

Open data for public policy: legal and political challenges

Legal and Licensing Workshop 2022 (LLW-2022)

Online event from the Free Software Foundation Europe (FSFE) Legal Network (LN)

Wednesday 18 May 2022

18:15–18:35 CEST

Robbie Morrison

robbie.morrison@posteo.de

Schillerstraße 85, 10627 Berlin, Germany

Chatham House Rule (CHR) waived

Copyright (c) 2022 Robbie Morrison

This work is licensed under a Creative Commons Attribution 4.0 International (CC-BY-4.0) License

Release 01 ▪ 18 May 2022

Git information: 5b1ed04 ▪ 2022-05-18 14:30:46 +0200 ▪ ~/synk/fsfe/online-2022/fsfe-llw-2022-beamer

Generated file: 2022-morrison-fsfe-llw-open-data-slidedeck.01.pdf

Disclaimer

The author has no legal training and the material presented should be interpreted within that caveat.

Abstract

Suitably licensed open data together with relevant open standards can contribute to better public policy for the much needed transition to carbon-neutral energy systems — through open science, improved public engagement, less opaque analysis, and potentially greater public trust.

This talk covers the legal context for the data needed — for instance, system and cost information, reserves and consumption data, econometric data, market clearance information — much of it privately held and subject to statutory reporting, the role of public licensing for both the primary data and the associated metadata, and issues related to license interoperability and data siloing.

The talk looks at the legal and political challenges of achieving usable open data for the given transition — including legally encumbered statutory reporting, non-open data standards and semantics, and frequent confusion about the scope of data copyright and database rights.

The proposed Data Act, currently before the European Parliament, seeks to introduce a novel Data Producers Right that would automatically attach to machine-generated data. This new right is discussed and the risks posed to the usability and reusability of information of public interest.

The talk is limited to non-personal data that may be made public legitimately, thereby avoiding matters of personal and commercial privacy. The context is limited to Europe where the legal issues are among the more challenging in any case.



Key themes

- focus is the **European Union**
- novel **data producers right** in the proposed Data Act
- **definition of “re-use”** in the Open Data Directive 2019/1024
- legal status of **information under statutory reporting**
- special needs of **information of public interest** — with legal uncertainty being especially corrosive in this context
- **AI processing** is intentionally omitted in the interests of simplicity

Energy transition as an example of public interest context

- climate protection
 - net-zero by 2050
- energy transition
 - 300 PJ/annum renewables potential
 - reduced demand on current levels required
 - negative emissions technologies required
- none of the current public policy analysis touches the public sphere
 - at least not in the United Kingdom and Germany
 - conducted entirely between ministries and research institutes
 - typically with closed source models, in-house data, many undocumented assumptions
 - analysis not reproducible, results are invariably sensitive to inputs
- many of the datasets needed are potentially legally (and technically) encumbered
 - European emissions trading timeseries, for instance

European Union digital single market

- European Union **digital single market**
 - flagship initiative and central to the European digital economy
 - complex and wide-ranging
 - provides the political context for this presentation
 - my interpretation is that the digital single market is underpinned by the proposition that information sanctioned as private property and subject to well-regulated trading and common standards will incentivize socially optimal outcomes
 - no carve-out for **information of public interest**
- in comparison
 - United States has a strong doctrine of public interest information in the public domain
 - United Kingdom is opting for concepts like “open data triage” and “presumed open”
- my recent posting expands on the material here and provides references
 - <https://forum.openmod.org/t/3214>

Legal encumbrance of information of public interest

- **copyright held in collections of data**

- some threshold of originality required (post-Brexit, UK law includes intellectual effort)
- legal uncertainty prevails

- **Database Directive 96/9/EC**

- enacted 1996 and subsequently transcribed into national law
- everything an analyst touches classes as a 96/9/EC database
- legal uncertainty prevails

