

INSEMPTIVES

Incentives for Semantics

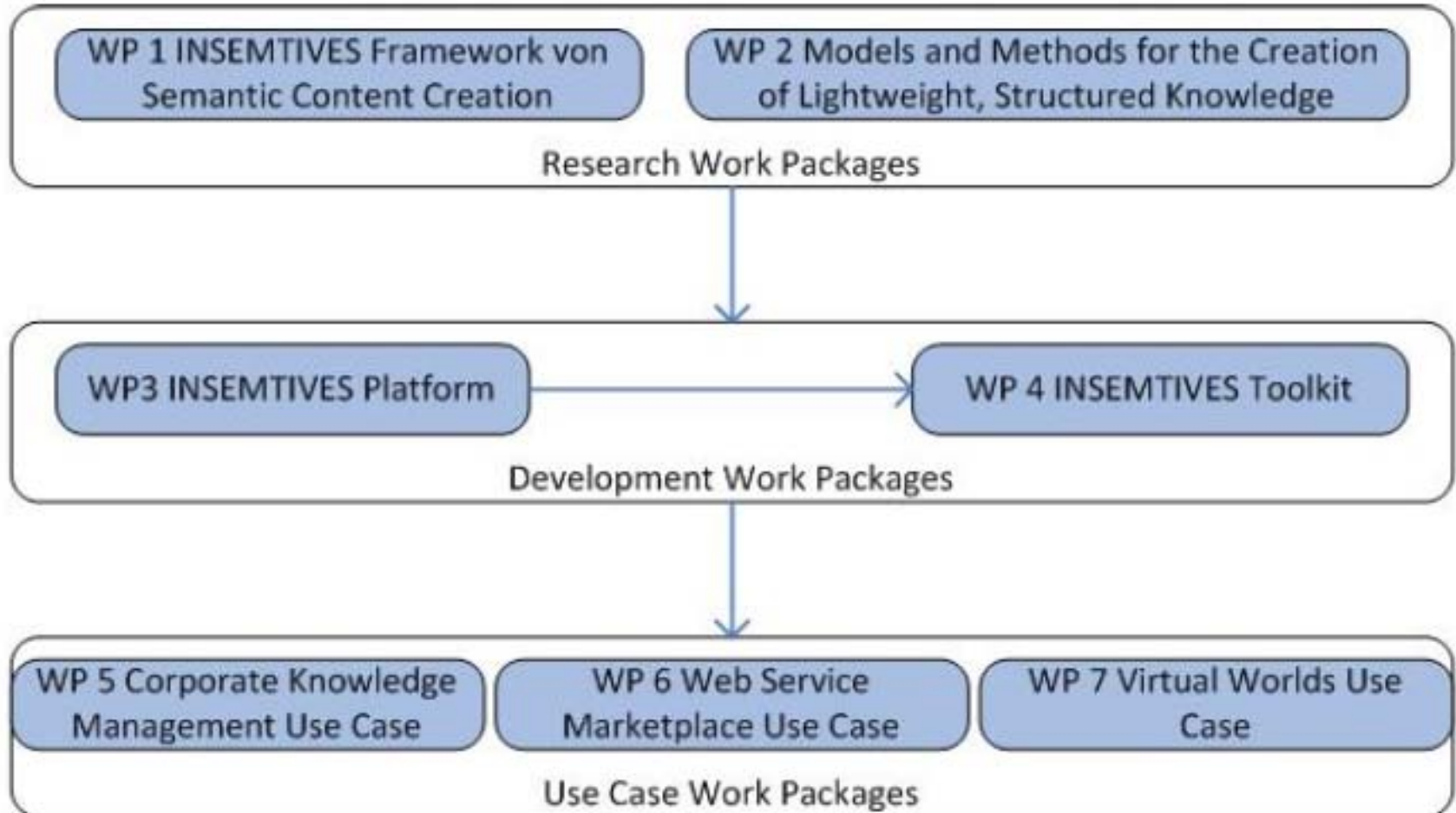
WP 2 - Models and Methods for the Creation and Usage of Lightweight, Structured Knowledge

Pierre Andrews, Ilya Zaihrayeu
UNITN





Work package dependencies





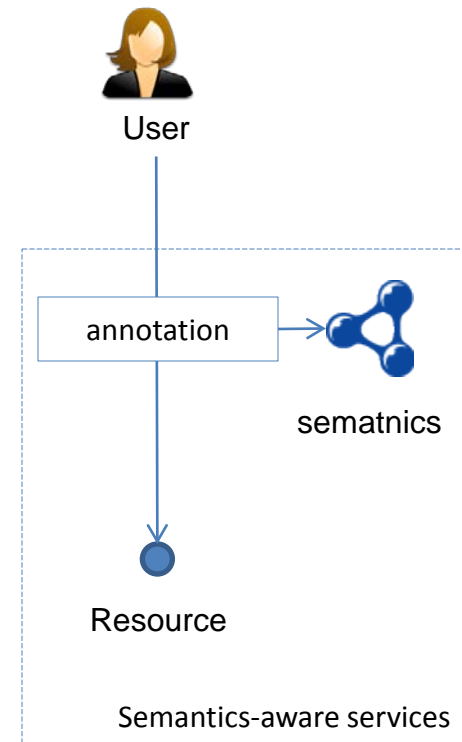
Motivation

- The current Web 2.0 annotation model lacks formal semantics and, therefore, suffers from several shortcomings, e.g.:
 - Searching for “images” should (not) return resources annotated with “picture” (synonymy problem)
 - Searching for “java” (books) should (not) return resources annotated with “java” (drink) (polysemy problem)
 - Searching for “animals” should (not) return resources annotated with “dogs” (specificity gap problem)
- These problems have a negative effect on the QoS for the end user (e.g., correctness, completeness)
- An annotation model with formal semantics can address these (and other) problems and enable new services



