

---

# Lizenzierungspraxis in Daten-Clouds - Empirische Befunde und Folgeprobleme

Tassilo Pellegrini

Department Medien & Wirtschaft, Fachhochschule St. Pölten

Ivan Ermilov

Institut für Angewandte Informatik, Universität Leipzig

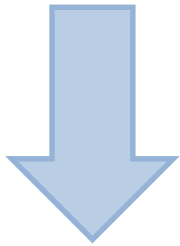
Digitale Bibliothek 2016

Universität Graz

26.02.2016

# Daten als ökonomische Güter medien & wirtschaft

## Datenwerke



## Daten-Assets



<b>Instanz-Daten</b>
<b>Metadaten</b>
<b>Ontologie</b>
<b>Content</b>
<b>(Services)</b>
<b>(Technologie)</b>

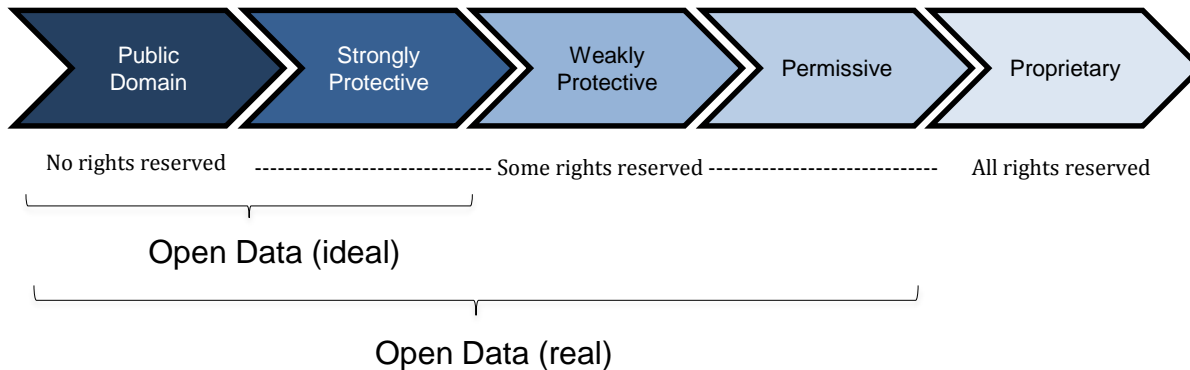
## Schutzinstrumente



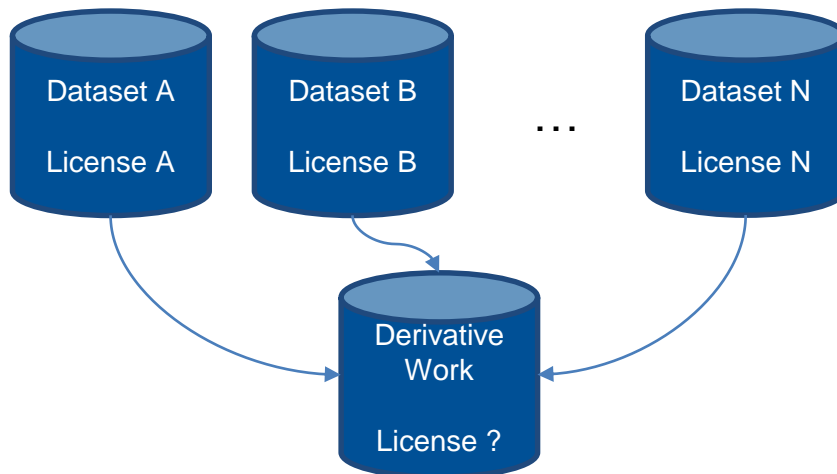
	<b>Urheberrecht</b>	<b>Datenbank-recht</b>	<b>Wettbew. Recht</b>	<b>Patentrecht</b>
<b>Instanz-Daten</b>	Case by Case	Ja	Ja	Case by Case
<b>Metadaten</b>	Case by Case	Ja	Ja	Case by Case
<b>Ontologie</b>	Ja	Ja	Ja	Case by Case
<b>Content</b>	Ja	Nein	Ja	Nein
<b>(Services)</b>	Ja	Nein	Ja	Ja
<b>(Technologie)</b>	Ja	Nein	Ja	Ja

# Problem derivativer Datenwerke



Lizenzspektrum:



- Derivative Datenwerke setzen sich (meist) aus mehreren Datensätzen zusammen
  - Datensätze können abweichende Lizenzbestimmungen aufweisen, die Konflikte erzeugen
  - Aufwand der Rechtklärung steigt mit jeder zusätzlichen Datenquelle:  $f(n) = n \cdot (n-1) / 2$
- Welche Nutzungsformen / Anwendungen sind zulässig?
- Welche Lizenzform ergibt sich für das derivative Werk?



