CE876 - Information Security Mng. & Eng.

Lecture 12: Data Protection

Seyedeh Atefeh Musavi / Mehdi Kharrazi
Department of Computer Engineering
Sharif University of Technology
Spring 1400

S4Lab

Acknowledgments: Some of the slides are fully or partially obtained from other sources. A reference is noted on the bottom of each slide to acknowledge the full slide or partial slide content.

Different kinds of human data

- Data which are generated.
    - Reveal by choice, such as through social media and e-mail.
    - Reveal via compulsory disclosure, as a condition, for example, of banking or traveling.
- Data which are collected.
    - By surrounding sensors.
- Data which are calculated or inferred .
    - based on demographic information, census data, and past behavior.
    - Those data are created, not collected.

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

What is the concern?

- So we are going to talk data governance , but what are the concerns?
- Privacy?
- Is there any thing else?

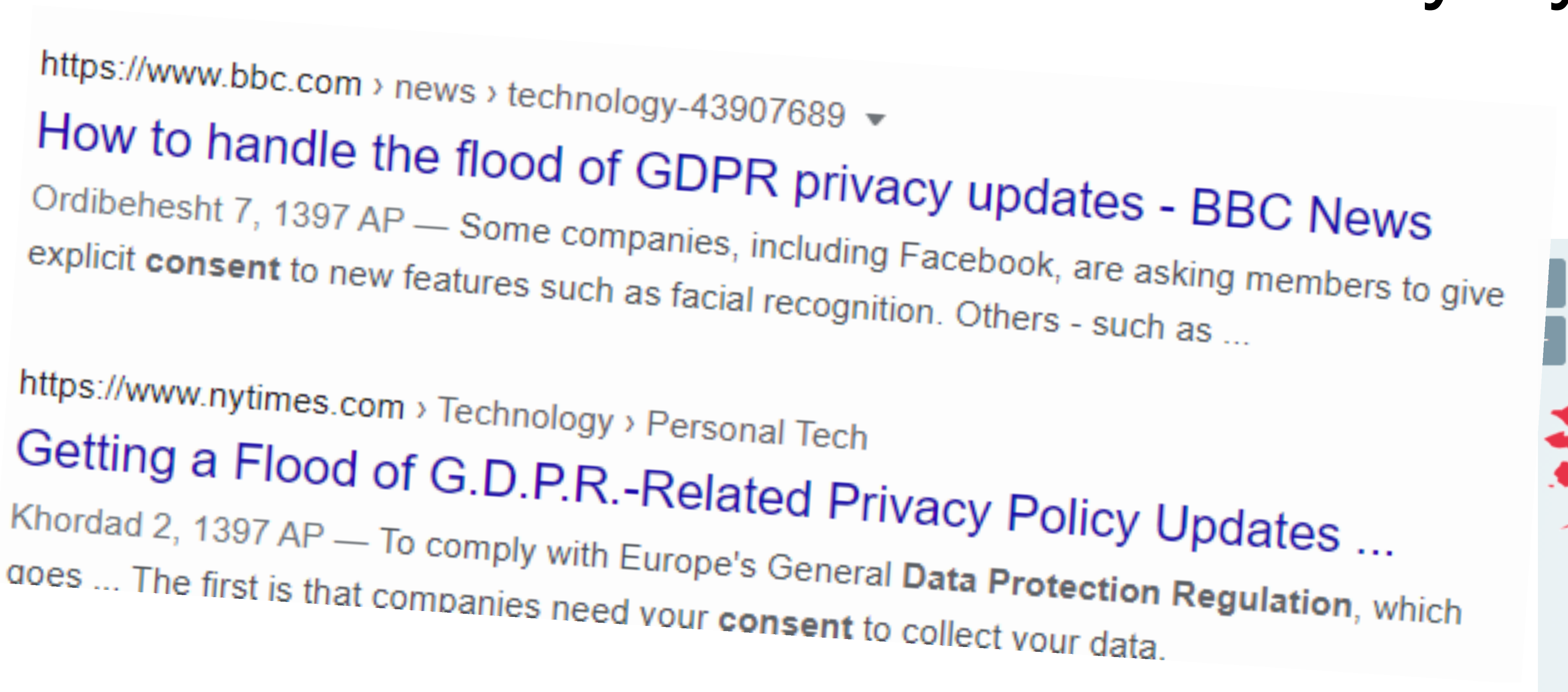
What is the concern?

- So we are going to talk data governance , but what are the concerns?
- Privacy?
- There are plenty of unintended consequences of big data, which might be controllable in some senses:
    - Data re-analysis and bias.
    - Ignoring the power of small data, undervaluing the power of data curation.
    - Not thinking to look out the window!!
    - “In the Wild” systems are open for abuse and manipulation.
    - Ineffectiveness of traditional methods and views (redundancy, simple verification, …).
- Let’s start with these concerns, we will add some others to the list soon.

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

How these concerns are handled at national level?