Aims and outcomes

- Key problem: how to enrich the Web 2.0 user-annotation-resource model with semantics and semantics-aware services that an ordinary user can comprehend and make use of?
- Models for annotations:
 - Structural complexity (e.g., tags vs. attributes)
 - Vocabulary support (e.g., free tags vs. controlled tags)
 - Collaboration level (e.g., single user vs. shared vocabularies)
- Services:
 - Annotation bootstrapping (help the user at the initial phase)
 - Vocabulary evolution (help users “talk” the same annotation language)
 - Annotation evolution (keep annotations synced with the vocabulary)
 - Semantic search
 - Annotation linkage (enable cross-platform services)



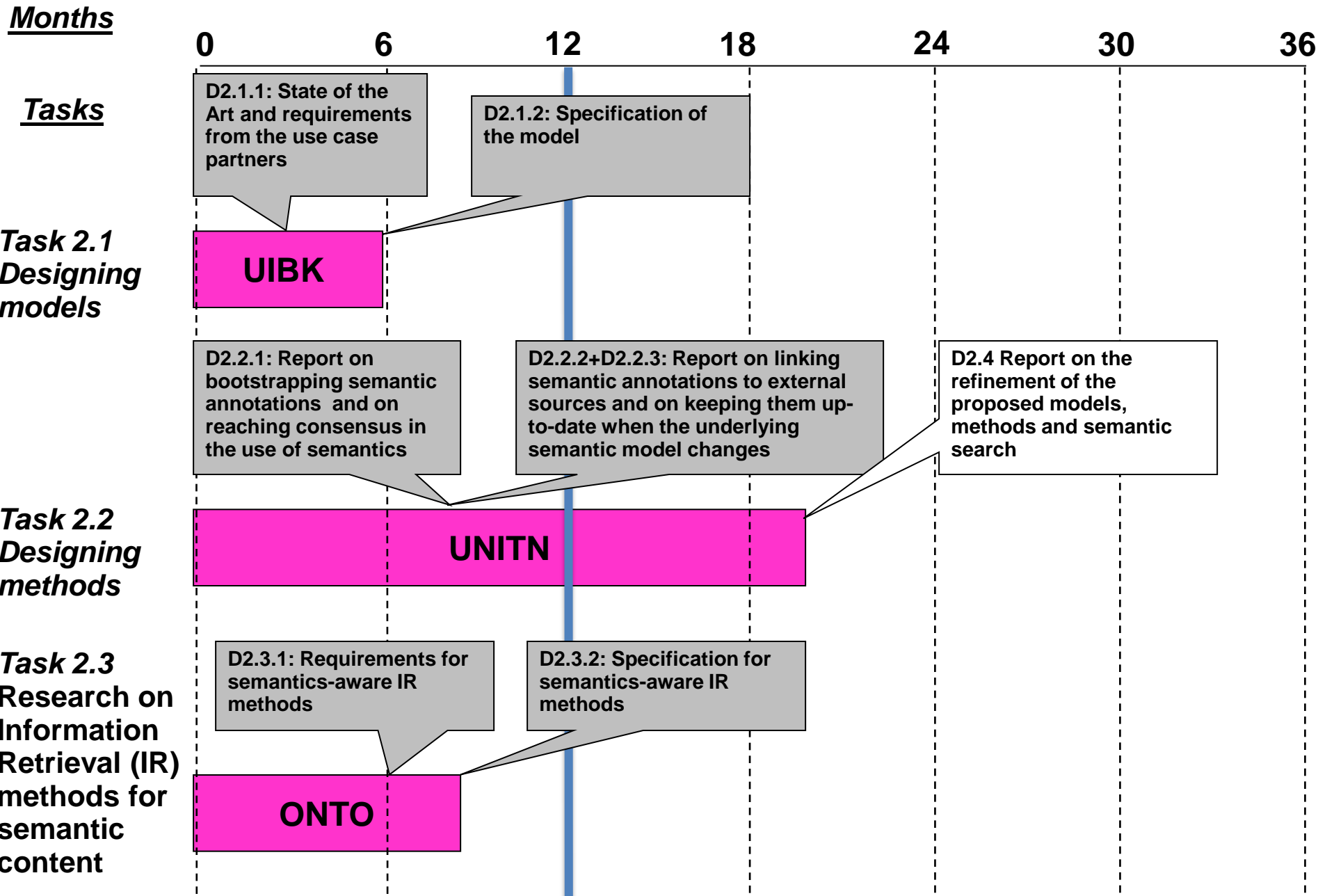


Research – key challenges

- What is the **right** level of complexity of semantics to let the **ordinary user** generate semantic annotations and to provide the user with useful new services?
- How to (semi-)automatically extract annotations from user generated contents and link them to the underlying model?
- How to support the users in (re)using the same semantics for annotations?
- How to keep semantic annotations up-to-date when the underlying semantic model changes?



