Call: HORIZON-HLTH-2021-TOOL-06
Topic: HORIZON-HLTH-2021-TOOL-06-03
Funding Scheme: HORIZON Research and Innovation Actions (RIA)

Grant Agreement no: 101057062

AIDAVA

AI powered Data Curation & Publishing Virtual Assistant

Deliverable No. 4.1
Annotations guidelines, tools & training

Approval by the European Commission Pending

Contractual Submission Date: 31/12/2022

Actual Submission Date: 05/01/2023

Responsible partner: P7-Medical University of Graz (MUG)

Funded by
the European Union

Grant agreement no.	101057062
Project full title	AIDAVA - AI powered Data Curation & Publishing Virtual Assistant

Deliverable number	**D4.1**
Deliverable title	**Annotation guidelines, tools & training**
Type[1]	DEM
Dissemination level[2]	PU
Work package number	WP4
Work package leader	P7-MUG
Author(s)	Markus Kreuzthaler (MUG), Kris Collins, Stefan Schulz (AVER), Kristian Kankainen (NEMC), Isabelle de Zegher (b!lo)
Keywords	manual annotation tools, manual annotation guidelines

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA).

Neither the European Union nor the granting authority can be held responsible for them.

Document History

Version	Date	Description
V01	05.01.2023	First submission

[1] **Type**: Use one of the following codes (in consistence with the Description of the Action):

R:	Document, report (excluding the periodic and final reports)
DEM:	Demonstrator, pilot, prototype, plan designs
DEC:	Websites, patents filing, press & media actions, videos, etc.

[2] **Dissemination level**: Use one of the following codes (in consistence with the Description of the Action)

PU:	Public, fully open, e.g. web
SEN:	Sensitive, limited under conditions of the Grant Agreement

Table of Contents

List of Abbreviations and definitions

The abbreviations and definitions used in the deliverable are based on the AIDAVA Glossary.

Abbreviation	Full Name
AI	Artificial Intelligence
API	Application Programming Interface
ATC	Anatomical Therapeutic Chemical
CDISC	Clinical Data Standards Consortium
CEN	Committee European de Normalisation
CRO	Contract Research Organisation
CVD	Cardio-vascular disease
DGA	Data Governance Act
DI	Data Intermediary
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DMP	Data Management Plan
DPIA	Data Protection Information Assessment
DPR	Data Portability Request
EAB	Ethics Advisory Board
EHDS	European Health Data Space
EHR	Electronic Health Record
EMA	European Medicine Agency
ETL	Extraction Transformation Load
FAIR	Findable Accessible Interoperable Reusable
FHIR	Fast Healthcare Interoperability Resources
GA	General Assembly
GDPR	Global Data Protection Regulation
GP	General Practitioner
HCP	Health Care Provider
HDI	Health Data Intermediary
HIMSS	Health Information Management System Society
HIPAA	Health Insurance Portability and Accountability Act
HL7	Health Level 7

Abbreviation	Full Name
ICD	International Classification of Disease
ICF	Informed Consent Form
IPR	Intellectual Property Rights
IPS	International Patient Summary
KER	Key Exploitable Results
KG	Knowledge Graph
LOINC	Logical Observation Identifiers Names and Code
MDR	Medical Device Regulation
ML	Machine Learning
NLP	Natural Language Process
NLP	Natural Language Processing
OMOP	Observational Medical Outcomes Partnership
PDEC	Plan for the Dissemination and Exploitation including Communication
PHD	Personal Health Data
PHI	Protected Health Information
PHKG	Personal Health Knowledge Graph
PII	Personal Identifiable Information
RDF	Resource Description Framework
RPA	Remote Process Assistant
RWD	Real World Data
SAB	Sustainability Advisory Board
SC	Steering Committee
SDLC	Software Development Life Cycle
SDTM	Study Data Tabulation Model
SHACL	Shapes Constraint Language
ShEx	Shape Expressions
SNOMED CT	Systematised Nomenclature of Medicine Clinical Terms
TA	Therapeutic Area
TRL	Technology readiness levels
TTP	Trusted Third Party
UMLS	Unified Medical Language System

1. Executive Summary

Manual annotations of clinical narratives are crucial for the adoption and evaluation of NLP tools, which support an overall AI assisted data curation approach within the AIDAVA project. In the preparation phase - in scope of this deliverable - for the Task “T4.3 Manual Annotation of text documents in 3 languages”, and based on the data elements identified for the use cases cross border breast cancer patient registries, and longitudinal individual health records for patients at risk of sudden cardiac arrest, requirements for the manual annotation tool have been formulated. Grounded on the requirement analysis, INCEpTION was chosen to support the manual annotation task. A first manual annotation schema was developed and tested, with a focus on the use of SNOMED CT and FHIR for the normalized form of the entity types of interest. A first version of the annotation guidelines is drafted in this document and will be revised in close cooperation with the manual annotators at the three different clinical sides (Med Uni Graz with MUG, Northern Estonian Medical Center with NEMC, Maastricht Medical University Center with UM), AVER and ONTO during the piloting phase until Q1 2023.

2. Introduction

AIDAVA pursues the goal to represent health data of an individual in a consistent semantic model, rooted in international standards for electronic health records (EHRs) committed to the FAIR (findable, accessible, interoperable, reusable) principles of data stewardship.

The challenging fact that large parts of EHR content is only available in narrative form in the local language, e.g., as findings reports or discharge summaries, is addressed by AIDAVA’s focus on diverse artificial intelligence technologies, aiming at interoperable and reusable health records.

In particular, this requires the use of existing and new tools for natural language processing (NLP) and machine learning (ML). Making narrative EHR content interoperable means to relate pieces of text to representational units (codes, relations) from controlled vocabularies and information models and attaching them to the patient data in the form of a personal knowledge graph. This is commonly understood as a combination of three common NLP techniques:

1. Entity recognition (ER): a text span (“entity”) is identified to represent an entity of meaning (commonly referred to as a concept) [1].
2. Entity normalization (EN): the concept referred to by the entity is identified in a controlled vocabulary and mapped to a language-independent code (from a standard ontology or thesaurus) [2].
3. Relation extraction (RE): semantic or linguistic relations are identified between entities [3].

AIDAVA aims at training deep learning models and combining them with existing NLP pipelines in order to perform the tasks ER, EN and RE. This requires annotation of pseudonymized [4] clinical narratives to be used as training and testing data sets. Pre-annotation will play a major role and is expected to speed up the manual annotation process. It will, however, depend on the availability of existing NLP pipelines and lexicons.

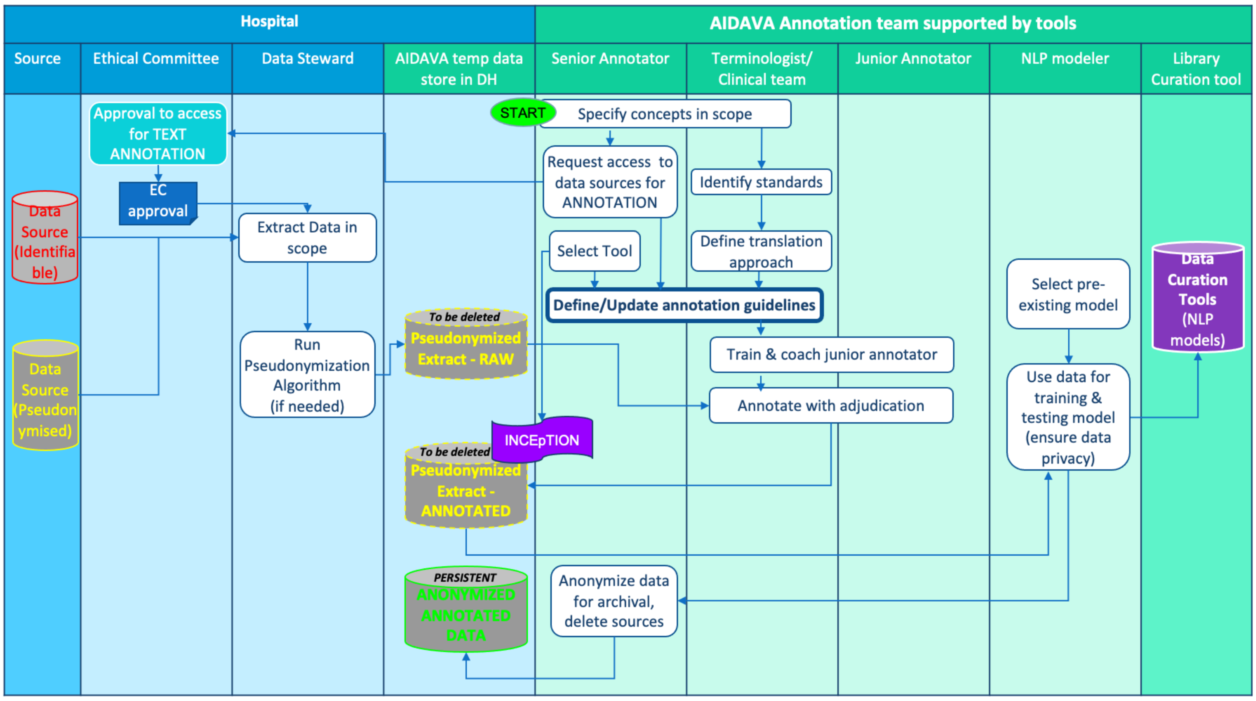

Hospital
AIDAVA Annotation team supported by tools
Source
Ethical Committee
Data Steward
AIDAVA temp data store in DH
Senior Annotator
Terminologist/ Clinical team
Junior Annotator
NLP modeler
Library Curation tool
START
Specify concepts in scope
Approval to access for TEXT ANNOTATION
Request access to data sources for ANNOTATION
Identify standards
EC approval
Data Source (Identifiable)
Extract Data in scope
Select Tool
Define translation approach
Define/Update annotation guidelines
Select pre-existing model
Data Curation Tools (NLP models)
To be deleted
Pseudonymized Extract - RAW
Run Pseudonymization Algorithm (if needed)
Train & coach junior annotator
Data Source (Pseudonymised)
Annotate with adjudication
Use data for training & testing model (ensure data privacy)
INCEpTION
To be deleted
Pseudonymized Extract - ANNOTATED
PERSISTENT
ANONYMIZED ANNOTATED DATA
Anonymize data for archival, delete sources

Figure 1. Overview of the annotation process.

As visualized in Figure 1 above, the annotation of clinical narratives is a multistep process that consists of the following steps with respect to the *preparation phase* (see Section 3):

Specify concepts in scope. Concepts and data elements that are expected to be extracted from the narratives. In our context, data elements are related to Breast Cancer and the risk of myocardial infarction in patients with Cardiovascular disease (see Section 4.1). It is expected to have a minimal set of entities identified, which are needed for both use cases (see Section 4.1).

Request access to data sources. Send a request to access clinical narratives out of the hospital information system to the local Ethical Committee (MUG, NEMC, UM). For the *preparation phase* (see Section 3) it is expected to have the approved local Ethical Committee votes (**Approval to access for text annotation, EC approval**) and some narrative data in scope of the use cases can be used for the first iteration of the creation of the manual annotation guidelines.

Extract data in scope. Once granted, extracts of (potentially identifiable) data can be provided by the hospital; technical identifiers of these extracts must be pseudonymized to limit the risk of a data privacy breach (see Section 4.2). If needed, a de-identification tool (see Section 4.2) will remove PHI data (**Run pseudonymization algorithm if needed).**

Data source - identifiable. Heterogenous (standardization, structure) EHR data which is identifiable. **Data source - pseudonymized.** Heterogenous (standardization, structure) EHR data. Technical identifiers have been assigned a pseudonym. **Pseudonymized extract - raw.** Containing narrative data which can be used for the manual annotation process.

