Keynote address:

Annotating clinical narratives with SNOMED CT: The thorny way towards interoperability of clinical routine data
"Classical" AI workflow
"Classical" AI workflow

Data Acquisition → Representation

- Reasoning A → Output A
- Reasoning B → Output B
"Classical" AI workflow

Data Acquisition → Representation A → Reasoning → Output A

Data Acquisition → Representation B → Reasoning → Output B
"Classical" AI workflow

- Data Acquisition A
- Data Acquisition B
- Representation DA
- Representation DB
- Reasoning
- Output A
- Output B
Data reliability $\rightarrow$ Data interoperability
Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more unstructured representation

Interpretation A → $D_A$

Interpretation B → $D_B$

structured representation

$D_A = D_B$

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more unstructured representation

Interpretation A → $D_A$

Interpretation B → $D_B$

structured representation

$D_A = D_B$

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more unstructured representation

Interpretation A → $D_A$

Interpretation B → $D_B$

structured representation

$D_A = D_B$

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more unstructured representation

Interpretation A → $D_A$

Interpretation B → $D_B$

structured representation

$D_A = D_B$
Focus of the talk

- Structured extracts from unstructured clinical data: reliability and interoperability
- Empirical study on inter-annotator agreement
- Analysis of examples for inter-annotator disagreement
- Mechanisms to improve agreement
  - better data reliability
  - better interoperability
  - better training data
  - better gold standards
Annotating clinical narratives with SNOMED CT
Annotating clinical narratives with SNOMED CT

**Coding**

observation → map

phenomena observed → metadata configurations

**Vocabulary**

map

symbols (configurations) → metadata configurations

**Annotation**

symbolic representation
Annotating clinical narratives with **SNOMED CT**

- Huge clinical reference terminology
- eHealth standard, maintained by transnational SDO
- ~300,000 "concepts"

SNOMED CT represents:

- Preferred terms and synonyms in several languages
- Covers disorders, procedures, body parts, substances, devices, organisms, qualities...
- Multiple hierarchies
- (quasi-)ontological definitional and qualifying axioms
- Representable as OWL EL
Annotation: Sources of complexity

Contrary to popular belief, Lorem Ipsum is not simply random text. It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old. Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia, looked up one of the more

Clinical narrative
- sequence of Tokens
- syntactic structures
- relations at various levels

Ontology
- entities, codes
- relations
- logical constructors
- axioms

Terminology
- preferred terms
- synonyms
- definitions

• Compactness
• Agrammaticality
• Short forms
• Implicit contexts

best text span to annotate?
Naïve or analytic annotation?

• Ill-defined concepts
• Similar concepts
• Pre-coordination vs. post-coordination

Complex annotations (> 1 concept)
Degree of formality?
### Clinical text

<table>
<thead>
<tr>
<th>Clinical text</th>
<th>SNOMED CT concepts (FSNs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;... the duodenum. The mucosa is...&quot;</td>
<td>- 'Duodenal structure (body structure)'</td>
</tr>
</tbody>
</table>
| "...Hemorrhagic shock after RTA..." | - 'Traffic accident on public road (event)'
| "...travel history of suspected dengue..." | - 'Suspected dengue (situation)'

| | - 'Mucous membrane structure (body structure)'
| | - 'Duodenal mucous membrane structure (body structure)'
| | - 'Traffic accident on public road (event)', 'Renal tubular acidosis (disorder)'
| | - 'Traffic accident on public road (event)' or 'Renal tubular acidosis (disorder)'
| | - 'Suspected (qualifier value)'
| | - 'Dengue (disorder)'

### Examples

- "... the duodenum. The mucosa is..."
- "...Hemorrhagic shock after RTA..."
- "...travel history of suspected dengue..."
Coding / Annotation guidelines

Examples:

1. German coding guidelines for ICD and OPS, 171 pages
2. Using SNOMED CT in CDA models: 147 pages
3. CHEMDNER-patents: annotation of chemical entities in patent corpus: annotation manual 30 pages
4. CRAFT Concept Annotation guidelines: 47 pages
5. Gene Ontology Annotation conventions: 7 pages

Complex rule sets, requiring intensive training

Annotation experiments in ASSESS-CT
Annotation experiments in ASSESS-CT

- EU project on the fitness of purpose of SNOMED CT as a core reference terminology for the EU: www.assess-ct.eu
- Scrutinising clinical, technical, financial, and organisational aspects of reference terminology introduction
- Summary of results: brochure published, scientific papers to appear

