



---

# Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes

Lena Enqvist

---

## Abstract

This chapter explores aspects of the relationship between technology, transparency, and accountability in public decision-making. It addresses how technological advancements have increased accessibility and automation while complicating decision process reviewability. It explores transparency as a relational concept and focuses on legal obligations on documentation and records-keeping, such as in the EU General Data Protection Regulation and the upcoming EU Artificial Intelligence Act, as a means to bolster transparency and improve reviewability. In particular it also discusses the feasibility of gathering and analysing ‘paradata’—data pertaining to data processes—as a means to safeguard legality and transparency in automated decision-making, notably within the public sphere.

---

## 1 Introduction

Although various public sector bodies around the globe have been using technologies to assist their tasks and decision-making for decades, such uses have advanced and intensified greatly especially in the last decennary. While these developments hold many promises of highly serviceable and efficient public sectors, concerns have also been raised over the risks of placing too much trust in the technologies’ capacities to produce reasoned recommendations or decisions that aligns with the law. Real-life examples, such as the Australian government’s use of the so-

---

L. Enqvist (✉)  
Department of Law, Umeå University, Umeå, Sweden  
e-mail: [lena.enqvist@umu.se](mailto:lena.enqvist@umu.se)

© The Author(s) 2024  
I. Huvila et al. (eds.), *Perspectives on Paradata*,  
Knowledge Management and Organizational Learning 13,  
[https://doi.org/10.1007/978-3-031-53946-6\\_13](https://doi.org/10.1007/978-3-031-53946-6_13)

233

called RoboDebt system, illustrate how flawed system designs or inappropriate applications can have effects on the legality of the public exercise of power at a large scale. The system was found to have miscalculated hundreds of thousands welfare recipients' incomes and related rights to benefits—and as a consequence automatically issued a vast number of faulty debt collection decisions to citizens often part of socially vulnerable groups (Carney, 2019). Examples like these have contributed to intensified political as well as academic discussions on the effects of technologies on the public exercise of power. At the core of these discussions lies the question of how to ensure transparency and accountability when public power is exercised via technological proxies. That governments are transparent in their exercise of power against the public is, namely, a foundational principle of the 'Rule of law' (Jamar, 2001). Otherwise, the prospect to review whether public powers have been exercised within their limits is hampered. Transparency is, however, not a fixed concept and relates to other similar concepts such as openness, explainability, interpretability, accessibility, visibility, and reason-giving (Felzmann et al., 2019a). There is no obvious or infallible solution to ensuring and safeguarding transparency. This contribution will, however, focus on public authorities' use of technology to assist their decision-making as it has proven to be principally challenging from a public transparency perspective. It will also, in particular, discuss the collection and use of 'paradata' as one possible and advantageous tool and building block of transparency in this context. The formalisation of data on processes that the collection of paradata implies may, namely, prove useful for enabling qualitative reviews of whether automated systems are operating lawfully.

Before continuing to outline this contribution, some introductory definitions are in place. Firstly, 'technologically assisted decision-making' is here used as a broad term including any use of technologies by public authorities to prepare, recommend, or make decisions. This means that fully as well as partially automated decision-making procedures are included. Secondly, 'paradata' is not a legal concept and therefore, naturally, also lacks a legal definition. This contribution will use a wide definition as *any* 'data-on-data related processes and practices' extending beyond its original survey domain (Couper, 2017). Importantly, the definition includes fixed design decisions on system processes, as well as data on how these design decisions have been employed in particular applications. As relevant to the context of public decision-making, 'paradata' here also includes descriptions of the procedural aspects of how a system is designed to run, its authorisation and constraints. Narrow distinctions from neighbouring or partly overlapping concepts such as provenance metadata or contextual metadata, for example, will not be made (Bentkowska-Kafel et al., 2012; Reilly et al., 2021).

The contribution will have the following structure. First, Sect. 2 places the 'law' and legal practice in an information and knowledge management context, and link to how technological developments in legal information and knowledge management have made 'the law' more accessible and automatable at the same time as it has obscured parts of the decision-making procedure and affected its reviewability. Section 3 deepens the analysis by focusing on the general requirements for transparency in public decision-making and combines it with an argument for an increased

need to analyse data-on-data related processes, ‘paradata’, as part of ensuring that automated decision-making processes are lawful and transparent. Section 4 then discusses how requirements on documentation and recordkeeping of data-related processes can contribute to increased qualitative transparency in connection with public automated decision-making, focusing on the GDPR and EU’s upcoming Artificial Intelligence Act as examples. In Sect. 5, the merits of legal standards for documentation and recordkeeping on data-related processes are discussed as one important affordance for the utilisation of such data in legal analysis. The chapter is finally concluded through Sect. 6, which discusses the potential benefits and challenges to utilising paradata analysis within the legal domain.

