance of human actions in designing or using the system. The narrow focus on technology and lack of normative debate will not be resolved if the awareness of socio-technical components of public algorithmic systems is not increased. Increasing awareness is possible. Interviewees gave examples of regulations (e.g., GDPR), political events and, incidents that increased the urgency to deal with privacy and transparency. However, these are also exemplary for the reactive practice discussed in Sect. 4.2.

4.4 Caught in a Linear Design Process

The final design practice relates to how public organization standardize the design process. The common linear conception of design processes conflicts with the idiosyncratic nature of those processes; public organizations consider and structure design processes chronologically or sequentially (cf. waterfall). Governance documents or procedures

often contain schemes with one-way arrows between discrete design process steps. Furthermore, in discussing the practical reorganization of design processes, the consortium tended to abstract the process to a series of subsequent steps.

Similarly, public organizations experience difficulties because they make sharp distinctions between the development and the use of algorithmic systems. The responsibility for design and use tends to be assigned to different teams and functions. However, the consortium often discussed the difficulties of transferring systems from the development phase to the use phase. Organizations try to overcome this divide between the two phases by, for example, letting designers interact with street-level bureaucrats that will use the system. However, these interactions are mostly about training the users instead of involving practice and practitioners in earlier design stages.

Although most actors still considered the design process to be linear, a shift in thinking could be observed in the consortium and its participating organizations. Two municipalities were experimenting with new design practices. One municipality developed complex algorithmic systems in an iterative and lean fashion, so-called agile-scrum sprints. The other municipality experimented with interdisciplinary teams in which all important actors were involved from the start of the design process. Nevertheless, in talking about a governance framework, the consortium often came to the conclusion that a linear abstraction of processes is needed to make it "workable". One interviewee stated that their team is constantly asked for new iterations of developed or implemented systems. In that case, the "business" asks to add new (technical) features to the system. In other words, implicitly there are already some iterative processes occurring; however, this brings difficulties. The iterations are mostly approached from a technical perspective and an overview of what iterations are commonly made within a project is absent. These ad-hoc iterations may obscure a clear overview of the project. For example, the goal of the algorithmic system may be lost out of sight.

More importantly, the design process misses correction and detection mechanisms of flaws in algorithmic systems. Public organizations are familiar with correction mechanisms, such as due process and objection procedures, or court cases. The consortium also tried to improve these procedures by enhancing the information position of civil servants that need to handle the objection to a decision made by or with the support of algorithms. However, traditional correction mechanisms mostly focus on individual cases. Failures ingrained in the algorithmic systems may be corrected for an individual, but there is no process yet to determine whether an individual incorrect or undesired outcome necessitates redesign on a system-level. Such system-level corrections also require detection of failures or changes in a system. The consortium was often confronted with the need to define a substantial or significant change in a public algorithmic systems. For example, determining the magnitude of changes in a system that must result in the alteration of a description in the algorithm register. However, the consortium was unable to operationalize a substantial change in algorithmic systems.

Public organizations have created design processes that do not align with how public algorithmic systems are developed in reality. This is mostly because the starting point of the design process – i.e., linearity – conflicts with the nature of development – i.e., iterative. Public organizations lack the swiftness to react to changes in systems and their perspective focuses on the internal organization – external forces changing the system

are overlooked. Moreover, it is hard for public organizations to change their ways of designing, as it means that other processes within the organization have to be changed as well. Changing the design process is a complex operation.

As a final observation, we examine the efforts of the consortium to enhance the design process using their four policy instruments. Their approach focused on instruments has its advantages. For example, the instruments helped to raise awareness within public organizations for the need for transparency in using algorithmic systems. However, the instruments have their limitations. The instruments clearly show that the current organization of the design process is the starting point for reforms. The consortium sustains a fragmented and narrow perspective on designing public algorithmic systems – e.g., singular impact assessments for different effects or an algorithm register that only gives insight what type of system is used. The instruments do not show a deliberate approach to deal with structuration – i.e., they ignore the interactions between the socio-technical components and apply instruments to very specific aspects of the design process.

5 Discussion

The four design practices show that institutions and human actions in the design process have distinctive influence on the development of public algorithmic systems. In this section, we extrapolate those design practices to the emergence of digital cages. The four practices indicate (1) that institutions and human actions in design processes need to be considered when examining the emergence of digital cages and their unintended consequences. We will elaborate on the implications of all four practices.

