# DATA COALITIONS & ESCROW AGENTS

RADICALXCHANGE

# DATA COALITIONS & ESCROW AGENTS

**Report by**

Jack Henderson & Matt Prewitt

**With support from**

Omidyar Network

**Date published**

May 2023

TABLE OF CONTENTS

# DATA COALITIONS & ESCROW AGENTS

> *"Computer system and network security should be viewed as the other side of the coin of information sharing. What is needed is a systematic technical solution to the problem of sure and convenient access by individuals and groups to the resources they have selectively been authorized to use, at the same time denying access to individuals and groups not so authorized. The solution must include audit trails, authorization channels, and facilities for continuous testing and evaluation. The problem of determining who should be authorized to use what and how is, of course, a separate matter, since it depends on the context."*
>
> – J.C.R. "Lick" Licklider, founder of ARPANET, *Computers and Government (1979)*

Lick's words above are from an essay highlighting what he felt were crucial elements missing from the original TCP/IP protocol for the internet. Alongside open protocols for identity, communication, and payments, he was calling for a protocol for governing information flows. The essay was hauntingly prescient, detailing the colonization by monopolists that would result if not for concerted multi-sectoral investment and public-private partnerships to proactively build such fundamental social infrastructure. Today, however, there is yet renewed hope, as we just come to grips with what the next steps in this agenda might require.

A long list of sociotechnical thinkers like Lick, Helen Nissenbaum, and danah boyd have developed a more nuanced and networked understanding of what it means to be private or public than the standard binary. If we consider the ways information flows organically, its public goodness is best served by complex, socially determined mixtures of revelation and concealment. Thus what we're really after in governing information flows is governing the contextual boundaries of shared information.

We need digital ways to *defend* those boundaries (holding data in the social context in which they exist rather than in alienated centralized repositories or in atomized financial relationships) as well as ways to *gatekeep* those boundaries (co-determining shared permissioning standards and procedures about exactly what is revealed to whom and how it gets revealed).

*Defining* those boundaries, moreover, requires detailed and sound judgment, something like the work of a thoughtful moderator. The data economies of the future have a narrow needle to thread: they must scale up their capacity to exercise this kind of subtle, context-specific judgment, without sacrificing a commitment to democracy.

**THE STATE OF DATA DIGNITY**

The movement for data dignity has come far in the last decade. It was in 2013 when Jaron Lanier published *Who Owns the Future?*,[1] sparking the calls that came later, in 2018, for data as labor and data intermediaries. Then in 2020 RadicalxChange Foundation released the Data Freedom Act, an outline of a regulatory framework that would establish and oversee data coalitions, a new kind of legal and fiduciary entity, with the necessary regulatory scaffolding to support the emergence of data cooperatives, unions, trusts, and other collective forms.

The broad vision influenced a set of policymakers, researchers, and entrepreneurs, many of whom advanced core aspects of the agenda. There is academic work on the relationality of data and why private property is the wrong conception, most notably because it leads to races to the bottom (if someone's going to sell it, you might as well sell it first, and sell it for less). There is work on data trusts that leverage existing trust law, which flexibly manages many kinds of rights and helps ensure institutional accountability and fiduciary duties.

But the core point – that data can't be coherently thought of as a matter of individual control, because it acquires value in massive recombination; that

[1] There was other scholarship on the idea around the same time.

it should be the subject of shared, democratic decisions rather than individual, unilateral ones – remained on the periphery of the public imagination. That is no longer true.

Generative foundation models (GFMs) are resounding proof that there is meaningful power and value, dependent on the public's inputs, for the public to tap into and govern. But getting there will take a few steps, and it starts by getting unstuck from the usual distinction between private data and open data.

**PRIVATE, SILOED DATA**

Breast cancer is the most common cancer in the world and the most deadly among women in 2020, yet it's almost never dangerous when found early. The problem is we struggle finding it: mammograms have a 1 in 5 false-negative rate.

This is puzzling, because breast cancer detection should be the kind of image classification problem that machine learning models are so good at. So what's the issue? A likely part of the problem is the largest available breast cancer datasets are only 3 to 5 million images, which is tiny compared to the many billions of images needed to effectively train GFMs.

Fortunately, there are many billions of mammogram images in hospitals and labs around the world, and if they could pool them all together, they might end this disease. But they have not found ways to do that without compromising medical privacy. They have had little choice but to play it safe, keeping the data locked up in protected siloes.

**OPEN, ALIENATED DATA**

Meanwhile, the only alternative to siloed data seems to be completely open data. When we choose to share data over the Internet, we actually send a copy of the original. And in doing so, we lose control of the information.

Now, arguably, open data has always been a problem. It has reduced our ability to communicate, because we can't be sure that the information we share in a particular setting won't be shared outside that setting. It has also allowed the most powerful actors to easily aggregate and exploit the information.

But the problem has gotten weirder and scarier lately, as just a few seconds of audio can now be used, out of context, to convincingly simulate our voices and scam us. We need to protect against this and analogous attacks. We need new ways to authenticate and protect the contextual integrity of information.

# PART 1. VALUE

To chart a path forward, let's first step back. Where does the value and power of data come from? This is key to understanding how we should govern it.

Think, for a moment, about a natural resource like a river, with a man-made dam built across it, using a hydroelectric turbine to generate electricity. Does the value generated come from the river, or the construct of the dam and turbine?