---

## 2 Legal Knowledge Management as the Nexus of Legal Practice

Law is a knowledge-based profession and its core, ‘legal practice’, is about providing specialised knowledge, expedited through expert services and the exercise of power (du Plessis & du Toit, 2006). More specifically, legal knowledge concerns the law and its application and is used to produce and manage legal work. Therefore, legal research is and has always been central to any legal practitioner or scholar to find solutions to particular legal questions. This is often a time-consuming task, as legal knowledge is acquired from the internalising of information gathered during legal studies, legal research and legal experience. The primary sources include statutes, preparatory works and case law, etcetera, and the secondary sources include legal reference works, digests, indexes, law reviews, legal periodicals, commentaries, books, and articles from specialised law publications (Roos et al., 1997). Unsurprisingly, functional legal information management is thus imperative to the acquisition and internalising of legal knowledge.

While legal knowledge management traditionally has been intimately tied to human carriers/intermediaries—technologically driven transformations have partially challenged this premise. Early legal information and knowledge management research was primarily concerned with libraries and their roles in aiding legal research through structuring legal information carriers such as statutes, case law or academic writing, etcetera. The introduction of new technologies shifted much of this focus to search engines and the build-up of legal databases to aid the work flows for those performing legal research (Berring, 1994; Foster & Kennedy, 2000). Such systems have functioned as technological drivers for a transformation in the methods that lawyers use to access, retrieve, and process information in order to solve legal problems (du Plessis & du Toit, 2006; Merwe, 1986; Susskind, 2000). The expected promises of the technologically mediated legal information retrieval have been high and span all the way to discussions on whether the new potential efficacy of legal information management might even render lawyers and their tacit knowledge superfluous in some cases (Davis, 2020; Susskind & Susskind, 2016). Because when technology is sufficiently capable of imitating a ‘cognisance’ of the

law, there is less need for human intermediaries to translate the law into directives on how to act.

In relation to human decision-makers, automated processes are intended to ‘embody the specialised knowledge and experience of a human expert in a chosen domain’ and provide a ‘mechanism for applying this knowledge to solve problems in that domain’ (Kidd, 1985). The aim is, thus, to augment human decision-making through knowledge management. Focusing on automated decision-making or recommender systems, they are more advanced in their operations than performing mere ‘legal research’. Their aim is not just to assist in determining what the law is, but also to determine how the law applies in a given situation. They therefore need to combine the (identified as) relevant set of rules with case specific data input, to produce a case specific data output in the form of an individualised decision or recommendation.

A growing credence to automated procedures is signalled through a nearly world-wide general tendency towards national public administrations deploying different types of technology (ranging from simpler pre-programmed ‘if-then’ statements to more dynamic AI and machine learning applications) to make or support their decision-making. Underlying assumptions are that automation ‘done right’ will help streamlining decision-making procedures, reducing the need for tacit knowledge, risks of skill-inadequacies and human bias. The expectations also include that automation will be able to provide a more accurate as well as speedy justice for those subject to the public exercise of power, as well as a more cost-efficient administration. Automated decision or recommendation systems are, thus, seen as media for knowledge management to scale up the capacity and effectiveness of knowledge distribution. On the other hand, there are also associated risks. The assuring that public automated processes work satisfactory and lawfully will require that human intermediaries exercise oversight and control of their functioning. Knowledge of how the system operates, and why, is crucial. What information governance regimes are in place thus plays an important role in elevating how the delivery of knowledge occurs to those tasked with overseeing the proper functioning of automated processes. The next section will, therefore, discuss requirements of transparency in relation to automated decision-making, and the possible utilisation of ‘paradata’ in this context.

---

### **3 Transparency in Public Decision-Making and the Growing Need for Analysis of Data-Related Processes**

As put by Oswald, algorithmic decision-making may come with the risk of creating substantial or genuine doubt as to why decisions were made and what conclusions were reached (Oswald, 2018). To mitigate these risks, the assisting technologies must serve several and sometimes counterbalancing objectives at the same time. They must not only aid the more obvious aims such as correct and more efficient decision-making procedures. The technologies must also serve several other legal (and ‘rule of law’) values such as the supremacy of law, equality before the law,

accountability to the law, fairness, and legal certainty (Zalnieriute et al., 2019). And, as already introduced, they must also secure sufficient transparency to enable scrutiny of the public exercise of power.

