



Classification and reference vocabulary in Linked Environment Data

Joachim Fock, Maria Rüther, Thomas Schultz-Krutisch,
Federal Environment Agency (UBA), Germany

Thomas Bandholtz,
innoQ Deutschland GmbH

Outline

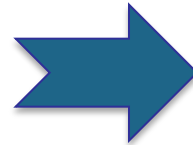
- Classification and thesaurus of the environmental library
- From class assignment to inference
- a break of patterns in 2005
- From library to Linked Environment Data
- Summary and conclusion



Classification and thesaurus

of the library
of the Federal Environment Agency
(Umweltbundesamt, UBA),
Germany

The UBA and its library



Berlin, 22.07.1974 - 2. Mai 2005, then moved to Dessau-Roßlau,



- > 400,000 books
- > 1,000 periodicals
- > 14,000 research reports
- countless press releases
- different media types

Library catalogue

Katalogangaben	
Medienart	Beitrag [Beitrag]
Signatur	UBA UM450123
Autor	<ul style="list-style-type: none"> ▸ Seyrich, Andreas [Groeditzer Umwelt Dienst] ▸ Friedrich, Martina [Groeditzer Umwelt Dienst]
Titel	Errichtung einer Behandlungsanlage fuer oelhaltige Abwaesser. Darstellung des Aufbereitungsbetriebes und Erlaeuterung der Forschungstaetigkeit in der Groeditzer Umwelt Dienst GmbH / Andreas Seyrich ; Martina Friedrich
Körperschaft	<ul style="list-style-type: none"> ▸ Groeditzer Umwelt Dienst [Affiliation]
Umfang / Seiten	3 Abb.
Sprache	Deutsch
Kongress	<ul style="list-style-type: none"> ▸ Innovative Technologien des Umweltschutzes (Deutsch-Internationale Fachtagung anlaesslich des 'Tages der deutschen Umwelttechnik')
In	<ul style="list-style-type: none"> ▸ Innovative Technologien des Umweltschutzes : deutsch-tschechische Fachtagung anlaesslich des 'Tages der deutschen Umwelttechnik' als Teil des Begleitprogrammes der 7. Internationalen Fachmesse fuer Umweltschutz und Umweltgestaltung EnviBrno und der WoodTec 98 ... unter Beteiligung deutscher und tschechischer Unternehmen auf dem Bruenner Messegelaende am 21. Oktober 1998 / W. H. Hauthal [Hrsg.] ; Wilhelm Kulke [Hrsg.]. - Duesseldorf - (1998), 33-40
	UBA UM450123
Autorendeskriptoren	<p>Oelhaltige-Abwasser Oel-Wasser-Gemisch</p>
Deskriptoren	<ul style="list-style-type: none"> ▸ Abwasserbehandlung ▸ Kläranlage ▸ Ultrafiltration ▸ Trennverfahren ▸ Grenzwert ▸ Kohlenwasserstoff ▸ Schadstoffelimination ▸ Abwasserreinigung ▸ Reinigungsverfahren ▸ Verfahrenstechnik ▸ Membranverfahren ▸ Abwassereinleitung ▸ Aufbereitungsanlage ▸ Öl ▸ Abwasser
Klassifikation	<ul style="list-style-type: none"> ▸ WA52 Wasser: Abwasserbehandlung, Abwasservermeidung, Abwasserverwertung

Classification

- + SR Strahlung
- + UA Allgemeine und übergreifende Umweltfragen
- + UR Umweltrecht
- + UW Umweltökonomie
- WA Wasser und Gewässer
 - + WA1 Wasser: Belastungen (Einwirkungen) durch Entnahme, V
 - + WA2 Wasser: Auswirkungen von Wasserbelastungen
 - WA3 Wasser: Methodische Aspekte der Informationsgewinnung (physikalisch, biologisch)
 - WA4 Wasser: Qualitätsfragen (Gütekriterien, Richt- und Grenzwerte)
 - WA5 Wasser: Vermeidung, Minderung oder Beseitigung von Belastungen (Gewässerschutz), Abwasserbehandlung und -entsorgung
 - WA51 Wasser: Aufbereitung
 - WA52 Wasser: Abwasserbehandlung, Abwasservermeidung, Abwasserverwertung
 - WA53 Wasser: Schutz und Sanierung oberirdischer Binnengewässer
 - WA54 Wasser: Meeresgewässerschutz
 - WA55 Wasser: Schutz und Sanierung des unterirdischen Wassers
 - WA6 Wasser: Planungsverfahren und -vorschriften der Wasserwirtschaft

