LibreOffice
The Document Foundation

FOSS and Open Standards for Digital Sovereignty

Italo Vignoli

What is Digital Sovereignty?

Digital sovereignty is the ability of a state or a federation of states to provide the digital technologies it deems critical for its welfare, competitiveness, and ability to act, and to be able to develop these or source them from other economic areas without one-sided structural dependency

How to Achieve Technology Sovereignty?

Influence open standards to direct international markets towards European technologies, as well as patent pools or Open Source software and hardware, to prevent the structural dependency from tech monopolies.


Google
FILTERS YOUR THOUGHTS
Google filters your search results and YouTube recommendations to analyze your reactions and confine you to its filter bubble
laquadrature.net
Apple
KNOWS WHERE YOUR MOM IS
If you have an iOS or Android smartphone, Apple and Google track, collect and analyze your location without telling you or giving you a choice
laquadrature.net
facebook
CHOOSES WHAT YOU READ
Facebook selects the contents of your newsfeed to analyze your reactions and confine you to its filter bubble
laquadrature.net
amazon
KNOWS WHAT PRESENTS YOU ARE GETTING
Analyzed on a sufficient scale, behaviours that might seem trivial can reveal a lot about your personality, your expectations and your peers
laquadrature.net
Microsoft
FORMATS YOUR KIDS
Microsoft is buying its way into schools and universities by striking deals to train teachers and students
laquadrature.net

Subservience to Microsoft

Is that true? Is that real? Unbelievable, Microsoft rules

Human Genes Renamed To Please Excel

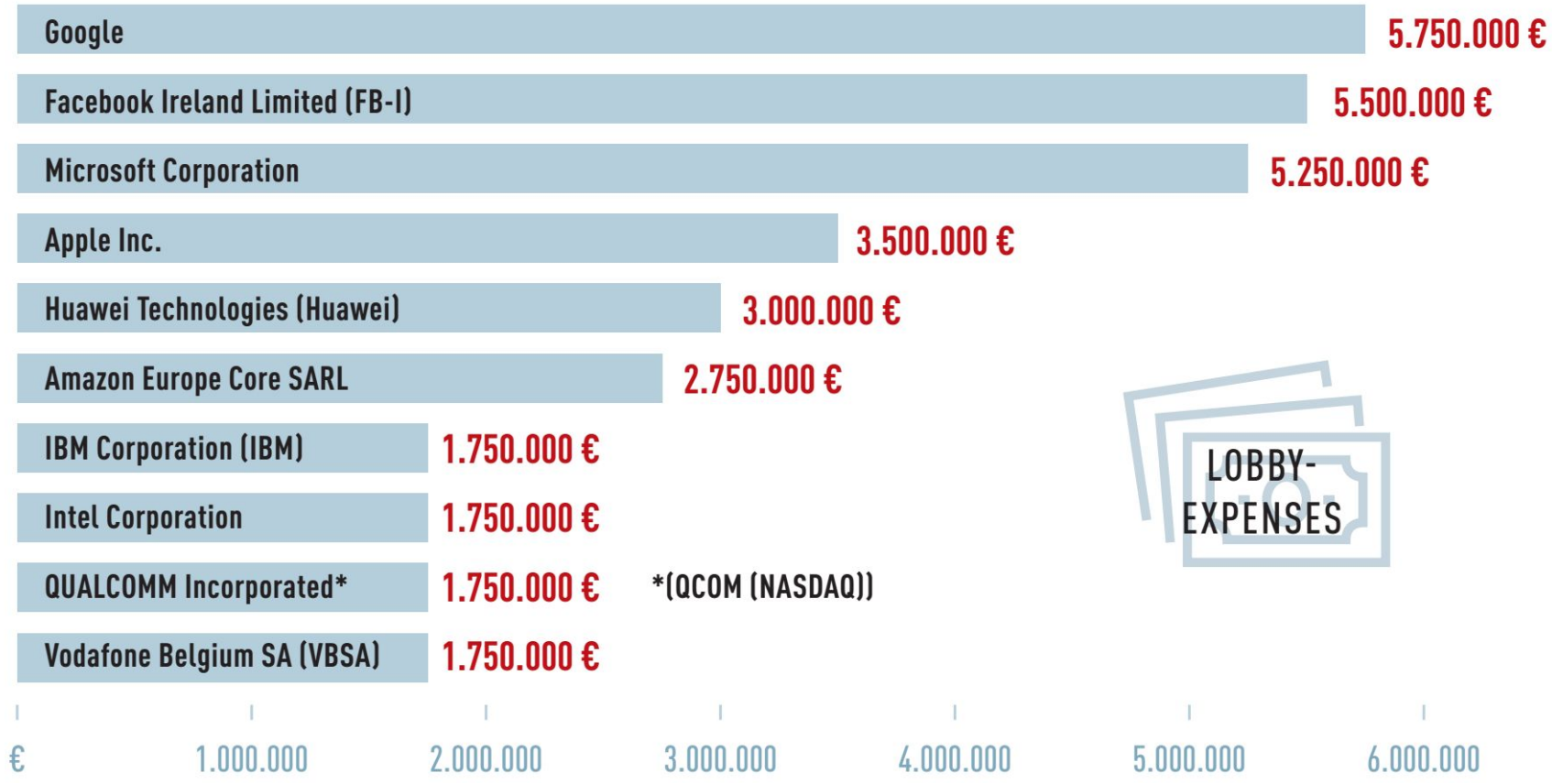
Written by Janet Swift

Friday, 07 August 2020

More than two dozen human genes have been renamed so that they can be typed into a spreadsheet without being formatted as dates. New guidelines for standardized gene naming explicitly allow for renaming genes to avoid problems with data handling.

- As Big Tech's market power has grown, so has its political clout.
- Just as the EU tries to rein in the most problematic aspects of Big Tech, digital giants are lobbying hard to shape new regulations.
- They are being given disproportionate access to policy-makers and their message is amplified by a wide network of think tanks and other third parties.
- Corporate Europe Observatory and LobbyControl profile Big Tech's lobby firepower, given it is now the EU's biggest lobby spending industry.

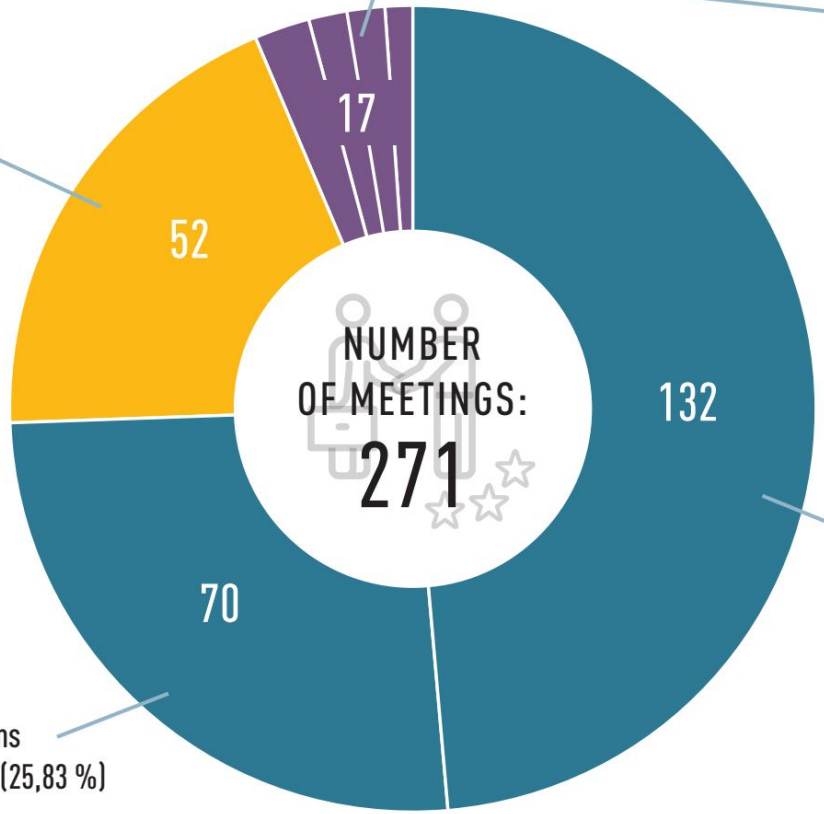
TOP 10 LOBBY SPENDERS OF THE DIGITAL INDUSTRY[17]



Impact of Lobby Budgets

- These huge lobbying budgets have a significant impact on EU policy-makers, who find digital lobbyists knocking on their door on a regular basis (more than 140 lobbyists work for the largest ten digital firms in Brussels and spend more than € 32 million on making their voice heard)
- Big Tech companies don’t just lobby on their own behalf, but they also employ an extensive network of lobby groups, consultancies, and law firms representing their interests, not to mention a large number of think tanks and other groups financed by them

