

# CLASSIFICATION AND LIBRARY LINKED DATA: the case of UDC

Aida Slavic

Editor-in-Chief, UDC

UDC Consortium, The Hague

# LINKED DATA – PROGRAMMATIC ACCESS TO DATA

Automatic linking of data on the Web by computer programs – programmatic access or *machine-to-machine* (m2m)

Driven by 3 basic Web standards: HTTPS, RDF and URI

- XML/RDF technology (Resource Description Framework) - data presentation which enables precise definition, i.e. statements about properties of a resource and extension/expansion of the the resource description by (meta)data “nesting” and pointing to remote data sources

RDF depends on the following principles

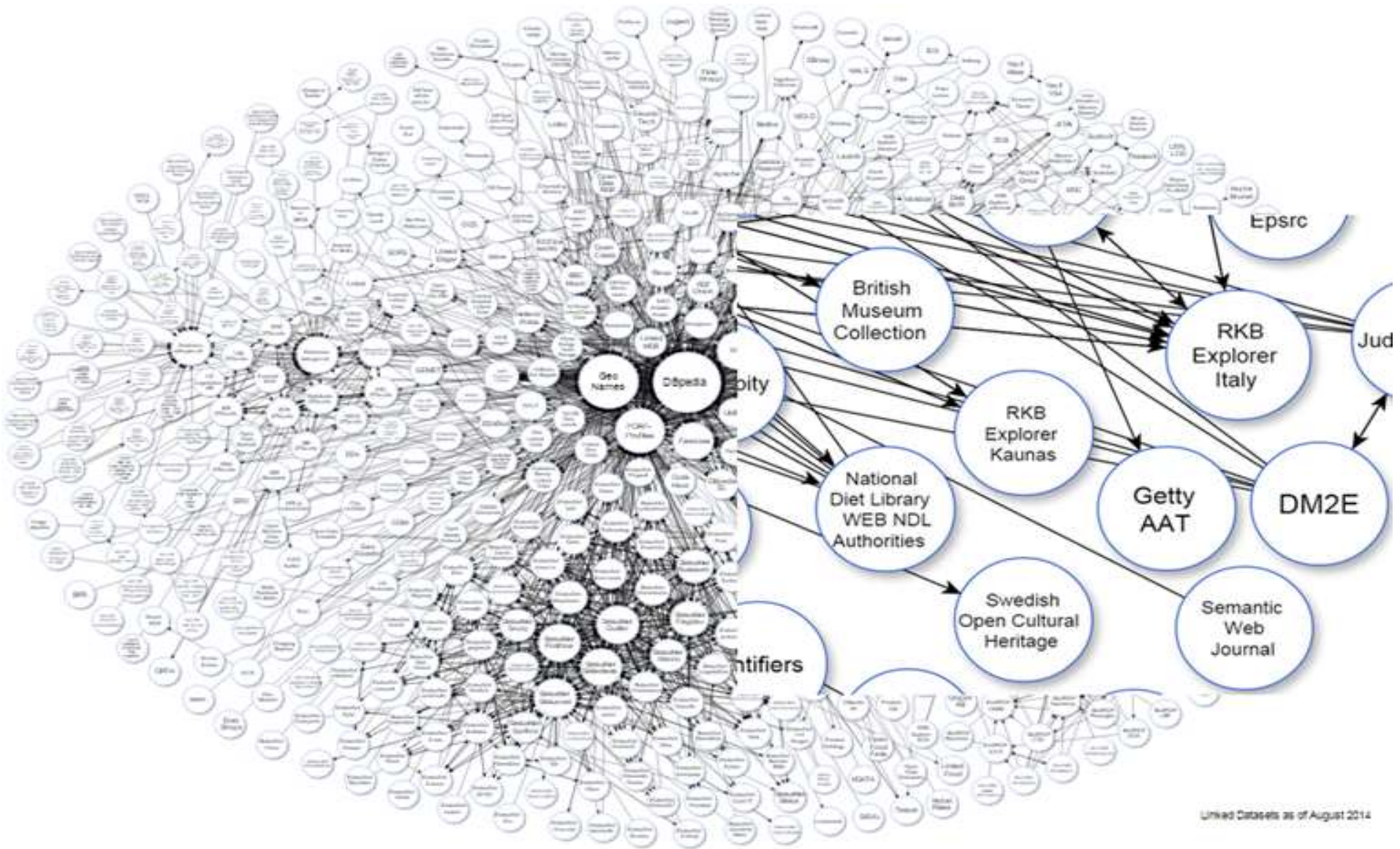
- each semantically significant unit or data set on the Web is described as a resource (a “thing”)
- each resource has a unique resource identifier (URI) which can be a Web address
- everything that can be identified can be linked and pointed to (from one resource to another)

