



The discussion wove through topics such as provenance, authenticity, valuation, and the complex web of reputations, processes and contracts that make up the market.

At the time I did a quick search on Amazon to see what books had been published about the space.

I flicked through a couple of samples and that was that.



A day later @art\_collector\_a— a twitter marketing account— had started to follow me.

What is unusual, before doing the search I had cleared my browser; I use different email addresses, and take various other steps to hide and normally use a VPN.

There were no followers in common either.

Except for name and an IP address geolocated to where I live there isn't much tying my Amazon and Twitter activities together.

Yet it appeared some quality matching was going on.

Of course it could just be coincidence but I'll never know.



### “more data begets more data”

The nature of data makes the antitrust remedies of the past less useful. Breaking up a firm like Google into five Googlelets would not stop network effects from reasserting themselves...A radical rethink is required:

- antitrust authorities need to move from the industrial era into the 21st century.
- loosen the grip that providers of online services have over data and give more control to those who supply them.

This month’s Economist issue is about— “The World’s Most Valuable Resource...”— and makes the case that a Networked Data Economy is not a goods economy.

That there are economies of “data” scale in quantity and quality that data companies enjoy.

Quality here is not your typical data wrangling. It points to the diversity and

scope of gathered data.

The Amazon and Twitter scenario I just mentioned.

The gist is “More data begets more data.” The author therefore argues, as a matter of principle, Regulators need new perspectives on antitrust concerns.

Aside from the idea of a new lens, the author also makes the case that you can't break-up these networked data businesses.

Others would emerge pretty quickly.

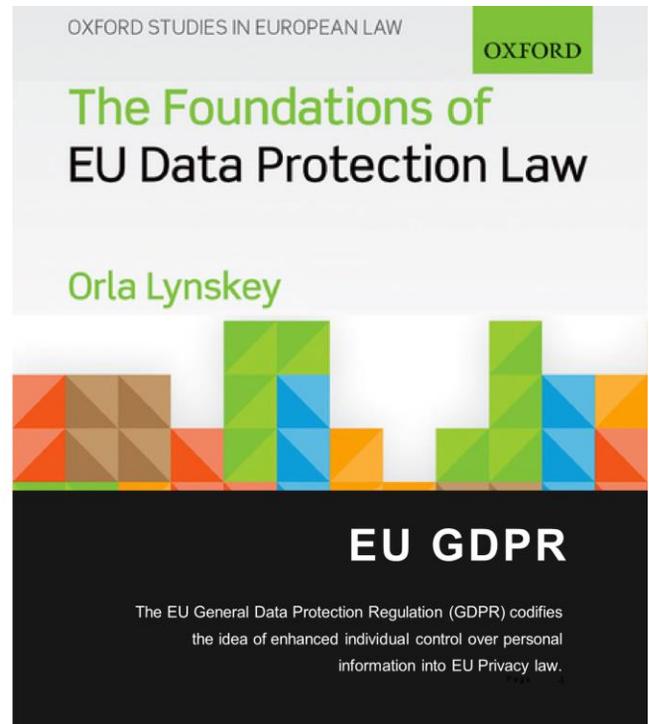
Instead he argues, another antitrust principle “is to loosen the grip that providers ... have over data.”

Moving control to those who provide it.

I'll presume for the moment he means the individuals whose behaviour is the source of the gathered data. Us!

- CONSENT
- DATA PORTABILITY
- RIGHT TO BE FORGOTTEN
- RIGHT FOR EXPLANATION

We don't architect the "ethics of data" into our technology stacks yet, so the practice needs to evolve.



The idea of putting control of data into those who co-produce it is nothing new.

The principles of informed consent and control have a long history in privacy, health ethics, and scientific protocols.

On the Internet it is “Whatever!”

The EU General Data Protection Regulation codifies the idea of

“\*\*Enhanced\*\* individual control over personal information” into EU Privacy law.

It does not grant “property rights” over personal data per-se but it does recognise individuals alone cannot be responsible for the protection of their personal data in an era of data proliferation.

It is not clear who would “own” the data, the environment is too complex, and there is a lot of good value to be gained integrating that data.

As such in addition to consent, the General Data Protection Regulation introduces new concepts into law; \_Data portability\_, a \_right to be forgotten\_, and a \_right for explanation\_.

These are big ideas; Important to privacy, competition, and democracy

I doubt a company that handles data in any reasonable volume will know how to instrument systems accordingly.

We don't know how to build "ethics into software." Yet?

# Information Paradox

As our environment becomes more complex we need more information to find the signals, to produce few bits of knowledge.

But sharing, pooling and reusing data comes with many risks.

GREAT FEATURES



Kafka's Trial



Bentham's Panopticon



Orwell's Newspeak



Page 5

Quality data is valuable. Quality information is valuable.

Information is used to produce knowledge that with tacit understanding— knowhow— we or some agent can act upon it.