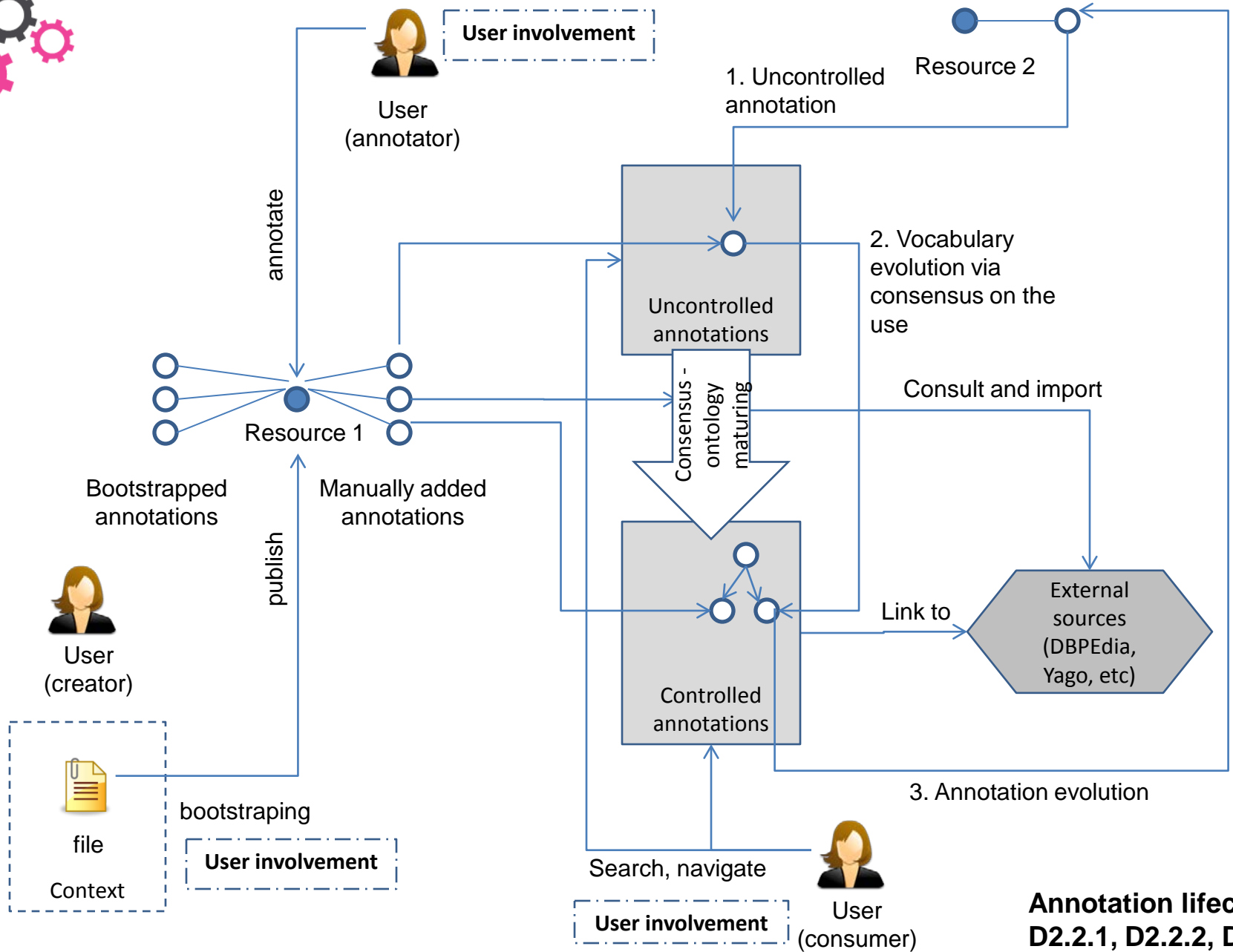
WP 2 TIMELINE AND DELIVERABLES





Outcomes

- Semantic annotation model
- Annotation bootstrapping algorithm
- Consensus reaching algorithm
- Algorithm for supporting the evolution in time of annotations
- Algorithms for semantic search and faceted navigation
- Algorithms for linking semantic annotations to external sources



Annotation lifecycle
D2.2.1, D2.2.2, D2.2.3



Model

Flickr:

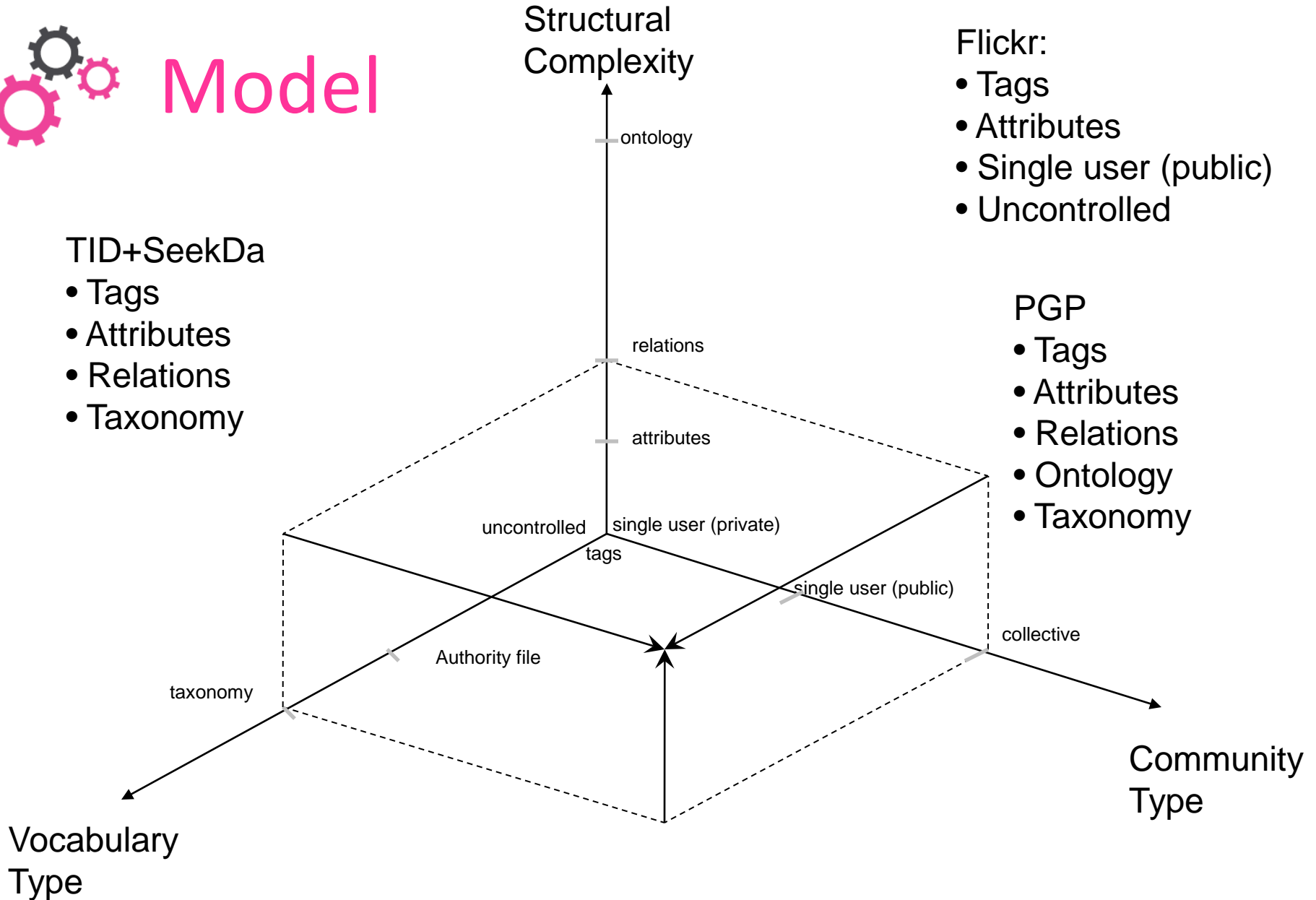
- Tags
- Attributes
- Single user (public)
- Uncontrolled