- 258 classes
- 14 root nodes
- three levels max
- notation:

└─┬─┘			
two-letter root code		one digit level 2	one digit level 3
- German/English

Thesaurus UMTHES®

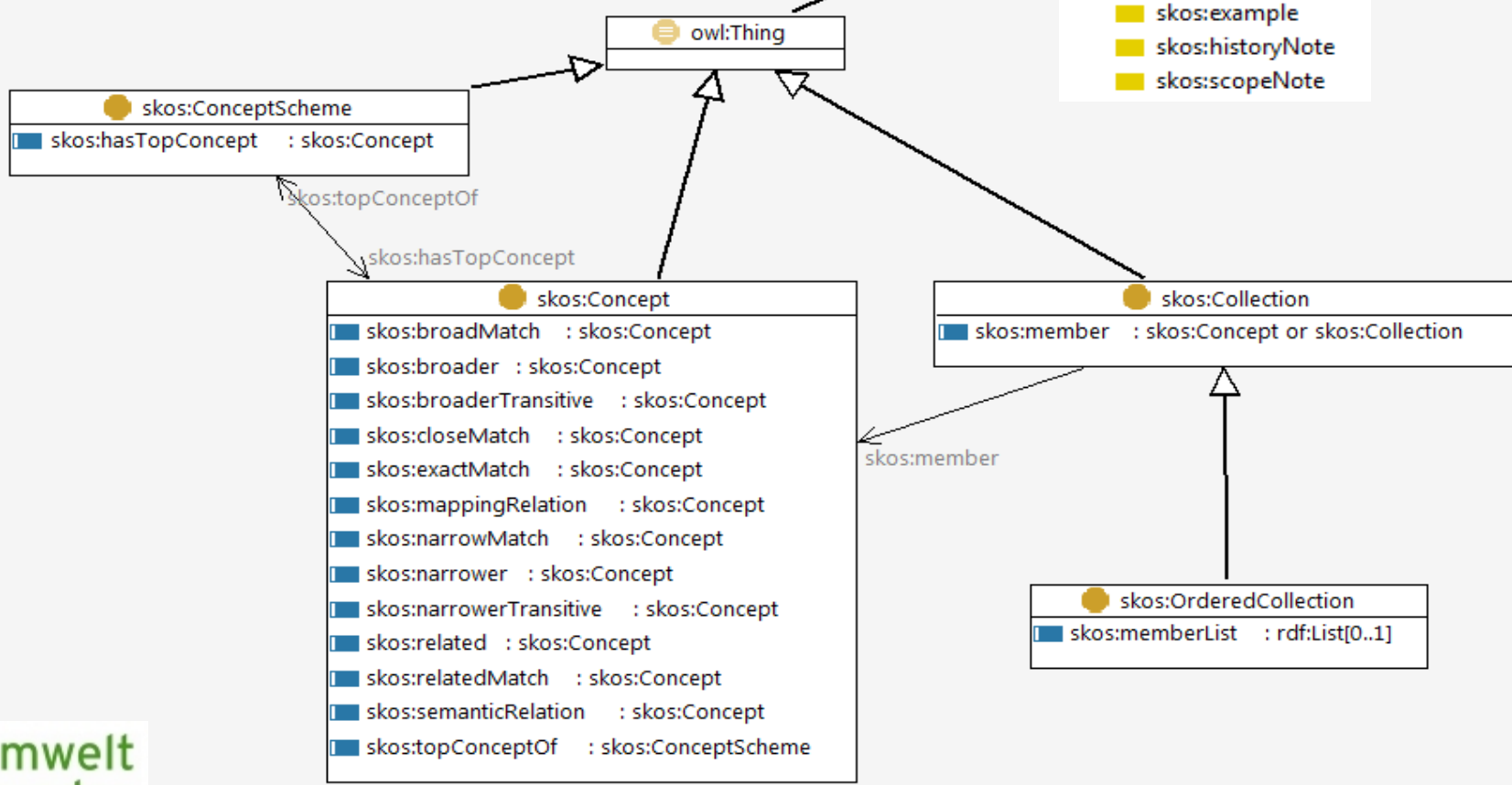
- 10,000 descriptors
- 25,000 non-descriptors
- Poly-hierarchy
- Inflectional forms of terms
- Term compositions
- Homograph qualifiers
- German centric with many English translations



