

## EMP-E 2021: Re-Energising Sustainable Transitions in Europe

Energy System Modelling, Methods & Results to  
support the European Green Deal

26th to 28th October • online

# What energy system modellers should know about open data and software licences

Robbie Morrison • open energy modeling community  
26 October 2021 • 16:00–17:30 CEST • Skills workshop

Release: 01 • Status: Presentation version • Date: 26 October 2021



Copyright (c) 2021 Robbie Morrison <[robbie.morrison@posteo.de](mailto:robbie.morrison@posteo.de)>  
This work is licensed under a Creative Commons Attribution 4.0 International (CC-BY-4.0) License

hosted in cooperation  
with the European Commission



# Opening remark

My generation has left our planet in a terrible state and very few of our underpinning technologies, systems, and policies are remotely fit for purpose.

The IPCC First Assessment Report in 1990 was entirely clear on the need for deep and rapid decarbonization.

Let's hope we can now understand, articulate, and transition sufficiently quickly to a world that will provide those who follow with a viable future.

The first Fridays for Future school strike held 14 December 2018, Berlin, Germany  
*Photograph:* Robbie Morrison



# My background

- **system modeling:** *deeco*
  - started in 1995 (consortium models like MARKAL were well established then)
  - ported *deeco* high-resolution framework from Hewlett-Packard to Intel hardware
  - added GPL-2.0 to *deeco* codebase in 2003 and attempted to build a user/developer community
  - maintainer for 7 years, abandoned because proprietary programming library became stranded
- **system modeling:** *xeona*
  - later wrote *xeona* framework, open-licensed but not published, 58k source-lines of C++
  - some experience with another C++ model
  - contributed to GLPK solver and MathProg algebraic modeling language project for 4 years
- **Open Energy Modelling Initiative** (openmod)
  - learned of openmod some 18 months after its formation and became active
  - am lead admin on the openmod forum after Tom Brown stepped aside at end of 2020
- **recent activities** of relevance:
  - contributed to the Icebreaker One Open Energy project via phase 2 advisory group 2 on policy, legal, and regulatory
  - participate in the Linux Foundation Energy (LFE) data architecture special interest group
  - am assisting the Open Energy Ontology (OEO) project as part of their steering committee
  - contributed to the recent EERAdata process on energy sector metadata
  - participate in the Free Software Foundation Europe (FSFE) Legal Network (LN) community since October 2017
  - coordinated four submissions on European Union data law reform, gave oral evidence in Brussels on one occasion

# Abstract

The workshop will review the default legal status of source code, datasets, and databases. Then cover public licensing, the attributes of selected open licenses, and how these facilitate transparency, reproducibility, inclusion, and collaboration. And finally how public interest analysis might be better served under these conditions.

The following objectives are sought:

- an understanding of the legal contexts for closed and open modelling
- working definitions for open source code and open data
- some understanding of public licensing, the attributes of various license types, license interoperabilities, and matters surrounding choice of license
- practicalities related to identifying licenses, applying license notices, and assuring conformance
- an improved understanding of the benefits of open modelling and those circumstance where open modelling is indicated or otherwise
- some insight into the advantages of using open models for public policy development and issues surrounding institutional uptake and reluctance

The workshop should be accessible to anyone. Participants are not expected to possess any special knowledge of public licensing or the underlying law. This workshop is not skills-driven nor focused on advocacy, so the workshop objectives listed earlier (point 2) apply here essentially. The session will mostly comprise presented material and group discussion.

Short quizzes will provide engagement and feedback and perhaps even light relief.

All are welcome to participate, no prerequisite skills or software necessary

# Orientation

- focus here is the legal aspects of open code and open data
- mostly in the context of **European Union law**
- "**software**" and "**code**" mean open source software for public interest analysis
- "**data**" means non-personal open data which can be or has been made public legitimately
- proprietary licenses will not be covered

# Open source software for public interest analysis

- discussion limited to software for **public interest analysis**
- earlier excuses: code too messy, no documentation, no overhead for servicing inquiries
- dominant software development model — estimates of 90% for new projects
- software as a service — cloud delivery drove much of the commercial adoption