MEETINGS WITH THE EU-COMMISSION ON THE DSA/DMA



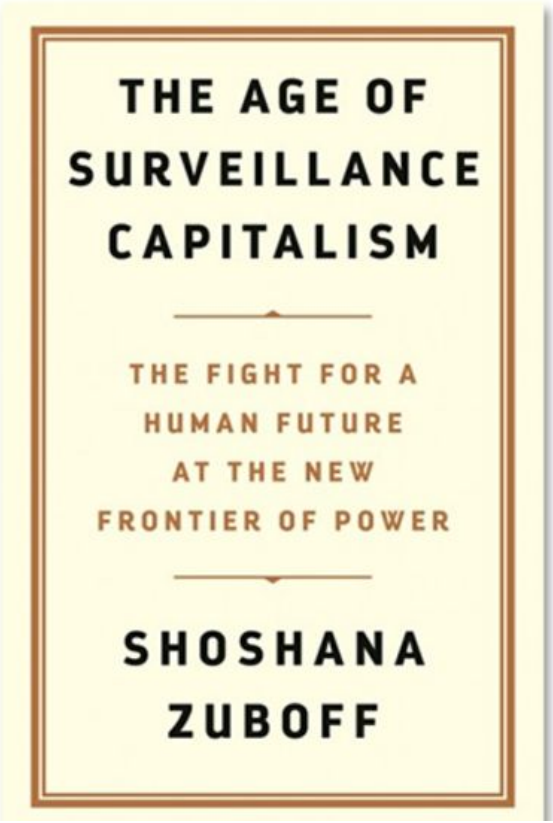
- 6 Professional consultancies (2,21 %)
- 4 Other organisations (1,48 %)
- 4 Think tanks and research institutions (1,48 %)
- 3 Other public or mixed entities, created by law whose purpose is to act in the public interest (1,11 %)

Companies & groups (48,71 %)

NGOs, consumer organisations and trade unions (19,19 %)

Trade and business associations and professional associations (25,83 %)

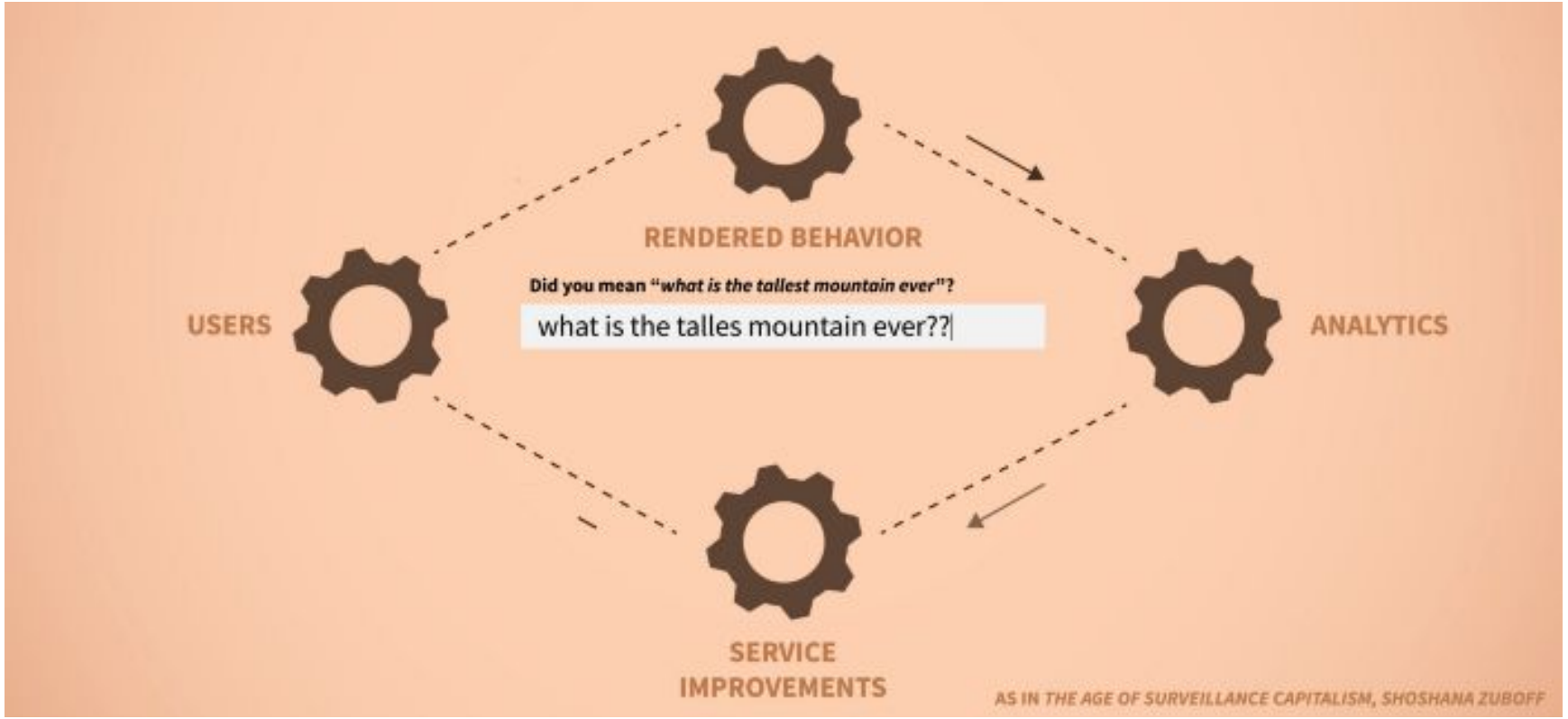
Surveillance Capitalism



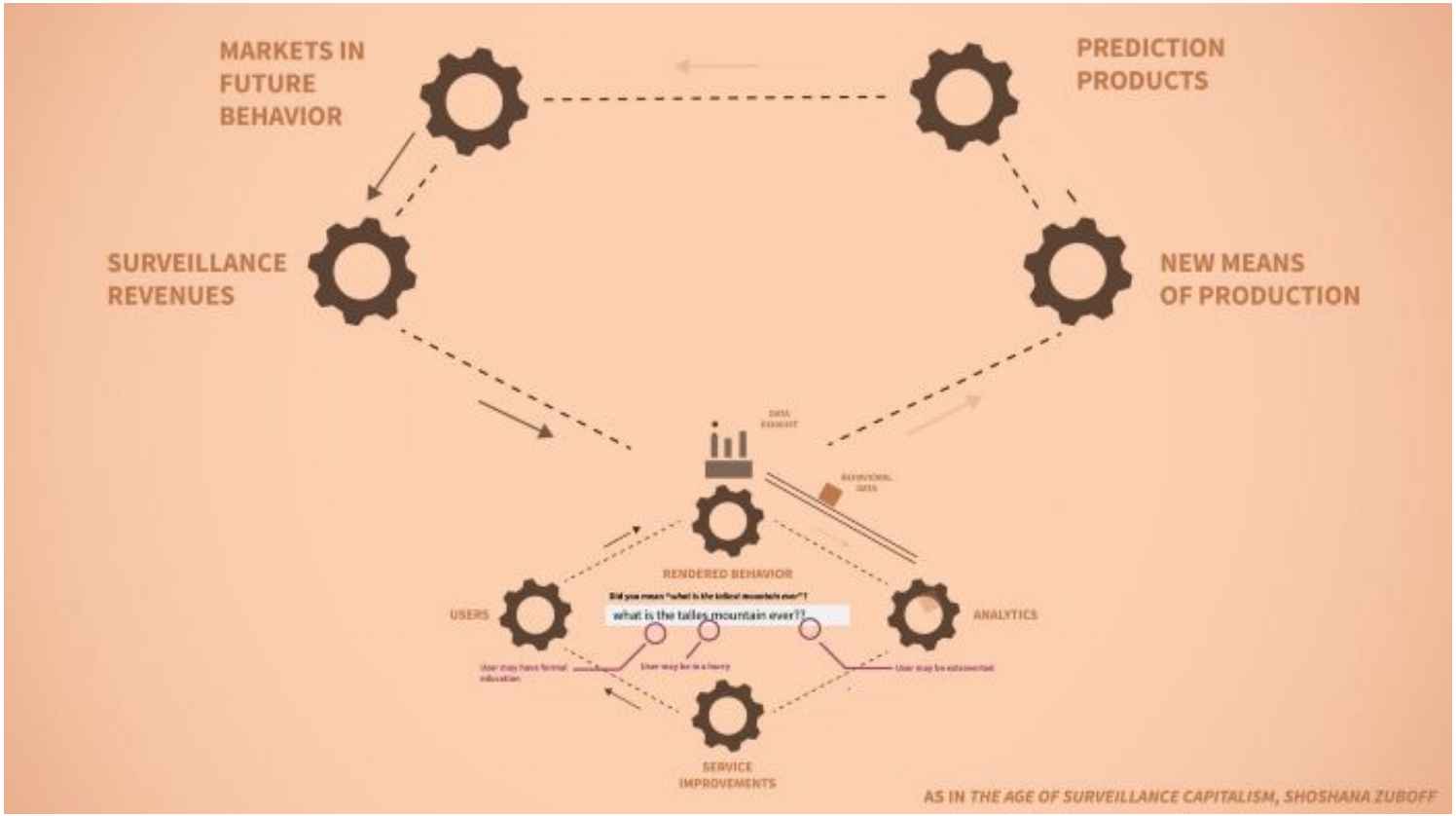
Defining Surveillance Capitalism

- Surveillance capitalism commandeered the wonders of the digital world to meet our needs for effective life, promising the magic of unlimited information and a thousand ways to anticipate our needs and ease the complexities of our harried lives.
- Under this new regime, the moment our needs are met is also the moment at which our lives are plundered for behavioral data, and all for the sake of others' gain.

