



ANNUAL CONFERENCE 2017
Budapest, Hungary

Implementation of an Open Science Policy in the Context of Management of CLARIN Language Resources: A Need for Changes?

Aleksei Kelli, Krister Lindén, Kadri Vider,
Penny Labropoulou and Erik Ketzan



CLARIN

Framework

- Open Science Policies in Europe: open access, sharing data, connecting disciplines, connecting data, etc. → European Open Science Cloud
 - CLARIN & Open Science:
 - not a stand-alone facility, but well-embedded in the European research infrastructure landscape at large, and as such fully committed to the European Open Science Policies, including re-use of data
- CLARIN Value Proposition 2016*
- but how is this **reflected** in the Management of Language Resources in CLARIN?

What is Open Science?

- based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools

EC 2016

- Open Access criteria:

- cover research results, raw data and metadata, source materials, digital pictorial and graphical materials, etc.;
- right holders grant to all users a license to use, distribute, and to make and distribute derivative works;
- a complete version of the work and all supplemental materials in an appropriate standard electronic format is deposited

Berlin Declaration on Open Access

- Openness

- Availability and access
- Reuse and redistribution
- Universal participation

Open Knowledge International

Licensing categorization s overview

- Main objective
 - for indexing in data catalogues
 - to facilitate search for users
 - a first hint on what users can do wi
 - NOT replacing licenses!
- Approaches to licensing categorization
 - "dissecting" licenses into features
 - e.g. CLARIN conditions of use, META-SHARE con
 - mainly used in license selectors and other [helper](#), [public license selector](#), [CLARIN lic](#)
 - in some repositories (e.g. clarin:el)
 - grouping licenses based on a feature de
 - CLARIN license categories

The screenshot displays a user interface for license categorization. It is divided into two main sections: 'Availability' and 'CAN I USE IT?'. The 'Availability' section includes a list of license levels with checkboxes: Public (26928), Academic (47), Restricted for, and Unspecified (1). Below this list is a note: 'These levels provide resources and tools a the specific condition end up using.' The 'CAN I USE IT?' section is a dropdown menu with several categories, each with a count: Free Re-use (910), Public Domain Marked (824), CC BY (85), CC0 (3), Limited Re-use (724), CC BY-NC-ND (716), CC BY-NC-SA (5), CC BY-NC (3), No Re-use (1,379), In Copyright (1,270), Copyright Not Evaluated (106), and Rights reserved - Free access (3). Below this menu is a section titled 'Restrictions of Use' with a list of features: Attribution (382), Academic - Non Commercial Use (209), Share Alike (52), Other (9), No Redistribution (8), and No Derivatives (6). The word 'less' is visible at the bottom of the interface.

COAR controlled vocabulary of access rights

- Confederation of Open Access Repositories
- mainly scholarly publications, but also other research data (datasets, software, etc.)
- "concepts to declare the access status of a resource"
 - **open access**: Open access refers to a resource that is **immediately and permanently online**, and **free for all** on the Web, **without financial and technical barriers**
 - **embargoed access**: Embargoed access refers to a resource that is metadata only access until released for open access on a certain date. Embargoes can be required by publishers and funders policies, or set by the author (e.g such as in the case of theses and dissertations).
 - **restricted access**: Restricted access refers to a resource that is available in a system but with **some type of restriction for full open access**. This type of access can occur in a number of different situations. Some examples are described below: The user must log-in to the system in order to access the resource The user must send an email to the author or system administrator to access the resource Access to the resource is restricted to a specific community (e.g. limited to a university community)
 - **metadata access**: Metadata only access refers to a resource in which access is limited to metadata only. The resource itself is described by the metadata, but is not directly available through the system or platform. This type of access can occur in a number of different situations. [...]

Rights statements from rightsStatements.org

- joint initiative of Europeana and the Digital Public Library of America
- any kind of digital cultural object (or its image) [*cultural heritage institutions and online cultural heritage aggregation platforms*]
- "high-level summaries of the underlying rights status"
- 3 categories (based on copyright) – 12 statements