# LINKED DATA AND BIBLIOGRAPHIC CONTROL

Publishing bibliographic resources as linked data (XML/RDF) enables the following functionalities in relation to information access:

- identification (as Web resources) of any of the following: an entire library collection, a library catalogue, each bibliographic record, each element of the bibliographic record;
- linking individual data within a bibliographic description, such as author's name, title, publisher, place of publishing, content description to additional information on a remote server;
- free access (by computer programs) to each data element of a bibliographic description and unlimited linking with semantically related resources on the Web.

# OPEN LINKED DATA (LOD)



# RATIONAL

Library classifications have the power to:

- improve and extend semantics and increase number of access points in the process of resource discovery
- improve subject retrieval in large numbers of collections in different languages / scripts in which a classification is used in content description

Two conditions

- Classification has to be available on the Web and open for m2m, i.e. programmatic access
- Library catalogues have to be available on the Web and open for m2m access



# CLASSIFICATION ENABLES ...

- Information retrieval of document collections in different languages and management of class notations and their verbal representations

519.8 Operational research (OR): mathematical theories and methods

Ricerca operativa (RO) teorie e metodi matematici

Mathematische theorieën en methoden van operationeel onderzoek

Математичні теорії та методи дослідження операцій

परिचालन शोध. गणितिय सिद्धांत और पद्यतियां

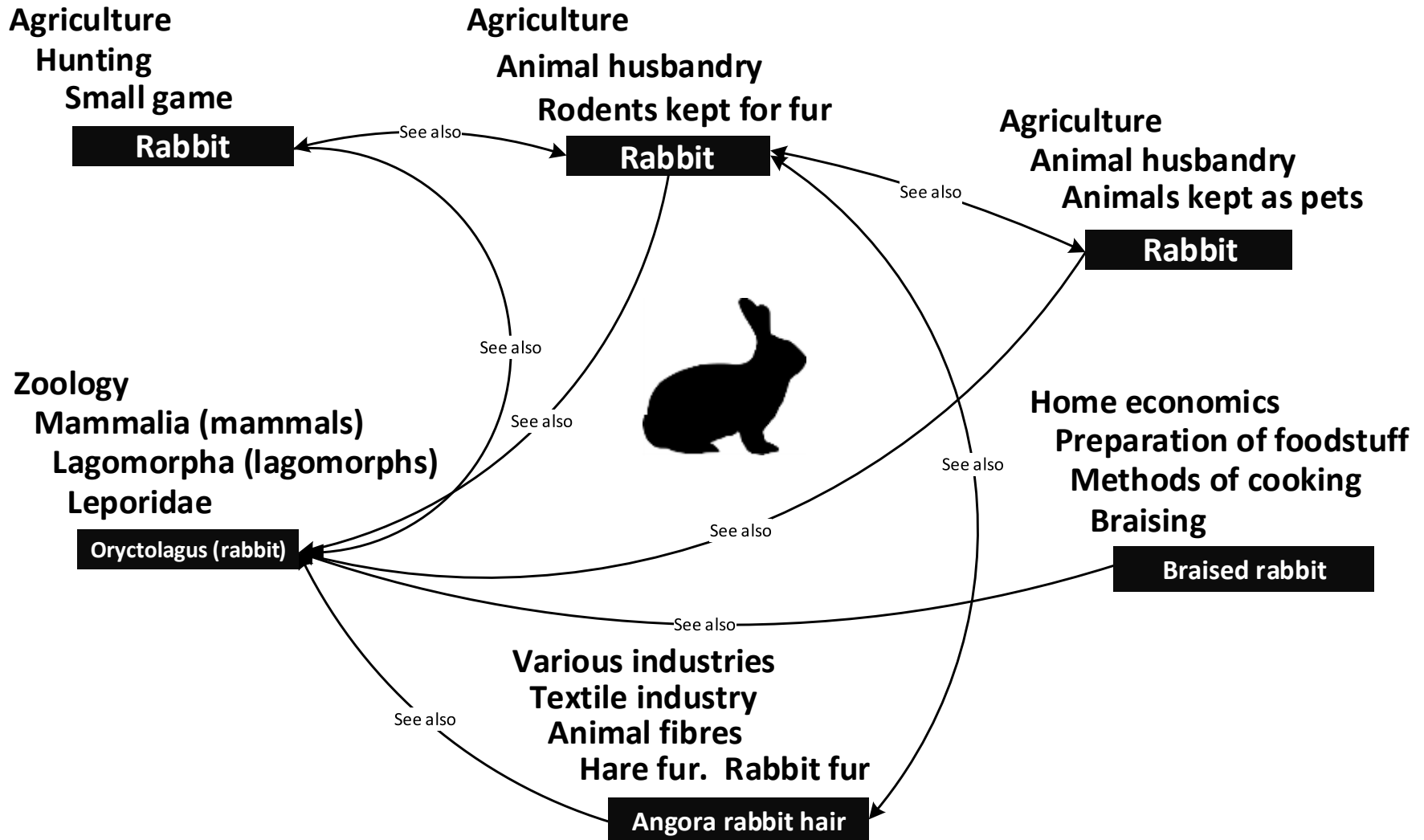
Τελεστική έρευνα, μαθηματικές θεωρίες και μέθοδοι

