

# Open Research Data

Jakub Míšek

Institute of Law and Technology

Faculty of Law

# Data and Information

- Data v. Information
- Many theories:
  - Claude Shannon: Information as a way to lower entropy
  - Luciano Floridi: Information consists of structured data
  - Michael Keeble Buckland: Information

BUCKLAND, Michael Keeble. Information as a Thing. *Journal of the American Society for Information Science and Technology*. 1991, Vol. 42, No. 5

- Information as knowledge: Knowledge (Intangible, Entity)
- Information as process: Becoming informed (Intangible, Process)
- Information as thing: Data, document (Tangible, Entity)
- Information processing: Data processing (Tangible, Process)

# Data and Information

- Norbert Wiener: Information as a basic premise of life
  - Assumption: Everything decays into entropy
  - Information is a way, how this decay slow or divert
  - The true information = information which has an organising effect
  - Basic notions:
    - Information is an opposite of entropy
    - Living organisms have more than critical amount of information
    - Living organisms (or their parts) react on a change of conditions with production of information
  - Examples: Ants, Justice
- The Law as an Information system
  - The law as a way how to lower entropy in the society

# Law and Information

- Presumption: The centrality of a human
- Information rights - 3 basic rights
  - Obtain information
    - To get the information from outside
  - To have information
    - Have control over information
    - Process information
  - Create, share and communicate information
- Regulation of Information
  - Duties correlating to these rights
  - Regulation of the environment

# Data as a Commodity

- What is “Data” in the light of Law?
  - Is it a thing?
    - Data ownership?

Recapitulation– Database Protection

# Database copyright protection

- Directive 96/9/EC
- Database = collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means

# Database copyright protection

- Databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation
- Protection of the structure of the database
- Rights:
  - Reproduction, translation, adaptation, arrangement and any other alteration, distribution to the public, communication to the public
- Generally analogous to the copyright protection



# Database sui generis protection

- Protection of the investment in the database
  - The content, but not specifically
- When?
  - database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents
- Maker of the database – the one who invests and initiates creation of the DB

# Database sui generis protection

- Rights:
  - **Prevent extraction and re-utilization of the whole or of a substantial part, evaluated qualitatively or quantitatively, of the contents of that database**
- Extraction = permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form
- Re-utilization = making available to the public all or a substantial part of the contents of a database
- No moral / property rights difference

# Database sui generis protection

- Limitations:
  - Lawful user of publicly available database may:
    - Extract & re-utilise insubstantial parts of contents, evaluated qualitatively and/or quantitatively, for any purposes
- Exceptions (non-compulsory)
  - Extract for private purposes from non-electronic databases
  - Education exception
  - Public security
  - Administrative or judicial procedure

Is data protected?



# IP data protection

- Non-direct protection thanks to the database protection
  - Database copyright – Structure
  - Sui generis database right – content / investment
- Data as such – not protected
  - CJEU Ryanair Case, C-30/14



# Data as a Commodity

- What is “Data” in the light of Law?
  - Is it a thing?
    - Data ownership?
- Generally no – Data itself not protected
  - What about “my data”?
    - “Whose data is that?”
    - Other means of protection come into play – Database protection
  - But! Provision as service?
    - CJEU Ryanair Decision



# Public Sector Information (PSI) & Open Data

- Public sector – a vast amount of data
  - Examples:
    - Statistics
    - Geospatial data
    - Meteorology data
    - Legal data
    - Transportation data
    - Data from research
    - ...