First, the narrow focus on technology obstructs considering the role of other socio-technical components in the emergence of digital cages. Section 4.1 shows that technology and engineers – although considered facilitative – have gained a strong position in the design process, reinforcing the false assumption that algorithms are neutral or objective. Similar to practice, research focuses on technology in automated decision-making [28]. The pivotal role of technology in design processes results in a focus on formalizing processes in public administration through algorithms (with emphasis on, for example, optimizing accuracy). This formalization is the, now strengthened, driving force behind digital cages. By prioritizing the technical artefact, the possibilities of public organizations to address digital cages and their unintended consequences are reduced; designers ignore institutions and human actions as ways to address digital cages. For example, institutions that protect the discretion of civil servants, tend to be unexplored by public organizations. Actors that could provide those perspectives lack the means to intervene.

Second, the unclear normative basis inhibits anticipating to and assessing possible unintended consequences within the design process. The digital cage is inherently normative, as it emerges from political and ideological choices about the level of formalization, the relationship between citizen and government, and the role of technology in executing public services. Dencik et al. [4] show how austerity policies have resulted in automated assessments of citizens based on risks; thereby shifting responsibilities to citizens and obstructing the engagement in scrutinizing government decisions. The flaws in the public debate could result in difficulties at the side of both citizens as the public organization. Since citizens are unable to participate in the debate, several perspectives,

needs, and interests are currently not involved in the design process. These perspectives, when included, could provide crucial or critical insights on the emergence of specific unintended consequences. Similarly, the lack of public debate can also impede deliberation on the assessment of unintended consequences. In the design process, (un)intended consequences of public algorithmic systems should be assessed on their impact and whether and how this impact should be prevented or mitigated. Finally, there is the possibility that a digital cage is intentionally or strategically implemented, making the consequences less unintended. The lack of public debate may result in withholding information about the rationales behind public algorithmic systems and digital cages.

Third, since a narrow focus on technology is undesirable, designers need to have awareness of socio-technical interactions within public algorithmic systems. When public organizations do not acknowledge that their public algorithmic systems can evolve into digital cages, they will not feel the urge to (re)design their systems to tackle the unintended consequences of digital cages. Similarly, designers that are unaware of the socio-technical nature of their systems and the digital cage, are unaware of the influence of human actions and institutions on digital cages and therefore cannot adequately address their unintended consequences. Furthermore, the lack of awareness may result in resistance to compliance activities in the design process. The first design practice already showed a need for situating systems in their social and institutional context. However, the importance of awareness shows that it is not only about changing the organizational structure. The culture or shared meanings within organizations, formed by the accumulation of human actions, also has to change.

Finally, organizing the design process linearly conflicts with the emergent nature of digital cages. In a linear process, unintended consequences will quickly appear for those who are harmed. However, those that need to correct the consequences will only acknowledge the detrimental impact of digital cages in late stages of developing and using the algorithmic system. For example, correction mechanisms are based on individual cases and therefore do not provide insight in more systemic failures or do not provide citizens with a voice to reveal problems in public algorithmic systems [33]. More sensitivity and response to the emergence of material harms for citizens is needed. Those should be picked up and acknowledged earlier and lead to corrective – i.e., redesign – and compensating actions – to alleviate harm. Public organizations are familiar with that sensitivity in or control on their processes, because citizens – the locus in public administration – need protection. However, guaranteeing the generation of public value and the possibility for public control on the process often conflicts with facilitating an innovative environment – i.e., providing flexibility and room for experimentation [16]. That flexibility is also needed when dealing with unintended consequences, as they cannot always be predicted or can be overlooked. Still, this does not necessarily mean that one cannot have control over the design process. One could, for example, think of dissent channels to give citizens a voice [5]. This asks for new institutional environments for the design process.

Our observations provide only few indications that public organizations are changing their design practices to address digital cages more effectively. We observed that instruments sustain the current rationale behind design processes. For example, public

organizations obtain broader perspectives, but only on specific topics such as transparency. Moreover, policy makers clearly consider the political debate as an activity for politicians, but they also do not provide politicians with possibilities to enhance the communication between politicians and designer. Finally, the instruments are based on the common linear way of thinking about design. Still, the instruments are useful tools to create awareness about socio-technics within public organizations.

However, this explorative research to identify design practices contributing to the emergence of digital cages in public algorithmic systems has its limitations. Only a select group of public organizations was observed and interviewed. And although the organizations represented a diverse set of public organizations, they were all situated in the Netherlands and therefore in Dutch governmental practices. Furthermore, the participants of the consortium were already thinking about how to address harms of public algorithmic systems. They may not fully represent public organizations that lack the capacity to contemplate on their use of algorithmic systems – for reasons of, for example, being too small (e.g., small municipalities that outsource algorithm development). Finally, the research focused on discussions of policy-makers about the design process; actual design processes were not observed.