Select tool. Out of a set of available manual annotation tools, the best one fitting the needs of this project has to be chosen. The selection of a manual annotation tool *and* its technical deployment at all three clinical partner sites is expected. The motivated selection approach (see Section 4.5) identified **INCEpTION** as the best choice, taking project driven requirements under consideration in combination with the support of pre-annotations (see Section 4.4)

Identify standards. To support the FAIR principles of data handling and the interconnected use cases between the clinical sites, the right choice of standard is of utmost importance. With respect to the manual annotation task, the manual annotation model (see Section 5.1.5) has to support the normalized form of an entity of interest. Standards considered so far in the *preparation phase* are SNOMED CT, ICD-10, LOINC and HL7 FHIR, acting as target information models.

Define translation approach. To have a full language support of the three languages of interest (German, Dutch, Estonian) and how to deal with multilingualism has to be considered in preparing the annotation guidelines. For the task "T4.3 Manual Annotation of text documents in 3 languages" the enrichment of terminologies with synonyms will be supported. A more detailed description can be found in Section 4.3.

Define / update annotation guidelines. Define (or update existing) annotation guidelines to ensure that all teams across the involved sites extract concepts in the same way. These annotation guidelines are used to **train & coach junior annotators** and to monitor the annotation process (**annotate with adjudication**) using narratives from **pseudonymized extract - annotated**. The detailed description of the guideline in its first version is described in Section 5.

The annotations are used for named entity recognition and named entity normalization (**use data for training & testing model - ensure data privacy**), therefore supporting the adoption of existing NLP pipelines (Averbis Health Discovery (HD), OntoText, **select pre-existing model**) in the generation of knowledge graphs which are leveraged by **data curation tools**. If needed by the local clinical sites from the ethics vote, the procedure is to **anonymize data for archival,** resulting in a persistent **anonymized annotated data** store.

3. Description of Activities

This guideline is the first release of an iterative process; an updated version will be released in April 2023 (Deliverable 4.3). We are working through the following phases:

Preparation phase. Contextualize the interplay between human / machine annotation with the overall knowledge ecosystem to support the defined use cases. Confirm the main components needed for the annotation process and design a first draft of the annotation guidelines, populated within this document, with a focus on the use of SNOMED CT. Submit a request to access the documents of relevance to the corresponding ethics committees (MUG, NEMC, UM) and get approval. Prepare secure IT infrastructure according to manual annotation tool deployment and test its functionality. Preparation / planning of the human annotators based on a first draft of the annotation guidelines. The preparation phase lasts from 09 2022 - 12 2022 and is confirmed with deliverable "D4.1 Annotation guidelines, tools & training".

Piloting phase. According to the approved ethics at the clinical sites, a first set of clinical narratives is extracted from the corresponding hospital information systems. The manual annotation process is started based on the first draft of the annotation guidelines, the process monitored qualitatively (throughput and inter-rater agreement) and disagreements adjudicated. The annotation guidelines are iteratively revised and manual annotators re-instructed in accordance with the updated guidelines. The preparation phase lasts from 01 2023 - 04 2023 and is confirmed with deliverable "D4.3 Update to Annotation guidelines, tools & training".

Productive phase. All relevant documents needed for the manual annotation process are exported from the clinical partners and imported to the manual annotation tool. The manual annotation process is started based on the updated annotation guidelines of the piloting phase. This process is monitored qualitatively (throughput and inter-rater agreement) and disagreements adjudicated. If still needed, the annotation guidelines are iteratively revised and manual annotators re-instructed in accordance with the updated guidelines. The productive phase lasts from 05 2023 - 04 2024 and is confirmed with deliverable "D4.7 Annotated datasets (3 languages/2 TA) with report".

4. Prerequisites to Annotation

4.1. Define domain and scope

Medical experts in breast cancer and cardiovascular diseases (CVD) identified a set of data elements that they need to gather from medical records to perform further analysis: 110 and 150 data elements were identified respectively for breast cancer and CVD use case, and mapped with SNOMED codes. Out of these, some data elements were selected for the preparation phase to cover different aspects (key symptoms, data elements with proprietary code list, lab value, etc.):

- T, N, M stages and type of surgery for breast cancer
- blood pressure (arterial hypertension), cardiac arrest and total cholesterol for CVD

4.2. Ethical committee approval and data handling

Access to clinical narratives from patients requires approval from an Ethical Committee (Institutional review board). Each participating site (NEMC, UM, MUG) filed a request to their local EC based on generic information provided by the AIDAVA ELSI team and information required locally. Approval was granted at each site within 3 to 6 weeks. A key element in the EC request is the description of data handling, as described below.

It is expected that 4,000 documents with **de-identified documents**[3] per use case (breast cancer, cardiac arrest) will be needed for each language. Given the common copy & paste practice in patient notes [5], several documents per patient are often not that useful for machine learning; we instead suggest using a number close to 1,000 patients with 4 documents rather than, for instance, 200 patients with 20 documents each. The training documents should be maximally representative regarding the attributes and values specified by the knowledge graph models as specified in Task 1.1, such as pTNM, grading and staging for breast cancer, as well as medications, therapies, and diagnoses. It is expected to use full documents with narrative content such as discharge summary, pathology report, doctor’s letter to patient, and referral.

The extraction, transform, and load (ETL) process of the clinical narratives is performed by a data curator who has institutional rights and technical access to the clinical information system and who ensures data extracts are stored in a secure data-lake on premise (see below) as data sink.

Some institutions have access to already pseudonymized data, others have to ensure proper pseudonymization and protection of data privacy. In the case where pseudonymization is needed, the de-identification process is part of this initial ETL process, as described below.

Data de-identification process. Out of the 18 Health Insurance Portability and Accountability Act (HIPAA) criteria [6] for Protected/Personal Health Information (PHI), a focus is set on names (HIPAA identifier 1), E-mail addresses (HIPAA identifier 7), and medical record numbers (HIPAA identifier 8). Two possibilities for the de-identification process are feasible: either exclusively manual or a toolset-supported process. In the tool-supported approach, PHI-related information in the narrative is identified and correspondingly masked (see Figure 1); this process can be complemented by a manual cross-check of the masked PHI.

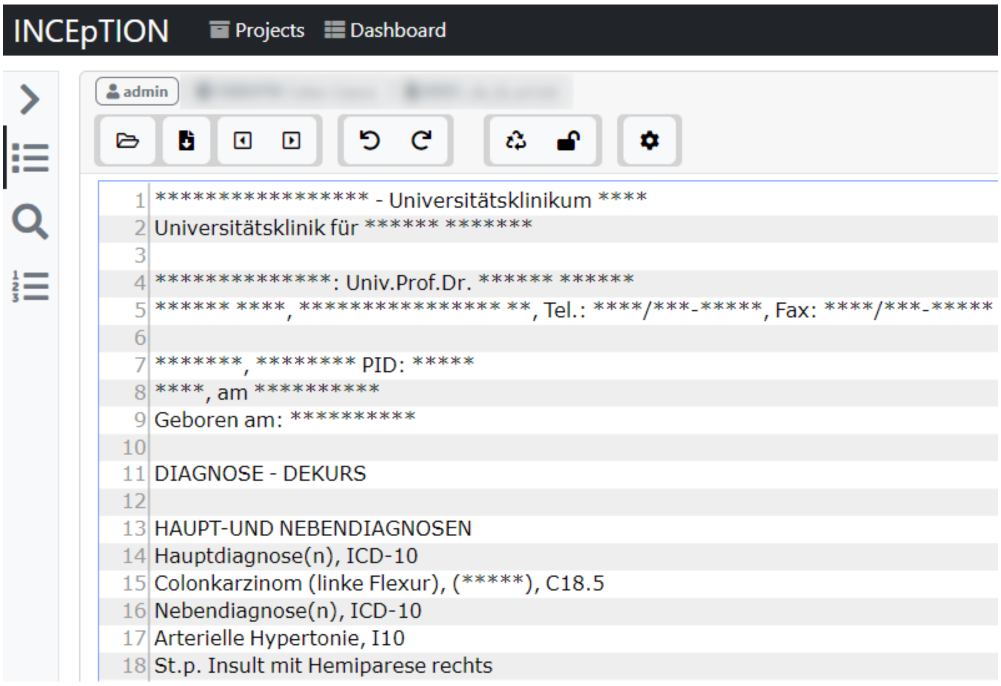
If the size of needed narrative extracts is too big for a full manual cross-check, a representative sample size of 100 will be used to check the risk of re-identification. The tool-set supported de-identification process is assessed by a qualitative ratio (number of documents correctly de-identified / number of

[3] While 4,000 documents is a reasonable estimate, it is difficult to give a general number of documents because it always depends on the concrete NLP task. Roughly, one needs about 50–100 examples for each label that one wants to predict. For yes/no decisions (e.g., smoker yes/no) 100 examples (=50 docs each about smoker/non-smoker) may be sufficient. To predict the complete ICD-10 with its more than 10,000 codes, we would need around 500,000 examples. With about ten diagnoses per document, this would correspond to 50,000 documents.

documents). This ratio shall reach a level of 95% percent, or the automated de-identification process has to be refined until a random sample of 100 documents reaches this level.

Identifiable data, including case and hospital numbers, can be assigned a pseudonym which can only be linked back by personnel who already have access to the identifiable information. Additionally, all personnel accessing the narratives must sign a policy and disclaimer and must operate under their existing contractual obligations, where re-identifying the data is expressly forbidden. Therefore, the operational data sets in the project scope cannot be linked to an individual without breaching legal and contractual obligations or requiring special legislation.

Between the technical and organizational measures, the risks of re-identification are therefore minimized at an operational level. The risks of re-identification are further minimized by ensuring that data is de-identified and processed without leaving its host institution.



[redacted]

***************** - Universitätsklinikum ****

Universitätsklinik für ****** *******

**************: Univ.Prof.Dr. ****** ******

****** ****, **************** **, Tel.: ****/***-*****, Fax: ****/***-*****

*******, ******** PID: *****

****, am **********

Geboren am: **********

DIAGNOSE - DEKURS

HAUPT-UND NEBENDIAGNOSEN

Hauptdiagnose(n), ICD-10

Colonkarzinom (linke Flexur), (*****), C18.5

Nebendiagnose(n), ICD-10

Arterielle Hypertonie, I10

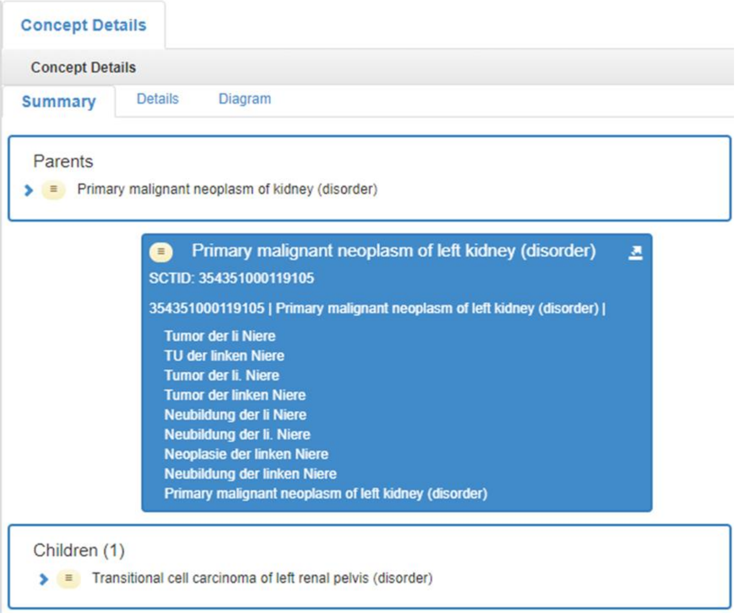
St.p. Insult mit Hemiparese rechts

Figure 2. Result of a de-identified document, imported to INCEpTION.