The basic idea is that transparency increases the chances to detect wrongdoings, uncover abuses of power, and scrutinise public activities (Matheus et al., 2021). In this respect, transparency is foremost a supporting value to the realisation of other pertinent values—and essential to establishing trust in the public administration. As put by Jamar, transparency refers to a cluster of related ideas, including governmental action in the open, the availability of information (particularly relating to the law), as well as accuracy and clarity of information (Jamar, 2001). There is no common or comprehensive definition of transparency in the legal sense. Focusing on the aim of transparency, Mock's definition is, however, a useful starting point:

Transparency is a measure of the degree to which the existence, content, or meaning of a law, regulation, action, process, or condition is ascertainable or understandable by a party with reason to be interested in that law, regulation, action, process, or condition. (Mock, 1999, p. 1082)

Notably, this definition expands from a purely 'informational' perspective on transparency (where open access to data would equal transparency)—into a 'relational' one, which takes the recipient's end into consideration (Felzmann et al., 2020). Transparency is thus not only understood as a quality of being open and overt, but also as a quality of being identifiable and understandable. As the latter aspects indeed depend on the recipient's knowledge base and need for awareness, a key question is for whom the automated decision-making processes are supposed to be transparent (Larsson & Heintz, 2020). This topic is discussed and debated as a matter of achieving 'meaningfully' transparent decision-making (Edwards & Veale, 2017; Felzmann et al., 2019b), where the discussions hook into relating concepts such as 'explainability' (Deeks, 2019) or 'reason-giving' (Ng et al., 2020). Creating and maintaining the transparency of technologically assisted decision-making is therefore, indeed, a process in itself that requires the repeated consideration of the recipients' end of process or data-oriented information.

Complicating the matter is also the fact that not all manifestations of 'transparency' are helpful for the cause of fair and lawful public decision-making. Full transparency into the processes of an algorithmic system may disservice its reviewability by overloading the receiver with information that at least partially requires special expert competence to decode and interpret. The recipient, or overseer, needs to be able to quickly translate the data into knowledge that is useful for identifying whether the system is somehow flawed, and whether a decision or a recommendation is lawful or not. Because even if there were full openness regarding the input data as well as regarding the algorithmic method used, it is primarily the interplay between the two that yields the complexity—and thus opacity (Burrell, 2016).

Transparency can also have negative effects on other legitimate objectives in public decision-making. The limited human control of automated systems makes them susceptible to risks if they are 'too' understandable, as this might open the door

to misuse by stakeholders trying to ‘game’ the system. Particular output objectives of transparency, such as the possibility to identify and correct biases embedded in, or reproduced by, an automated system might contrast with privacy protections (Independent High-Level Expert Group on Artificial Intelligence, 2018; Larsson & Heintz, 2020). Moreover, there are of course also limits to the technical or economic feasibility of providing extensive transparency, as well as limits posed by obligations or wishes to protect intellectual property, trade secrets, national security and defence, as well as public security.

Transparency in relation to algorithmic decision-making is thus complicated. Even so, the obscuring effect that automated processes have in relation to public decision-making makes it clear that the mere disclosure of a system’s in- and output is not enough. The fact that the legal rules themselves are public and published is not sufficient to ensure a transparent handling of the input data, as the automated processes will not necessarily interpret or utilise this data in a way that replicates legal reasoning. A focus on the strictly informational aspect of transparency is therefore not enough to ensure efficient scrutiny of public automated or automatically supported decision-making.

Attending also to the relational aspect of transparency requires that the knowledge representation and problem-solving processes employed by the system are readily intelligible to the user. Only if this is true will the user both be able to interact competently and efficiently with the system during its reasoning process, and also be confident in the system’s reasoning and advice (Jamar, 2001). And only then can the lawfulness of a decision or recommendation be efficiently or substantively evaluated.

It is now time to bridge the discussion on transparency as an overarching legal value comprising benefits as well as risks to the exercise of fair and lawful public decision-making, over to the utilisation of ‘paradata’ in this context. Here, the argument is simple enough—that the collection of ‘paradata’ could, and should, be emphasised as a pro-transparency measure in relation to automated decision-making processes. Data on the data-related processes through which the system works, including on how data are collected and interpreted, is highly relevant for making sure that automated systems produce decisions or recommendations that are in accordance with the law. Analysis of ‘paradata’ could, for example, help answering important questions like: What processes are in place for the system to retrieve data (including what sources these data are collected from)? What processes does the system use to evaluate whether the collected data are accurate and sufficient to inform a decision, as well as whether further investigation is needed? Are there established feedback mechanisms in place, and how are they designed to work? Is the system equipped with precautionary security measures, such as set procedures for when to interrupt a decision-making process and when it is to be handed over for human review? Did these specific processes run in a particular case of using the system, and how did these different processes combine or feed into each other?

