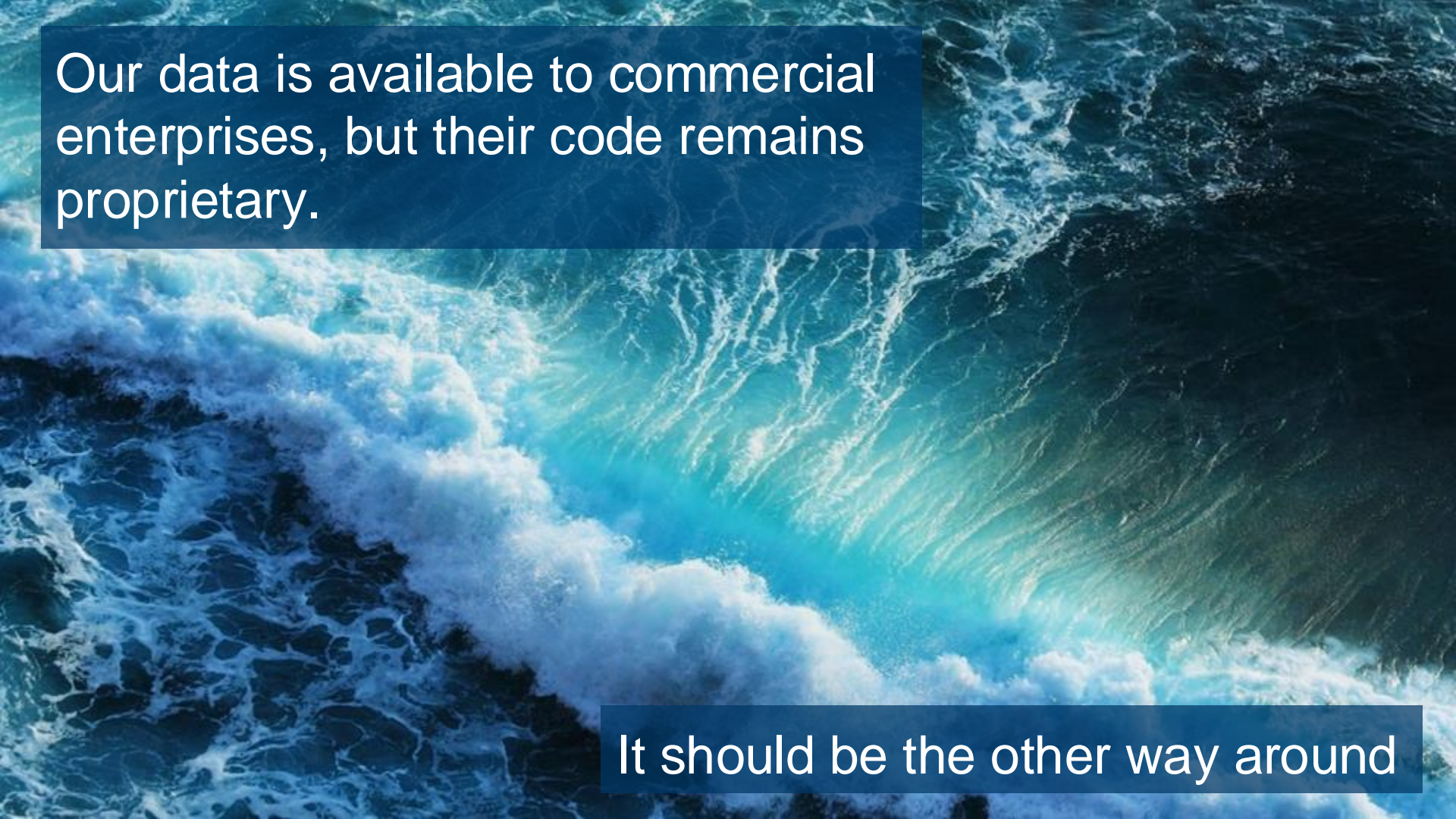




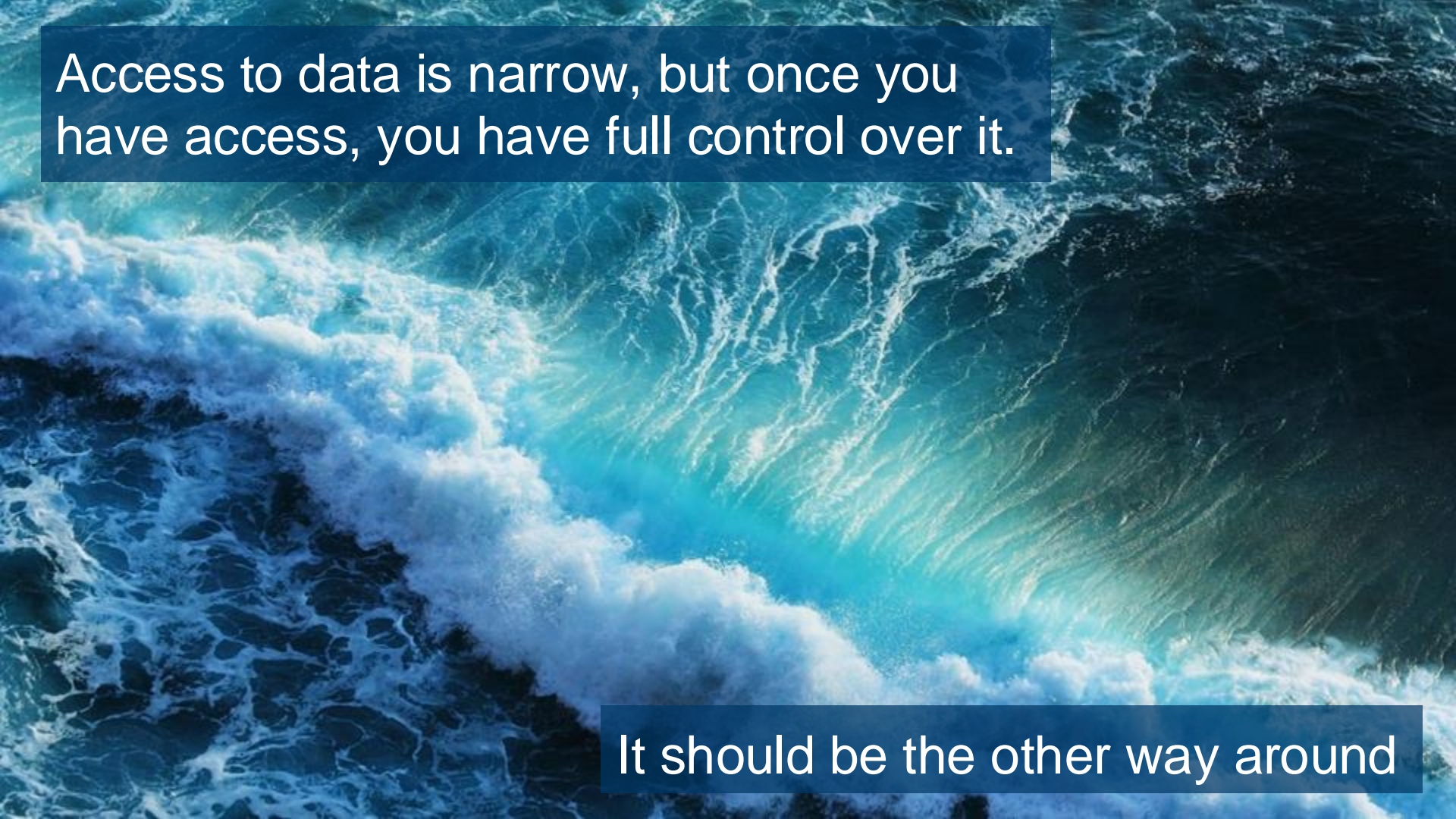
Creating Digital Property Rights Using Blockchain

Co-Pierre Georg
(Frankfurt School of Finance & Management)



Our data is available to commercial enterprises, but their code remains proprietary.