This applies equally to commercial, social, health and environmental knowledge as it does to personal knowledge. With better insight we can

orchestrate better outcomes.

Paradoxically as our environment becomes more complex we need more information to find the signals, to produce less but higher quality knowledge.

But sharing, pooling and reusing data comes with many risks.

We don't want to be indefinitely detained in Kafka's trial, or normalised under Bentham's Panopticon, or become mules in an Orwellian infowar lathered with trump-speak.



## Privacy xor Science

### Privacy xor Society

Should we acknowledge the risks and just abandon the effort?

No! We shouldn't stifle scientific endeavour, erase our historical evidence, or ignore social ills because sharing data is just too risky.



Equally we don't want to stifle scientific endeavour, erase our historical evidence, or ignore social ills because sharing data is just too risky.

Should we acknowledge the risks and just abandon the effort?

I don't want to; Privacy and science should not be mutually exclusive.

# DISTRIBUTED LEDGERS SMART CONTRACTS ORACLES

.....

Despite an experimental status, the associated hype, and a lot to workout, the ideas and technology of Distributed Ledgers, [Legal] Smart-Contracts and other enablers are important to data reuse and digital trust.

→



In late 2015, the Economists cover emblazoned with the headline “The trust machine” as the issue went on to discuss the possibilities introduced by the technology that under pins bitcoin. Despite the hype there is truth in the usefulness of this technology and strong momentum.



## TRUST IS A BET

an inter-temporal choice.

A bet about future outcomes and the risks associated with them. In our case, will others keep their promise to reuse data appropriately.

Trust is important because if we want to escape the privacy and complexity paradoxes, we need to trust each other to responsibly reuse the information that we co-create. This is the fundamental barrier to greater responsible reuse; the lack of control and with it, a lack of trust.

Page 8

My interest is in Digital Trust;

how we achieve trust in information and architect systems when every bit looks like any other.

Trust is important because if we want to escape the privacy and complexity paradoxes, we need to trust each other to responsibly reuse the information that we co-create.

This is the fundamental barrier to greater responsible reuse; the lack of control and with it, a lack of trust.

Not just one step from source but every downstream derivative form.

As you're all aware there is a strong and well established data market. And as a software engineer I have contributed to its construction. What I'm arguing for here is to bring transparency, accountability and ethics into the equation. Spanning the entire stack, not just the top layers.

Downstream use and reuse of data will cross many trust boundaries; Hence the way I'm using the term `_Trust_` is to mean an inter-temporal choice.

A bet about future outcomes and the risks associated with them.

In our case: will others keep their promise to reuse data appropriately.

There are a number of technology solutions that can help us but enabling trust is **\*\*NOT\*\*** just about implementing effective technology.

# Governance Architecture

## THE MARKET

Market fundamentalism is selfish, competitive and extractive.

## STATE MANDATED

State rule is inherently powerful and coercive.

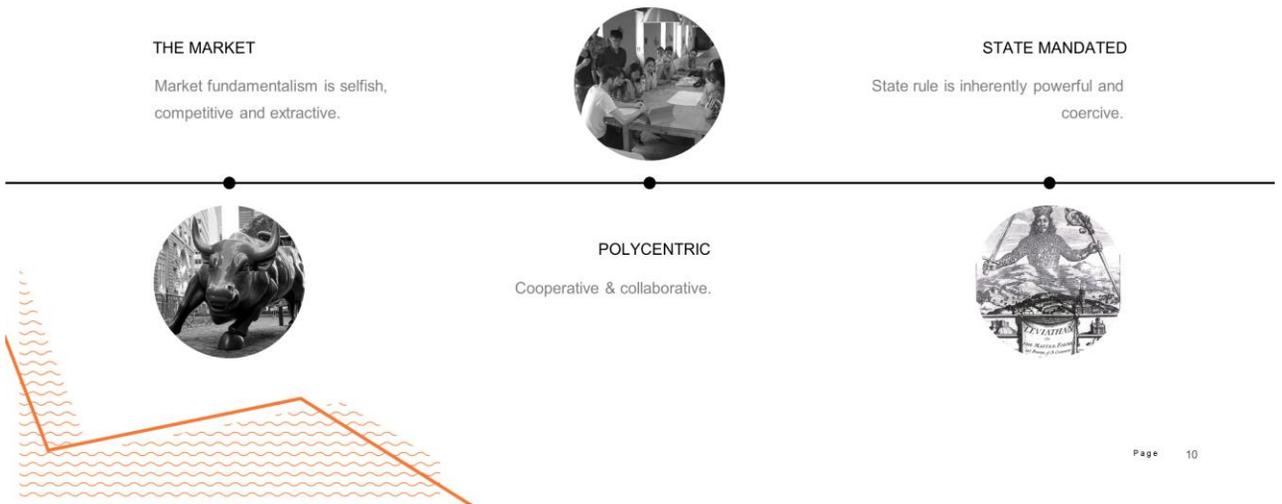


Existing modes of high-level institutional forms— the market or state are unlikely to be effective here.