TID+SeekDa

- Tags
- Attributes
- Relations
- Taxonomy

PGP

- Tags
- Attributes
- Relations
- Ontology
- Taxonomy



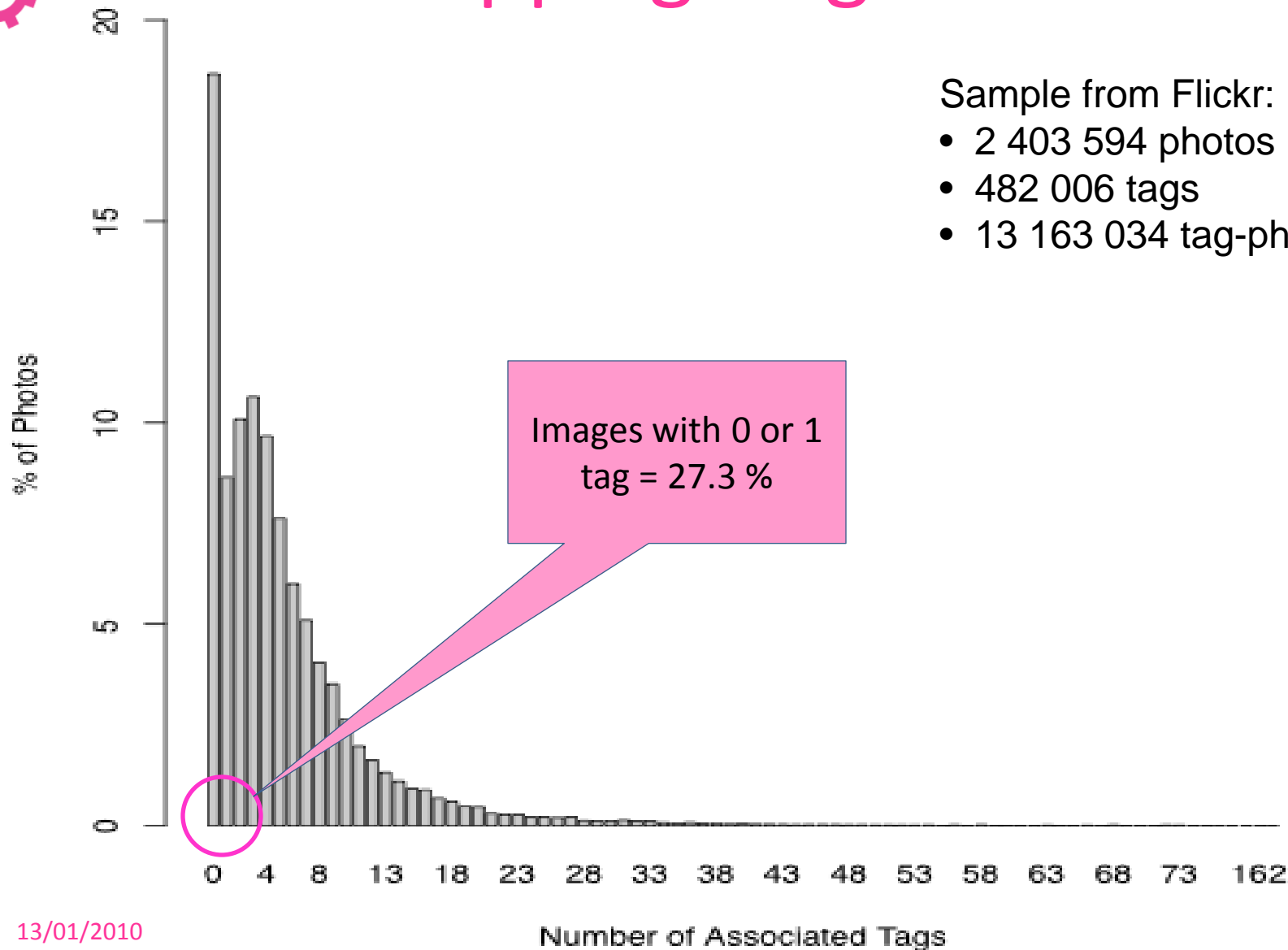


Bootstrapping: motivation

- Data on the web grows very fast:
 - 161 exabytes (108 TB) of information was created or replicated worldwide in 2006
 - 6X growth is predicted by 2010
- The largest source of data is:
 - user generated content with 4+ billion devices – cameras, phones, PCs, CCTVs – mostly multimedia!
 - will increase 50% by 2010
- However, at the publishing time, metadata encoded in the local context in which the multimedia items resided is lost



Bootstrapping: Tag Use at Flickr

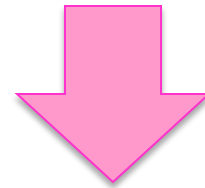


- Sample from Flickr:
- 2 403 594 photos
 - 482 006 tags
 - 13 163 034 tag-photo pairs



Bootstrapping Algorithm

Photos/conferences/Brain and Computer Science/java



Tokenisation
Lemmatisation
Multi Word Detection

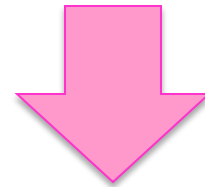
photo

conference

brain science

computer science

java



Word Sense Disambiguation

conference#1

Brain science#1

Java#1 (programming)