Specification of the data lake on-premises for the de-identified data sources.

To ensure that documents and data do not leave the institution, a dedicated server is installed. The server is integrated into the network of the respective institution and is thereby closed to ungranted access from outside and is subject to the respective security guidelines. Authorized users for processing the data are exclusively activated, and the activation of users is logged. All analyses (e.g., the manual annotation of entities of interest via INCEpTION) and results are located exclusively within this defined area.

4.3. Translation approach



Parents

Primary malignant neoplasm of kidney (disorder)

Primary malignant neoplasm of left kidney (disorder)

SCTID: 354351000119105

354351000119105 | Primary malignant neoplasm of left kidney (disorder) |

Tumor der li Niere
TU der linken Niere
Tumor der li. Niere
Tumor der linken Niere
Neubildung der li Niere
Neubildung der li. Niere
Neoplasie der linken Niere
Neubildung der linken Niere
Primary malignant neoplasm of left kidney (disorder)

Children (1)

Transitional cell carcinoma of left renal pelvis (disorder)

Figure 3. Example of additional synonyms for a concept defined in SNOMED CT.

As the project as well as the annotation guideline defined in this document should be applicable for three different languages (German, Dutch, Estonian), considerations about translations have to be done concerning clinical narrative data as well as the used terminologies ICD-10, LOINC, SNOMED CT, and the target information model FHIR. The following decision has been made with respect to “T4.3 Manual Annotation of text documents in 3 languages”.

Narrative data. The narratives will not be translated into a target language, but should be manually annotated according to the defined INCEpTION annotation model as defined in Section 5.1.5. The use of external automatic translation services, contradicts the on-premises setting for clinical data processing for the manual annotation process as well as the quality of the translations have to be assessed before using such a service due to the idiosyncratic nature of clinical narratives [7].

Terminology resources. Translation approaches of standardized descriptions of an entry of a terminology (ICD-10, SNOMED CT, LOINC) will not be part of Task 4.3, but the corresponding terminology versions in English will be used in the first iteration. The generation of layman expressions of terminology entries are not considered to be placed in Task 4.3. The enrichment of synonym expressions as used as part of clinical routine documentation, for example in the case of SNOMED CT corresponding to a preferred name, will be part of Task 4.3. Acronyms and abbreviations are treated in that case as synonyms, as shown in Figure 3.

4.4. Pre-annotation process

To speed up the manual annotation process under the condition of an unbiased stable inter-rater agreement [8], pre-annotations [9] shall be integrated into the manual annotation tool, which has to be manually confirmed by the user if correct or not. The pre-annotations have not to be necessarily involved in an active learning approach [10] but should support the manual annotator in finding and confirming entity types of interest regarding the specified use cases. As this process influences confirming true positives and identifying false positives, false negatives from the pre-annotations still have to be carefully considered. Pre-annotations should be supported by the already existing component repository available within the Averbis HD.

Available components within the **Averbis HD pathology pipeline**
Generic terminology annotator (SNOMED CT, requires custom vocabulary), ClinicalSections, Laterality, PatientInformation, TNM, LabValues, Diagnoses, Topography, Morphology, GleasonScore, Enumerations, Negations, TumorStage, Receptors, Specimen, DiagnosisStatus, Disambiguation, PathologyDocumentationClassification, HealthPostprocessing

Available components within the **Averbis HD discharge pipeline**
Generic terminology annotator (SNOMED CT, requires custom vocabulary), ClinicalSections, Laterality, PatientInformation, Organizations, PhysicalTherapies, LabValues, Diagnoses, Procedures, Medications, Enumerations, Negations, DiagnosisStatus, Procedure, Disambiguation, MedicationStatus, HealthPostprocessing

For the integration of pre-annotations, a type-system conversion from the Averbis HD to the project specific INCEpTION type system has to be done (layer and feature definitions). Averbis HD provides specific python libraries which support this conversion. The resulting filled data model can be imported into INCEpTION providing the user with already annotated entities of interest with their normalized form.

A pre-condition for the support of this feature is a clear definition of the consolidated annotation model (layer and feature definitions), which has to be used at all clinical partner sites for consistency, and a first version is populated within this document.

4.5. Selection of the tool

4.5.1. Key requirements for the manual annotation tool

Neves and Ševa [11] made an extensive review of 78 manual annotation tools according to the following four categories: publication, technical, data, and functional criteria, listed in more detail in Table 1.

Publication criteria

- P1 - Year of last publication
- P2 - Citations in Google Scholar (Sep 2019)
- P3 - Citations for corpus development (Sep 2019)

Technical criteria

- T1 - Date of the last version (Aug 2019)
- T2 - Availability of the source code
- T3 - Online availability for use
- T4 - Easiness of installation
- T5 - Quality of the documentation
- T6 - Type of licence
- T7 - Free of charge

Data criteria.

- D1 - Format of the schema;
- D2 - Input format of the documents
- D3 - Output format for annotations

Functional criteria.

- F1 - Support of multi-label annotations
- F2 - Support of document-level annotations
- F3 - Support for annotation of relationships
- F4 - Support for ontologies and terminologies
- F5 - Support for pre-annotations
- F6 - Integration with PubMed
- F7 - Suitability for full texts
- F8 - Support for saving documents partially
- F9 - Ability to highlight parts of the text
- F10 - Support for users and teams
- F11 - Support for inter-annotator agreement (IAA)
- F12 - Data privacy
- F13 - Support for various languages

Table 1. Evaluation criteria for manual annotation tools under investigation [11].

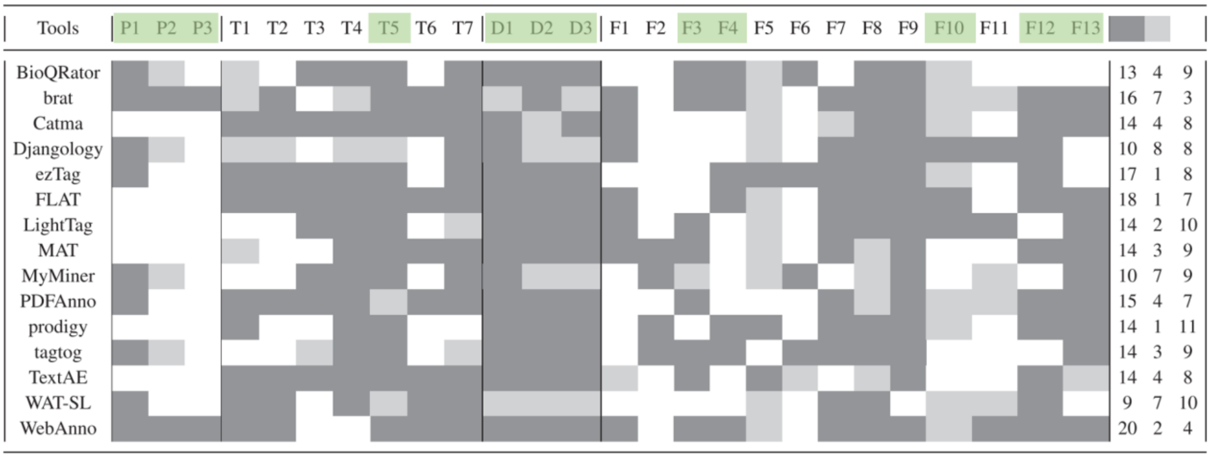
The following table lists requirements needed at a specific necessity level (**M**: must-have requirement. **S**: should-have requirement. **C**: could-have requirement. **W**: will not have requirements. [12]) for the manual annotation tool within AIDAVA and tries to map them to the specification from [11] listed before.

Requirement	Justification	Necessity	Criterion [11]
Parametrization of look and feel of annotations	Support of usability and joy of use.	C	-
Easy installation	The installation process should be straightforward and documented.	C	T4
Web-based deployment	The interaction of the tool should be supported via a web-based view.	M	F10
Collaborative working supported	Annotators can jointly work on a specific set of documents.	M	F10
Assignment of users and roles	User and role management should be supported by the tool. E.g., exclusive view on annotations for inter-rater agreement.	M	F10
Active maintenance and development	The tool should be under active maintenance and development, also due to security issues of the type of data used in the project.	C	T3
Documentation available	Proper written documentation has to be available for all functional aspects of the annotation tool.	M	T5
Definition of annotation schema	Entities of interest (types, e.g., diagnosis), together with specific features (e.g., normalized form according to a terminology in use - SNOMED CT), can be parametrized together with contextual patterns (e.g., negation, experiencer).	M	F4, D1, D2, D3
Sharing of annotation schema	The consolidated annotation schema should be used by all partners in three languages.	M	F10, D1, D3
Multi-language support	Narratives in different languages have to be manually annotated (Dutch, German, Estonian).	M	F13
Annotation of relations	The annotation of relations between entities has to be supported, as well as the type of the relation according to the entities involved.	M	F3
Annotation of entity	See “Definition of annotation schema”.	M	F4, D1, D2, D3
In use by other applied research projects	The use of the tool by others confirms its proven usefulness and applicability.	M	P1, P2, P3
Annotation of normalized form	See “Definition of annotation schema”.	M	F4, D1, D2, D3
Document level annotations	In the project scope, meta information of documents like document type e.g., discharge summary or medical speciality e.g., oncology are additional values of interest to be annotated. Most of the annotations will be on entity level with one document.	S	F2
Active learning functionality	Based on validated manual annotations, the tool suggests entities of interest in the document which shall be annotated. The tools learn from confirmed manual annotations.	C	F5
Pre-annotations available	Static pre-annotations can boost the manual annotation throughput. The suggested tool-based pre-annotations have to be confirmed by the user.	S	F5

Requirement	Justification	Necessity	Criterion [11]
Value set navigation	Values sets out of a terminology can be easily chosen and navigated within the manual annotation tool to support medical concept normalization of an entity of interest	S	F4
Open source	The annotation tool should be free of charge	M	F13
Consideration of data privacy	In the scope of the project, clinical documents will be annotated, therefore considerations of data privacy for the manual annotation tool should be taken into account.	M	F12
Checking annotation quality	Support for, e.g., inter-rater agreement statistics	S	F11

Table 2. Requirement mapping.

4.5.2. Comparison of tools


Tools
P1 P2 P3
T1 T2 T3 T4 T5 T6 T7
D1 D2 D3
F1 F2 F3 F4 F5 F6 F7 F8 F9 F10 F11 F12 F13

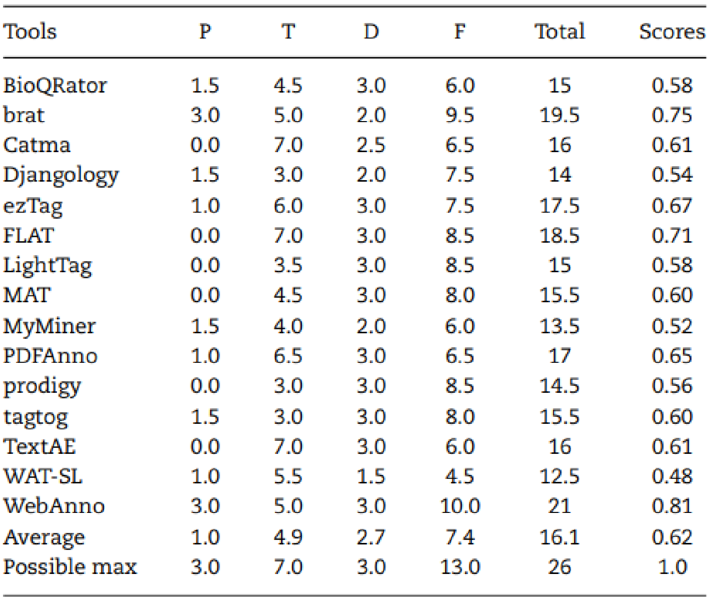
Tools	■	▒	
BioQRator	13	4	9
brat	16	7	3
Catma	14	4	8
Djangology	10	8	8
ezTag	17	1	8
FLAT	18	1	7
LightTag	14	2	10
MAT	14	3	9
MyMiner	10	7	9
PDFAnno	15	4	7
prodigy	14	1	11
tagtog	14	3	9
TextAE	14	4	8
WAT-SL	9	7	10
WebAnno	20	2	4

Table 1. Evaluation visualization according to [11]. Mandatory functionality for the project is highlighted in Green.