Market fundamentalism is selfish, competitive and extractive. State rule is coercive.

Rules-of-Form from high often appear detached from the Rules-of-Use found at ground level.

# Governance Architecture



In contrast Elinor Ostrom's research into polycentric governance showed that the development of rules **\*\*within\*\*** complex group arrangements and communities could out perform free-markets or state regulation.

# POLYCENTRIC

Governance

Cooperative & collaborative.

Development of rules within complex groups and communities  
can out perform free-markets or state rule.



Page 11

---

Rules developed by the participants.

Rules that were understood by the participants.

Rules that allowed for conflict resolution between participants.

Rules governed by the participants.



## Elinor Ostrom

Governing the Commons

.....

Research work showed that is groups could self-govern without tragedy in an efficient sustainable manner.

Page 12

Elinor Ostrom became the first and only women in 2009 to receive a Nobel prize for economics. Awarded for her analysis of governance models, especially the commons. Her work showed how firms and other communities could facilitate economic efficiency from within.

That is groups could self-govern without tragedy in an efficient sustainable manner.

Scale was the biggest challenge. But with newer technology that may not be such a problem.

The Common is where we should look to enable data reuse.



## KNOWLEDGE COMMONS

.....

Elinor Ostrom's work focused on natural resources but her ideas and principals are also applicable to knowledge.

Ostrom's work focused mostly on natural resources but her ideas and principals are also applicable to knowledge.

Five examples of these are seen in the Internet, the W3C, Open Source Software, Wikipedia and the Creative Commons.

The five examples are focused on technology protocol standards,

software, and cultural production; some situated within an existing regulatory structure of intellectual property.

Data reuse doesn't fit into that particularly well. But data reuse does fit within the scope of contract and arbitration regulations.

Without prejudice to possible direction there should be:

- A focus on building network trust
- With a commons co-design and commons governance model
- That functions with the use of contracts
- Where dispute resolution works through arbitration
- And uses a technology design that reduces transaction costs
- While also embedding consent directly into data management

- That provides an ability to monitor compliance.
- Adding versioning, process lineage, and provenance, for accounting and audit.

These are some of our Protocols to scale up safe reuse of data.

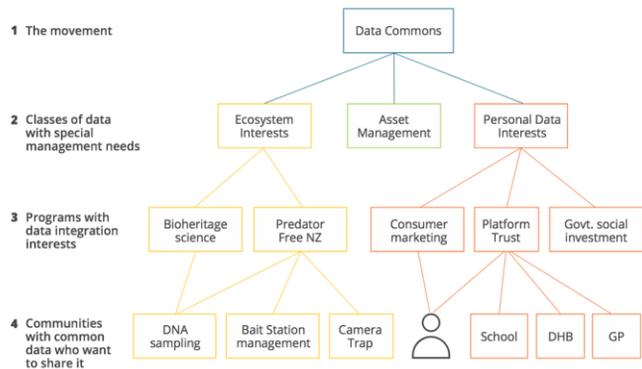
Turning a risk dilemma into a feedback loop.

# DATA COMMONS

.....

A Data Commons, simply put is a way that communities of practice can agree on how to reuse their data. Add to the value to data, and manage the risk of its reuse.

→ THERE ARE MANY DATA COMMONS



A Data Commons, simply put is a way that communities of practise can agree on how to reuse their data. Add value to data, and manage the risk of its reuse.

These could be privacy risks, or process risks, or use risks. Equally they can be publication risk.

The Data Commons approach focuses on solving risks across competing

interests and giving control to the participants.

In this sense \_Protocols\_ form the \_community contract\_.

In naming the Data Commons, I'm not talking about a singular entity either.

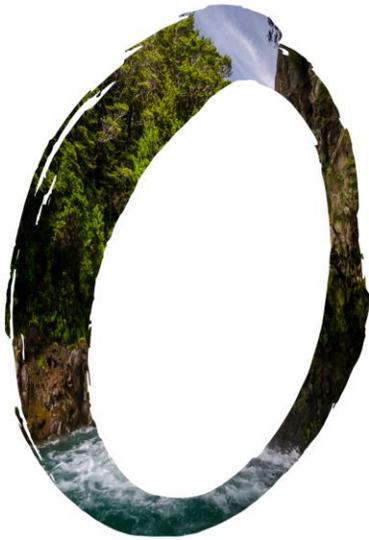
Rather the Data Commons are many. Organised around the conversations needed to define the \_Protocols\_.

We can frame the challenge into four points of focus:

- The movement.
- Classes of information.
- Specific programs that want to reuse datasets.

- Communities that want to share data, but aren't likely to reuse it.

We can think of it as an internet of agreements woven together.



