Stefan Schulz

Medical University of Graz (Austria)

purl.org/steschu

Fifth Annual Workshop of the Clinical and Translational Science Ontology Group
Amherst, NY, Sep 7 – 8, 2016

Keynote address:

Coding clinical narratives: Causes and cures for inter-expert disagreements

Purpose of the talk

- To report on empirical studies that scrutinized clinical terminologies / ontologies for EHR interoperability (in a European context)
- To expose typical examples and analyze reasons for disagreement between manual annotations with SNOMED CT
- To discuss how and whether ontology can support interoperability and mitigate the effects of inter-coder disagreement
- To defend empirical methods to guide terminology / ontology engineering

Benchmarking ontologies in action

KR-MED 2006

International Workshop - November 8, 2006 in Baltimore, MD, USA

"Biomedical Ontology in Action"

AMIA

- home
- call for papers
- program
- important dates
- committees
- location

Biomedical Ontology in Action

November 8, 2006, Baltimore, Maryland, USA

Workshop organized by the National Center for Ontology Research (NCOR) and the Working Group on Formal (Bio-)Medical Knowledge Representation of the American Medical Informatics Association (AMIA), co-sponsored by the AMIA Formal (Bio)Medical Knowledge Representation Working Group.

Collocated with FOIS 2006

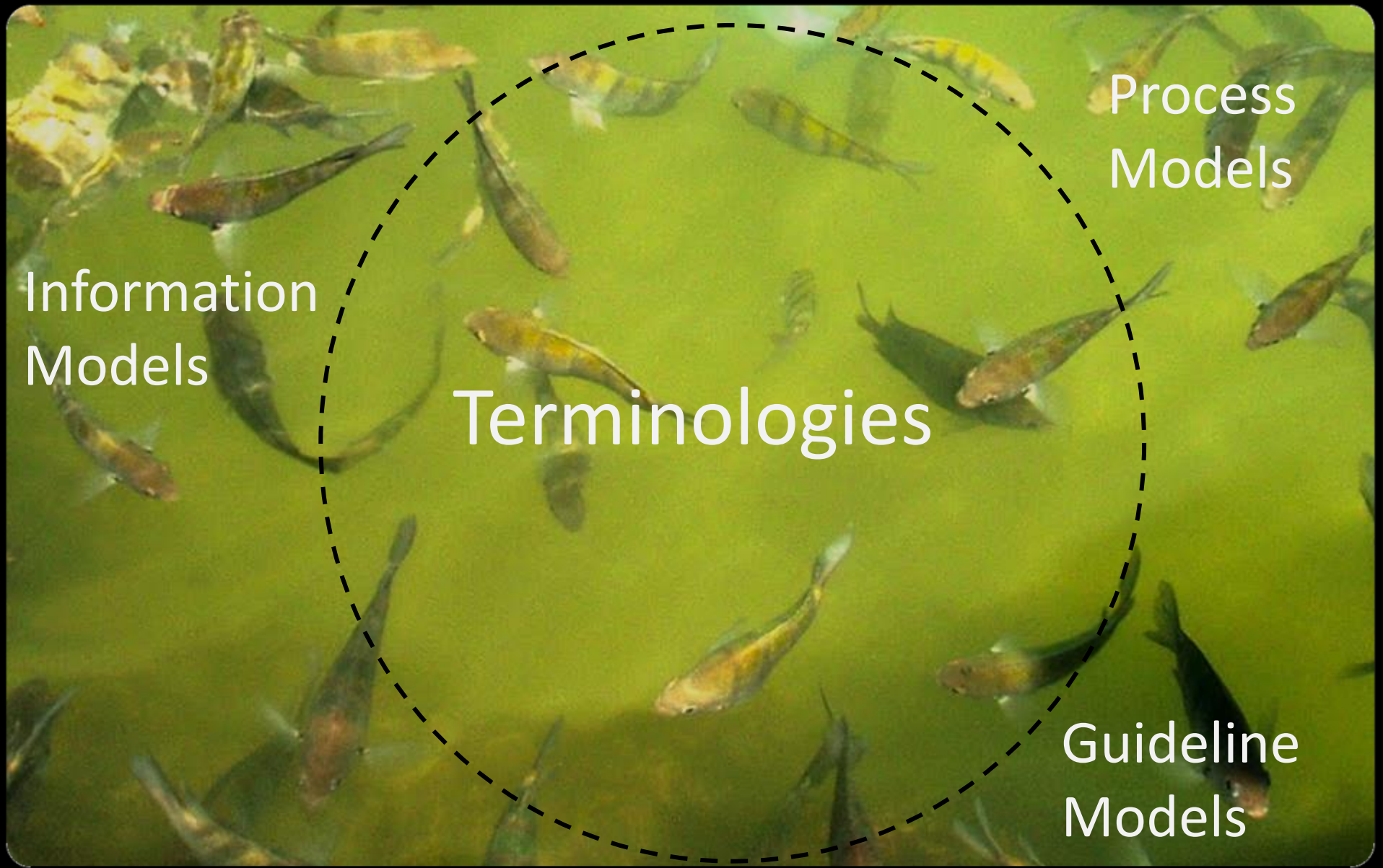
Context of study: ASSESS-CT

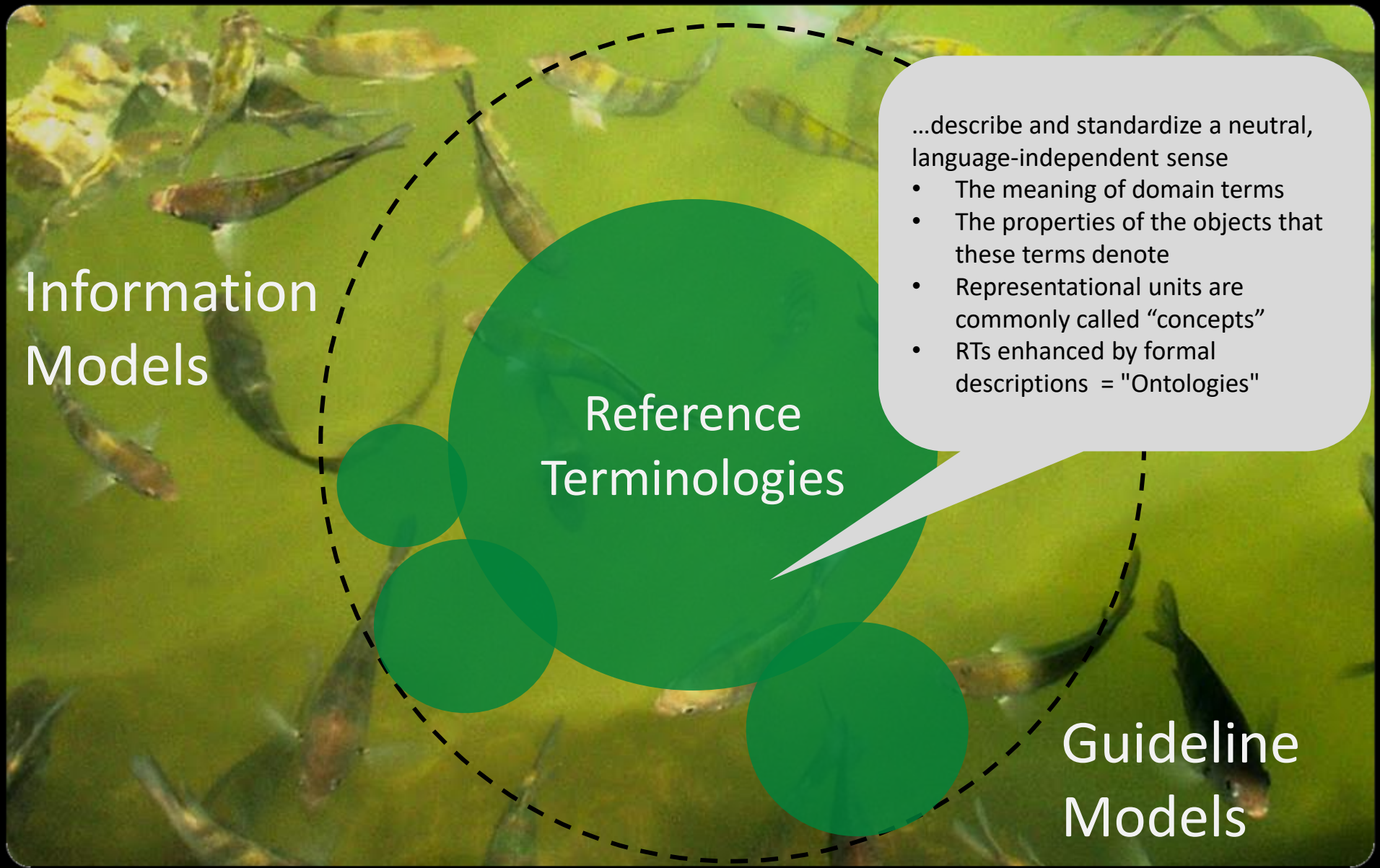
- European project on the fitness of purpose of SNOMED CT as a core reference terminology for the EU: www.assess-ct.eu
- Feb 2015 – Jul 2016
- Scrutinising clinical, technical, financial, and organisational aspects of reference terminology introduction
- Main recommendations:
  - "SNOMED CT is the best candidate for a core reference terminology for cross-border, national and regional eHealth deployments in Europe."
  - Must be part of an ecosystem of semantic assets

