



Chart Abstraction on VINCI: tools and services

Lacey Lewis, MPH

Veterans Informatics and Computing Infrastructure (VINCI)

VA Salt Lake City Healthcare System

1. VINCI Annotation Services
2. Overview of chart abstraction
3. Chart abstraction tools (ChartReview)
4. Questions and discussion

Outline

- What is your primary role in VA?
 - student, trainee, or fellow
 - clinician
 - researcher
 - Administrator, manager or policy-maker
 - Other

Poll Question #1

- Which best describes your research experience?
 - have not done research
 - have collaborated on research
 - have conducted research myself
 - have applied for research funding
 - have led a funded research grant

Poll Question #2

- What describes your experience with chart abstraction and/or annotation?
 - I have not conducted or collaborated on a project involved with chart abstraction or annotation.
 - I have collaborated on a project involved with chart abstraction or annotation.
 - I have led/conducted a chart abstraction or annotation project.

Poll Question #3

VINCIservices@va.gov

- Concierge
- Data Provisioning
- Compliance
- Feasibility
- Recruitment
- **Annotation / Chart Abstraction**
- Natural Language Processing
- Analytics and Data Services
- Application Development

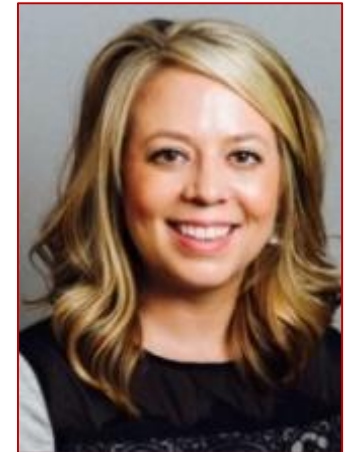
VINCI Services Team

VINCI Services Annotation team



Associate Director of VINCI:
Jeffrey Ferarro, PhD

Annotation Manager:
Lacey Lewis, MPH

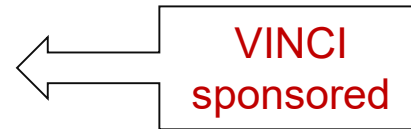


Team comprised of clinical and professional annotators.

VINCI Annotation Services

Education and Training

- Introduction to abstraction tools
- Training/Consultation on chart abstraction methods
- Consultation on tool set-up
- Provide examples of chart abstraction documents



Contracted Services

- Process definition and guideline development
- Contracting annotators
- Full chart abstraction process



VINCI Annotation Services

Chart Abstraction (Chart review, Medical record review, Chart annotation)

- A research methodology of data collection for retrospective investigation.

Chart Abstraction Methods:

- Single annotation of notes/patient charts
- Double annotation on % of notes/patient charts followed by single annotation
- Double annotation of notes/patient charts followed by adjudication
- Single annotation by more than one reviewer with % randomly double annotated
- **Double annotation on % of notes/patient charts followed with a % randomly double annotated throughout remaining notes/patient charts**
- For any method, what level is acceptable for agreement on defined concepts
 - Percentage agreement and Kappa calculated
 - Inter-rater reliability

Overview

- IRB Approval
- DART Approval
 - Request Real SSN
 - Request CDW production domains
 - Request TIU Text Notes
 - Ensure HIPAA Waiver
- ePAS (NDS approval for everyone that will require access to the database of interest)
- Contact VinciServices@va.gov for regulatory related questions.

Does your study require a HIPAA Waiver?