According to the review from [11] (see Table 2) WebAnno [13,14] got 0.81 points out of a maximum of 1 regarding the criteria stated in Section 4.5.1. Based on this, in combination with supporting 11 out of 12 mandatory requirements, one requirement fulfilling partially, **the choice to WebAnno's successor, INCEpTION** [15]**, was made** fulfilling the same required specifications. In addition, the maintainer of INCEpTION is employed at the NLP industry partner AVER, therefore supporting short possible additional requirement loops within the interplay of gold standard creation and NLP pipeline adaptation according to the use cases. As both solutions are built on top of UIMA and its standardized type system representation of entities of interest, the choice supports a common technology framework.

The brat rapid annotation tool [16] reaching a score of 0.75 (see Table 2), though still used in the scientific community and despite its easy-to-use and straight forward deployment process, was neglected as it is not maintained any more since 2012. Due to the fact that the project works with sensitive data, security updates and maintenance are of utmost importance. brat fulfilled 4 out of 12 mandatory requirements partially and the remaining 8 sufficiently.

Another tool in inspection because of its transformer based [17] active learning possibilities was prodigy [18], reaching an overall evaluation score of 0.56 (see Table 2). As it is not free of charge and mainly supports annotation at the level of named entities (e.g., diagnosis, medications), but also the normalized forms are needed in the manual annotation process, the tool was finally not taken into account. prodigy fulfilled 1 out of 12 mandatory requirements partially, 7 sufficiently and 4 requirements could not be fulfilled.



Tools	P	T	D	F	Total	Scores
BioQRator	1.5	4.5	3.0	6.0	15	0.58
brat	3.0	5.0	2.0	9.5	19.5	0.75
Catma	0.0	7.0	2.5	6.5	16	0.61
Djangology	1.5	3.0	2.0	7.5	14	0.54
ezTag	1.0	6.0	3.0	7.5	17.5	0.67
FLAT	0.0	7.0	3.0	8.5	18.5	0.71
LightTag	0.0	3.5	3.0	8.5	15	0.58
MAT	0.0	4.5	3.0	8.0	15.5	0.60
MyMiner	1.5	4.0	2.0	6.0	13.5	0.52
PDFAnno	1.0	6.5	3.0	6.5	17	0.65
prodigy	0.0	3.0	3.0	8.5	14.5	0.56
tagtog	1.5	3.0	3.0	8.0	15.5	0.60
TextAE	0.0	7.0	3.0	6.0	16	0.61
WAT-SL	1.0	5.5	1.5	4.5	12.5	0.48
WebAnno	3.0	5.0	3.0	10.0	21	0.81
Average	1.0	4.9	2.7	7.4	16.1	0.62
Possible max	3.0	7.0	3.0	13.0	26	1.0

Table 2. Summary of the evaluation with scores [11].

5. Annotation Instructions

5.1. Tool configuration and use

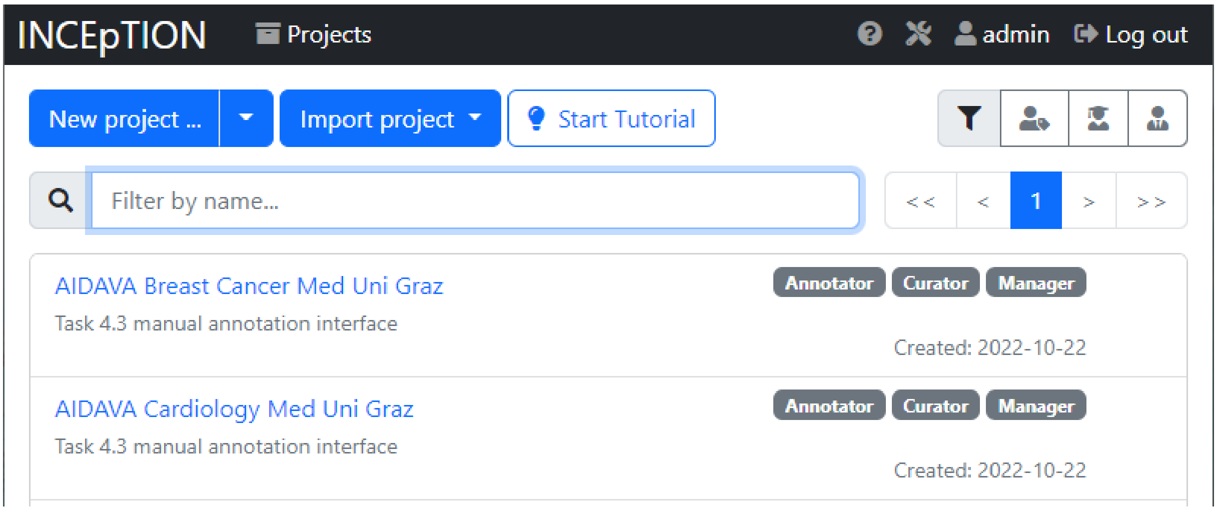
5.1.1. Deployment

Based on the decision motivated in Section 4.5 deployment of INCEpTION needs JAVA 11 or higher. As stated in "Download and start INCEpTION - Step 2b - Open via terminal" the program can be run via the following command line tool: `java -jar inception-app-standalone-x.xx.x.jar`

The deployment of the server on-premises has to fulfill local and approved regulations regarding data handling in the project context of AIDAVA.

5.1.2. Configuration and initial set-up

Project creation. After initial deployment of the tool, two projects are created according to the following naming conventions "AIDAVA Breast Cancer LOCATION SIDE" (e.g. AIDAVA Breast Cancer Med Uni Graz) and "AIDAVA Cardiology LOCATION SIDE" (e.g. AIDAVA Cardiology Med Uni Graz). Link to detailed documentation - Project Settings



AIDAVA Breast Cancer Med Uni Graz

Task 4.3 manual annotation interface

Created: 2022-10-22

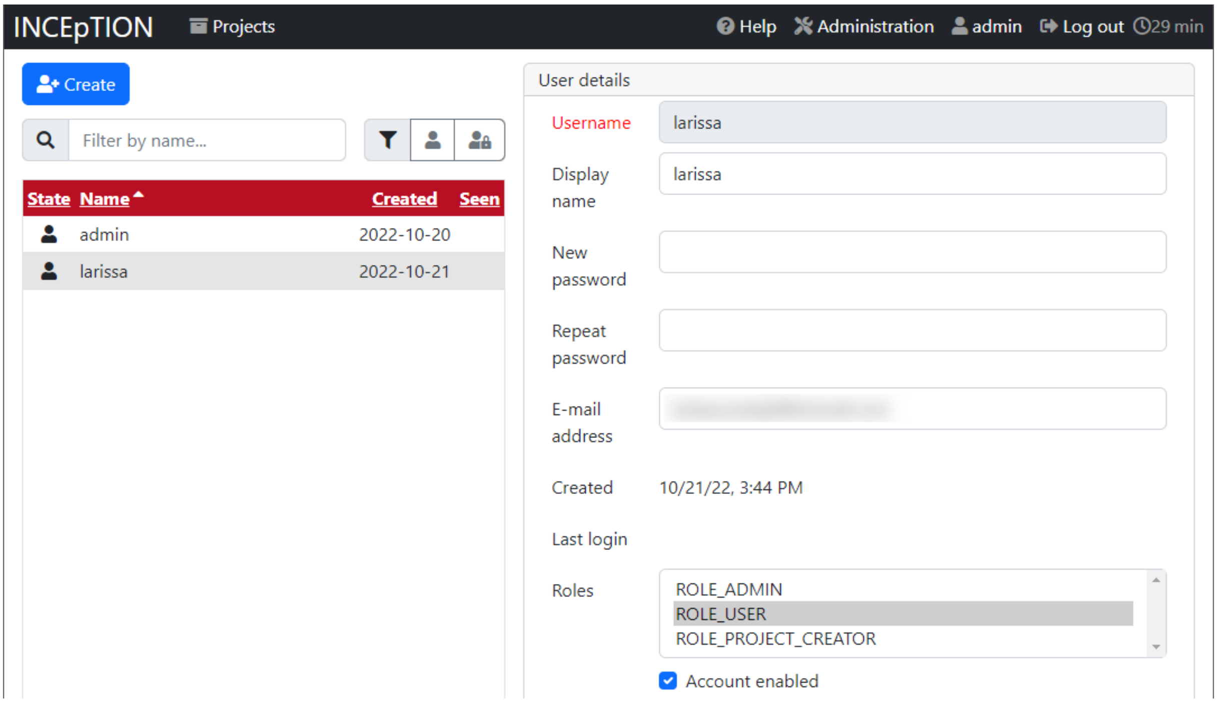
AIDAVA Cardiology Med Uni Graz

Task 4.3 manual annotation interface

Created: 2022-10-22

Figure 1. Initial project configuration - project creation.

User creation. Before one can add a user with a specific role to a project, the user has to be created. INCEpTION : Administration: Users. Link to detailed documentation - User Management



State	Name	Created	Seen
	admin	2022-10-20	
	larissa	2022-10-21	

User details

Username: larissa

Display name: larissa

New password:

Repeat password:

E-mail address: [redacted]

Created: 10/21/22, 3:44 PM

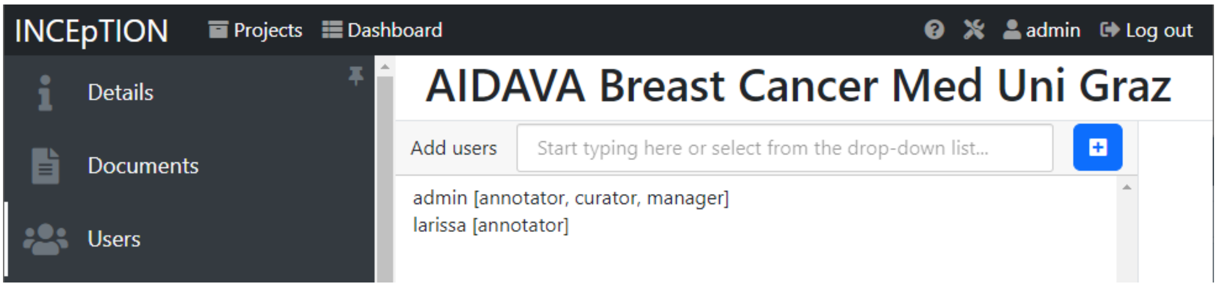
Last login:

Roles: ROLE_ADMIN, ROLE_USER, ROLE_PROJECT_CREATOR

☑ Account enabled

Figure 2. Initial project configuration - user creation.