# PSI

- Access to data
  - Fundamental political right
  - Principle of publicity
  - Exercise of public authority
- Reuse of data
  - Economic aspect
    - Economic reasoning – we have already paid once for the data
  - Accent in private law
  - Open data - high efficiency

# PSI – International context

- Primarily access
- The Universal Declaration of Human Rights (1948)
  - Part of freedom of speech
  - Art. 19 - Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.
- The European Convention on Human Rights (1950)
- The International Covenant on Civil and Political Rights (1966)

# PSI – International context

- Council of Europe Convention on Access to Official Documents (CETS No.205, Tromsø, 18/06/2009)
  - Access to administrative documents (= all information in any format)
  - Sets basic rules for access
- European Union
  - The Charter of Fundamental Rights of the European Union
    - Art 11
      - Right to freedom of expression.
    - Art 42
      - Any citizen of the Union, and any natural or legal person residing or having its registered office in a Member State, has a right of access to documents of the institutions, bodies, offices and agencies of the Union, whatever their medium.
  - Regulation (EC) No 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission documents

# Right to access the information

- Basic principles of the right to information
  - Every person has the right to receive information held by government authorities
  - The applicant for information shall not be obliged to prove the reasons for his request
  - Access to information should be applied on an equal footing
  - The information should be provided within a reasonable time
  - The public authority must give reasons for the rejection of information
  - Rejection of the application must be reviewable.

# Reuse of information

- Main focus of European PSI legislation
- Guidelines for improving the synergy between the public and the private sectors in the information market (1989)
  - *“Public organizations should, as far as is practicable and when access is not restricted for the protection of legitimate public or private interests, allow these basic information materials to be' used by the private sector and exploited by the information industry through electronic information services.”*
- Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information
  - Amended by Directive 2013/37/EU
  - Goal: Economical effectivity

# Directive 2003/98/EC as amended by directive 2013/37/EU

- Does not create an obligation to provide information
- General principle - where information is provided, this should be done in such a way as to be reusable as easily as possible
- “Documents”
  - Art. 2 (3): any content (or its part) whatever its medium (written on paper or stored in electronic form or as a sound, visual or audiovisual recording)
  - Exceptions:
    - documents which do not fall within the public domain of the entity
    - third party intellectual property rights
    - restricted documents for privacy reasons
    - documents held by cultural institutions other than libraries, museums and archives
    - Possible situation - the document is accessible but not reusable

# Directive 2003/98/EC as amended by directive 2013/37/EU

- Stipulates quite technical provisions
  - Machine-readable format
  - Open format
  - Formal open standard

# 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

- Recast of the directive 2003/98/EC, implementation until 7/2021
- Heavier accent on ensuring the possibility of reuse
- 3 main novelties:
  1. Applies to Public undertakings
  2. High Values Datasets
  3. Applies to research data



# PSI v. IP Rights

- Possible means of protection
  - Copyright protection of content
  - Copyright protection of database
  - Sui generis database rights
- Possible solutions
  - In some jurisdictions – exception of official (administrative) work
  - Otherwise licensing
    - Best solution – Public licences
      - Property based
      - Building on Copyright
      - Mind change – enforcing openness and freedom, not exclusivity

“In all cases, the Commission encourages authors to retain their copyright and grant adequate licences to publishers. **Creative Commons offers useful licensing solutions.** This type of licence is a good legal tool for providing open access in its broadest sense.”

- Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, p. 8

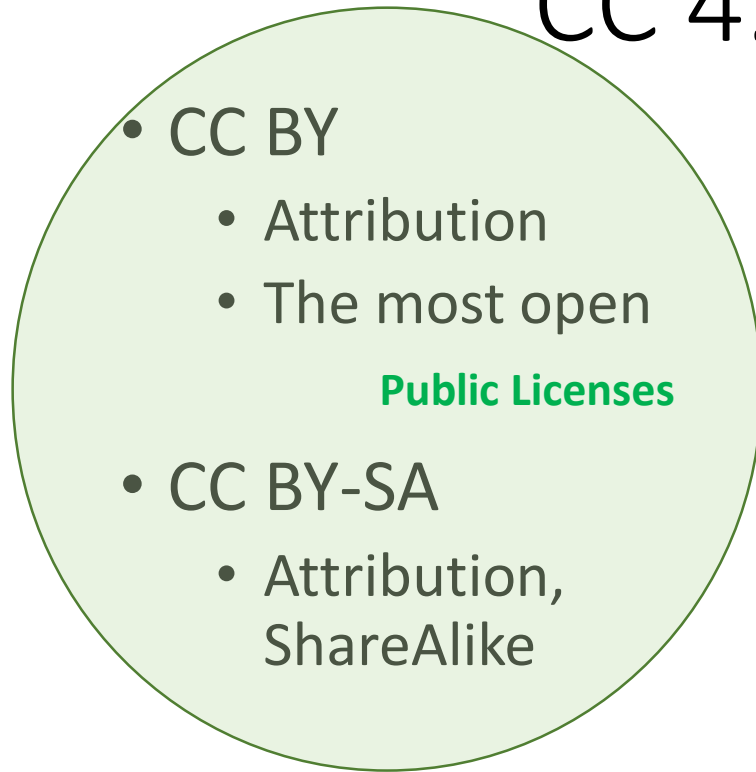
# Public licences

- Open Content Definition “5R”:
  - Retain
  - Reuse
  - Revise
  - Remix
  - Redistribute

# Public licenses – examples

- Creative commons (Attribution, Share Alike, Zero)
- Open Publication License
- Against DRM license
- GNU Free Documentation License
- Open Game License
- Free Art License
- Open Database Licenses (ODbL)

# CC 4.0 Licenses



- CC BY-ND
  - Attribution, NoDerivs

- CC BY-NC
  - Attribution, NonCommercial
- CC BY-NC-SA
  - Attribution, NonCommercial, ShareAlike
- CC BY-NC-ND
  - Attribution, NonCommercial, NoDerivs

# CC0 1.0 Universal

- Applicable on:
  - Copyright protected work
    - the right to reproduce, adapt, distribute, perform, display, communicate, and translate;
    - moral rights retained by the original authors;
    - publicity and privacy rights pertaining to a person's image or likeness depicted in a Work;
    - rights protecting the extraction, dissemination, use and reuse of data in a Work;
    - database rights;
    - Etc.
- A waiver – after publishing under CC0 the work becomes public domain
- Link: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>

# CC0 1.0 Universal

- A problem: Some jurisdictions do not allow to waive the rights
  - e.g. in Czechia copyright can be only licensed but Sui generis database right can be waived
- Solution:
  - Public Licence Fallback

Don't forget!

Nemo plus iuris ad alium transferre potest  
quam ipse habet.

No one can transfer more rights (to another)  
than he himself has.



# PSI v. Personal Data Protection

- It there is personal data, the protection framework applies
  - Both on provider and reuser
- Regulation 2016/679 (GDPR)
  - Art. 4 (1):
    - ‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name
  - Necessary legal grounds to process data
    - Consent
    - Legal duty
    - Legitimate interest