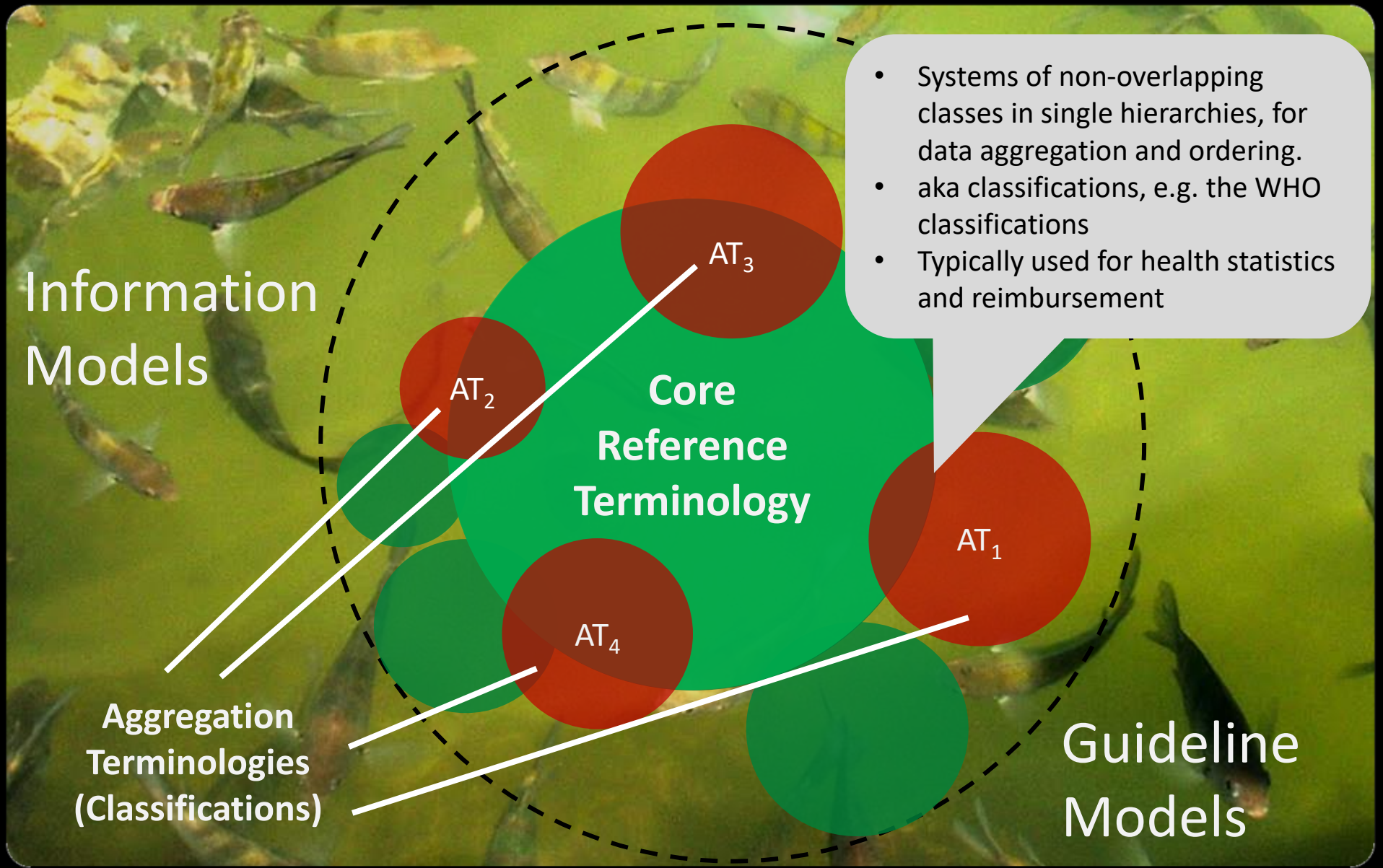
Context of study: ASSESS-CT

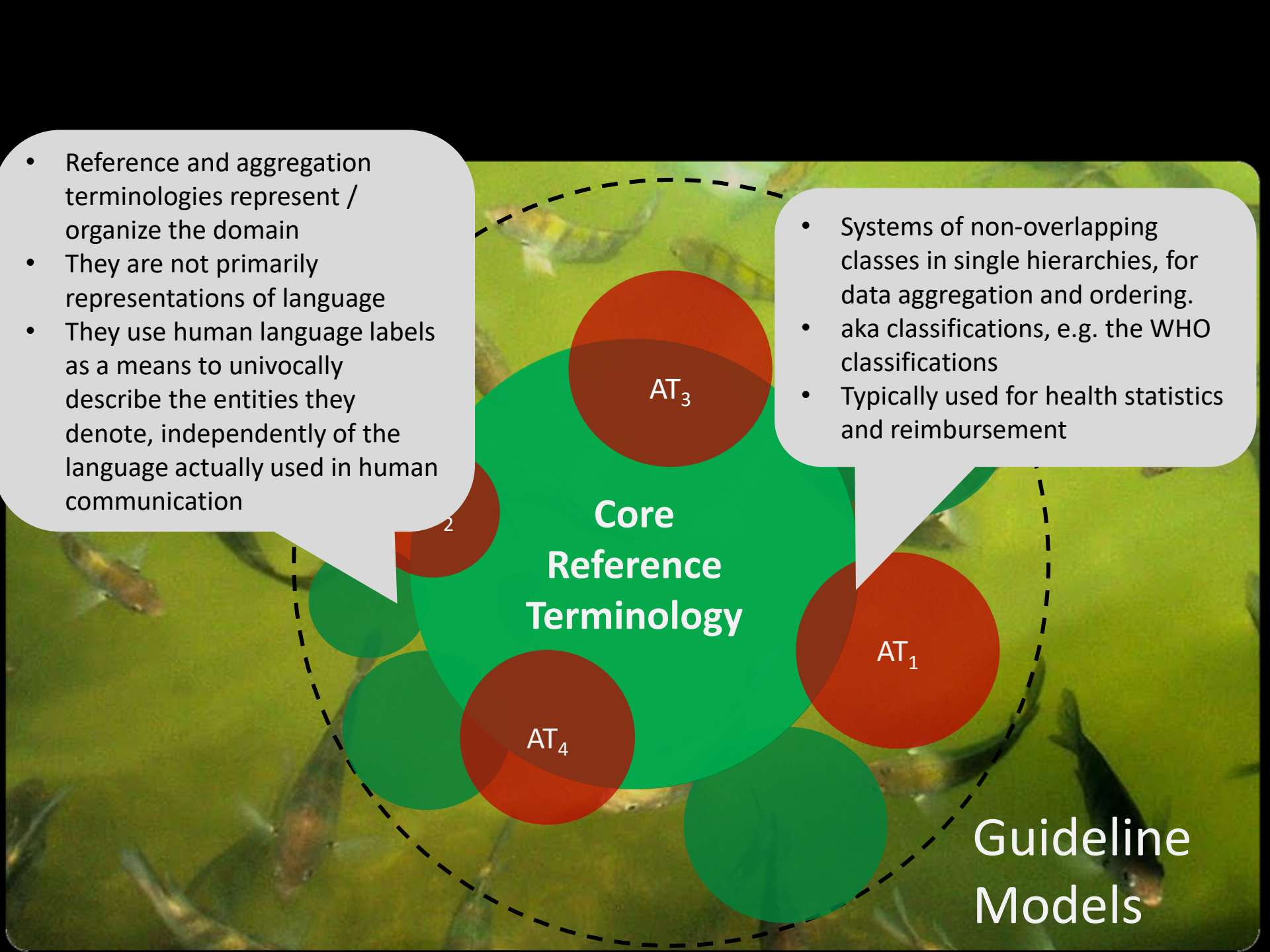
- European project on the fitness of purpose of SNOMED CT as a core reference terminology for the EU: www.assess-ct.eu
- Feb 2015 – Jul 2016
- Scrutinising clinical, technical, financial, and organisational aspects of reference terminology introduction
- Main recommendations:
  - "SNOMED CT is the best candidate for a core reference terminology for cross-border, national and regional eHealth deployments in Europe."
  - Must be part of an ecosystem of semantic assets