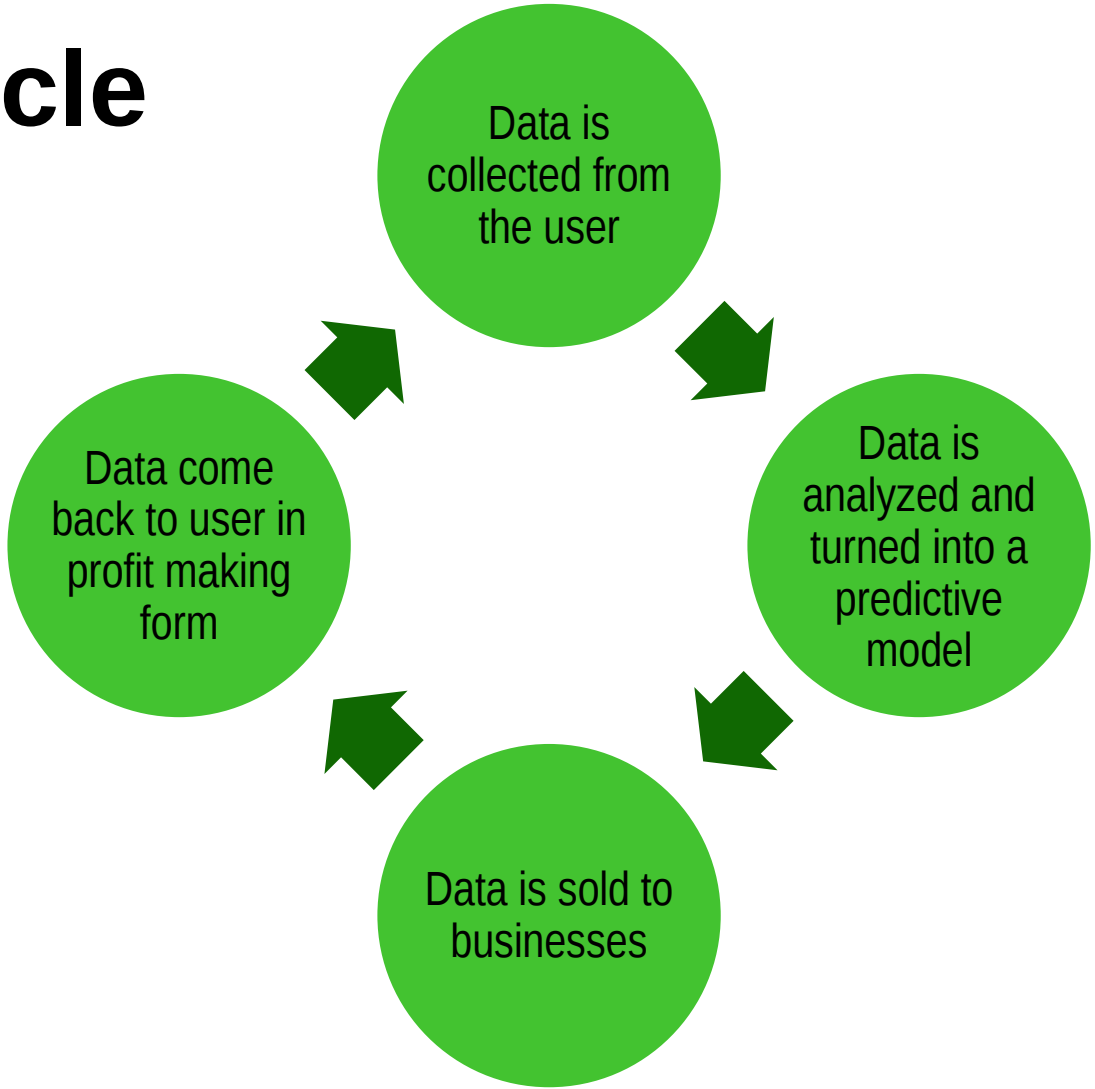
Behavioral Value Reinvestment



Discovery of Behavioral Surplus



The Cycle


Data is collected from the user
Data is analyzed and turned into a predictive model
Data is sold to businesses
Data come back to user in profit making form

What Can We Do?

- Collective action and awareness against surveillance capitalism
- New laws that protect citizens against surveillance capitalism
- Governments needs to stand-up and protect citizens

EU Digital Sovereignty: State of Play

The influence of non-EU tech companies is a concern for EU policy-makers, especially with regard to their impact on the EU's data economy and innovation potential, on EU privacy and data protection and on the establishment of a secure and safe digital environment