# Open Data & Open API

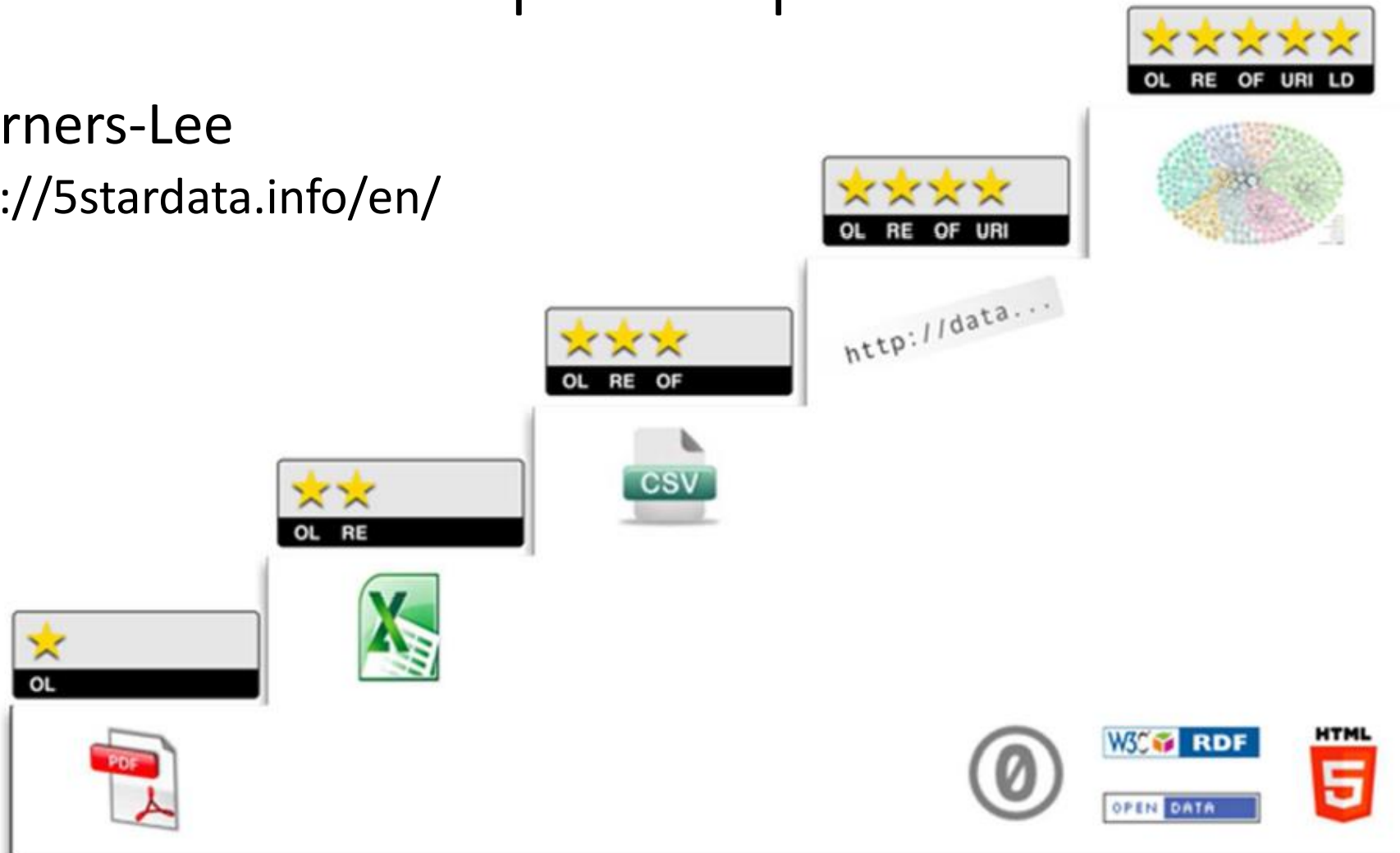
- An efficient way of PSI publication
  - We are not concerned on “what” is provided, but “how”
- Rufus Pollock (founder of Open Knowledge Foundation):
  - *“The best thing to do with your data will be thought of by someone else.”*
- The Open Definition:
  - “Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness).”
  - “Open data and content can be **freely used, modified, and shared** by **anyone** for **any purpose**”
    - <https://opendefinition.org/>

# General conditions on Open Data

- Data must be:
  - Complete
  - Easily accessible (online)
  - Machine readable
  - Using open standards
  - No (minimal) legal restrictions for reuse
  - Available to users at the minimum possible cost