If we think it comes from the dam and turbine, then all we really need are entrepreneurs and engineers to build them and make them more efficient. But in fact, they are just one way to harness value from the river's natural flow, often in ways that can be exploitative and destructive to the local environment.

If instead we think the value comes from the river, then those humans (and other life forms) who live along and steward its watershed (i.e., its natural polity) need ways to co-govern it as a shared resource, making joint judgments about how it should be used. The issue is most rivers today have dysfunctional governance; they aren't governed by their natural polity. They flow across jurisdictions, which gives multiple actors competing claims to act unilaterally in ways that affect everyone else, such as selling the rights to access the river and build a dam.

*Where does value come from?*

This pattern also applies to interpersonal data and data-dependent technologies. We might see the sources of information – human interactions and relationships – as the river, the crystallization of information into data as the dam, the algorithms and compute that process and apply the data as the hydroelectric turbine, and the resulting insights and intelligence as the electricity.

The insight that data and its value derive from human relationships is a critical starting point. Much work remains to ensure this insight becomes the basis of a better data economy, instead of just hoping we'll be able to clean up the mess in an unjust one. We need ways to protect data's value and harness its ability to act as a public good. Well-governed democracies like Taiwan do this already for managing river basins. We need to do the same for managing information.

**THE PARADOXES OF DATA'S "VALUE"**

Prices are supposed to be able to indicate the value that goods have to society. When the good is data, however, they fail to do this in spectacular fashion. Data that is easiest to use to the detriment of the public (say, real-time geolocation) is often of greatest market value; the value and the hazard are essentially linked. The reason is that the externalities of data disclosure are spectacularly hard to untangle.

Data is relational. It represents information that emerges from a social context, never from a single person; and never abstractable from that context. By trying to think of data as the property of individual actors, we collapse that context, and ignore the way that data affects relationships between actors. We are left with private prices that say nothing about the value the data provides to society as a whole, but only reflect the advantage it gives to one particular party.

This kind of thing is always a problem in markets, but for traditional, non-informational goods, it is less of a problem. Consider the examples below, where markets for pencils and cars help society move somewhat closer to an ideal public use of these goods.

| | Pencils | Cars |
|---|---|---|
| Market Sale | Pencils to those who need them | Cars to those willing to pay |
| Ideal Public Use | Pencils to all who want them | Cars to only those who seriously need them |

In the cases above, society would likely be worse off without the markets' allocative power. By contrast, the market sale of information does not help society approach an ideal use of the information.

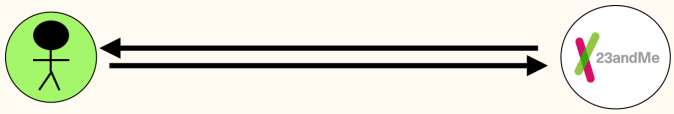
A few examples: Should information about the location of an oasis in a desert go to a lost hiker, or to the buyer who wants to enclose it? Should evidence of sexual preferences in an intolerant society go to a mutual aid and protection consortium, or to the secret police? Should data on individuals' dietary habits go to disinterested researchers, or shareholder-interested corporations, or the state? The market makes no moral judgment; it finds the information's value to the highest bidder, not the public.

| | Oasis Location | Sexual Preferences | Dietary Habits |
|---|---|---|---|
| Market Sale | Information used to enclose the oasis and extract profit | Information used to endanger people or extract ransoms | Information used to manipulate or reify people's consumption habits for private gain |
| Ideal Public Use | Share with thirsty hikers | Share with mutual aid org, or no one | Share with disinterested researchers |

The market price of information reflects uses that are wholly decoupled from any sensible, public-interested use of the information. To find these sensible uses, we cannot avoid making value judgments that escape market logic – like "we want lost people to find water in deserts" and "we don't want anyone to oppress others on the basis of sexual orientation."

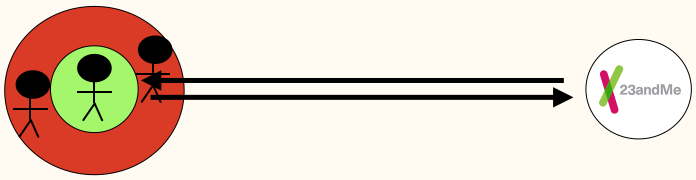
These are easy cases, but sometimes it is very hard to judge which use of information is best for the public. Imagine an individual considering a data transaction with a genetic testing company like 23andMe.

From the perspective of the individual who is interested in their genetic ancestry, this may seem like a good transaction. The individual would share their genetic data with 23andMe in exchange for analysis of that data, which might tell them interesting or useful things about their genetic ancestry. In economic terms, this could have positive social externalities.



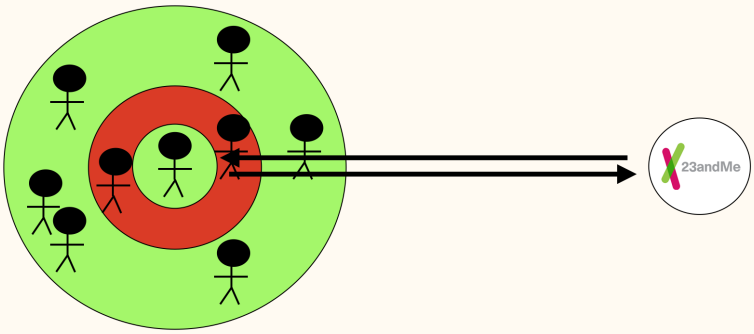

*Good for the individual interested in genetic ancestry.*

However, if we zoom out slightly, we begin to see the consequences of this transaction on a wider circle of people. If we now consider the interests of the individual's family members, who may want to hide their genetic ancestry from third-parties, the same transaction appears to have net-negative social externalities.