Mostly by data protection regulations



https://www.bbc.com › news › technology-43907689 ▾

How to handle the flood of GDPR privacy updates - BBC News

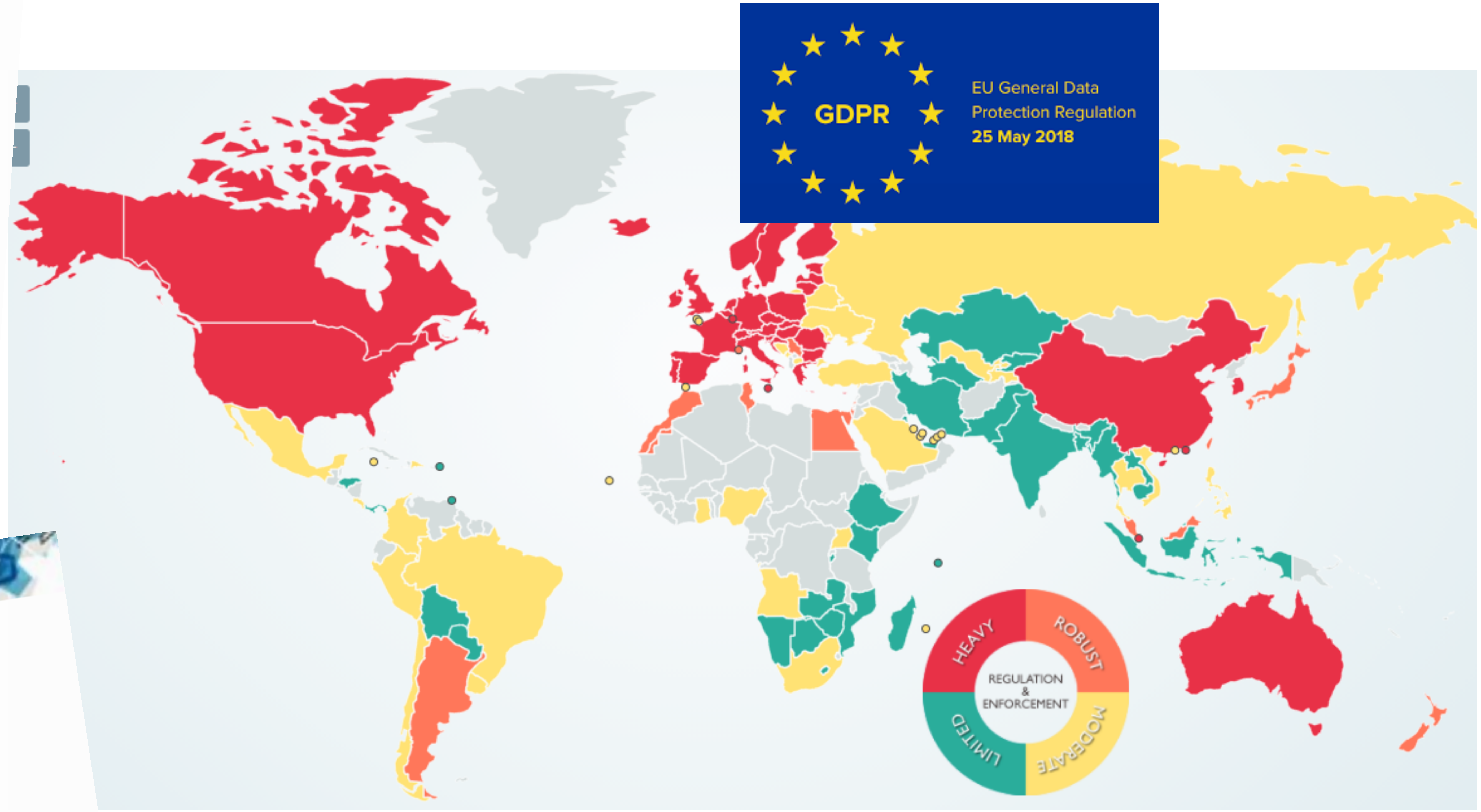
Ordibehesht 7, 1397 AP — Some companies, including Facebook, are asking members to give explicit **consent** to new features such as facial recognition. Others - such as ...

https://www.nytimes.com › Technology › Personal Tech

Getting a Flood of G.D.P.R.-Related Privacy Policy Updates ...

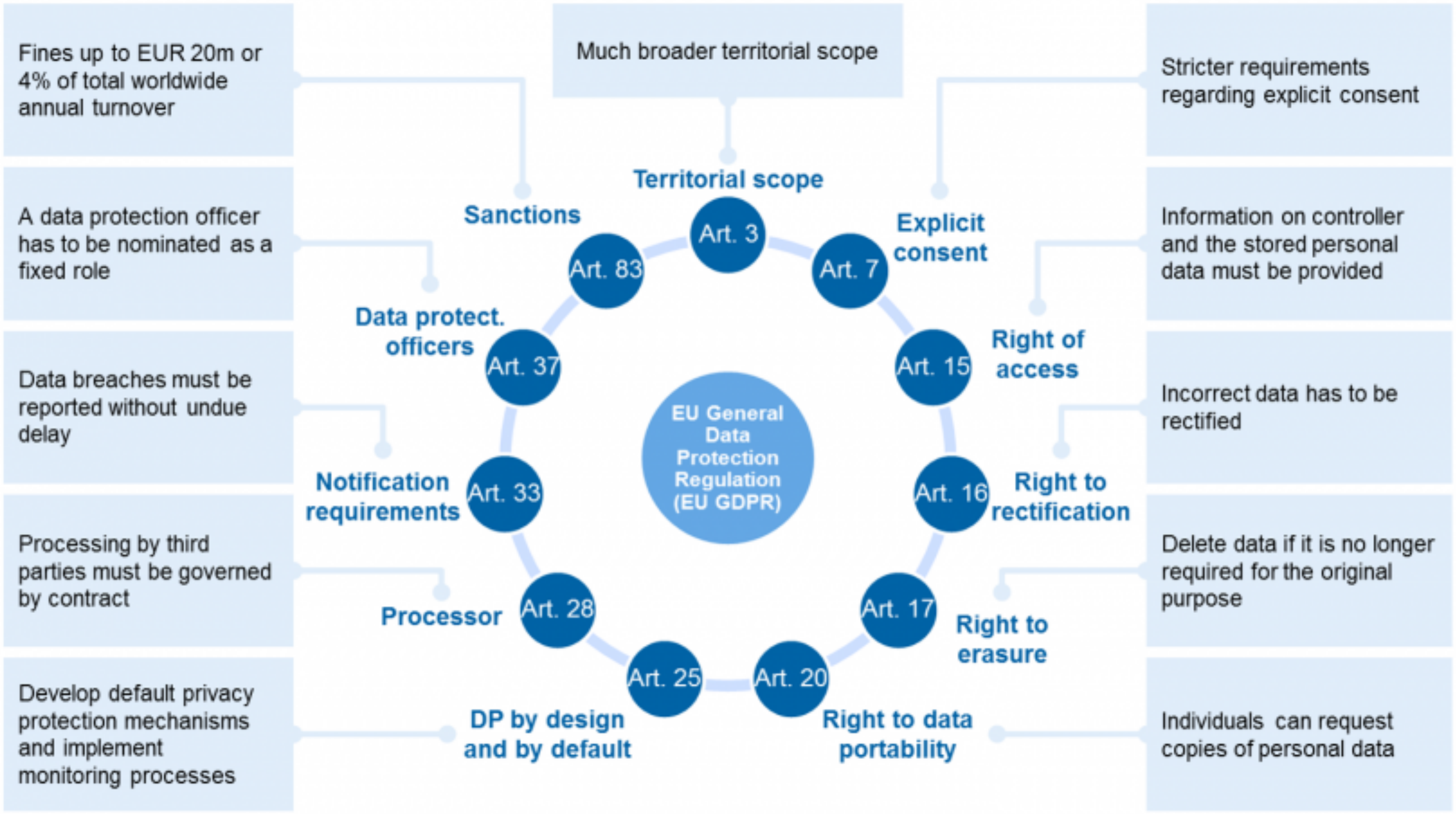
Khordad 2, 1397 AP — To comply with Europe's General **Data Protection Regulation**, which does ... The first is that companies need your **consent** to collect your data.