SKOS model

any resource can have these labels (~terms) and notes

- rdfs:Resource
- skos:altLabel : rdfs:Literal
- skos:hiddenLabel : rdfs:Literal
- skos:prefLabel : rdfs:Literal
- skos:note
 - skos:changeNote
 - skos:definition
 - skos:editorialNote
 - skos:example
 - skos:historyNote
 - skos:scopeNote



SKOS representation (1)

Concept

```
:_00000940 a skos:Concept;  
    skos:prefLabel "Abwasser"@de  
    skosxl:prefLabel :Abwasser;  
    skos:altLabel :wasteWater;@en  
    skosxl:altLabel :wasteWater;  
    (...)  
    skos:definition "Durch häuslichen,..."@de;  
    ext:memberOf coll:WA52;  
    ext:memberOf coll:WA77;  
    skos:narrower :_00022045;  
    skos:related :_00028954;  
    skos:broader :_00041343;  
    (...)  
    skos:closeMatch gemet:9127.
```

SKOS representation (2)

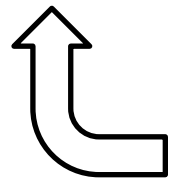
Labels and label relations

```
:Abwasser a skosxl:Label;  
    skosxl:literalForm "Abwasser"@de;  
    ext:inflectional "Abwassern";  
    ext:inflectional "Abwassers";  
    ext:inflectional "Abwässer".  
  
:wasteWater a skosxl:Label;  
    skosxl:literalForm "waste water"@en;  
    ext:lexicalVariant "wastewater";  
    ext:compoundFrom (:waste :water).  
  
:waste a skosxl:Label;  
    skosxl:literalForm "waste"@en.  
  
:water a skosxl:Label;  
    skosxl:literalForm "water"@en;  
    ext:inflectional "waters".
```

SKOS representation (3)

The classification as a set of nested collections.

```
coll:WA52 a skos:Collection;  
    skos:prefLabel "WA52 Wasser: Abwasserbehandlung, ..."@de;  
    skos:altLabel "WA52 Waste water treatment, ..."@en;  
    skos:definition "Maßnahmen zur Behandlung ..."@de;  
    skos:member :_00000940;  
    (...)  
ext:memberOf coll:WA5 .
```



broader class reference

iQvoc – open source tool

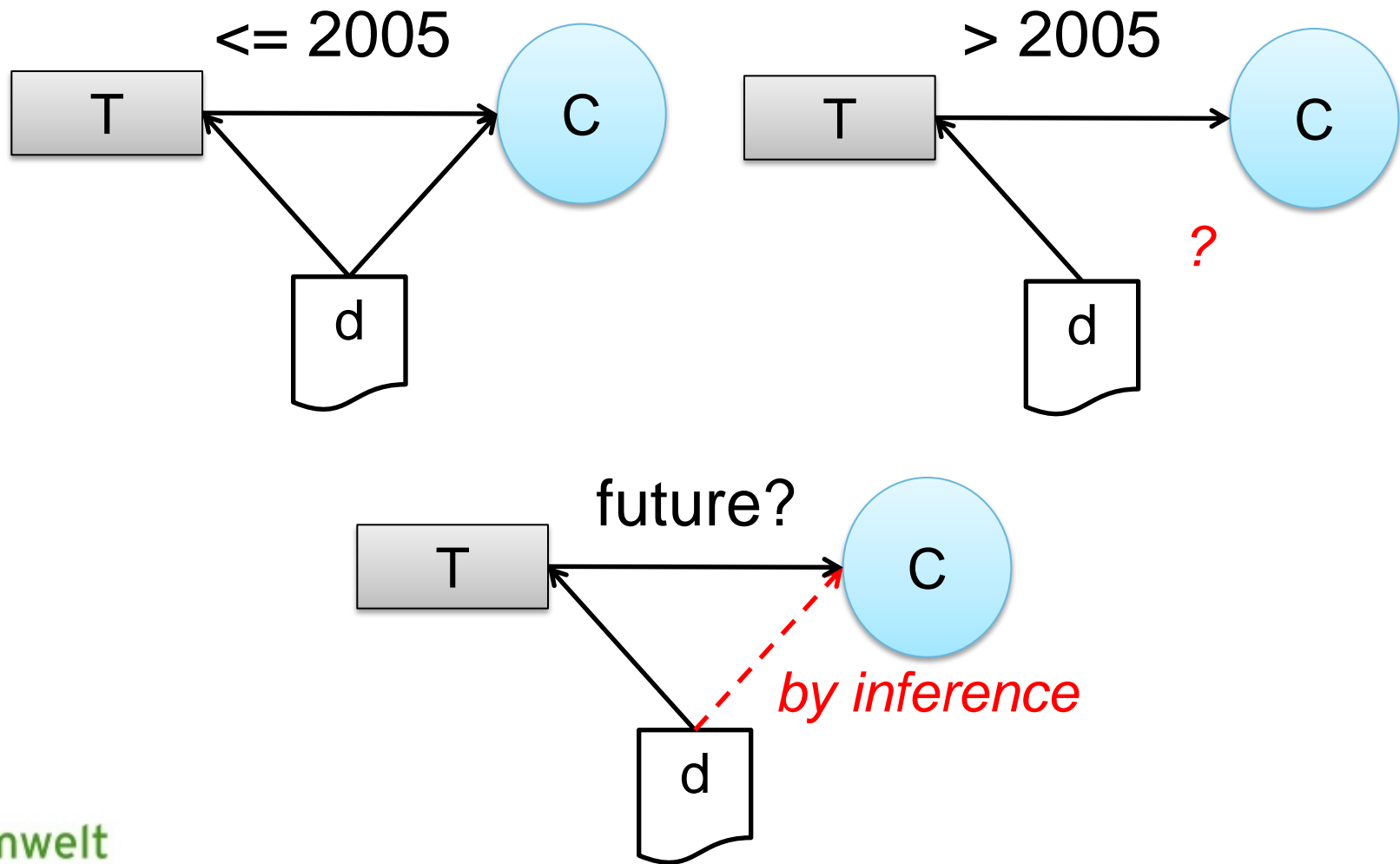
- **managing vocabularies** (classifications, thesauri, etc.) on the Web
 - *intuitive user interface*
 - *Linked Data compliant*
 - *import/export of existing SKOS vocabularies*
 - *editorial control and workflow*
- **full SKOS support:**
 - *multiple languages, both user interface and vocabulary*
 - *nested collections*
 - *notes and annotations*
 - *mapping properties*
 - *SKOS-XL extension for labels*
- licensed under the Apache 2.0 **open source** license.
- UMTHES live: <http://iqvoc-umt.innoq.com>
- get iQvoc: <https://github.com/innoq/iqvoc/wiki>