## FASHIONING OUTCOMES

### ✓ PRODUCTIVE USE OF INFORMATION

The strength of our economy can be determined our capability to put information to use.

### ✓ SOCIAL CAPITAL

Building the connections and trust they engender is a key factor in the health of economies.

### ✓ OUTCOMES FOCUSED

Value can be a rather confusing construct. So we're more interested in Outcomes. With more productive use of information we can orchestrate better results.

With more productive use of information we can orchestrate better outcomes.

To get that knowledge we need quality information. That means handling the challenges of Volume and Variety at Velocity.

This also implies the strength of our economy can be determined by our computational capacity. Our capability

to put information to use.

Coase's work on transaction costs in the fashioning of institutions explored this idea in the 1930s. More recently Social Capital — the connections and trust they engender— is recognised as a key factor in the health of economies.

We often discuss these challenges and principals in the context of Value. But what is value?

In reality a culture defines value, hence value can be a rather confusing and limiting construct. In the Data Commons we're interested in Outcomes.

Personal Outcomes, Health Outcomes, Environmental Outcomes, Social Outcomes, Policy Outcomes.



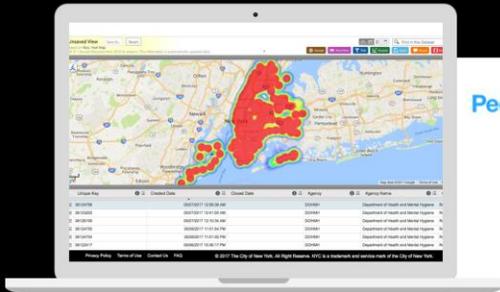
Many of the Smart health related devices— such as the iHealth gluco-monitor or a Fitbit— owe their market existence to the idea of achieving an outcome. We hire a Fitbit to monitor and coach our physical wellbeing.

While gathered within proprietary silos, Data Portability makes that information available.

Aside from human factors in usability,

consent and control issues, there is no technical reason preventing such data being used for broader health outcomes; For the individual and various cohorts.

Nike+ pioneered this, adding social dynamics to encourage achieving goals.



PeerJ

## Rat sightings in New York City are associated with neighborhood sociodemographics, housing characteristics, and proximity to open public space

Michael G. Walsh

Department of Epidemiology and Biostatistics, School of Public Health, State University of New York, Downstate, Brooklyn, NY, USA

### ABSTRACT

Rats are ubiquitous in urban environments and, as established reservoirs for infectious pathogens, present a control priority for public health agencies. New York City (NYC) harbors one of the largest rat populations in the United States, but surprising little study has been undertaken to define rat ecology across varied features of this urban landscape. More importantly, factors that may contribute to increased encounters between rats and humans have rarely been explored. Using city-wide records of rat sightings reported to the NYC Department of Health and Mental Hygiene, this investigation sought to identify sociodemographic, housing, and physical landscape characteristics that are associated with increased rat sightings across NYC census tracts. A hierarchical Bayesian conditional autoregressive Poisson model was used to assess these associations while accounting for spatial heterogeneity in the variance. Closer proximity to both subway lines and recreational public spaces was associated with a higher concentration of rat sightings, as was a greater presence of older housing, vacant housing units, and low education among the population. Moreover, these aspects of the physical and social landscape accurately predicted rat sightings across the city. These findings have identified specific features of the NYC urban environment that may help to provide direct control targets for reducing human-rat encounters.

**Subjects:** Epidemiology, Public Health  
**Keywords:** *Rattus norvegicus*, Rats, Landscape epidemiology, Urban health, Zoonotic diseases

### INTRODUCTION

Rats are a prominent feature of the human landscape and can compromise public health. *Rattus norvegicus*, commonly known as the brown or Norway rat, is ubiquitous in urban settings across the United States (US), while *Rattus rattus*, commonly known as the black rat (or roof rat), is now limited primarily to parts of the southeastern US. *Rattus norvegicus* is a burrowing species and thus often occupies the underground environs of subway and sewer systems (Cline, 1983). Given its preferred environment, *R. norvegicus* does especially well in large urban centers. These rats are reservoirs for significant human

Submitted: 13 June 2014  
 Accepted: 1 August 2014  
 Published: 28 August 2014  
 Corresponding author:  
 Michael G. Walsh,  
 mgwalsh@downstate.edu  
 Academic editor:  
 Clara Wilson  
 Additional information and  
 Declarations can be found on  
 PMP 13  
 DOI: 10.7717/peerj.1533  
 Copyright  
 © 2014 Walsh  
 Distributed under  
 Creative Commons CC-BY 4.0  
 OPEN ACCESS

Open Data, the publication of civic data, demonstrates the potential of collaborative reuse.

A relevant example is the coordinated action in New York City to manage the rat population using predicative analytics over datasets of community reported rat sightings, rubbish collection, and the physical geography of the city.