 **IN
COPYRIGHT**

 **EU ORPHAN
WORK**

  **EDUCATIONAL
USE PERMITTED**

  **NON-COMMERCIAL
USE PERMITTED**

 **UNKNOWN
RIGHTSHOLDER**

 **CONTRACTUAL
RESTRICTIONS**

 **NON-COMMERCIAL
USE ONLY**

 **OTHER LEGAL
RESTRICTIONS**

 **IN THE
UNITED STATES**

 **COPYRIGHT
NOT EVALUATED**

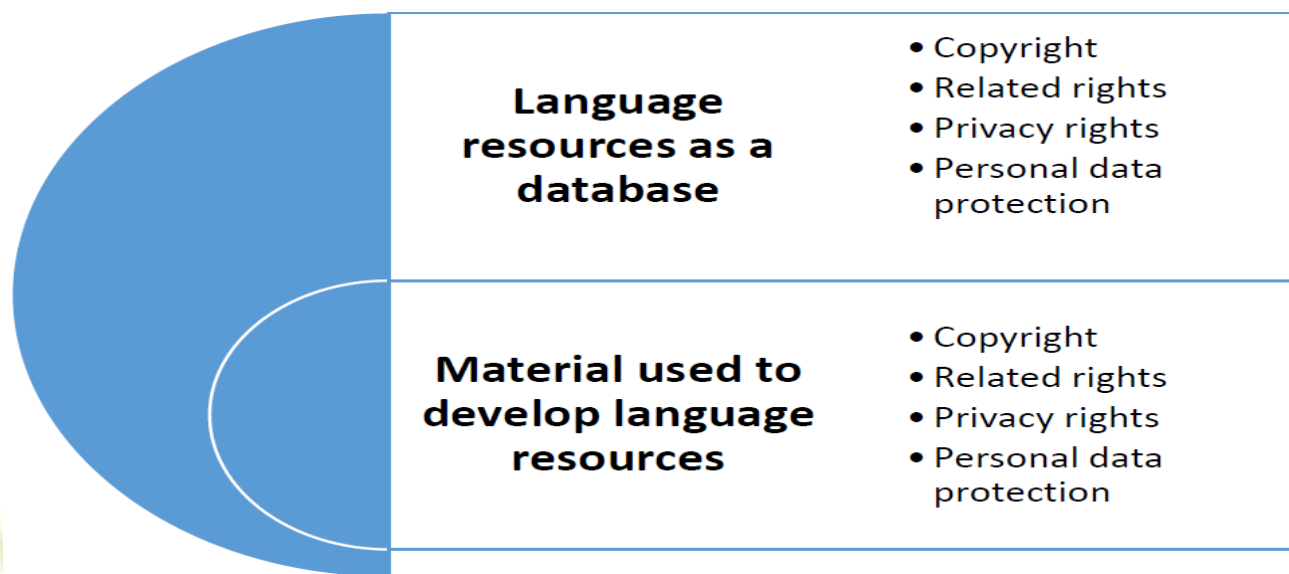
 **COPYRIGHT
UNDETERMINED**

 **NO KNOWN
COPYRIGHT**

Food for thought

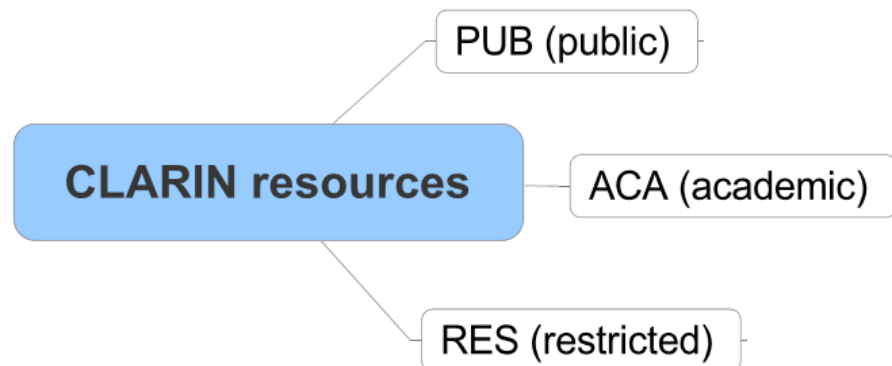
- controlled vocabularies with *clear* and *published* semantics
- possibility to link with concepts from other initiatives
- effort to standardize
- addressing the needs of their own user community
 - access vs. copyright
 - selected uses in rights statements: educational, non-commercial
- they do not replace licenses!

CLARIN licensing schemes: rights covering language resources

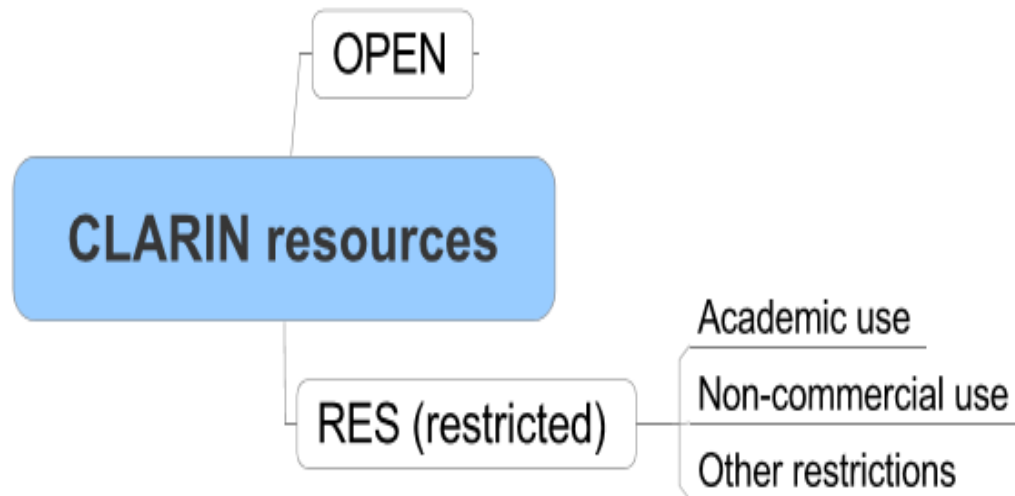


CLARIN licensing schemes

- Models used to create language resources:
 - The exception model (based on copyright exceptions and limitations);
 - The contract model (material is licensed)
- License categories (Oksanen, Lindén, Westerlund 2010):



The alternative categorization of CLARIN resources





Thank you

Aleksei Kelli & Penny Labropoulou