Hence, more research on the organization of the design process is needed. First, public organizations should obtain a broader perspective in the design process, but it is poorly understood how that can be achieved. Socio-technical design research provides some substantive starting points, such as making the design process integral [2], interdisciplinary [3], political [2], and iterative [20]; but it is unclear what these concepts entail for the design process of public algorithmic systems. Similarly, public organizations are faced with the challenge to change their organizational culture. Research could show how the deterministic and objectivistic perspective on technology can be replaced by a broader and situated view on technology. Finally, the need for a broader perspective also implies that public organizations have to reflect on their dependency on external developers. Therefore, more research is needed on the extent to which external parties make or influence critical design choices for the public domain.

Second, the need for flexibility in public design processes justifies more research. Studies on achieving flexibility could use unintended consequences as a starting point and examine processes to anticipate and react to these consequences. In this respect, an interesting avenue for research could be to examine the extent to which public organizations validate their algorithmic systems and what the value of validation could be. Next, research could show how flexibility can be combined with public control on design processes and the implementation of checks and balances in the design process.

Third, the role of the designer in the emergence of digital cages needs to be analyzed alongside the role of the designer. This research mostly focused on designers, but, users' actions are as important in the emergence of digital cages. For example, discretion is not only curtailed by the technical artefact, it is also affected by how users interact with or react to technology. More insight is needed in the role of users in the development of digital cages. This research seems to indicate that considering the role of the users has to go further than user experience, user participation, or user training.

Finally, more research is needed on integrating public debates in design processes and vice versa. Regarding this topic, most literature points towards participation. However,

there is much more to it (see, for example, [9]). The public debate comprises more complex forms of deliberation and democracy. Moreover, normativity asks for mapping the values that are touched upon in digital cages and how trade-offs between these values can be made in a democratic way.

6 Conclusions

The increasing use of public algorithmic systems can reinforce the emergence of digital cages. Although digital cages are familiar symptoms of algorithms used in public administration, there is a lack of knowledge in the design practices that result in the emergence of digital cages. Therefore, the goal of this paper was to provide insight in these design practices in order to support the prevention, mitigation, or correction of detrimental impact, following from digital cages, on citizens.

The diagnosis of the design process, presented in this article, shows that the current organization of design processes in public organizations ignores the fact that specific human actions and institutions contribute to the emergence of digital cages. Four practices that impede addressing digital cages were identified. Firstly, design processes emphasize technical aspects and effects of public algorithmic systems. Thereby, causes for digital cages in institutional or social parts of these systems are ignored. Secondly, politics are hidden in the design process and the discourse on public algorithmic systems remains narrow. Public deliberation on the normative impact of digital cages is missing. Thirdly, the consideration of socio-technics in design processes heavily depends on the awareness of designers. Finally, public organizations struggle with facilitating corrective iterations during a system's life-cycle. Design processes do not have the flexibility to deal with the complexity of a socio-technical design process and are only partly able to adequately react to unintended consequences.

In order to address digital cages, the design process of public algorithmic systems needs to be altered. However, public organization public organizations only show little attempts to reflect on their design practices. The instruments they implement to rearrange design processes continue the current organization of those processes. In other words, the techno-centric and sequential way of thinking is maintained, and the publicness and the impact on citizens of algorithmic systems is still neglected. However, the interviewees also showed the motivation to start such reflections. In order to support public organizations in their reflection on design processes, more research is needed to practices of anticipation and reaction to unintended consequences, the implementation of flexibility and checks and balances in the design process, and how design processes can become more democratic.

References

1. Burrell, J.: How the machine 'thinks': understanding opacity in machine learning algorithms. Big Data Soc. **3**(1), 1–12 (2016)
2. Clegg, C.W.: Sociotechnical principles for system design. Appl. Ergon. **31**, 463–477 (2000)
3. De Bruijn, H., Herder, P.M.: System and actor perspectives on sociotechnical systems. IEEE Trans. Syst. Man Cybern. – Part A: Syst. Hum. **39**(5), 981–992 (2009)