The first year of GDPR cost businesses over €359 million in fines



[Image: https://www.dlapiperdataprotection.com/]

GDPR summary



Fines up to EUR 20m or 4% of total worldwide annual turnover

A data protection officer has to be nominated as a fixed role

Data breaches must be reported without undue delay

Processing by third parties must be governed by contract

Develop default privacy protection mechanisms and implement monitoring processes

Much broader territorial scope

Stricter requirements regarding explicit consent

Information on controller and the stored personal data must be provided

Incorrect data has to be rectified

Delete data if it is no longer required for the original purpose

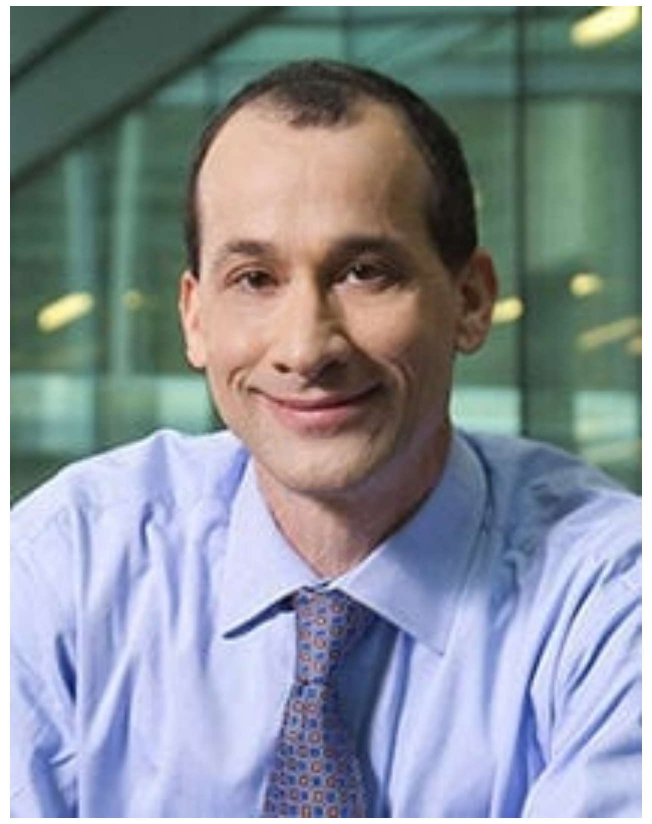
Individuals can request copies of personal data

[GDPR deep dive—how to implement the ‘right to be forgotten’, Daniel Crow, Banking Hub, 2017]

An elegy for consent-based data protection regulations

“Social and legal norms about privacy promise too much, namely data control, and deliver too little.”

Paul Schwartz ,UC Berkeley School of Law, 1999.



Consent at Time of Collection

- Most data protection laws place some or all of the responsibility for protecting privacy on individual data subjects through the operation of notice and consent.
- Not only FTC rules or GDPR, but also in older rules.
    - E.g. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980).
- Relevant and important in situations where choice about data collection is appropriate and meaningful
- There is mounting evidence that in many settings, individual choice is impractical and undesirable.

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

Complexity of notices

- Notices are frequently complex.
  - When PayPal's privacy notice is added to its other terms of use disclosed to consumers, the total word count is 36,275, longer than Hamlet (at 30,066 words), and iTunes' comes to 19,972 words, longer than Macbeth (at 18,110 words)☺️
  - One study calculated that to read the privacy policies of just the most popular websites would take an individual 244 hours—or more than 30 full working days—each year
- Regulators, industry groups, academics, and others have proposed a variety of ways of making notices more accessible:
  - Including shortened notices, layered notices, standardized notices, and machine-readable ones.
- Why it did not happen? or updates weren't successful?

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

Inaccessibility of notices

- Most data collection and creation take place without direct contact with data subjects.
- Data are collected or generated by sensors, such as surveillance cameras or microphones, of which the data subject may be only vaguely aware.
    - “CCTV in use” warnings found throughout London and other major cities.
- Neither meaningful notice (i.e., where are the cameras, what happens to the video….), nor effective options (e.g., avoiding or disabling the cameras).
- More often, there is no opportunity for notice or choice about collection at all.
    - Inferences, probabilities, predictions, and other data are also created by government and industry for hundreds of reasons ranging from marketing to tax audits.
    - I agree tick!

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

Inadequate Privacy Protection

- Reliance on consent ignores the fact that consent does not equal privacy:
    - Individuals can enjoy strong privacy protection without consent and can suffer serious incursions into their privacy with consent.
- Notice and consent do not protect us from our own bad, ignorant, unintentional, or unavoidable choices.
- Contrast consent in privacy with other types of consumer protection laws:
    - A consumer cannot consent to be defrauded, but they can consent to have their privacy violated.
- In addition, the energy of data processors, legislators, and enforcement authorities is often expended on notices and consent opportunities rather than on actions that could actually protect privacy.
    - Compliance with data protection laws is usually focused on providing required notices in the proper form at the right time and recording consent!

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

Other ineffectiveness factors

- False Dichotomy: The preoccupation with consent frequently sets up an artificial dichotomy between personally and non–personally identifiable information.
  - In a world of big data, where with sufficient, interconnected data, even de-identified or anonymized data may be rendered personally identifiable.
- Choice as a disservice to individuals and society: consider information about individuals' creditworthiness.
  - In the words of former FTC chair Muris (2001), the credit reporting system "works because, without anybody's consent, very sensitive information about a person's credit history is given to the credit reporting agencies."