As indicated by the example questions above, ‘paradata’ does not equal either the data that is fed into the decision-making or decision support system or the system outputs in the form of decisions or recommendations. The collection and analysis

of ‘paradata’ alone could therefore not answer, from the perspective of lawfulness, important questions such as if a particular recommendation or decision is legally compliant—unless the data reveals instructions on the system’s running-processes that are contrary to law (such as if the system takes legally irrelevant data into consideration). As ‘paradata’ is data-on-data related processes, its primary function in the context of transparent and scrutable public decision-making is that such data enables the taking of additional factors, other than the current representations of the input data, into consideration. In relation to legal analysis, ‘paradata’ is therefore primarily an auxiliary explanation tool that can help to provide context to analyses of whether there is lawful congruity between a system’s in- and output.

Now, although useful as a tool for analysis, the collecting of ‘paradata’ is not necessarily a straight-forward task. It might be that such data are only readily available in the form of system code, illegible or overwhelmingly technical and detailed to most. From the perspective that transparency is not just about the technical or practical availability of data, one could claim that consideration of the relational aspect of transparency necessitates a certain level of active control of what data are collected (selection) and how it is presented (information design). This makes regulated documentation standards and their design particularly interesting from a public transparency perspective. Not only because such regulated obligations make data retrievable, but also because they provide and give expression to modes of governance on how information on a system’s functioning is to be presented. As we will see in the next section, we can also glimpse a tendency towards specified and increased requirements to document data-on-data related processes.

---

## 4 Examples of Legal Requirements on Documenting and Keeping Records on Data-Related Processes

As introduced, ‘paradata’ is neither a legal concept nor a term that is used in regulatory practice. However, a recognition that the collection and review of this type of data can function as a safeguarding measure or tool in relation to automated decision-making processes can be discerned in some regulation.

One example of a regulation containing certain requirements on documenting and keeping records of data-on-data related processes is the EU General Data Protection Regulation (GDPR), which applies to the vast majority of all processing of personal data taking place within the EU (Article 2 GDPR). Personal data is defined as any information relating to an identified or identifiable natural person (Article 4(1) GDPR). The regulation explicitly requires controllers, meaning the natural or legal person which determines the purposes and means of the processing of personal data, to be able to demonstrate how they ensure compliance with the regulation (Article 5(2) GDPR). When personal data are handled via automated processes, this may include documentation on the processes used to ensure that the data are only processed when there is a lawful basis to do so, as well as the keeping of records on how these processes did in fact run in a particular case (to enable *ex post* review of their proper functioning).

Except from a few narrow exceptions, Article 30 GDPR lays down general and explicit requirements to keep and maintain records of any personal data processing activities. These include the keeping of records, for example, on the categories of recipients to whom the personal data have been or will be disclosed to, and a general description of the technical and organisational security measures. Although they include the documentation of some fixed design choices on the processes by which personal data are to be handled, as well as some records of their actual operations—these express requirements are rather limited regarding data-on-data related processes in particular. To demonstrate compliance the responsible controllers may, however, sometimes need or want to document such data irrespective of whether Article 30 GDPR explicitly requires it or not. Demonstrating compliance with requirements such as keeping the personal data accurate and up-to-date may be too complex without the help of different kinds of processing tools—such as data classification tools, data quality tools or data flow mapping tools to determine data lineage, etcetera (Libal, 2021; Wrobel et al., 2017). To opt for documenting the design and use of such processes or tools can therefore help protect controllers in the event of potential violations (Grow, 2018). The GDPR thus both directly and indirectly places obligations on (foremost) controllers of personal data to collect and document some data-on-data related processes and practices.

One potentially even more wide-reaching example of direct regulation prescribing the documentation and keeping of records on data-related processes within the EU is found in the upcoming EU Artificial Intelligence Act (AIA).

This regulation will, notably, only apply to those automated decision-making procedures that run on AI technology. The most extensive documentation requirements will also only pertain to those AI systems considered at risk of having an adverse impact on people's safety or their fundamental rights (so-called high-risk AI systems). These will most likely be relevant to the bulk of public automated decision-making procedures that are assisted by AI technology, as high-risk systems under the proposal, for example, include any AI system deployed in the areas of access to and enjoyment of essential private services and public services and benefits; law enforcement; migration, asylum and border control management and administration of justice and democratic processes (Article 6 and Annex III AIA).