Yes

No

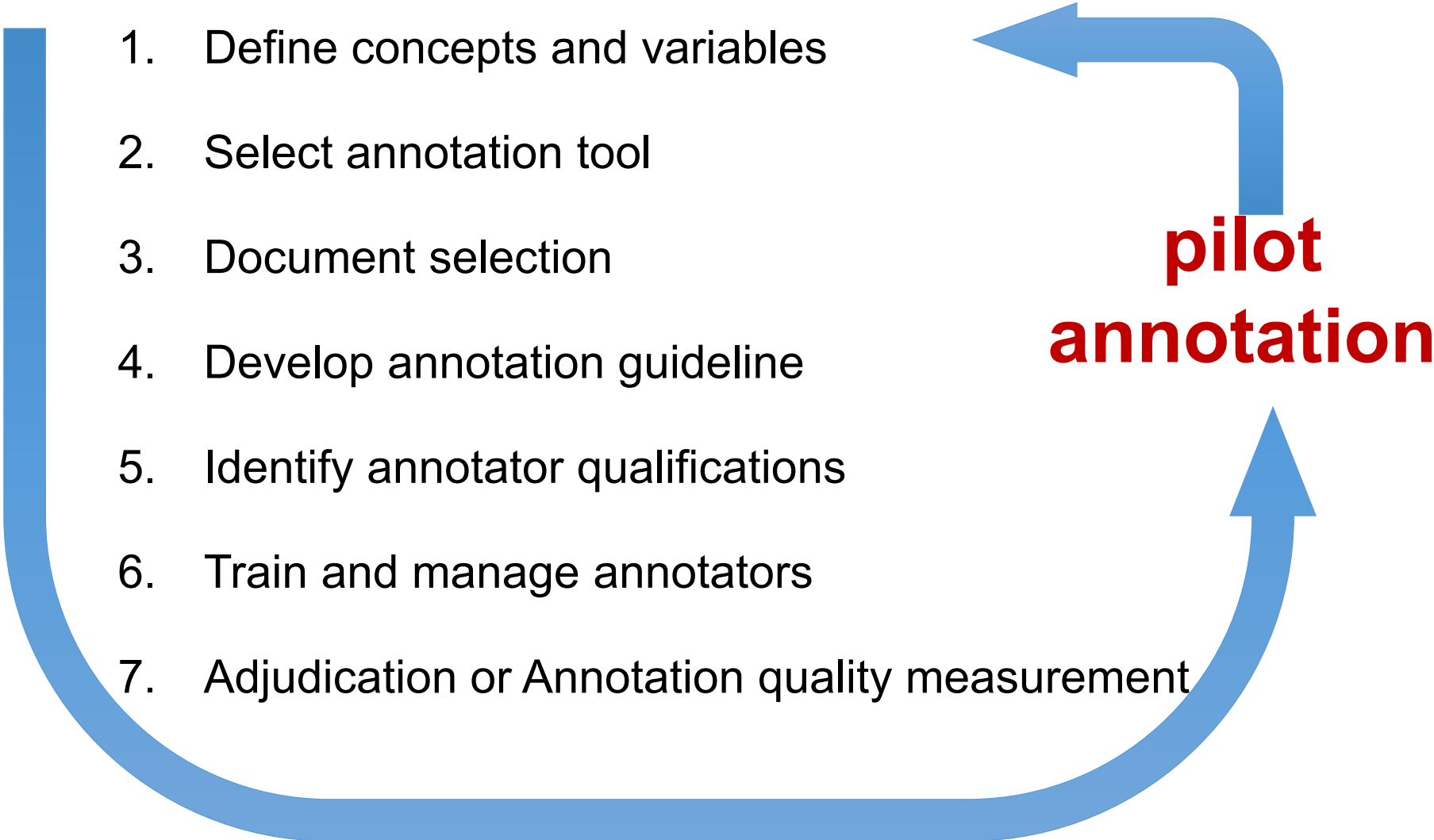
IDENTIFIERS [\(More about identifiers\)](#)

Real SSN

Scrambled SSN

Identifiable data but no real or scrambled SSNs

Regulatory Requirements

1. Define concepts and variables
 2. Select annotation tool
 3. Document selection
 4. Develop annotation guideline
 5. Identify annotator qualifications
 6. Train and manage annotators
 7. Adjudication or Annotation quality measurement
- 
- pilot
annotation**

Annotation Project Workflow

Defining a concept

1. Concept definition
 - Diagnosis, lab test, action, event...
2. Variable definition – values that the concepts can have
 - Diagnosis: explicitly mentioned or inferred
 - Lab test: exact numeric value or range or direction
 - Action: planned or occurred
 - Event: explicitly mentioned or inferred
3. Level of the annotation
 - patient, event, document, or instance
4. Annotation boundaries for instance annotations

Annotation type definition

Concept sheet - formalized concept/variable definition

Concept	Variable definition	Source	Range of values
Procedure indication	Explicitly stated indication for procedure.	Colonoscopy report	Screening High Risk/diagnostic Treatment

Example of a concept sheet for a study on quality of colonoscopy procedures

Operationalization !!!

No vague definitions!

Words to avoid:

“Any evidence of ...”

“... including but not limited to ...”

! Tip !