Adding users. Navigating to INCEpTION : Settings : Users, a user can now be assigned to a project. Link to detailed documentation - Adding users to a project

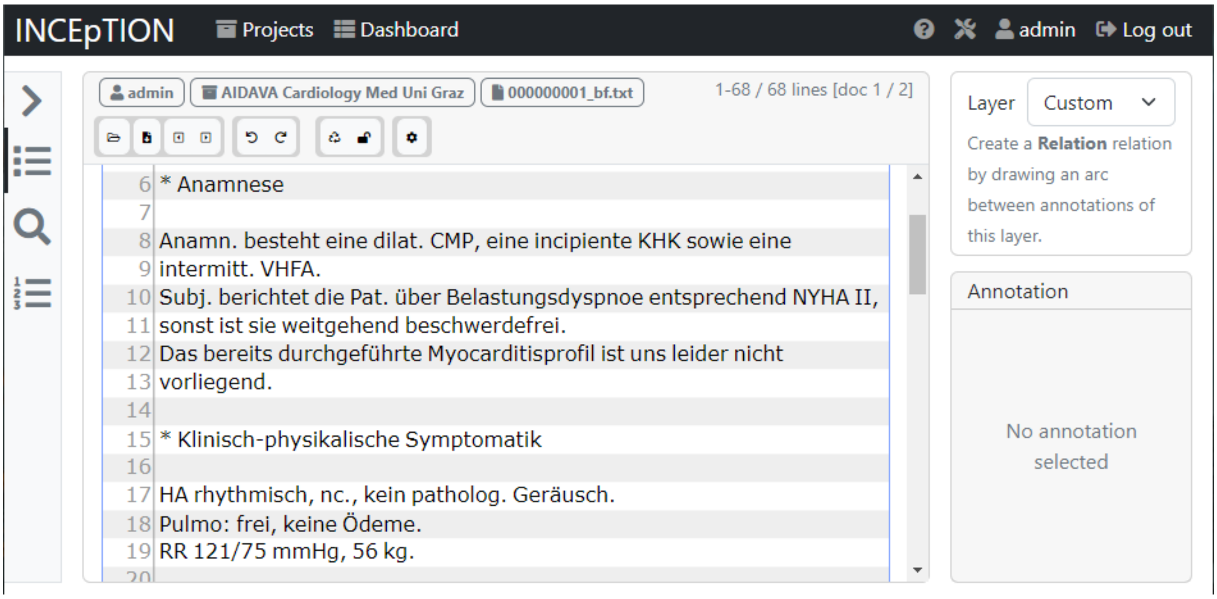


AIDAVA Breast Cancer Med Uni Graz

admin [annotator, curator, manager]
larissa [annotator]

Figure 2. Initial project configuration - assignment of a user to a project.

Import documents. At this step, the documents of the application domain (breast cancer and cardiology) have to be imported in the corresponding projects. The preferred format is .txt (Plain text from the menu). INCEpTION : Settings : Documents : Files to import. Link to detailed documentation - Documents



INCEpTION
Projects
Dashboard
admin
Log out
admin
AIDAVA Cardiology Med Uni Graz
000000001_bf.txt
1-68 / 68 lines [doc 1 / 2]
Layer
Custom
Create a Relation relation by drawing an arc between annotations of this layer.
Annotation
No annotation selected

* Anamnese

Anamn. besteht eine dilat. CMP, eine incipiente KHK sowie eine
intermitt. VHFA.
Subj. berichtet die Pat. über Belastungsdyspnoe entsprechend NYHA II,
sonst ist sie weitgehend beschwerdefrei.
Das bereits durchgeführte Myocarditisprofil ist uns leider nicht
vorliegend.

* Klinisch-physikalische Symptomatik

HA rhythmisch, nc., kein patholog. Geräusch.
Pulmo: frei, keine Ödeme.
RR 121/75 mmHg, 56 kg.

Figure 3. Imported plain text document.

Import annotation layers. The consolidated layers containing the annotation model are imported at this step. A first description of the annotation model can be found in Section 5.1.5. INCEpTION : Settings : Layers. Link to detailed documentation - Layers

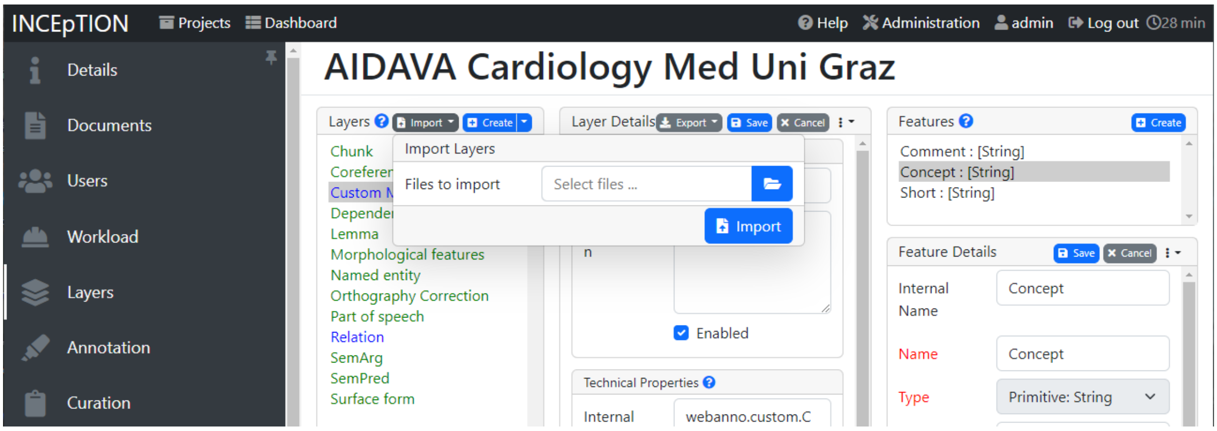
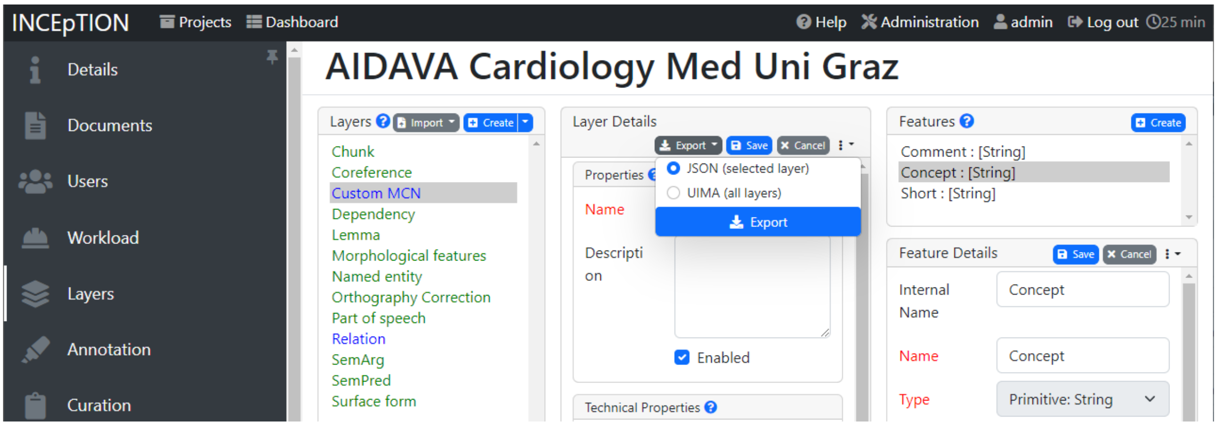


Figure 4. Import of the annotation model into INCEpTION.

5.1.3. Continuous update of configuration

Most important is the consolidated use of the annotation model (layers and features). This requirement is supported via the export and import functionality of the Layers' configuration side. INCEpTION : Settings : Layers. The annotation model versions will be distributed via the Med Uni Graz ownCloud functionality, with link sharing to the clinical sides. Link to detailed documentation - Layers



Chunk
Coreference
Custom MCN
Dependency
Lemma
Morphological features
Named entity
Orthography Correction
Part of speech
Relation
SemArg
SemPred
Surface form

Name
Description
☑ Enabled

Comment : [String]
Concept : [String]
Short : [String]

Internal Name: Concept
Name: Concept
Type: Primitive: String

Figure 5. Export of the annotation for consolidated distribution and updates.

5.1.4. Monitoring of annotation process

Monitoring of the manual annotation status is supported via INCEpTION and supports the following states: Done / In Progress or between New / Locked. In addition, statistical quality measurements of the manual annotations are supported [19]. Link to detailed documentation - Workload; Link to detailed documentation - Agreement.

5.1.5. INCEpTION annotation schema

The annotation schema in INCEpTION is split into layers. In the first iteration, we define two layers: a Custom MCN layer and a Relation layer.

Custom MCN layer. The Custom layer has the following features, with an emphasis on the possibility to annotate the normalized form according to the following resources: LOINC, SNOMED CT, FHIR. The extent to which ICD-10 and others will also be used still has to be defined. See detailed examples of use in the Annex of the document.

Feature	Description
Concept : [String]	Standardized description, e.g. SNOMED CT fully specified name (“241998008 |Cardiovascular decompression injury (disorder)|”). The line contains the ID, the preferred named and the semantic tag in parentheses, equivalent to the entity type.
Short : [String]	Short readable form for visualization within INCEpTION.
Comment : [String]	Additional comments for this concept annotation.

Table 3. Feature definitions of the custom layer.

Relation layer. Relations between manually annotated entity types of interest in combination with their normalized form. The following features are therefore defined within the Relation layer. See detailed examples of use in the Annex of the document.

Feature	Description
Relation : [String]	Standardized description, e.g. SNOMED CT fully specified name (“363698007 |Finding site (attribute)|”). The line contains the ID, the preferred named and the semantic tag in parentheses, equivalent to the entity type.
Short : [String]	Short readable form for visualization within INCEpTION.
Comment : [String]	Additional comments for this relation annotation.

Table 4. Feature definitions of the relation layer.

5.2. Annotation process

The annotation process is directed by an annotation guideline, currently in use at Med Uni Graz and will be additionally informed by the Google annotation guideline [20]. The testing and iterative refinement of the annotation process will be the focus in Q1 2023 and result in an updated version of the annotation guidelines presented in this document and the INCEpTION annotation schema.

5.2.1. Basic principles of annotation

Standards like SNOMED CT, LOINC, and FHIR promise internationally interoperable and computable representations of clinical content. This is the reason why the AIDAVA annotation process uses these resources. Bridging between human language and interoperable representations requires leveraging state-of-the-art technology in text mining and artificial intelligence, particularly machine learning and deep learning. This requires significant amounts of annotated clinical corpora. Good annotation should aim at:

- With the same input text, different human annotators produce the same target representation.
- With different paraphrases of the same clinical content, different human annotators produce a target representation for which semantic equivalence could be stated, e.g. by logical reasoning powered by logical axioms or by link predictions learnt from large knowledge graphs.
- With translation of the same clinical content to different human languages, different human annotators produce target representation for which semantic equivalence could be stated.

AIDAVA annotation will also depend on the reference knowledge graph schema as elaborated in WP2 Task 2.2 for the two use cases. Annotators should comply with this schema in terms of scope and granularity.

Annotation strategies have been highly diverse and not comparable, regarding aspects like:

- Whether spans to be annotated are defined by the annotators, by automatic named entity recognition techniques, or whether annotation happens at a token level
- Whether only entities (spans, words) are annotated or also relations between entities.
- Whether entity and relation types follow existing ontologies or are created ad-hoc, inspired by natural language predicates.
- The extent to which an annotation should take context into account (e.g. span “procedure” in: “after the procedure the patient was instructed to avoid…”), which refers to a more specific concept introduced before.