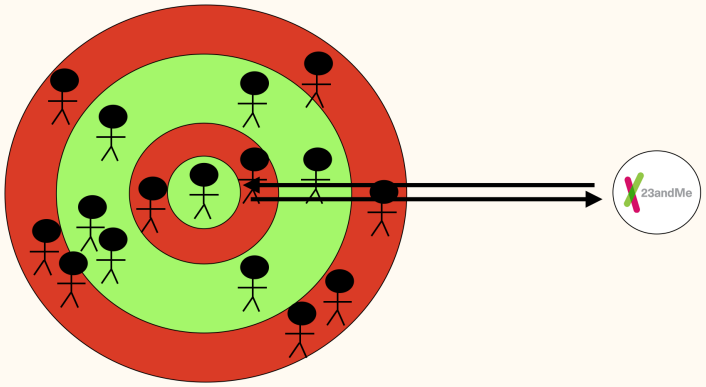

*Bad for the family interested in hiding their genetic ancestry.*

If we zoom out again, we see that this transaction could support a national R&D program analyzing diseases that are correlated with genetics, helping advance medical innovations that may lead to more effective disease diagnoses and overall better medical care for an even wider circle of people. Now the transaction seems to have net-positive social externalities.




*Good for people with genetic diseases diagnosed with this data.*

Zoom out further, and we see that insurance companies may use this information to globally discriminate against people who are vulnerable to similar genetic diseases. The transaction turns net-negative again.

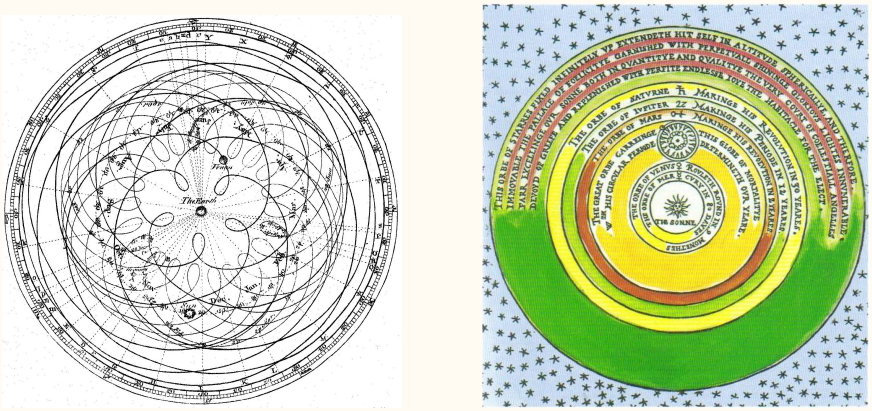



*Bad for people vulnerable to similar genetic diseases who then get discriminated against by insurance companies.*

With a focus on any particular level of context, if we then zoom in or out, the net social utility of the transaction can not only fluctuate, but the actual sign (positive or negative) can change. Simple bilateral transactions in data, such as the one represented above by the two arrows, cut blindly across many other layers of context with social externalities that impact broader circles of people.

Ignoring these effects, and judging the “value of data” based on the utility to any particular buyer, is akin to Ptolemaic astronomy: in ancient Egypt the astronomer Ptolemy tracked the motions of the heavenly bodies, under the assumption they all rotated around the Earth. In trying to predict where they would be in the sky, he added more and more epicycles to try to account for all their anomalous movements. But when we shift the frame,

and assume things orbit the Sun, we get a much simpler picture of what's going on.



*Left: Planets orbit the Earth. Complicated!*
*Right: Planets orbit the Sun. Much simpler.*

What data economies actually orbit is human relationships. Data concerning a person contains deep and predictive insights about other people with whom they associate, so whenever one person unilaterally discloses or withholds it, countless others are affected in important ways.

Indeed if anyone has absolute control over "their own" data, no one does. Strict individual control simply pretends that this problem does not exist. Instead of wishing the complexity away, we need to begin to simplify it, and we can only do that by discerning agents larger than the individual – communities, polities, groups – to drive the data economy. To use data well, data economies need to become *political* economies.

**BUNDLED PUBLICS**

Markets between individual actors can't reflect the value or harm that data has to society as a whole. But states don't have a way of representing the complex networks and tangled group interests that could. The idea behind

data coalitions (DCs) is to represent these complex interests in a way that could anchor the data economy.

DCs try to strike a balance between integrity to distinct contexts and accountability to shared worlds. Each DC would represent a unique bundle of contexts, interests, and purposes, which individuals could opt in or out of.[2] And each DC would be able to take a broad view, accounting for social context but grappling with hard normative questions, to deliver good decisions on behalf of a particular layer of "the public". By bundling publics into intermediate layers – something like a neural network – society could learn to make better shared decisions about who gets information and how they can use it.