# Anforderungen an Datenlizenzierung in der Cloud

- Datensätze sollten eine Lizenz ausweisen (Provenance + Usage Rights)
- Lizenzinformation sollte verfügbar sein in
  -  Menschen-lesbarer Form (Text)
  -  Maschinen-verarbeitbarer Form (Serialisierung)
- Maschinen-verarbeitbare Lizenz sollte auf offenen Standards basieren und logische Operationen erlauben (Reasoning)
- Maschinelle Repräsentationen müssen rechtlich valide sein

# Empirische Erhebung zur Lizenzierungspraxis: Untersuchungsgegenstand

Followers

0

## Organization



## ie-ckan-net

There is no description for this organization

## Social

Google+

Twitter

Facebook

## License

Other: License Not Specified

## Psychiatric Services

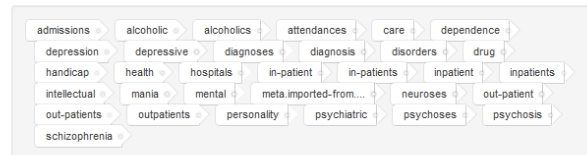
Statistics of Psychiatric Services from [www.statcentral.ie](http://www.statcentral.ie) under the theme People and Society - Health from the Department of Health and Children

Classifications:

- Health Board/Regional Authority
- Marital status
- Diagnosis

## Data and Resources

This dataset has no data



## Additional Info

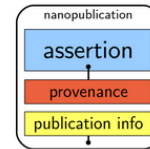
Field	Value
Source	<a href="http://www.statcentral.ie/viewStat.asp?id=59">http://www.statcentral.ie/viewStat.asp?id=59</a>
Author	<a href="mailto:info@health.gov.ie">info@health.gov.ie</a>
Last Updated	November 27, 2014, 20:51
Created	November 27, 2014, 20:51

<https://datahub.io/dataset/psychiatric-services>

Followers

1

## Organization



## Nanopublications

Datasets formatted as nanopublications:  
<http://nanopub.org> [read more](#)

## Social

Google+

Twitter

Facebook

## License

Open Data Commons Open Database License (ODbL)

[View ODbL](#)

## Nanopubs extracted from DisGeNET v2.1.0.0

940034 nanopublications. These nanopubs were automatically extracted from the DisGeNET dataset. See also the [main DisGeNET data on Datahub](#).

Download the content of this set of nanopublications from the server network using [nanopub-java](#):

```
$ np get -c -o nanopubs.trig RAWY3S2hufPKpm/PC-wQ3AT7g4Vla-Q48D1pr29L108Q
```

## Data and Resources

[Datahub](#) **Access to nanopublication index** [More information](#) [Go to resource](#)  
Download the content of this set of nanopublication:

[Datahub](#) **Access to nanopublication index** [More information](#) [Go to resource](#)  
Download the content of this set of nanopublication:

[Datahub](#) **Nanopublications** [More information](#) [Go to resource](#)  
Gzipped TriG file

biology | biomedicine | diseases | disgenet | gene-disease-assoc... | genes | life sciences | text mining

## Additional Info

Field	Value
Author	<a href="#">Núria Queralt Rosinach</a>
Maintainer	<a href="#">Núria Queralt Rosinach</a>
Version	v2.1.0.0
Last Updated	November 24, 2015, 15:50
Created	July 10, 2015, 12:11
content size	940034 nanopubs
index size	941 nanopubs