# Five steps of openness

- Tim Berners-Lee
  - <http://5stardata.info/en/>

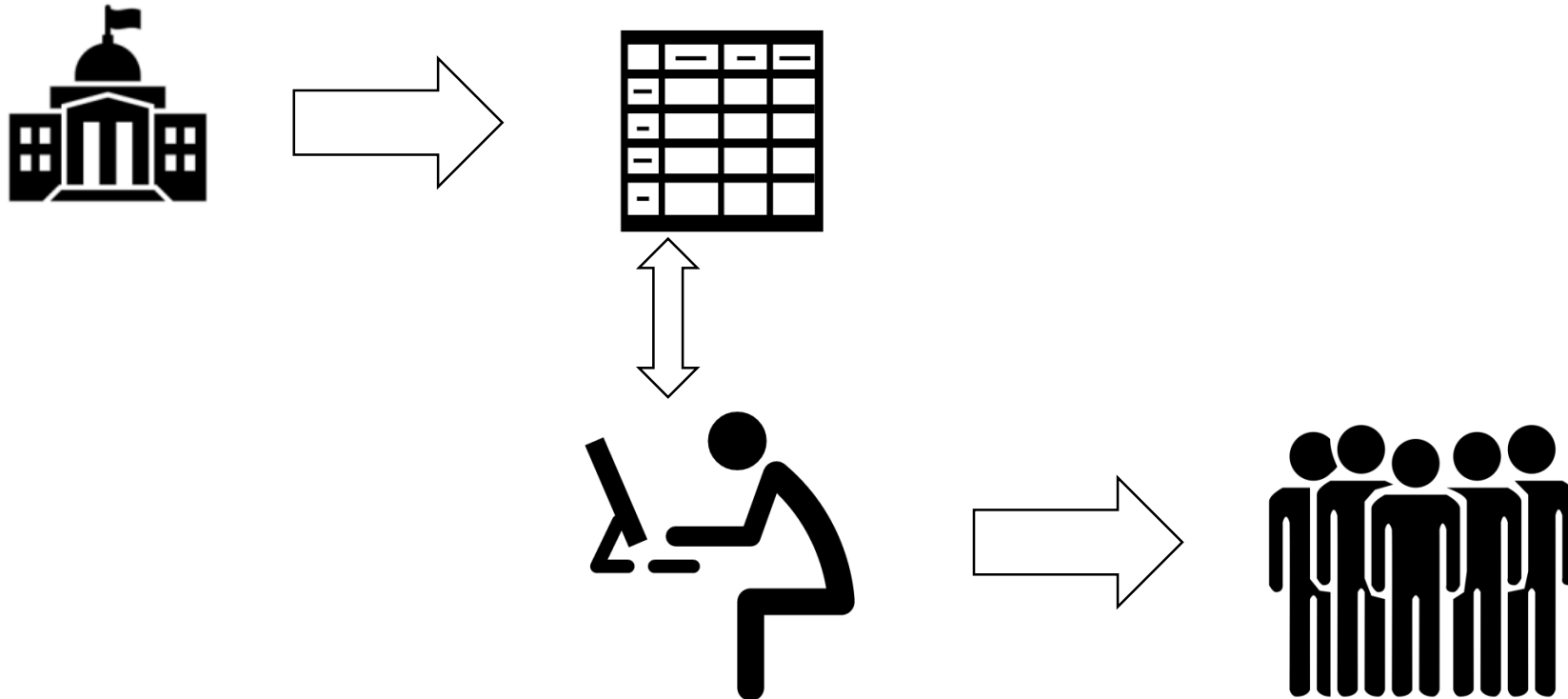


Zdroj: HAUSENBLAS, Michael. 5 star Open Data. Dostupné z: <http://5stardata.info/>

# How does it work

PSI provided as Open Data (Open API)

A third party developer creates an application



# Open Data & Open API

- Open Data:
  - Bulk download
- Open API (Application Programming Interface)
  - Machine to machine access to the database
  - Specifications freely published online
  - Possible condition: API Key
  - Example: IoT



...and now for something completely different

# Research Data

- Data coming from research in a digital form
  - Facts or numbers
  - Basis for reasoning, discussion or calculation
  - E.g.:
    - Statistics
    - Results of experiments
    - Measurements
    - Observations resulting from fieldwork
    - Survey results
    - Interview recordings
    - Images