From class assignment to inference

A break of patterns in 2005

Triade documents, thesaurus and classification



Property chain

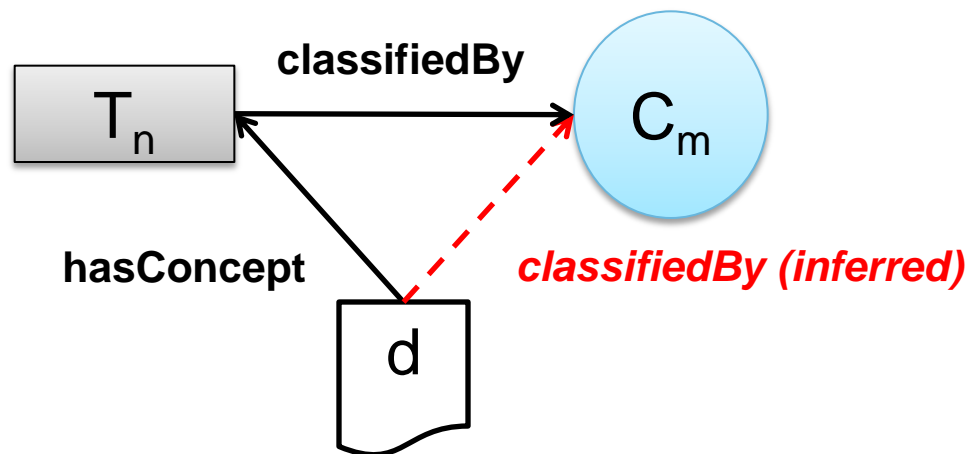
Given the RDF/OWL representation of the library catalogue and the thesaurus, a simple property chain (OWL2) may be interpreted by a reasoner to detect implicit class assignments of documents.

A common property chain example is:

:hasUncle owl:propertyChainAxiom (:hasParent :hasBrother) .