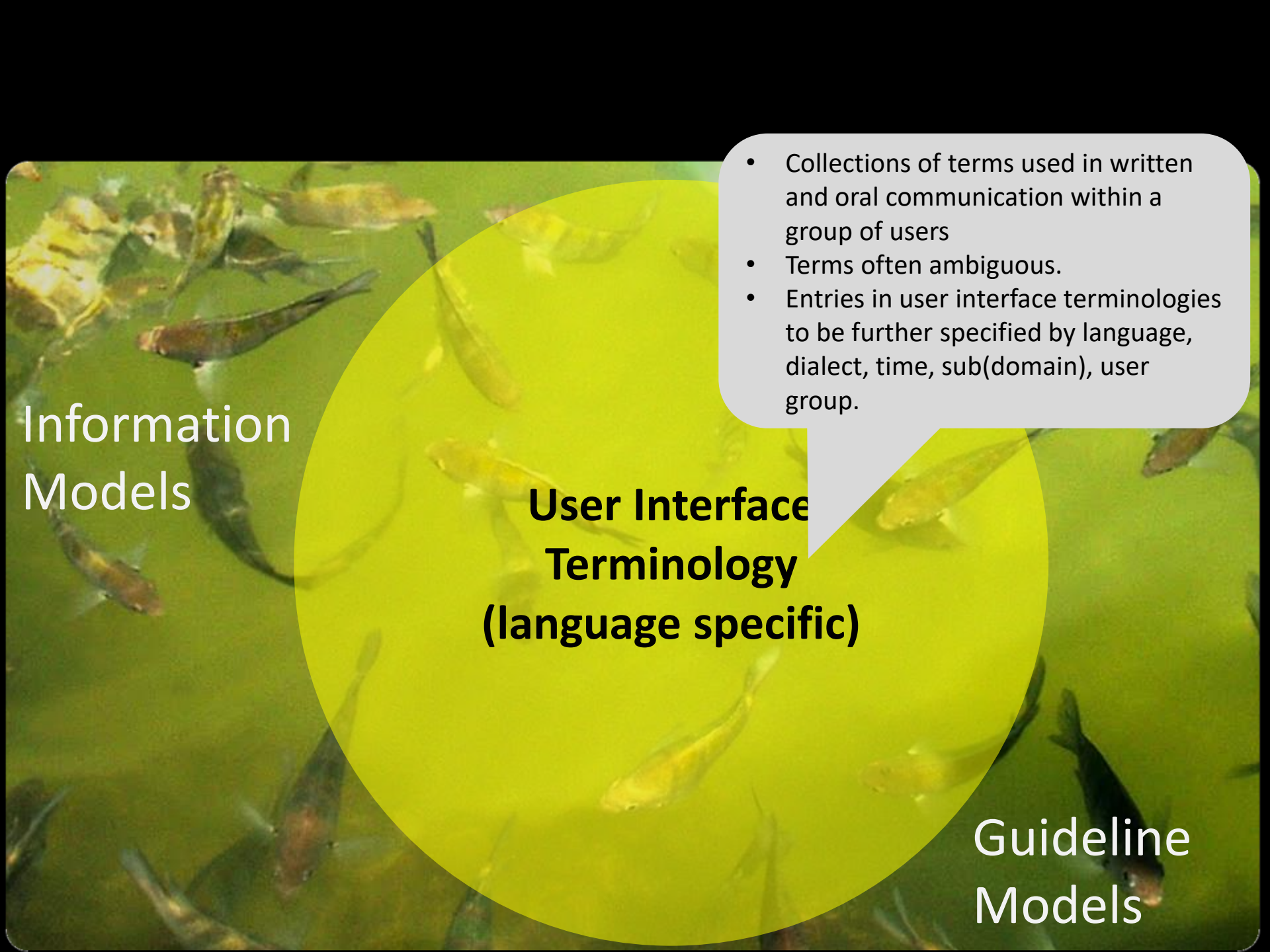
Ecosystem of semantic assets

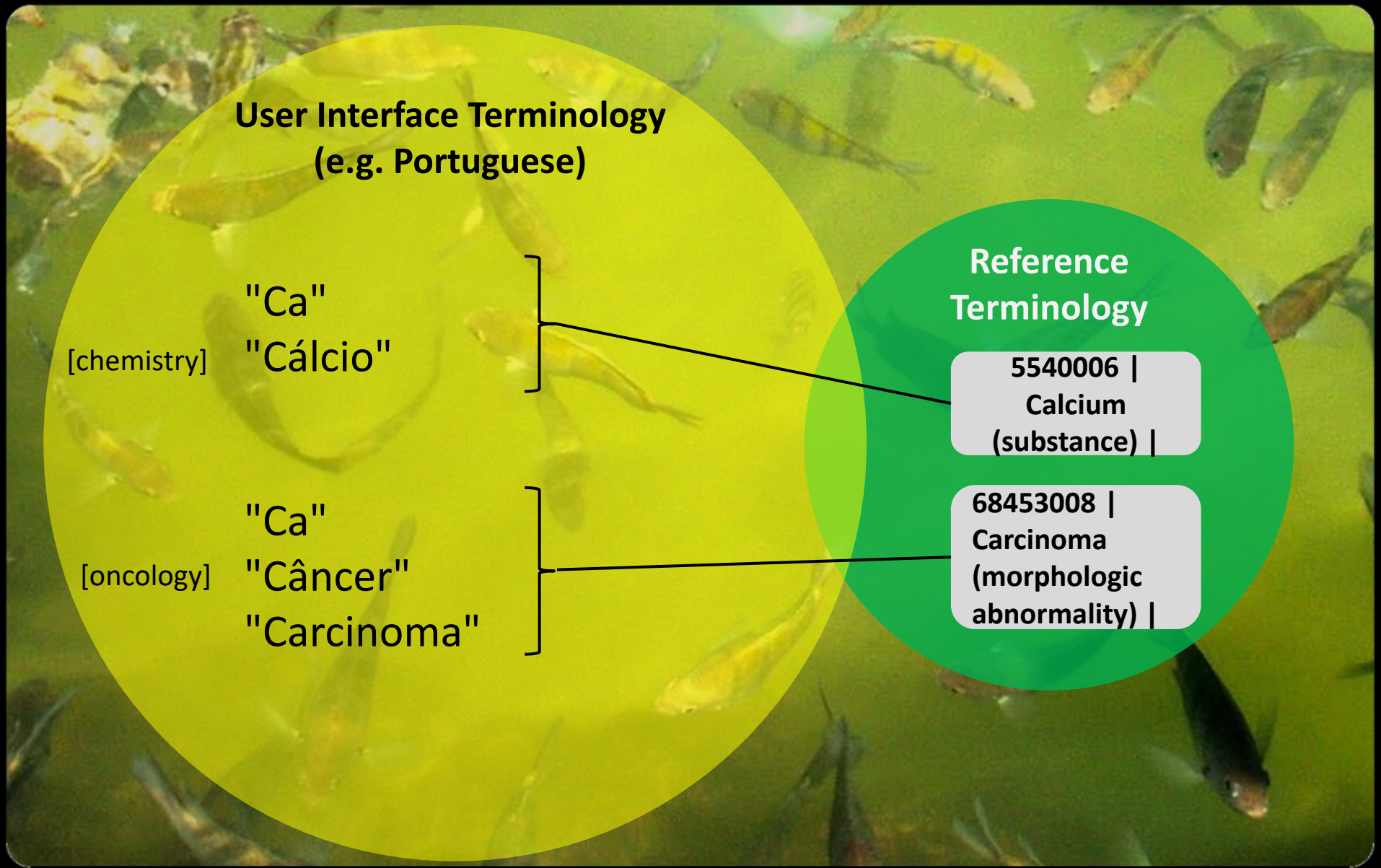


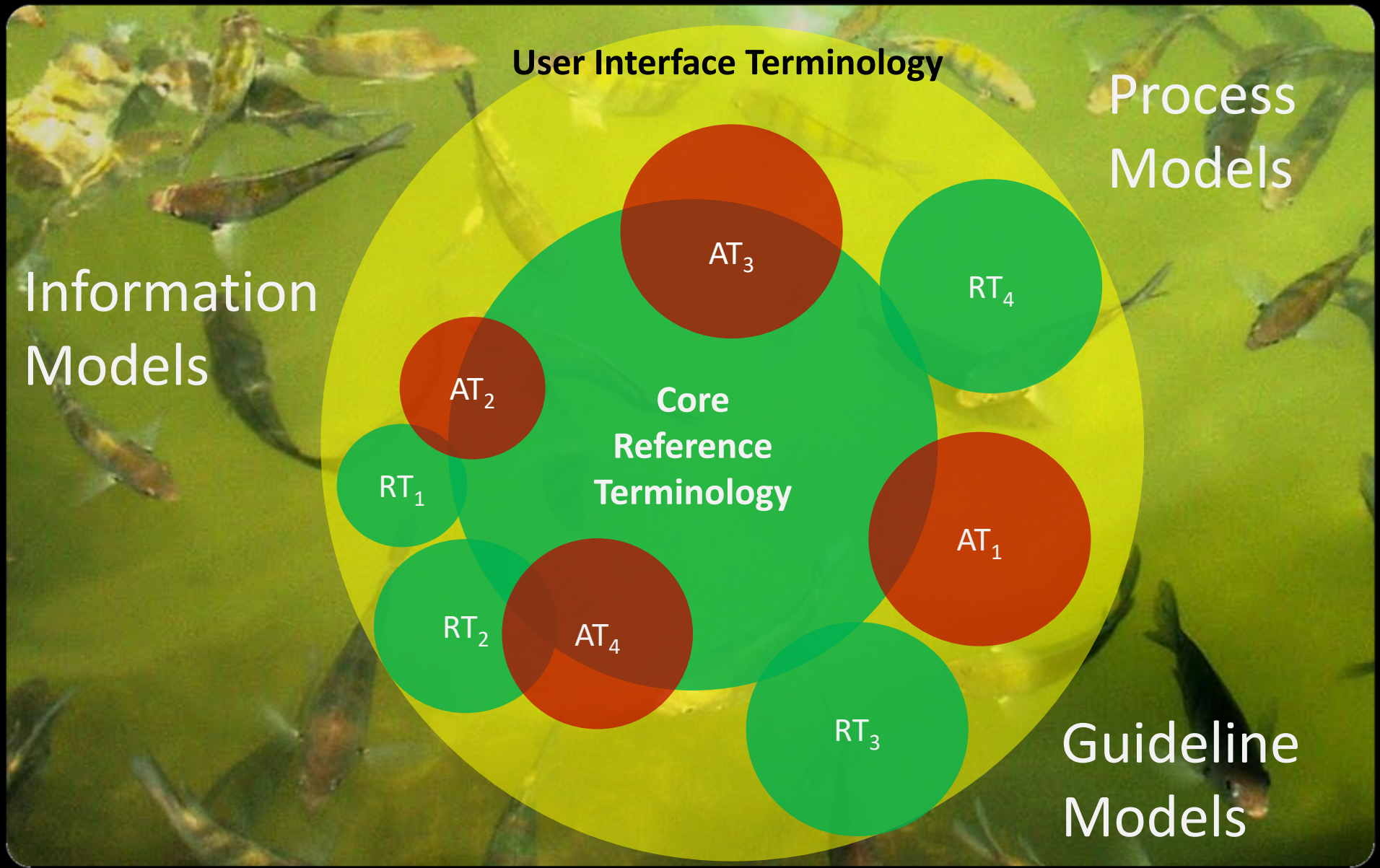

Information Models
Reference Terminologies
Guideline Models
…describe and standardize a neutral, language-independent sense
• The meaning of domain terms
• The properties of the objects that these terms denote
• Representational units are commonly called “concepts”
• RTs enhanced by formal descriptions = "Ontologies"


Information Models
Core Reference Terminology
AT3
AT2
AT1
AT4
Aggregation Terminologies (Classifications)
Guideline Models
• Systems of non-overlapping classes in single hierarchies, for data aggregation and ordering.
• aka classifications, e.g. the WHO classifications
• Typically used for health statistics and reimbursement


• Reference and aggregation terminologies represent / organize the domain
• They are not primarily representations of language
• They use human language labels as a means to univocally describe the entities they denote, independently of the language actually used in human communication