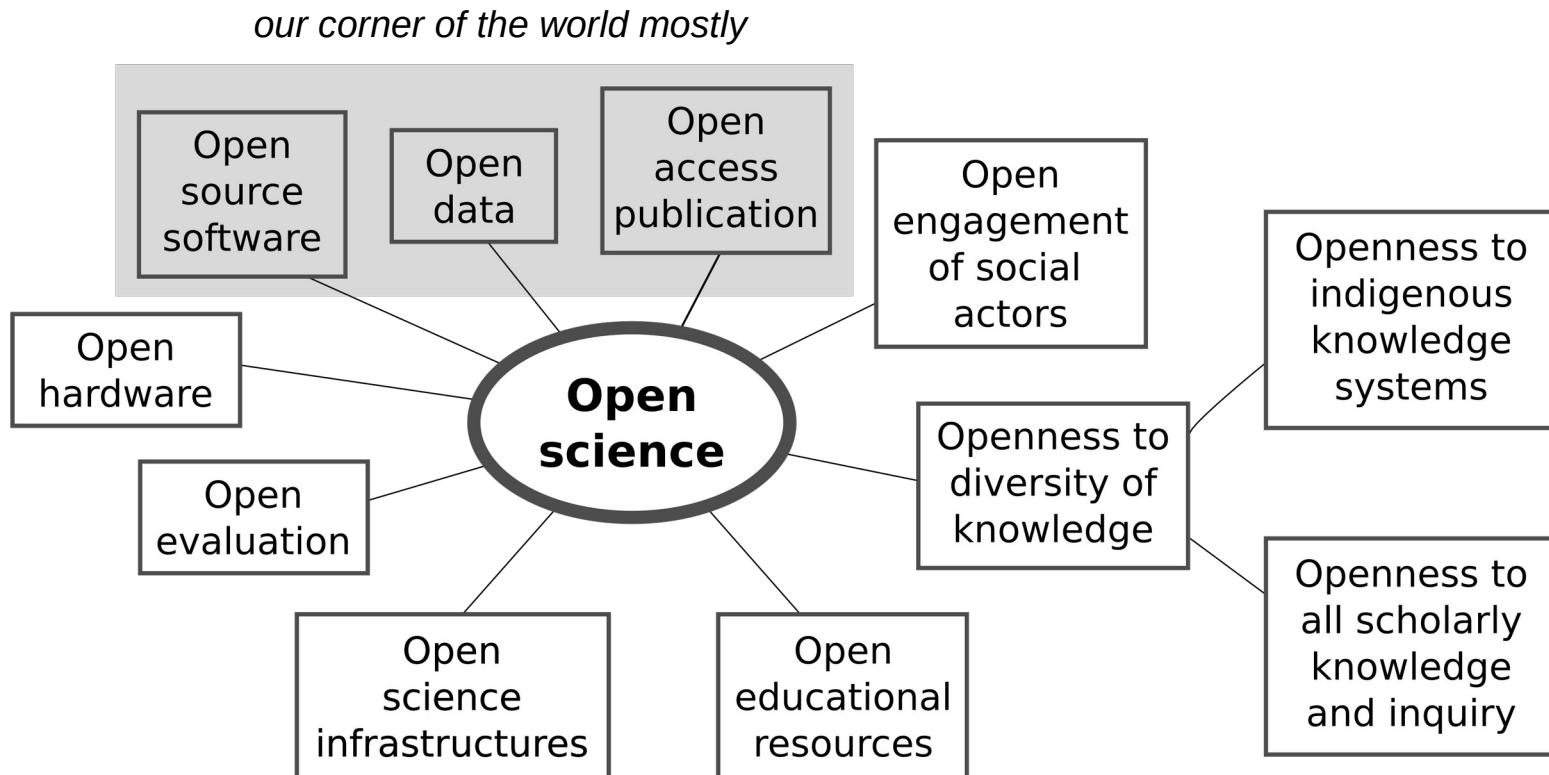
*Further reading:* Raymond, Eric S (2001). *The cathedral and the bazaar: musings on Linux and open source by an accidental revolutionary.* Sebastopol, California, USA: O'Reilly Media. ISBN 978-0-596-00108-7.  
Extended version of his original essay published in 1998.

# Non-personal open data

- discussion limited to **non-personal data** that can be or has been made public legitimately
- current data "tsunami" enabled by hardware developments and internet technologies
- interest for openness driven equally by transparency and opportunity

# Open science

Providing some context

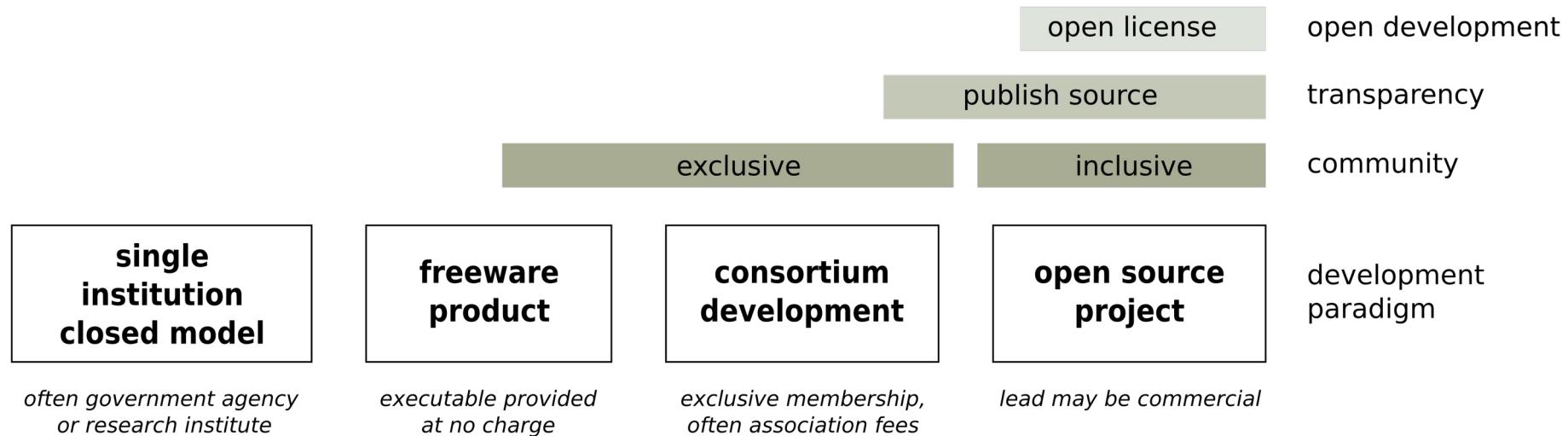


Adapted from: UNESCO draft open science recommendation presented at the Open Science Conference 2021 on 17 February 2021

# Development paradigms

## Energy system models providing public interest analysis

- not all projects map directly to this scheme, especially historical projects



# Privately-held information of public interest

- some disciplines can generate the primary data they need
- energy system analysts are **highly-reliant** on what the European Commission describes as:  
**"privately-held information [of] public interest"**
- two examples of problematic information under **statutory reporting** with neither open licensed (more later):
  - ENTSO-E Transparency Platform database
  - mandated public disclosures by the EEX