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

People trade their privacy simply!

- Users are much too prone to casually sacrificing their personal data to any interested company offering a marginal incentive.
- This is where GDPR will likely be amended in the future. If the goal is to protect user privacy, policymakers cannot escape a coming showdown with the inconvenient fact that people will eagerly trade their privacy for a chance to look at "10 Celebrities Who Didn't Age Well."

[The Future of Data Protection Law, Spencer Kimball, Cockroach Lab, 2019]

Other data protection approaches

“[T]he value of big data means we must directly control use rather than using notice and consent as proxies.”

Susan Landau, Fletcher School of Law & Diplomacy, Tufts University, 2015.

Alternatives to consent

- Shifting the Focus from individual consent to data stewardship.
  - The effective governance of big data requires shifting more responsibility away from individuals and toward data collectors and data users, who should be held accountable for how they manage data rather than whether they obtained individual consent.
- A Greater Focus on Data Uses.
  - Assessing the risk to individuals posed by those data almost always requires knowing the context in which they will be used.
  - Data used in one context, or for one purpose or subject to one set of protections, may be both beneficial or dangerous.
  - A use-focused approach is especially important in the context of big data because the analysis of big data doesn't always start with a question or hypothesis but rather may reveal insights that were never anticipated.

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

Why collected vs used? (1)

- One of the most pronounced changes that will result from the evolution toward a greater focus on use is to diminish the role of the purpose for which data were originally collected.
- Why diminish? Why not to make initial purpose consistent with further use purposes?

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

Why collected vs used? (2)

- OECD (1980) guidelines explicitly provide for a purpose specification principle that requires that “the purposes for which personal data are collected should be specified not later than at the time of data collection,” and then limits subsequent use to “the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.”
- This principle is problematic for many reasons,
  - Precisely because of it, data processors usually specify exceptionally broad purposes that offer little meaningful limit on their subsequent use of data.
  - In addition, because data increasingly are generated in ways that involve no direct contact with the individual, there is never a purpose specified.
  - Personal data may have substantial valuable uses that were wholly unanticipated when the data were collected.
- Some modern data protection systems have dealt with these problems by creating broad exceptions to this principle, interpreting “not incompatible” so expansively as to undermine the principle, or simply ignoring it altogether.

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

Better view is the risk assessment

- The GDPR introduces stricter requirements for high-risk processing.
- Three examples of high-risk activities are provided, including:
  - “Systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the individual or similarly significantly affect the individual.”
  - “Processing on a large scale of special categories of data.”
  - “A systematic monitoring of a publicly accessible area on a large scale.”

[The Risk-Based Approach in the GDPR: Interpretation and Implications, Gabe Maldoff, IAPP, 2016]

Better view is the risk assessment (con’t)

- Controllers that engage in processing that poses a high risk for data subjects face three additional obligations:
  - Article 33 requires controllers to conduct a data protection impact assessment for high-risk processing activities.
  - When high risk, Article 34 requires the controller to consult the relevant supervisory authority before conducting the activity.
  - Under Article 32, controllers are required to notify individuals in addition to the competent authorities of a security incident if “the personal data breach is likely to result in a high risk” to their rights and freedoms.
- GDPR encourages risk-based compliance.

[The Risk-Based Approach in the GDPR: Interpretation and Implications, Gabe Maldoff, IAPP, 2016]

Society-wide responses for Data protection

The building of a new social and economic order based on the extraction of value from human life through data relations is not something that individuals can resist, or even manage, by themselves…Society-wide responses are needed to such society-wide transformations.

[Making data colonialism liveable: how might data's social order be regulated?, Couldry, N. & et al., Internet Policy Review, 2019]

When everything reveals everything!

- What if everything reveals everything!
  - Floor protectors reveal creditworthiness.
  - Shopping habits reveal pregnancy.
- Does everything reveal everything?
  - Stated so starkly, this cannot literally be true.
- But as the state of the art of big data advances, we will learn that many different modifiers will, independently of one another, lead to true claims.
- So what are the affective factors?

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

The good, the bad, and the ugly

- What will it mean if even a weak version of everything reveals everything turns out to be true?
- The Good: you will control your behavior ☺️
- The bad: One concern is that we will begin to modify our behavior in undesirable ways to avoid these consequences.
- The ugly: Instead of behavior modification being the problem, there is also the distinct possibility of a quite opposite threat: the inability to modify one’s behavior to respond adequately to big data’s inferences.

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

What about ourselves?

- A consumer is unlikely to know the relevant inputs, algorithms, or weightings among variables. Alongside being revealed, then, there is the added threat of being demoralized.
  - “They” know you are likely to drive poorly, even if you don’t feel that way. Driving better may not help, and it might be unclear how “they” are arriving at their predictions.
  - This could produce paranoia, anger, or at least a rising sense of claustrophobia.
- Instead of discovering that we are so consistent as to have little room for creative play in our sense of who we are, we could be led to believe that about ourselves by the pervasive use of big data analytics to sort, categorize, and price us.
- This numbing inner consistency might not be true, but we could come to believe it about ourselves anyway.

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

How laws protect our data?

- Privacy Laws draw 4 types of lines:
  - Lines between personally identifiable and “non–personally identifiable” information
  - Lines based on the purported sensitivity of information
  - Lines drawn is by industry sector
  - Lines Drawn is by actors
- Anti-discrimination law
  - Defines protected classes, such as (but not limited to) race, ethnicity, sex, religion, and age. Advances in statistical inference may reveal the otherwise-unspecified. Values for some of these categories, opening the door to intentional, invidious discrimination.
- Consumer Protection Law
  - Related to both privacy and anti-discrimination law, some laws draw lines around categories of information that may not legitimately be used for certain important life decisions.

Because big data blurs the lines between contexts, our laws will by necessity become more under-inclusive and over-inclusive.

[Big data is not a monolith, Sugimoto, C. R., Ekbia, H. R., & Mattioli, M., MIT Press, 2016]

Semantic discontinuity

- New concepts may be required, e.g semantic discontinuity vs older principle of “contextual integrity”.
- To limit the possibility of separate data sets being combined so as to generate inferences of a sort that data subjects did not consent to being made.
- But maybe few chance to enter these concepts into regulations!
  - How can semantic discontinuity be made effective as a legal principle when it contradicts the stated purposes of countless corporations who seek access to personal data?

