Disclosure by Design: Designing information disclosures to support meaningful transparency and accountability

Chris Norval
chris.norval@cst.cam.ac.uk
Compliant & Accountable Systems Group
University of Cambridge
Cambridge, UK

Kristin Cornelius
krisbcorn@g.ucla.edu
Informatics Department
University of California
Irvine, USA

Jennifer Cobbe
jennifer.cobbe@cst.cam.ac.uk
Compliant & Accountable Systems Group
University of Cambridge
Cambridge, UK

Jatinder Singh
jatinder.singh@cst.cam.ac.uk
Compliant & Accountable Systems Group
University of Cambridge
Cambridge, UK

ABSTRACT

There is a strong push for organisations to become more transparent and accountable for their undertakings. Towards this, various transparency regimes oblige organisations to **disclose** certain information to relevant stakeholders (individuals, regulators, etc). This information intends to empower and support the monitoring, oversight, scrutiny and challenge of organisational practices. Importantly, however, these disclosures are of limited benefit if they are not *meaningful* for their recipients. Yet, in practice, the disclosures of tech/data-driven organisations are often highly technical, fragmented, and therefore of limited utility to all but experts. This undermines a disclosure's effectiveness, works to disempower, and ultimately hinders broader transparency aims.

This paper argues for a paradigm shift towards reconceptualising disclosures as *'interfaces'* – designed for the needs, expectations and requirements of the recipients they serve to inform. In making this case, and to provide a practical way forward, we demonstrate Document Engineering as one potential methodology for specifying, designing, and deploying more effective information disclosures. Focusing on data protection disclosures, we illustrate and explore how designing disclosures as interfaces can better support greater oversight of organisational data and practices, and thus better align with broader transparency and accountability aims.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**; • **Human-centered computing**;

KEYWORDS

transparency, accountability, GDPR, document engineering, interfaces, data rights, usability

FAccT '22, June 21–24, 2022, Seoul, Republic of Korea
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9352-2/22/06.
https://doi.org/10.1145/3531146.3533133

ACM Reference Format:
Chris Norval, Kristin Cornelius, Jennifer Cobbe, and Jatinder Singh. 2022. Disclosure by Design: Designing information disclosures to support meaningful transparency and accountability. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3531146.3533133

1 INTRODUCTION

There are calls greater transparency around organisational data practices [37]. Information disclosures, whereby organisations provide particular information to an interested stakeholder, typically individuals (users), as one way of potentially increasing transparency over the often opaque nature of organisational practices. Generally, disclosures seek to inform and balance information asymmetries [29], and support accountability and recourse [13]. Importantly, *disclosures will be of limited benefit if they are not meaningful to their intended recipients* – yet, current disclosure practices typically entail highly technical 'data dumps' [5, 8, 44, 50], often with little to no supporting information to assist the recipients with interpretation [30]. This hinders *meaningful* transparency [2, 7, 8, 41], as it becomes increasingly challenging for all but the most expert recipients to navigate and comprehend the information which has been disclosed.

Better disclosures are not only possible, but crucial for supporting broader organisational transparency and accountability. Towards this, we make the case for a general *reconceptualisation of disclosures* towards *user interfaces* as a means for pursuing more effective transparency. This is because interface design provides a perspective where better understanding users' aims, needs, and expectations can lead to more appropriate and useful creations [25], which can be leveraged (alongside the wealth of HCI insights, methodologies, and evaluation techniques [1, 16, 46]) to design disclosures which are more suitable, usable, and meaningful for their intended recipients.

1.1 Supporting accountability

Disclosures often form part of a transparency regime, where transparency typically seeks to support accountability [13]. Fundamental to accountability, however, is that information is disclosed such that the recipient can effectively engage with, understand, and take steps to act on that information [9]. Transparency does not necessarily provide much benefit in and of itself [3, 9, 35]. Indeed, too much information can even counter-productively hinder understanding of what is happening, obscuring useful material behind or within other information (the so-called 'transparency paradox' [47]); current disclosure practices appear an unfortunate case in point (§2.2). Instead, meaningful accountability requires the disclosure of *contextually appropriate* information [13]: that *(i) relevant* to the forms of accountability likely to be involved, *(ii) accurate* in that it is correct, complete and properly representative, *(iii) proportionate* to the level of knowledge needed, and *(iv) comprehensible* by those to whom an account is owed. Such information, in turn, supports the broader accountability mechanisms of deliberation, review, interrogation, contestation, challenge, recourse, etc. [13, 32].

1.2 Toward meaningful disclosures

This paper argues that better disclosures—which act to meet the needs of their recipients—are possible. In particular, we challenge the overly-technical and fragmented nature of current disclosures, and highlight the potential and ways for realising better disclosures. This topic warrants urgent attention; as we have seen with other forms of mandated information (e.g. privacy policies [39]), the typical efforts and approaches for implementing such disclosures—whether effective or otherwise—tend to cement themselves as the 'status quo' over time. Current public discourse around the practices of tech/data-driven organisations along with new and emerging regulations that often include transparency provisions (§2.1) together offer an opportunity to challenge and reshape the expectations of information disclosures. This could prevent the current (inadequate) approaches to disclosures from being considered acceptable and falling into a stasis of (ineffective) 'checkbox exercises' as a result.

Specifically, we demonstrate the potential for **disclosure interfaces**, where disclosures—conceptualised as user interfaces—are designed around the needs, interests and expectations of their recipients. We illustrate one potential methodology (*Document Engineering* [23]) for their construction and evaluation, practically demonstrating the process by elaborating the methodological steps (and by leveraging a selection of real-world disclosures) for creating a disclosure interface. We then indicate the potential for disclosure interfaces through a user-study, showing that participants found a disclosure interface more usable, and were more confident in understanding its information. In all, we highlight the need for more attention to be brought to the practical dimensions of disclosures (and transparency regimes in general), showing that considering disclosure design can help in making organisational transparency measures more effective.

2 THE IMPORTANCE OF EFFECTIVE DISCLOSURES

An effective disclosure works to support a transparency regime, as the information disclosed can assist in monitoring and holding organisations to account for their actions (or, indeed, inactions) [13]. Disclosures are important for addressing power imbalances and information asymmetries, potentially helping to bring about more empowered and informed stakeholders [4] by providing a means to reveal or oversee organisational practices of concern or interest. For example:

Users, or those otherwise affected by a technology, potentially have much to gain from being able to understand more about an organisation's internal practices, particularly those that impact themselves. This may be a sole individual with an interest in how organisations use their information, or it could involve support from activists, civil society or public interest groups, etc. Having greater oversight over organisational practices can assist in better understanding the potential implications of engaging with an organisation (directly or indirectly), in determining whether the organisation behaves in line with expectations, and supporting further action. That is, disclosures can provide information that enables users to understand, identify, and push back against unscrupulous practices in an informed way [45, 51].

Regulators (i.e. independent public authorities responsible for monitoring and enforcing the application of the law) will also have an interest in the practices of organisations, and effective disclosures can assist them in carrying out their role. For example, disclosures can assist regulators' monitoring and oversight activities, ensuring that organisations are behaving appropriately and in accordance with their legal obligations [13]. Regulators play a particularly important role, given that they can take action in ways that individuals cannot; in investigations, disclosures can act as evidence, from which enforcement steps can be taken – be it guidance, warnings, or more serious punitive measures.

Organisations (including representative groups and trade organisations) themselves also stand to benefit from more transparency over their actions. Effective disclosures give opportunities for interested external parties to probe their actions, perhaps helping to identify practices leading to issues or outcomes that may be unexpected or unintentional (i.e. verifying that processes are operating as intended). Better disclosure practices can also help facilitate improved user relationships through demonstrating that they take their responsibilities seriously [34, 42].

2.1 Disclosure obligations within data protection law

Given the scale of personal data processing, disclosures within data protection regimes are particularly important. Data protection law generally seeks to empower those whose personal data is processed by organisations by providing means to oversee and such processing. The provision of disclosed information can support a range of interested stakeholders in understanding and challenging the organisational practices that concern them, and can assist in bringing organisations using personal data to heel [15]. The role of disclosed information in this regard has long made it (alongside organisational transparency and accountability more broadly) an important focus for policymakers.