Annotation type definition

Defining annotation level

Instance Level

- A specific phrase in a document is being identified
- A document may have 0 to many instances (mentions, phrases)
- Instances in a single document may not always agree

Document Level

- The contents of the whole document are being labeled

Event Level

- Information in several documents is used to label patient on a specific date

Patient Level

- The total of all instances and/or documents are combined to reach a final patient classification
- A patient may have 0 to many documents, events, and instances

Annotation type definition

Instance boundary delimitation

Example: Annotation type = Polyp removal method
What is the phrase to annotate?

Polyp was removed with cold snare

Polyp was removed with cold snare

Polyp was removed with cold snare

Annotation type definition

Tool selection

Many tools can be used for chart abstraction and annotation. The tool most useful will be dependent on project specific needs.

A wide variety of tools available for chart abstraction and annotation. Many may be used within the VINCI environment:

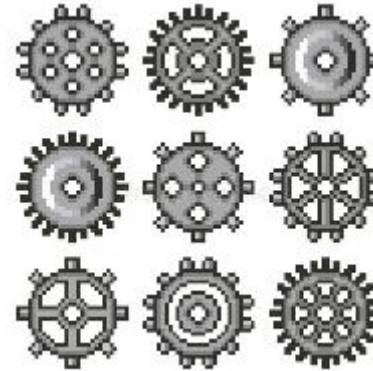
- eHost - <https://code.google.com/p/ehost/>
- ChartReview - <http://department-of-veterans-affairs.github.io/ChartReview/>
- BRAT - <http://brat.nlplab.org/>
- RapTAT – <https://code.google.com/p/raptat/>
- Knowtator – <http://knowtator.sourceforge.net/>
- GATE Teamware - <https://gate.ac.uk/sale/tao/splitch25.html>
- WebAnno - <https://webanno.github.io/webanno/>
- Prodigy - <https://prodi.gy>
- XTrans - <https://www ldc.upenn.edu/language-resources/tools/xtrans>
- WordFreak - <http://wordfreak.sourceforge.net>
- Djangology - <https://sourceforge.net/projects/djangology/>
-

Annotation tools

Annotation **complexity** inversely proportional to annotation **consistency**

- Working memory limit is 4* to 7**
- Consistency of annotation is the key!

Project size vs number of projects



* Cowan N. The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Curr Dir Psychol Sci.* 2010;19(1):51-57.

** Miller GA. The Magical Number Seven, Plus or Minus Two Some Limits on Our Capacity for Processing Information. *Psych Review.* 1956.

Scope of the annotation project

• Annotation

- **Class** - assigned meaning to data

Label = Concept = annotation class = annotation type \approx semantic type

- **Span** - a pointer to start and stop points in a text
- **Features** - attributes of the Class and their **values**

- Generated by human, machine, or human+machine.