运筹学理论与方法

オペレーションズ・リサーチ(OR)の数学的理論

- Establishing hierarchical relationships between concepts or groups of concepts: which is useful in the graphical presentation of a knowledge space and browsing through a knowledge space using conventional knowledge forms
- Connecting concepts dispersed in knowledge and information universe (associative relationships )
  - 'rabbit' in zoology, sport, cooking, textile & fur industry, agriculture

# ASSOCIATION OF CONCEPTS IN KNOWLEDGE UNIVERSE



# LEGACY DATA TO TAKE INTO ACCOUNT

## Example of Universal Decimal Classification

- long-lived and dynamic system
- Translated into 57 languages (to various levels of depth), recorded use in collections in 138 countries: 2012 count of libraries in Europe and Asia only: at least 250,000 libraries
- 20 standard UDC “editions” since 1992.
  - 1992: 60,000    2019: over 71,000 classes
  - over 12,000 cancelled (deprecated) classes
  - over 22,000 new classes



# UNDERSTANDING CLASSIFICATION CONTEXT

- from 2009 UDC Summary – over 2.600 classes (6% UDC Master Reference File) published under Creative Commons licence (57 languages) – average use daily use figures 20,000 accesses
- from 2011. UDC Summary linked data (SKOS XML/RDF) to support research and experimenting with UDC data
- UDC translation databases
  - UDC Summary (57 languages)
  - Complete UDC MRF (13 languages)
  - Standard abridged UDC, 11,000 classes (17 languages)
- 2013 -2019 UDC Online Multilingual hub – complete UDC schedules online in 7 (9) languages - end-user tool for indexing and classification <http://www.udc-hub.com>

# UDC SKOS PRESENTATION (XML/RDF)

UDC Summary Linked Data English

URI <http://udcdata.info/050820> RDF

Notation 63

Caption Agriculture and related sciences and techniques. Forestry. Farming. Wildlife exploitation

Application Note The 62-1/-9 auxiliaries are also applicable throughout 63, with the exception of -1... and -2... in 633/635