<https://datahub.io/dataset/disgenet-v2-1-0-0-nanopubs>

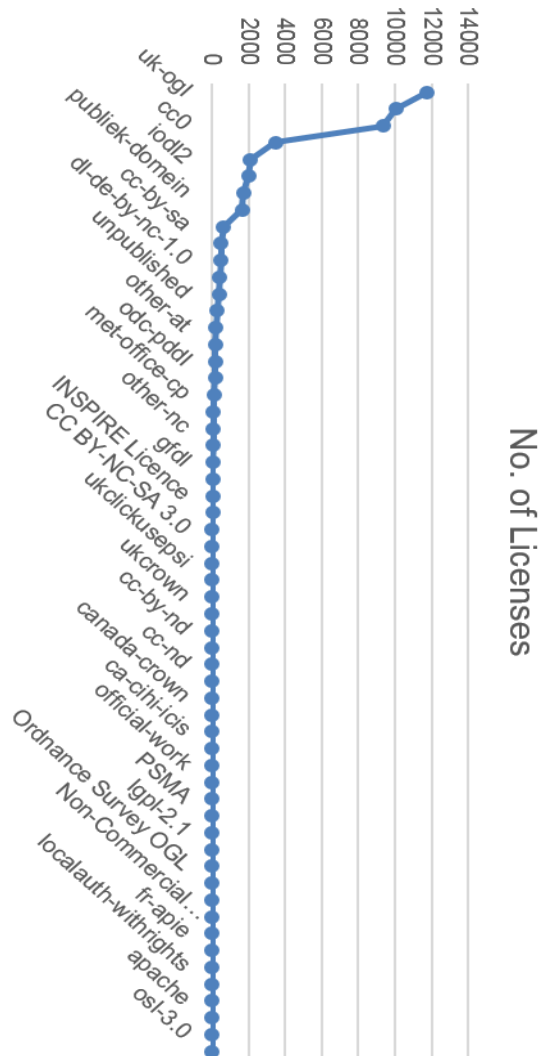
# Empirische Erhebung zur Lizenzierungspraxis: Methode

- Sind die Datensätze mit Lizenzinformation ausgestattet?
- Welche Lizenzen kommen zum Einsatz?
- Wie hoch ist der Anteil maschinen-verarbeitbarer Lizenzinformation?
- Existieren Kompatibilitätskonflikte innerhalb und zwischen den Daten-Clouds?
- Extraktion von Lizenzinformation mittels Crawler
  - CKAN aggregator:  
<https://github.com/AKSW/ckan-aggregator-py>
- Vier Datenkataloge
  - Europe: Publicdata.eu
  - USA: Data.gov
  - Canada: Open Data Canada
  - Open Knowledge Foundation: Datahub.io
- 441.315 Datensätze
- April 2015

# Ergebnisse

Data Licenses on the Cloud				
	Datagov	Open Canada	Public Data	Datahub
<b>Datasets</b>	132 206	244 257	55 481	9 371
<b>License Types</b>	10	3	50	33
<b>Not Specified</b>	99.6%	0.0%	24.3%	59.1%
<b>CC</b>	0.4%	0.0%	35.3%	17.1%
<b>ODC</b>	0.0%	0.0%	0.5%	4.8%
<b>Other</b>	0.0%	100.0%	39.9%	19.0%
<b>Deref. Link</b>	0.4%	100.0%	43.2%	23.1%
<b>Mach. Read.</b>	0.0%	0.0%	2.6%	2.2%

