

Introduction to Clinical Text Processing

Olga Patterson, PhD

¹ VINCI, VA Salt Lake City Health Care System

² Division of Epidemiology, University of Utah



Disclaimer

The contents of this presentation do not represent the views of the Department of Veterans Affairs or the United States Government.

Acknowledgement

Some work described in this presentation was supported using resources and facilities at the VA Salt Lake City Health Care System with funding from VA Informatics and Computing Infrastructure (VINCI), VA HSR RES 13-457.

Poll Question #1

What is your primary role in VA?

- A. Researcher/Investigator
- B. Research staff (coordinator, statistician, analyst...)
- C. Administrator, manager, or policy-maker
- D. Other
- E. Not affiliated with VA

Poll Question #2

Which statement best reflects your thoughts about clinical text processing?

- A. Too much ado about nothing... NLP is overrated.
- B. I have heard good things about NLP, but do not know what I can do with it
- C. I believe NLP can help me in my research - I want to know more
- D. NLP is so exciting! I want to collaborate with NLP experts and do cool projects!
- E. I am an expert in text processing and I want to collaborate with clinical researchers.

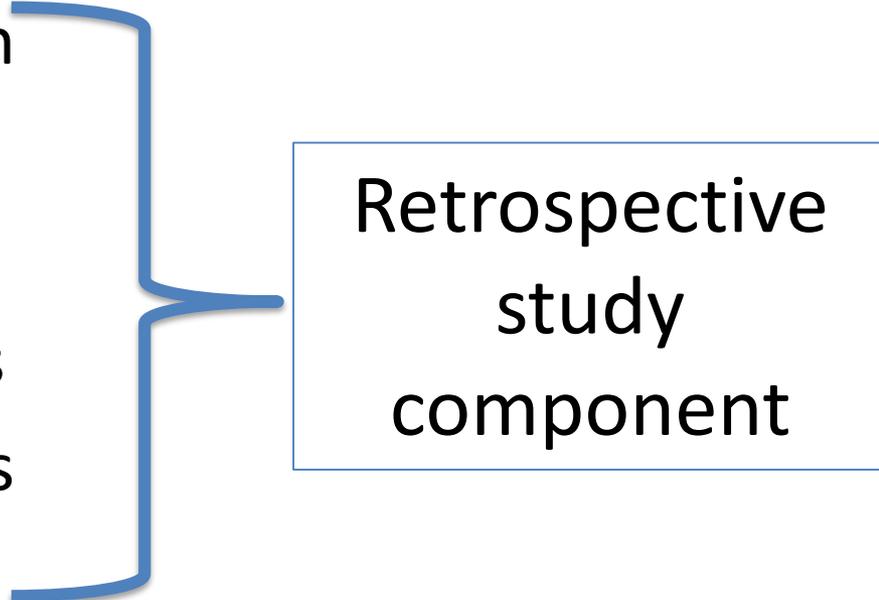
Overview

- Using EHR data in clinical research
- Clinical narrative as data source
- What is NLP?
- Where do I start with NLP?



Research areas in Clinical domains

- Clinical outcomes research
- Health services research
- Disease modeling
- Comparative effectiveness
- Prospective clinical studies
- ...