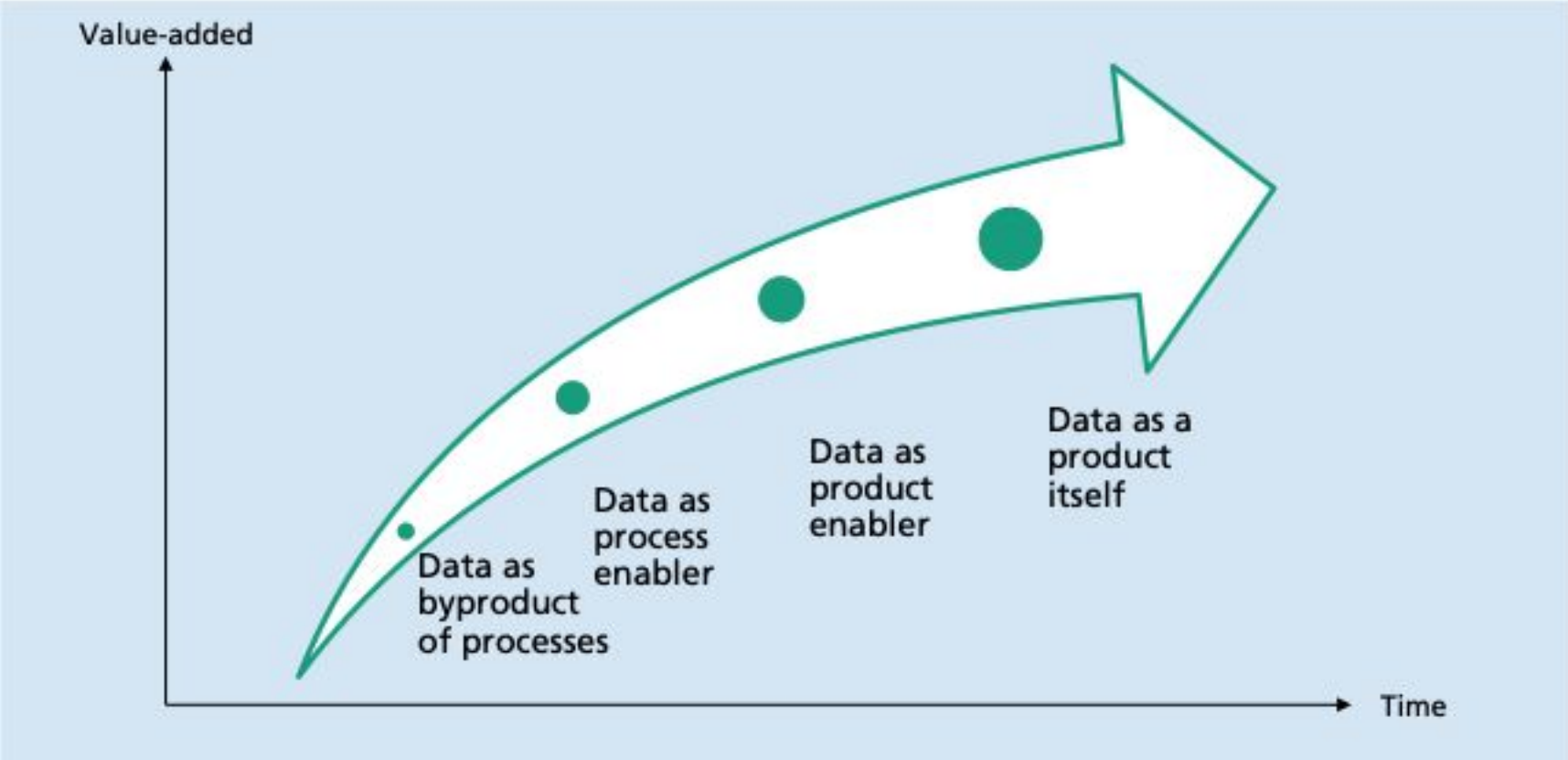
EU Digital Sovereignty: Initiatives

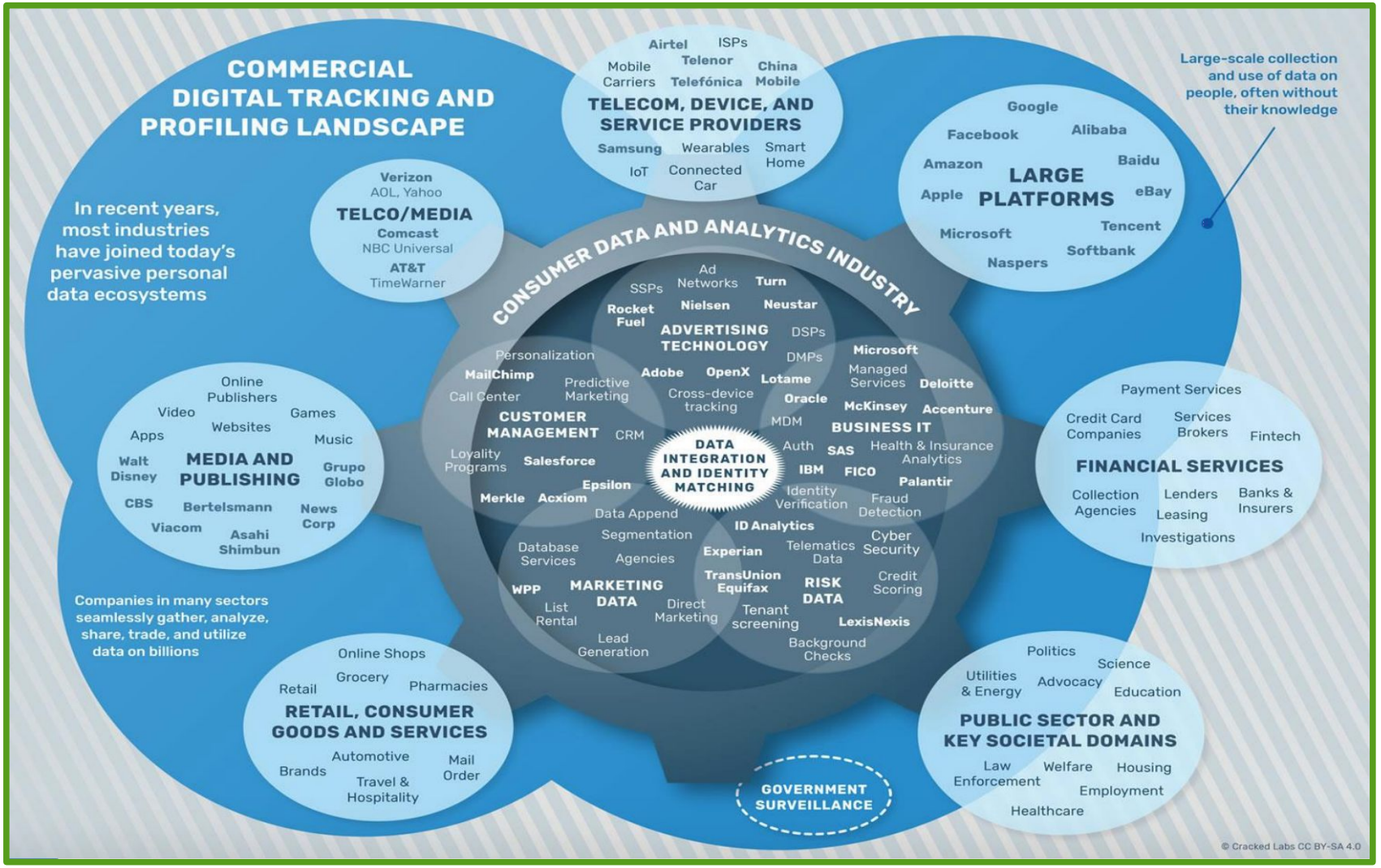
Reliable digital infrastructure and services are critical in today's society. A range of initiatives have been proposed or are already under discussion at EU level to accelerate the digitalisation process and enhance Europe's autonomy in the digital field around three building blocks of (i) building a data framework, (ii) promoting a trustworthy environment, and (iii) adapting competition and regulatory rules

EU Digital Sovereignty: Solutions

Building a secure pan-European data framework and adopting new standards and practices to provide trustworthy and controllable digital products and services would ensure a safer digital environment, in line with EU values and principles. Furthermore, in the competition and regulatory framework, a shift towards more defensive and prudential mechanisms, including new rules to address foreign state ownership and large tech companies' distortive practices, would seem desirable to achieve more technological autonomy.