# Ergebnisse



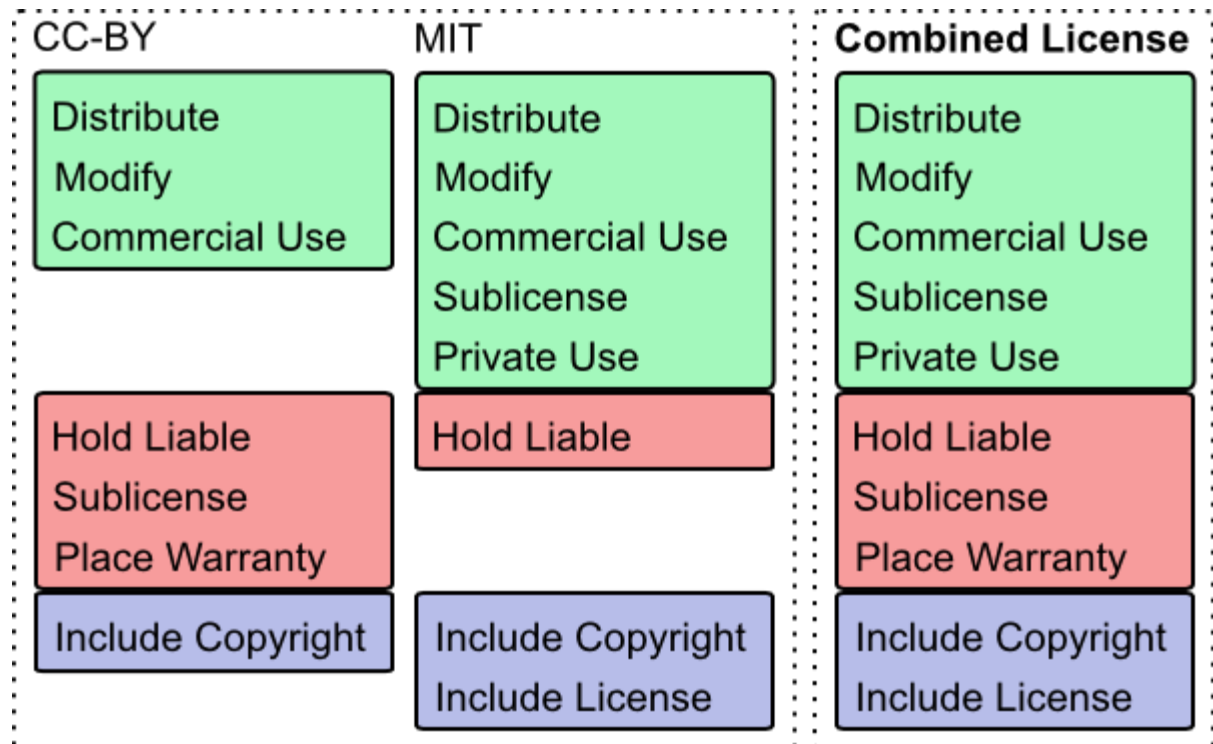
Data Licenses on the Cloud				
	Datagov	Open Canada	Public Data	Datahub
<b>Datasets</b>	132 206	244 257	55 481	9 371
<b>License Types</b>	10	3	50	33
<b>Not Specified</b>	99.6%	0.0%	24.3%	59.1%
<b>CC</b>	0.4%	0.0%	35.3%	17.1%
<b>ODC</b>	0.0%	0.0%	0.5%	4.8%
<b>Other</b>	0.0%	100.0%	39.9%	19.0%
<b>Deref. Link</b>	0.4%	100.0%	43.2%	23.1%
<b>Mach. Read.</b>	0.0%	0.0%	2.6%	2.2%

- Spezifische regionale Unterschiede
- Hohe Lizenzheterogenität
- Unzureichende Informationslage
- Kompatibilitätskonflikte
- Kaum maschinen-verarbeitbare Versionen



# Folgeprobleme

- Rechtklärung sehr aufwendig und fehleranfällig
- De facto keine Grundlage für Automatisierung gegeben
- Automatisch kompillierte Lizenzen entbehren Rechtsgrundlage



# Limitationen und Future Work

- Weitere Untersuchungen notwendig:
  - Art und Umfang von Kompatibilitätskonflikten
  - Domänen-abhängige Lizenzierungsprobleme (→ Klassifikationsproblem)
  - Untersuchung zeitlicher Veränderungen (→ insb. Permissive Licensing)
- Maschinelle Repräsentation von Lizenzen bedarf juristischer Validitätschecks
- Bisher existieren keine belastbaren Tools, die Reasoning über Lizenzinformation erlauben
- Rechtssicherheit in der Lizenzierung ist Voraussetzung für Industrie 4.0 bzw. Internet der Dinge

# Vielen Dank für Ihre Aufmerksamkeit!

This work has been carried out within the project NoLDE (Network of Linked Data Excellence) funded by the Austrian Promotion Research Agency under grant number 3592880. We acknowledge support from GeoKnow project, GA number no. 318159, as well as BMBF project SAKE.

[tassilo.pellegrini@fhstp.ac.at](mailto:tassilo.pellegrini@fhstp.ac.at) Tassilo Pellegrini  
[iermilov@informatik.uni-leipzig.de](mailto:iermilov@informatik.uni-leipzig.de) Ivan Ermilov