It should be the other way around

An aerial photograph of a turbulent ocean. The water is a mix of deep blue and bright turquoise, with white foam from breaking waves creating a complex, swirling pattern. The perspective is from above, looking down into the churning water.

Access to data is narrow, but once you have access, you have full control over it.

It should be the other way around

Companies struggle to control their data

once companies share data, they cannot control what happens with it; as a result, data is siloed and only 0.5% of it is used

Example: A botched partnership



A Wall Street Journal article revealed in 2019 that Google was collaborating with hospital chain Ascension, which operates more than 2,600 hospitals in the US.



This immediately raised privacy concerns about possible HIPAA violations and caused massive public backlash.



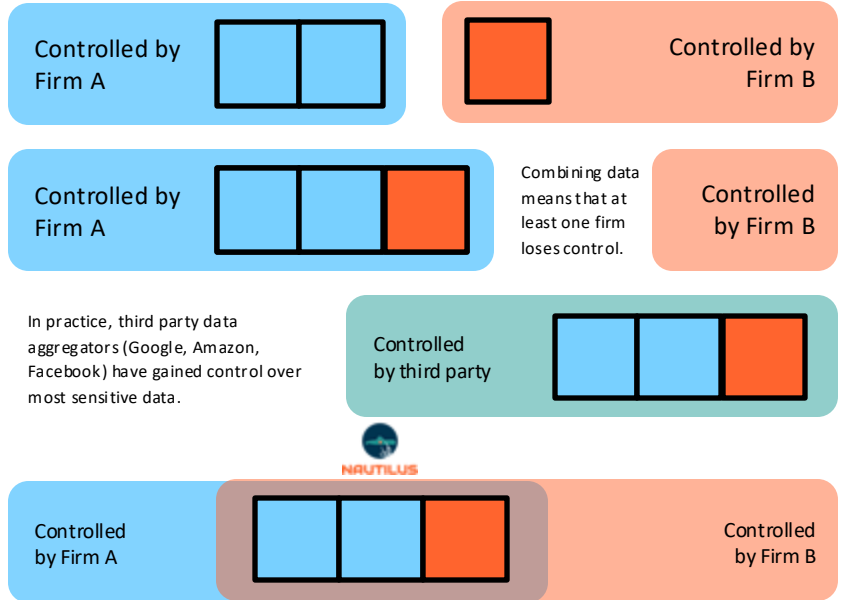
Congressional Democrats demand details on Google's use of patient data by Dec. 6

Published: 11/15/2019 10:00 AM EST | Updated: 11/15/2019 10:00 AM EST

Google health-data scandal spooks researchers

Scientists fear the controversy over the Nightingale project will undermine trust in research.

Today, users can control their data only until they want to combine it with someone else's data. Either one user must share their data with the other, or both must share it with a third party.



Using the Nautilus platform, users can combine their data without losing control over how it is used in the future and without loss of privacy.

The lack of control over data impedes research

the inability to access sensitive data is one of the biggest impediments to academic research today



Open data improves visibility

In many fields, data is confidential. This **limits reproducibility** and diminishes trust in empirical results.

By contrast, open data is associated with higher citation rates and better research visibility.^{5,6}



Data management is complex

Restricted-use data requires complex access management involving **long application and vetting processes**.⁷

These processes could be simplified if, instead of limiting access, data management systems could ensure confidentiality by limiting use.



Data silos limit research, reproducibility

While most researchers support data sharing, most data remains in **closely guarded silos**.⁸

This delays scholarly research, hinders innovation, and adds huge unnecessary costs for society.



Data silos limit collaboration

Researchers do not have the tools to **share and combine research data** with others without loss of control.