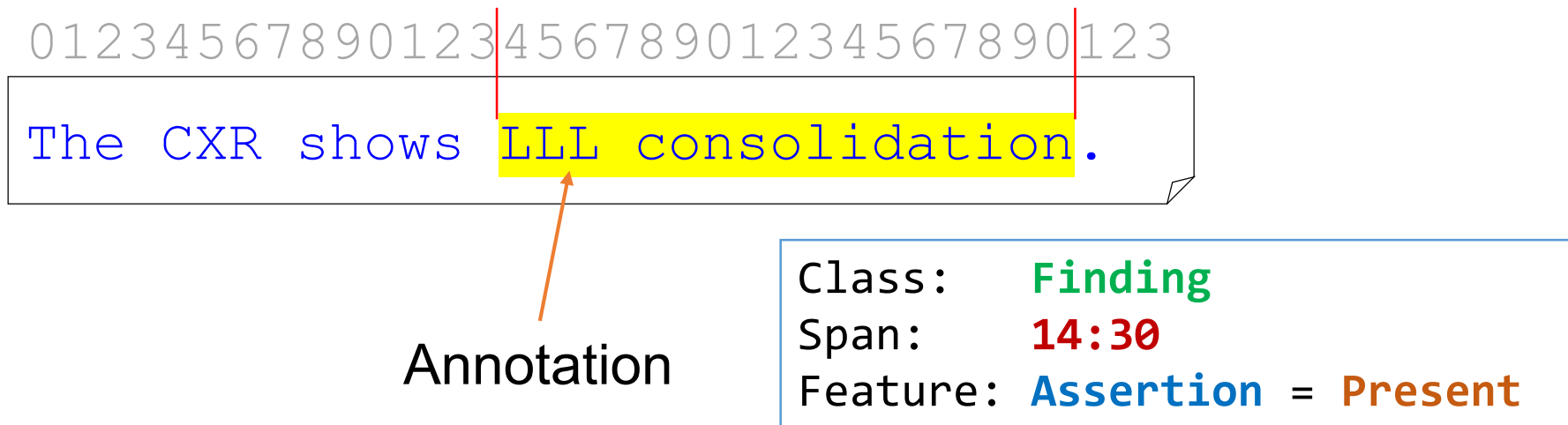


Chart Abstraction through Annotations

- Markables – elements to be annotated
 - Annotation type = label to be assigned to a segment of text
 - Relationship = link between instances of annotations
 - Features = Attributes of annotation types and relationships
- Annotation Schema
 - Complete list of all markables for the project

What to annotate

Once concepts are defined, the information can be compiled into an annotation guideline.

Other information to include in a guideline:

- Annotation schema
 - Annotation types = classes, labels
 - Attributes = features, modifiers
 - Relationships
- Formal step-by-step description of the annotation process
 - Must have:
 - Background information
 - Annotation instructions with examples specific to the annotation tool and to the documents in the corpus
 - Examples
 - what to annotate
 - what not to annotate

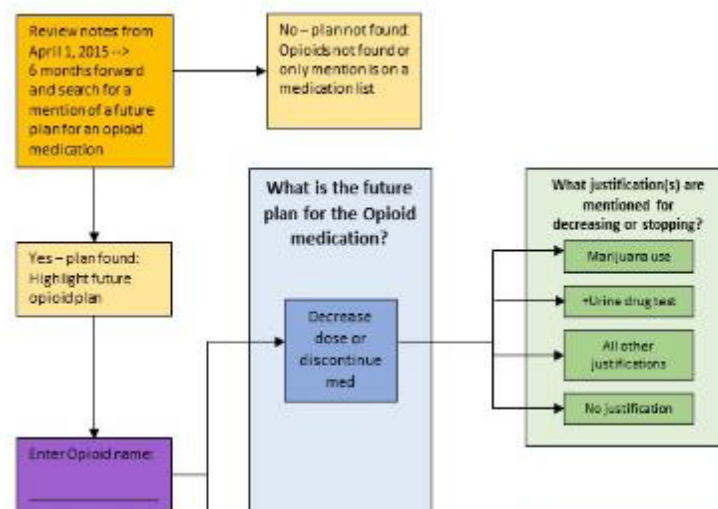
Annotation guideline

a. You will review the following notes and focus in on the health maintenance section of the notes: PC (primary care/women's health clinic; not to include PC pharmacy, nursing etc.), mammography notes, consults and orders related to mammograms/breast, radiology reports.

b. Highlight/annotate if the patient has the following information/risk factors [all at the document level]:

1. Patient has **Personal_History** of breast cancer.
 - Yes=annotate
 - No
2. Patient has a **Family_History** of breast cancer.
 - Yes= annotate
 - No
3. Patient has a **Genetic_Mutation** linked to breast cancer risk (e.g., BRCA 1, BRCA 2).
 - Yes=annotate
 - No
4. Patient has a history of **Previous_Radiotherapy** to the chest at a young age (e.g., most commonly as part of treatment for Hodgkin's Lymphoma).
 - Yes=annotate
 - No

*Please annotate a Yes or a No for each of the risk factors. If the patient has NO risk factors for any, 1-4, then see step 6.



- Include detailed information on the concepts
- Steps in utilizing the tool of interest
- Steps for the reviewer to follow to find the concepts of interest

Annotation Guideline

Training consists of:

- Defined concepts
- Guideline
- Tool specifics
- Any other project related information

Important to:

- Communicate consistently amongst annotators
- Have concept definitions and guideline very specific to task

Train annotators/reviewers

- Review of Kappa from double annotation
 - Cohen's kappa coefficient (κ) is a statistic that is used to measure inter-rater reliability (and also Intra-rater reliability) for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation, as κ takes into account the possibility of the agreement occurring by chance.
- Percent of accuracy found for each concept
- Review for annotator drift/quality check of annotation