In addition to being democratically accountable to the individuals that comprise them, DCs would need to be mutually accountable, to each other. Their decisions would impose complex externalities onto other DCs, in the same way that individual data disclosures impose externalities on other individuals. This is why an overarching regulatory framework is ultimately needed.[3]

**POLICY SUPPORT**

The Data Freedom Act (2020) proposed a policy solution to this problem, mapping a legal framework that would create avenues for conflict mediation, auditing and accounting, and profit-sharing between DCs. It would make DCs important civic actors – a special kind of fiduciary that would help steer the data economy from a position of democratically legitimate power. Corporate actors would need to work with DCs in order to access the best recent data for valuable tasks, such as targeted advertising or AI training.

---

[2] Individuals could choose multiple DCs as partial stewards of their data, even diversifying across sectors that have different incentives and policies for sharing data.
[3] While democratic innovations like citizens' assemblies, deliberation technology, and plural voting can help DCs find areas of consensus and form emergent networks, there is a limit to what these can achieve *between* DCs. The plurality of DCs we should expect will inevitably represent conflicting interests, impose complex externalities on one another, and thus face persistent disagreements and disputes.

While certain data regulations have taken some steps in the right direction, none have gone far enough. The EU's Data Governance Act allows for the formation of a special class of data intermediaries, an important step. But it treats them as marketplaces, not fiduciaries, thus failing to move beyond the traditional, individualistic data control paradigm, or to address the heart of the problem. And the EU's new AI Act, which is focused on safety issues, does not directly address concentration and distortion of power – that is, the subtler sociopolitical hazards created by GFMs.

Unfortunately, without comprehensive policy support, DCs have struggled to gain momentum. They lack ways to prevent races to the bottom, to deliver diverse but collective representation, and especially to get along with other collective data entities. And because DCs would represent a profound change to the legal landscape, we simply can't assume the support will materialize soon. The necessary legal changes would run contrary to major entrenched interests unless watered down unacceptably.

Ultimately, DCs will reach a scale that will require state-level support, but to get traction in the political world, we need to do whatever we can to establish and scale DCs to a level that will earn state attention.

We survey these possibilities next. What would DCs operating without the support of the state need to do? What technical problems would they need to solve? How can we start building them?

# PART 2. DATA COALITIONS WITHOUT THE STATE

DCs need clever ways to defend and gatekeep contextual boundaries of shared information. Fortunately, there have been huge advances in technical tools and architectures that can help enable this. The work of, for example, the OpenMined community applying privacy-enhancing technologies (PETs) has the potential to upend how information flows online – introducing the possibility of single-copy data that can still be approved for use by others, who are only able to learn concise answers to specific questions.

## DEFENDING BOUNDARIES BY WITHHOLDING CREDIBILITY

When it flows frictionlessly across social boundaries, information loses its context. This violation of context explains why information flows that should be useful can instead become exploitative. And it's core to why bundling data interests is so important: we need ways to define and defend the boundaries of shared information, such that the democratic authority over what happens to every datum is aligned with the social context it comes from.

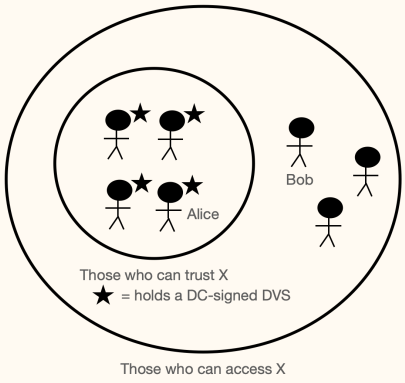
A key challenge for DCs here actually comes from within. How can they dissuade members from unilaterally disclosing data outside the context of the coalition? Preventing disclosure is, in a strict sense, impossible: If someone can access data, they can communicate it to others. And even if their access is tightly controlled (e.g. through DRM), there's always a way to copy and paste, or take a screenshot, or record with another device, and smuggle information across the contextual boundary.[4]

[4] This is called the copy problem, and it helps create disclosure races to the bottom.

But the problem is not hopeless. There are emerging methods by which DCs might prevent credible disclosure (at least where the information is not independently verifiable). This matters because data gets much of its value from being credible – indeed, a primary role of DCs is guaranteeing the credibility of the data they govern.

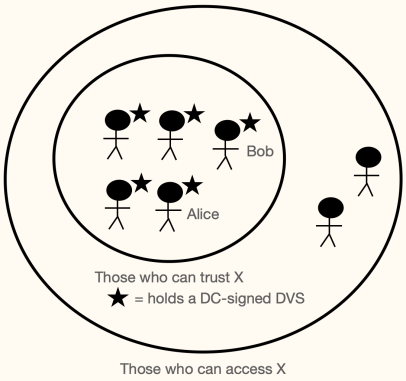
Credibility also operates with contextual boundaries. People don't directly verify most of the information they come across, but rather trust its accuracy insofar as they trust its provenance and attestations. Such access and trust is usually permissioned socially by one's peers.

There are powerful ways to imitate this digitally (thus decreasing our reliance on massive one-to-many verifiers like big tech platforms). One is with designated verifier signatures (DVSs), which can enable DCs to control who can access credible proofs about data provenance. DVSs work like this: if a DC wants to communicate to one of its members (Alice) that a datum (X) is true, it would actually issue Alice a zero knowledge proof of a compound statement: "Either X is true, or I am Alice." Alice knows she did not make the proof, so she knows X is true. But if she conveys the same proof to Bob, he won't know which condition is true. Her attempt at unilateral disclosure therefore isn't credible.