• Systems of non-overlapping classes in single hierarchies, for data aggregation and ordering.
• aka classifications, e.g. the WHO classifications
• Typically used for health statistics and reimbursement
Core Reference Terminology
AT3
AT1
AT4
Guideline Models


Information Models
User Interface Terminology (language specific)
Guideline Models
• Collections of terms used in written and oral communication within a group of users
• Terms often ambiguous.
• Entries in user interface terminologies to be further specified by language, dialect, time, sub(domain), user group.


User Interface Terminology
(e.g. Portuguese)
[chemistry]
"Ca"
"Cálcio"
[oncology]
"Ca"
"Câncer"
"Carcinoma"
Reference Terminology
5540006 | Calcium (substance) |
68453008 | Carcinoma (morphologic abnormality) |


User Interface Terminology
Process Models
Information Models
Guideline Models
Core Reference Terminology
AT3
RT4
AT2
RT1
AT1
RT2
AT4
RT3

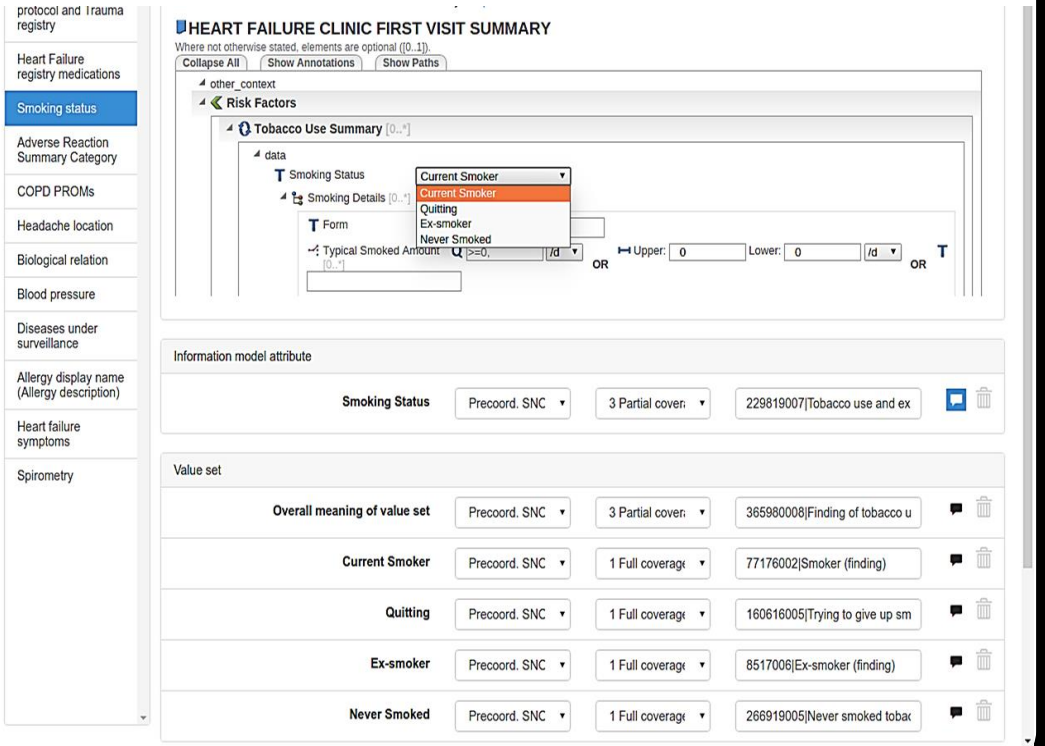
ASSESS CT investigations

ASSESS CT investigations

- Performance of human experts for
  1. Terminology binding to clinical models
  2. Annotating of clinical narratives
- Quality of annotation of clinical narratives using natural language processing
- End points
  - Concept coverage, inter-annotator agreement (1.,2.)
  - Term coverage (2.)

Terminology binding to clinical models

- Resources
  - 12 information model extracts, 101 elements
  - Full SNOMED CT vs. set of ICD-10, ATC, LOINC, and MeSH
  - 6 experts from 6 countries (5 EU + US)
- Method
  - SNOMED CT vs. compilation of other international terminologies (English interface terminology)
  - Complete annotation by each expert



Annotation of clinical narratives

- Resources
  - Parallel corpus: 60 clinical text samples from 6 languages, translated to all languages, representing clinical disciplines, document types and document sections
  - For each language: 2 annotators * 40 samples → 20 samples annotated twice
- Comparing
  - SNOMED CT vs.
  - UMLS - (SNOMED - Read – inactive sources -U.S. terminologies) + nonUMLS translations
    (artificial alternative core terminology as required by EU call)

Höger njure: inget anmärkningsvärt.
Röngten av buk
Bild är lätt att bedöma, bra belysning. Lite gas i tarmen. Vänster njure är en aning förstorad.