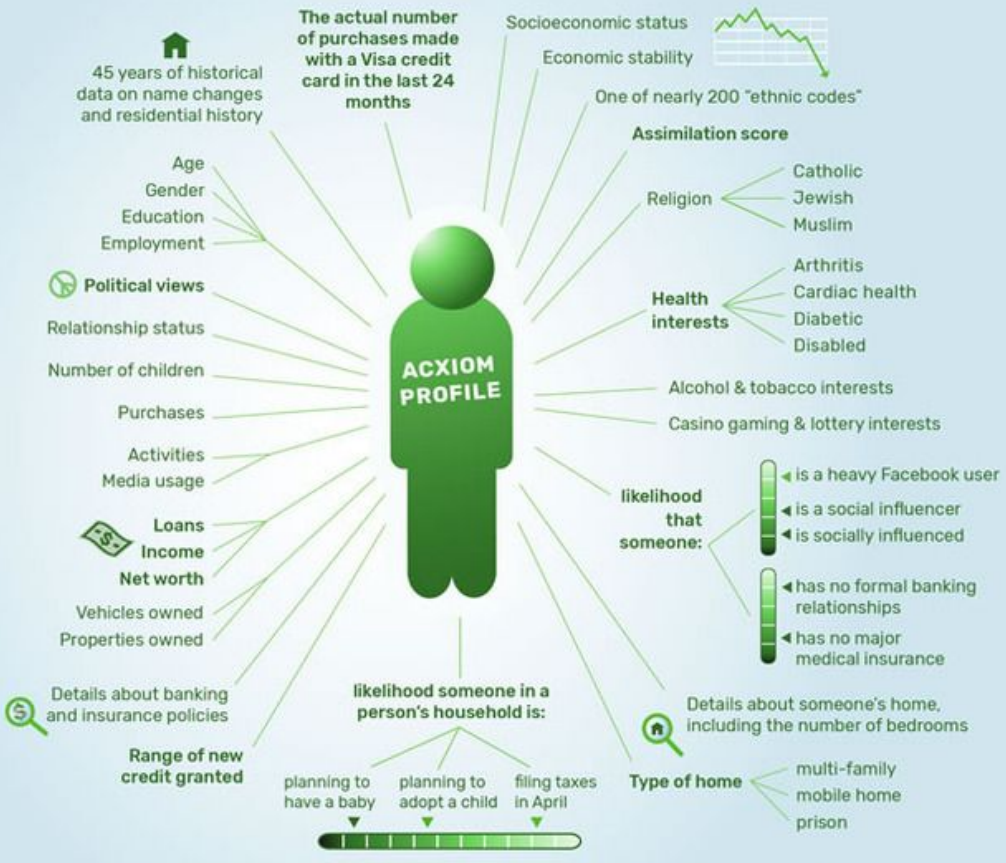
Evolution of Data



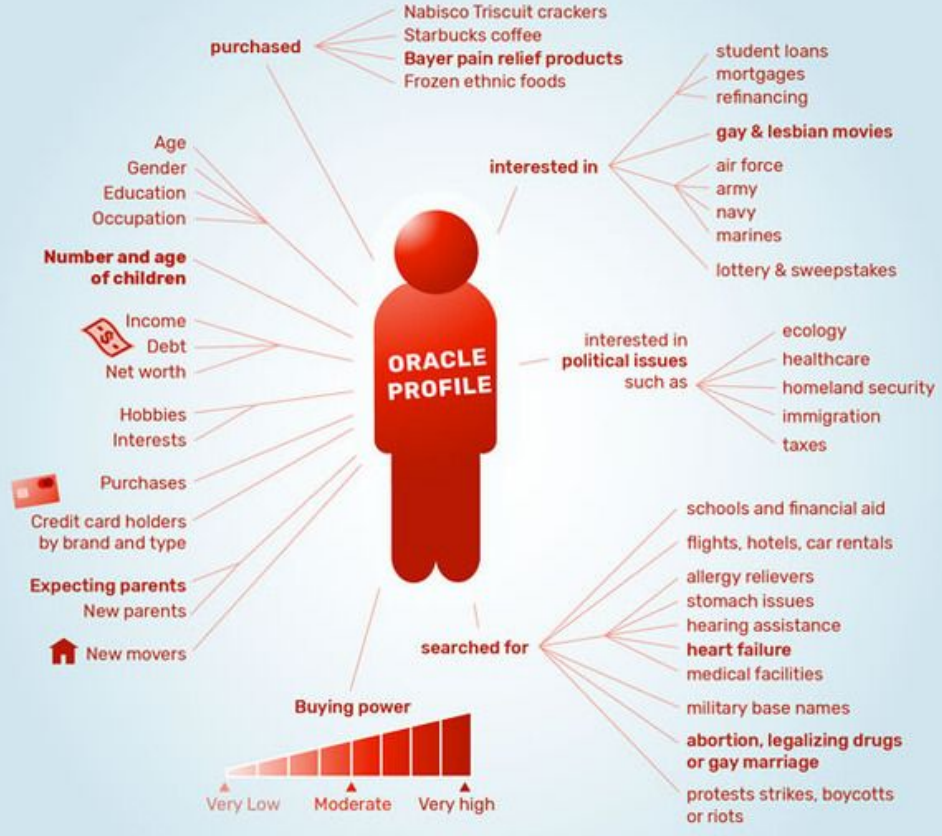

COMMERCIAL DIGITAL TRACKING AND PROFILING LANDSCAPE
In recent years, most industries have joined today's pervasive personal data ecosystems
Companies in many sectors seamlessly gather, analyze, share, trade, and utilize data on billions
Large-scale collection and use of data on people, often without their knowledge
TELECOM, DEVICE, AND SERVICE PROVIDERS
Airtel
ISPs
Telenor
Mobile Carriers
Telefónica
China Mobile
Samsung
Wearables
Smart Home
IoT
Connected Car
TELCO/MEDIA
Verizon
AOL, Yahoo
Comcast
NBC Universal
AT&T
TimeWarner
LARGE PLATFORMS
Google
Facebook
Alibaba
Amazon
Baidu
Apple
eBay
Microsoft
Tencent
Softbank
Naspers
CONSUMER DATA AND ANALYTICS INDUSTRY
ADVERTISING TECHNOLOGY
Ad Networks
SSPs
Turn
Rocket Fuel
Nielsen
Neustar
DSPs
DMPs
Personalization
MailChimp
Predictive Marketing
Adobe
OpenX
Lotame
Call Center
Cross-device tracking
Oracle
CUSTOMER MANAGEMENT
CRM
Loyality Programs
Salesforce
Epsilon
Merkle
Acxiom
DATA INTEGRATION AND IDENTITY MATCHING
Microsoft
Managed Services
Deloitte
McKinsey
Accenture
MDM
BUSINESS IT
Auth
SAS
Health & Insurance Analytics
IBM
FICO
Palantir
Identity Verification
Fraud Detection
Data Append
ID Analytics
Segmentation
Database Services
Agencies
Experian
Telematics Data
Cyber Security
WPP
MARKETING DATA
List Rental
Direct Marketing
TransUnion Equifax
RISK DATA
Credit Scoring
Tenant screening
LexisNexis
Lead Generation
Background Checks
MEDIA AND PUBLISHING
Online Publishers
Video
Games
Websites
Apps
Music
Walt Disney
Grupo Globo
CBS
Bertelsmann
News Corp
Viacom
Asahi Shimbun
FINANCIAL SERVICES
Payment Services
Credit Card Companies
Services Brokers
Fintech
Collection Agencies
Lenders
Banks & Insurers
Leasing
Investigations
RETAIL, CONSUMER GOODS AND SERVICES
Online Shops
Retail
Grocery
Pharmacies
Brands
Automotive
Mail Order
Travel & Hospitality
PUBLIC SECTOR AND KEY SOCIETAL DOMAINS
Politics
Science
Utilities & Energy
Advocacy
Education
Law Enforcement
Welfare
Housing
Employment
Healthcare
GOVERNMENT SURVEILLANCE
© Cracked Labs CC BY-SA 4.0

DATA BROKERS HAVE EXTENSIVE PROFILE INFORMATION ON ENTIRE POPULATIONS

Examples of data on consumers provided by Acxiom and Oracle



Acxiom provides of up 3,000 attributes and scores
on 700 million people in the US, Europe, and other regions.



Oracle sorts people into thousands of categories
and provides > 30,000 attributes on 2 billion consumer profiles

Current vs Intended Paradigm

CURRENT

- Data is a fictitious commodity, that can be sold and traded in markets
- Data handling & monetization is opaque
- Even if there is regulation, there is no possible enforcement
- We produce the data, they own it
- Benefits of data sharing are privatized (surveillance capitalism)

INTENDED

- Data is a common resource & infrastructure where to build upon (new services)
- Data handling and usage is transparent & privacy-friendly
- Data is shared according to rules set by common (enforceable) governance
- Shared benefits of data sharing
- New political, economic, and legal regime that recognize social and communal rights to data

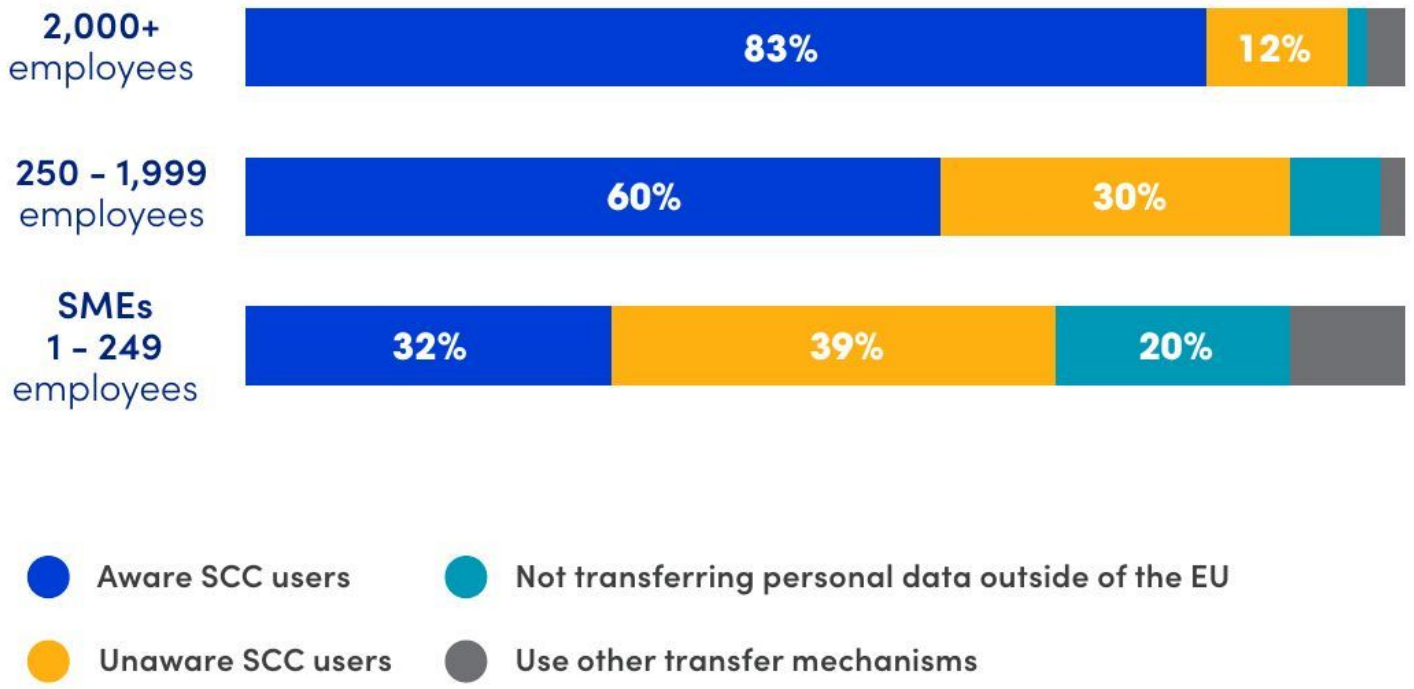
Enabling the Data Commons

- Can we regain control of our data?
- We want to decide who to share it with, under which rules, when and for what purpose
- One can decide in a democratic way based on trust but if data is a monopoly of a few, collective intelligence is lost…

Situation with Proprietary SW

- A large percentage of governments in Europe - at every level – rely on proprietary software for desktop productivity and cloud storage of data, independently from the level of confidentiality
- This puts citizen's personal data, including several extremely confidential information (health) at risk
- This is confirmed by the recent Schrems II sentence from the Court of Justice of the European Union

Awareness of SCC* (Schrems II)



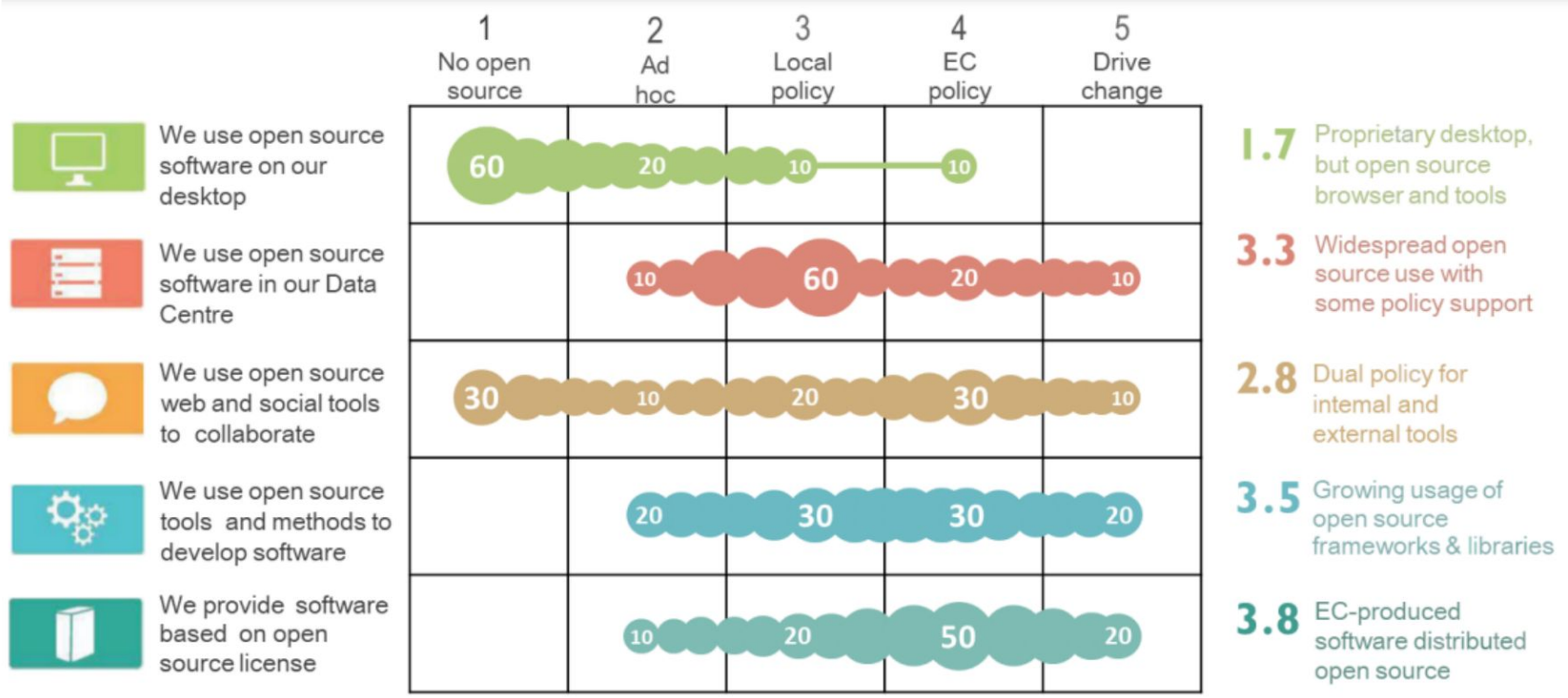
Source: DIGITALEUROPE I Base: All respondents (n = 292)[9]