5.2.2. Conceptual model and definitions

We understand by “concepts” all units of non-relational meaning provided by SNOMED CT (terminology), LOINC (terminology) and FHIR (value sets). We understand by “entities” passages (words, phrases) in a text that are annotated by a concept. This implies that the entity of individual meaning that is referred to by the entity is identified as an instance of the concept. For instance, if the entity “ductal invasive breast cancer” occurs in a clinical text and is annotated with the SNOMED CT code “408643008 |Infiltrating duct carcinoma of breast (disorder)|”, this means that the concrete tumor instance (the referent) of the woman who is subject of the record is identified as the instance of that concept. If, in addition, the entity “pT3N2M0” is annotated by “1229859000 |American Joint Committee on Cancer pT3 (qualifier value)|, and 1229957002 |American Joint Committee on Cancer pN2 (qualifier value)|”, these two concepts are instantiated by qualities of this cancer. A relation annotation would then assert a directed relation, such as ‘has quality’ between the tumor and its TNM concepts.

We understand SNOMED CT and LOINC as ontologies. This means they provide codes (with associated labels, definitions and axioms), which delineate classes of individual things in the clinical domain. FHIR, instead, provides a structure to represent the clinical reality of individual patients expressing related temporal and epistemic contexts. To this end, terminology codes are bound to FHIR elements, but FHIR also provides its own value sets. Some of them are from SNOMED CT, others are mappable to SNOMED CT, others are completely proprietary.

The standard interpretation of an entity annotated by a terminology concept (not set in any FHIR context) is being instantiated by concept (referred to the annotation) is that in the portion of reality described by the document an instance of the concept exists during the episode of treatment and related to the subject of care, i.e., the patient the document is about.

Any deviation from this requires representational elements from FHIR, particularly regarding events that precede this episode of care, that are related to another subject, or express uncertainty or negation. We pursue the following rules:

1. There is no predetermined annotation span. The minimal span corresponds to one token. Spans have to be defined by the annotator. The span definition follows the identification of mentions that can be expressed by one clinically-relevant concept.
2. In case of content that can be represented by more than one terminology or value set, preference is given to FHIR. Where content can be expressed by SNOMED and LOINC, preference is given to SNOMED. (If mapping between two terminologies of value sets is necessary, this is done in a post-processing step.). Lab procedures are used as proxies for non-existing lab observables.
3. No external terminology is bound to FHIR other than SNOMED CT or LOINC.
4. Gaps in SNOMED CT have to be identified and recorded, as well as the need for post-coordination, which goes beyond the logical combination of codes.

5.3. Preferences regarding SNOMED content

5.3.1. Core concepts

Whenever possible, preference is given to SNOMED core concepts. We understand by “core concepts” those that are ideally fully expressive when they stand alone and do not modify other concepts (called “supportive concepts”).

Core concepts come typically from the following hierarchies:

- Clinical conditions (SNOMED findings / disorders / events); we prefer the term condition because of its use in FHIR. Most of them are partly or fully defined using a set of relations and concept types as prescribed by the SNOMED concept model.
- Procedures. Many of them are also fully defined.
- Observables, together with qualitative values or numbers (and units).
- Procedures that substitute observables (wherever concepts for lab observables are missing).
- Products.

Clinical conditions (SNOMED CT clinical findings) are typically modified by morphology, body structure, organism and device concepts, qualifier values; Procedures by body structure and device.
Core concepts are typically related to supportive concepts via outgoing relations (following the SNOMED concept model), such as:

Clinical condition → Causative agent → Organism
Administration of drug or medicament → Direct substance → Substance
Product containing only codeine → Has active ingredient → Substance

5.3.2. Modifier concepts

Concepts from other hierarchies should only be used in case they are clinically important and not expressible with the above hierarchies, and when the interpretation of other parts of the text depends on them. Qualifier values and units of measurement are used only when related to other concepts.

For those context attributes where pre-selections are available (from FHIR), no SNOMED CT codes are necessary.

5.3.3. Substances and products

Concepts from the SNOMED CT “pharmaceutical product” sub-hierarchy should be given preference over substances, particularly with combined drug products, e.g. “Folsan”. Substance concepts are used where in the document only the substance but not the specific product is used.

5.3.4. Time and dates, values of measurement, units

Numeric values are annotated using primitive data types. Units of measurement use values from UCUM.

5.3.5. Excluded SNOMED CT content

Concepts from the hierarchies Situation with explicit context (situation), Special concept (special concept) should be avoided, due to overlap with HL7 FHIR. From the hierarchy SNOMED CT Model Component (metadata) only the relations below Concept model object attribute (attribute) are used

5.3.6. SNOMED CT concepts that require values

Concepts of the hierarchy Observable entity are only used with a quantitative or qualitative value. Wherever this hierarchy does not provide a concept to express the measurement of something, use subconcepts of “785673007 |Measurement of level of substance in blood (procedure)|” instead.

A current drawback of Observables is that they are not related to their defining concepts, e.g. “446089006 |Volume of lower limb (observable entity)|” is not related to the lower limb. It is therefore undefined how to refine observables via post coordination, such as Volume of left lower limb. We here suggest that for laterality, the relation laterality is used in the same way as for body parts.

5.3.7. Relations, predicates and operators

If possible, only those SNOMED relations should be used that also occur in SNOMED concept definitions (below Concept model object attribute (attribute)). In case of doubt which relation to choose, look up similar concepts in SNOMED CT and follow the pattern they are defined.

Short	Long
Finding site	363698007 |Finding site (attribute)|
Finding method	418775008 |Finding method (attribute)|
Direct substance	363701004 |Direct substance (attribute)|
Temporally follows	363708005 |Temporally follows (attribute)|
Procedure site	363704007 |Procedure site (attribute)|
Clinical course	263502005 |Clinical course (attribute)|
Unit	767525000 |Unit (qualifier value)|
Laterality	272741003 |Laterality (attribute)|
Associated morphology	116676008 |Associated morphology (attribute)|
Due to	42752001 |Due to (attribute)|

Apart from SNOMED relations, the following FHIR relational elements should be used:

Short	Long
verificationStatus	Condition.verificationStatus (confirmed (default), differential, provisional, refuted)
clinicalStatus	Condition.clinicalStatus (active (default) recurrence, relapse, inactive, remission, resolved)

Short	Long
Severity	Condition.severity (mild, moderate, severe)
Name	FamilyMemberHistory.name
asserter	asserter.RelatedPerson.relationship
onsetAgeFam	FamilyMemberHistory.condition.onset
Name	RelatedPerson.name
Condition	FamilyMemberHistory.condition
relationship	familyMemberHistory.relationship
value	Quantity.value
unit	Quantity.unit
Comparison (>, <, >=, <=)	Quantity.comparator
Procedure status	Procedure.status (planned, stopped, on-hold, not-done, in-progress, default, preparation, completed)
onsetAge	Condition.onset
MedicationRequest.performer	MedicationRequest.performer
PlanDefinition.action	PlanDefinition.action
PlanDefinition	PlanDefinition.action.code
DeviceRequest	DeviceRequest.code
ServiceRequest	ServiceRequest.category
Goal.description	goal.description.CodeableConcept.coding

The default elements can be omitted. I.e. SNOMED clinical finding annotation without any modifying relation is interpreted as occurring within an affirmative relational statement.

5.3.7. Relations, predicates and operators

Concepts with a negative meaning (e.g., “162062008 |No vomiting (situation)|”) should be avoided. Whenever possible, they should be expressed by combining the positive meaning with the value “refuted” in Condition.verificationStatus. The use of concepts with negative meaning is limited to those cases where there is no alternative, e.g., “249695006 |Absence of rib (finding)|” because there is no “presence of rib”. (The latter ones are common finding / disorder concepts).

5.3.8. Boolean operators

The operators AND and OR express the way two or more SNOMED concepts that annotate the same span are to be interpreted. There is no use of a Boolean operator NOT (because negation cannot be expressed at SNOMED level): the FHIR attributes Condition.verificationStatus or Procedure.status are used instead.

- OR expresses ambiguity: only one concept is the correct annotation; which one cannot be decided by the annotator. This might be typical in the case of ambiguous acronyms.
- AND expresses a logical conjunction of the meaning of two or more concepts of the same sub-hierarchy. Particularly with procedure and condition concepts, the meaning may also be additive, because the implicit interpretation of “A” is always “patient having A” or “procedure with A”. (which explains, that “X-ray of radius and ulna (procedure)” is under “Procedure on ulna (procedure)” and under “Procedure on radius (procedure)”.

5.4. Monitoring annotation quality

The quality of the manual annotations can be monitored as described in Section 5.1.4. Endpoints of interest are statistical agreement measurements of the manual annotation process specifically supported by INCEpTION, as well as timing aspects of how long it takes to annotate all entities of interest in the document. It is planned to compare these outcomes with the pre-annotation process included, as specifically the normalization process of an entity of interest is especially time-consuming. We hypothesize that with an enabled pre-annotation process as described in Section 4.4 supporting medical concept normalization, the manual annotation process can be significantly reduced.

5.5. Annotator staff

Within the preparation and piloting phase until 31.12.2022, the main focus is set on “train the trainers” at the local sites to get familiar with the functionalities of INCEpTION, the annotation guidelines and the corresponding needed terminologies for normalization. The following table specifies the local INCEpTION trainers at their clinical side, as well as domain experts in the selected tooling (Averbis HD, INCEpTION).

Name	Institution	Role
Markus Kreuzthaler	MUG	Local trainer
Kristian Kankainen	NEMC	Local trainer
Petros Kalendralis	UM	Local trainer
Kris Collins, Stefan Schulz	AVER	Clinical NLP experts
Richard Eckart de Castilho	AVER	INCEpTION maintainer

Table 5. Role definitions for the manual annotations on the clinical sides.

Medical students (4th-6th year) and resident-physicians will be selected based on the following criteria:

- **Intrinsic motivation** to deal with clinical documents and medical content is the best skill set for the start of the manual annotation task;
- **Language skills** in English are necessary due to the use of English terminology resources for the normalization approach of the entity types of interest within the manual annotation process (see Section 4.3).

They will be instructed by the local trainers (see Table 5) following the annotation guidelines on an initial set of use case oriented clinical narratives. Two main aspects have to be assessed in the piloting phase for the manual annotation task:

- **Manual annotation throughput.** Due to the fact that the entities of interest have to be normalized to a specific terminology in combination with relations between them, the creation of an annotated set of documents can be very time-consuming. Therefore, the impact of pre-annotations can have a critical impact on manual annotation speed, but also the INCEpTION annotation model has to be adapted from its complexity to realistically fulfil the annotation task in time.
- **Manual annotation quality.** Inter-rater agreement will be assessed using INCEpTION's internal evaluation capabilities. A more detailed description can be found in Section 5.1.4.

A strong level of agreement is expected to be on the entity level of the manual annotations. A moderate level of agreement due to the complexity of this task [21] is expected to be for the entity normalization task. If the levels of agreement are too low, the annotation guidelines, the INCEpTION annotation model and terminology-specific value sets have to be revised for the productive annotation phase starting in Q2 2023.

Value of Kappa	Level of Agreement	% of Data that are Reliable
.40–.59	Weak	15–35%
.60–.79	Moderate	35–63%
.80–.90	Strong	64–81%
> 0.90	Almost perfect	82–100%

Table 6. Interpretation of Cohen's kappa according to [19].