One outcome we want to help orchestrate with the Data Commons is to support NZ's biodiversity, via Predator-Free NZ. We'd love to enable predictive trapping and bait management of rats and other predators; Particularly with broader community involvement of sightings, trapping, monitoring and awareness.

While data risks would seem absent in this case, they are still there.

Academic credit goes to the first person to present an idea clearly in public. Therefore no researcher wants to have their research publication scooped and datasets are an asset in this regard.

Possum sightings on a farm may increase the odds of TB being discovered in a herd and that is commercially sensitive.

# COLLABORATIVE SOCIAL SERVICES



## KNOWING THE FULL STORY

Being informed about a client's life story can be a significant contributor to better outcomes



## INTEGRATING CLIENT MANAGEMENT

NGO Service delivery can be improved with integrated and accountable data.



## IMPROVING OPERATIONAL EFFECTIVENESS

NGO collect the same data over and over again. A process that is time-consuming and expensive, and for a client tiring and humiliating.



Analytics is only one use case for reusing data.

In many social outcome situations being informed about a client's life story can be a significant contributor to lifting an individual out of a trap.

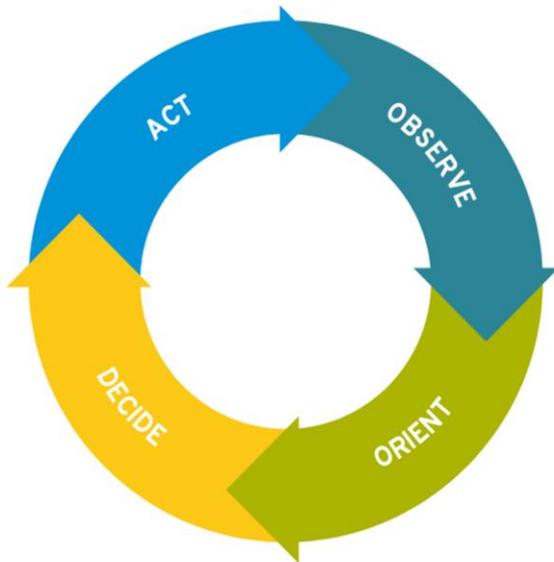
At present the information landscape of the NGO sector is a patchwork of information systems. It is widely agreed in the sector that service

delivery can be improved with integrated and accountable data. Information that tells a client's story.

Operational effectiveness would improve too.

For example NGO staff must collect the same data over and over again. A process that is time-consuming and expensive for the organisation; Tiring and humiliating for the client.

Risks abound in this domain. Nonetheless it is a challenge worth addressing and provides the framing to deeply consider the implications of reusing sensitive identifying information.



# POLICY FEEDBACK

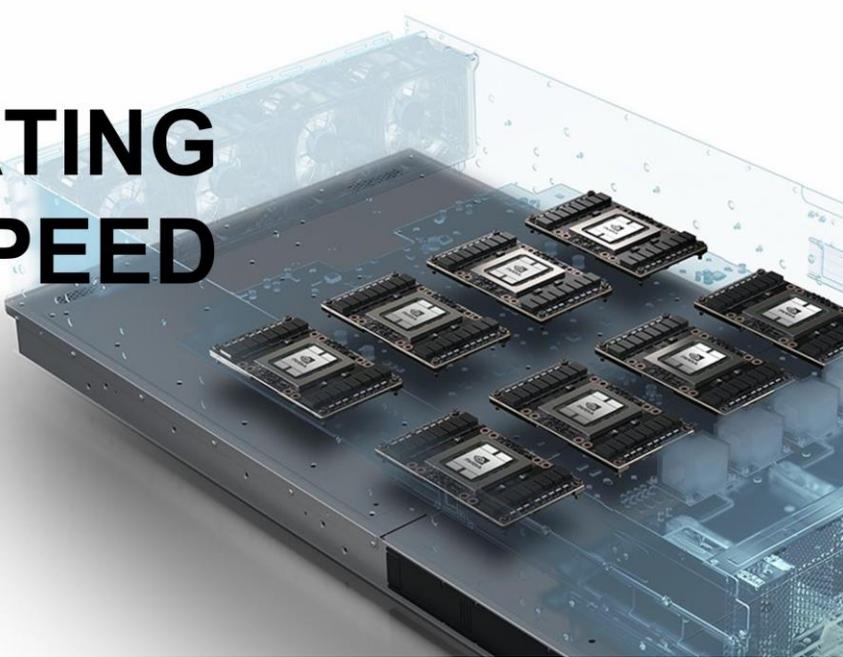
.....

Integrating data, by reusing it under community contract, policy development could be more responsive and agile.

The assumption is that by integrating data, by reusing under community contract, we'll get higher quality information. Not just information about events, but information about interventions, process, or policy effectiveness.

That in-turn will have a profound effect on policy development, making a more responsive and agile way of governing.

# REGULATING AT SPEED



In a data fuelled, algorithm driven society the only way to govern will be with other algorithms.