- **Open Data Directive 2019/1024**

- covers public sector information
- defines “re-use” as “use” — with both being copyright verbs (more later)

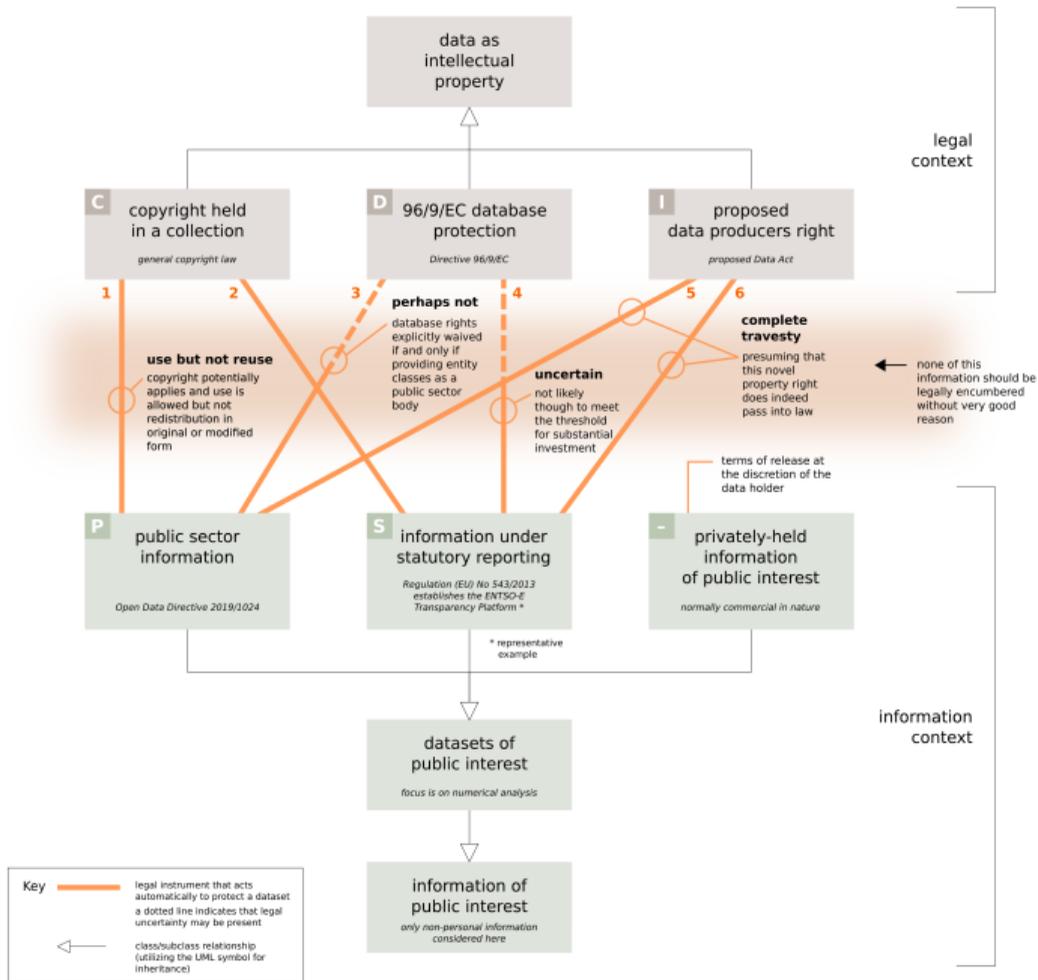
- novel **data producers right** (DPR)