Results

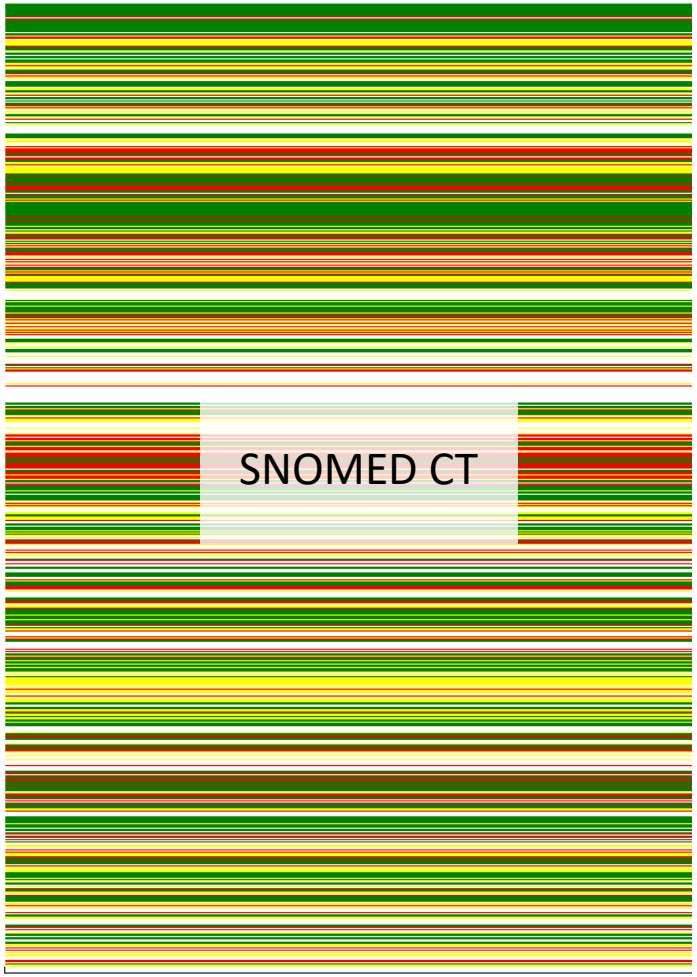
Concept coverage [95% CI]	SNOMED CT	Alternative
Clinical model annotations	.79 [.76-.82]	.51 [.57-.55]
Text annotations	.86 [.82-.88]	.88 [.86-.91]

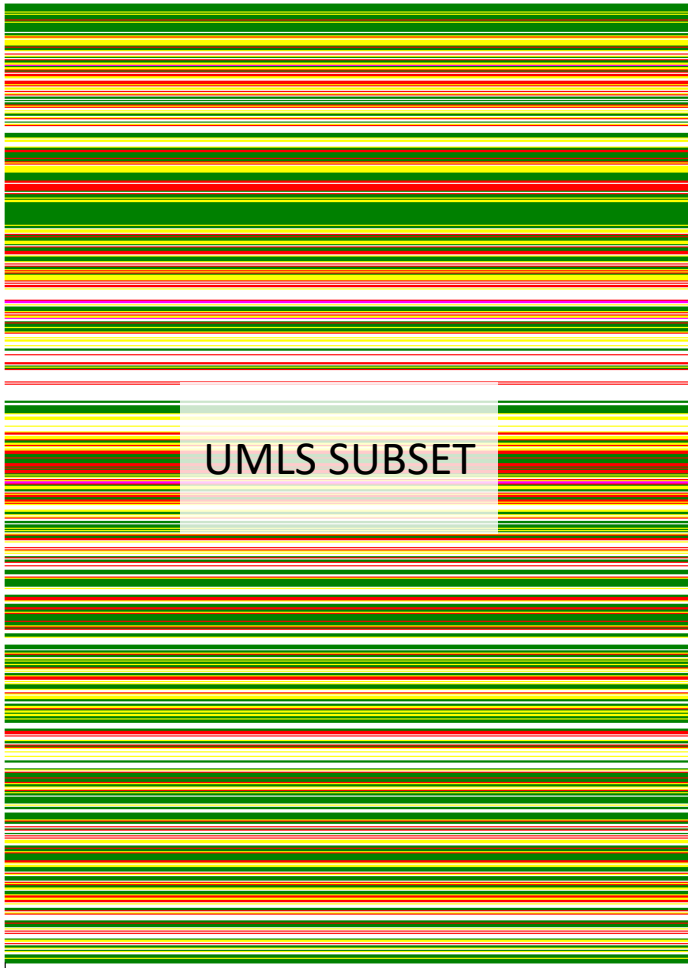
Term coverage [95% CI]	SNOMED CT	Alternative
Text annotations – English	.68 [.64; .70]	.73 [.69; .76]
Text annotations – Swedish	.47 [.44; .52]	.35 [.32; .40]

Inter annotator agreement Krippendorff's Alpha [95% CI]	SNOMED CT	Alternative
Clinical model annotations	.61 [.55-.66]	.47 [.41-.54]
Text annotations	.37 [.33-.41]	.36 [.32-.40]

Krippendorff, Klaus (2013). Content analysis: An introduction to its methodology, 3rd edition. Thousand Oaks, CA: Sage.

Agreement map: text annotations

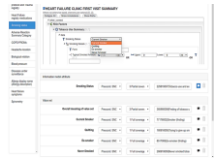




green: agreement – yellow: only coded by one coder – red: disagreement

Interoperability Scenario 1

Structured /
Semi-structured
Representation

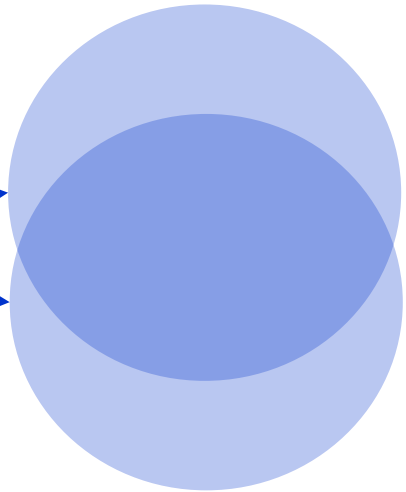


Höger njure: inget anmärkningsvärt.
Röngten av buk
Bild är lätt att bedöma, bra belysning. Lite gas i tarmen. Vänster njure är en aning förstorad.