The prominent example of data protection law is the EU's General Data Protection Regulation (GDPR) [18]. The GDPR governs the processing of personal data, strengthening the rights of those whose personal data is being processed ('data subjects'), while reinforcing the responsibilities of the organisations or entities responsible for that processing ('data controllers'). A foundational principle of the GDPR is the transparency of information about processing (GDPR

Art. 5(1)(a), Arts. 12–14). The GDPR includes requirements concerning record-keeping and documentation around data processing that controllers must comply with (Art. 30). Data controllers are under various transparency obligations to disclose information to data subjects about processing involving their personal data and about their rights in relation to that processing (Arts. 13–14). Such is the importance of these obligations that the GDPR attaches potential fines of up to the greater of 20M€ or 4% of the data controller's global turnover for non-compliance (Art. 83(5)(b)).

One key right under the GDPR's transparency regime is data subjects' 'right of access' to personal data (Art. 15). This right allows individuals to request a copy of any personal data relating to them that is being processed by the controller, and obliges controllers to respond in a timely manner. Such 'data disclosures' should include additional information, including (amongst others) the purposes of processing, the categories of personal data concerned, and the recipients or categories of recipient to whom the personal data has been shared (Art. 15(1)).

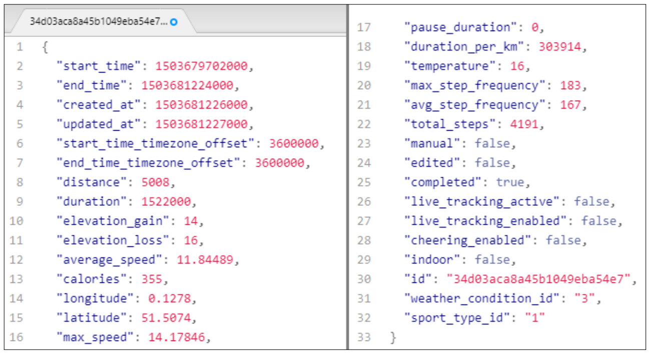
The transparency of organisational practices also serves as an important mechanism for regulators. Under the GDPR, regulators are granted a range of investigatory, advisory, and corrective powers (Art. 58). These include the ability to carry out investigations in the form of data protection audits (Art. 58(1)(b)), to obtain access to all personal data and to all information necessary for the performance of its tasks (Art. 58(1)(e)), and to issue reprimands where processing operations have infringed provisions of the Regulation (Art. 58(2)(b)). Importantly, regulators also have the power to set standards, and intervene if current disclosure practices are deemed unsatisfactory (Art. 58(3)). In this way, they have the ability to influence the practices and procedures of organisations for the better, and to shape these going forward.

While but one example, the GDPR is fast becoming the global standard in data protection, influencing subsequent regulations around the globe. Several similar regulatory efforts have since (or are due to) come into effect, including California's CCPA [12], Brazil's LGPD [40] and India's PDPB [36], each of which offers comparable (to a degree) rights, record-keeping and transparency obligations for data controllers. Moreover, EU law requires that personal data can only be transferred to non-EU jurisdictions in certain circumstances, including where that jurisdiction has a similarly adequate level of data protection (Art. 45), thereby effectively working to 'export' certain standards.

2.2 Limitations of disclosure practices: A need for attention

While there are various benefits to having meaningful and effective transparency over organisations that process data, current disclosure practices leave much room for improvement. For one, it has been shown that data disclosures can be difficult to understand; they are often fragmented, decontextualised, and highly technical in nature [5, 8, 44]. Moreover, there is much inconsistency in how organisations handle disclosures and in the information returned [5, 8, 44, 50], making it challenging for more general tooling or guidance to assist. Indeed, in obtaining disclosures for this research (§4.1), we found that they were highly inconsistent, technical, fragmented, and decontextualised from the applications at hand.

Figure 1: A file returned as part of a disclosure from an exercise tracking application. Line-breaks were added for readability, and some values perturbed for anonymity, though data types and formats have been preserved.



```
{
  "start_time": 1503679702000,
  "end_time": 1503681224000,
  "created_at": 1503681226000,
  "updated_at": 1503681227000,
  "start_time_timezone_offset": 3600000,
  "end_time_timezone_offset": 3600000,
  "distance": 5008,
  "duration": 1522000,
  "elevation_gain": 14,
  "elevation_loss": 16,
  "average_speed": 11.84489,
  "calories": 355,
  "longitude": 0.1278,
  "latitude": 51.5074,
  "max_speed": 14.17846,
  "pause_duration": 0,
  "duration_per_km": 303914,
  "temperature": 16,
  "max_step_frequency": 183,
  "avg_step_frequency": 167,
  "total_steps": 4191,
  "manual": false,
  "edited": false,
  "completed": true,
  "live_tracking_active": false,
  "live_tracking_enabled": false,
  "cheering_enabled": false,
  "indoor": false,
  "id": "34d03aca8a45b1049eba54e7",
  "weather_condition_id": "3",
  "sport_type_id": "1"
}
```

Fig. 1 shows an indicative example from one such disclosure we received, corroborating the literature.

Though the GDPR itself does not prescribe specific formats or structures to disclosures,[1] the GDPR does specify that any disclosed information should be provided "in a concise, transparent, intelligible and easily accessible form, using clear and plain language" (Art. 12(1)). Moreover, the Article 29 Working Party (which comprised representatives from the data protection authority of each EU member state) further elaborated on the importance of "the quality, accessibility and comprehensibility of the information", noting that it "is as important as the actual content of the transparency information" [4]. In other words, current organisational disclosure practices appear inadequate, though the GDPR grants a range of powers to regulators to require higher standards as deemed appropriate in within the law (§2.1). Data protection regulators therefore have the impetus and the remit to act towards better practices.

In short, the technically-oriented 'data dumps' that many have identified as common disclosure practice appear to fall short of the expectations and requirements set out in data protection regimes – which, given the role disclosures play in addressing asymmetries, is to society's detriment. While regulators have tools at their disposal, there is a need for a better understanding of how more 'effective' disclosures might be realised. This includes, amongst other things, work which strengthens understanding of the challenges facing various stakeholders and their informational needs, as well as methods and approaches for better developing and evaluating disclosures. Work such as ours helps to both argue the case and demonstrate what is possible, thereby providing a concrete way forward in this important area.

3 RECONCEPTUALISING DISCLOSURES AS INTERFACES

There is a real opportunity to improve current disclosure practices, to bring about disclosures that are more effective in supporting user

[1] Note the right to data portability that specifies that responses must be "in a structured, commonly used and machine readable format" (Art. 20(1)). However, this is not the GDPR's only transparency obligation, and is separate from the right of access (Art. 15).

understanding and oversight. Towards this, we argue that *reconceptualising disclosures as user-centric* ***interfaces***—such that they are *designed around the needs of their intended recipients*—represents one practical way forward. Importantly, the aim of such is not simply for 'prettier' interfaces, as these won't necessarily translate to more meaningful transparency; nor should interface design act as a replacement for the provision of 'raw data', which can allow for deeper analysis and validation, while helping support compliance (§2.1). Rather, disclosures must first and foremost meet their purpose of *informing* recipients and supporting effective transparency and oversight.

3.1 Reshaping disclosure practices

Different avenues of study contribute to our suggested paradigm shift towards disclosures as organisational interfaces. For example, Drucker's proposed approach of 'graphesis', the "study of the visual production of knowledge", includes a close, critical analysis of the conventions of the graphical user interface (GUI) that "encode knowledge through [its] visual structures and rhetorics of representation" [17]. In other words, certain graphical forms carry with them epistemological assumptions about the information they present, and can ultimately steer a user's interpretation of that data. This is particularly relevant for the presentation of disclosure data; for instance, Drucker notes how information arranged "statically" in tabular form (as seems common with current data disclosures; §2.2) can give the impression that it has been produced according to a "strict distinction of content types" that are now unchangeable. Without a critical methodology, this form will then come to be expected by users and its arbitrariness will come to seem the only way that this information can be presented. Moreover, this theory illustrates a limited window of opportunity to work towards reshaping disclosures into something more meaningful to their recipients, ultimately fostering a new genre of data disclosure which utilises interface design to deliver information. Towards this, we will now show one such methodology, *Document Engineering* [23], which appears particularly promising in helping to create interface disclosures that are more contextually appropriate, meaningful, and effective.