7. Next steps

Short term

- Deploy the INCEpTION tool at the three sites (MUG, NEMC, MUMC), Q4 2022
- Recruit annotators at the three sites (MUG NEMC, MUMC), Q1 2023
- Organize a train the trainer session at the 3 sites, Q1 2023
- Start first annotation process and revise guidelines, Q1 2023

Midterm

- Assessment of process (including usage of tools) - benefit and concerns and identify areas of improvement.
- Improve the process, e.g., improve / extend the pre-annotation pipeline to increase automation and support for annotators.
- Evaluate the most appropriate approach to keep annotated datasets after the project, while ensuring compliance with data privacy.
- Update guidelines as of Q1 2023 after 4 months of practice.

8. References

[1] Bose P, Srinivasan S, Sleeman WC, Palta J, Kapoor R, Ghosh P. A Survey on Recent Named Entity Recognition and Relationship Extraction Techniques on Clinical Texts. Appl Sci 2021;11:8319. https://doi.org/10.3390/app11188319.

[2] Luo Y-F, Henry S, Wang Y, Shen F, Uzuner O, Rumshisky A. The 2019 n2c2/UMass Lowell shared task on clinical concept normalization. J Am Med Inform Assoc JAMIA 2020;27:1529-e1. https://doi.org/10.1093/jamia/ocaa106.

[3] Alimova I, Tutubalina E. Multiple features for clinical relation extraction: A machine learning approach. J Biomed Inform 2020;103:103382. https://doi.org/10.1016/j.jbi.2020.103382.

[4] Dernoncourt F, Lee JY, Uzuner O, Szolovits P. De-identification of patient notes with recurrent neural networks. J Am Med Inform Assoc JAMIA 2016;24:596–606. https://doi.org/10.1093/jamia/ocw156.

[5] Steinkamp J, Kantrowitz JJ, Airan-Javia S. Prevalence and Sources of Duplicate Information in the Electronic Medical Record. JAMA Netw Open 2022;5:e2233348. https://doi.org/10.1001/jamanetworkopen.2022.33348.

[6] Ivers D, Mitchell M. A HIPAA primer. J Ark Med Soc 2002;99:139.

[7] Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical Natural Language Processing in languages other than English: opportunities and challenges. J Biomed Semant 2018;9:12. https://doi.org/10.1186/s13326-018-0179-8.

[8] Lingren T, Deleger L, Molnar K, Zhai H, Meinzen-Derr J, Kaiser M, et al. Evaluating the impact of pre-annotation on annotation speed and potential bias: natural language processing gold standard development for clinical named entity recognition in clinical trial announcements. J Am Med Inform Assoc JAMIA 2014;21:406–13. https://doi.org/10.1136/amiajnl-2013-001837.

[9] Mikulová M, Straka M, Štěpánek J, Štěpánková B, Hajic J. Quality and Efficiency of Manual Annotation: Pre-annotation Bias. Proc. Thirteen. Lang. Resour. Eval. Conf., Marseille, France: European Language Resources Association; 2022, p. 2909–18.

[10] Chen Y, Lasko TA, Mei Q, Denny JC, Xu H. A study of active learning methods for named entity recognition in clinical text. J Biomed Inform 2015;58:11–8. https://doi.org/10.1016/j.jbi.2015.09.010.

[11] Neves M, Ševa J. An extensive review of tools for manual annotation of documents. Brief Bioinform 2021;22:146–63. https://doi.org/10.1093/bib/bbz130.

[12] Kravchenko T, Bogdanova T, Shevgunov T. Ranking Requirements Using MoSCoW Methodology in Practice. In: Silhavy R, editor. Cybern. Perspect. Syst., Cham: Springer International Publishing; 2022, p. 188–99. https://doi.org/10.1007/978-3-031-09073-8_18.

[13] Yimam SM, Gurevych I, Eckart de Castilho R, Biemann C. WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations. Proc. 51st Annu. Meet. Assoc. Comput. Linguist. Syst. Demonstr., Sofia, Bulgaria: Association for Computational Linguistics; 2013, p. 1–6.

[14] Eckart de Castilho R, Mújdricza-Maydt É, Yimam SM, Hartmann S, Gurevych I, Frank A, et al. A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. Proc. Workshop Lang. Technol. Resour. Tools Digit. Humanit. LT4DH, Osaka, Japan: The COLING 2016 Organizing Committee; 2016, p. 76–84.

[15] Klie J-C, Bugert M, Boullosa B, Eckart de Castilho R, Gurevych I. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. Proc. 27th Int. Conf. Comput. Linguist. Syst. Demonstr., Santa Fe, New Mexico: Association for Computational Linguistics; 2018, p. 5–9.

[16] Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J. brat: a Web-based Tool for NLP-Assisted Text Annotation. Proc. Demonstr. 13th Conf. Eur. Chapter Assoc. Comput. Linguist., Avignon, France: Association for Computational Linguistics; 2012, p. 102–7.

[17] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Vol. 1 Long Short Pap., Minneapolis, Minnesota: Association for Computational Linguistics; 2019, p. 4171–86. https://doi.org/10.18653/v1/N19-1423.

[18] Prodigy · Prodigy · An annotation tool for AI, Machine Learning & NLP. Prodigy n.d. https://prodi.gy/ (accessed November 28, 2022).

[19] McHugh ML. Interrater reliability: the kappa statistic. Biochem Medica 2012;22:276–82.

[20] GitHub - google/healthcare-text-annotation n.d. https://github.com/google/healthcare-text-annotation (accessed December 14, 2022).

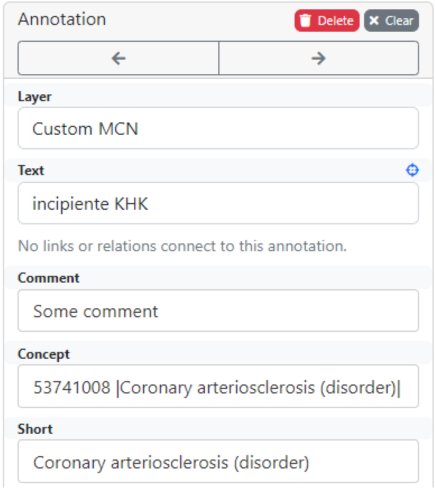
[21] Miñarro-Giménez JA, Cornet R, Jaulent MC, Dewenter H, Thun S, Gøeg KR, et al. Quantitative analysis of manual annotation of clinical text samples. Int J Med Inf 2019;123:37–48. https://doi.org/10.1016/j.ijmedinf.2018.12.011.

9. Annex - Examples for annotation with INCEpTION

9.1. Inception examples and recommendations

The following examples highlight recommendations that were elaborated in past annotation experiments at the Medical University of Graz. They can be considered as mostly consolidated, which however does not preclude revisiting them, particularly in the light of the use of LOINC along with SNOMED CT.

[Layer] mark the word or phrase and use the layer “Custom MCN“ copy and paste the SNOMED concept into “Concept” and the text of the concept again in “Short”.



Layer
Custom MCN

Text
incipiente KHK

No links or relations connect to this annotation.

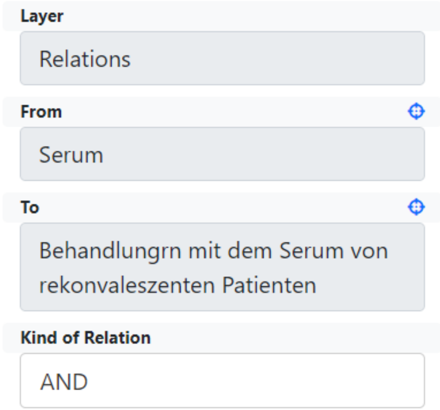
Comment
Some comment

Concept
53741008 |Coronary arteriosclerosis (disorder)|

Short
Coronary arteriosclerosis (disorder)

[AND/OR] if there are two fitting SNOMED terms for one concept (OR) or you need to express a refined meaning as the conjunction of two concepts [AND]:

Mark the word and use the layer “Custom MCN” and



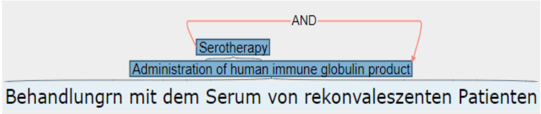
Layer
Relations

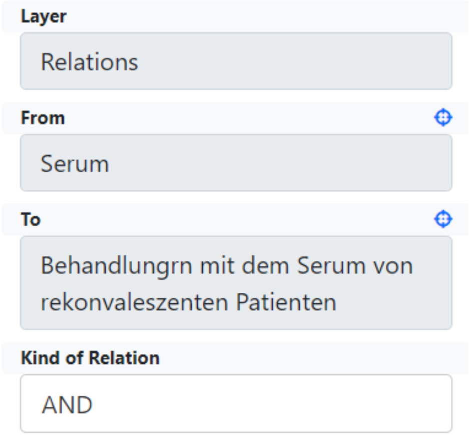
From
Serum

To
Behandlungrn mit dem Serum von rekonvaleszenten Patienten

Kind of Relation
AND

copy/paste the first SNOMED ID as in [Layer] and then mark the same word again use the layer “Custom MCN” and copy/paste the second SNOMED ID as in [Layer]. Connect the two concepts with “Relation” and define it with “AND”/”OR”.


AND
Serotherapy
Administration of human immune globulin product
Behandlungrn mit dem Serum von rekonvaleszenten Patienten



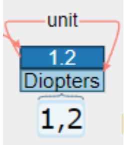
Layer
Relations

From
Serum

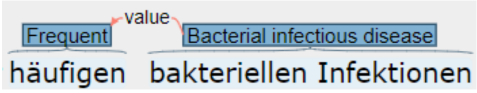
To
Behandlungrn mit dem Serum von rekonvaleszenten Patienten

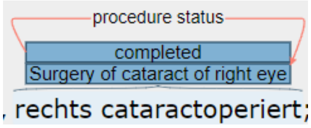
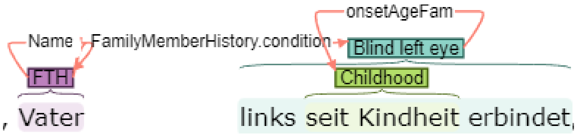
Kind of Relation
AND

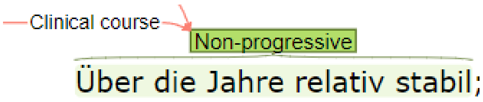
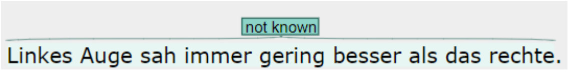
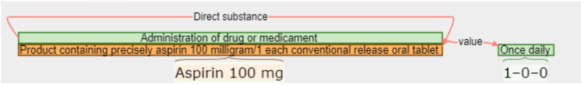
[RELATION] if a word has two meanings, e.g. 100 for 100 mg, mark the word two times as in AND/OR and connect them with “Relation”. Define it, e.g. with Unit.