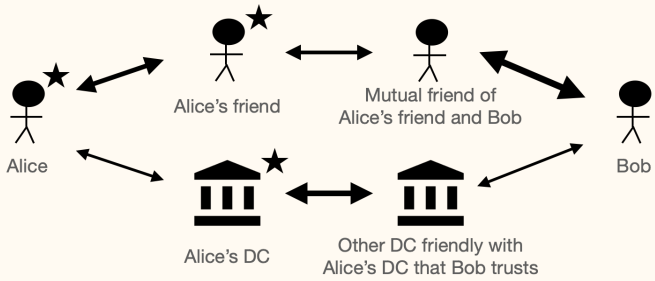



Instead, for Alice to credibly disclose to Bob that X is true, she'd need democratic approval from the DC, who would then sign and issue a new DVP to Bob that "Either X is true, or I am Bob." Bob would thus be brought

inside the boundary as another “designated verifier” who now believes the claim is true.




Alice and the DC could also use social attestations, deploying their own reputations and social network to further prove the claim to Bob, since a claim is more credible if those attesting to it are more trusted or independent. They could seek out paths of trust to Bob, and even stake or put a “lien” on their trustworthiness.[5]




New designs for community currencies could prove to be important complementary systems. If one must acquire and hold a DC’s exclusive, closely-traded currency in order to gain smooth technical access to its information, unauthorized unilateral disclosures that allow outsiders inside the boundary would become much costlier.[6]

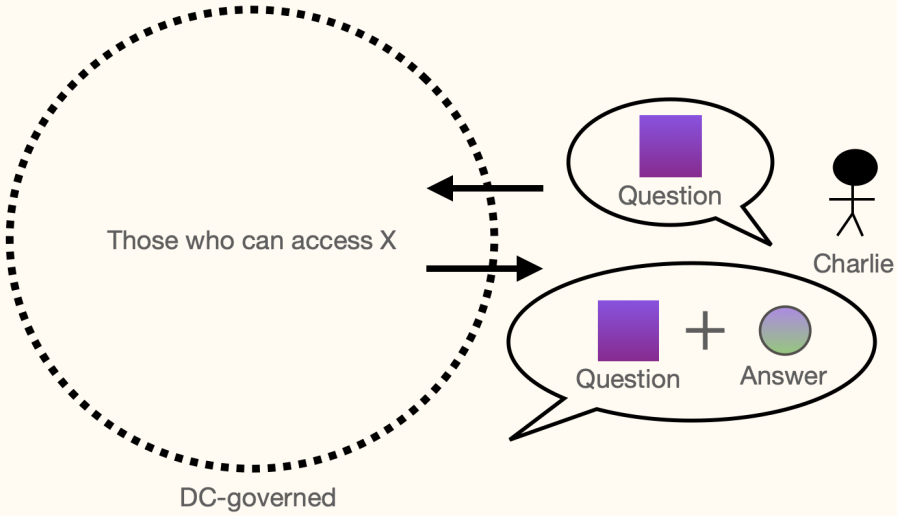
[5] This could even be done at a quadratic rate with correlation coefficients. They would be vulnerable to a “burn” if their claim is proven untrue; otherwise, the stake would be gradually released, plus interest to reward the deepening of trust.
[6] A DC’s data could be the assets backing its community currency.

## GATEKEEPING BOUNDARIES WITH PETs

The DC trusts Bob. They brought him inside the boundary and gave him total access to the raw and verified datum, X. But what about Charlie, who they don't trust as much; do they let him inside or not?

Using various privacy enhancing technologies (PETs), the DC can make the boundary semipermeable. So depending on the context and how much they trust Charlie, the DC can programmably adjust the granularity of his access permissions.

Say Charlie wants to use X to help train his machine learning model. First, he doesn't actually need to come inside the boundary to do that. He can interact with the DC through federated learning, where Charlie sends his model to some remote machine that is governed by the DC; and they can use fully homomorphic encryption or actively-secure multiparty computation to allow Charlie's model to run specifically approved computations over encrypted data. Charlie never gets inside the boundary. His model is allowed in, but it's largely "blindfolded" and only able to ask a specific question; all that comes out with the model is a specific answer.




However, even though Charlie never sees X, we can't really say the DC maintains full control of it. The concise answer that left the boundary

represents a limited “leak” that went to Charlie and might now be used to answer other, hard-to-foresee questions. It’s the start of a new information flow downstream of X. The DC will want to mitigate this to the extent possible.

One risk is that if they share too many answers about X, Charlie or someone else might find a way to use those answers to reverse engineer X itself. With differential privacy they can quantify this probability of privacy loss and programmably lower it by adding a degree of random noise to Charlie’s answer. In other words they can meter the amount of privacy erosion, or sensitive information leaking across their boundary.

And to ensure that the leaked information is put to the best use, they could auction off use-licenses using an innovative property regime like Partial Common Ownership. For example, if the DC really values privacy, they might choose to grant only one use-license at a time, with strict bounds on possible uses, and only among a particular set of trusted data scientists. If Charlie were among that set, he would be able to purchase the license.

**USING JUDGMENT**

PETs can help mitigate information leakage, but after information leaves the DC, all bets are off. Consequently, PETs can’t solve the entire problem alone: they don’t automatically enable the DC, or Charlie, to make informed value judgments about the risks and benefits.