3.2 Creating more meaningful documents with DocEng

Document Engineering (DocEng) is an established methodology for specifying, designing, and deploying the information models that enable document-centric applications [23]. It involves describing disparate information and data sources with "new document models," reusing "common or standard patterns to make documents more general and robust". DocEng is said to assist in making information flows more consistent, structured, compatible, and visible across stakeholders; it was a crucial methodology in the transition from people-oriented documents to technical schemas.

DocEng works to find commonalities amongst forms and transaction documentation so as to streamline and automate document exchange within and across organisations. Organisational relationships are conceptualised as a chain of document exchanges, where the parties understand each other's documents. In this way, documents are framed as exposing the inputs and outputs of business processes, where they serve as the public interfaces to the organisation [23]. The DocEng methodology contains six steps, involving the deconstruction of existing information sources (steps 2–3) into a reconstruction of related documents (steps 4–6), based on requirements of the stakeholders (step 1).

Given that the aims of DocEng are to to facilitate a more consistent, structured, and understandable approach to document exchange, it follows that the methodology is applicable to the context of creating user-oriented disclosures. Fundamentally, disclosures *are* documents, and we can therefore look to 'deconstruct' examples of current disclosure practices (i.e. technical data dumps) before 'reconstructing' them as user-facing documents—*interfaces*—that are more intuitive and which better align to the requirements of their recipients. DocEng therefore works in concert with design methodologies (such as user-oriented development), documentation approaches (e.g. datasheets [21], model cards [27]), or any number of other potential methods to build and tailor the appearance of these disclosures once the substance of those documents are established. While the context of use changes—i.e. we flip the typical application of DocEng from deriving schemas from user-facing interfaces to the reverse—we maintain the integrity and order of DocEng's six steps.

Again, DocEng is but one (promising) approach to creating more meaningful disclosures. In outlining how we might apply this methodology, our aim is not to argue that this is how disclosures *should* be created, nor that DocEng will be applicable in every situation (such factors being highly contextual). Rather, we illustrate one such way in which interfaces offer a new way forward, towards information disclosures that are designed around the needs of recipients.

4 METHOD: DISCLOSURE INTERFACES IN PRACTICE

To illustrate the types of disclosures that an interface-conceptualisation might facilitate, and their potential, we now explore how one might engineer and evaluate disclosure interfaces. Our focus is on data disclosures aimed at individuals, i.e. the user or someone otherwise affected by the technological or organisational practices. This is because in a personal data context, individuals are the target of the transparency regime and are afforded related rights (§2.1). Moreover, design practices are particularly relevant for individuals c.f. other stakeholders who often have access to greater expertise.

We begin by developing an exemplar disclosure interface—grounded in actual disclosures from real applications—illustrating the DocEng methodology's six key steps in a disclosure context (§5). We evaluate the output against an example representing current practice through a user study with 123 participants (§6).

4.1 Grounding the process

We demonstrate the process from the perspective of a hypothetical organisation which has developed a mobile app. This app tracks and records exercise sessions (running, cycling, etc.), and streams music to the user as they work out. It collects a broad range of data in the process, including user fitness information, GPS data, and music streaming activity. It uses this wealth of data to make inferences about each user (e.g. their shopping and eating preferences, and their state of physical and mental well-being) to supply targeted advertising (ads) based on user profiles (i.e. groups of users with similar inferences). This app was designed to cover several aspects of common functionality (i.e. exercise tracking, music streaming, advertising) and be derivative of existing and widely used apps and practices, while remaining realistic in the data processing that occurs (we derived the nature of the data and its processing from real organisational disclosures).

To ground our work, we obtained a collection of data disclosures from 20 popular tech-oriented services. We did this by first identifying a range of different application sectors, some of which were directly related to our app in question (e.g. music streaming services, exercise tracking apps, advertising platforms), while others to gain a broader view over the disclosure landscape (e.g. news and media, food delivery services, communication platforms; see [33]). We then selected apps that were highly prominent within that sector (number of users, app store ranking, etc.), using that service for a period of time, and obtaining disclosures through two methods: exercising our GDPR right of access, and using 'download your data' tools.[2] Note that our intention was not to collate a representative or exhaustive sample of organisational disclosure practices (as others have studied; §2.2), nor was our aim to single out any particular organisations or sectors. Rather, our goal was to obtain a sample of disclosures from indicative organisations, so to illustrate current disclosure practices, and to provide a grounded basis for deriving a disclosure interface reflective of real-world disclosure practices.

5 DEMONSTRATING 'DOCUMENT ENGINEERING' – STEP-BY-STEP

DocEng [23] comprises six steps in total, which we now demonstrate being adapted to a disclosure context.

5.1 Descriptive analysis

The first three steps involve selecting disclosures of interest, identifying commonalities, and analysing current practices.

Step 1 – Analyzing the contexts of use: The first step of DocEng entails analysing the contexts in which the outcome of the process, here the disclosure, will be used. In the context of disclosures, this step involves identifying the potential stakeholders and what they might wish to gain from the disclosure, thus acting as requirements for the later steps of analysis. It is also important to consider the legal requirements that the disclosure should meet (satisfying the organisation's regulatory obligations). Often the outcome of this step will be generally applicable across a range of disclosures, and thus the contexts of uses may be similar across a range of organisations and sectors. As discussed, a general requirement for disclosed information is that it should be contextually appropriate (§1.1), providing information which is (i) *accurate*, (ii) *relevant*, (iii) *proportionate*, and (iv) *comprehensible* [13]. The presentation of disclosure data will naturally impact across all of these dimensions, and can assist in identifying ways in which disclosure interfaces can better support stakeholders in meaningfully engaging in how an organisation is processing personal data.

Step 2 – Analysing business processes and patterns: The second step involves exploring current business (data) processes. Here, we used the range of disclosures we obtained to consider on the 'categories' (i.e. types) of data presented, alongside how this data is grouped, structured, and represented within current disclosures (Table 1). Such patterns allowed us to start to build an indicative model of the types of information that our exemplar disclosure should likely contain, and how this information might be grouped and presented to form a more contextually appropriate disclosure. For example, we observed that all exercise apps contained data that matched the 'user details', 'activity data' (exercise sessions), 'geographic data' (exercise coordinates), and 'device information' categories; correspondingly, music streaming apps all contained user details, activity data (music streamed), purchase information (subscriptions), and usage logs.

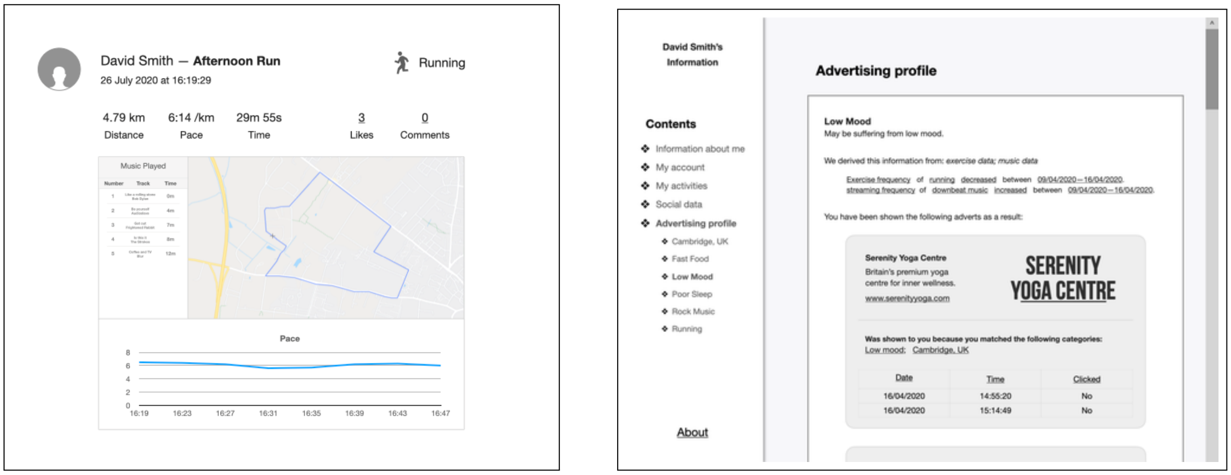
Table 1: A combined list of data categories identified from analysing disclosures.