*Source of quote:* Public sector information (PSI) law reform consultation documents

ENTSO-E = European Network of Transmission System Operators for Electricity

EEX = European Energy Exchange

ACER = European Union Agency for the Cooperation of Energy Regulators

# Open Energy Modelling Initiative

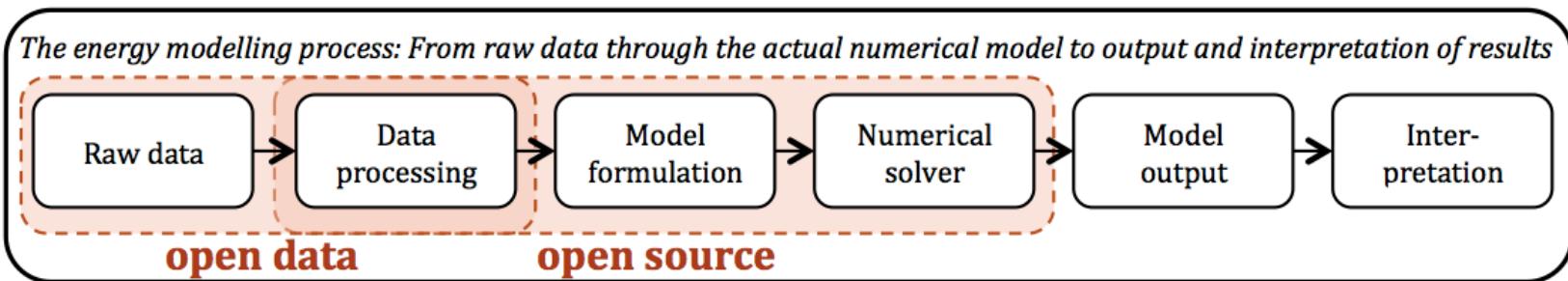
Also known as "openmod"

- began **September 2014** in Berlin, Germany
- 2 day meeting in German language with 28 attendees
- produced a manifesto
- Berlin open source lawyer Till Jaeger spoke at that first meeting
- the openmod does not endorse projects or take specific positions
- 190 attendees at last physical meeting before Covid-19
- 800 accounts on discussion forum, 900 participants on mailing list

*Further reading:* Jaeger, Till and Axel Metzger (6 February 2020). Open Source Software: Rechtliche Rahmenbedingungen der Freien Software (5. Ausgabe) [Open source software: legal framework for free software (5th ed)] (in German). Munich, Germany: CH Beck. ISBN 978-3-406-73497-7.