Broader class [6](#) APPLIED SCIENCES. MEDICINE. TECHNOLOGY

Narrower classes

- [630](#) Forestry
- [631/635](#) Farm management. Agronomy. Horticulture
- [636](#) Animal husbandry and breeding in general. Livestock rearing. Breeding of domestic animals
- [637](#) Produce of domestic (farmyard) animals and game
- [638](#) Keeping, breeding and management of insects and other arthropods

```
<skos:ConceptScheme rdf:about="http://udcdata.info/udc-schema">
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#ConceptScheme"/>
<dcterms:title>UDC Summary</dcterms:title><skos:note/>
</skos:ConceptScheme>

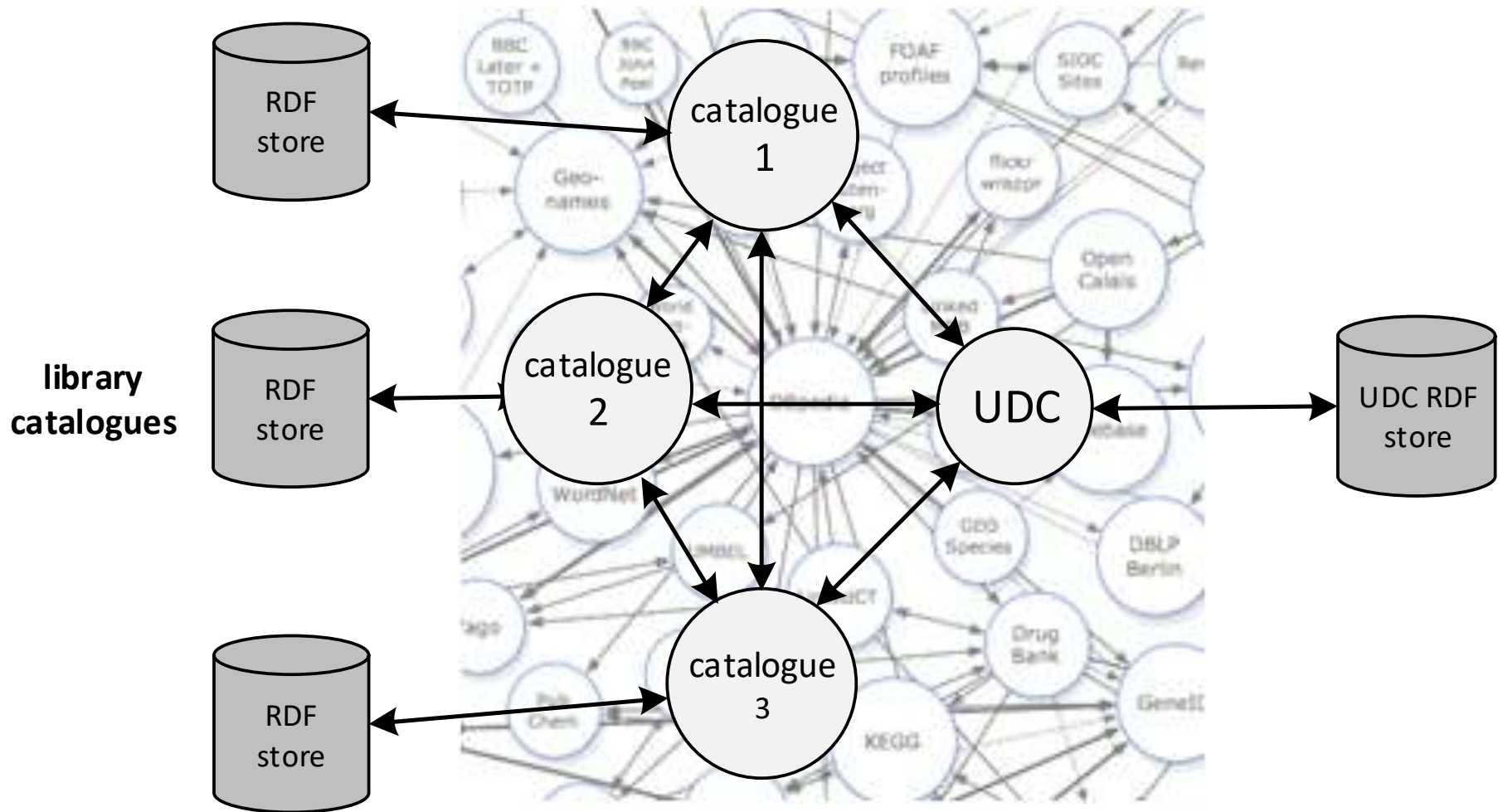
<skos:Concept rdf:about="http://udcdata.info/rdf/050820">
<skos:inScheme rdf:resource="http://udcdata.info/udc-schema"/>
<skos:broader rdf:resource="http://udcdata.info/rdf/037275"/>

<skos:notation rdf:datatype="http://udcdata.info/UDCnotation">63</skos:notation>

<skos:prefLabel xml:lang="en">Agriculture and related sciences and techniques. Forestry. Farming.
Wildlife exploitation</skos:prefLabel>

<udc:applicationNote xml:lang="en">The 62-1/-9 auxiliaries are also applicable throughout 63, with the
exception of -1... and -2... in 633/635
</udc:applicationNote><skos:narrower rdf:resource="http://udcdata.info/rdf/050821"/>
<skos:narrower rdf:resource="http://udcdata.info/rdf/050934"/>
<skos:narrower rdf:resource="http://udcdata.info/rdf/052242"/>
<skos:narrower rdf:resource="http://udcdata.info/rdf/052535"/>
<skos:narrower rdf:resource="http://udcdata.info/rdf/052779"/>
<skos:narrower rdf:resource="http://udcdata.info/rdf/052911"/>
</skos:Concept>
```