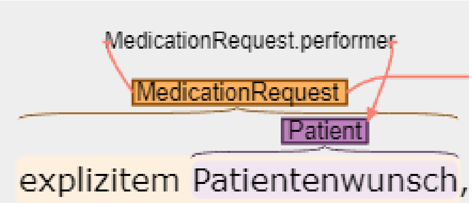
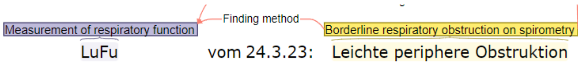
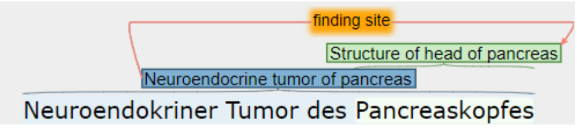
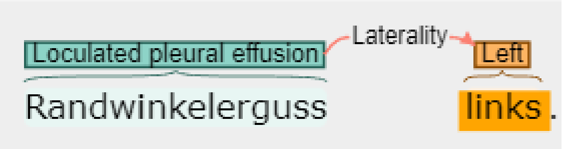


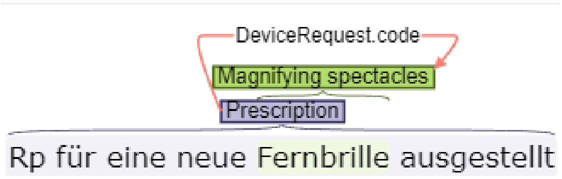
[value] if there is a procedure/finding with a qualifier value/number, mark the procedure/finding use the layer “Custom MCN” and copy/paste the SNOMED ID and mark the value/number and copy/paste the SNOMED ID. Connect the two concepts with “Relation” and



define it with the FHIR ID "Quantity.value" short "value".	
[procedure status] if there is a procedure with a procedure status-mark the procedure as in [Layer] and mark the procedure status use the layer "Custom MCN" and copy/paste a FHIR ID from procedure status. Connect the two concepts with "Relation" and define it with the FHIR ID "Procedure.status".	 procedure status completed Surgery of cataract of right eye , rechts cataractoperiert;
[FamilyMemberHistory.condition] if there is a finding in the family history - mark the family member use the layer "Custom MCN" and copy/paste the FHIR ID relationship as in [Layer] and mark the finding and copy/paste the SNOMED ID as in [Layer]. Connect the two concepts with "Relation" and define it with "FamilyMemberHistory.condition".	 onsetAgeFam Name FamilyMemberHistory.condition Blind left eye FTH Childhood , Vater links seit Kindheit erbindet,
[FamilyMemberHistory.name] links the "Relation" [Name] with relationship and the name of the relative, if there is no name given, link it again with the relationship.	
[Family history]: Grandfather had glaucoma and macular degeneration FamilyMemberHistory.relationship --> grandfather FamilyMemberHistory.condition --> glaucoma FamilyMemberHistory.condition --> macular degeneration There is the same span in text for both person and role: same domain and range node (use Shift in INCEpTION)	
[onsetAgeFam] if there is a finding in the family history with an onset age, mark the finding as in [Layer] and the onset Age as in [Layer]. Connect the two concepts with "Relation" and define it with the FHIR ID "FamilyMemberHistory.condition.onset" short "onsetAgeFam".	
[onsetAge] if there is a finding/condition with an onset Age - mark the finding as in [Layer] and the onset Age as in [Layer]. Connect the two	

concepts with “Relation” and define it with the FHIR ID “Condition.onset” short “onsetAge”.	
[Clinical course] if there is a finding with a clinical course, mark the finding as in [Layer] and the clinical course as in [Layer]. Connect the two concepts with “Relation” and specify the SNOMED ID “263502005 |Clinical course (attribute)|” short “Clinical course”	 Über die Jahre relativ stabil;
[not known] if there is a concept for which no SNOMED ID fits: mark the word as in [Layer] and write “not known” in concept A and B short. If there is a relation to the not known concept, connect it to the relation concept and define it with a relation.	 Linkes Auge sah immer gering besser als das rechte.
[Medication] whenever there is a SNOMED-ID with administration, use it. If there is just the substance, mark the word and add a Layer. In the second layer, you copy and paste “386359008 |Administration of drug or medicament via oral route (procedure)|” and in the first the SNOMED ID for the substance. Connect both with “Relation” and define it with “363701004 |Direct substance (attribute)|”. **[Regime]** if there is an administration regime like 1-0-0 or 0-0-1 mark the term as in “Layer” and copy and paste a SNOMED ID (like Once daily (qualifier value)|). Connect it with the medication concept with “Relation” and define it with the FHIR ID “Quantity.value” short “value”.	 Direct substance Administration of drug or medicament Product containing precisely aspirin 100 milligram/1 each conventional release oral tablet value Once daily Aspirin 100 mg 1–0–0
[Third party medical history] In case of a third party medical history (E.g. the mother is doing all the medical history for her daughter) mark the person (E.g. mother) use the FHIR code of the asserter (E.g. MTH) and mark the condition as in [Layer]. Connect both concepts with “relation” and define it with the FHIR code asserter in the short version and with “asserter.RelatedPerson.relationship” in the long version. Then link the “Relation” [RelatedPerson.name] with the relationship and the name of the relative, if there is no name given link it again with the relationship.	

[MedicationRequest.performer] if there is a person with a request for a medication, mark the medication request and use the FHIR ID "MedicationRequest" and mark the person who is requesting use the SNOMED ID as in Layer. Connect both concepts with "Relation" and define it with the FHIR ID "MedicationRequest.performer". Connect the "MedicationRequest" with the medication which is requested. Use "Relation" and define it with "363701004 |Direct substance (attribute)|".	 MedicationRequest.performer MedicationRequest Patient explizitem Patientenwunsch,
[Finding method] If there is a finding and its finding method (procedure)- mark the finding as in [Layer] and copy and paste the SNOMED ID. Then mark the finding method and copy and paste the SNOMED ID. Connect the two concepts with "Relation" and define it with the SNOMED ID "finding method".	 Measurement of respiratory function Finding method Borderline respiratory obstruction on spirometry LuFu vom 24.3.23: Leichte periphere Obstruktion
[Finding site] - if there is a finding with a location of the finding (anatomic structure) mark the finding and add the SNOMED ID. Mark the location and add the SNOMED ID. Connect both concepts with "Relation" and define it with the SNOMED concept "finding site".	 finding site Structure of head of pancreas Neuroendocrine tumor of pancreas Neuroendokriner Tumor des Pancreaskopfes
[Comparison] use the FHIR code (>, <, >=, <=) as a relation (just like value or unit) in order to make comparisons (E.g. 3kg > P97 in 6th)	
[Dropping issues] If there is only one mention of a "problem" which is too complex, it will be ignored, but if there are two or more we have to find a solution. As with "recommendations" which occur from time to time in medical texts, but the task to include them properly is quite difficult and hence there is only one recommendation in all the texts, we decided to ignore this issue.	
[Laterality] if there is a disorder/clinical finding/procedure/body structure with a laterality (left/right/both) mark the disorder/clinical finding/procedure/body structure as in [Layer] and mark the laterality and use the ID from SNOMED for left/right/left and right. Connect both concepts with [Relation] and define it with "Laterality" for the short term and	 Loculated pleural effusion Laterality Left Randwinkelerguss links.

with “272741003 |Laterality (attribute)|” for the long term.	
[Due to] if there is a reason given for a clinical finding or procedure, mark the clinical finding/procedure and use the SNOMED ID as in [Layer] and mark the “reason”, use the SNOMED ID as in [Layer]. Connect both concepts with [Relation] and define it with “Due to” for the short term and with “42752001 |Due to (attribute)|” for the long term.	
[DeviceRequest.code] if there is a device request e.g. “a device request for glasses”, mark the device as in [Layer] and mark the request as in [Layer]. Connect both concepts with [Relation] and define it with “DeviceRequest” for the short term and with the FHIR ID “DeviceRequest.code” for the long term.	 Rp für eine neue Fernbrille ausgestellt
[ServiceRequest.category] if there is a service request e.g. “the service of an otorhinolaryngologist was requested (Überweisung) “ mark the requested service as in [Layer] and mark the request e.g. “germ: Überweisung” as in [Layer] e.g. “103320006 |Request for (contextual qualifier) (qualifier value)|”. Connect both concepts with [Relation] and define it with “ServiceRequest” for the short and with the FHIR ID “ServiceRequest.category” for the long term.	
[PlanDefinition.action.code] - planned procedures [goal.description.CodeableConcept.coding] Recommendations for lifestyle (weight, smoking, sport) according to FHIR also recommendations.	

9.2. Limitations and workarounds

The following examples highlight persisting issues that came up in ongoing annotation experiments at the Medical University of Graz. They are subject to ongoing discussion.

[Planned procedures]: “NTx geplant”; “NTx” annotated with “70536003 |Transplant of kidney (procedure)|” ; “planned” annotated with “405613005 |Planned procedure (situation)|”; “70536003 |Transplant of kidney (procedure)|” --- > [PlanDefinition.action.code] --- > “405613005 |Planned procedure (situation)|”; Potentially refine by purpose (indication) and goal (intended state after the action).

[Conditional recommendations (like in clinical guidelines)]: “conditional” : if… then… else not represented.

[goals vs. plan]: Clinical narrative does not reveal all background discussions / decisions: therefore, our baseline is: Goal: if a specified state/condition of the patient shall be achieved; Plan: if a specified intervention (diagnostic / therapeutic) is planned / scheduled (PlanDefinition.action).

[Ambiguities]: Example “patient was recommended to seek therapy by community surgery service” PlanDefinition.action vs goal; Better: ServiceRequest.category -> FHIR points to surgical procedure in SNOMED; “275146006 |Refashioning of ingrowing toenail (procedure)|”; Rule of thumb: choose the FHIR resources that require the least that you have to take decisions not grounded in the text.

[Asserter]: E.g. mother who informs about a condition of her child; Again, no separate mention of person and role in text, therefore recursive link like in Family history; Mother (RelatedPerson.Relationship = MTH) - Relation: asserter.RelatedPerson.- condition (from condition to Person); RelatedPerson - FHIR v4.3.0 (hl7.org).

9.3. Domain and range constraints

This table gives an overview of the current domain / range constraints and therefore high-level annotation patterns. It will facilitate the decisions during the annotation process and will be further refined during annotation piloting. The table will also serve to count the instances of these annotation patterns.

Domain type		Relation	Range type		Count
FHIR	SNOMED		FHIR	SNOMED	
Condition	Event				
Condition	Clinical Finding/Disorder	Finding site		Body part	
Condition	Clinical Finding/Disorder	Finding method		procedure	
	Procedure E.g. Administration of drug	Direct substance		substance/ clinical drug	
Condition	Disorder/ Clinical Finding	Due to		Procedure/ finding	
	Disorder/ Clinical finding/ Procedure/body structure	Laterality		Laterality (left, right, both)	
	Procedure	Procedure site		Body part	
Condition	disorder	Associated morphology		morphologic abnormality	
		Temporally follows			
Condition	Disorder	Clinical course		finding	
Condition	Disorder/ Clinical Finding	Causative agent		Substance/ organism	
Medication Request		MedicationRequest. performer	performer	person	

Domain type		Relation	Range type		Count
familyMemberHistory.relationship		FamilyMemberHistory.name	familyMemberHistory.relationship/Name		
RelatedPerson.Relationship		RelatedPerson.name	RelatedPerson.Relationship/ Name		
RelatedPerson.Relationship		asserter.RelatedPerson	Condition	finding/disorder	
familyMemberHistory.relationship		FamilyMemberHistory.condition	Condition		
finding/ procedure/		Quantity.value		qualifier value/Number	
Number		Quantity.unit		qualifier value	
Condition	Disorder/ finding	Condition.verificationStatus	ConditionVerificationStatus		
Condition	Disorder/ finding	Condition.clinicalStatus	ConditionClinicalStatusCodes		
	procedure	Procedure.status	EventStatus		
Condition	Disorder/ finding	Condition.severity	Condition/DiagnosisSeverity		
		>, <, >=, <=			
		AND			
		OR			
Condition	Disorder	FamilyMemberHistory.condition.onset		qualifier value/ Number	
		DeviceRequest.code	Device	Physical object	
		ServiceRequest.category			
		PlanDefinition.action.code		procedure	
		goal.description.CodeableConcept.coding			
		MedicationRequest.performer			
Condition	Disorder/finding	Condition.onset		qualifier value/ Number	
Others					