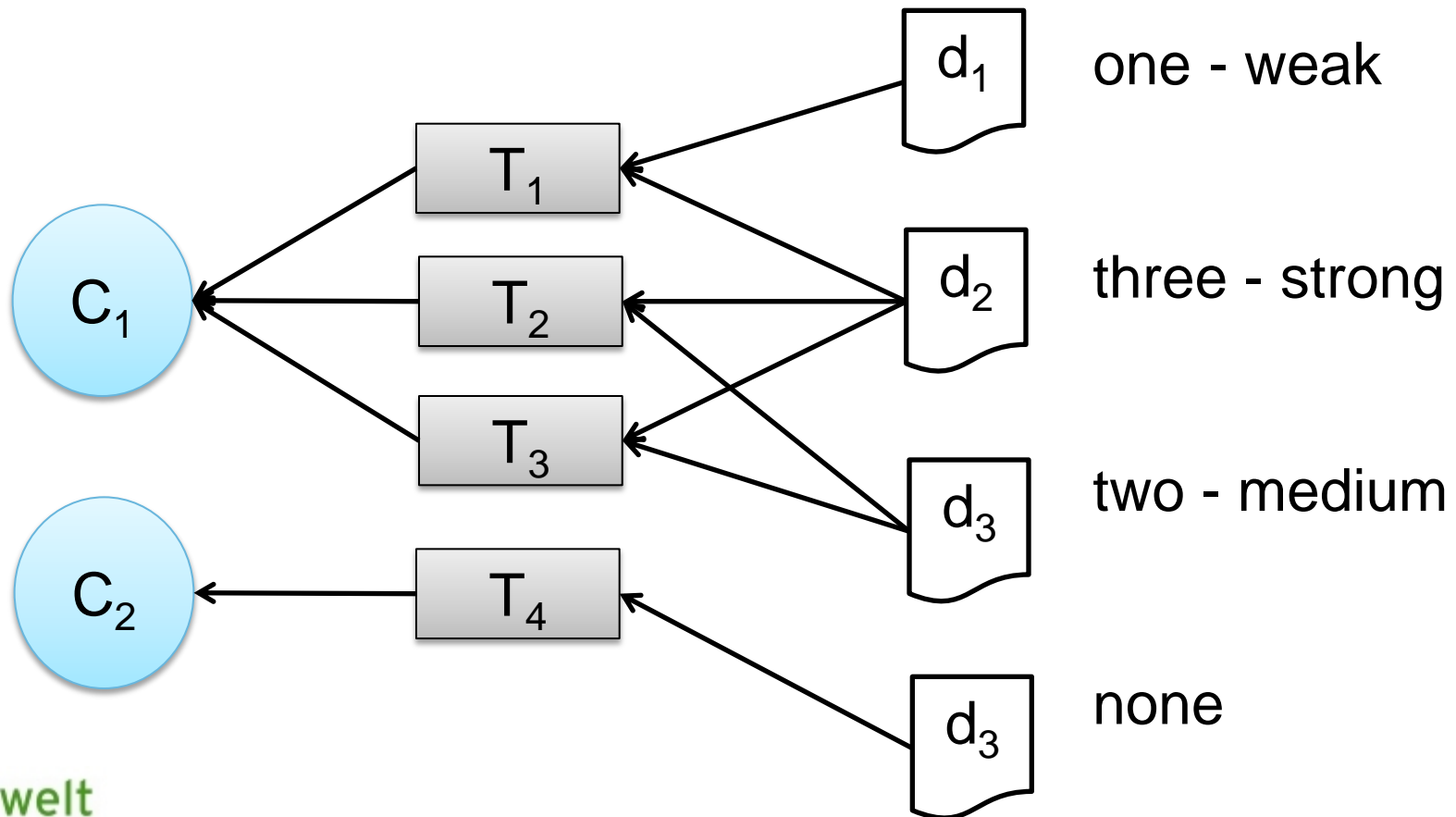
In our case we may add a similar statement:

:DOC4711 :classifiedBy owl:propertyChainAxiom (:hasConcept :isMemberOf) .