First, PETs don’t enable DCs to interpret the possible downstream consequences of answering Charlie’s question, before deciding whether to answer it. DCs can only “exercise judgment” in advance, pre-setting automated policies that approve or deny data uses based on their stated interests in things like privacy, compensation, and social goods.

Second, PETs don't enable Charlie to interpret whether the data helps answer his question, before deciding whether to pay for it. He can only search for signals that suggest X is broadly relevant.[7]

This leads to either paralysis or profligacy: either the parties (the DC and Charlie) will not agree to any information exchange because they don't know what information is included in it, or they will agree without actually knowing exactly what they're agreeing to.

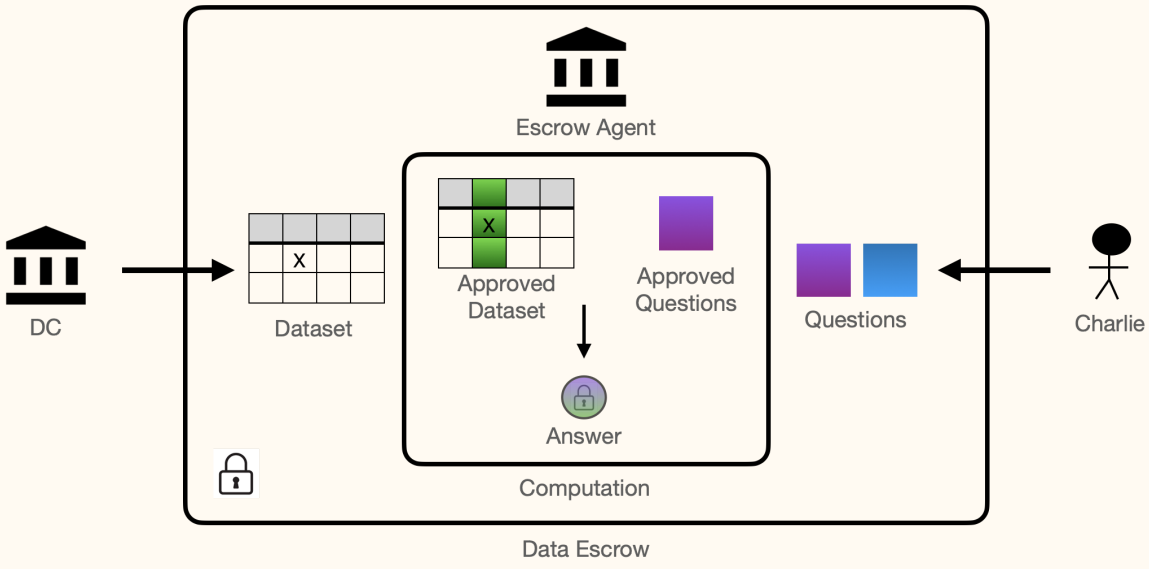
This doesn't mean we should throw up our hands. It means we need to supplement the affordances of PETs by maximizing the role of detailed and nuanced judgment about the likely consequences of information sharing. This is possible, but it needs to be institutionalized. Here's how.

[7] Charlie might see that the DC has a cryptographic signature proving it's related to a legitimate institution; or a DVP that proves its provenance; or regulatory certifications; or reputation-based reviews that X helped other models test well on validation data sets.

# PART 3. ESCROW AGENTS

An escrow agent (EA) is a neutral third-party, which could be human or automated, that operates a data escrow. Alone inside the escrow, the EA can securely surface all relevant information and facilitate responsible value judgments about data sharing.

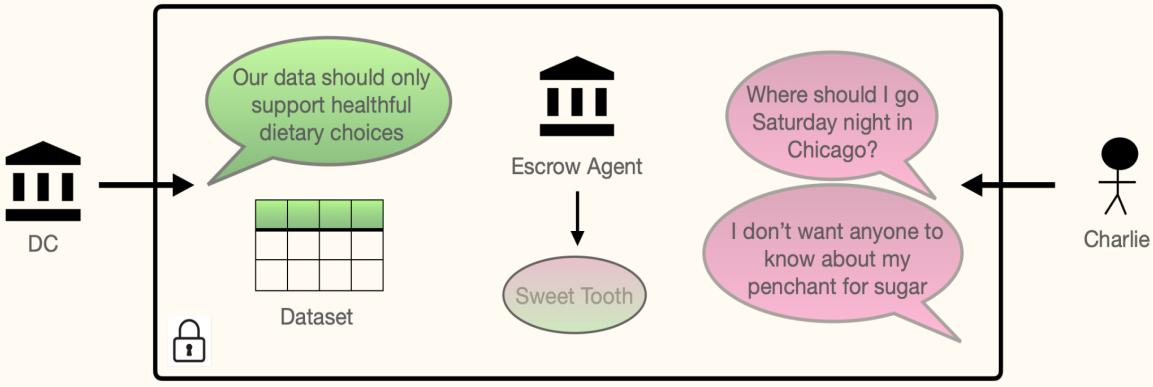
To set one up, DCs can upload data to an escrow in "enclave mode," enabling the EA to run approved computations over them, but never leak the answers without explicit permission. Likewise, data users can submit their models and questions to the EA without leaking information or "tipping their hand" to DCs.