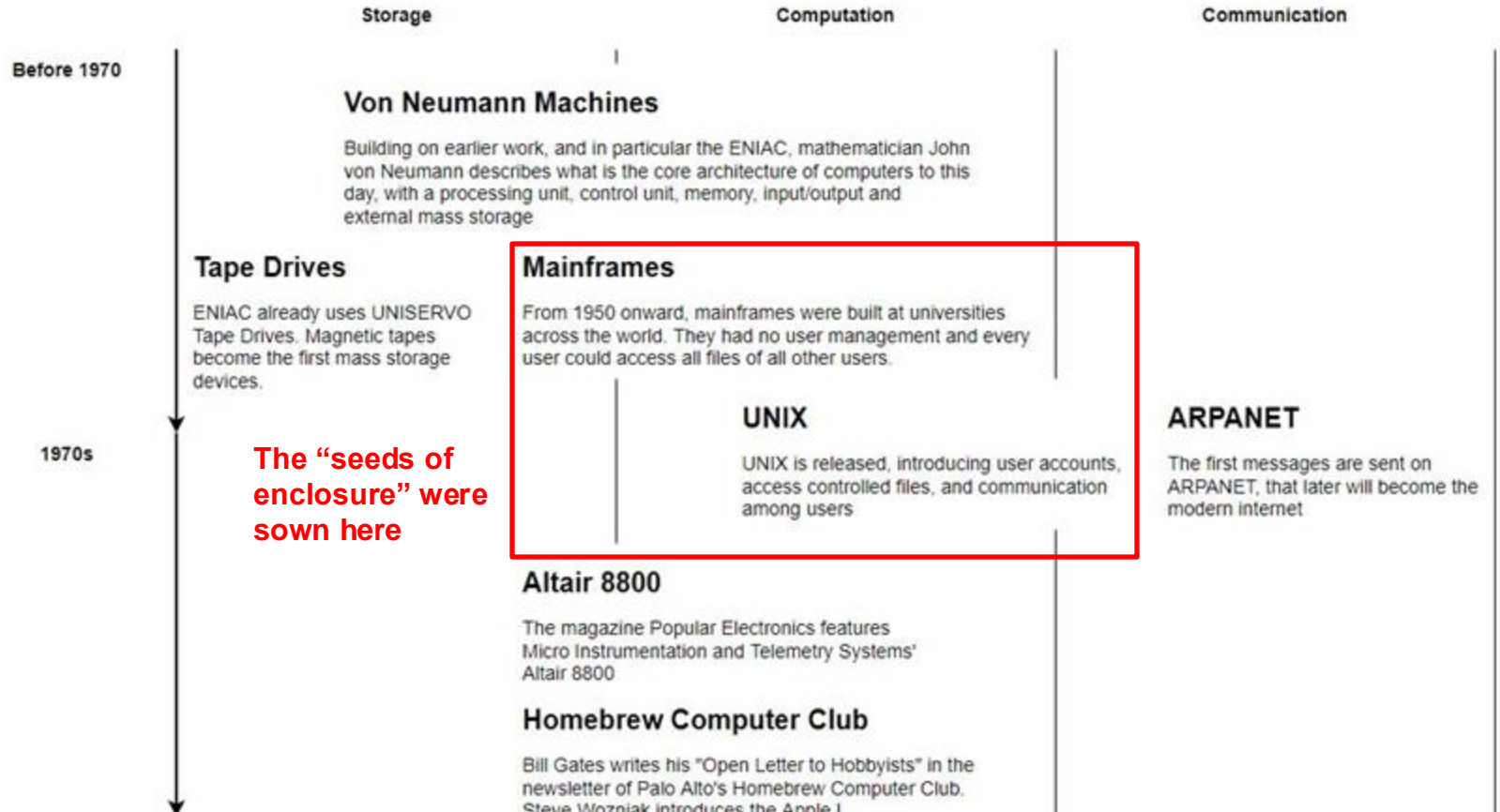
This limits the ways in which researchers can collaborate.