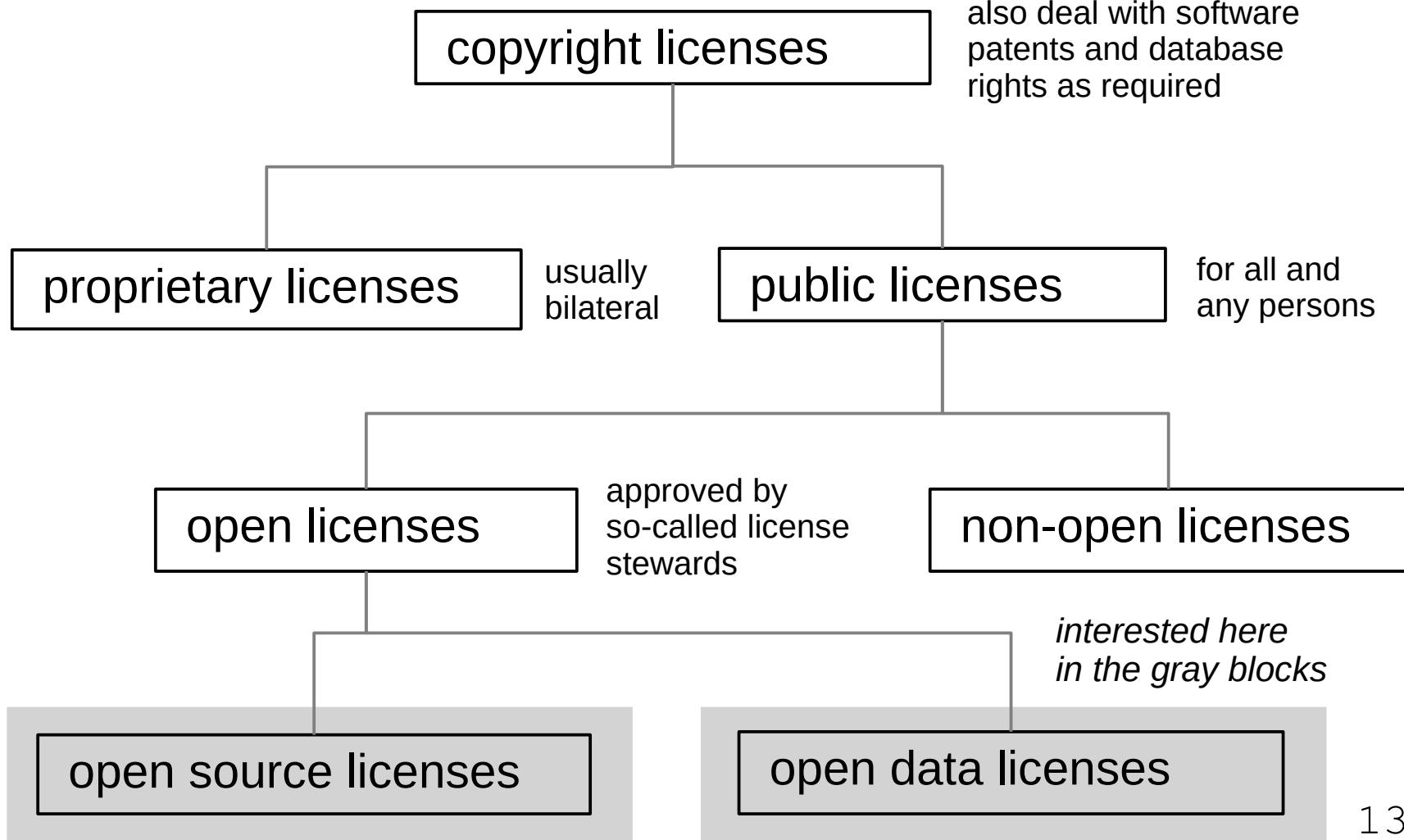
# openmod manifesto

Dated 2014



- **community** : sharing data, code, ideas
- **quality** : reproducibility, transparency, accuracy
- **efficiency** : model interoperability, raw data and metadata standards
- **insights** : modeling, collaboration
- **information** : knowledge sharing, choice of license

# License typology



# A need for hybrid licenses

The code/data split just shown is indeed fuzzy

- particularly for **AI applications**
- but also for any energy system models in which the codebase embeds hardcoded data
- hybrid licensing is currently an academic research question

# Contributor agreement (CA)

Apply to software projects but not widely used

- also known as contributor license agreements (CLA)
- contributor agreements are complementary to the prevailing software license and provide the project with greater legal certainty regarding individual contributions. Quoting from the Contributor Agreements project:

"To put it in a nutshell, contributor agreements avoid as far as possible any future legal issues regarding the individual contributions, such as disputes over origin or ownership of respective rights."

- CLAs are used by GNU projects
- but not deployed by any energy system modeling projects to my knowledge

# Developer certificate of origin (DCO)

Affirm corporate developers can indeed contribute

- developer certificates of origin (DCO) were first introduced by the Linux Foundation to affirm that a contributor is legally able to submit code contributions

# SPDX identifiers

Software Package Data Exchange project

- scheme overseen by the Linux Foundation
- standard identifiers for public licenses, some examples:
  - software, permissive: MIT
  - software copyleft : AGPL-3.0-or-later
  - data and content, attributive: CC-BY-4.0
  - data and content, public domain or close: CC0-1.0

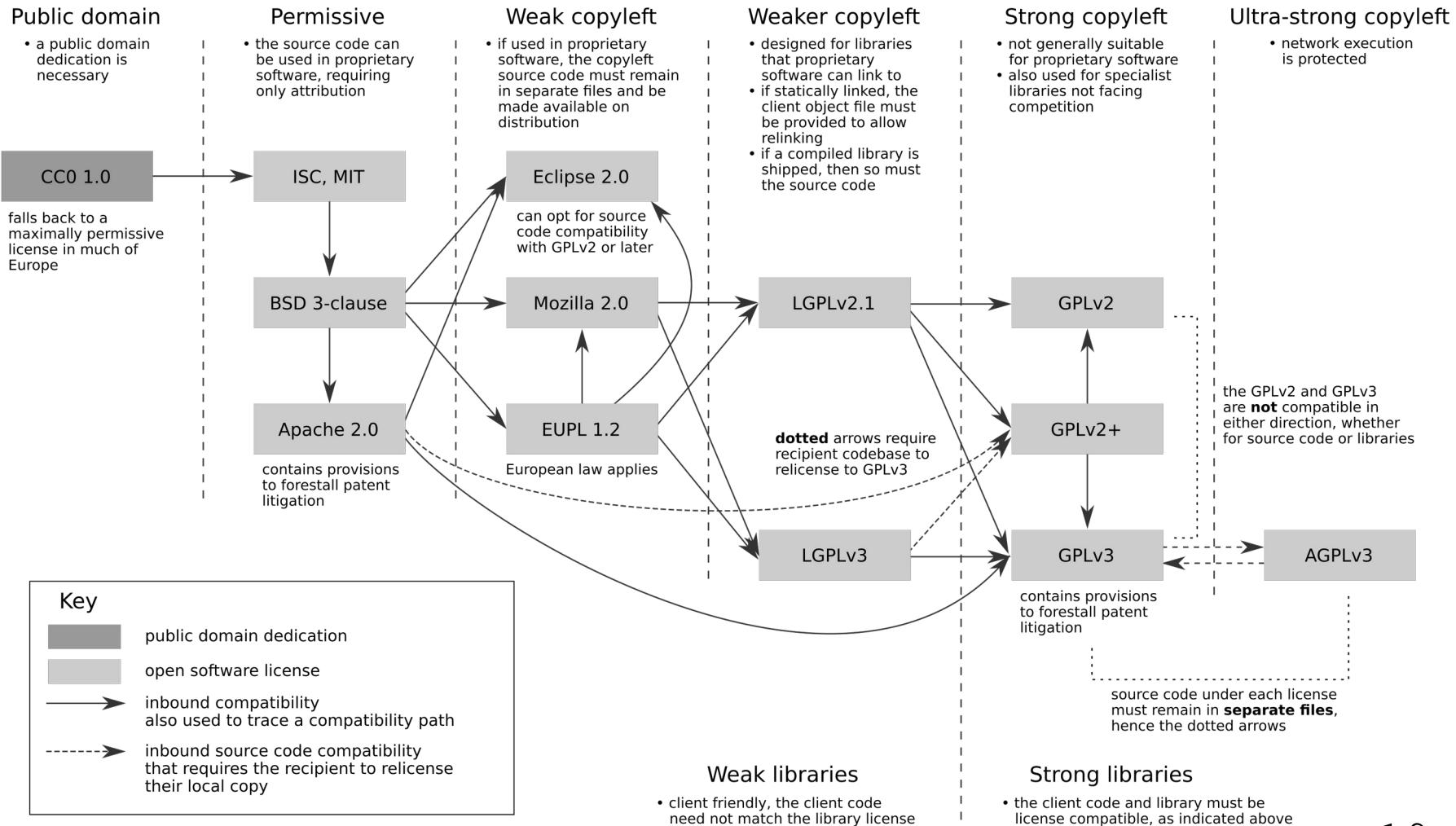
## QUICK QUIZ

How many entries in the full version of the SPDX License List of public licenses?