# Open Research Data

- Research data published as open data
- Reasons:
  - Public funding
  - Validation of results
  - Possible reuse

# Open Research Data - H2020

- H2020 beneficiaries should make their research data findable, accessible, interoperable and reusable (FAIR)
- Art. 29 of the H2020 Model Grant Agreement
  - Obligation to disseminate results
  - (3) Research Data
- Make possible to for third parties to:
  - Access
  - Mine
  - Exploit
  - Reproduce
  - Disseminate
- Possibility to opt-out
  - during the application phase
  - during the grant agreement preparation (GAP) phase and
  - after the signature of the grant agreement.

# Open Research Data - H2020

- Data Management Plan

- should include information on:

- the handling of research data during & after the end of the project
    - what data will be collected, processed and/or generated
    - which methodology & standards will be applied
    - whether data will be shared/made open access and
    - how data will be curated & preserved (including after the end of the project).

- More info: [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)

# Open Research Data - Open Data Directive (2019/1024)

- Rec. 27:
  - ... Research data includes statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. It also includes meta-data, specifications and other digital objects. Research data is different from scientific articles reporting and commenting on findings resulting from their scientific research. For many years, the open availability and re-usability of scientific research data stemming from public funding has been subject to specific policy initiatives. Open access is understood as the practice of providing online access to research outputs free of charge for the end user and without restrictions on use and re-use beyond the possibility to require acknowledgement of authorship. Open access policies aim in particular to provide researchers and the public at large with access to research data as early as possible in the dissemination process and to facilitate its use and re-use. ...

# Open Research Data - Open Data Directive (2019/1024)

- Rec. 28:
  - For the reasons explained above, it is appropriate to set an obligation on Member States to adopt open access policies with respect to publicly funded research data and ensure that such policies are implemented by all research performing organisations and research funding organisations. ... Open access policies typically allow for a range of exceptions from making scientific research results openly available. ... Additionally, the conditions, under which certain research data can be re-used, should be improved. For that reason, certain obligations stemming from this Directive should be extended to research data resulting from scientific research activities subsidised by public funding or co-funded by public and private-sector entities. Under the national open access policies, publicly funded research data should be made oHowever, in this context, concerns in relation to privacy, protection of personal data, confidentiality, national pen as the default option. security, legitimate commercial interests, such as trade secrets, and to intellectual property rights of third parties should be duly taken into account, according to the principle ‘as open as possible, as closed as necessary’.

# Open Research Data - Open Data Directive (2019/1024)

- Art. 10
  - Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'), following the principle of 'open by default' and compatible with the FAIR principles. In that context, concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests, shall be taken into account in accordance with the principle of 'as open as possible, as closed as necessary'. Those open access policies shall be addressed to research performing organisations and research funding organisations.

# Providing open access to research data

- If content of the database is protected by copyright
  - Publish it under a public licence (CC-BY or CC-BY-SA preferably)

# Providing open access to research data

- If the database is protected by copyright
  - Publish it under a public licence (CC-BY or CC-BY-SA preferably)



# Providing open access to research data

- If the database is protected by Sui generis database right and you:
  - a) Wish attribution
    - Publish it under a public licence (CC-BY or CC-BY-SA preferably)
  - b) Want to open it as much as possible
    - Wave the rights by CC0

# Providing open access to research data

- It can combine
  - Copyright protection of content & Sui generis right
  - Copyright protection of content & Copyright protection of Database
  - Copyright protection of Database & Sui generis right
  - All three of them
  - None of them

# Providing open access to research data

- If the database is not protected either by Copyright or Sui generis database right:
  - State it
  - It is not a licence – there is nothing to license
- Possibility to provide content under contract

Thank you for your attention.

Questions?

@jkb\_misek