WordNet Knowledge Base

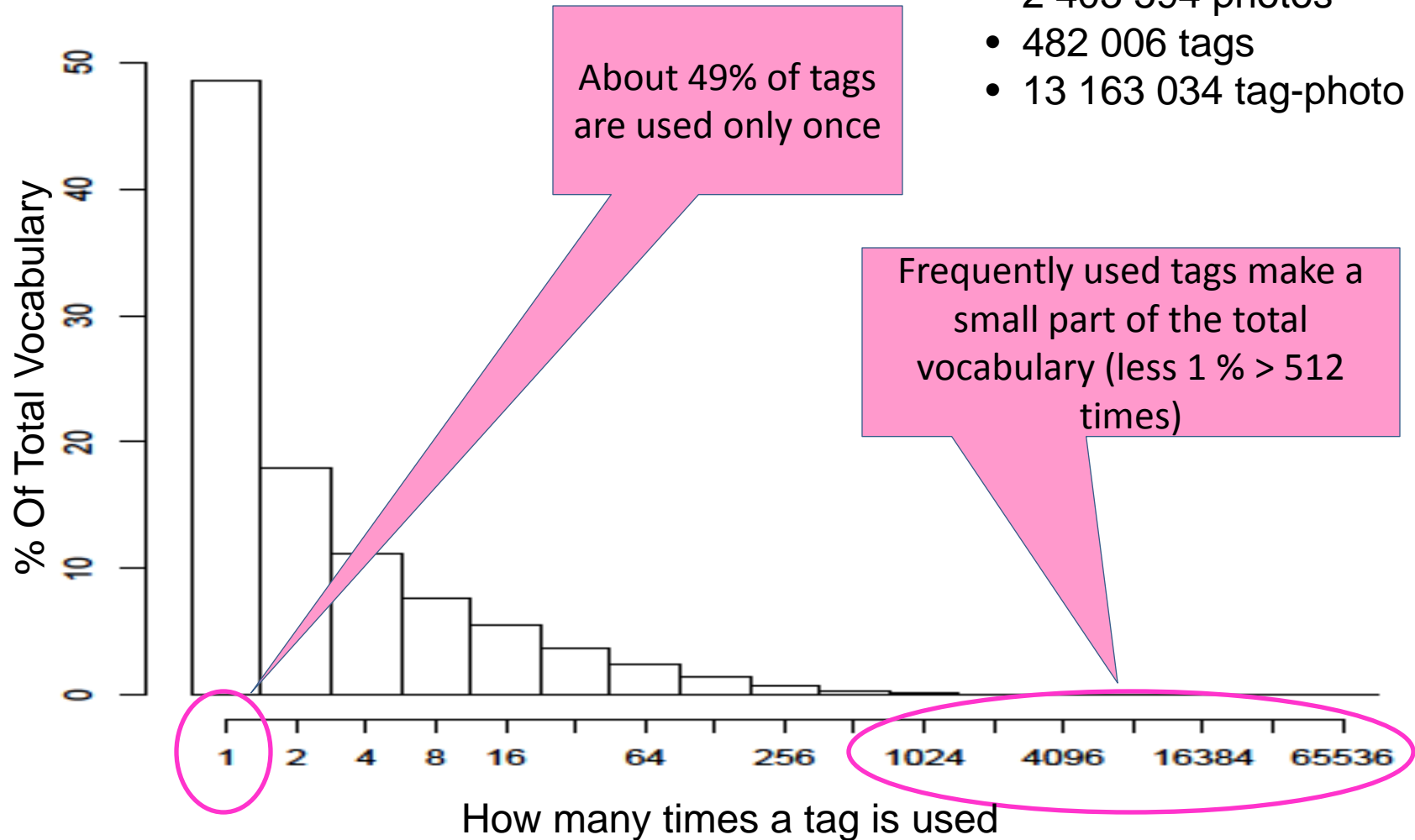
Computer science#1

Reaching Consensus: Tag reuse in Flickr



Sample from Flickr:

- 2 403 594 photos
- 482 006 tags
- 13 163 034 tag-photo pairs





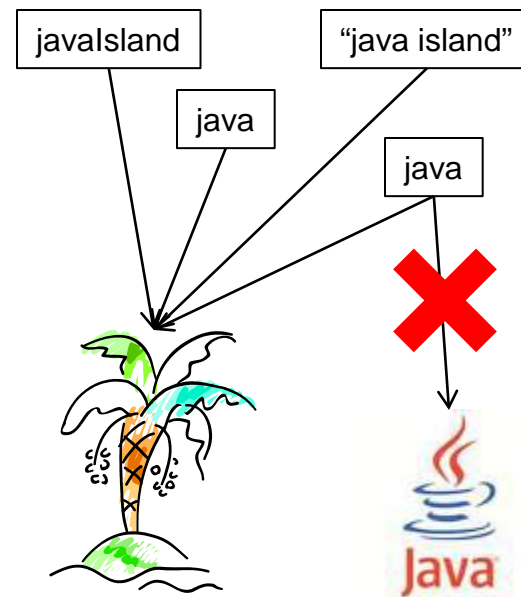
Consensus through User Interaction

- Tag Recommendation

Sea#water mass
Java#island
Island#land mass
Java#coffee
Java#language

Jav

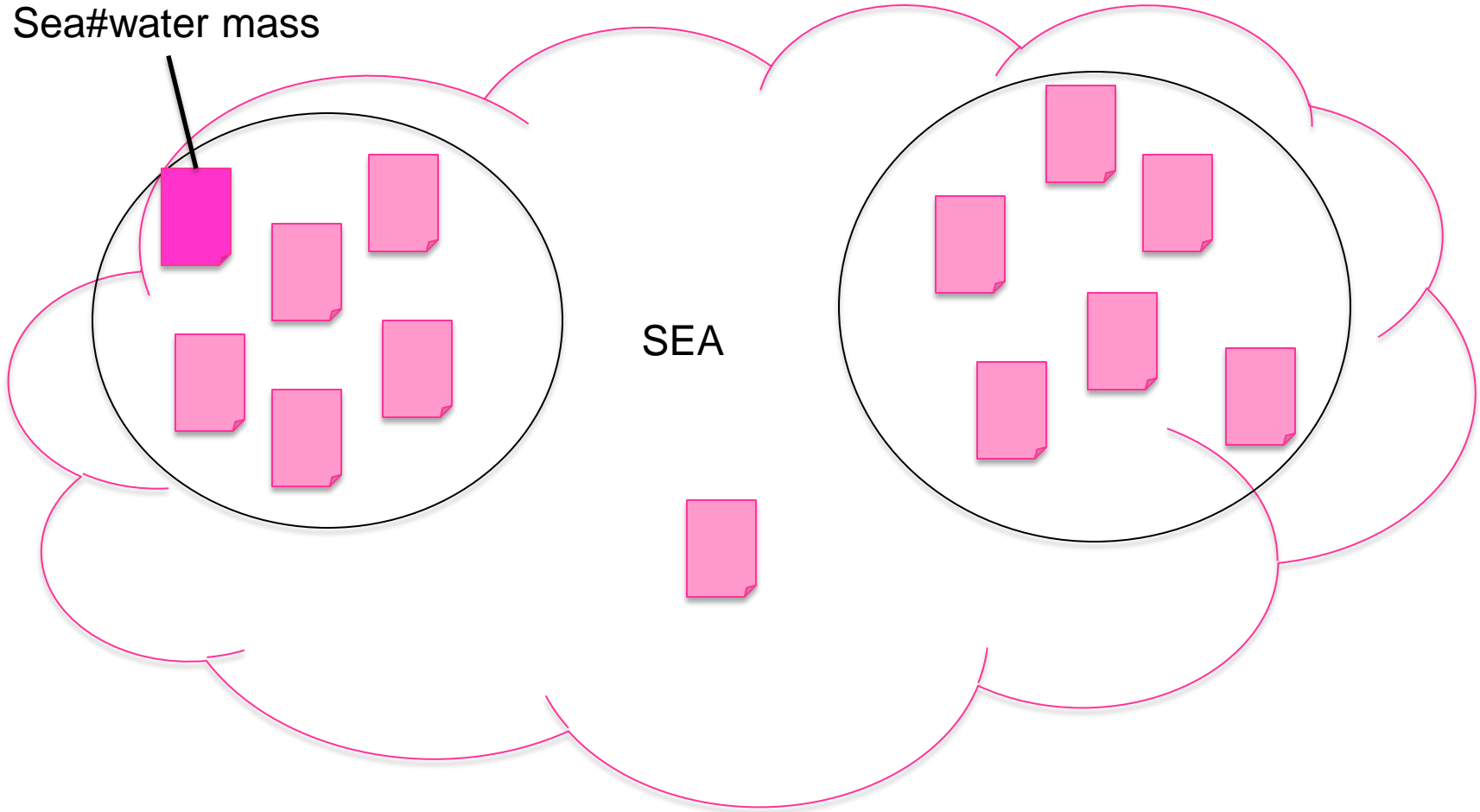
- Java#island
- java#coffee
- Javanese#citizen
-
-