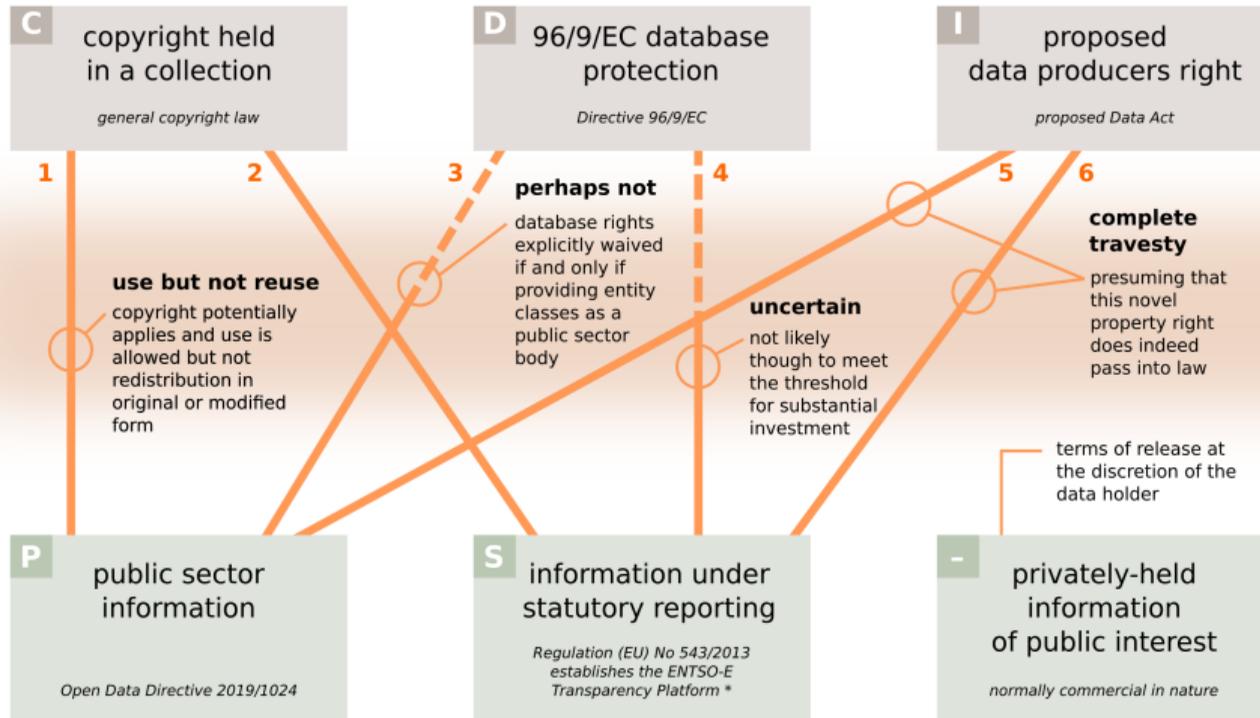
- proposed Data Act currently before the European Parliament
- raw machine-generated data automatically protected
- almost everything an analyst will want to touch will be subject to this right

Data producers right (DPR)

Technical details are provided in European Commission (2017:9–10) — while noting that these details may have been subsequently modified during legislative development:

- “Machine-generated data is created without the direct intervention of a human by computer processes, applications or services, or by sensors processing information received from equipment, software or machinery, whether virtual or real.”
- “Machine-generated data can be personal or non-personal in nature. Where machine-generated data allows the identification of a natural person, it qualifies as personal data with the consequence that all the rules on personal data apply until such data has been fully anonymised (e.g. location data of mobile applications).”
- “Raw machine-generated data are not protected by existing intellectual property rights since they are deemed not to be the result of an intellectual effort and/or have any degree of originality.”