- 200
- 800
- 17 000

*Refer:* <https://github.com/spdx/license-list-XML>

# Software license compatibilities



# Which software license?

## My suggestions

- in the absence of other considerations ..
- aim **high** to maintain codebase integrity (similar to Linux):  
AGPL-3.0-or-later
- or **low** to facilitate third-party uptake (say for link libraries)  
MIT
- if high, can normally relicense low as circumstances dictate

*More information:* Morrison, Robbie (April 2018).

"Energy system modeling: public transparency, scientific reproducibility, and open development".  
*Energy Strategy Reviews.* **20:** 49–63. ISSN 2211-467X. doi:[10.1016/j.esr.2017.12.010](https://doi.org/10.1016/j.esr.2017.12.010). Open access.

# Core legal concepts / 1

## Intellectual property

- any sufficiently creative and substantial work is automatically protected by **copyright**
  - these thresholds vary markedly between jurisdictions
- other intellectual property rights (IPR) may apply to code:
  - individual **software patents** following application and examination (United States)
- other intellectual property rights may apply to databases under their legal definition:
  - 96/9/EC **database protection** (European Union)
- doctrines like **fair use** (United States) and **fair dealing** (countries with British legal traditions) may apply — but less so in continental Europe

# Core legal concepts / 2

## Public licensing

- **public licenses** provide additional copyright permissions to any and all persons provided those persons also abide by the terms and conditions of the license in question
- failure to do so typically results in the license being **summarily revoked**
- conversely, if no IPRs are indeed present then the license notice can be **legitimately ignored**
- as indicated, jurisdiction-specific legal doctrines like fair use may also apply
- most public license texts are based on United States legal traditions
- unlike private property, there is **no statutory support** in this realm for what might be termed "public property"
- very many unanswered legal questions around open licensing

# Core legal concepts / 3

## Jurisdictional differences

<b>Concept</b>	<b>United States</b>	<b>European Union</b>	<b>United Kingdom</b>
governing law for public licenses	intellectual property rights (IPR) law	either IPR law or contract law	
public domain	✓ 1	✗	
moral rights	✗	✓	
threshold for copyright	sufficient creativity 2	sufficient creativity	sweat-of-the-brow
database protection	✗	✓	✓

Notes: 1: work by Federal employees is public domain by default • 2: US Copyright Office states most datasets will fail test

# Core legal concepts / 4

## European legislation

- **software directive** 2009/24/EC (officially the directive on the legal protection of computer programs) — and noting that software patents do not exist in the European Union
- **copyright directive** (EU) 2019/790
- **database directive** 96/9/EC
- **open data directive** (EU) 2019/1024 — covering public sector information (PSI)

## Currently under consultation

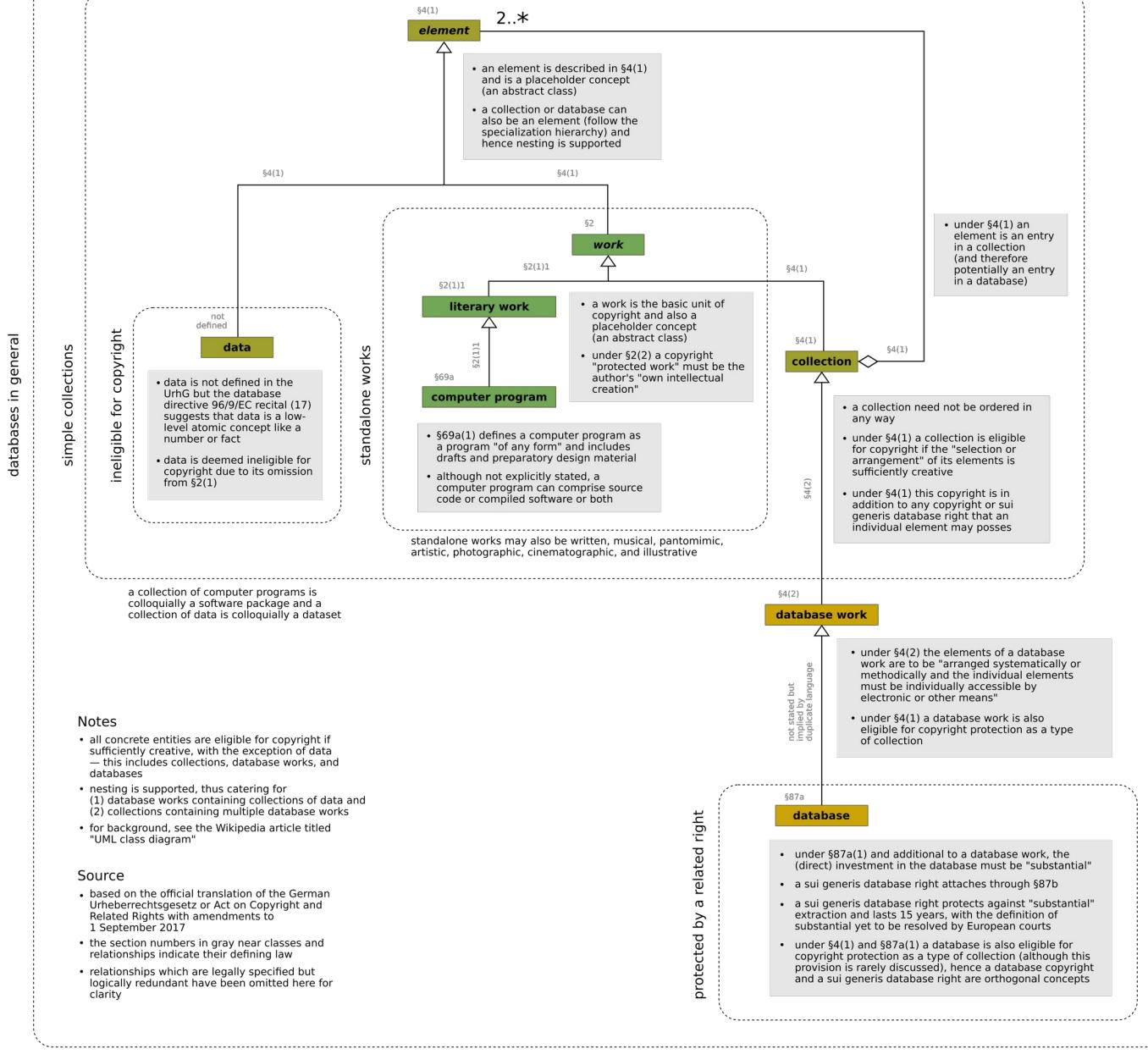
- proposed **data act** covering public interest information — no draft legislation as yet
- proposed **data governance act**