* **SCC** = Standard Contractual Clauses (for data protection) when transferring data to the US

Proprietary versus FOSS

- Proprietary software protects the user by obfuscating algorithms and information, but in this way they also obfuscate the way they handle end user data
- FOSS protects the user with transparency, by sharing source code and all information about methodologies used by projects to manage end user data

EC OSS Adoption Maturity Index



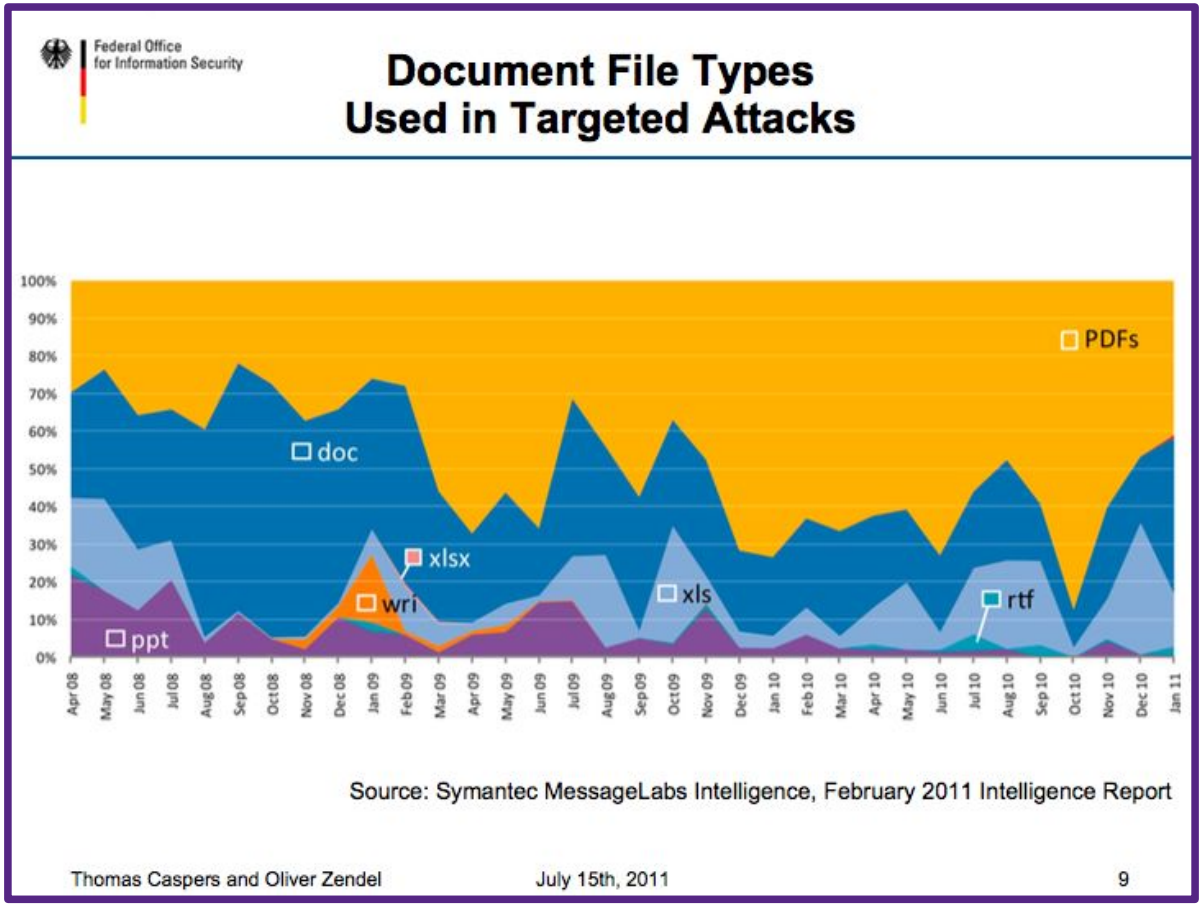
What Would Change with FOSS

- By switching to FOSS for desktop productivity and cloud storage, European governments would regain control of citizen's personal data and manage them according to their confidentiality
- In addition, switching to FOSS would include moving from proprietary to standard document formats, with a significant advantage in term of interoperability

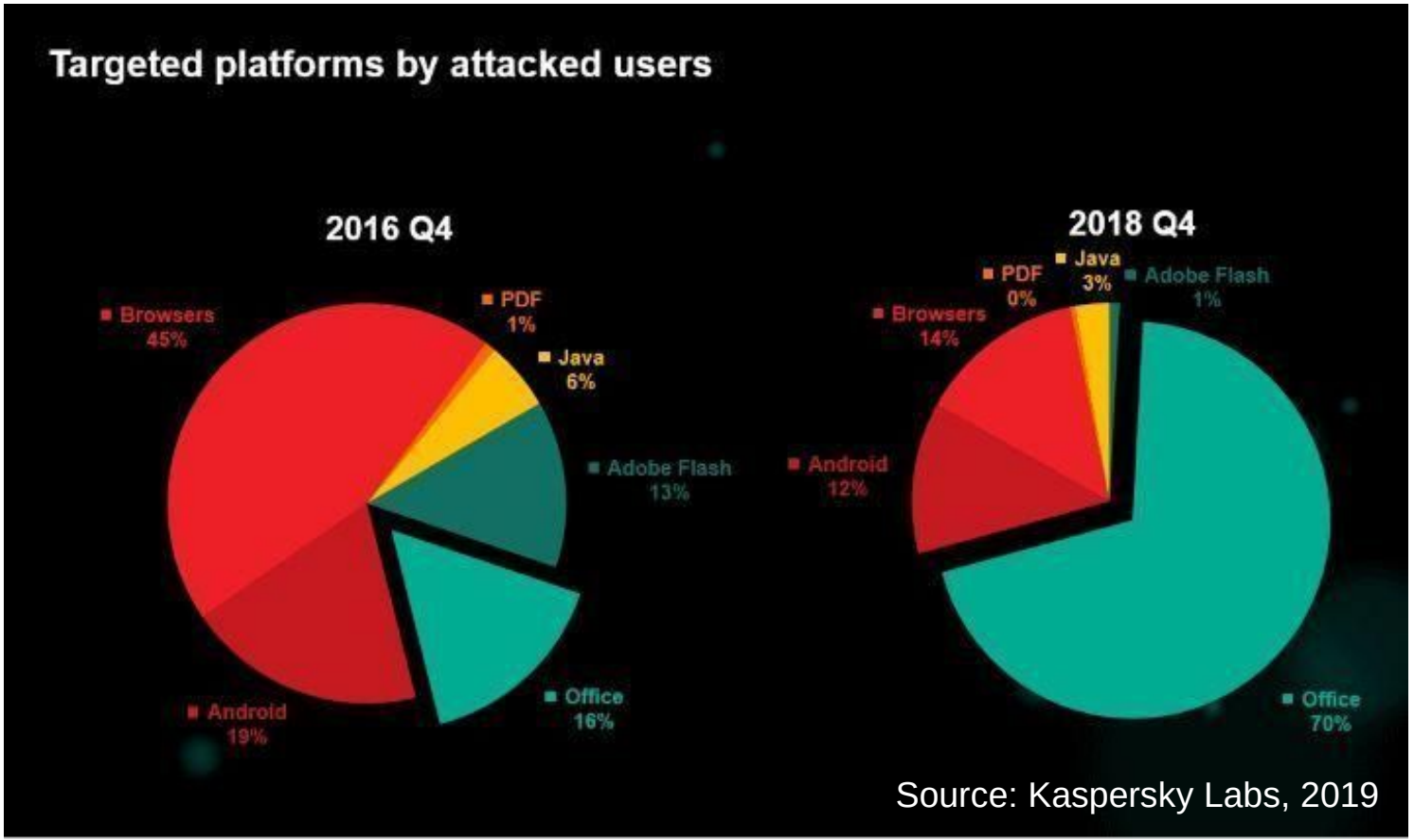
Apparently a No Brainer, but...