KAPPA CALCULATION

$$\mathbf{\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}}$$

Pr(e) Calculation

$$\begin{aligned} \text{Expected Agreement} &= \frac{\left(\frac{\text{cm}^1 \times \text{rm}^1}{n}\right) + \left(\frac{\text{cm}^2 \times \text{rm}^2}{n}\right)}{n} \\ \text{Expected Agreement} &= \left(\frac{157 \times 150}{222}\right) + \left(\frac{65 \times 72}{222}\right) \\ \text{Expected Agreement} &= \frac{106.08 + 21.08}{222} = \underline{\underline{.57}} \end{aligned}$$

$$\mathbf{\text{Kappa} = \frac{.94 - .57}{1 - .57} = .85}$$

Biochemia-medica.com

Annotation analysis

VINCI ChartReview



Annotation

Task

Done Save Hold/Next Submit/Next Task Info

Task Name: CR_TEST_Synth_Pres (197617)
Principal: CR_SYNTH_Smoking_Patients_20170717
Description: no detailed description

Items

- 1.1. Smoked Current_Smoker
1.2. gender
1.3. age
2. Smoking intensity
2.1. frequency
3. Cigar Use
3.1. Does the patient have a history of cigar Use?

Previous Skip Next/Complete

1.1. Has the patient ever smoked:

- Yes
No
(Clear Answer)

Annotations

All Annotations Wrap

Table with columns: Span, ClinicalElement, Class, Type, Creation Date, ClinicalElement. Row: TOBACCO USE DISORDER, CR_SYNTH_Sr 800000085674, Current_Smoke, New, 2017-08-09T16:42:12Z, CR_SYI...

Certainty: High

Tobacco product type:

Cigarettes

CR_SYNTH_Smoking_Patients_20170717

Summary

Done

Classify

ScrSSN: 111000111
Race: White
Gender: M
DOB: 1950-01-01

CR_SYNTH_Smoking_Documents_20170717

List (5 of 5)

Detail: Right Description Wrap Auto-select Search Detail View

Table with columns: ReferenceDateTime, TIUDocumentSID, TIUDocumentDefinition. Rows: ADMIN_NOTE, Clinic Visit Note, Nursing Shift Note Assessment, Followup Summary, Followup Summary.

View in new tab Done Popup Annotation Text

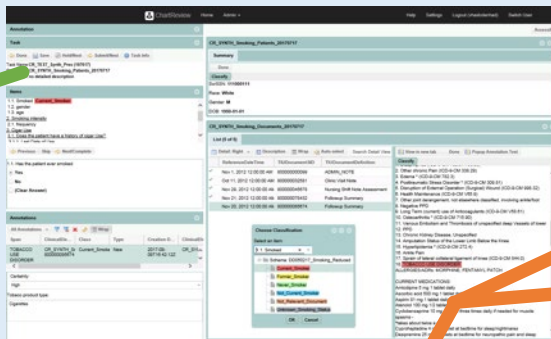
Classify

- 2. Other chronic Pain (ICD-9-CM 338.29)
3. Edema * (ICD-9-CM 782.3)
4. Posttraumatic Stress Disorder * (ICD-9-CM 309.81)
5. Disruption of External Operation (Surgical) Wound (ICD-9-CM 998.32)
6. Health Maintenance (ICD-9-CM V85.9)
7. Other joint derangement, not elsewhere classified, involving ankle/foot
8. Negative PPD
9. Long Term (current) use of Anticoagulants (ICD-9-CM V58.61)
10. Osteoarthritis * (ICD-9-CM 715.90)
11. Venous Embolism and Thrombosis of unspecified deep Vessels of lower
12. PPD
13. Chronic Kidney Disease, Unspecified
14. Amputation Status of the Lower Limb Below the Knee
15. Hyperlipidemia * (ICD-9-CM 272.4)
16. Ankle Pain
17. Sprain of lateral collateral ligament of knee (ICD-9-CM 844.0)
18. TOBACCO USE DISORDER
ALLERGIES/ADRS: MORPHINE, FENTANYL PATCH
CURRENT MEDICATIONS:
Amlodipine 5 mg 1 tablet daily
Ascorbic acid 500 mg 1 tablet daily with iron - uses twice daily
Aspirin 81 mg 1 tablet daily
Atenolol 100 mg 1/2 tablet daily
Cyclobenzaprine 10 mg 1 tablet three times daily if needed for muscle spasms -
*takes about twice a day
Cyproheptadine 4 mg 1 tablet at bedtime for sleep/nightmares
Desipramine 25 mg 5 tablets at bedtime for neuropathic pain and sleep

Choose Classification dialog box. Select an item: 1.1. Smoked. Schema: DD050217_Smoking_Reduced. Options: Current_Smoker, Former_Smoker, Never_Smoker, Not_Current_Smoker, Not_Relevant_Document, Unknown_Smoking_Status.

