Dissemination of language resources

Aleksei Kelli
Professor of IP Law
aleksei.kelli@ut.ee
Background:

• Developing open science & open data strategy for Estonia;
• Analysis of CLARIN LR categories;

Relevant publications:


Open science & open data:

“If you have an apple and I have an apple and we exchange these apples then you and I will still each have one apple. But if you have an idea and I have an idea and we exchange these ideas, then each of us will have two ideas.”

— George Bernard Shaw

Source: http://www.goodreads.com/quotes/23088-if-you-have-an-apple-and-i-have-an-apple
Opportunities:

• Builds public domain knowledge base;
• Saves taxpayers’ money (no need to repeat the study);
• Allow verification of research results;
• Enhances development of new services;
• Facilitates cooperation;
• Creates new possibilities (e.g. TDM);
• Supports democratic processes.
Data:

- OECD (2015): research data as factual records used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings;
- Personal data: “any information relating to an identified or identifiable natural person” (GDPR Art. 4 (1));
- Public sector information (PSI): information documented in any manner which is obtained or created upon performance of public duties (PIA § 3).
Open science & data:

- European Commission (2016): open science moves towards sharing and using all available knowledge at an earlier stage in the research process.
- OECD (2015) defines open data as “data that can be used by anyone without technical or legal restrictions. The use encompasses both access and reuse”;
- Open Knowledge International: “open data and content can be freely used, modified, and shared by anyone for any purpose”.
Legal rights covering data:

Related rights

Future development: data producer rights covering raw data?

Copyright does not protect fact/data

Database protection (collection of data)

Personal data protection

...
Example of language data:

Two tiers of rights covering language resources

| Language resources as a database | · Copyright  
|                                  | · Related rights  
|                                  | · Privacy rights  
|                                  | · Personal data protection |

| Material used to develop language resources | · Copyright  
|                                           | · Related rights  
|                                           | · Privacy rights  
|                                           | · Personal data protection |
The key features of openness (Open Knowledge International):

- Availability and access;
- Reuse and redistribution;
- Universal participation

Source: https://okfn.org/opendata/ (4.11.2019)
Dissemination:

Categorization of data (e.g. CLARIN model)

Quality & liability (dissemination on "as-is" and "as-available" bases)

Dissemination of open data

FAIR principles (FORCE11)

Time (pre-publication dissemination?)

FINDABLE
ACCESSIBLE
INTEROPERABLE
RE-USABLE
Open data & entrepreneurial university:

- Entrepreneurial university
  - Open data
  - Socio-economic impact
  - Accessible and reusable data
  - IP protection
- Patenting
  - Trade secret protection
- Compatibility of policies
- Mission of the university
- Knowledge transfer
  - Spin-off creation
  - Licensing
  - Assignment of IPRs
  - Consultancy

Interaction of open data and entrepreneurial university policies
Implementation:

Open data policy

- Complicated to define research data (e.g., language data)
- Dissemination
  - Time
  - Quality
  - Responsibility (access on "as-is" bases)

Competing policies
- IPR
- Entrepreneurial university
- Reuse of PSI

Incentives for researchers
- Copyright reform
- Personal data protection reform
- Data producer’s rights covering machine-generated raw data

EU initiatives
Implementation:

Open data policy implementation

State level
- Policies and regulations
- Research funding

Organisational level
- Administrative support
- Technical support

Individual level
- Monetary incentives
- Career incentives
Implementation:

- Researcher as a key player
  - Incentives to make data available
    - Research money
    - Career
    - Helping others
    - Make the world a better place
  - Incentives to keep data
    - Get the maximum number of publications
    - Do not waste time on making data available
    - Fair of misuse of data
    - Fair that errors become evident
CLARIN licensing scheme:

CLARIN resources
- PUB (public)
- ACA (academic)
- RES (restricted)

OPEN
- Academic use
- Non-commercial use
- Other restrictions

Language technology development:

Case of folklore:

- Previously unpublished works
- Personal data protection
  - Private use
  - Research & education
  - Creation of derivative works

Sui generis database

Database protection

Exceptions & limitations

- Enforcement
  - State property (e.g. Bahrain)
- Performance
- Derivative work
- Recordings (phonograms)
- Orphan works
Tänan