[Making data colonialism liveable: how might data’s social order be regulated?, Couldry, N. & et al., Internet Policy Review, 2019]

Right to offline alternatives

- In the near future, it will be impossible to buy a car not equipped with the mandatory ecall system and already today it is difficult to find a state of the art TV without networking capabilities
- Individuals will be forced into allowing the processing of their data in order to participate in society.
- Should be any legal support to let you be offline?
  - A discourse emerges around a right to network-free or “offline” alternatives.
- Offline technology would reduce risks to privacy and also evade the imposition – if at all possible – of having to navigate complex system settings to disable services one never wanted in the first place.
- Governmental regulation provides balance between fundamental rights and the public interest, in other cases the market decides whether offline alternatives are worth maintaining.
- Rather than letting the market decide, and make offline alternatives a luxury for the few, one could argue for a right to have an alternative.

[Is There a Right to Offline Alternatives in a Digital World?, Karaboga, M., et al., In Data Protection and Privacy: (In) visibilities and Infrastructures, 2017]

Data ownership concern

- Today’s paradoxical but real question:
  - Who has the ownership of your data!
- This comment resonated with how many DNA testing companies talk about data ownership. It reflects an intuitive feeling that many share that I should own information about myself, but as President Obama suggested, this intuition is not necessarily congruent with legal understanding of data ownership.

[Who Owns the Data in a Medical Information Commons?, Amy L. McGuire, The Journal of Law, Medicine & Ethics, 2019]

Data ownership concern

S4Lab

- Today’s paradoxical bu
  - Who has the owners
- This comment resonat about data ownership. I should own information about myself, but as President Obama suggested, this intuition is not necessarily congruent with our legal understanding of data ownership.

> “I would like to think that if somebody does a test on me or my genes, that that's mine, but that's not always how we define these issues.” Barack Obama

[Who Owns the Data in a Medical Information Commons?, Amy L. McGuire, The Journal of Law, Medicine & Ethics, 2019]

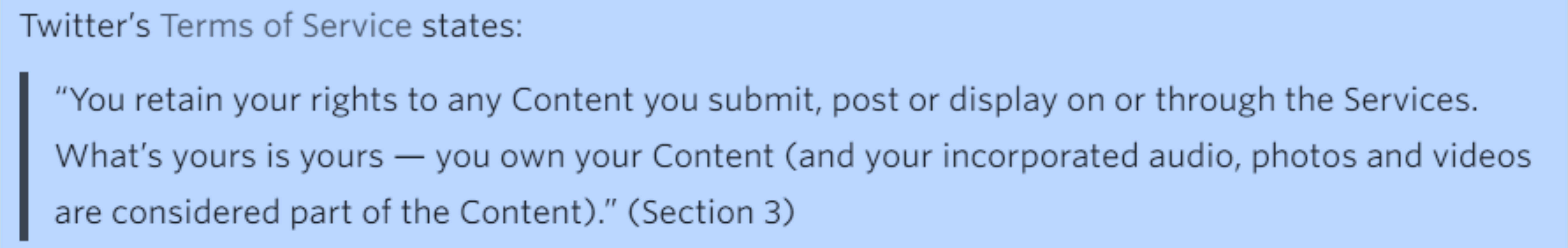
Data ownership concerns (1)

- Ownership of content in your digital life is one of the most legal/economical discussed issues.
- Also tagged as Content Licensing, Data Capitalism, Data Colonialism, and etc.
- It is important to distinguish between user-generated content and user-generated data.
    - Content is separate from user-generated data as its authorship and ownership can be defined by copyright licensing within a specific platform.
    - Some data is created by the act of a user browsing a website, such as their web traffic data that includes their browser type, public IP address, time logs, and cookies.

[Ownership of Content in Your Digital Life – Social Media, Jason Cheung, UBC, 2018]

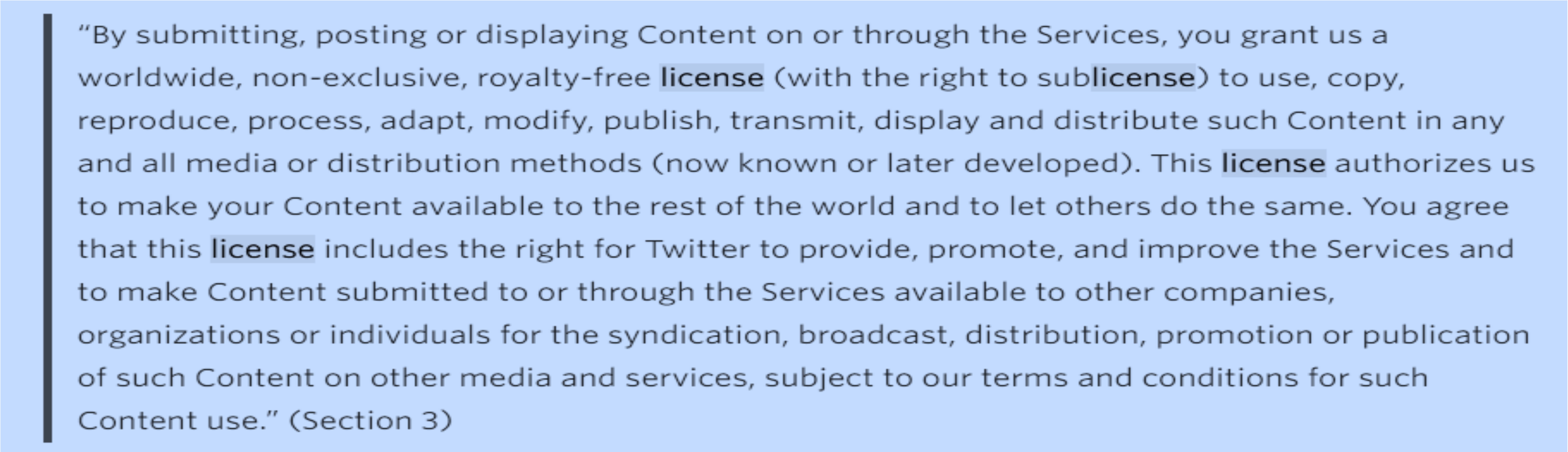
Data ownership concerns (2)