Category	Brief Description
User details	Information about the user and their account
Account preferences	Preferences and settings about the user's account
Activity data	Records produced while using the service
Purchases/transactions	Purchase history
Library information	Playlists, libraries, and account inventories
Geographic data	Location-based data
Inferences	Inferences about the user
Advertising	Records of the adverts that the user has seen
Search history	What the user has searched for
Community/social	Groups joined, public messages, etc.
Private messages	Private messages between users
Media	Photos, videos, etc.
Health data	Data relating to the user's health
Support messages	Messages between the user and the organisation
Usage logs	Logs generated while using the service
Device information	Information about the user's device(s)

Step 3 – Document analysis: The third step involves analysing the documents. This entailed analysing the content and structure of the disclosures we obtained, including looking at the structure of disclosures of particular sectors (and more broadly) to explore how related data is grouped, and whether visualisations or other modalities might better suit the context of the recipient's requirements. We also contrasted the disclosures against the interfaces of the apps in question, to identify where the disclosure could better align with the experiences of the user. While only a few organisations presented data in HTML (though even much of this was still tabular), most comprised several files of structured data (CSV, JSON, etc.) which were typically highly technical in nature (often containing database keys, hashed values, etc.). Furthermore, closely-related information (e.g. exercise sessions and their corresponding GPS coordinates) were often separated, requiring technical knowledge and effort to join and meaningfully interpret them. Many data types were generally uninterpretable or required some conversion – dates and times were frequently presented as UNIX timestamps, location data was usually in lat/lng, durations were typically in seconds (regardless of length), and some values lacked measurement units

[2] While there is a distinction between such tools and GDPR rights requests, a number of organisations instructed us to use such tools when exercising our data protection rights. Moreover, these tools are common mechanisms by which this information is disclosed by many (larger) organisations, and therefore directly reflect how individuals gain access to their data in the real world.

Figure 2: Left: A prototype component for the ‘activity data’ category, showing our recontextualisation of disparate exercise, geographic, and streaming datasets. Right: A prototype disclosure interface containing the ‘advertising’ component.



David Smith — **Afternoon Run**
26 July 2020 at 16:19:29

Running

4.79 km	6:14 /km	29m 55s	3	0
Distance	Pace	Time	Likes	Comments

Music Played

Number	Track	Time
1	Like a rolling stone Bob Dylan	0m
2	Be yourself Audioslave	4m
3	Get out Frightened Rabbit	7m
4	Is this it The Strokes	8m
5	Coffee and TV Blur	12m

David Smith's Information

Contents

- Information about me
- My account
- My activities
- Social data
- Advertising profile
  - Cambridge, UK
  - Fast Food
  - Low Mood
  - Poor Sleep
  - Rock Music
  - Running

About

Advertising profile

Low Mood
May be suffering from low mood.

We derived this information from: *exercise data; music data*

Exercise frequency of running decreased between 09/04/2020 – 16/04/2020.
streaming frequency of downbeat music increased between 09/04/2020 – 16/04/2020.

You have been shown the following adverts as a result:

Serenity Yoga Centre
Britain's premium yoga centre for inner wellness.
www.serenityyoga.com

Was shown to you because you matched the following categories:
Low mood; Cambridge, UK

Date	Time	Clicked
16/04/2020	14:55:20	No
16/04/2020	15:14:49	No

for interpreting the data (e.g. “distance: 5008"); see Fig. 1 for an indicative example.

5.2 Constructing our disclosure

The next three steps of DocEng involve constructing and implementing the document (disclosure interface). Importantly, DocEng does not prescribe any particular design approach, such that various design methods can be employed as appropriate. Here we opted to undertake rapid prototyping to create our interface [49], though other approaches (e.g. co-design) could be relevant. Again, we reiterate that the interface we develop is to illustrate what is possible with the methodology, and does not represent the only way that such interfaces can be created, nor their outcome.

Step 4 – Component assembly: The fourth step involves starting to prototype *components* (groups of related data which, when collated, present a bigger picture of the information at hand). This involves exploring how disclosure data might be re-assembled into more meaningful structures (i.e. re-imagining the ‘substance’ of what the document should contain), and how these may be presented. Importantly, we consider how users might interpret and understand the data itself; for our exemplar, we aggregated information from different files to maintain the contextual nature of what they represent (e.g. joining exercise summaries and GPS coordinates). We considered the presentation of this data, opting to plot geographic data on a map, pace information as time series, etc. Where data was presented in numerical formats, technical data types (e.g. UNIX timestamps) were converted into more familiar date formats, units were presented where relevant, and snippets of information about how these data were being used were included. We explored this process for three prototype components (relating to the ‘user details’, ‘activity data’, and ‘advertising’ categories) to support a demonstration – though, in practice, this would involve all applicable categories identified during Step 2.

Step 5 – Document assembly: The fifth step of DocEng entails assembling the components into a document model, which works to provide a complete understanding of what the final document should look like, and how relevant components are grouped. In demonstrating this step, we aimed for a simple design, breaking up related components into sections that the user can navigate. Structurally, we opted for an HTML-based disclosure, given the potential for pagination, interactive components, and compatibility. We created pages for groups of related information (i.e. components), and added hierarchical navigation links that the recipient could traverse. In short, we brought together the components of Step 4 (Fig. 2; Left) into a cohesive prototype of our disclosure (Fig. 2; Right).

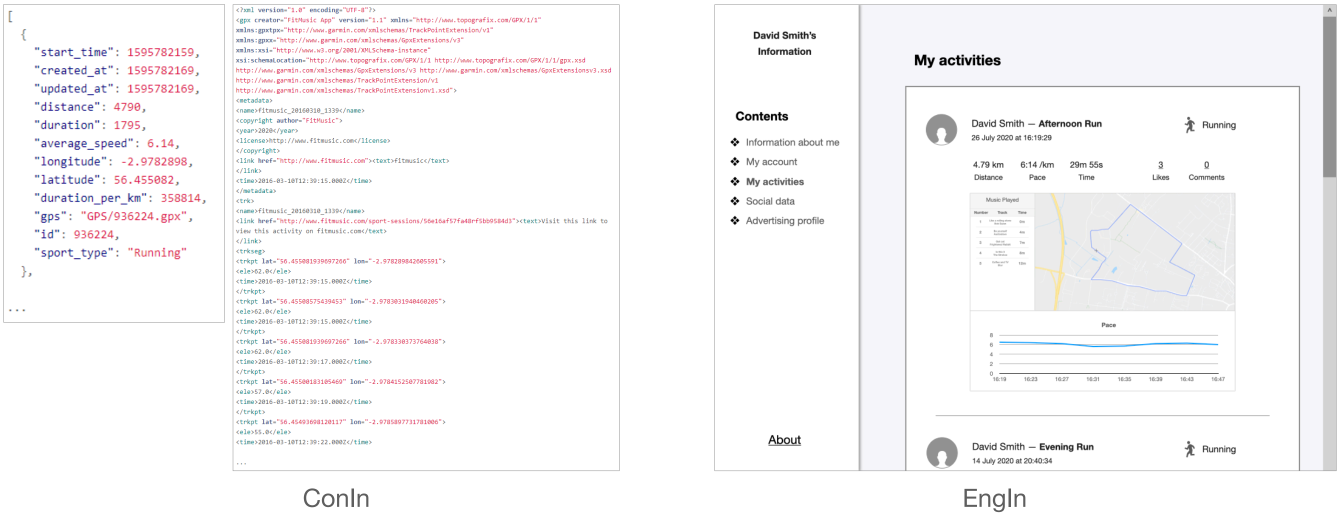
Step 6 – Implementation and beyond: This final step of DocEng concerns operationalising and implementing the new interface. This involves the organisation putting the engineered disclosure into practice (e.g. for our demonstrator, creating and deploying the HTML/CSS/Javascript for our interactive browser-based disclosure). Given the requirements of the GDPR, all files should be obtainable by the recipient allowing for local (offline) archival and access – though organisations may consider providing an online version through their app as well.

Importantly, the DocEng process does not end here; much like the applications themselves, disclosure interfaces entails an iterative approach – refining and adapting practices as the systems, usage contexts and requirements change, and as the wider landscape of transparency regimes continues to evolve.