# UDC SKOS EXPORT (XML/RDF)



# UDC AND LIBRARY LINKED DATA

- National Szechenyi Library (Hungary)

```
<bibo:Document rdf:about="http://nektar.oszk.hu/resource/manifestation/2645471">  
<dcterms:subject>  
  <rdf:Description> <dcam:memberOf rdf:resource="http://purl.org/dc/terms/UDC"/>  
    <rdf:value>894.511-32</rdf:value>  
  </rdf:Description>  
</dcterms:subject>
```

- Trondheim - Library of Norwegian University of Science And Technology (NTNU) – TEKORD <http://ckan.net/package/tekord>

- UDC numbers can be found in the following LLD bibliographic datasets
  - AGRIS Journals database
  - ISSN Centre

# REALITY – PROBLEMS IN LINKING

## UDC in bibliographic control

- bibliographic records contain UDC notations cancelled 30-50 years ago;
- UDC data in bibliographic records are not accompanied by information about UDC source (edition/provenance);
- UDC is a synthetic classification and bibliographic records may contain complex UDC classmarks that are not present in a standard UDC edition;
- libraries lack resources or expertise to maintain subject authority files which would provide additional semantics and access points to UDC classmarks.

# ON THE OTHER HAND...

We have data

- UDC MRF database and UDC archive contains historical UDC data and record of UDC changes;
- complex UDC classmarks can be automatically parsed with 100% accuracy and there are algorithms/programs ready for use;
- UDC Consortium gave permission to publish the complete UDC data as linked data providing this would not jeopardise sustainability of the scheme (UDC is self-funded).



# THINKING BEHIND UDC LINKED DATA

- libraries should not need to worry about resolving the semantics or parsing the components of UDC codes
- UDC linked data should be supported by a front-end service (number look-up/resolution service) – which would enable parsing, validating and resolving URI for UDC codes
- UDC as linked data, i.e. UDC RDF triple store should contain all data necessary to resolve and interpret strings coming from library catalogues (including historical UDC data)