S4Lab



Twitter's Terms of Service states:

> "You retain your rights to any Content you submit, post or display on or through the Services. What's yours is yours — you own your Content (and your incorporated audio, photos and videos are considered part of the Content)." (Section 3)

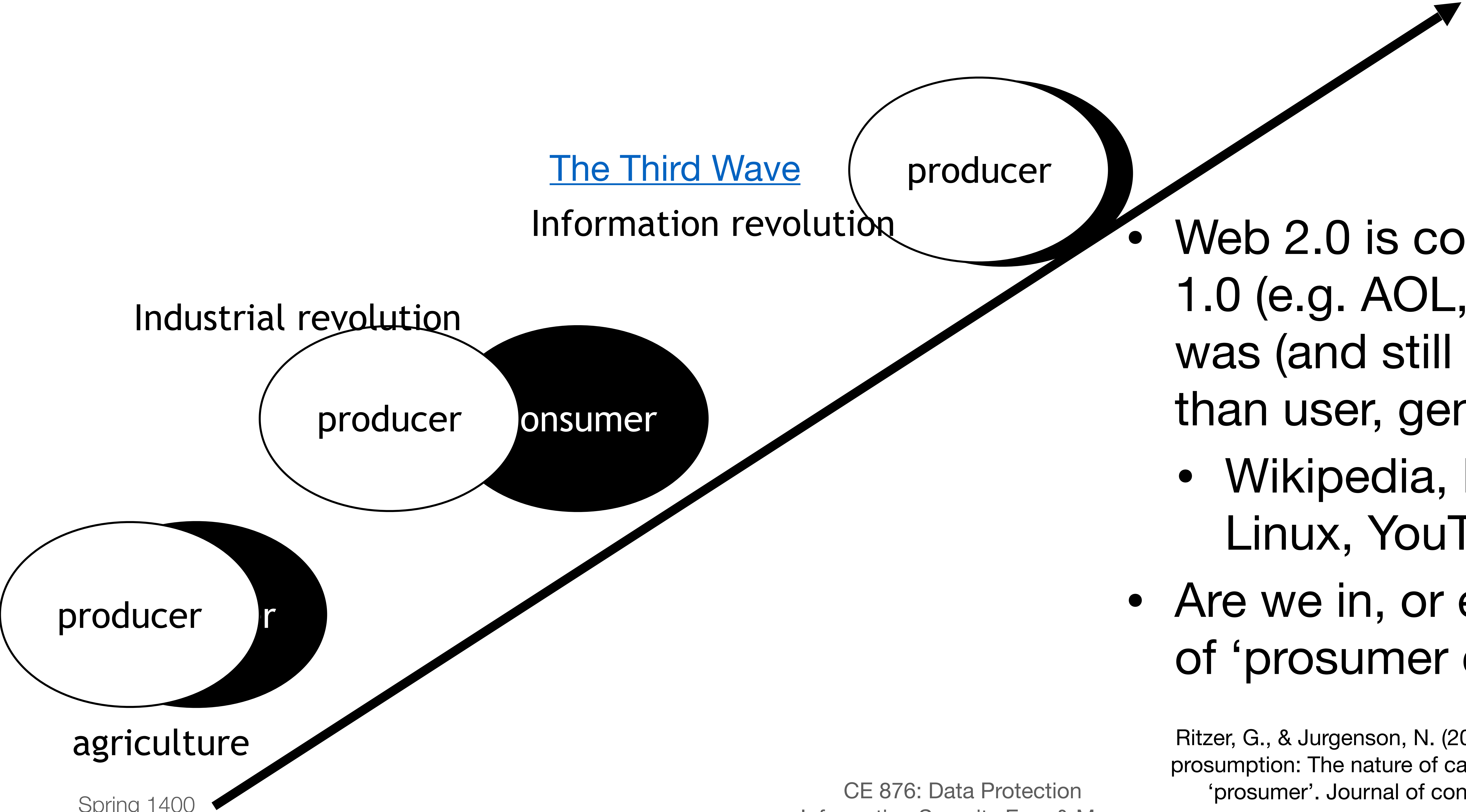


> "By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods (now known or later developed). This license authorizes us to make your Content available to the rest of the world and to let others do the same. You agree that this license includes the right for Twitter to provide, promote, and improve the Services and to make Content submitted to or through the Services available to other companies, organizations or individuals for the syndication, broadcast, distribution, promotion or publication of such Content on other media and services, subject to our terms and conditions for such Content use." (Section 3)

[Ownership of Content in Your Digital Life – Social Media, Jason Cheung, UBC, 2018]