# Core legal concepts / 5

## Choice of law issues

- choice of law provisions indicate which **national law** should govern license interpretation and related activities
- the **location of any allied server** may determine the choice of law in the first instance
- public licenses may elect a choice of law, but the trend is for open licenses to be **international** — in other words, the governing law is intentionally not specified in the license text
- that said, the majority of open licenses have been drafted using United States legal traditions

# German Copyright Act (UrhG) definitions



## Key

- contains ("has a") relationship with each entity existing independently
- specialization ("is a") relationship
- two-to-many multiplicity



Copyright (c) 2021 Robbie Morrison  
This work is licensed under a Creative Commons Attribution 4.0 International (CC-BY-4.0) License

Urheberrechtsgesetz (UrhG)  
Software and data related definitions  
depicted using UML class modeling

release 10

# Entr'Ouvert v Orange

GPL infringement case currently being litigated in France under contract law

Two parties are **Entr'Ouvert**, a small software company developing a GPL-licensed SAML authentication library called Lasso since 2003, and **Orange Business Services** (previously part of France Telecom) who made use of that library when developing a nationwide authentication service for France since 2005.

- GPL is the **GNU Public License** for software (with various versions and variants in use)
- being prosecuted as breach of contract under **contract law** and not as the civil wrong (or tort) of **copyright infringement**

*Further reading (in the absence of academic analysis): The HFT Guy (30 August 2021). French Appeal Court affirms decision that copyright claims on GPL are invalid, must be enforced via contractual dispute. Read this blogged analysis with caution.*

# Many unresolved issues

Limited case law and meager statutory support means few robust conclusions

For instance:

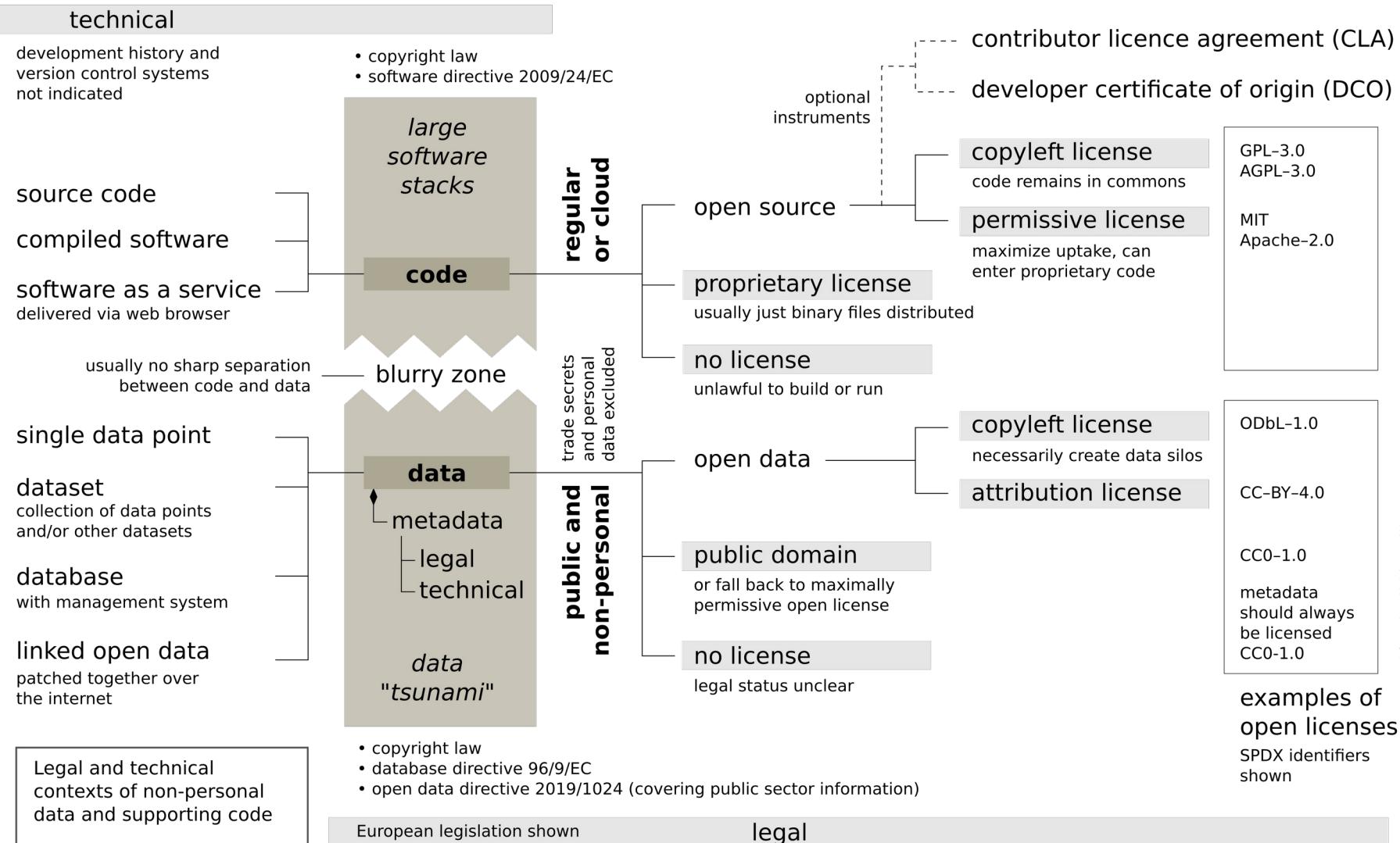
- the Google v Oracle America API case in which the US Supreme Court said fair use applied but expressly avoided the question of whether APIs are indeed protected under copyright
- does each contributor hold the copyright to their own contributions (often assumed so) or does some form of joint authorship doctrine apply (see next slide)
- the legal lines concerning various thresholds are vague in this context, namely:
  - the degree of creativity required for copyright
  - the line between a derived work and new work
  - what constitutes fair use and fair dealing
  - what constitutes substantial investment and substantial extraction regarding a legal database
  - what is the legal status of public sector information under European law