[5] *Privacy-preserving data sharing infrastructures for medical research: systematization and comparison*, [PLoS ONE](#), 2007 | [6] *Creating value through open data*, [European Data Portal](#), 2015 | [7] *Reported Individual Costs and Benefits of Sharing Open Data*, [BioScience](#), 2021 | [8] *Open-access policy and data-sharing practice in UK academia*, [Journal of Information Science](#), 2019

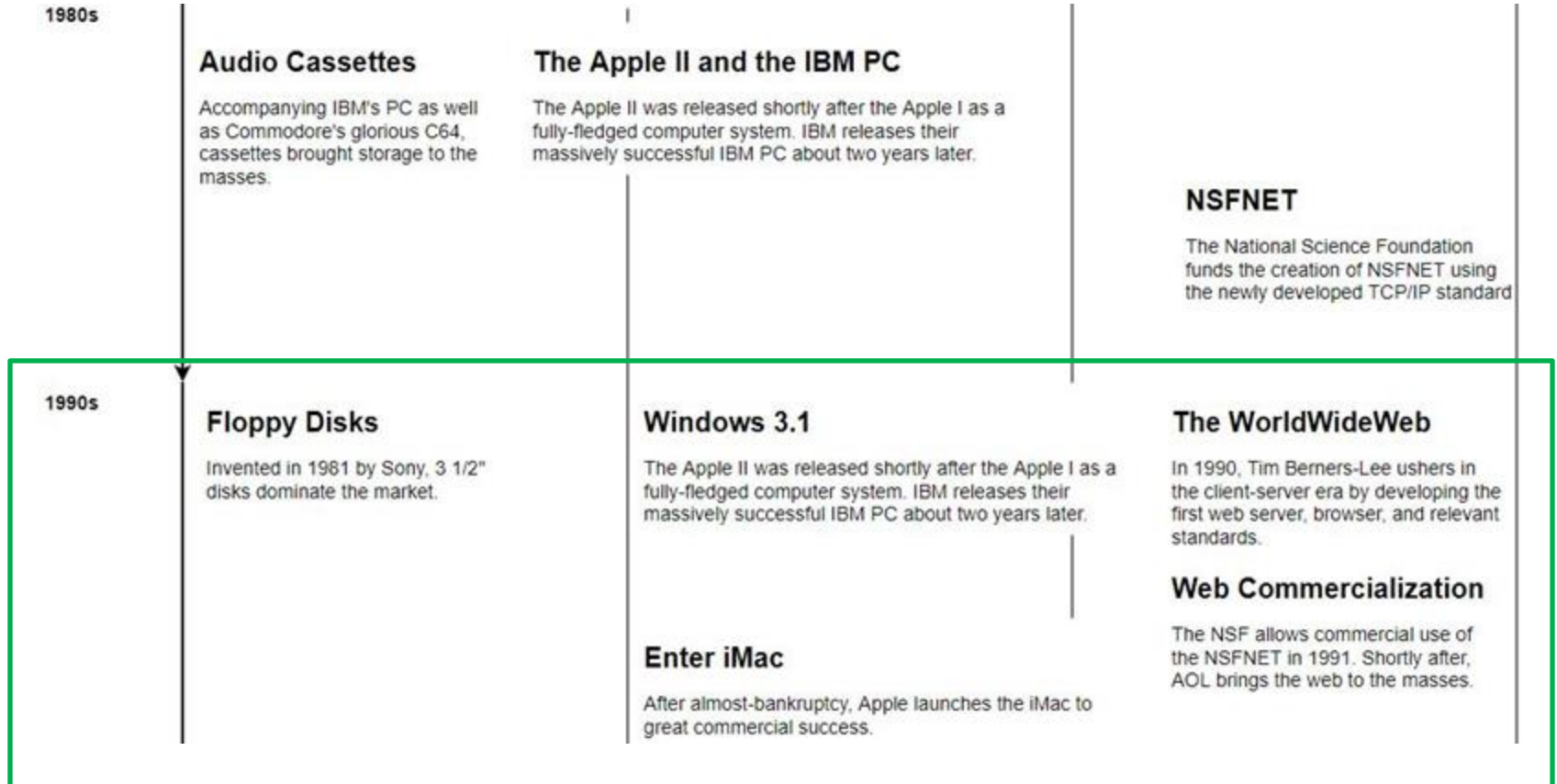
Precursor: Data is relational, but data control is absolute

(The aggregation of data results in externalities between data creators)