This also suggests the important role of transparency and accountability.

In a data fuelled, algorithm driven society the only way to govern will be with other algorithms.

Transparency, accountability, data portability, rights to be forgotten, and right of explanation become critical principles for underpinning a civil data-driven society.



The World's most valuable resource is not oil but data

#### NETWORKED DATA ECONOMY

Personal Control as antitrust intervention. Informed Consent. Free exchange and competition in the face of algorithmic knowhow. Algorithmic accountability

This month's Economist issue I mentioned earlier has a number of articles discussing the implications of the Networked Data Economy.

- Personal Control as antitrust intervention
- Informed Consent
- Free exchange and competition in the face of algorithmic knowhow, and
- Algorithmic accountability

All issues that motivate a commons approach to data reuse.



The World's most valuable resource is not oil but data

#### NETWORKED DATA ECONOMY

Personal Control as antitrust intervention. Informed Consent. Free exchange and competition in the face of algorithmic knowhow. Algorithmic accountability



Source and guides for principled design of a commons

#### INSTITUTIONAL ECONOMICS

Ostrom provides the framework for thinking about governance.  
Coase provides the basis for transaction costs.

I've also mentioned two institutional economists whose work provides the basis for principled thinking into the design of a commons. And specifically the role trust and transaction costs play in fashioning our institutional forms.

Ostrom provides the framework for thinking about the structure, rules, disputes, and governance models.

Coase provides the basis for thinking about contracts and their performance in reducing the costs of interaction.



The World's most valuable resource is not oil but data

#### NETWORKED DATA ECONOMY

Personal Control as antitrust intervention. Informed Consent. Free exchange and competition in the face of algorithmic knowhow. Algorithmic accountability



Source and guides for principled design of a commons

#### INSTITUTIONAL ECONOMICS

Ostrom provides the framework for thinking about governance.  
Coase provides the basis for transaction costs.

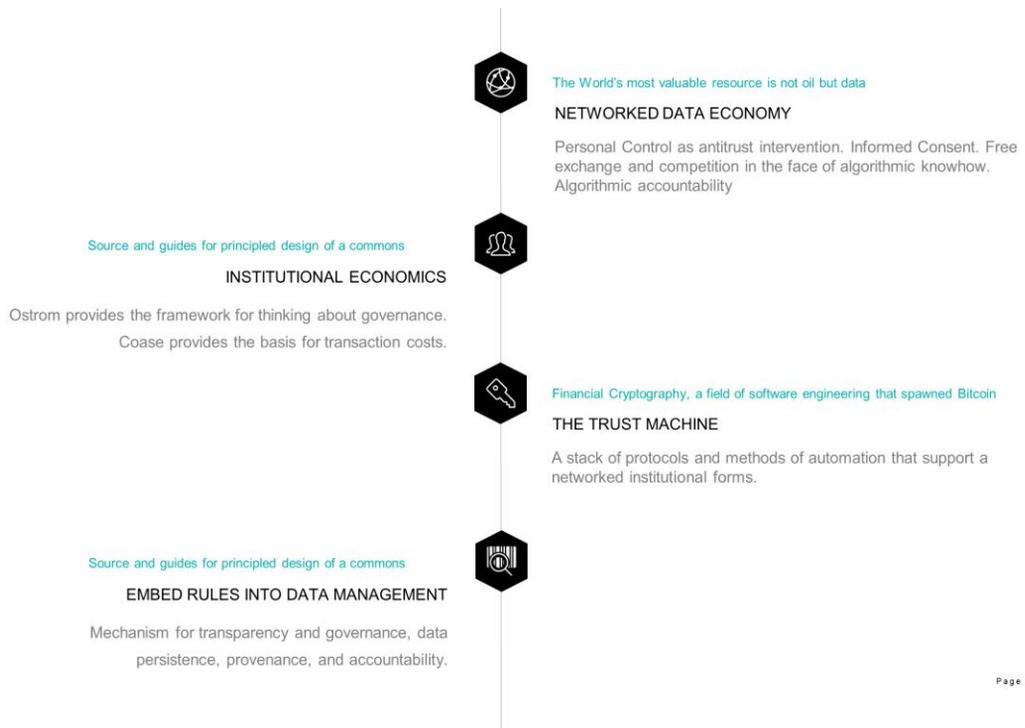


Financial Cryptography, a field of software engineering that spawned Bitcoin

#### THE TRUST MACHINE

A stack of protocols and methods of automation that support a networked institutional forms.

Financial Cryptography, a field of software engineering that spawned Bitcoin, point to a stack of protocols and methods of automation that support a networked institutional form.



Each embed notions of clear rules, mechanism to enhance transparency in operations and governance, building provenance—the lineage, the history of interactions—and accountability.