For example, let's say Charlie has a model trained on his past preferences, and he's in Chicago for the weekend, so he wants his model to learn about the Chicago bar and restaurant scene to tell him what he's likely to enjoy.

Charlie shares his model with an EA and asks: “Where should I go Saturday night in Chicago?” His model knows how much he loves sugary soda, but Charlie thinks this preference could be used against him. So he also tells the EA: “I don’t want anyone to find out about my penchant for sugar.”

The EA searches for good Chicago data on behalf of Charlie, and runs his query on a dataset that would tell him to go to a place called Sweet Tooth. However, the dataset is controlled by a DC committed to ensuring its information is only used to support healthful dietary choices.




Should the answer be conveyed to Charlie? Who decides? Paradoxically, both parties should influence the decision – but they can’t make the decision themselves without seeing information from the other party, which would render the decision moot.

Before Charlie will choose to run his query, he wants to know whether the answer to his query will be genuinely useful to him. But he shouldn’t see the answer before the decision is made, because it might reveal information the DC doesn’t want to reveal (that Charlie should go to Sweet Tooth, which would conflict with their mission).

For the DC to approve the query, they want to know the information they’ll be providing aligns with their values. But they shouldn’t see their answer together with Charlie’s question, because it might reveal information

Charlie doesn't want to reveal (they might infer that Charlie likes sugary soda).

If either of them were to see the answer, only to decide the answer shouldn't be shared, it would be too late. We can't uncrack a cracked egg.

**THE ESCROW AGENT MUST DECIDE**

Thus, the only way for Charlie to ask the question, and for the DC to exert control over whether the question is answered – without information escaping contextual boundaries – is for a third-party to decide if the question should be answered.

The EA would be able to learn the values and preferences of the DC and Charlie, and have access to both the question and the answer in escrow. They would be positioned to weigh the possible costs and benefits downstream, and make a fair and informed judgment on behalf of everyone.

Today, stewards of sensitive data lack institutions that can navigate such analyses, which leaves them in the dark about the data's social value. They either play it safe, keeping the data in silos; or play it loose, pursuing self-interest without regard for social consequences. EAs can make responsible sharing a viable possibility for more data.[8]

**ACCOUNTABILITY**

EAs could become quite powerful, and trustworthy appeals processes and quality controls should do most of the work of legitimating them. But when an EA is reasonably suspected of an unjustified exercise of authority, their work may need to be opened up to public scrutiny.

To even enable the possibility of suspicion, the public must know that authority has been exercised in the first place. Dorota Mokrosinska, a

[8] This also heightens incentives for data to be cleaned, curated, and well-documented, as it leads to more matches with data users like Charlie and thus more benefits. Such cleaning could be a service DCs provide their members, the quality of which may be another way to differentiate them.

political philosopher, calls these shallow secrets – secrets that are at least known to exist, even though the content is unknown. Deep secrets, those not known to exist, are impossible to raise suspicion about and hold to account.

By design, EAs must keep a tamper-proof log of everything that happens inside the escrow.[9] This means oversight bodies, such as other EAs or regulators, could be granted discrete access to a log in order to audit an EA's decision. The audit would determine whether the EA's actions were within the bounds of its authority, with respect to the interests of all the parties it was authorized to represent. To mitigate the recursive enforcement problem, such oversight should also be subject to gatekeeping, such as permissioning remote access to proofs about the log but not to the log itself.

**MINIMUM DISCLOSURE**

So would the EA let Charlie know about Sweet Tooth? It might depend on the EA's duty to disclose information to the DC about its data uses.

For example, if the DC requires disclosures about the asker, the question, and the answer, then the EA might decide not to run the query – knowing it would compromise Charlie's interest in privacy. If the DC only needs to know the asker and the answer, then the EA might decide to run the query – because without knowing the question, the disclosed information (Charlie, Sweet Tooth) does not necessarily reveal that Charlie likes sugary soda.[10]

This example is fairly simple, so a well-calibrated automated EA might be able to make the right judgment; they are certainly going to be needed to help navigate and make sense of complexity. But their judgments must be backed by institutions that can take responsibility for bad decisions.

---

[9] The log tracks inter alia all access permissions and attempts to access data, ask questions, and give answers.

[10] Once Charlie is linked to places like Sweet Tooth enough times, DCs will start to infer his preference for sugar. This is similar to privacy erosion exceeding a certain threshold; EAs will need tools to help them monitor and protect against this.

## ESCROW AGENTS CAN HELP GOVERN DATA COALITIONS

One way judgments may become complex is if Charlie's question depends on a combination of data from multiple DCs. This may not seem much of a problem: it might be harder for Charlie to navigate, but we can still assume all the DCs would fairly represent their members' unique contexts and interests.

Yet, as the 23andMe example illustrates, the normative evaluation of a disclosure depends on the scale of the "public" that evaluates it. No single DC would be in a good position to zoom out and grapple with the aggregated externalities of all the transactions together, making the joint exercise of good judgment a tall order.

This is why any framework for DCs requires a higher authority to oversee them and arbitrate conflicts. While states naturally have the broadest reach and deepest suite of enforcement powers, EAs could also play an important role in governing relations between DCs.[11]

DCs could agree to convey their interests and priorities to an EA, and give it the authority to exercise judgment on behalf of everyone. In fact, rather than a single actor like the state trying to regulate complex data ecosystems, a network of EAs jointly regulating them with varied scope and expertise should go a long way.