Any high-risk AI system will be subject to far-reaching technical documentation standards, much focused on the documentation of system processes. A detailed description is not expedient here, but the regulation includes that the provider of a high-risk AI system should provide for documentation of how the AI system will or could interact with external hardware or software, as well as of the system elements and process for its development. The requirements also include the description of the system's general logic—where key design choices such as the rationale and assumptions made, main classification and optimisation choices as well as the relevance of different parameters (etcetera) are to be documented. The same is true for the system architecture explaining how software components build on or feed into each other and integrate into the overall processing, as well as the computational resources used to develop, train, test, and validate the AI system. The regulation will also require that relevant datasheets describing the training

methodologies, techniques, and training data sets used are to be provided. These sheets should include detailed information about the provenance of those data sets, their scope and main characteristics, how they were obtained and selected, any labelling procedures or data cleaning methodologies. Further examples are descriptions of pre-determined changes to the AI system and its performance, detailed information about used validation and testing procedures as well as those relating to monitoring, functioning, and control or risk management. Any changes made to the system through its lifecycle, as well as on the system in place to evaluate the AI system performance are also to be included (Articles 11, 12 and Annex IV AIA).

Notably, all the above-mentioned documentation and recordkeeping requirements pertain to fixed design choices made by the system provider (including by sub-contractors of the provider) that relate to the process operations of the systems. The regulation will, however, also require providers to ensure certain logging capabilities while the system is operating. In contrast to the rather detailed documentation requirements, the required scope and contents of these logging capabilities are not very elaborated. They should, however, ensure a level of traceability of the AI system's functioning throughout its lifecycle (to an extent that is appropriate to the intended purpose of the system). The regulation also states that these logging records should in particular enable the monitoring of certain risks to health or public safety, etcetera, or substantial modifications of the system, and that they should facilitate the 'post-market monitoring' of the system (Article 61 and 3(25) AIA).

As seen, these requirements comprise fixed process design decisions as well as the collection of data on the particular application of these processes, and therefore captures aspects of 'paradata' collection under the definition used in this contribution. One limitation to the regulation is that it is primarily aimed at establishing requirements on providers of AI systems, whereas it is vaguer on extent to which public authorities in the capacity as users, deployers, of high-risk systems is to monitor the systems workings by analysing available data, such as data-on-data related processes. It is clear that system deployers should follow the system instructions and have access to certain information on its functioning (Article 13 AIA). It is also clear that any documentation and recorded data are to be made available to competent supervisory authorities upon request, and that it therefore is meant to facilitate supervisory scrutiny (Article 23 and Recital 46 AIA). In all, it seems that the documentation and recordkeeping requirements of the regulation are primarily geared towards providing for the informational transparency of AI systems, and less around what these data are to be used for and by whom.

It is clear from the upcoming AIA that documentation and the keeping of records—not only on what data these AI systems run on, but also of the data-related processes they premise or perform—has been emphasised. And even if there might remain certain (intentional or inadvertent) gaps in the regulation regarding the utilisation of this data, it is still clear that the overarching aim for the detailed standards, however cumbersome to realise, is to provide an adequate basis for ensuring and monitoring the safe and proper functioning of AI systems.

## 5 Utilising ‘Paradata’ for Increased Transparency of Technologically Assisted Public Decision-Making

Automation carries the potential for delivering speedy and more cost-efficient public decision-making but does not reduce the complexity of the law itself. The responsibilities of public authorities to ensure that any decisions they make are in accordance with the law therefore continues to require intermediaries—which now also have to decipher technical information (despite individual differences their aptness to do so) (Čyras & Lachmayer, 2015; Felzmann et al., 2020). At the same time, new technical tools also complement the legal order by offering new means to monitor the side effects of automated decision-making procedures (Fule & Roddick, 2004; Tamò-Larrieux, 2021). Technology may enable the keeping of larger and different sets of records and at lower costs if compared to manual records. While the importance of keeping records in relation to transparent public decision-making procedures is apparent, the problem is rather how to ensure that value is generated through these records (for example, in the form of better reviewability of the public exercise of power). This requires an understanding of what types of analysis the data is meant to support, that there are measures in place to ensure that the relevant data are collected and presented in a way that is legible to human intermediaries.

Different forms of data documentation (electronic or other) have always been imperative to the legal practice. The same is true for data analysis, as the evaluating and assessment of whether a fact—data or sets of data—is relevant, accurate, and substantiated enough to form the basis of a particular decision lies at the core of legal analysis. In addition, this analysis must also capture whether that particular decision came about in a systematic and formal way following certain procedural requirements—ultimately to safeguard the integral structure of the legal system. In relation to both these aspects of legal analysis, the growing use of automated decision-making procedures have somewhat changed the playing field. Where automated processes aim to assist the application of law, the subsumption of legal facts under legal criteria is accomplished by the system (Čyras & Lachmayer, 2014).