- Politicians – who are not technology experts – see GAFAMs as part of the global system, and therefore consider their issues as blockers for the entire digital transformation process (and try to help them)
- On the contrary, politicians – because of their limited understanding of technology – do not see FLOSS as part of the global system, and as a consequence do ignore FLOSS as a potential solution

Document Vulnerabilities in 2011



Document Vulnerabilities in 2018



Time to FOSS Vulnerability Fixes

FSFE Project

Public Money

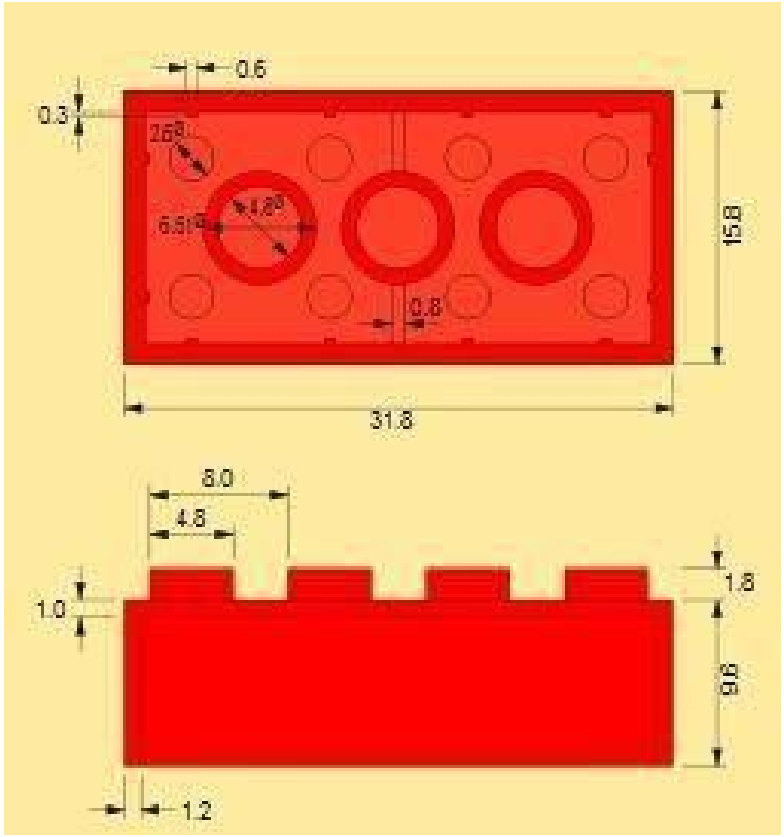
Public Code

publiccode.eu

Standard and Interoperability

Interoperability is the ability of information and communication technology (ICT) systems, as well as of the business processes they support, to exchange data and enable the sharing of information and knowledge.

European Interoperability Framework, IDABC



Importance of the HTML Standard

- It was the standardization of the HTML format that allowed the web to take off. And not just the fact that it's a standard, but the fact that it's open and royalty-free...
- Had HTML not been free and open, and a proprietary technology, the business of selling HTML and competing products would have been born...
- This means we need standards, because this avoids competition over technology, and fuels the value-added business built on the platform...

Tim Berners-Lee, CERN
world wide web inventor

Document Format as a Hindrance?

- Government should be platform independent and allow only true document standards, as pseudo standards can be tweaked in a way not visible to users to prevent document interoperability
- In fact, tweaked standards force citizens to pay a fee to create documents (purchase of a proprietary license), or to accept the intrusive license / spying conditions of a cloud based platform
- Only standards associated to FOSS can solve this problem

Open Document Format

the true document standard
which offers freedom of choice

Open Document Format

- **Independent** from a single product: anyone can write a software that handles an open format
- **Interoperable**: allows the transparent sharing of data between heterogeneous systems
- **Neutral**: it does not force the user to adopt – and often buy – a specific product, but leaves a wide choice based on features/quality vs price ratio
- **Perennial**: protects user developed contents from the “evolution” based obsolescence of technology

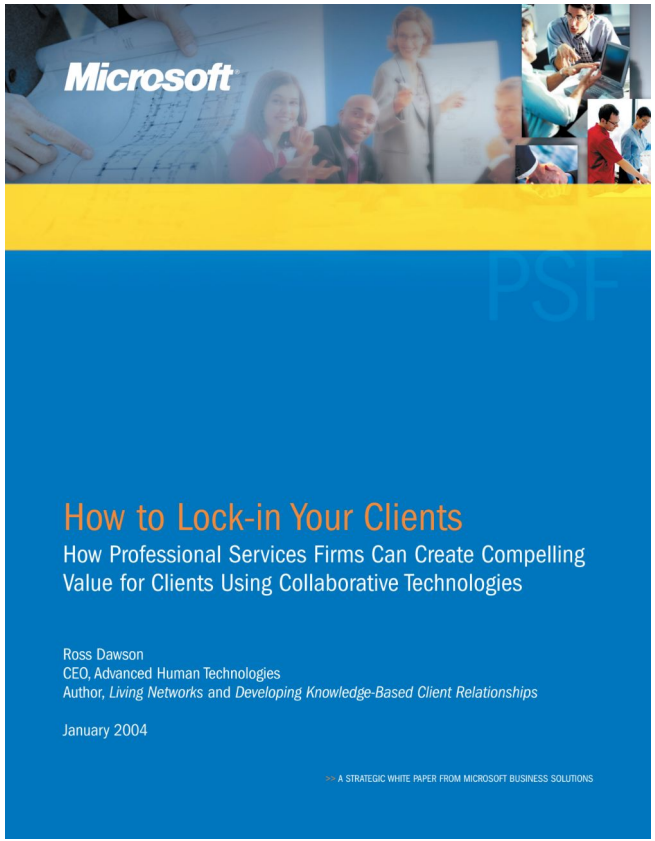
Basic Concepts

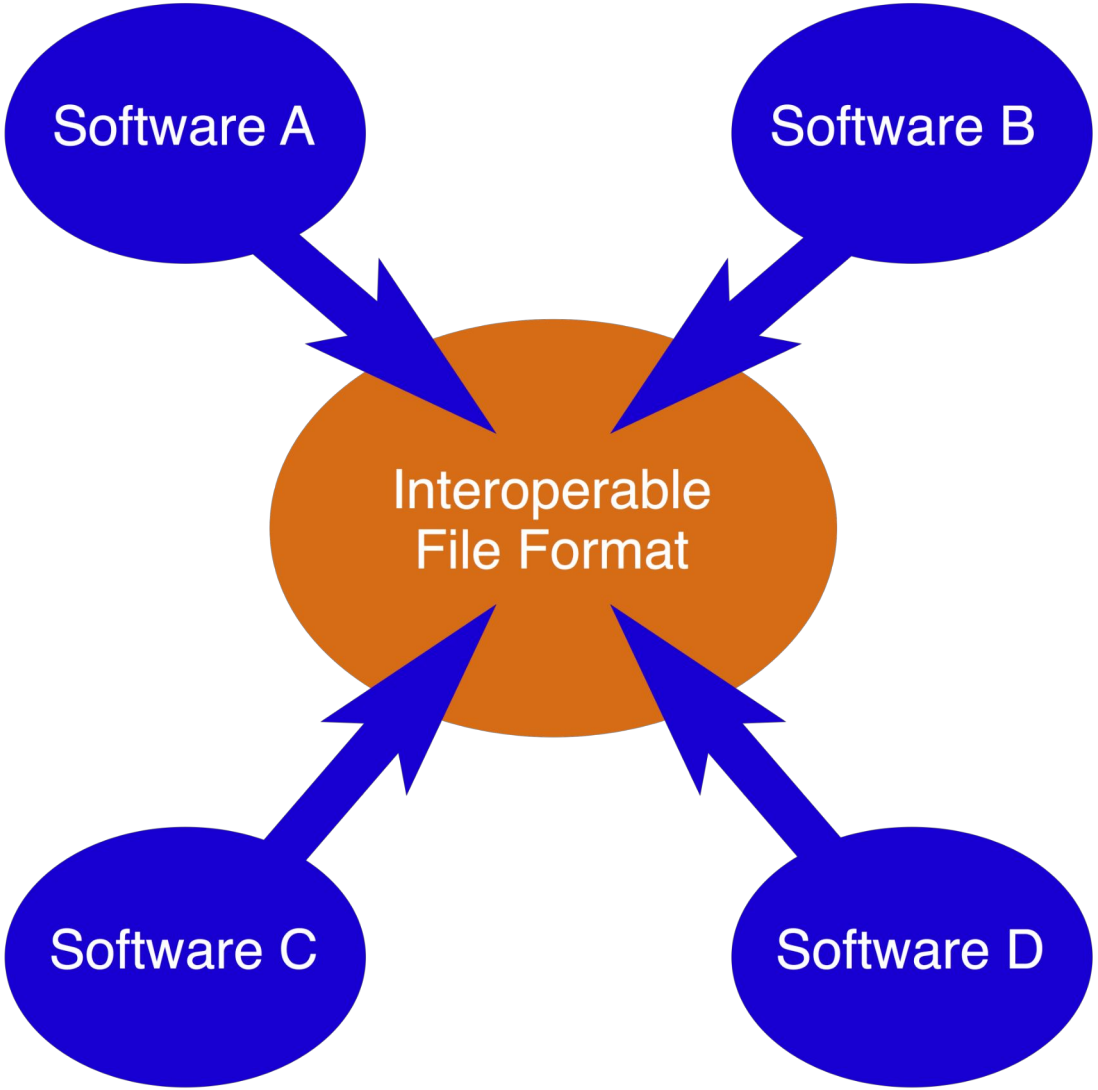
- ODF is solid and robust
- ODF is consistent across OS
- ODF is truly interoperable
- ODF is predictable
- **ODF is the best standard file format for users of personal productivity SW**

Lock In

WE CANNOT READ YOUR DOCUMENTS

DOCUMENTFREEDOM.ORG




Software A
Software B
Interoperable
File Format
Software C
Software D

Digging into Document Formats

OOXML Transitional and Strict

- As of 2020, the Office default for .docx, .xlsx and .pptx is Transitional OOXML, a proprietary document format which was created as a bridge from legacy MS Office formats and the approved ISO Standard.
- OOXML Strict is the ISO approved open standard, but being the non publicized last option on MS Office “file, save as…” menu has not been adopted, so 100% of existing OOXML files we are referring to are proprietary (non standard).