Annotation of clinical narratives

- Comparing
  - SNOMED CT vs. UMLS derived terminology
- Resources
  - Parallel corpus: 60 clinical text snippets from 6 languages, high diversity
  - For each language: 2 annotators * 40 samples → 20 snippets annotated twice
- Annotators
  - trained by webinars
  - follow annotation guideline (10 pages)

- e.g.
- chunking into noun phrases
- annotation of chunks by sets of codes
- give preference to maximally pre-coordinated codes
- understanding text and assign maximally specific codes

<table>
<thead>
<tr>
<th>Medical Term</th>
<th>Concept ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitroglycerin pump spray as required</td>
<td>387404004;385074009;225761000</td>
</tr>
<tr>
<td>Amantadine bds</td>
<td>372763006;229799001</td>
</tr>
<tr>
<td>Allopurinol 300 ½ tablet every other day (last dose on)</td>
<td>387135004;385055001;225760004</td>
</tr>
<tr>
<td>Mefenamic acid 500 mg up to 3x daily for pain in conjunction with</td>
<td>387185008;258684004;229798009;22253000</td>
</tr>
<tr>
<td>Pantoprazole 40mg.</td>
<td>395821003;258684004</td>
</tr>
<tr>
<td>Torasemide bds</td>
<td>318034005;229799001</td>
</tr>
<tr>
<td>Melperone 50 mg p.m.</td>
<td>442519006;258684004;422133006</td>
</tr>
<tr>
<td>§ 7 Intact teeth are in the mouth.</td>
<td>11163003;245543004;123851003</td>
</tr>
<tr>
<td>Fractures are visible on the medians of Mandible and Maxilla</td>
<td>263172003;263156006;260528009</td>
</tr>
<tr>
<td>the fragments are dislocated.</td>
<td>123735002</td>
</tr>
<tr>
<td>Normal mucous membranes in mouth, pharynx and on the larynx.</td>
<td>17621005;33044003;71248005</td>
</tr>
<tr>
<td>Hyoid and thyroid cartilage are intact.</td>
<td>21387005;52940008;11163003</td>
</tr>
<tr>
<td>Fragmental fractures of the two upper vertebrae of the cervical spine.</td>
<td>13321001;207984009;207983003</td>
</tr>
<tr>
<td>Otherwise the cervical spine is intact.</td>
<td>122494005;11163003</td>
</tr>
<tr>
<td>Oesophagus as well as trachea are torn at the lower end of the neck.</td>
<td>262793000;282459005;261122009;123958008</td>
</tr>
</tbody>
</table>
## Principal quantitative results (English)

<table>
<thead>
<tr>
<th></th>
<th>SNOMED CT</th>
<th>Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Concept coverage [95% CI]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text annotations – English</td>
<td>.86 [.82-.88]</td>
<td>.88 [.86-.91]</td>
</tr>
<tr>
<td><strong>Term coverage [95% CI]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text annotations – English</td>
<td>.68 [.64; .70]</td>
<td>.73 [.69; .76]</td>
</tr>
<tr>
<td><strong>Inter annotator agreement Krippendorff's Alpha [95% CI]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text annotations</td>
<td>.37 [.33-.41]</td>
<td>.36 [.32-.40]</td>
</tr>
</tbody>
</table>

Agreement map: text annotations (English)

- Green: agreement
- Yellow: only annotated by one coder
- Red: disagreement
Systematic error analysis

- Creation of gold standard for SNOMED CT
  - 20 English text samples annotated twice → 208 NPs
  - Analysis of English SNOMED CT annotations by two additional terminology experts
  - Consensus finding, according to pre-established annotation guidelines
- Inspection, analysis and classification of text annotation disagreements
- Presentation of some disagreement cases for SNOMED CT
Reasons for disagreement
Human issues

- Lack of domain knowledge / carelessness

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;IV&quot;</td>
<td>'Structure of abductor hallucis muscle (body structure)'</td>
<td>'Abducens nerve structure (body structure)'</td>
<td>'Abducens nerve structure (body structure)'</td>
</tr>
</tbody>
</table>

- Retrieval error (synonym not recognised)

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Glibenclamide&quot;</td>
<td>'Glyburide (substance)'</td>
<td>–</td>
<td>'Glyburide (substance)'</td>
</tr>
</tbody>
</table>

- Non-compliance with annotation rules
## Ontology issues (I)

- **Polysemy ("dot categories")**

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Lymphoma'</td>
<td>'Malignant lymphoma (disorder)'</td>
<td>'Malignant lymphoma - category (morphologic abnormality)'</td>
<td>'Malignant lymphoma (disorder)'</td>
</tr>
</tbody>
</table>