The principal merit of analysing ‘paradata’ in the legal context is that it could help reduce the level of opaqueness and abstraction that these systems display. The interaction taking place between human intermediaries and systems could, however, transpire in relation to different aspects of the system’s workings, as well as be performed by different categories of ‘humans’ with different authorisations as well as knowledge bases. Naturally, the aptness to identify, understand, and make use of relevant data-on-data related processes will also depend on whether the ‘human in the loop’ is a lawyer, data scientist, or other. The potential to reduce the need for highly qualified personnel, as well the potential to reduce mundane and labour-intensive human administration has been one main reason for the growing deployment of automated systems in performing or assisting public decision-making (Tamò-Larrieux, 2021). Overall, it is therefore unrealistic to expect that everyone involved at all stages of a decision-making process would have the

mandate, time, and know-how to utilise any collected ‘paradata’ to the same degree. The information needed to oversee decision-making processes may hold different degrees of granularity depending on the context, the particular user, and the likely weight of the outcome that the system informs (Oswald, 2018).

On the general level, however, knowledge about the processes in place to evaluate what data a decision is to be based on, or data on the actual process elements that made up the particular procedure by which a decision or recommendation was made may, for example, help the assessment of whether a case has been decided on sufficient grounds. And knowledge on the processes by which the input data has been collected and processed may help the assessment of whether there is reason to question the accuracy of that data in relation to a particular decision. Data on the processes in place to trigger safety-measures such as fall-outs to manual administration, or knowledge on the selection profiles that determines the more specific arrangement of particular process elements, could also help the evaluation of whether the process practices align with procedural requirements and thus the law. ‘Paradata’ documentation is, thus, one measure to increase the transparency of a system’s normative features—improving the reviewability of the process as such, as well as of individual decisions.

Having established that ‘paradata’ might be useful for legal analysis, this points to the need for making such data readily accessible and useable to human overseers (with different competencies and at different levels of the decision-making procedure). And this is where the design and scope of recordkeeping standards come in. Cobbe argues that the difficulties associated with understanding, overseeing, or reviewing automated decision-making processes often not only suffer from the opaqueness arising from the meeting between technology and people who lack sufficient technical know-how (illiterate opacity), or from the complexity and difficulty of interpreting the system irrespective of technical know-how (inherent opacity). She argues that these systems might also display a type of ‘unwitting’ opacity stemming from that those responsible for designing, developing, deploying, and using systems simply don’t think to record relevant organisational aspects of the system processes (Cobbe et al., 2021). While it should be stressed that the mere recording of different types of process-related data would not help efficiently overcome the opaqueness of automated decision-making procedures, and that ‘paradata’ documentation requirements certainly is no single or ‘silver bullet’ solution in this respect—they importantly help to mitigate unwitting opaqueness as a first step towards serving the informational aspect of transparency.

From the perspective of legal analysis, one advantage of documentation standards is that they to some extent require that legal and technical knowledge is meshed and presented in a way that better help the knowledge distribution to users of these systems. Requirements on this type of meshing is seen in the AIA draft requirements, as many of the required entries presuppose the active articulation and augmentation of decision-making processes, rather than the mere disclosure of technical system process data. Although any documentation standards represent a type of selection and prioritisation of certain data or information over other, and although this means that they will create proxies through which a more complex

reality of a system's procedure is presented in a comprehensive format—'paradata' documentation and recordkeeping standards is one way to facilitate a common ground for the conversation and conceptions about data-related processes. They may thus also serve the relational aspect of transparency.

---

## 6 Conclusions

In the light of all things discussed, it is relevant to question whether 'paradata' as a particular terminology contributes something specific to the legal domain? Here, the strictly formal answer is simple—it does not. The term is not used in any regulation and does not provide any formal or substantive guidance to the content scope of legal recordkeeping standards, either in the GDPR, the upcoming AIA or elsewhere. It is quite possible to analyse data-on-data related processes, as well as to set requirements which include keeping records on data-related processes, without specifically framing this as 'paradata' analysis or 'paradata' collection.

From the perspective of legal analysis, however, 'paradata' as terminology could serve a pedagogical function in distinguishing different types of data from each other and aid better cognisance of what types of analysis on decision-making procedures that it may support. 'Paradata' is not the input data that is used to feed an automated decision-making procedure. 'Paradata' is also not that type of data that describes and gives information about other data, such as information on when and by whom a certain data was collected (metadata). 'Paradata' is data-on-data related *processes*. It may therefore provide information on the procedural aspects of automated decision-making processes (as particularly relevant in relation to the public exercise of power).

The here used definition of 'paradata' is intentionally broad and includes data on fixed design decisions on system processes, as well as data on how these design decisions have been applied in particular applications (that is, in individual decisions). There are overlaps with similar terminology such as contextual metadata, or statistical and process data, in that they include process-related data. More important than the specific definition or choice of terminology is, however, to point to the functional need to collect and analyse data-on-data related processes. The context-creating merit of attributing certain types of data specifically to 'paradata' lies, at least from the legal analytical point of view, in that it conflates data with procedural properties into a cohesive category of information—around which awareness and knowledge on data-related processes could be more effectively managed.