# LINKED DATA AS A SOLUTION

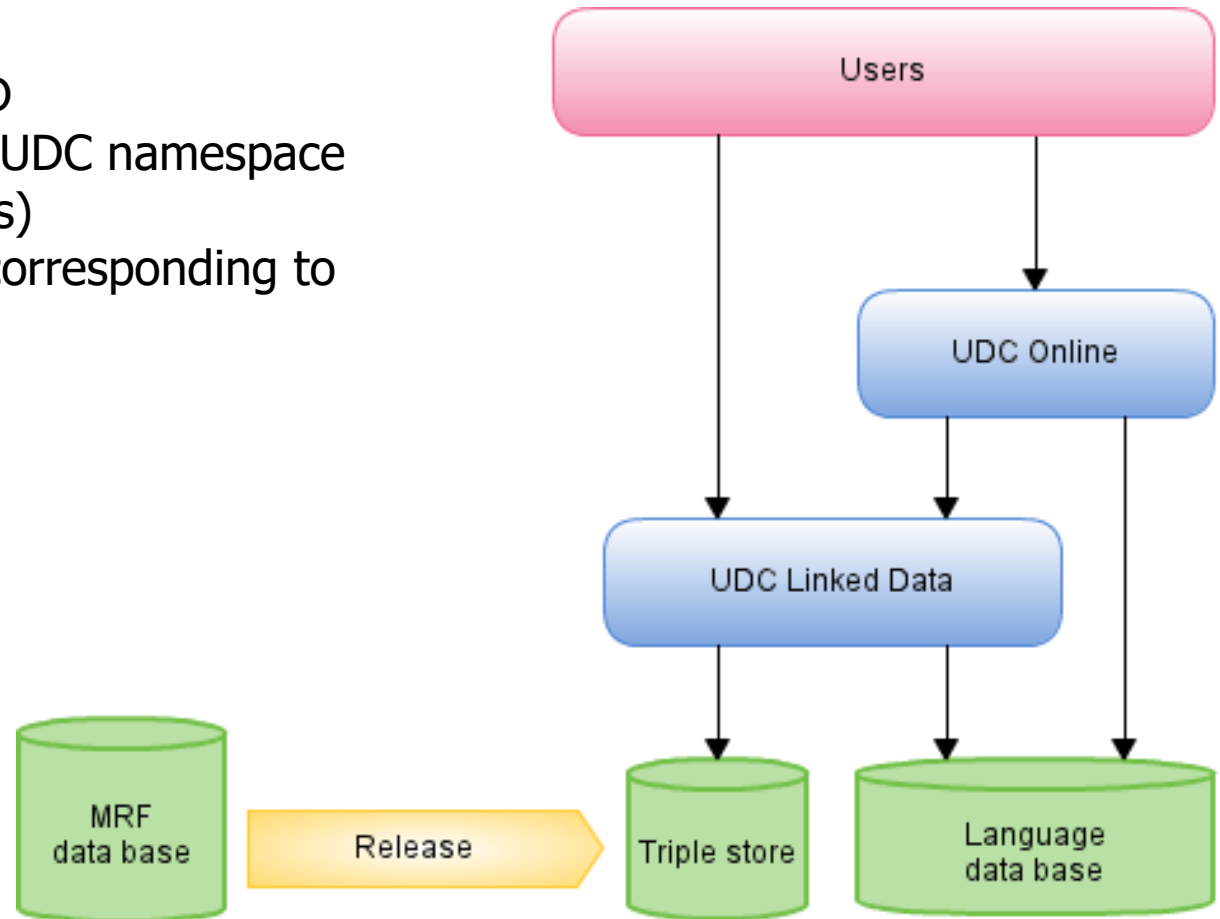
Publishing UDC as linked data and providing a look-up service to be accessed by programmes can resolve the following:

- providing automatic alignment of UDC data contained in library records with the UDC namespace by
  - parsing and validating UDC strings
  - returning correct URIs for each UDC query
  - returning redirection data for cancelled records
  - returning additional semantic and language data
  
- managing access to various level of UDC data (scope, depth, and languages) based on licence scheme

# ARCHITECTURE OF UDC DATA SERVICES

Plan for 2019:

- managing LD & LOD
- managed access to UDC namespace (SPARQLE templates)
- different data sets corresponding to licence rights