4. Dencik, l., Redden, J., Hintz, A., Warne, H.: The 'golden view': data-driven governance in the scoring society. Internet Policy Rev. **8**(2), 1–24 (2019)
5. Dobbe, R., Krendl Gilber, T., Mintz, Y.: Hard choices in artificial intelligence. Artif. Intell. **300**, 103555 (2021)
6. Giddens, A.: New Rules of Sociological Method. Basic Books, New York (1976)
7. Giddens, A.: Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis. University of California Press, Berkeley (1979)
8. Giddens, A.: The Constitution of Society: Outline of the Theory of Structure. University of California Press, Berkeley (1984)
9. Himmelreich, J.: Against "Democratizing AI". AI Soc. (2022). https://doi.org/10.1007/s00146-021-01357-z
10. Hildebrandt, M.: Law for Computer Scientists. Oxford University Press, Oxford (2019)
11. Hodson, R., Martin, A.W., Lopez, S.H., Roscigno, V.J.: Rules don't apply: Kafka's insights on bureaucracy. Organization **20**(2), 256–278 (2012)
12. Hodgson, G.H.: What are institutions? J. Econ. Issues **40**(1), 1–25 (2006)
13. Janssen, M., Kuk, G.: The challenges and limits of big data algorithms in technocratic governance. Gov. Inf. Q. **33**, 371–377 (2016)
14. Kroll, J.A.: The fallacy of inscrutability. Philos. Trans. A **376**, 20180084 (2018)
15. Lorenz, L., Meijer, A., Schuppan, T.: The algocracy as a new ideal type for government organizations: predictive policing in Berlin as an empirical case. Inf. Polity **26**, 71–86 (2021)
16. Meijer, A., Thaens, M.: The dark side of public innovation. Public Perform. Manag. Rev. **44**(1), 136–154 (2021)
17. Mulligan, D.K., Bamberger, K.A.: Saving governance-by-design. Calif. Law Rev. **106**(3), 697–784 (2018)
18. Mulligan, D.K., Bamberger, K.A.: Procurement as policy: administrative process for machine learning. Berkeley Technol. Law J. **34**(3), 773–851 (2019)
19. OECD: Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449 (2021)
20. Orlikowski, W.J.: The duality of technology: rethinking the concept of technology in organisations. Organ. Sci. **3**(3), 398–427 (1992)
21. Ostrom, E.: Understanding Institutional Diversity. Princeton University Press, Princeton (2005)
22. Peeters, R.: The agency of algorithms: understanding human-algorithm interaction in administrative decision-making. Inf. Polity **25**, 507–522 (2020)
23. Peeters, R., Widlak, A.: The digital cage: administrative exclusion through information architecture – the case of Dutch civil registry's master data management system. Gov. Inf. Q. **35**, 175–183 (2018)
24. Ranchordas, S., Scarcella, L.: Automated government for vulnerable citizens. William Mary Bill Rights J. (2021)
25. Van der Voort, H.G., Klievink, A.J., Arnaboldi, M., Meijer, A.J.: Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? Gov. Inf. Q. **36**, 27–38 (2019)
26. Van Schie, G., Smit, A., Coombs, N.L.: Racing through the Dutch governmental data assemblage: a postcolonial data studies approach. Global Perspect. **1**(1) (2020)
27. Van Zoonen, L.: Data governance and citizen participation in the digital welfare state. Data Policy **2**(10), 1–17 (2020)
28. Vydra, S., Klievink, B.: Techno-optimism and policy-pessimism in the public sector big data. Gov. Inf. Q. **36**, 101383 (2019)
29. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harvard J. Law Technol. **31**(2), 841–887 (2018)

30. Wachter, S., Mittelstadt, B., Russell, C.: Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI. Comput. Law Secur. Rev. **41**, 105567 (2020)
31. Wagner, B.: Liable, but not in control? ensuring meaningful human agency in automated decision-making systems. Policy Internet **11**(1), 104–122 (2019)
32. Weber, M.: Economy and Society. University of California Press, Berkeley (1978). Translated by G. Roth and C. Wittich, Original work published in 1925
33. Widlak, A., Peeters, R.: Administrative errors and the burden of correction and consequence: how information technology exacerbates the consequences of bureaucratic mistakes for citizens. Int. J. Electron. Gov. **12**(1), 40–56 (2020)
34. Zouridis, S., Bovens, M., Van Eck, M.: Digital discretion. In: Evans, T., Hupe, P. (eds.) Discretion and the Quest for Controlled Freedom. Palgrave/MacMillan (2019)
35. Zuiderwijk, A., Chen, Y., Salem, F.: Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda. Gov. Inf. Q. **38**(3), 101577 (2021)
36. Zuurmond, A.: From bureaucracy to infocracy: are democratic institutions lagging behind? In: Snellen, I.T.M., van de Donk, W.B.H.J. (eds.) Public Administration in an Information Age: A Handbook. IOS Press, Amsterdam (1998)