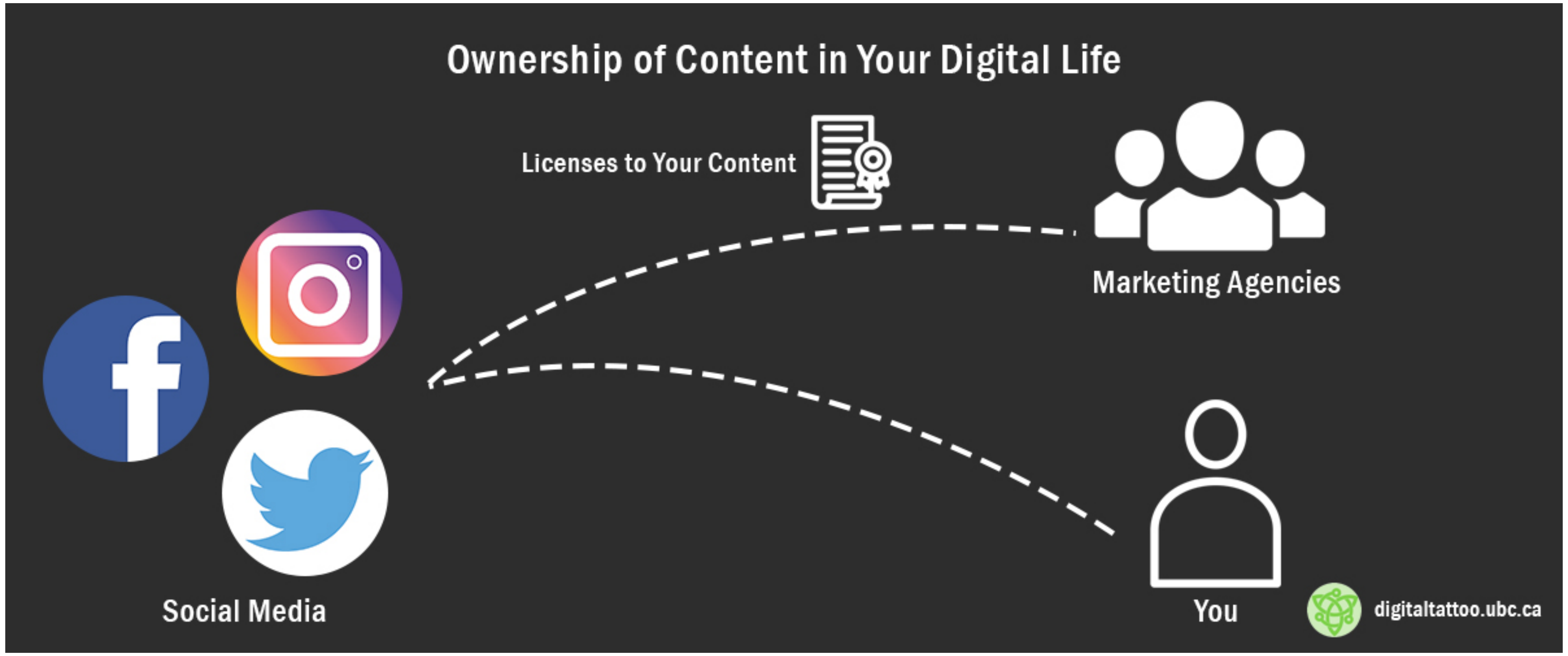
Digital prosumer model

S4Lab



- Web 2.0 is contrasted to Web 1.0 (e.g. AOL, Yahoo), which was (and still is) provider, rather than user, generated.
  - Wikipedia, Facebook, Ebay, Linux, YouTube
- Are we in, or entering, the age of ‘prosumer capitalism?

Ritzer, G., & Jurgenson, N. (2010). Production, consumption, prosumption: The nature of capitalism in the age of the digital ‘prosumer’. Journal of consumer culture, 10(1), 13-36.



- Toffler argues that prosumption is the natural evolution of Marxist capitalism, as capital requires continuous growth to capture additional surplus value.
- Prosumption not only allows capital to increase corporate profit margins through 'overcharging' consumers above the cost of production and 'underpaying' employees to extract surplus value from labour, but also makes it so that "prosumers seem to enjoy, even love, what they are doing and are willing to devote long hours to it for no pay"

[Ownership of Content in Your Digital Life – Social Media, Jason Cheung, UBC, 2018]

“Data is the new oil”?

- The public is often told that “data is the new oil”.
- The evocative idea of “new oil” might recall the benefits (for some) of historic colonialism.
- For sure, capitalism has always sought to commodify everything and control all inputs to its production process.
- Platforms become software-constructed spaces that produce the social for capital. Social life is thereby transformed into an open resource for extraction that is somehow “just there” for exploitation.

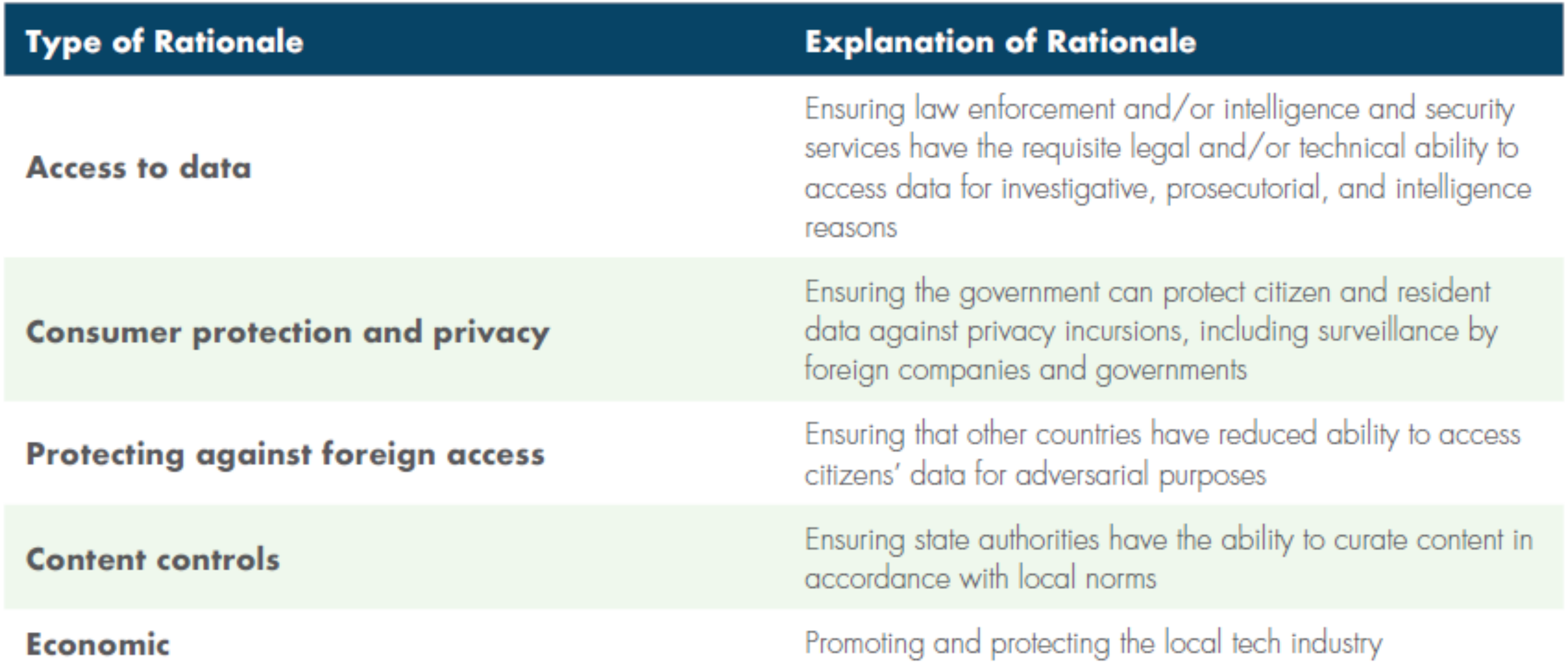
[Making data colonialism liveable: how might data’s social order be regulated?, Couldry, N. & et al., Internet Policy Review, 2019]

Data nationalism/localization

- Not-so-long ago there were a vision of a free, open, and globally interconnected internet which was consistent with US push of “one internet, one global community, and a common body of knowledge that benefits and unites us all”.
- Once-forceful calls for a “free and open” internet are increasingly being replaced by efforts to protect, control, and manage the internet and its related risks, often in a geographically bordered way.