OOXML Strict Standard Support

- MS Office 2010: NO
- MS Office 2013: YES, but default is Transitional
- MS Office 2016: YES, but default is Transitional
- MS Office 2019: YES, but default is Transitional
- MS Office macOS: NO
- MS Office 365: NO
- According to Microsoft statements in 2007, OOXML Strict should have been the default since Office 2010

OOXML Philosophy

- The OOXML pseudo-standard document format appears to be designed by Microsoft for Microsoft products, and to inter-operate with the Microsoft environment
- Little thought appears to have been exercised for interoperability with non-Microsoft environments or compliance with established vendor-neutral standards

ODF Philosophy

- The philosophy behind the ODF standard document format was to design a mechanism in a "vendor neutral" manner from the ground up using existing standards wherever possible
- Although this means that software vendors would need to tweak their individual packages more than if they continued down their original routes the benefits for interoperability were important enough to justify the move

ODF vs OOXML Strategic Difference

- ODF has been designed as a document standard for the next 20-50 years, to liberate users from the lock-in strategy built into yesterday's and today's proprietary formats, and foster interoperability
- OOXML has been designed as a pseudo-standard document format to propagate yesterday's document issues and lock-in strategy for the next 20-50 years, to the detriment of users and interoperability

LibreOffice as Shakespeare (ODT)

2018

<text:p text:style-name="P1">To be, or not to be, that is the question</text:p>

2019

<text:p text:style-name="P1">To be, or not to be, that is the question</text:p>

2020

<text:p text:style-name="P1">To be, or not to be, that is the question</text:p>

MS Office as Shakespeare (DOCX)

2018

<w:t>To be</w:t>
<w:t>, or</w:t>
<w:t xml:space="preserve"> not to be</w:t>
<w:t>,</w:t>
<w:t xml:space="preserve"> that is the</w:t>
<w:t xml:space="preserve"> question</w:t>

MS Office as Shakespeare (DOCX)

2019

<w:t>To be</w:t>

<w:t>,</w:t>

<w:t xml:space="preserve"> or not to be, that </w:t>

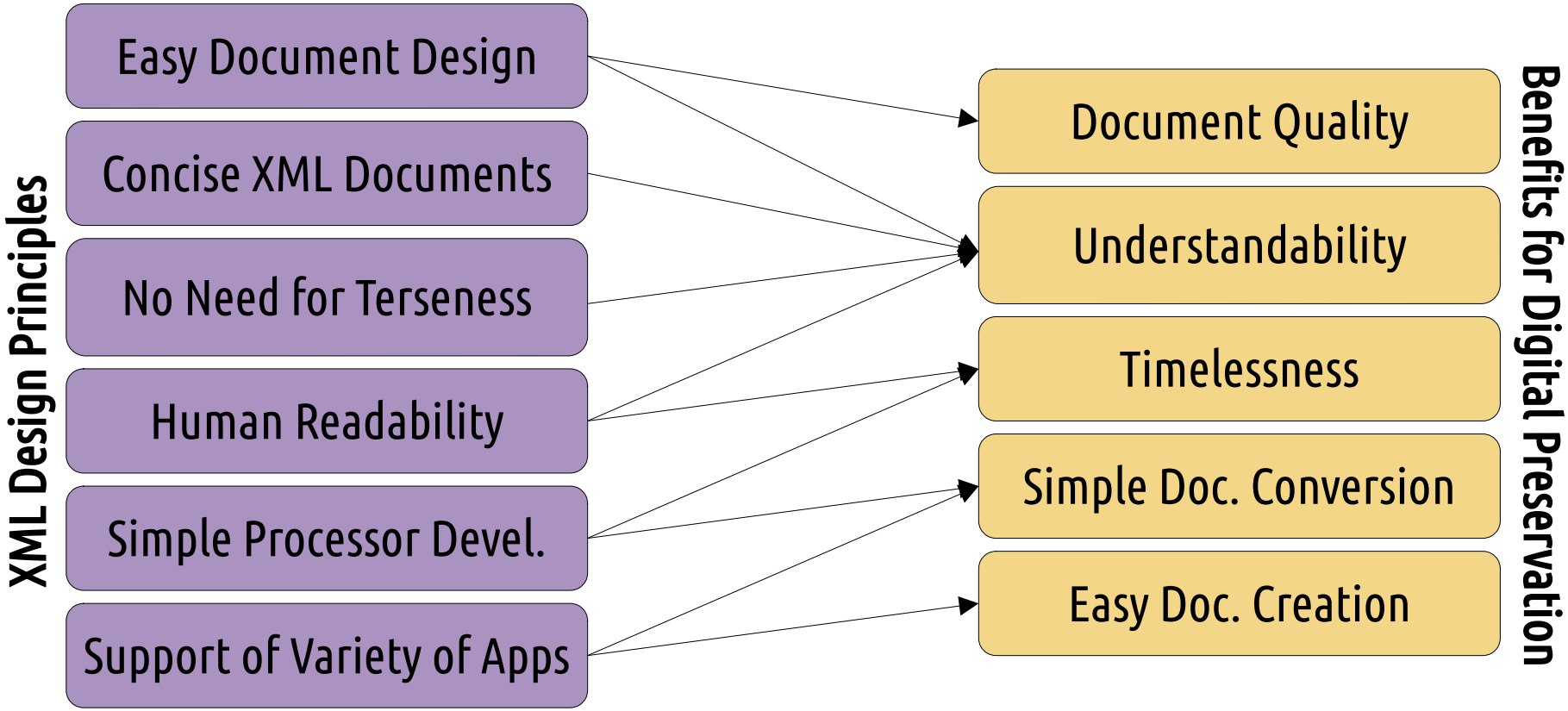
<w:t>is the question</w:t>

MS Office as Shakespeare (DOCX)

2020 (Office 365)

<w:t xml:space="preserve">To be, or </w:t>
<w:t>not</w:t>
<w:t xml:space="preserve"> to be, </w:t>
<w:t>that</w:t>
<w:t xml:space="preserve"> </w:t>
<w:t>is</w:t>
<w:t xml:space="preserve"> the question.</w:t>

XML Design Advantages



OOXML Poor XML

- Poor names and inconsistent naming conventions for elements and attributes
- Ecma 376 contradicts the goals of XML which are
  - XML documents should be human-legible and reasonably clear
  - Terseness in XML markup is of minimal importance
- Instead, Ecma 376 often uses unclear names and inconsistent naming conventions
  - These include unnecessary vowel removals, name truncations, and unusual abbreviations, as described in next slide

Differences in Tags and Tag Naming

- The OOXML has shorter tag names, which save file space and facilitates an increase in the speed used of “parsing” the data to convert it to the internal structures the application needs, but increases the number of tags needed in that format.
- The ODF naming is longer and more wordy as it follows the XML convention for naming tags, to ease interoperability when implementing the standard, while file space and slower parsing are offset by the fact there are fewer tags required in this format.

OOXML Reinvents the Wheel

- More than 80% of OOXML's huge documentation (over 7K pages) is used to “reinvent the wheel”:
  - Describe proprietary Microsoft formats adopted to replace available open standards (i.e, VML over SVG)
  - Describe OOXML's extremely "convoluted" XML Schema, which is not following any XML convention (i.e., text for “text”, strong for “bold”, etc.)
  - Describe many proprietary elements of legacy MS Office formats, which are not part of the ISO standard

“Naive” Deductions

- All LibreOffice developers are genius
- All Microsoft Office developers are just i****s

“Real” Deductions

- Microsoft Office XML files are artificially filled with unnecessary content to reduce the chances that software other than Microsoft Office can open them correctly
- Microsoft has a clear commercial interest in opposing interoperability based on standard and open formats, to protect a market that is still worth more than $25 billion
- So, documents created with Microsoft Office are standard on paper, but in reality they are built to fool users (and convince them that interoperability cannot exist)

Simplicity vs Hidden Complexity

- **ODT / LibreOffice**
  - Reduced, very low or non existing complexity
  - XML files are human readable (as they should be)
- **OOXML / Microsoft Office**
  - Highest possible complexity vs technology
  - XML files are not human readable (contrary to what the XML standard language mandates)

Complexity as a Strategy

- Complexity is the deliberate distribution of ambiguous, confusing or misleading information, to interfere with digital sovereignty and data ownership
- OOXML deliberate complexity and false statements about the standard status have the objective of making content sharing difficult for end users and at the same time disqualify the idea of document standards, as inefficient and cumbersome

Future of Surveillance Capitalism

- It is a market strategy, not a technology
  - Its continuation/expansion is not inevitable
- Its business model is based on misuse of our personal data, in ways concealed from us
- Regulation can change the business model
  - It has to prohibit key objectionable practices
  - It has to be ‘dissuasively’ enforced, and global
- The EU’s GDPR is starting to lead the way

What is at stake here
is the human expectation of sovereignty
over one’s own life and authorship
of one’s own experience

Shoshana Zuboff, The Age of Surveillance Capitalism

Sources of Information

- K. Blind, M. Böhm, P. Grzegorzewska, A. Katz, S. Muto, S. Pätsch & T. Schubert, **The Impact of OS Software and Hardware on Technological Independence, Competitiveness and Innovation in the EU Economy** (final study report), Brussels, May 2021
- M. Bank, F. Duffy, V. Leyendecker & M. Silva, **The Lobby Network: Big Tech‘S Web of Influence in the EU**, Corporate Europe Observatory and LobbyControl e.V., Brussels and Cologne, August 2021
- S. Zuboff, **The age of Surveillance Capitalism: the Fight for a Human Future at the new Frontier of Power**, Public Affairs, New York, 2018

Thanks

Italo Vignoli
The Document Foundation
italo@libreoffice.org
italo@vignoli.org