6 THE POTENTIAL FOR ENGINEERED DISCLOSURES

The goal is for disclosures that better meet their purpose of effectively communicating information. Here we explore this dimension, as well as ways that disclosures might be compared—outlining a way forward, not only for the creation of disclosure interfaces, but for their evaluation and iterative development—focusing on principles of usability and information retrieval (over aesthetics alone; see §7.2).

Figure 3: Disclosure representations of the same information about an exercise session. Left: part of the Control Interface (ConIn), based on current disclosure practices. Right: part of our Engineered Interface (EngIn), prototyped through DocEng.



```
[
  {
    "start_time": 1595782159,
    "created_at": 1595782169,
    "updated_at": 1595782169,
    "distance": 4790,
    "duration": 1795,
    "average_speed": 6.14,
    "longitude": -2.9782898,
    "latitude": 56.455082,
    "duration_per_km": 358814,
    "gps": "GPS/936224.gpx",
    "id": 936224,
    "sport_type": "Running"
  },

...
```

```
<?xml version="1.0" encoding="UTF-8"?>
<gpx creator="FitMusic App" version="1.1" xmlns="http://www.topografix.com/GPX/1/1"
xmlns:gpxtpx="http://www.garmin.com/xmlschemas/TrackPointExtension/v1"
xmlns:gpxx="http://www.garmin.com/xmlschemas/GpxExtensions/v3"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.topografix.com/GPX/1/1 http://www.topografix.com/GPX/1/1/gpx.xsd
http://www.garmin.com/xmlschemas/GpxExtensions/v3 http://www.garmin.com/xmlschemas/GpxExtensionsv3.xsd
http://www.garmin.com/xmlschemas/TrackPointExtension/v1
http://www.garmin.com/xmlschemas/TrackPointExtensionv1.xsd">
<metadata>
<name>fitmusic_20160310_1339</name>
<copyright author="FitMusic">
<year>2020</year>
<license>http://www.fitmusic.com</license>
</copyright>
<link href="http://www.fitmusic.com"><text>fitmusic</text>
</link>
<time>2016-03-10T12:39:15.000Z</time>
</metadata>
<trk>
<name>fitmusic_20160310_1339</name>
<link href="http://www.fitmusic.com/sport-sessions/56e16af57fa48rf5bb9584d3"><text>Visit this link to
view this activity on fitmusic.com</text>
</link>
<trkseg>
<trkpt lat="56.455081939697266" lon="-2.978289842605591">
<ele>62.0</ele>
<time>2016-03-10T12:39:15.000Z</time>
</trkpt>
<trkpt lat="56.45508575439453" lon="-2.9783031940460205">
<ele>62.0</ele>
<time>2016-03-10T12:39:15.000Z</time>
</trkpt>
<trkpt lat="56.455081939697266" lon="-2.978330373764038">
<ele>62.0</ele>
<time>2016-03-10T12:39:17.000Z</time>
</trkpt>
<trkpt lat="56.45500183105469" lon="-2.9784152507781982">
<ele>57.0</ele>
<time>2016-03-10T12:39:19.000Z</time>
</trkpt>
<trkpt lat="56.45493698120117" lon="-2.9785897731781006">
<ele>55.0</ele>
<time>2016-03-10T12:39:22.000Z</time>

...
```

David Smith's Information

Contents

- Information about me
- My account
- **My activities**
- Social data
- Advertising profile

About

My activities

David Smith — **Afternoon Run** Running
26 July 2020 at 16:19:29

4.79 km	6:14 /km	29m 55s	3	0
Distance	Pace	Time	Likes	Comments

David Smith — **Evening Run** Running
14 July 2020 at 20:40:34

We now present a user study that explores to what extent individuals might benefit from an 'engineered interface' *(EngIn)* compared to that of a 'control interface' *(ConIn)* (Fig. 3; Left) that that reflects current practice. EngIn was our final version of the interface (developed in §5, see Fig. 2 & Fig. 3; Right), and comprised three main sections; 'My account', 'My activities', and 'Advertising profile'. ConIn comprised real-world disclosure responses we received from organisations, fusing together the information for advertising, music streaming, exercise tracking, and account information categories to reflect our app's functionality. Importantly, both ConIn and EngIn contained the same underlying information, and were reflective of the information contained within actual disclosures. This allowed our control interface to be representative of current practices while still acting as a point of comparison for our engineered disclosure.

Again, our focus was on individuals (those using the app; see §6.1), and we focus on three main points of measurement: *user preference* as a data point reflecting which disclosure participants preferred; *usability* through the System Usability Scale (SUS) [11]; and *task certainty* reflecting the participants' perceived confidence in being able to meaningfully interrogate the data held within the disclosure. We also asked open-ended questions to probe participants on their views regarding disclosures. By comparing EngIn to ConIn, we explore the potential for what disclosures could become when designed as interfaces. Our approach also illustrates how such disclosures might be compared and contrasted, though the exact nature of what makes a disclosure 'effective' will often relate to its particular context.

6.1 Method

We recruited participants through Mechanical Turk, requiring those having a high rate (>95%) in successfully completing tasks on the platform (such criteria in line with guidance from the literature [43]). All participants were UK-based, given we sought native speakers from territories with established data protection regimes. We ended up with 123 participants (63% male; 37% female) with a fairly broad level of technical expertise (26% claimed to have 'Expert' or 'Advanced' technical knowledge; 41% had a 'Good level'; and 32% had 'Some' or 'No' knowledge). Participants were compensated by an amount reflecting the UK's 'living wage' and we received approval from our Department's ethics committee.

Participants were first introduced to the concept of disclosures and given an overview of the study. After consenting and providing some optional demographic information, they were presented with a description of our app and asked to imagine that they had requested a copy of their personal data. Participants were given no indication as to how or why the two interfaces were designed (until the end of the study), to help mitigate 'good-subject' effects [31].

Participants were randomly allocated one of the two interfaces to start with, and asked three questions relating to the content within (on what date they accepted the Terms of Service; the distance they ran on a particular session, and why a particular advert was shown to them). After providing each answer, they were then asked how certain they were ("Very certain"; "Uncertain"; or "Couldn't answer"). At the end of the three questions, the participants then completed an SUS questionnaire for that interface [11], before repeating these steps on the other interface (half moving from ConIn to EngIn, the other half from EngIn to ConIn). Finally, participants were asked which interface they preferred, followed by a series of open-ended questions about their experiences with the two interfaces, including what they liked and disliked about the interfaces, and what advice they would give to designers of disclosures going forward.

Please see our supplementary materials for further details of the interfaces, study and the analysis [33].

Table 2: Evaluation results. The disclosure derived through DocEng (EngIn) received higher usability scores, higher task certainty scores, and was preferred more frequently than the control (ConIn).

	Comparison	Data	ConIn (mdn)	EngIn (mdn)	Statistical Test
1	SUS score	All responses	25	87.5	Z = -9.57, $p < .001^{***}$, r = .86
2	SUS score	First system evaluations only	32.5	80	Z = -8.17, $p < .001^{***}$, r = .74
3	SUS score	First system evaluations & non-technical participants	25	72.5	Z = -4.91, $p < .001^{***}$, r = .78
4	SUS score	First system evaluations & technical participants	43.75	80	Z = -3.27, $p = .001^{**}$, r = .58
5	# tasks answered	First system evaluations only	2	3	Z = -8.43, $p < .001^{***}$, r = .76
6	# tasks answered	First system evaluations & 'Very certain'	1	3	Z = -8.99, $p < .001^{***}$, r = .81

6.2 Results

Our study evaluates the effectiveness of our engineered interface (EngIn) in comparison to current practices (ConIn).

6.2.1 Participants preferred EngIn to ConIn: When directly asked which interface they preferred, participants near unanimously selected EngIn over ConIn (98%). This was further supported when comparing the proportions of those that prefer EngIn using a Chi-Squared test;[3] χ^2 (1, N=123) = 115.130, $p < .001$, $\phi = .97$.

6.2.2 Participants found EngIn more usable than ConIn: We also compared SUS scores for each interface. The SUS is a 0–100 score "that can be used for global assessments of systems' usability" [11]. A higher score indicates a more 'usable' interface, with prior work determining that a "poor" average SUS score is 35.7, "okay" is 50.9, and "good" is 71.4 [6].