# Ontology issues (I)

- **Polysemy ("dot categories")**

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Lymphoma'</td>
<td>'Malignant lymphoma (disorder)'</td>
<td>'Malignant lymphoma - category (morphologic abnormality)'</td>
<td>'Malignant lymphoma (disorder)'</td>
</tr>
</tbody>
</table>

- **"Pseudo-polysemy"**

- **Incomplete definitions**

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Former&quot;</td>
<td>'In the past (qualifier value)'</td>
<td>'History of (contextual qualifier) (qualifier value)'</td>
<td>'Ex-smoker (finding)'</td>
</tr>
<tr>
<td>Smoker&quot;</td>
<td>'Smoker (finding)'</td>
<td>'Smoker (finding)'</td>
<td></td>
</tr>
</tbody>
</table>
## Ontological issues (II)

### Incomplete definitions

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Motor: &quot;Skeletal muscle structure (body structure)&quot;</td>
<td>'Muscle finding (finding)'</td>
<td></td>
<td>'Skeletal muscle normal (finding)'</td>
</tr>
<tr>
<td>normal bulk and tone&quot;</td>
<td>'Normal (qualifier value)'</td>
<td>'Normal (qualifier value)'</td>
<td></td>
</tr>
</tbody>
</table>

![Diagram](image.png)
## Ontological issues (II)

### Normal findings, incomplete definitions

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Motor:</td>
<td>'Skeletal muscle structure (body structure)'</td>
<td>'Muscle finding (finding)'</td>
<td>'Skeletal muscle normal (finding)'</td>
</tr>
<tr>
<td>normal bulk and tone</td>
<td>'Normal (qualifier value)'</td>
<td>'Normal (qualifier value)'</td>
<td></td>
</tr>
</tbody>
</table>

### Fuzziness of qualifiers

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Significant</td>
<td>'Significant (qualifier value)'</td>
<td>'Severe (severity modifier) (qualifier value)'</td>
<td>'Moderate (severity modifier) (qualifier value)'</td>
</tr>
<tr>
<td>bleeding&quot;</td>
<td>'Bleeding (finding)'</td>
<td>'Bleeding (finding)'</td>
<td>'Bleeding (finding)'</td>
</tr>
</tbody>
</table>
## Interface term (synonym) issues

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Blood&quot;</td>
<td>'Blood (substance)''</td>
<td>'Hemorrhage (morphologic abnormality)'</td>
<td>'Hemorrhage (morphologic abnormality)'</td>
</tr>
<tr>
<td>extravasation</td>
<td>'Extravasation (morphologic abnormality)'</td>
<td>'Hemorrhage (morphologic abnormality)'</td>
<td>&quot;extravasation of blood&quot;</td>
</tr>
</tbody>
</table>

![Hemorrhage (morphologic abnormality) SCTID: 50960005](image)
## Interface term (synonym) issues

<table>
<thead>
<tr>
<th>Tokens</th>
<th>Annotator #1</th>
<th>Annotator #2</th>
<th>Gold standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Blood&quot;</td>
<td>'Blood (substance)'</td>
<td>'Hemorrhage (morphologic abnormality)'</td>
<td>'Hemorrhage (morphologic abnormality)'</td>
</tr>
<tr>
<td>extravasation</td>
<td>'Extravasation (morphologic abnormality)'</td>
<td>'Extravasation of blood'</td>
<td></td>
</tr>
<tr>
<td>&quot;anxious&quot;</td>
<td>'Anxiety (finding)'</td>
<td>'Worried (finding)'</td>
<td>'Anxiety (finding)'</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&quot;anxious cognitions&quot;</td>
</tr>
</tbody>
</table>
Language issues

- Ellipsis / anaphora
  - "Cold and wind are provoking factors."
    (provoking factors for angina)
  - "These ailments have substantially increased since October 2013" (weakness)
  - "No surface irregularities" (breast)
  - "Significant bleeding" (intestinal bleeding)

- Ambiguity of short forms
  - "IV" (intravenous? Fourth intracranial nerve?)