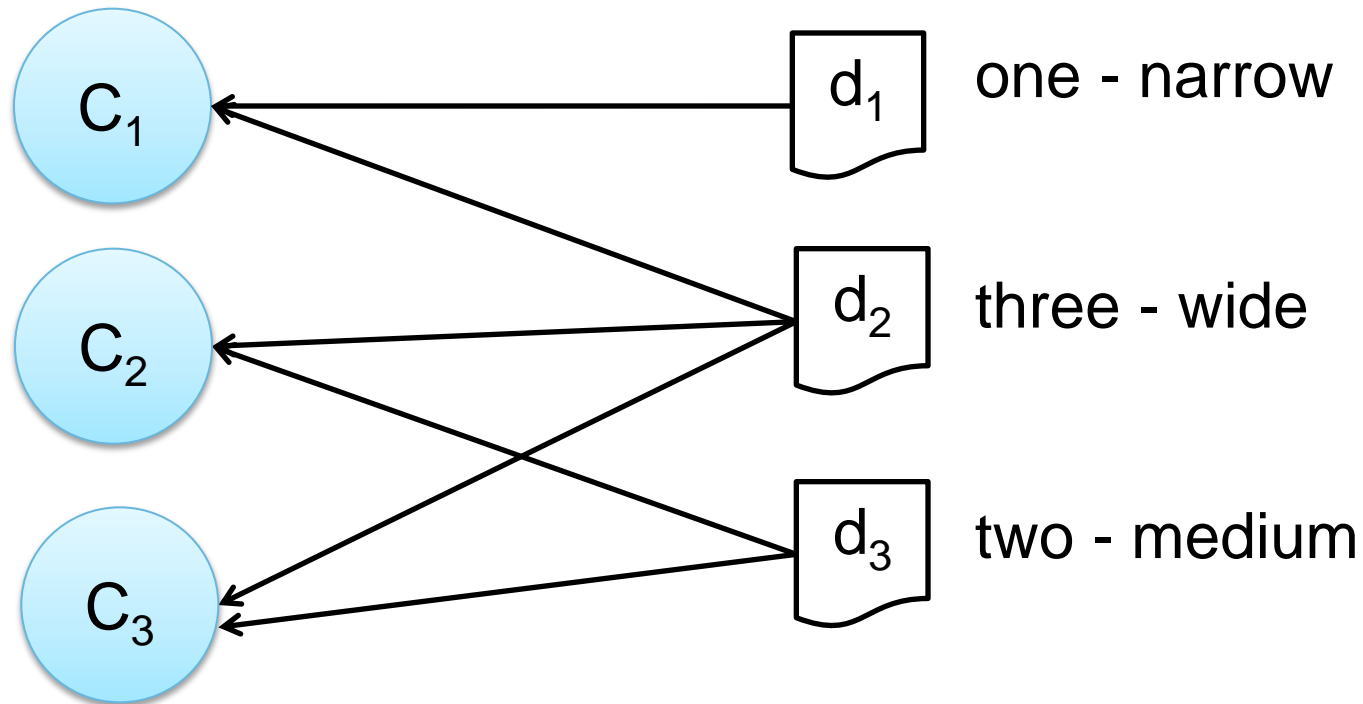
Significance by descriptors

How many descriptors T_n of document d_m refer to the same class?



Significance by focus

How many classes C_n are referred by the same document d_m



Example

classification & indexing
of a document

- manually assigned classe vs.
- classes inferred by descriptors

△ = both methods

○ = different levels of
hierarchy

title out of OPAC (literature)	Dioxins and dioxin-like PCBs in different fish from the river Elbe and its tributaries, Germany / B. Stachel, ..., C. Schroeter-Kermani						
descriptors with classes	2,3,7,8-Tetrachlorodibenzo-p-dioxin	CH7	CH1				
	eel	NL72	WA25	CH7	NL2		
	biological tissue			CH1	NL2		
	biphenyl			CH7	CH1		
	bream		WA72	NL2	CH1		
	fish	NL72	WA72	NL2	LF7		
	river	WA7	NL7				
	golden orfe	WA25	CH2	WA3	CH1		
	flood	WA6	WA75	NL1			
	muscle		CH1				
	polychlorinated dibenzofuran			CH7	CH1		
	polychlorinated biphenyls			CH7	CH1		
	flooding	WA6	NL1	BO1			
	polychlorinated dibenzodioxin			CH7	CH1		
	pollutant						
	load	WA1	BO1	LU1	WA2	BO2	LU2
	Statistical evaluation			NL3	WA3	BO3	LU3
classification of literature	<p>Water pollution: Effects on surface waters – WA21</p> <p>Origin, behaviour, distribution and fate of substances in the environment – CH1</p> <p>Impact of burdens on nature, landscape and landscape elements – NL2</p> <p>Water pollution: Effects on aquatic plants, animals and microorganisms – WA25</p>						