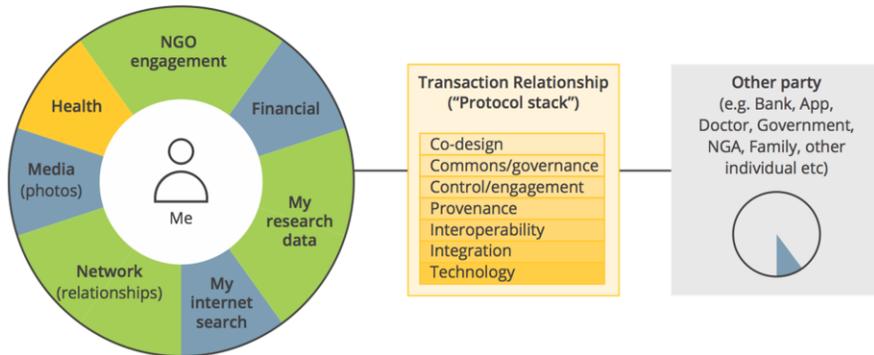
Persistence of Data even. Bits are more fragile than ink. Encrypted bits more so.

How do we take these ideas and concerns, and fashion a Data

Commons that serves our purpose?

# PROTOCOL STACK

.....



PROTOCOL is used in four ways: the social, the contractual aspects, the building of an accountable record, and the technology standards. At different levels of the commons abstraction. Hence the Data Commons Protocol Based.

The Data Commons community needs to define the contracts for data reuse across an often bewildering array of interest groups, and several levels of abstraction.

We're referring to this as being Protocol Based.

We're using the term Protocol in the sense of the social, the contractual aspects, the building of an accountable

record, and the technology standards involved at different levels of the commons abstraction.

And this is where much of the work will be done.

# OPEN DATA



## DATA COMMONS FOCUSED ON REUSE

Sharing, integration, and reuse. These are different things. Collectively termed Data Reuse and is the focus of the Data Commons



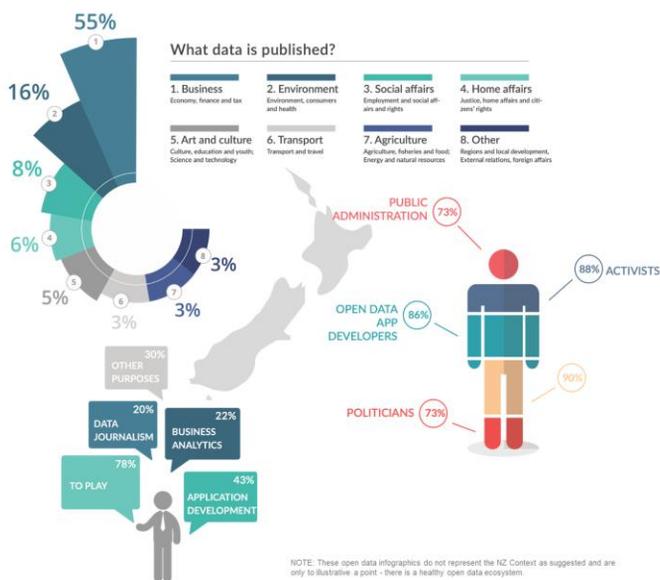
## WIDER CLASS OF DATA

The Data Commons sets out to enable the sharing, integration and reuse of a wider class of gathered data.



## BUILD UPON OPEN DATA

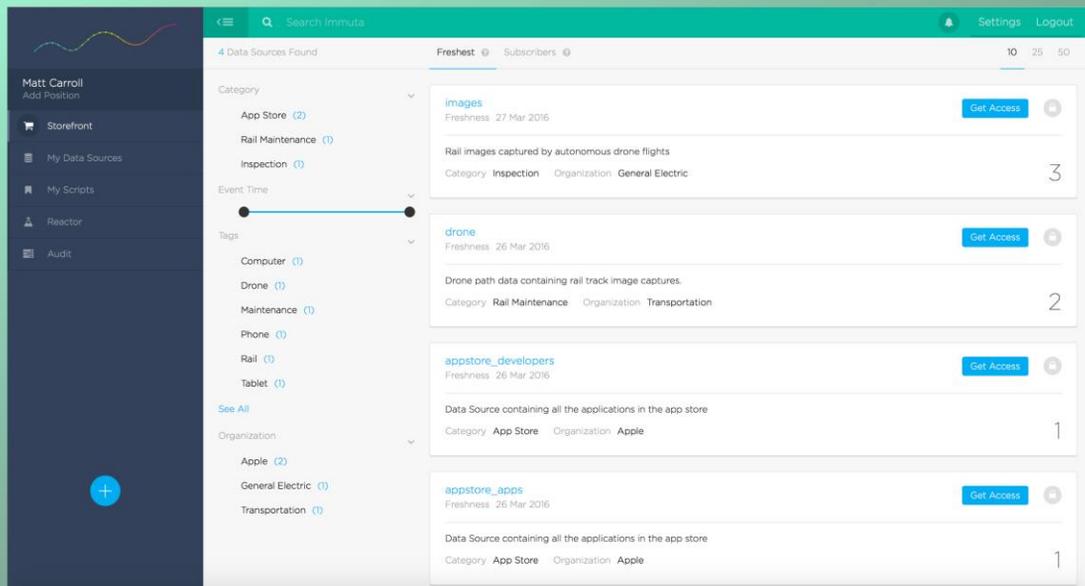
We can learn from the approach, the lessons, the communication, and the outreach strategies. Using Open Data as a starting point to add protocols for wider reuse.