# Joint authorship

Legal treatment unclear for both code and data

- commonly believed that each contributor retains the copyright to their individually identifiable contributions
- for alternative view based on United States law: Chestek, Pamela S (2017). [\*\*"A theory of joint authorship for free and open source software projects"\*\*](#). *Colorado Technology Law Journal*. **16**: 285–326. Open access.
- the issue is material in relation to license compliance

# Code/data landscape



# Open licensing of source code

Please refer to recent Free Software Foundation Europe (FSFE) good practice recommendations — in summary:

- use SPDX identifiers
- add the **full license text** using a `LICENSE.txt` file in the root directory of your public repository
- add **license notices** to all header files (compiled languages) and source files (all languages) while noting that these notices need not extend the copyright year range on each change of year
- the legal status of **version control information** (say from git) in this context remains uncertain

# Open licensing of data

An emerging challenge with very many unresolved issues

- use SPDX identifiers
- add license information via legal metadata where possible
- note the **Dublin Core** provides the following fields as Resource Description Framework (RDF) properties:
  - **creator**: an entity primarily responsible for making the resource
  - **rights**: information about rights held in and over the resource
  - **contributor**: an entity responsible for making contributions to the resource
  - **license**
  - **rightsHolder**
- related support within the Open Knowledge Foundation (OKF) Frictionless Data concept

# A touchstone definition for open data

Recital 16 from the 2019 open data directive:

"Open data as a concept is generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose."

Source: European Commission (26 June 2019).

"Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information — PE/28/2019/REV/1"

Official Journal of the European Union. L 172: 56–83. The directive entered into force on 16 July 2019. **Recital 16** (page 58) quoted above.

See also: Morrison, Robbie and contributors (22 February 2019). [Definitions for open](#). openmod forum. Germany.

# Database Directive 96/9/EC

Covers the European Economic Area (EEA)

- the legal definition of a "database" is very broad:
  - includes printed maps
- a database is protected where both:
  - the **direct investment** in the database system is substantial
  - the **extraction** is substantial
  - the substantiality principle derives from copyright law
- the directive was intended to support a database industry within Europe
  - instead material gets harvested and used to stock servers in the United States
  - the provision creates legal uncertainty for risk-averse researchers in the absence of suitable open licenses

# Public sector information

## In relation to 96/9/EC database protection

- public sector information (PSI) is unlikely to attract 96/9/EC database protection based on arguments advanced by Giannopoulou (2018)
- but again this result is nonetheless uncertain

*Reference:* Giannopoulou, Alexandra (2018). *Chapter 6: Understanding open data regulation: an analysis of the licensing landscape.* In Bastiaan van Loenen, Glenn Vancauwenberghe, and Joep Crompvoets (editors) (2018). *Open data exposed.* The Hague, the Netherlands: TMC Asser Press. Pages 101–125. ISBN 978-94-6265-261-3. doi:10.1007/978-94-6265-261-3\_6. Downloadable PDF available.

# US Copyright Office (2017)

For comparison with the United States

"The notion of technical databases being creative is largely mutually exclusive. Orthodox database are highly structured, but they are not much selected and arranged. Nonorthodox databases, while not highly structured, are similarly even less likely to be selected and arranged."

Hence in the United States:

- copyright in a compilation (or collection) of data is highly unlikely
- database protection never entered US law (despite three attempts)

Source: §727 from US Copyright Office (November 2017).

The Compendium of US Copyright Office Practices — Third edition: Chapter 700. US Government.