Example

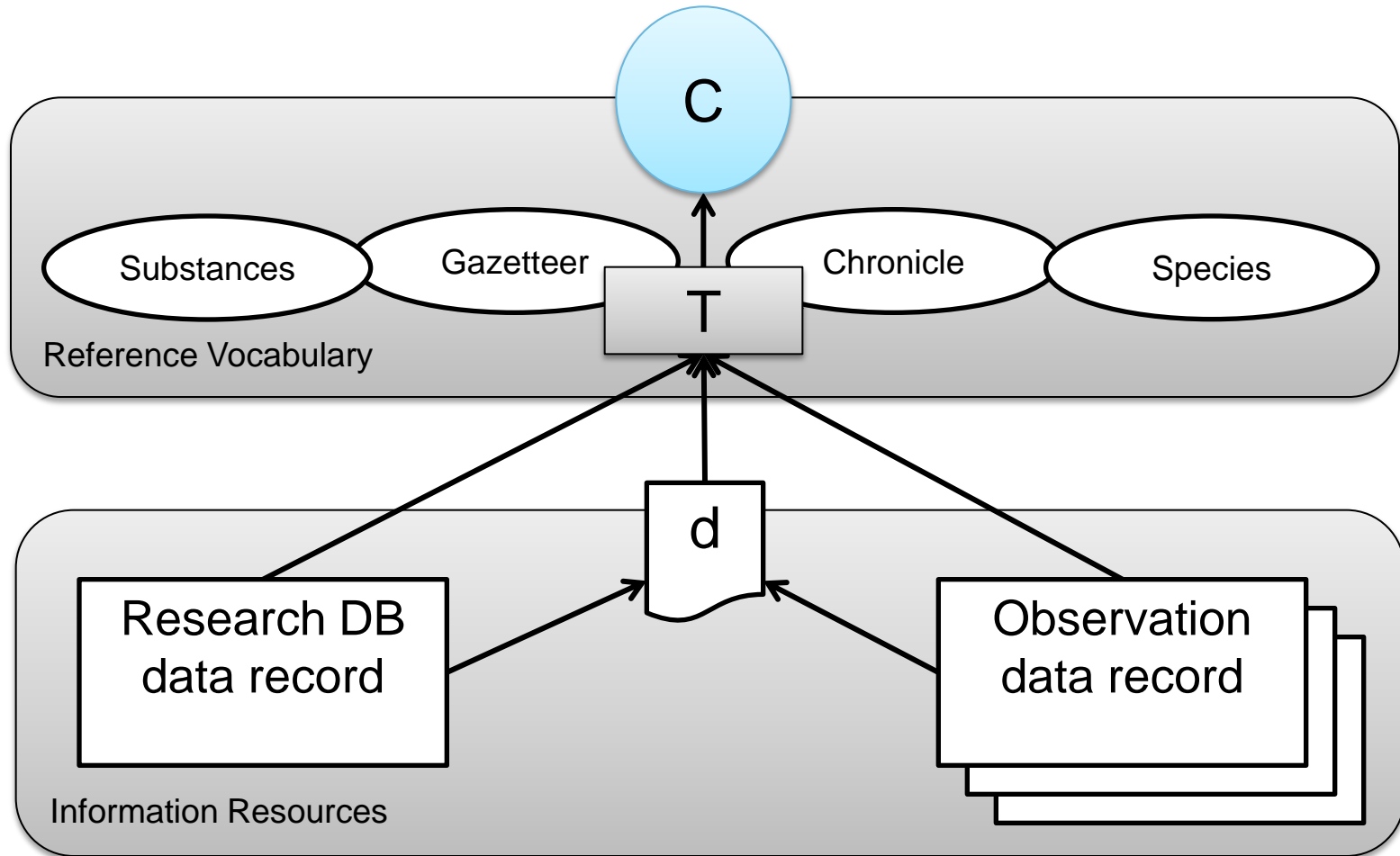
Class	Significance by descriptors	focus 4 manually	inferred	
			focus 10 frequency aware (>1)	focus 4 frequency aware (>2)
CH1	XXXXXXXXXX	X	X	X
CH7	XXXXXX		X	X
NL2	XXXX	X	X	X
WA72	XXX		X	X
BO1	XX		X	
NL1	XX		X	
NL72	XX		X	
WA25	XX	X	X	-
WA3	XX		X	
WA6	XX		X	
BO2	X			
CH2	X			
LF7	X			
LU1	X			
LU2	X			
LU3	X			
NL3	X			
NL7	X			
WA1	X			
WA2	X			
WA7	X			
WA75	X			
WA21		X	-	-