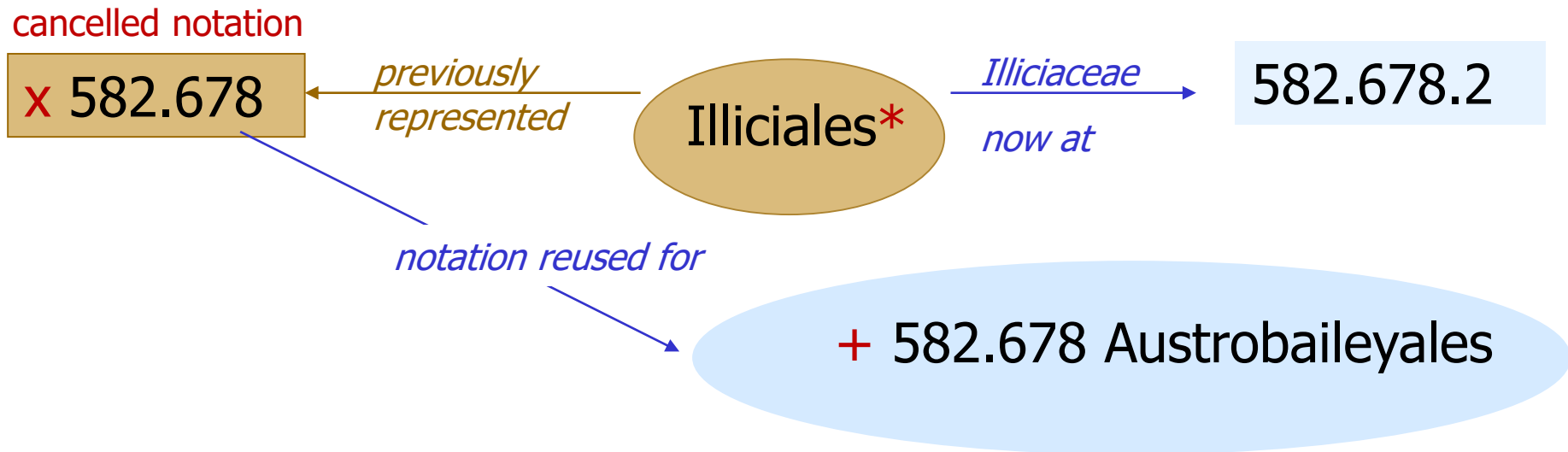
Graph by Christophe Gueret

# WE HAVE DATA – HOW TO PRESENT THEM?

- In the case of UDC in 2019 we are moving on from the previous experimental UDC Linked Data publishing (UDC Summary as linked data). We plan changes in the following:
  - URI – moving towards structured URI expression MRF versions
  - new data element schema
  - adding historical UDC data
  - flexible approach to LD publishing UDC data (licence driven access and “paywall” as well as open linked data access)
- KOS publishing standards (e.g. SKOS) do not have an adequate solution for expressing historical and vocabulary evolution data that would secure semantic alignment between collections and KOS;
- Extending SKOS with DC terms, OWL or LOD2 appears to be less suited than using OWL (and OWL Lite)

# ISSUE OF CLASSIFICATION NOTATION REUSE

## Reuse of notation for different meaning (broadening)



\* an order not recognized by the current systems of plant classification, its families placed in Austrobaileyales and Magnoliales

# NOTATION HISTORY AND CONCEPT HISTORY

- Whenever UDC notation is re-used we record this change as follows

! 582.62 Fagales

Notation History:      Notation previously used for Hamamelididae, now at 582.62/.63.

Concept History:      Fagales were previously at 582.632

Semantics:              The scope of class narrowed

- Data is held in a more structured way in the MRF database



# CANCELLATION DATA

- After 1992 - UDC number may be cancelled but its record and its ID stays permanently in the database
  - **cancellation date** (date of cancellation)
  - **cancellation source** (issue of Extensions & Corrections in which this is published)
  - **replaced by:** ID of the record to which the UDC number is redirected
  - **replacement type**  
controlled list of types, expressing what the cancelled number is replaced with:  
new class, colon combination; combination with common auxiliary;  
combination with special auxiliary; other
  - **replacement (semantic) alignment**  
controlled list: exact match, to broader, to narrower, approximation
- Additional task: incrementally digitize and add from the UDC archive pre-1992 data

# FINAL THOUGHTS

- potential of linked data technology may be more far reaching than originally thought for both libraries and KOS communities
- it can help align library subject data with the latest knowledge organization tools without libraries having to do re-classifications / re-indexing or re-organizations of collections
- it can serve as a vehicle for interpreting, validating and enriching and serving back subject data to libraries
- it can help heritage collections being mined for knowledge via subject data

MANY THANKS



ISKO LOW COUNTRIES

[www.isko-lc.org](http://www.isko-lc.org)

Join us in Brussels, 20-21 June 2019

“Morsels of Knowledge” 1<sup>st</sup> ISKO LC Conference