So, although 'paradata' is neither a fixed term nor a fixed legal concept—and irrespective of whether it will ever permeate into the legal vocabulary—it has a clear utility function in relation to legal knowledge management and the data analysis imperative to ensure that automated decisions or recommendations align with the law. It is evident that the mere keeping of records that include 'paradata' does not solve the problems of opaque decision-making procedures. The clear challenges to utilising 'paradata' in legal analysis are distinct and undeniable. These

include competence issues (the technical knowledge to decipher the data and relate this information to legal requirements). Challenges also include the organisational conditions within the public authorities regarding whom and how oversight is to be performed. Some of these challenges could be addressed by legislative measures that do not relate to documentation or the keeping records. Records are, however, still at the core of the legal infrastructure, and perhaps even more so in the age of technology. As put by Iacovino, recordkeeping lies at the heart of some of the fundamental assumptions of how and why legal systems develop and is not only supported by—but also supports—the practice of the law (Iacovino, 1998). The establishment of recordkeeping regimes that are able to assist the structural and substantive qualities of legal system is therefore a topic deserving of much more in-depth attention in the age of accelerated technological assistance in public decision-making.

**Acknowledgements** This work was supported by the Swedish Research Council under Grant number 2020-02278.

---

## References

- AIA. At the time of writing, the final text of the AIA has not yet been adopted. References to the AIA in this chapter are based on the February 2024 text of the provisional agreement resulting from interinstitutional negotiations between the European Parliament and the EU Council of Ministers. This text outlines the content of the Regulation but may undergo minor, primarily editorial changes before final adoption. Accessed May 15, 2025, from [https://www.europarl.europa.eu/meetdocs/2014\\_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf](https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf).
- Bentkowska-Kafel, A., Denard, H., & Baker, D. (2012). *Paradata and transparency in virtual heritage*. Ashgate.
- Berring, R. C. (1994). Collapse of the structure of the legal research universe: The imperative of digital information. *Washington Law Review*, 69(1), 9–34.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data and Society*, 3(1), 205395171562251. <https://doi.org/10.1177/2053951715622512>
- Carney, T. (2019). Robo-debt illegality: The seven veils of failed guarantees of the rule of law? *Alternative Law Journal*, 44(1), 4–10. <https://doi.org/10.1177/1037969X18815913>
- Cobbe, J., Lee, M. S. A., & Singh, J. (2021). *Reviewable automated decision-making*. ACM. doi:<https://doi.org/10.1145/3442188.3445921>.
- Couper, M. P. (2017). Birth and diffusion of the concept of paradata. *Advances in Social Research*, 2017(18), 14–26.
- Čyras, V., & Lachmayer, F. (2014). Program transparency for legal machines. In *Jusletter IT* (Vol. 2014, pp. 47–54). Presented at the *Transparency, Proceedings of the 17th international legal informatics symposium, IRIS 2014*, 20–22 February 2014, Universität Salzburg, Wein: Österreichische Computer Gesellschaft. <https://doi.org/10.38023/2d0f8625-a63b-411d-93d9-f1262a443811>.
- Čyras, V., & Lachmayer, F. (2015). Towards multidimensional rule visualizations. In M. Araszkievicz, P. Banaś, T. Gizbert-Studnicki, & K. Pleszka (Eds.), *Problems of normativity, rules and rule-following* (pp. 445–455). Springer International Publishing. [https://doi.org/10.1007/978-3-319-09375-8\\_33](https://doi.org/10.1007/978-3-319-09375-8_33)
- Davis, A. E. (2020). The future of law firms (and lawyers) in the age of artificial intelligence. *Revista Direito GV*, 16(1). <https://doi.org/10.1590/2317-6172201945>