Annotator #1
(human / machine)

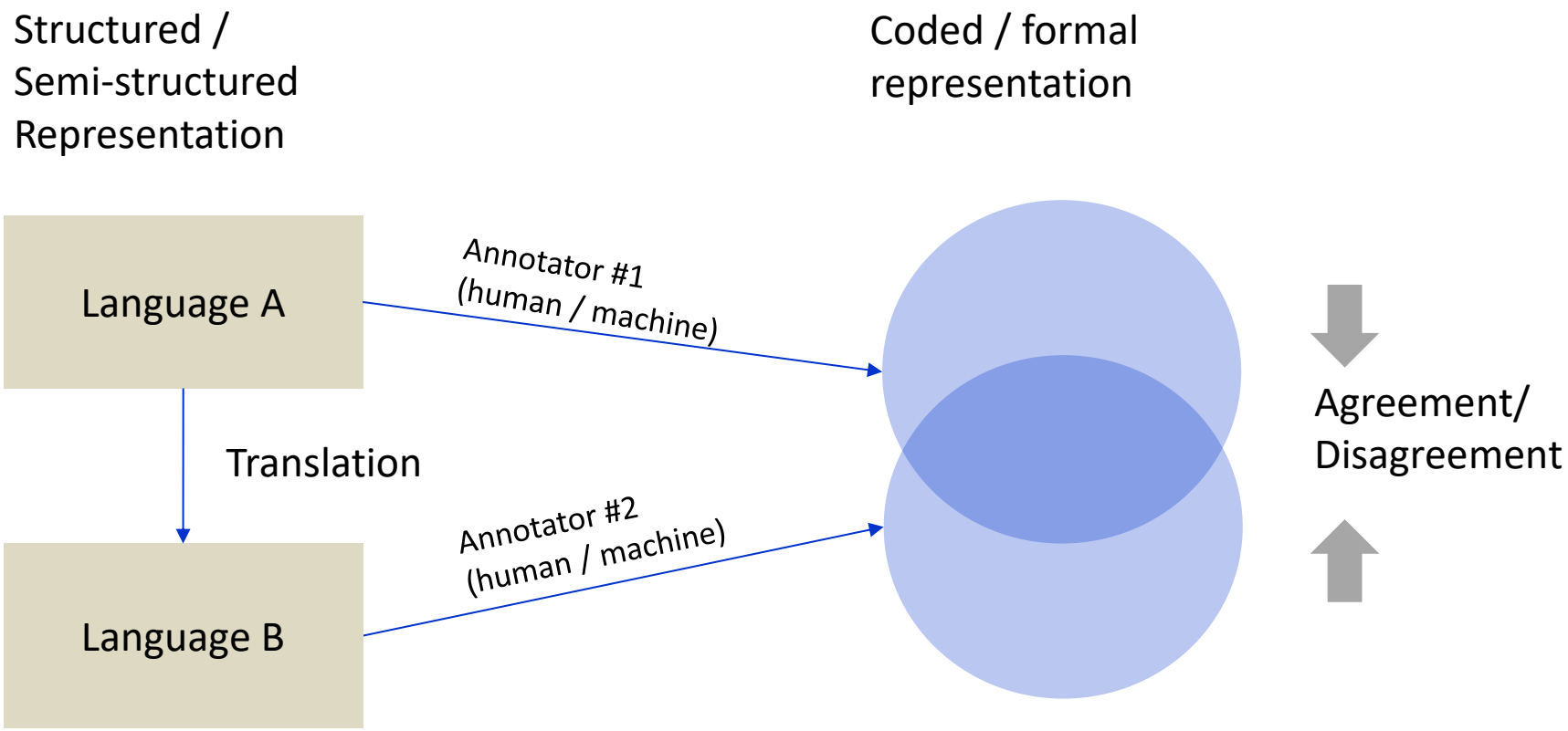
Annotator #2
(human / machine)

Coded / formal
representation



Agreement/
Disagreement

Interoperability Scenario 2



Further Analysis

- Creation of gold standard
  - 20 text samples annotated twice →208 NPs
  - Analysis of English SNOMED CT annotations by two additional terminology experts
  - Consensus finding where disagreements, following pre-established annotation guidelines
- Inspection and analysis of text annotation disagreements
- Inspection and analysis of disagreements in the clinical model annotation example

Reasons for disagreement

Human issues (I)

- Lack of domain knowledge / carelessness

Tokens	Annotator #1	Annotator #2	Gold standard
IV	53240008 |Structure of abductor hallucis muscle (body structure)	80622005 |Abducens nerve structure (body structure)	80622005 |Abducens nerve structure (body structure)

- Disregard of annotation guideline

Tokens	Annotator #1	Annotator #2	Gold standard
No	2667000 |Absent (qualifier value)	–	–
ptosis	11934000 |Ptosis of eyelid (disorder)	11934000 |Ptosis of eyelid (disorder)	11934000 |Ptosis of eyelid (disorder)

Human issues (II)

- Retrieval error (no interface term found)

Tokens	Annotator #1	Annotator #2	Gold standard
Glibenclamide	384978002 | Glyburide (substance)	–	384978002 | Glyburide (substance)

- Others:
  - Editing (mistyping)
  - Disregard of terminology specific constraints

Annotation guideline issues

- Underspecification
  - e.g. put anatomy concept always in procedure or disorder context

Tokens	Annotator #1	Annotator #2	Gold standard
IV	39322007 |Trochlear nerve structure	171674005 |Exploration of trochlear nerve (IV) (procedure)|	171674005 |Exploration of trochlear nerve (IV) (procedure)|

  - more general: avoid isolated primitive concepts
- Contradictions within annotation guidelines
  - absence of preference rules

Ontology issues (I)

- Polysemy ("dot categories")*

Tokens	Annotator #1	Annotator #2	Gold standard
Lymphoma	118600007 |Malignant lymphoma (disorder)|	115244002 |Malignant lymphoma - category (morphologic abnormality)|	118600007 |Malignant lymphoma (disorder)|

- Incomplete definitions / pseudo-polysemy

Tokens	Annotator #1	Annotator #2	Gold standard
Former	410513005|In the past (qualifier value)	77176002|Smoker (finding)	8517006|Ex-smoker (finding)
Smoker	77176002|Smoker (finding)	392521001 |History of (contextual qualifier) (qualifier value)|	8517006|Ex-smoker (finding)

* A. Arapinis, L. Vieu: Complex categories in ontologies, FOIS 2014 Workshop on ontology and linguistics Applied

Ontology issues (II)

- Incomplete definitions

Tokens	Annotator #1	Annotator #2	Gold standard
Diabetes	73211009 |Diabetes mellitus (disorder)|	170742000 |Diabetic monitoring (regime/therapy)|	170742000 |Diabetic monitoring (regime/therapy)|
monitoring	360152008 |Monitoring - action (qualifier value)|	170742000 |Diabetic monitoring (regime/therapy)|	170742000 |Diabetic monitoring (regime/therapy)|

- Navigational concepts (not for coding)

Tokens	Annotator #1	Annotator #2	Gold standard
palpebral fissure	301916005 |Finding of measures of palpebral fissure (finding)|	595000 |Structure of palpebral fissure (body structure)|	363934008 |Measure of palpebral fissure (observable entity)|

Ontological issues (III)

- Normal findings, no full definitions

Tokens	Annotator #1	Annotator #2	Gold standard
Motor:	127954009 |Skeletal muscle structure (body structure)|	106030000 |Muscle finding (finding)|	298300008 |Skeletal muscle normal (finding)|
normal bulk and tone	17621005 |Normal (qualifier value)|	17621005 |Normal (qualifier value)|	298300008 |Skeletal muscle normal (finding)|

- Fuzziness of qualifiers

Tokens	Annotator #1	Annotator #2	Gold standard
Significant	386134007 |Significant (qualifier value)|	24484000 |Severe (severity modifier) (qualifier value)|	6736007 |Moderate (severity modifier) (qualifier value)|
bleeding	131148009 |Bleeding (finding)|	131148009 |Bleeding (finding)|	131148009 |Bleeding (finding)|