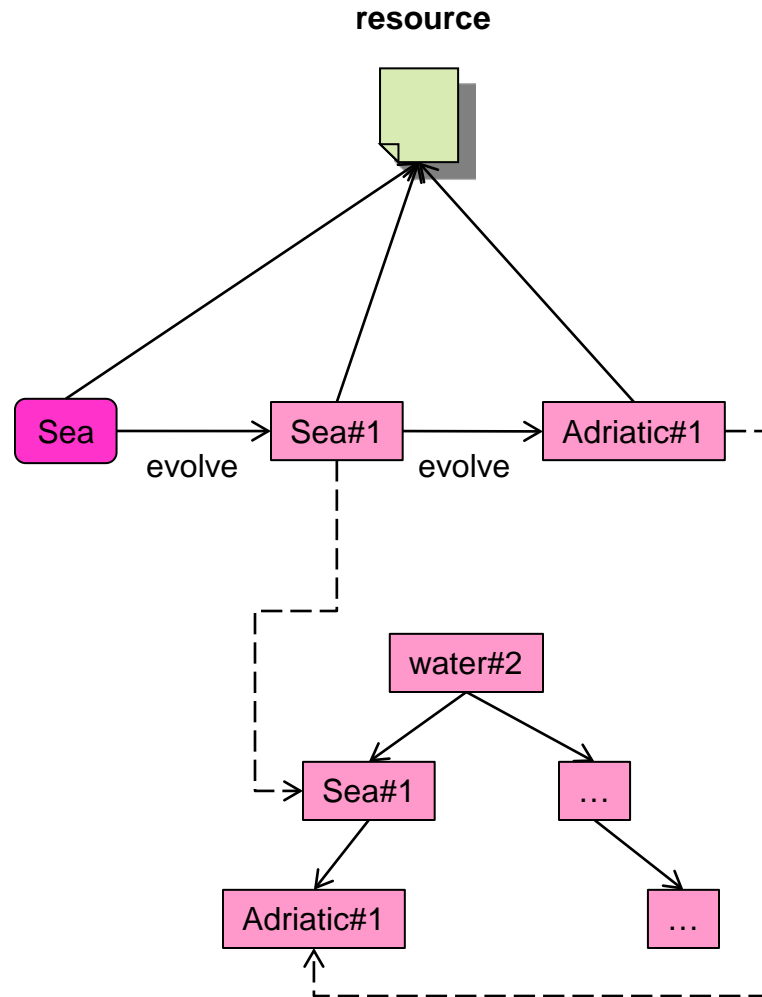
Consensus through clustering

Sea#water mass





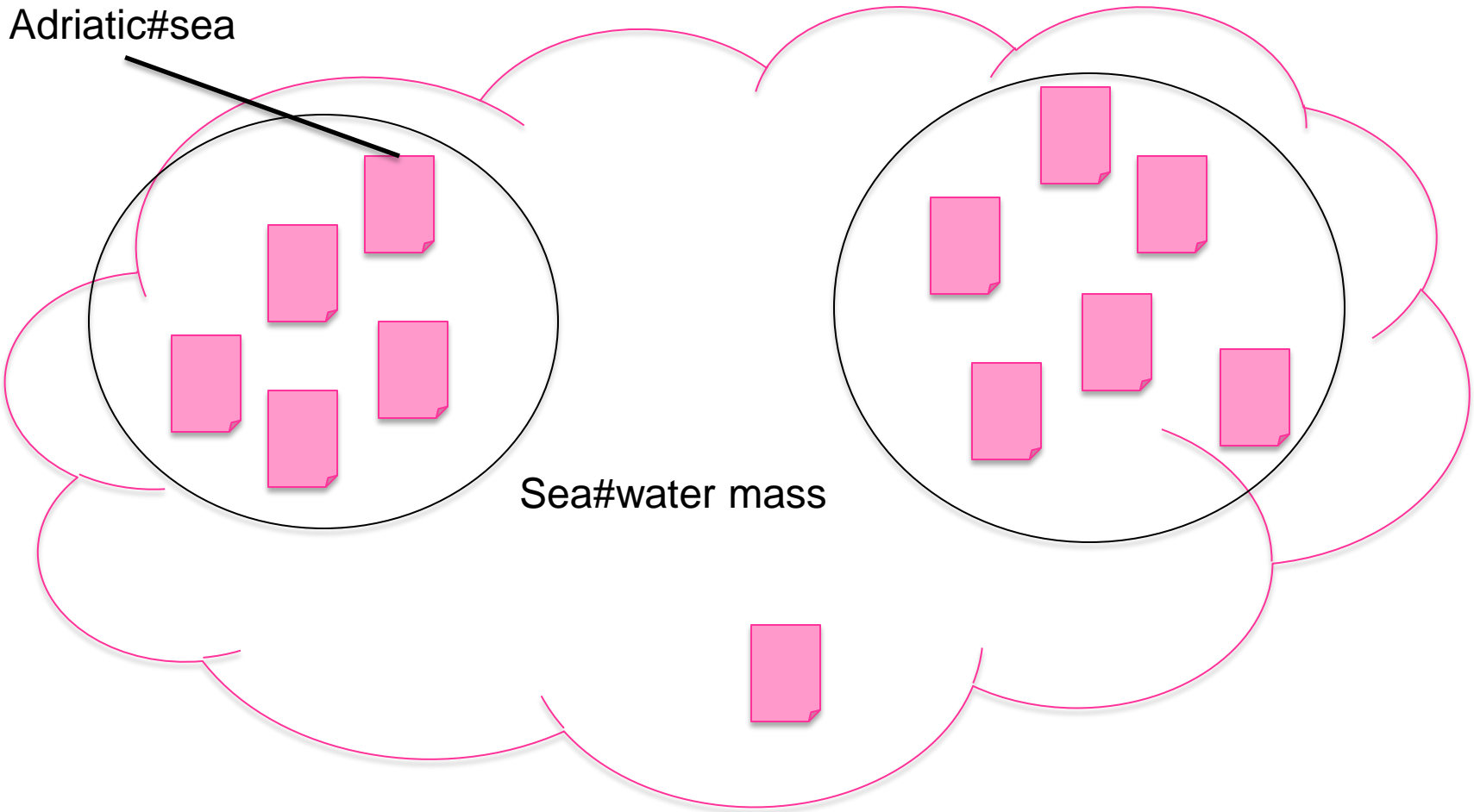
Evolution in Time





Evolution in time through clustering

Adriatic#sea





Outlook

- Further specification and refinement of models, algorithms, and semantic IR methods (to be reported in D2.4 due on M20)
- Know-how transfer and integration with WP3 and WP4 as well as with use case WPs
- Work on the implementation of models and algorithms (part of WP3)

INSEMTIVES

Incentives for Semantics

www.insemtives.eu

<http://blog.insemtives.eu>





Requirements from the Use Case Partners

Partner	Structural Complexity				Vocabulary Type		User Involvement		
	Tags	Attributes	Relations	Ontologies	Free	Taxonomy	Vocabulary construction	Read	Write
TID	X	X	X		X	X	Experts + Users	controlled	Controlled
SEEKDA	X	X	X		X	X	Experts	Controlled	Controlled
PGP	X	X	X	X		X	Experts	Everyone	Everyone
Generic	X	X	X			X	Experts	Everyone	Everyone



Flickr as an Example: GeoTags

Proportion of Geotagged photos uploaded
per week