Considering the SUS scores, we found a statistically significant difference between the two interfaces in EngIn's favour (Table 2; comparison 1). To control for any potential ordering effects [38]), we also compared SUS scores for only the first interface that participants were shown; as ordering was randomised, which gave two groups of (unpaired) SUS responses, each comprising half of the participants. Again, findings indicated a statistically significant difference between the two interfaces in EngIn's favour (Table 2; comparison 2). We also controlled for the impact of the participants' technical expertise on their scores, repeating two further SUS comparisons (both controlling for ordering effects); one containing only those who self-identified as having *limited* or *no* technical knowledge (n = 40), and another with only those with *advanced* or *expert* technical knowledge (n = 32). In both cases, EngIn received statistically higher SUS scores than ConIn (Table 2; comparisons 3 & 4). Though EngIn being preferred is perhaps unsurprising, these results show the potential for designed interfaces to provide a more usable form of information disclosure cf. current practices.

6.2.3 Participants were more confident in interrogating the engineered disclosure: We also explored the degree to which participants could meaningfully interrogate the data, focusing on their perceived confidence in answering questions about it,[4] comparing (i) the number of questions each participant could answer, and (ii) the confidence in their response. Again, we only analyse data relating to the first interface that each participant saw, thereby mitigating order effects. In both cases, EngIn had participants answering more questions and demonstrating more certainty about the answers they gave over ConIn (Table 2; comparisons 5 & 6). This suggests that EngIn was more closely aligned with the aims of a disclosure, instilling a greater degree of confidence in what was presented in terms of insight and understanding.

6.2.4 Participants saw disclosures as a way of providing organisational oversight: In prioritising a list of reasons why they might wish to obtain their data (i.e. seek a disclosure), some of the most frequently prioritised reasons related to reasons of organisational oversight regarding either particular or more general concerns (see Table 3). For example, to 'ensure that no unwanted data is being stored' was within the top-three reasons of almost 2/3 of our respondents. Other prominent reasons including distrust of the organisation, and to verify and find particular information, showing the role for disclosures playing as an informative function. Of much lower priority were 'to transfer that data to another service' and 'archival purposes', which is interesting given that these reasons better suit the (technically-oriented) nature of disclosure approaches common today. We also found that 18% (22) of participants had previously requested a disclosure from an organisation, a greater number than we had anticipated. This shows that disclosures are already being used, and having real-world effects and implications, thus highlighting the urgency of considering this topic.

Table 3: The list of reasons that participants' reordered to indicate their interests in obtaining a disclosure. This shows the percentage of respondents that thought each reason was high- or low-priority (i.e. in their top-three or bottom-three reasons respectively).

% of participants to list this reason in their:	Top-three	Bottom-three
To ensure that no unwanted data is stored	64.2%	16.3%
Concern or distrust of the organisation	49.6%	12.2%
General interest	43.1%	17.9%
Find particular information	42.3%	5.7%
Verify correct information	40.7%	22.8%
Legal Reasons	31.7%	25.2%
To transfer that data to another service	13.8%	51.2%
Archival purposes	9.8%	48.0%

[3] We opt for nonparametric tests (Chi-Squared for proportions; Wilcoxon for paired samples; Mann-Whitney U for independent samples) given they make fewer assumptions about the underlying data [22].

[4] Given that the purpose of a disclosure is to help inform, we focused on participants' perceived confidence, which allowed us to differentiate uncertain guesses from confident answers. This therefore indicates whether they felt able to interpret and understand the information presented to them.

6.2.5 Participants had a range of attitudes towards engineered disclosures: We also collected a number of open responses from participants, and used thematic analysis [10] with inductive coding to uncover several interesting insights:

Usability: One of the most frequent themes to emerge was a desire for disclosures to be easier to use and to understand. EngIn was described as *"much easier to use in terms of picking out the data you were looking for"*, whereas ConIn was *"jumbled and incomprehensible, very confusing"*. ConIn was said to be *"impossible to use practically by an average person, unless they are computer and technical experts"* – a rebuke of current disclosure practices.

Insightfulness: Several comments from participants indicated that EngIn helped them gain new insights into data collection practices in general. Some expressed unease at the inferences (*"I disliked the fact that they had a lot of data about me that they had guessed"*; *"Scary to think that adverts are based on my exercise and music data – but I'm assuming this must be true (or will be soon!)"*), and the amount of data was described as *"spooky"* and *"invasive"*. One participant specifically reflected on their relationship with technology: *"I'm not getting an Apple Watch for sure"*. Eliciting these types of responses is indicative of an effective disclosure – facilitating the scrutiny of organisational practices, and better informing users as to how their personal data is being used, enabling them to appropriately respond [13].

Familiarity: A recurring suggestion was for disclosures to have a familiar design (e.g. reflecting that of common websites or apps), as one way of making the recipient *"feel comfortable and at ease"*, as *"the user already knows how to navigate it."* While the appropriate disclosure design will typically depend on the context, note that many organisations will have already designed user-facing interfaces for the service's app/website itself; therefore, designing their disclosures to take a similar form could be a straightforward way for communicating in a familiar manner. However, we again stress that a more familiar (or appealing) interface doesn't necessarily translate to an effective information disclosure (§7.2), and DocEng appears a promising way toward grounding the design of such disclosures in the requirements of recipients.

Support: A further insight was the desire for tutorials or support accompanying the disclosure. Some suggested *"help pop-ups"* or *"extra tips to be able to decipher exactly what you're looking at"*. Again, these highlight the need for disclosures to be built to support the particular stakeholders, rather than technical experts. As some responses outline, designers of disclosure interfaces should *"think about your user"*, and *"make it accessible to non-technical people"*. One response emphasised that developers should *"test designs on users before implementing, and seek feedback"*, supporting our wider call for a paradigm shift towards treating disclosures as interfaces which can be evaluated and improved.

In all, there was a clear sense that ConIn was insufficient for informing users of how their data was being used, undermining the purpose of disclosures as a tool to inform. One respondent noted that *"most people would ... think [ConIn] was some kind of coding and not what they asked for"*, another stating that it would *"likely get deleted without me even making any effort to understand it"*. Such responses show the need for disclosures that better target their recipients.

7 DISCUSSION

It is clear that organisational data practices can have significant real-world consequences for wider society [28]. Information disclosures represent an important mechanism to help hold organisations to account by providing transparency over often-complex organisational practices. Again, transparency will not alone *'solve'* accountability concerns, as transparency in and of itself does not necessarily support effective oversight and accountability [47]. But as we and others have argued, meaningful information is important for holding entities to account (§1.1), providing a means to oversee, review, and challenge the practices of organisations. In this way, it is crucial that such disclosures can be properly obtained, comprehended, and acted upon by the relevant parties.

Despite the importance of disclosure mechanisms (§2), current disclosure practices within a data protection context appear well-short of expectations. Through obtaining disclosures from actual organisations (§4.1), we found that responses were typically highly technical and fragmented, corroborating findings from the literature [5, 8, 44, 50]. While some disclosures appeared better than others, they overwhelming seemed a far cry from the requirements set out in the GDPR (concise, intelligible, easily accessible, etc.). And our user study findings showed that disclosures representing current approaches were more poorly received, across various dimensions, when compared to our engineered disclosure interface. Put short, current practices are often sub-par, and without new methods and ways of thinking about the formation of disclosures, these ineffective documents risk becoming cemented as 'the norm' (as we have seen with other ineffectual forms of mandated disclosure, such as privacy policies [39]) – much to the detriment of society more widely.

We have set out to both (i) draw attention to the area of disclosures, and (ii) show that *better is possible.* As a way forward, we described how disclosures can be conceptualised as interfaces, so to help produce disclosures that are more meaningful, useful, and tailored around the needs of recipients, exploring DocEng as one method amenable for such.