Interface term issues

Tokens	Annotator #1	Annotator #2	Gold standard
Pain	406189006 |Pain observable (observable entity)|	22253000 |Pain (finding)|	22253000 |Pain (finding)|
		"pain observations"	

Tokens	Annotator #1	Annotator #2	Gold standard
Blood	87612001 |Blood (substance)|	50960005 |Hemorrhage (morphologic abnormality)|	50960005 |Hemorrhage (morphologic abnormality)|
extravasation	76676007 |Extravasation (morphologic abnormality)|		
		"extravasation of blood"	

Tokens	Annotator #1	Annotator #2	Gold standard
anxious	48694002 |Anxiety (finding)|	79015004 |Worried (finding)|	48694002 |Anxiety (finding)|
		"anxious cognitions"	

Language issues

- Ellipsis / anaphora
  - "Cold and wind are provoking factors as well." (provoking factors for angina)
  - "These ailments have substantially increased since October 2013" (weakness)
- Context
  - "No surface irregularities" (breast)
  - "Significant bleeding" (gastrointestinal bleeding)
  - "IV" (intravenous? Forth cerebral nerve? Type 4)
- Co-ordination:
  - "normal factors 5, 9, 10, and 11 "
- Negation
  - "no tremor, rigidity or bradykinesia"

Prevention of annotation disagreements

Prevention of annotation disagreements

- Users (humans, text processing algorithms)
  - Training
  - Tooling
    - Guideline enforcement by appropriate tools
    - Post-co-ordination
  - Machine-processable annotation rules
  - Context awareness, scoping (e.g. looking back for anaphora resolution, identification of content of text passages)
  - Support by comprehensive, well-curated interface terminologies, tailored to the specific sublanguage to be analyzed

Preventive measures (SNOMED CT structure)

- Fill gaps
  - equivalence axioms (reasoning)
  - Self-explaining labels (FSNs)
  - Scope notes where necessary (e.g. what means "entitic")
- Remove unnecessary ambiguity
- Flag concepts that should not be used (navigational concepts, anatomic "entire" concepts)
- Strengthen ontological foundations
  - Upper-level ontology alignment
  - Formalize constraints (SNOMED CT concept model)
  - Ontology / information model boundary
  - Overhaul problematic subhierarchies, especially qualifiers

Preventive measures (SNOMED CT content maintenance)

- Include large-scale analysis of real data in routine maintenance process
  - Harvest notorious disagreements between notorious text passages and value sets with concepts
  - Compare concept frequency across institutions and users to detect imbalances
- Stimulate community processes for ontology-guided content evolution:
  - SNOMED CT ontological content
  - Interface terminologies for languages, specialties, users
  - Linking interface terminologies / value sets with SNOMED CT codes or expressions

Remediation of annotation disagreements

Remediation of annotation disagreements

- Dependencies / Inferences

Concept A	Concept B	Dependency
Mast cell neoplasm (disorder)	Mast cell neoplasm (morphologic abnormality)	A subclassOf AssociatedMorphology some B
Isosorbide dinitrate (product)|	Isosorbide dinitrate (substance)	A subclassOf HasActiveIngredient some B
Palpation (procedure)|	Palpation - action (qualifier value)	A subclassOf Method some B
Blood pressure taking (procedure)	Blood pressure (observable entity)	No connection
Increased size (finding)	Increased (qualifier value)	No connection
Finding of heart rate (finding)|	Heart rate (observable entity)	A subclassOf Interprets some B
Electrocardiogram finding (finding)|	Electrocardiographic procedure (procedure)	A subclassOf Interprets some B
Electrocardiogram finding (finding)|	Electrocardiogram finding (observable entity)	No connection

Experiment

- Gold standard expansion:
  - Step 1: include concepts linked by attributive relations:
    - A subclassOf
      Rel some B
  - Step 2: include additional first-level taxonomic relations:
    - A subclassOf B
  - Apply to results from English and Swedish annotator

Result

Language of text sample	Gold standard expansion	F measure
English	no expansion	0.28
	expansion step 1	0.28
	expansion step 2	0.29
Swedish	no expansion	0.14
	expansion step 1	0.15
	expansion step 2	0.15

- Minimal improvement
- Side observation (English vs. Swedish):
  - Translation effects
  - Interface terminology effects

Work in progress (I)

- Transformation of code groupings in plausible post-coordinated expressions:
  - Source group:
    - 24 Hour electrocardiogram (procedure)
    - Cardiac arrhythmia (disorder)
  - Pattern:
    Procedure (procedure) -> {Has focus (attribute)-[Clinical finding (finding)]}
  - Pattern frequency in SNOMED CT : 748 (frequent)
  - Suggested representation:
    24 Hour electrocardiogram (procedure) -> {Has focus (attribute)-[Cardiac arrhythmia (disorder)]}
- Limitations: ambiguities (e.g. substance - disorder)

Work in progress (II)

- Enrichment of reference standard by maximally post-coordinated expressions

Tokens	Gold standard codes	Gold standard post-coordinated expression
wounds	416462003 |Wound (disorder)|	"262749000 |Open wound of eyelid (disorder)|: { 116676008 |Associated morphology (attribute)| = 59091005 |Open wound (morphologic abnormality)|, 363698007 |Finding site (attribute)| = (51360009 |Skin structure of eyelid (body structure)|: 272741003 |Laterality (attribute)| = 7771000 |Left (qualifier value)|) } + 313261004 |Open wound of chin (disorder)|: { 116676008 |Associated morphology (attribute)| = 59091005 |Open wound (morphologic abnormality)|, 363698007 |Finding site (attribute)| = (30291003 |Chin structure (body structure)|: 272741003 |Laterality (attribute)| = 7771000 |Left (qualifier value)|) }"
to		
the		
left	7771000|Left side	
eyelid	262749000|Open wound of eyelid;313261004|Open wound of chin	
and		
chin	262749000|Open wound of eyelid;313261004|Open wound of chin	

Conclusion

- Lack of inter-annotator agreement impairs successful use of clinical terminologies /ontologies
  - SNOMED CT slightly better than alternative scenario
- Prevention:
  - Education, tooling, annotation / coding guidelines
  - Content quality improvement: labelling, scope notes, ontological clarity, full definitions, community processes, large-scale clinical data analysis
  - Importance of interface terminologies, dealing with ambiguity
- Mitigation
  - Classical language understanding challenges
  - Resolution of agreement issues still speculative, e.g. machine-supported post-co-ordination
  - Research required

- Acknowledgements: ASSESS CT team: Jose Antonio Miñarro-Giménez, Catalina Martínez-Costa, Daniel Karlsson, Kirstine Rosenbeck Gøeg, Kornél Markó, Benny Van Bruwaene, Ronald Cornet, Marie-Christine Jaulent, Päivi Hämäläinen, Heike Dewenter, Reza Fathollah Nejad, Sylvia Thun, Veli Stroetmann, Dipak Kalra
- Contact: stefan.schulz@medunigraz.at