VINCI

<http://vhacdwdwhweb04:8080/chart-review>



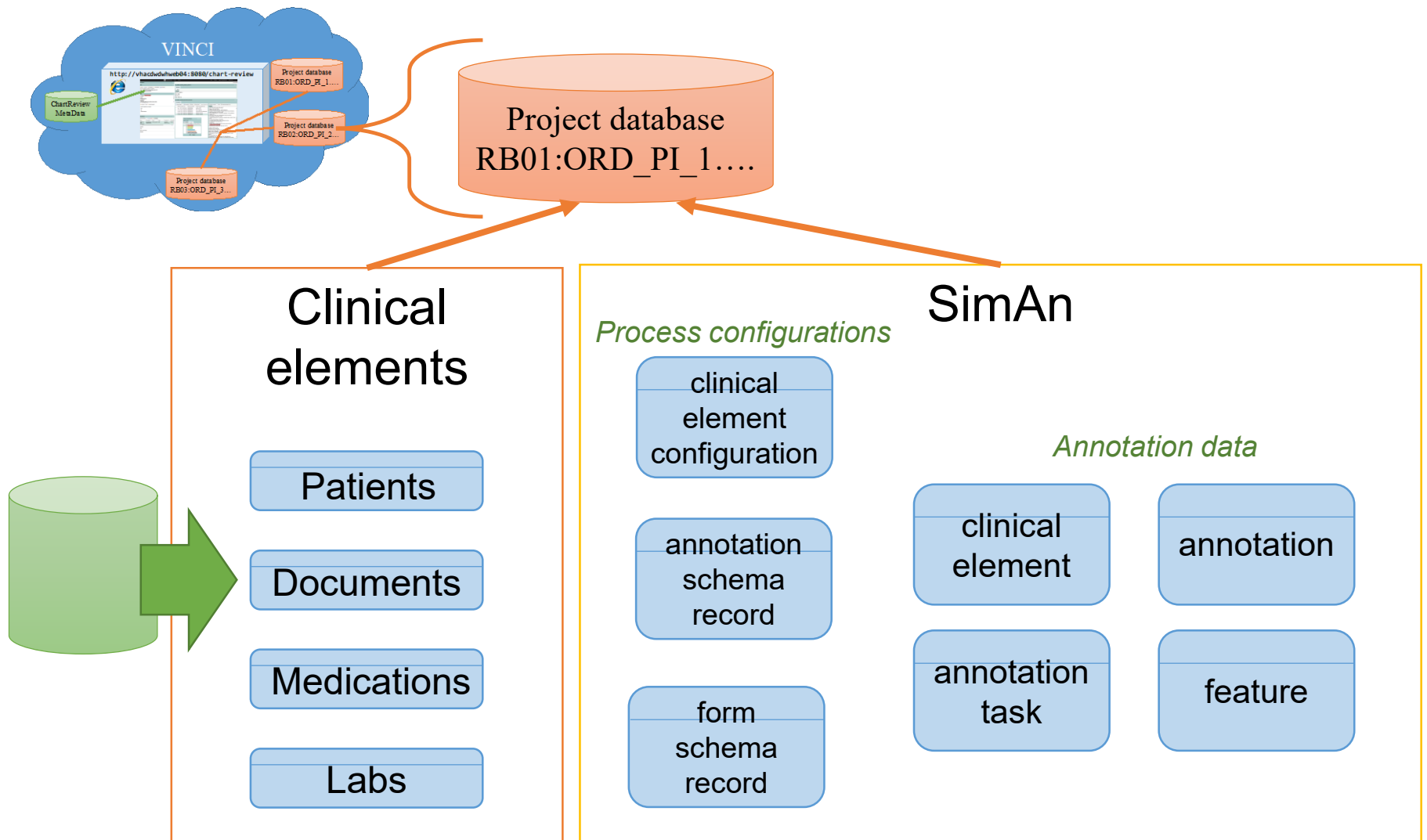
ChartReview
MetaData

Project database
RB01:ORD_PI_1....

Project database
RB02:ORD_PI_2...

Project database
RB03:ORD_PI_3...