In the early days, there were no users, just data

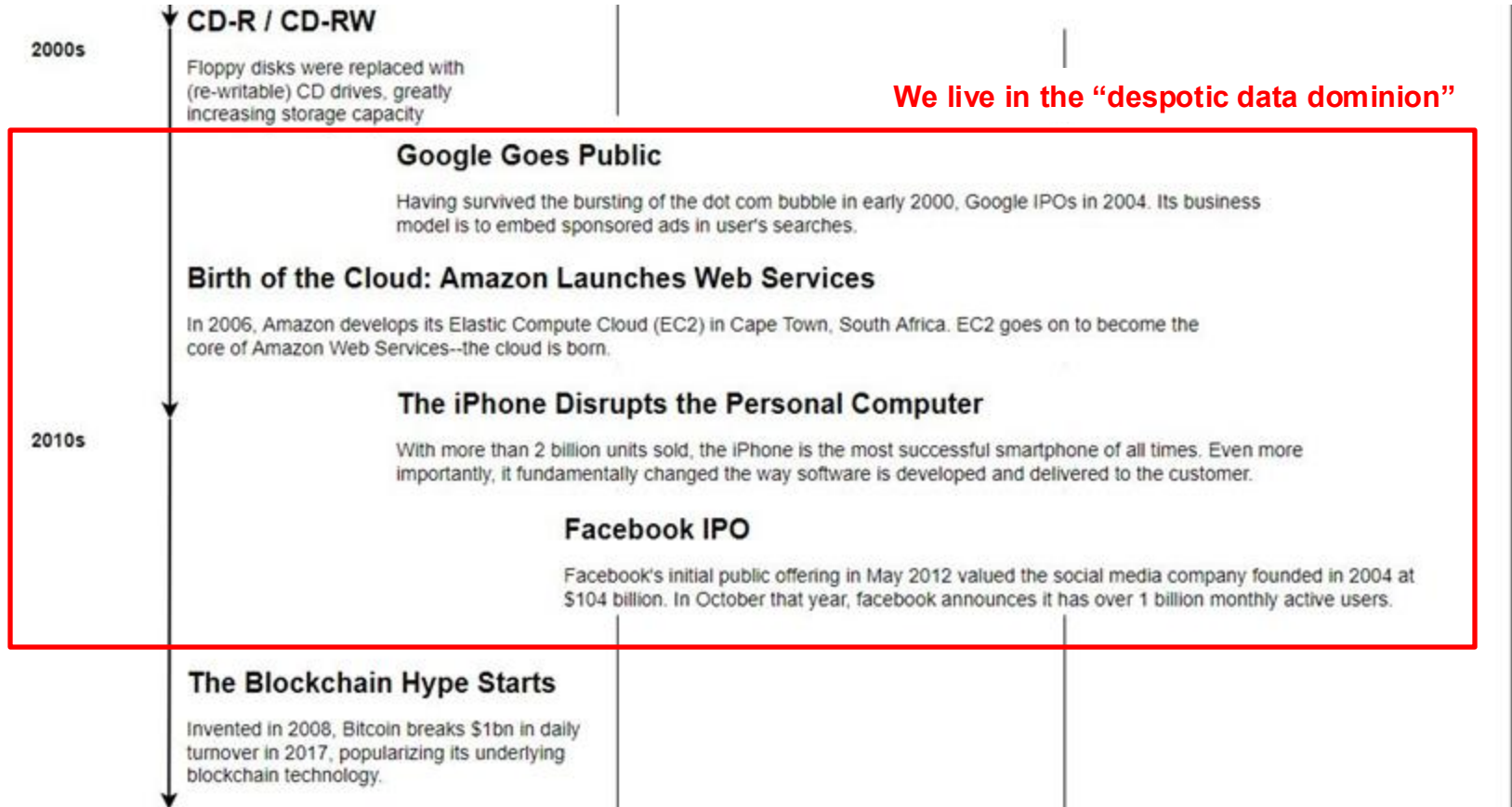


The battle starts once data and code are enclosed



The “golden age” of data control

What happened in the early 2010s?



