Legal framework to linguistic data

Jakub Michálek

member of the Assembly of the capital city Prague

Nothing in this presentation is to express the opinions

or statements of the capital city Prague.

What data are we talking about?

- Big data
- Parallel data (translations)
  - Official registry collection of documents
  - International treaties
  - European union documents
  - Official translations of Czech government documents
  - Possibly books – obligatory print (+ license issues) or public domain

Securing rights for translated work

- Translation is a **derivative work**, the use of which is also the use of the translated work (§ 2/4)

- The contract should include an **exclusive license** for the translation with the right to issue sublicenses (if not by law in case of works for hire)

- If the translation is to be an official work (§ 3 a), the contract should include agreement on this

- For linguistics, the contract should include a sufficiently wide public license (such as the Creative Commons) or this should be granted by the exclusive copyright holder

Ways to get the data

- **text harvesting online** – existing collections, subtitles, official work § 3 a), public domain § 28, political speeches § 34 d …
- **FOIA request** – according to the freedom of access of information act (PSI directive) – even university libraries providing public information services
- **book translations** – obligatory print, digitazation, quotation exception § 31/1 b,c – for purpose of research, only necessary scope – either non-commercial or honest practice

Ways to get the data

- **archives** – university can have copy any works to its non-commercial archive (§ 37/1 a)

- **orphan works** (§ 37a) – register of coll. soc.

- **works for hire** (§ 58) where the copyright holder is a public entity, in document registry

- **students' works** § 35/3 – university library, for instance student translations at FHS UK

Publishing data

- To publish data is a **legally permitted** (§ 5/7 FOIA) – with regard to exceptions of the law

- When publishing there is legal duty to include **all language versions** created (§ 4b FOIA)

- Remove <u>protected</u> personal data under 101/2000 Coll.

- Clarify status and licensing conditions (official work, work for hire, licensed work), grant a public license if necessary

- § 5/3 FOIA – publish requested data