But EAs would also benefit from an overarching framework: harmonizing their policies to some extent seems important in order to avoid their own races to the bottom that could occur if powerful interests are able to shop for favorable regulatory regimes.

[11] We suspect the problems created by the lack of an overarching regulator will eventually prompt states to consider DCs worthy of attention. In the meantime, other powerful institutions can establish DCs within their walled garden. In the closed universe of a monopoly platform, for example, the platform itself could play a state-like role, regulating the relationships between DCs operating within it.

## ESCROW AGENTS CAN HELP GOVERN GENERATIVE FOUNDATION MODELS

Similar to us moving beyond a simple choice between private and open data, EAs can help us move beyond a similar divide with regard to GFMs.

The divide is something like this: many don't want these models to be concentrated in so few hands; and they want transparency into these powerful forces starting to shape our lives and livelihoods. But others think open-sourcing is unsafe because the technologies are dangerous; that it's also futile because powerful actors will always have an advantage in compute power; and they also think these models are inherently opaque, that it's part of the deal, and we can't understand their inner workings no matter how hard we try.

As philosopher Seth Lazar argues, this black-box nature makes it virtually impossible to justify their exercise of power, because we can't see *how* they are exercising power or *who*/*what* is behind them. So the choice seems to be either demand impossible or unsafe transparency, or simply accept illegitimate opacity.

This came up on a recent podcast with Ezra Klein and Danielle Allen when they were discussing alignment assemblies:

> EZRA KLEIN: *… If the assembly comes up and says, hey, we want this to be legible and we want to be able to know what the system is doing. And the people say — who are making them say, that's not really possible. We don't know how to do that in these models. To do that is to basically break the whole approach. Who's right and who's wrong there? ...*
>
> DANIELLE ALLEN: *… if assemblies say, we want to know what's going on, and then technologists say, well, you can't if we have systems of this kind, then there's a thing to negotiate, then that's the work to do: what's the relationship then between that desire to know — that need to know — and the fact that technology can't deliver it? … the fact that there's a*

> *discrepancy between the values people may articulate and what technologists think they can deliver is the beginning of the work, not the end of it.*

EAs could help resolve this discrepancy. If they can develop a legitimate capacity to interface between people working out their values and aspirations around these technologies and the more technical process of executing on them, we can have not just democratic input but democratic representation with regard to GFMs and their interactions with the world.

As third-party auditors with special permission to see inside GFMs – to see what data they're ingesting, what's going on inside them, and why – EAs would have the necessary context to effectively interpret democratic inputs. They could audit complex questions about which data are being used, and how, and whether a model aligns or not with public values.[12] They could track precisely how diverse data from DCs improve GFMs, ensuring that DCs are fairly compensated and maintain a fair stake in those models. They could even govern the interactions between GFMs and community fine-tuned models.

Operators of GFMs will hopefully see it's in their interest to invite EAs into their systems. EAs would represent them too, concealing enough information about how their models work to ensure security and safety. They could also unlock rich datasets, anticipate and reduce social risks, and even help publicly legitimate those models.

[12] When they don't align, the EA would deny the data flow; they might also conceal from the model operator whether the DC has good data for them or not, which could otherwise create perverse incentives for the operator to acquire the data by other means. In addition, if EAs are not provided sufficient documentation of a model and its datasets, they may decide to automatically deny its use.

# NEXT STEPS

The moment demands experimenting and investing along these lines. So how could we get started?

DCs could emerge naturally from data stewards that already exist, such as data trusts, unions and cooperatives, government agencies with citizen data, universities with research data, or businesses with customer data.

To establish themselves as DCs, they would need to define a clear mission on behalf of their members as well as the tradeoffs they would make between privacy, monetization, control over downstream uses, and other member interests.

They could begin using PETs by deploying a domain server or network node with OpenMined. Some government agencies are already at this point, and many data users already use OpenMined's PySyft library to learn from private data.

EAs, however, should be data-focused but non-data holding civil society organizations. To avoid conflicts of interest, they should be wholly separate and independent from any DCs or data users. Like a jury of peers, they should be neutral enough to be fair and proximate enough to understand the context.

The Open Data Institute and MyData Global are two possible examples: they hold certain values, could assemble the relevant expertise, and might reasonably fit it under their missions to create a new EA arm of their organizations. Other third-parties could also emerge as EAs for-hire.

Establishing a data escrow would require them setting up a secure enclave, but more work is needed to pair designs for data escrows with PETs infrastructure like OpenMined; eventually EAs might be able to register with organizations like OpenMined to deploy a kind of meta-network node, which DCs could then join.

The exact governance, social permissioning, and accountability frameworks that EAs will need is also an important direction of research. For instance, it seems promising to develop them with an ethic of prudent vigilance, but there are many questions here.

Please reach out if you're interested in working through them with us.

**ACKNOWLEDGMENTS**

Thanks to Raul Castro Fernandez, Andrew Trask, Divya Siddarth, Nick Vincent, Glen Weyl, Shrey Jain, Paula Berman, Saffron Huang, Kyoko Eng, Danielle Allen, Sille Sepp, Puja Ohlhaver, Sebastian Benthall, Maria Savona, Anouk Ruhaak, Daron Acemoglu, Sylvie Delacroix, Neil Lawrence, Kate McCall-Kiley, and Travis Moore, with support from Omidyar Network.