- Deeks, A. (2019). The judicial demand for explainable artificial intelligence. *Columbia Law Review*, 119(7), 1829–1850.
- du Plessis, T., & du Toit, A. S. A. (2006). Knowledge management and legal practice. *International Journal of Information Management*, 26(5), 360–371. <https://doi.org/10.1016/j.ijinfomgt.2006.06.003>
- Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for. *Duke Law and Technology Review*, 16(1), 18.
- Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019a). Robots and transparency: The multiple dimensions of transparency in the context of robot technologies. *IEEE Robotics & Automation Magazine*, 26(2), 71–78. Presented at the IEEE Robotics & Automation Magazine. doi:<https://doi.org/10.1109/MRA.2019.2904644>.
- Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019b). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1), 205395171986054. <https://doi.org/10.1177/2053951719860542>
- Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*, 26(6), 3333–3361. <https://doi.org/10.1007/s11948-020-00276-4>
- Foster, L., & Kennedy, B. (2000). Technological developments in legal research the evolution of research. *Journal of Appellate Practice and Process*, 2(2), 275–304.
- Fule, P., & Roddick, J. F. (2004). Detecting privacy and ethical sensitivity in data mining results. In *Proceedings of the 27th Australasian conference on Computer science* (Vol. 26, pp. 159–166). Australian Computer Society,. Accessed 14 September 2022.
- Grow, G. (2018, March 28). *Understanding the role of data quality in GDPR article 5 compliance. Best Data Management Software, Vendors and Data Science Platforms*. Accessed September 14, 2022, from <https://solutionsreview.com/data-management/understanding-the-role-of-data-quality-in-gdpr-article-5-compliance/>
- Iacovino, L. (1998). The nature of the nexus between recordkeeping and the law. *Archives and Manuscripts*, 26(2), 216–246. <https://doi.org/10.3316/informit.294192853020461>
- Independent High-Level Expert Group on Artificial Intelligence (Trans.). (2018). *Ethics Guidelines for Trustworthy AI*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Jamar, S. D. (2001). The human right of access to legal information: Using technology to advance transparency and the rule of law. *Global Jurist*, 1(2), 6. <https://doi.org/10.2202/1535-167X.1032>
- Kidd, A. (1985). Chapter 14 - Human factors problems in the design and use of expert systems. In A. Monk (Ed.), *Fundamentals of human-computer interaction* (pp. 237–247). Elsevier. <https://doi.org/10.1016/B978-0-12-504582-7.50023-8>
- Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, 9(2), 1–16. <https://doi.org/10.14763/2020.2.1469>
- Libal, T. (2021). Towards automated GDPR compliance checking. In *Trustworthy AI - Integrating learning, optimization and reasoning* (pp. 3–19). Springer International Publishing. [https://doi.org/10.1007/978-3-030-73959-1\\_1](https://doi.org/10.1007/978-3-030-73959-1_1)
- Matheus, R., Janssen, M., & Janowski, T. (2021). Design principles for creating digital transparency in government. *Government Information Quarterly*, 38(1), 101550. <https://doi.org/10.1016/j.giq.2020.101550>
- Merwe, D. P. V. D. (1986). *Computers and the Law*. Juta.
- Mock, W. (1999). On the centrality of information law: A rational choice discussion of information law and transparency. *The John Marshall Journal of Computer and Information Law*, 17(4), 1069–1100.
- Ng, Y.-F., O’Sullivan, M., Paterson, M., & Witzleb, N. (2020). Revitalising public law in a technological era: Rights, transparency and administrative justice. *University of New South Wales Law Journal*, 43(3), 1041–1077. <https://doi.org/10.53637/YGTS5583>
- Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: Framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions of*

- the Royal Society of London. Series A: Mathematical, Physical, and Engineering Sciences*, 376(2128), 20170359. <https://doi.org/10.1098/rsta.2017.0359>
- Reilly, P., Callery, S., Dawson, I., & Gant, S. (2021). Provenance illusions and elusive paradata: When archaeology and art/archaeological practice meets the phygital. *Open Archaeology (Berlin, Germany)*, 7(1), 454–481. <https://doi.org/10.1515/opar-2020-0143>
- Roos, J., Roos, G., Dragonetti, N. C., & Edvinsson, L. (1997). *Intellectual capital: Navigating the new business landscape*. Macmillan.
- Susskind, R. E. (2000). *Transforming the law: Essays on technology, justice and the legal marketplace*. University Press.
- Susskind, R., & Susskind, D. (2016, October 11). Technology will replace many doctors, lawyers, and other professionals. *Harvard Business Review*. Accessed September 19, 2022, from <https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals>
- Tamò-Larriex, A. (2021). Decision-making by machines: Is the ‘Law of Everything’ enough? *The Computer Law and Security Report*, 41, 105541. <https://doi.org/10.1016/j.clsr.2021.105541>
- Wrobel, A., Komnata, K., & Rudek, K. (2017). *IBM data governance solutions* (pp. 1–3). IEEE. <https://doi.org/10.1109/BESC.2017.8256387>
- Zalnieriute, M., Moses, L. B., & Williams, G. (2019). The rule of law and automation of government decision-making. *The Modern Law Review*, 82(3), 425–455. <https://doi.org/10.1111/1468-2230.12412>

**Lena Enqvist** is an Assistant Professor of Law at the Department of Law, Umeå University. She is currently engaged in research related to digitalisation and automation in the public sector. This research encompasses inquiries into the necessity, function, and conditions of legal frameworks in fostering transparency and enabling accountable exercise of authority.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