ChartReview architecture

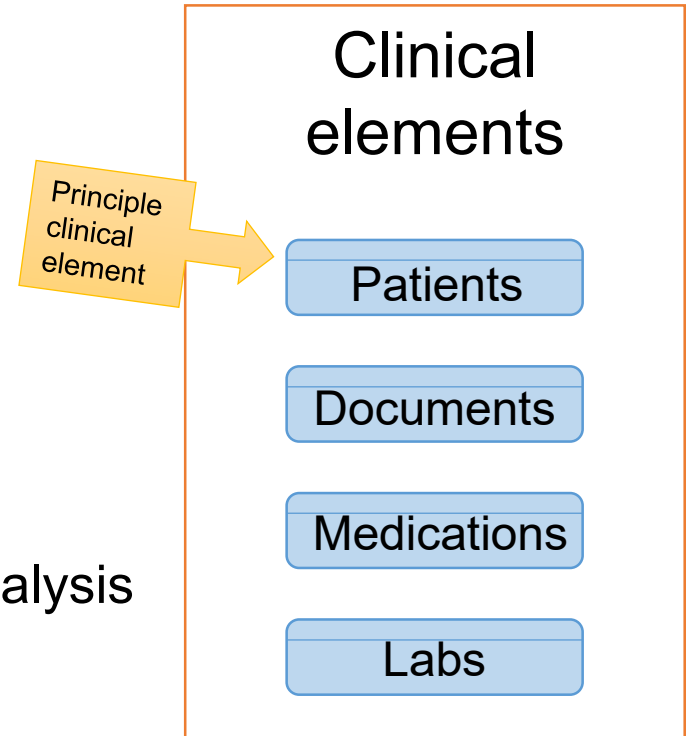


* SimAn - Simple Annotation Schema

ChartReview architecture

Clinical element

- Data element to be viewed during chart abstraction and/or annotation.
- Examples:
 - Lab values
 - Notes
 - Radiology reports
 - ICD-9/10 Codes
 - Medications
 - Vital status
 - Etc.



Principle clinical element

- Clinical element that serves as a unit of analysis
- Most frequently -- Patient
- Links all other clinical elements

ChartReview terminology

Project

- Assigned workspace
- Associated with VINCI research database
 - i.e. (ORD_PI_201801020D)

Process

- Individual abstraction/annotation undertaking for a specific research project
- Defined by: schema, clinical elements

Task

- The individual unique item to be reviewed within the process.
- Corresponds to [principle clinical element](#)

ChartReview Terminology

Annotation Schema

- Defines markables
- Example:

Annotation type= Smoking

- Attribute = smoking status
- Value data type = Option (current, past, never)

Annotation type = Patient Date of Birth

- Attribute = Date
- Value data type = date in range (01/01/1910 - 01/01/2001)

Form Schema

- Abstraction instrument
- Example:

1. Does the patient have associated co-morbidities: YES/NO

If Yes, then select from the following:

- Chronic Heart Failure
- Chronic Kidney Disease
- Diabetes
- Etc.

ChartReview Terminology

If you have a question that was not answered, or suggestion, please email it to:

VinciServices@va.gov

Subject: Annotation Services

Discussion

1. South BR, Shen S, Barrus R, DuVall SL, Uzuner O, Weir C. Qualitative analysis of workflow modifications used to generate the reference standard for the 2010 i2b2/VA challenge. *AMIA Annu Symp Proc.* 2011;2011:1243-1251.
2. Forbush TB, Shen S, South BR, Duvall SL. What a catch! traits that define good annotators. *Stud Health Technol Inform.* 2013;192:1213.
3. Warner JL, Anick P, Drews RE. Physician inter-annotator agreement in the Quality Oncology Practice Initiative manual abstraction task. *J Oncol Pract.* 2013;9(3):e96-e102.
4. Matt V, Matthew H. The retrospective chart review: important methodological considerations. *J Educ Eval Health Prof.* 2013;10:12.
5. Liu K, Chapman W, Hwa R, Crowley RS. Heuristic Sample Selection to Minimize Reference Standard Training Set for a Part-Of-Speech Tagger. *J Am Med Informatics Assoc.* 2007;14(5):641-650.
6. Ogren P V, Savova G, Chute CG. Constructing Evaluation Corpora for Automated Clinical Named Entity Recognition. In: *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco; 2008.
7. Chapman WW, Dowling JN, Hripcsak G. Evaluation of training with an annotation schema for manual annotation of clinical conditions from emergency department reports. *Int J Med Inf.* 2008;77(2):107-113.

Suggested reading