[Data Nationalism on the Rise, Daskal, J. & Sherman, J., Data Analyst, 2020]

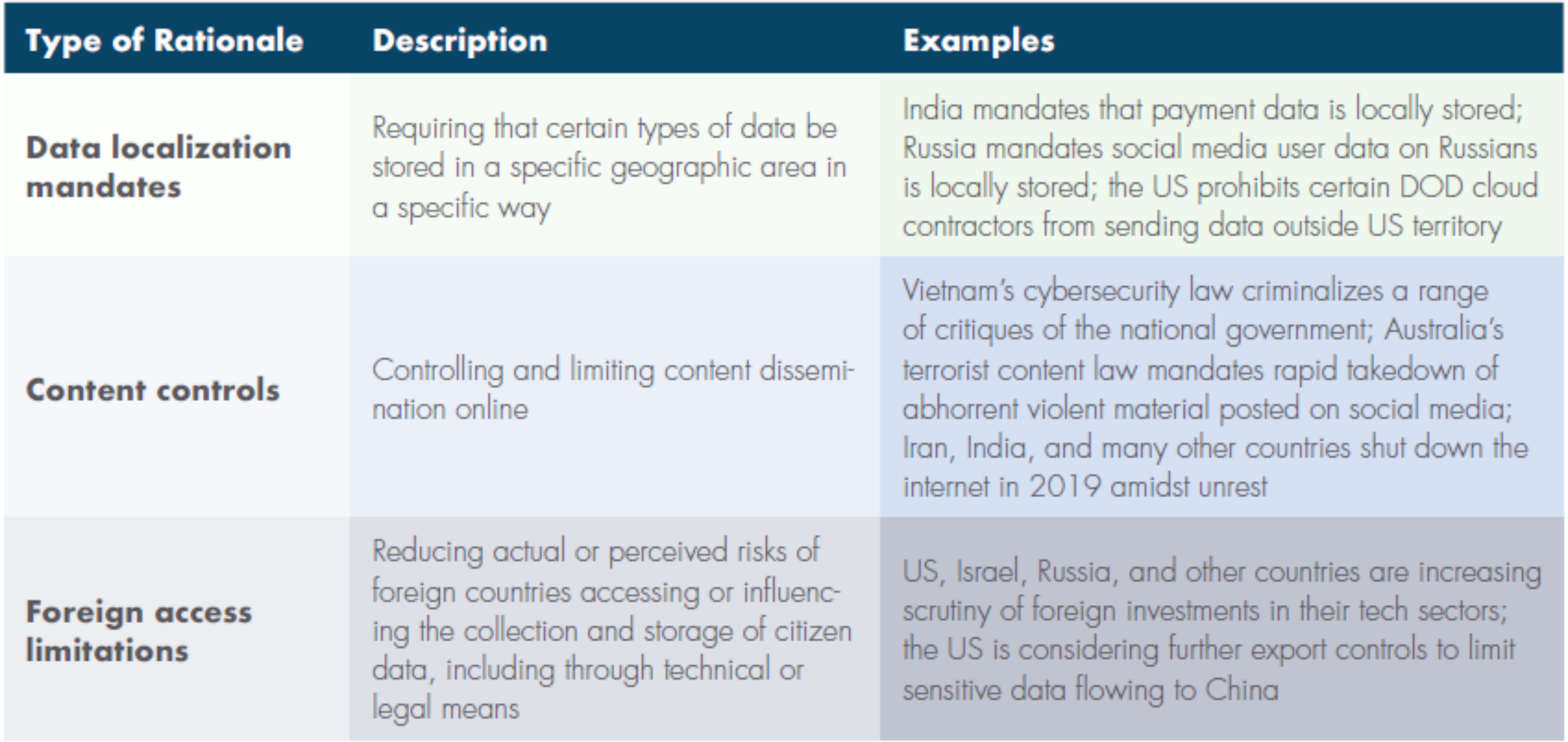
Data nationalism/localization



Type of Rationale	Explanation of Rationale
Access to data	Ensuring law enforcement and/or intelligence and security services have the requisite legal and/or technical ability to access data for investigative, prosecutorial, and intelligence reasons
Consumer protection and privacy	Ensuring the government can protect citizen and resident data against privacy incursions, including surveillance by foreign companies and governments
Protecting against foreign access	Ensuring that other countries have reduced ability to access citizens' data for adversarial purposes
Content controls	Ensuring state authorities have the ability to curate content in accordance with local norms
Economic	Promoting and protecting the local tech industry

[Data Nationalism on the Rise, Daskal, J. & Sherman, J., Data Analyst, 2020]

Forms of data nationalism



Type of Rationale	Description	Examples
Data localization mandates	Requiring that certain types of data be stored in a specific geographic area in a specific way	India mandates that payment data is locally stored; Russia mandates social media user data on Russians is locally stored; the US prohibits certain DOD cloud contractors from sending data outside US territory
Content controls	Controlling and limiting content dissemination online	Vietnam's cybersecurity law criminalizes a range of critiques of the national government; Australia's terrorist content law mandates rapid takedown of abhorrent violent material posted on social media; Iran, India, and many other countries shut down the internet in 2019 amidst unrest
Foreign access limitations	Reducing actual or perceived risks of foreign countries accessing or influencing the collection and storage of citizen data, including through technical or legal means	US, Israel, Russia, and other countries are increasing scrutiny of foreign investments in their tech sectors; the US is considering further export controls to limit sensitive data flowing to China

[Data Nationalism on the Rise, Daskal, J. & Sherman, J., Data Analyst, 2020]

QA