From library to LED

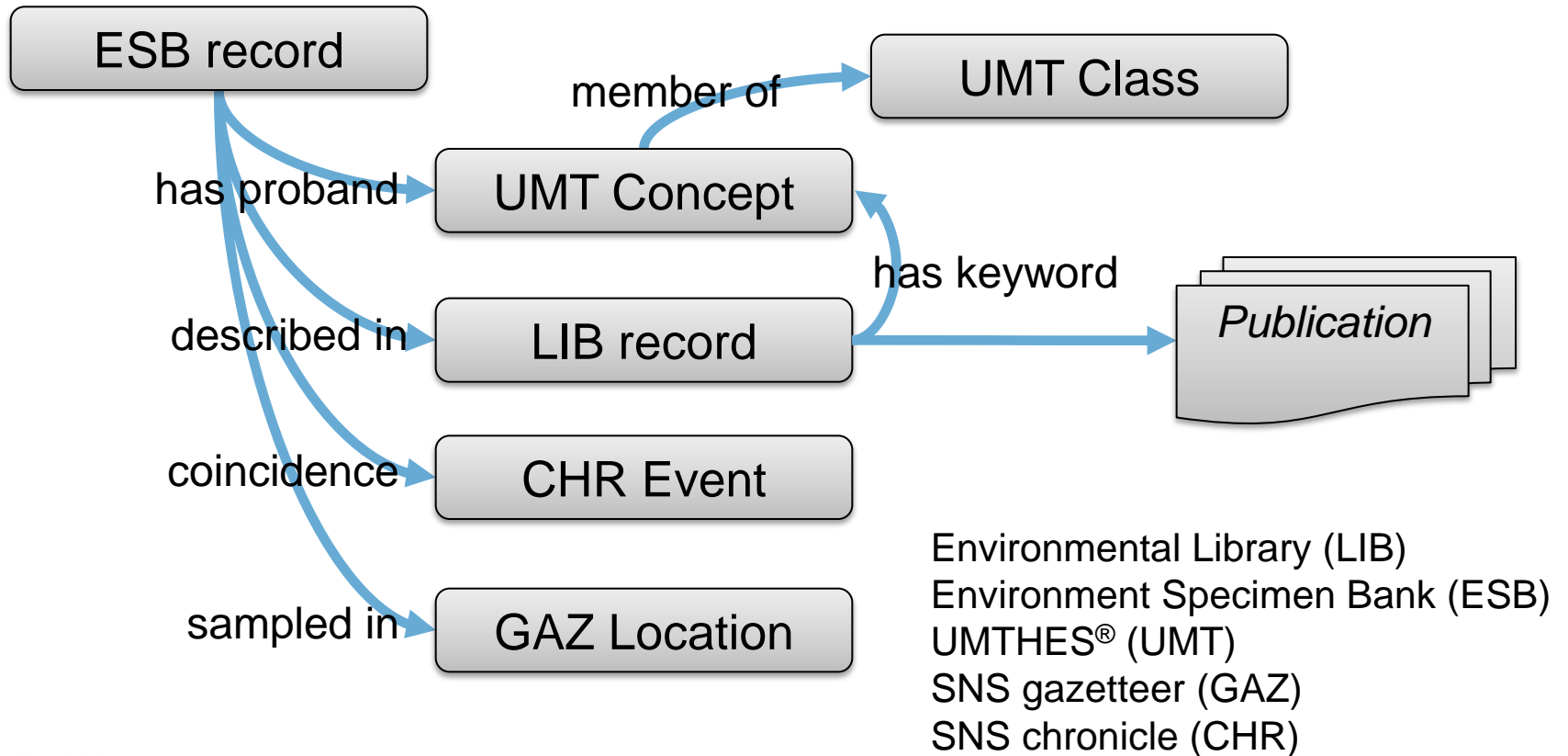
Linked Environment Data

Linked Environment Data



Library Linked Data use case

Crosslinking Environment Data and the Library





Summary and conclusion

Classes and concepts are complementary,
but sometimes data refer to concepts only ...

Classified information spaces

When you have

- a classified thesaurus and
 - data records referring to concepts (“descriptors”)
- classifying these records by inference might be a solution.

RDF representations enable

- crosslinking vocabularies, bibliographic records and scientific data on the Web,
- property chaining by standard reasoners, but this needs further refinement,
- finally, classified information spaces.



Thank you for listening!

Joachim Fock

joachim.fock@uba.de

Maria Rüther

maria.ruether@uba.de

Thomas Schultz-Krutisch

thomas.schultz-krutisch@uba.de

Federal Environment Agency

(UBA), Germany

<http://www.umweltbundesamt.de>

Phone: +49 340 2103 0

Thomas Bandholtz

thomas.bandholtz@innoq.com

innoQ Deutschland GmbH

<http://www.innoq.com>

Phone: +49 178 404 9387