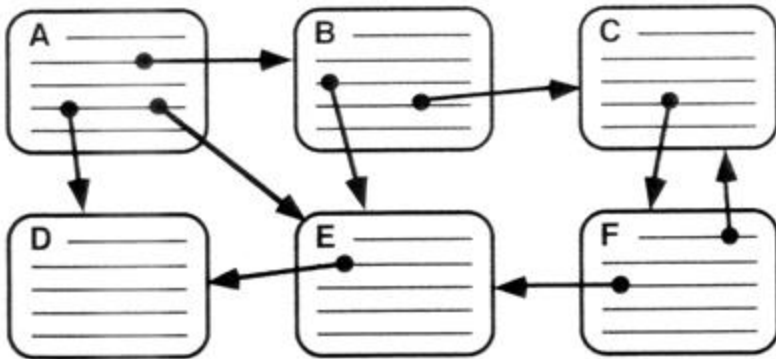
Focus: The Battle Over the Provenance of Data



UNIX implements copy as “write-to”



Ken Thompson and Dennis Ritchie build the first UNIX



Hypertext uses one-way linking to connect documents



Tim Berners-Lee, inventor of the WorldWideWeb

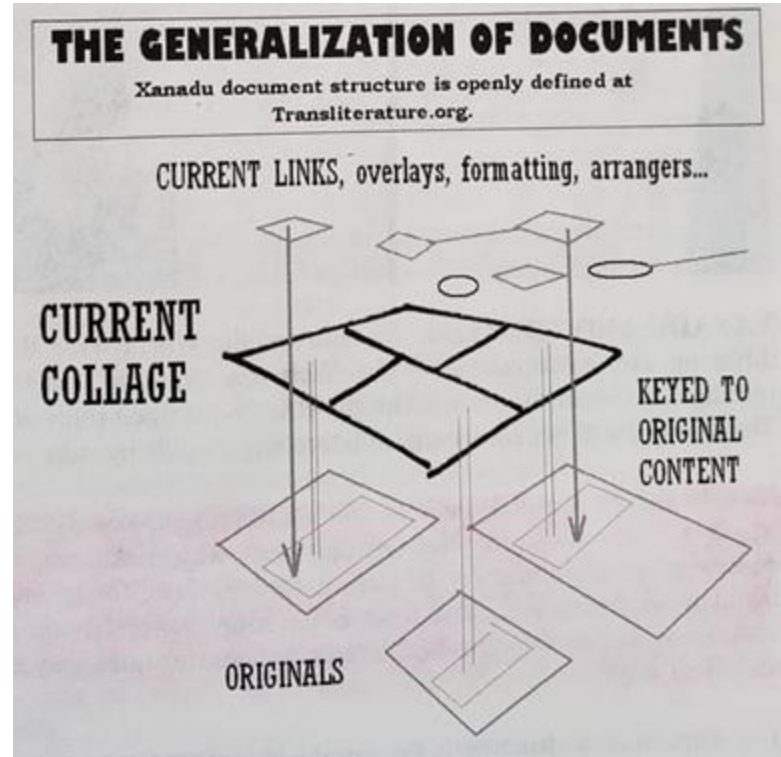
Focus: The Battle Over the Provenance of Data

The alternative

In 1967, Ted Nelson worked with Andries van Dam and others at Brown to implement the first Hypertext.

He leaves the project over a disagreement about the bi-directional linking and proposes **Xanadu**.

Xanadu was **never finished** and Wired calls it *“the longest-running vaporware story in the history of the computer industry”*



**Conclusion: Ted Nelson was
right**

**Data needs provenance
(and hypertext should be bi-directional)**

Three examples where **information provenance** exists today



Academic Research

Academics **attribute credit** by citing all relevant research

Provenance enforced as a **social norm**



Open Source Software

OOP encourages **re-use** of existing code via libraries and modules

Provenance enforced as a **technical necessity**



Digital Art

Digital artists often create mash-ups by re-using existing work

Provenance possible, but not enforced -> NFTs?



How to Create Data Provenance for Private Data



Digital Rights Tokens

Digital Rights Tokens encode users' rights

Tradability of tokens ensures efficient allocation of rights

What to include in a DRT?