Limitations: We reiterate that the exact nature of our user study and interface are illustrative; we recognise that our study contains limitations, including that our sample is likely unrepresentative of the broader population (as is the case with any self-selecting group [19]), and that the disclosures we obtained, which form the basis for our engineered interface, represent those of a limited set of organisations. Despite this, our DocEng process resulted in a disclosure that participants significantly favoured and were better able to navigate, even when accounting for differing levels of technical expertise and prior exposure. Further, some aspects might not be as amenable to interface design, as more complex information (e.g. that about ML model specifics, should that be needed) may be more difficult to represent. However, DocEng as a method enables the range of design techniques, documentation and evaluation methods, etc., to assist and be integrated into disclosures where appropriate; e.g. in an ML context, datasheets [21], model cards [27], and decision provenance [45] might all work to increase greater organisational transparency over complex information.

In all, we demonstrate that there are ways to develop, evaluate, and refine disclosures that can better serve the needs of those that receive them. There is a strong need for more attention on this topic – not just to push organisations toward implementing better disclosure practices, but also regarding how these might be brought about in practice. To that end, we next outline some key considerations that arise from our work.

7.1 Incentives and drivers for change

While effective data disclosures bring various benefits (§2), organisations do not appear to be taking steps to improve the status-quo. Indeed, while many companies extensively process user data as part of their business models, it is arguably not in their interests to provide details about their activities and internal processes. Nevertheless, there are various factors that could incentivise (and mandate) better disclosure practices.

First is that disclosures are either already required or being proposed in existing and emerging data-/technology-relevant regulations. As part of this, there is scope for regulators to encourage better practices. Under the GDPR, for instance, regulators have powers to take corrective actions, such as issuing warnings or reprimands where disclosures do not meet the standards required by law (Art. 58(2)). That said, though data protection regulators already provide much guidance as to what is appropriate on a range of topics, the disclosure aspect currently appears under-considered. However, we are already seeing regulators scrutinise the design practices of organisations (e.g. against dark patterns; [14, 26, 48]), and regulators more actively bringing attention on data disclosures could help drive positive changes. There thus appears to be real scope for regulators to actively engage and provide input into what constitutes an 'effective' disclosure, and there is much opportunity for work which further supports the interplay between regulation and design.

Second, industry bodies and trade associations could also assist by defining appropriate transparency-related behaviours for their members, both to ensure certain standards of conduct are maintained and to demonstrate that the industry is proactive in taking their responsibilities seriously. This could also help to reassure their members that they are taking defensible positions surrounding their disclosure practices. In practice, this might take the form of codes of conduct (which in a GDPR context can be approved by data protection regulators (Art. 58(3)(d))), or certification procedures (with certificates granted by regulators or accredited bodies (Art. 58(3)(e–f))).

Third, civil society groups, researchers, and individuals (e.g. the 'informed minority') can also help to shape public policy, by demanding disclosures which better align with their missions and concerns. Community action can apply pressure on organisations (i.e. by critiquing their disclosure practices), and we are starting to see this happen [41].

There are many directions from which stakeholders can assist in realising better data disclosure (transparency) practices. Moreover, note that many organisations will already have the means and capability for designing disclosures, given that usability and comprehensibility is already considered when designing the (user-facing) app/website itself. Indeed, larger organisations often have staff dedicated to designing their product interfaces (though not their disclosures, it would appear), equipped with knowledge of various methods for designing and evaluating user-facing artifacts.

7.2 More than aesthetics

Considering disclosures as *interfaces* opens up many opportunities for how they can be designed, implemented, and evaluated to cater to a wider variety of recipients. As the volume of personal data being processed continues to grow, and as the inner workings of technologies become ever-more complex and opaque, ensuring that data protection disclosures are built to support *useful and meaningful* transparency is a key step toward greater organisational accountability.

Important to note, however, is that making disclosures more aesthetically pleasing will not in and of itself make them more effective (nor was our argument for necessarily 'prettier' interfaces). While we have made the case for disclosure interfaces, it is crucial that their design aligns with the interests of recipients; a key consideration (and indeed, risk) is that they are not misused to whitewash, or otherwise distract from, questionable practices (i.e. 'dark patterns'; [24]). It was for this reason we explored the efficacy of our engineered disclosure across a number of evaluation metrics, including usability, the degree to which users could meaningfully engage with the underlying data, and participant opinions and attitudes. Further, while it may seem intuitive for their design to mirror the interface of the app/website itself—or, indeed, providing the disclosure interface *within* the application—relying too heavily on existing designs might limit the disclosure's potential to provide the useful transparency that disclosures should provide. That is, regardless of their form of presentation, the requirements for disclosures should be directly considered, and methodologies supporting disclosure processes have a clear role to play here. There is also a risk that the design of disclosure interfaces could be misused to steer users away from information they may wish to find (which one could argue the application interfaces often do), or provide too high a level of abstraction.

For this reason, *disclosure interfaces should support, rather than replace more technically-oriented forms of disclosure.* Providing the 'raw' data in addendum to disclosure interfaces can act to support a broad range of aims and objectives that recipients may have. For example, the raw data might act to enable more detailed and expert analysis, and different forms of oversight (e.g. by being machine readable and thus facilitating data analysis or processing). Such data also allows for the monitoring and validation of particular disclosures as well as general disclosure practices, e.g. to ensure the organisations are being forthright (not obfuscatory) in their design. That is to say, disclosure interfaces entail much more than just aesthetics, but rather, are about ensuring that disclosures first and foremost meet their regulatory purpose of supporting meaningful oversight and transparency.

7.3 Broader applicability

Regardless of the various drivers for change, transparency is increasingly on the public agenda. Transparency regimes are a feature of a number of emerging laws and regulations, and regulators are showing an increasing awareness of transparency-hindering design [26, 48]. This looks to be a trend set to continue, as regulators come to determine what designing 'effective' transparency mechanisms should entail. Importantly, this will drive the need for better disclosure practices, ensuring that such regimes are operating as they should: to meaningfully inform.

Importantly, though this work focuses mainly on data protection requirements, disclosure obligations exist in a variety of domains, including areas such as health, finance, and consumer protection [20]. Moreover, record keeping and transparency regimes are increasingly prominent within emerging regulatory frameworks, including the EU's proposed AI Act, Digital Services Act, Digital Markets Act, and ePrivacy Regulation. As such, there will be mounting pressures for more meaningful and useful access to such information. Not only does this mean that disclosure interfaces are of broad significance, but also that there are many possibilities for further research which focuses on a wide variety of different contexts, sectors, purposes, user groups, etc. In this way, one goal of our work is to highlight the need for more attention to be brought towards more effective disclosures in general.

8 CONCLUDING REMARKS

Disclosures are a key means for holding data-driven organisations to account. They can provide information that notifies and informs, supporting review, interrogation, and collective challenge – an ever-important task given the pervasiveness of technology and the issues of power this raises. We have argued the importance of ensuring that such disclosures are *meaningful* for their recipients, to bring about more effective transparency, oversight, and accountability.

Within a technology context, current practices for data disclosures generally appear unfit for this purpose. Towards this, we have demonstrated the potential in treating disclosures as *interfaces*. Our findings show that engineering an interface can lead to disclosures that are better received and more easily understood compared with current practices. By conceptualising disclosures as interfaces to the data (and the underlying organisational processes they represent), there are a myriad of ways forward toward making them more communicative, informative, and useful to their recipients.

Disclosures represent an important transparency mechanism. As concern around organisational practices grows, there is a real opportunity for shaping the design and conventions of disclosures to make them more effective for their recipients. However, this window is limited as disclosure practices are being shaped right now, and urgent action is needed to shift the paradigm towards better supporting transparency and accountability. The time to act is now, before insufficient and inappropriate conventions develop, which ultimately work to hinder and limit the key aim of disclosures – which is to meaningfully inform.

ACKNOWLEDGMENTS

We acknowledge the financial support of the UK Engineering and Physical Sciences Research Council (EP/P024394/1 and EP/R033501/1) and Microsoft through the Microsoft Cloud Computing Research Centre.

REFERENCES

[1] Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, et al. 2004. User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications* 37, 4 (2004), 445–456.

[2] David Alpert. 2020. Beyond request-and-respond: Why data access will be insufficient to tame big tech. *Columbia Law Review* 120, 5 (2020), 1215–1254.

[3] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 3 (2018), 973–989. https://doi.org/10.1177/1461444816676645

[4] Article 29 Working Party. 2018. Guidelines on transparency under Regulation 2016/679. WP260 (11 April 2018).