← none of this information should be legally encumbered without very good reason

Technical barriers to information under statutory reporting

- web-based statutory reporting by the European Energy Exchange (EEX) can be **viewed but not extracted**
 - enforced by server-side CSS or JavaScript
 - European regulator (ACER) says this behavior is compliant

Tarkowski and Vogelezang (2021) on the DPR

- Tarkowski and Vogelezang (2021:7–8) highlight **dangers posed to science**:
 - *The new right would strongly affect freedom of expression and information as well as freedom of scientific research and services, given that it would greatly reduce overall information availability. In this light, the European legislator would have to prove that a new property right would be socially and economically justifiable for information access by citizens and researchers.*
- And that generating data is increasingly a **joint undertaking** rather than an individual activity (p8):
 - *Data is increasingly seen as [socially] relational and co-generated. Salomé Viljoen (2021) proposes that the relational character of data means that for any exchange of data there are collective—even population-level—interests that cannot be reduced to individual interests. . . . These characteristics of data imply competing interests among various actors in the data economy. Exclusive property rights can therefore easily be questioned by other parties, asking for recognition of their rights in data.*

Creative Commons CC-BY-5.0 ?

- Creative Commons CC-BY-4.0 license will doubtless need to be modified to reflect any enacted **data producers right**
- version number provocatively bumped here to **5.0** to signal a major revision

Copyright verbs regarding the Open Data Directive

- §2.11 of the Open Data Directive 2019/1024 defines data “re-use” as simply “use”
- one can argue that “**use**” is a well established concept under copyright law and that this concept does not clearly permit the free **redistribution of data and datasets in original or modified form**
- or alternatively, why would open licenses need to be so explicit and specific when a simple grant for “use” would have been sufficient? (see next slide)

Copyright verbs from established open licenses

- the notion of **copyright verbs** comes from the wider approach of grammatical and textual analysis
- some copyright verbs from established open licenses follow — although important anding relationships are omitted with this first cut

phrase	generally permitted by all rights reserved
copy	× but some provisions for personal copies, scientific usage, and so forth
distribute	×
extract	× but some provisions for limited passages
modify	×
prepare derivative works	×
publish	×
redistribute	×
reuse	×
share	×
share adapted material	×

Knowledge commons

- projects within the **energy domain** now feeding towards a knowledge commons
 - assembling data is increasingly a collective undertaking
 - significant support from science funders
 - https://en.wikipedia.org/wiki/open_energy_system_databases
- inputs must be genuinely open and legally signaled — my recommendations
 - primary data: CC-BY-4.0, CC0-1.0, or something inbound compatible
 - associated metadata: CC0-1.0
 - the attribution overhead is not an issue — provenance needs tracking in any case
- key requirements
 - community curation
 - open and freely usable technical standards
 - consensus on domain-specific semantics — best encapsulated using ontologies

Some issues not traversed in this presentation

- the **fuzzy boundary** between data and code
 - where do scripts that interrogate remote databases sit legally?
 - analysis on computer-aided manufacturing files might be applicable
- transmission of intellectual property rights along **data processing pipelines**
 - where do derived works end and new works begin?
- treatment of **joint ownership**
 - how best to deal with multiple authorship?
- provisions for **science** versus public interest
 - legal exceptions for education and science not sufficiently broad for public debate
- data processing practices and **attribution-based licensing**
 - the alleged combinatorial explosion versus good data management

Closing remarks

1. information of public interest needs to be prioritized over the marketization of public sector information and information under statutory reporting
2. freely usable and re-usable data and datasets of public interest — which in turn facilitate data catalogs, data portals, and ultimately various domain-specific knowledge commons — are central to solving humanity's problems
3. the current and projected legal status of data of public interest is problematic and legal uncertainty is corrosive
4. the only real solution is for analysts to push for Creative Commons CC-BY licensing on primary data of public interest
5. what can be ported from free software experiences — or does everything regarding genuinely open data need to be discovered afresh in this new context?

References

- European Commission (10 January 2017). *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: “Building a European data economy” — COM (2017) 9 final*. Brussels, Belgium: European Commission.
- Morrison, Robbie (May 2022). *Proposed data producers right for Europe*. Frankfurt, Germany: Open Energy Modelling Initiative. Posting. <https://forum.openmod.org/t/3214>
- Tarkowski, Alek and Francesco Vogelezang (10 December 2021). *The argument against property rights in data — Policy brief #1*. Europe: Open Future.