- Co-ordination:
  - "normal factors 5, 9, 10, and 11"

- Scope of negation
  - "no tremor, rigidity or bradykinesia"

- Addressed by annotation guideline
- Manageable by human annotators
- Known challenges for NLP systems
Prevention and remediation of annotation disagreements
Prevention: annotation processes

- Training with continuous feedback
  - Early detection of inter annotator disagreement triggers guideline enforcement / guideline revision
- Tooling
  - Optimised concept retrieval (fuzzy, substring, synonyms)
  - Guideline enforcement by appropriate tools
  - Postcoordination support (complex syntactic expressions instead of grouping of concepts)
  - Anti-patterns, e.g. avoid unrelated primitive concepts (?)
Prevention: improve terminology structure

- Fill gaps
  - equivalence axioms (reasoning)
  - Self-explaining labels (FSNs), especially for qualifiers
  - Scope notes / text definitions where necessary
- Manage polysemy
- Flag navigational and modifier concepts
- Strengthen ontological foundations
  - Upper-level ontology alignment
  - Clear division between domain entities and information entities
  - Overhaul problematic subhierarchies, especially qualifiers
Prevention: improve content maintenance

- Analysis of real data to support terminology maintenance process
  - Harvest notorious disagreements between text passages and annotations from clinical datasets
  - Compare concept frequency and concept co-occurrence between comparable institutions and users to detect imbalances
- Stimulate community processes for ontology-guided content evolution:
  - Crowdsourcing of interface terms by languages, dialects specialties, user groups (separation of interface terminologies from reference terminologies is one of the ASSESS-CT recommendations)
Remediation of annotation disagreements
Remediation of annotation disagreements

- Exploit ontological dependencies / implications

<table>
<thead>
<tr>
<th>Concept A</th>
<th>Concept B</th>
<th>Dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Mast cell neoplasm (disorder)'</td>
<td>'Mast cell neoplasm (morphologic abnormality)'</td>
<td>A subclassOf AssociatedMorphology some B</td>
</tr>
<tr>
<td>'Isosorbide dinitrate' (product)'</td>
<td>'Isosorbide dinitrate (substance)'</td>
<td>A subclassOf HasActiveIngredient some B</td>
</tr>
<tr>
<td>'Palpation (procedure)'</td>
<td>'Palpation - action (qualifier value)'</td>
<td>A subclassOf Method some B</td>
</tr>
<tr>
<td>'Blood pressure taking (procedure)'</td>
<td>'Blood pressure (observable entity)'</td>
<td>A subclassOf hasOutcome some B</td>
</tr>
<tr>
<td>'Increased size (finding)'</td>
<td>'Increased (qualifier value)'</td>
<td>A subclassOf isBearerOf some B</td>
</tr>
<tr>
<td>'Finding of heart rate (finding)'</td>
<td>'Heart rate (observable entity)'</td>
<td>A subclassOf Interprets some B</td>
</tr>
</tbody>
</table>
Experiment

- **Gold standard expansion:**
  - Step 1: include concepts linked by attributive relations:
    - $A \text{ subclassOf } \textbf{Rel} \text{ some } B$
  - Step 2: include additional first-level taxonomic relations:
    - $A \text{ subclassOf } B$

<table>
<thead>
<tr>
<th>Language of text sample</th>
<th>Gold standard expansion</th>
<th>F measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>no expansion</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>expansion step 1</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>expansion step 2</td>
<td>0.29</td>
</tr>
</tbody>
</table>

- only insignificant improvement
- possibly due to missing relations in SNOMED CT, e.g. haemorrhage - blood
Conclusion (I)

- Low inter-annotator agreement limits successful use of clinical terminologies / ontologies
  - for manual annotation scenarios
  - for benchmarking of NLP-based annotations
  - for optimised training data for ML
- Structured data essential for many intelligent systems, but unreliable information extracted from clinical narratives raises patient safety issues when used for decision support
Conclusion (II)

- **Prevention of disagreements**
  - Education, tooling, guideline support
  - Terminology content improvement: labelling, scope notes, ontological clarity, full definitions, community processes
  - High coverage interface terminologies

- **Remediation of disagreements**
  - So far no clear evidence of ontology-based resolution of agreement issues
  - Big data approaches?
R & D required:

- "Learning systems" for improvement terminology content / structure / tooling. Clinical "big data" underused resource
- Harmonization of annotation guideline creation and validation efforts
- Formulate and enforce good quality criteria for clinical terminologies used as annotation vocabularies
- Better ontological underpinning of clinical terminologies
- Ontologically founded patterns for recurring clinical documentation tasks: Information extraction rather than concept mapping*

*Martínez-Costa C et al. Semantic enrichment of clinical models towards semantic interoperability. JAMIA 2015 May;22(3):565-76
Thanks for your attention

- Slides will be accessible via at purl.org/steschu
- 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