[5] Jef Ausloos and Pierre Dewitte. 2018. Shattering One-Way Mirrors – Data Subject Access Rights in Practice. *International Data Privacy Law* 8, 1 (2018), 4–28.

[6] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies* 4, 3 (2009), 114–123.

[7] Omri Ben-Shahar and Carl E Schneider. 2014. *More than you wanted to know: The Failure of Mandated Disclosure.* Princeton University Press.

[8] Greg Bensinger. 2020. So far, under California's new privacy law, firms are disclosing too little data – or far too much. https://www.washingtonpost.com/technology/2020/01/21/ccpa-transparency/. Accessed: 2022-05-09.

[9] Mark Bovens. 2006. Analysing and assessing public accountability. A conceptual framework. *European Governance Papers* C-06-012006 (2006).

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[11] John Brooke. 1996. SUS: A quick and dirty usability scale. In *Usability evaluation in industry.* Taylor and Francis.

[12] California State Legislature. 2018. California Consumer Privacy Act of 2018. *Cal. Civ. Code* §1798.100 (24 September 2018).

[13] Jennifer Cobbe, Michelle Seng Ah Lee, and Jatinder Singh. 2021. Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21).* Association for Computing Machinery, New York, NY, USA, 598–609. https://doi.org/10.1145/3442188.3445921

[14] Commission Nationale de l'Informatique et des Libertés. 2022. Cookies: the CNIL fines Google a total of 150 million euros and Facebook 60 million euros for non-compliance with French legislation. https://www.cnil.fr/en/cookies-cnil-fines-google-total-150-million-euros-and-facebook-60-million-euros-non-compliance. Accessed: 2022-05-09.

[15] Akber Datoo. 2018. Data in the post-GDPR world. *Computer Fraud & Security* 2018, 9 (2018), 17–18. https://doi.org/10.1016/S1361-3723(18)30088-5

[16] Alan Dix, Janet Finlay, Gregory D Abowd, and Russell Beale. 2003. *Human-computer interaction.* Pearson Education.

[17] Johanna Drucker. 2014. *Graphesis: Visual forms of knowledge production.* Harvard University Press.

[18] European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* L119 (4 May 2016), 1–88.

[19] Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society* 4, 1 (2018), 1–14. https://doi.org/10.1177/2056305118763366

[20] Financial Services Authority. 2008. Transparency as a Regulatory Tool. https://www.fca.org.uk/publication/discussion/fsa-dp08-03.pdf. Accessed: 2022-05-09.

[21] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for Datasets. *Commun. ACM* 64, 12, 86–92. https://doi.org/10.1145/3458723

[22] Jean D Gibbons and Jean D Gibbons Fielden. 1993. *Nonparametric statistics: An introduction.* Sage.

[23] Robert J Glushko and Tim McGrath. 2005. *Document Engineering: Analyzing and designing documents for business informatics and web services.* MIT Press.

[24] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18).* Association for Computing Machinery, New York, NY, USA, 1—-14. https://doi.org/10.1145/3173574.3174108

[25] Alan Kay. 1990. User interface: A personal view. *The art of human-computer interface design* (1990), 191–207.

[26] Le laboratoire d'innovation numérique de la CNIL (LINC). 2020. Shaping Choices in the Digital World. https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf. Accessed: 2022-05-09.

[27] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19).* Association for

Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596

[28] Martin Moore and Damian Tambini. 2018. *Digital dominance: the power of Google, Amazon, Facebook, and Apple.* Oxford University Press.

[29] Arvind Narayanan, Joanna Huey, and Edward W. Felten. 2016. A Precautionary Approach to Big Data Privacy. (2016), 357–385. https://doi.org/10.1007/978-94-017-7376-8_13

[30] Gabriel Nicholas. 2020. The New Portability: Designing Portability with Competition in Mind. https://www.law.nyu.edu/sites/default/files/The_New_Data_Portability.pdf. Accessed: 2022-05-09.

[31] Austin Lee Nichols and Jon K. Maner. 2008. The Good-Subject Effect: Investigating Participant Demand Characteristics. *The Journal of General Psychology* 135, 2 (2008), 151–166. https://doi.org/10.3200/GENP.135.2.151-166

[32] Chris Norval, Jennifer Cobbe, and Jatinder Singh. 2021. Towards an accountable Internet of Things: A call for 'reviewability'. In *Privacy by Design for the Internet of Things: Building accountability and security.* The Institution of Engineering Technology.

[33] Chris Norval, Kristin Cornelius, Jennifer Cobbe, and Jatinder Singh. 2022. Disclosure by Design: Supplementary Materials. https://github.com/cnorval/disclosure-by-design. Accessed: 2022-05-09.

[34] Chris Norval, Heleen Janssen, Jennifer Cobbe, and Jatinder Singh. 2021. Data protection and tech startups: The need for attention, support, and scrutiny. *Policy & Internet* 13, 2 (2021), 278–299. https://doi.org/10.1002/poi3.255

[35] Jonathan A. Obar. 2020. Sunlight alone is not a disinfectant: Consent and the futility of opening Big Data black boxes (without assistance). *Big Data & Society* 7, 1 (2020), 1–5. https://doi.org/10.1177/2053951720935615

[36] Parliament of India. 2019. The Personal Data Protection Bill. Bill No. 373 of 2019 (11 December 2019).

[37] Frank Pasquale. 2015. *The Black Box Society: The secret algorithms that control money and information.* Harvard University Press.

[38] William D. Perreault. 1975. Controlling Order-Effect Bias. *The Public Opinion Quarterly* 39, 4 (1975), 544–551.

[39] Irene Pollach. 2007. What's Wrong with Online Privacy Policies? *Commun. ACM* 50, 9 (sep 2007), 103–108. https://doi.org/10.1145/1284621.1284627

[40] Presidência da República Secretaria-Geral Subchefia para Assuntos Jurídicos. 2018. Lei Geral de Proteção de Dados Pessoais. Law No. 13,709 (14 August 2018).

[41] Privacy International. 2020. No, Facebook is not telling you everything. https://privacyinternational.org/long-read/3372/no-facebook-not-telling-you-everything. Accessed: 2022-05-09.

[42] Andrew K. Schnackenberg and Edward C. Tomlinson. 2016. Organizational Transparency: A New Perspective on Managing Trust in Organization-Stakeholder Relationships. *Journal of Management* 42, 7 (2016), 1784–1810. https://doi.org/10.1177/0149206314525202

[43] Frank M. Shipman and Catherine C. Marshall. 2020. Ownership, Privacy, and Control in the Wake of Cambridge Analytica: The Relationship between Attitudes and Awareness. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376662

[44] J. Singh and J. Cobbe. 2019. The Security Implications of Data Subject Rights. *IEEE Security & Privacy* 17, 6 (2019), 21–30. https://doi.org/10.1109/MSEC.2019.2914614

[45] Jatinder Singh, Jennifer Cobbe, and Chris Norval. 2019. Decision Provenance: Harnessing data flow for accountable systems. *IEEE Access* 7 (2019), 6562–6574.

[46] Brian Still and Kate Crane. 2017. *Fundamentals of user-centered design: A practical approach.* CRC press.

[47] Cynthia Stohl, Michael Stohl, and Paul M Leonardi. 2016. Managing opacity: Information visibility and the paradox of transparency in the digital age. *International Journal of Communication* 10, 2016 (2016), 123–137.

[48] Ben Wagner, Krisztina Rozgonyi, Marie-Therese Sekwenz, Jennifer Cobbe, and Jatinder Singh. 2020. Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 261–271. https://doi.org/10.1145/3351095.3372856

[49] James Wilson and Daniel Rosenberg. 1988. Rapid Prototyping for User Interface Design. In *Handbook of Human-Computer Interaction*, MARTIN HELANDER (Ed.). North-Holland, Amsterdam, 859–875. https://doi.org/10.1016/B978-0-444-70536-5.50044-0

[50] Janis Wong and Tristan Henderson. 2018. How Portable is Portable? Exercising the GDPR's Right to Data Portability. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 911–920. https://doi.org/10.1145/3267305.3274152

[51] Zoe Zwiebelmann and Tristan Henderson. 2021. Data Portability as a Tool for Audit. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers.* Association for Computing Machinery, New York, NY, USA, 276–280. https://doi.org/10.1145/3460418.3479343