1. Rights are specified as (open source) code, provided and persisted in a publicly accessible repository
2. Rights are specified in relation to a data pool, identified by a unique hash
3. A cryptographic hash of the data pool and the executable code

+
+

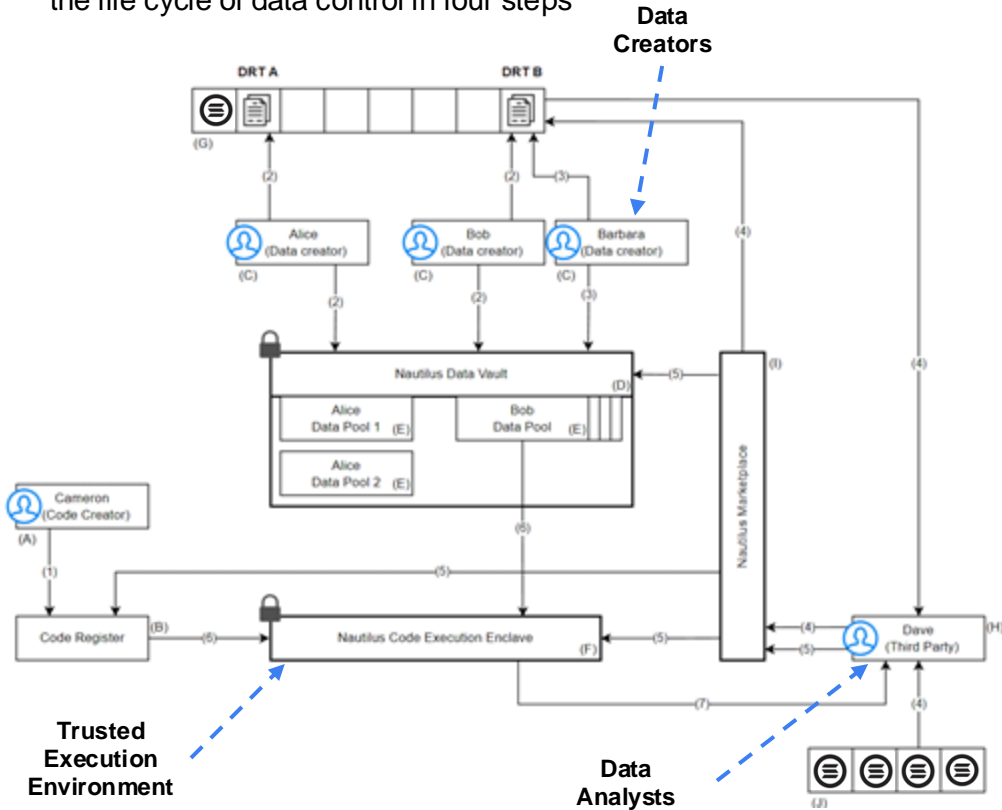
=

Secure Enclaves
Blockchain
Oracles

Data Control

a system to manage digital rights tokens

the life cycle of data control in four steps



Step 1: Set up data vault and issue DRTs

Alice and Bob create personal data vaults using a vault app. They are charged a storage fee per month and GB but can issue and sell digital rights tokens. DRTs include a reference to code that can be executed on the encrypted data.

Step 2: Turn data vaults into data pools

A specific DRT Bob could issue is one that allows other data creators to add encrypted data to the vault and turn it into a data pool—as long as appended data and original data have the same schema. Both Bob's original data and Barbara's additional data are encrypted and accessible only to the Trusted Execution Environment.

Step 3: Issue DRTs for registered code

Registered code can be referenced in DRTs, which ensures that data creators stay in control over which code can be executed on their data pools.

Step 4: Data analysts trigger code execution

Data analysts can acquire DRTs on a marketplace using crypto and redeem them at the Code Execution Enclave. The enclave will pull the code referenced in the DRT and instruct the enclave hosting the data pool to execute it. The result of the code execution is reported to the data analyst. Sigma charges a fee for code execution.

**For the first time in history, we have the tools to fight
against digital despotism**



Contact: co.georg@fs.de