# Open data licenses

## My personal picks

### Creative Commons CC-BY-4.0



- introduced 25 November 2013
- first data-capable license (that deals with the EU Database Directive 96/9/EC)
- requires **attribution and attribution tracking**
- material may be modified or mixed and licensed under more restrictive terms but the attribution requirement must remain

### Creative Commons CC0-1.0



- public domain dedication
- falls back to maximally permissive license in civil law jurisdictions like those in Europe
- no legal obligations for users
- metadata should always be licensed CC0-1.0

in most cases, open data licenses do not provide **permission** but rather offer **certainty**

# FAIR data principles

Please treat this slide as a comment and not a criticism

**F** = findable, **A** = accessible, **I** = interoperable, **R** = reusable

specifically: **R1.1**: (meta)data are released with a clear and accessible data usage license

the **R** is thus not necessarily global, hence consortium data and trusted brokerage schemes are also intended

note therefore the FAIR/O extension where **O** = open licensed

See also: Wierling, August, Valeria Jana Schwanitz, Sebnem Altinci, Maria Balazińska, Michael J Barber, Mehmet Efe Biresselioglu, Christopher Burger-Scheidlin, Massimo Celino, Muhittin Hakan Demir, Richard Dennis, Nicolas Dintzner, Adel el Gammal, Carlos M Fernández-Peruchena, Winston Gilcrease, Paweł Gladysz, Carsten Hoyer-Klick, Kevin Josho, Mariusz Kruczak, David Lacroix, Małgorzata Markowska, Rafael Mayo-García, Robbie Morrison, Manfred Paier, Giuseppe Peronato, and Mahendranath Ramakrishnan (15 October 2021). "Advancing FAIR metadata standards for low carbon energy research". *Energies*. 4 (20): 6692. ISSN 1996-1073. doi:10.3390/en14206692. Open access.

# Legal metadata for data

## Another raft of challenges

- as indicated, the **Dublin Core** for metadata includes fields for the **creator** and for the **rights** held, the DCMI extension includes a **provenance** field, and all are defined as Resource Description Framework (RDF) properties
- that may work for relatively static works like manuscripts but not for datasets that are widely combined and continually and subtly altered
- some data-specific **challenges** therefore:
  - data versioning
  - data license compliance and attribution tracking
  - the recording of provenance more generally
- these issues have been traversed for software but the solutions developed for code are unlikely to transfer readily to data

# Proposed EU Data Act

Speculative comments on my part

- current law reform from the European Commission
- I lead-authored two public submissions from the open energy modeling community (also on Zenodo)

Key aspects, as I read them:

- need to facilitate **business-to-government** (B2G) information transfers for public interest reasons
- the 96/9/EC database directive is under scrutiny
- little indication thus far of if and how B2G transfers will be made more generally available

# Statutory reporting issues

Two substantial cases of concern

**1** datasets on **ENTSO-E Transparency Platform** are not open licensed:

- long-standing dialog between openmod members and ENTSO-E
- the matter remains unresolved however

**2** **mandatory public disclosure** by the **EEX** remains numerically unusable:

- web-based reporting is limited to human viewing and highlighting and copying of numerical text explicitly disabled (using CSS and/or JavaScript functionality)
  - current practice is compliant according to ACER
- both services should offer important sources of numerical information to energy system analysts

ENTSO-E = European Network of Transmission System Operators for Electricity

EEX = European Energy Exchange

ACER = European Union Agency for the Cooperation of Energy Regulators

## QUICK QUIZ

Is the Creative Commons CC-BY-4.0 license:

- a copyright license under copyright law
- a contract under contract law
- it depends

# Some legal conclusions / 1

Relative to open energy system analysis

## Code

- no evident issues related to legal interpretation
- choice of license well resolved — in the absence of special considerations:
  - permissive: MIT
  - copyleft: AGPL-3.0-or-later
- practicalities of licensing sufficiently well resolved
- contributor agreements could be considered by projects

# Some legal conclusions / 2

Relative to open energy system analysis

## Data

- choice of license well resolved — in the absence of special considerations:
  - data: CC-BY-4.0
  - metadata: CC0-1.0
- interoperability with other licenses like the OGL-UK-3.0 needs analysis
- tooling to handle legal metadata and information provenance more generally still in their infancy
- substantial community work needed on data semantics, ontologies, and metadata standards
- information under statutory reporting and B2G-to-community data flow best addressed using legislative solutions in my view

# Some readings

Hirth, Lion (18 September 2014). *Workshop zur offenen Energiesystemmodellierung* [Open energy system modeling workshop] (in German), Berlin, Germany: Neon Neue Energieökonomik GmbH. Date from PDF metadata. Workshop program.

Hirth, Lion (1 January 2020). "Open data for electricity modeling: legal aspects". *Energy Strategy Reviews*. **27**: 100433. ISSN [2211-467X](#). doi:[10.1016/j.esr.2019.100433](https://doi.org/10.1016/j.esr.2019.100433). Open access.

Morrison, Robbie (April 2018). "[Energy system modeling: public transparency, scientific reproducibility, and open development](#)". *Energy Strategy Reviews*. **20**: 49–63. ISSN [2211-467X](#). doi:[10.1016/j.esr.2017.12.010](https://doi.org/10.1016/j.esr.2017.12.010). Open access.

Morrison, Robbie (20 November 2019). "[An open energy system modeling community](#)". Generation R blog. Hannover, Germany: Leibniz Research Alliance Open Science. doi:[10.25815/ff3b-d154](https://doi.org/10.25815/ff3b-d154). ISSN [2512-3815](#). Creative Commons CC-BY-4.0 license.