So far I've loosely referred to data sharing, data integration, and data reuse. These are different things, but I've mainly adopted the turn-of-phrase Data Reuse to infer all three.

What sets the Data Commons apart from just Open Data for example is a focus on Reuse. With the intent of enabling the sharing, integration and reuse of a wider class of gathered data.

We can use Open Data as a starting point and exemplar. Building identification, descriptive, provenance, and operational metadata standards that aren't strictly of direct concern to Open Data. Yet these are critical to accountability, transparency and persistence.



The concrete experience of using a Data Commons community will be realised through the use of "Apps" in the sense of datasets, APIs, services, derivatives, or Sensors and Smartphone Apps that help gather data.

The process of negotiating access to a dataset would be done through a marketplace. Not unlike a data portal like CKan, Socrata. The difference

maybe in the way you obtain access.

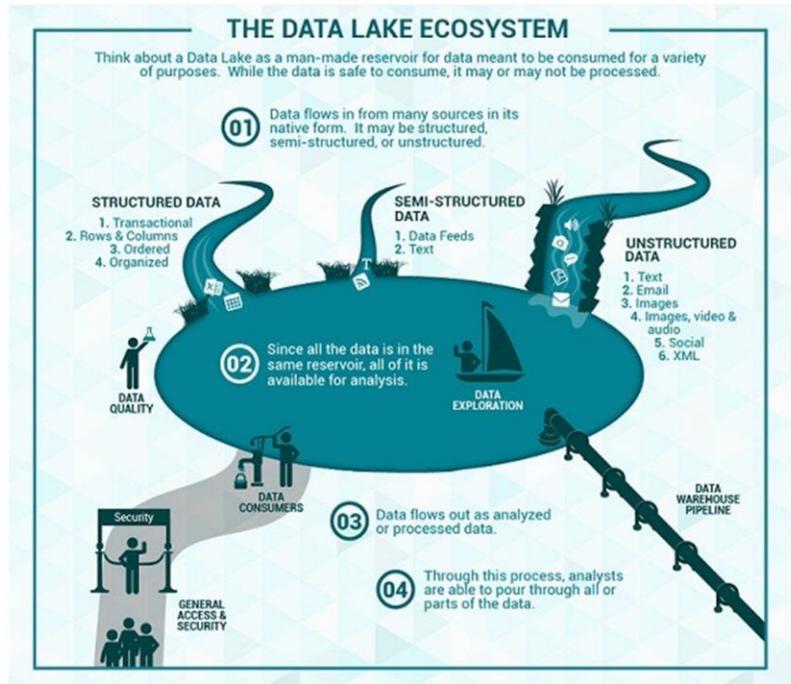
In the Data Commons instead of having open access, or purchasing access, you'd be negotiating it based on the specific access requirements, reuse, and compliance properties you need.

This could be a complex web of constraints, rights, promises, compensation, and participant reputations.

# ENTERPRISE DATA IS A COMMONS

.....

DATA LAKES: The various factors of a data commons are also concerns with an organisation's own internal data management.



There are a lot of factors to consider but as I've hinted at with the Economist articles and the EU General Data Protection regulation, these data governance and stewardship issues are coming into the foreground.

If you're dealing with data inside your organisation, if you are now discussing or implementing your own data lake and policy for data governance and stewardship that go

with it; then the various aspects I have discussed are your concerns too.

At least should be!



To date a group of people have come together, developed a Blueprint that digs into the topic of why, what and how of the Data Commons in more detail.

This work builds on the New Zealand Data Future Forum and has been sponsored by the Next Foundation, Bio-heritage New Zealand, and Inflection.

Our next steps are to progress with small end-to-end prototype(s) that help to anchor and inform more detailed thinking into what needs to be considered and what form these may take.



## CO-DESIGN

Working with the community as active participants in the design process, to create shared value and better outcomes.



As a Commons one of our core protocols is the process of Co-Designing with as many diverse voices as feasible. As such we're reaching out.

# DATA COMMONS

[datacommons.org.nz](http://datacommons.org.nz)



If you're interested, read the blueprint.  
Share it. Join into the discussion.

Billy our facilitator has printed copies  
of the Blueprint, so you can also grab  
a copy here.

And the QR Code provides the direct  
URL to the PDF.

Thank you.