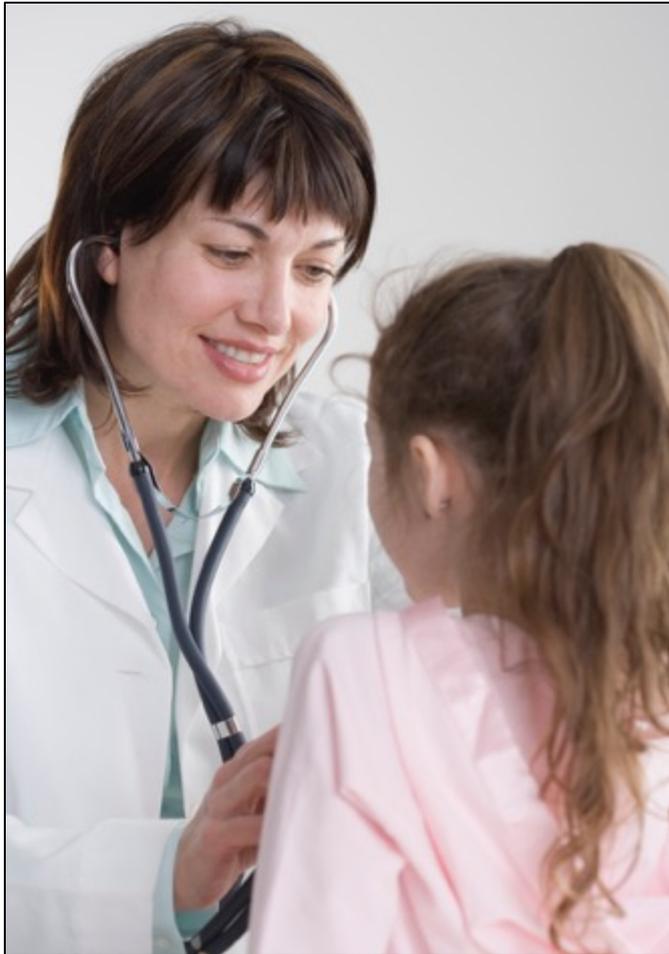
Retrospective
study
component



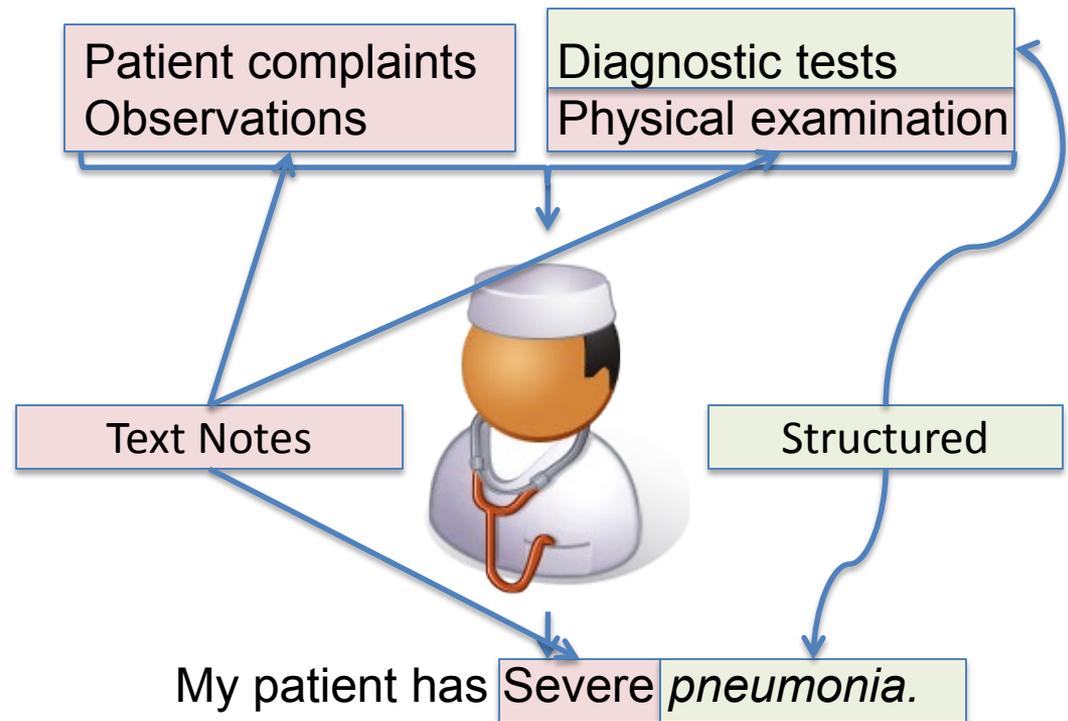
Secondary use of Electronic Medical Record

How is EMR Created?

Electronic Medical Record = EMR

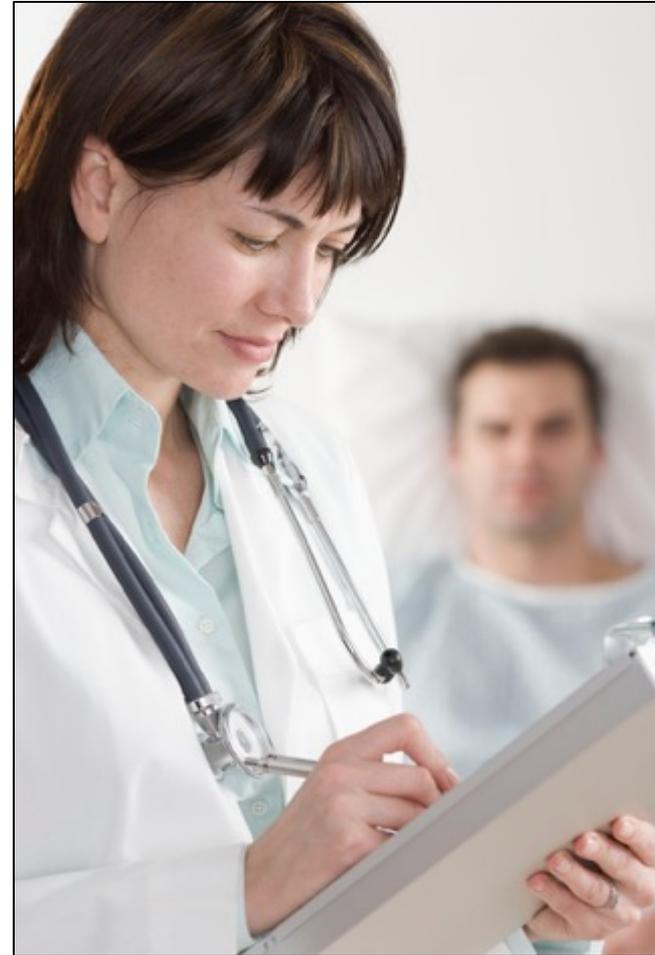
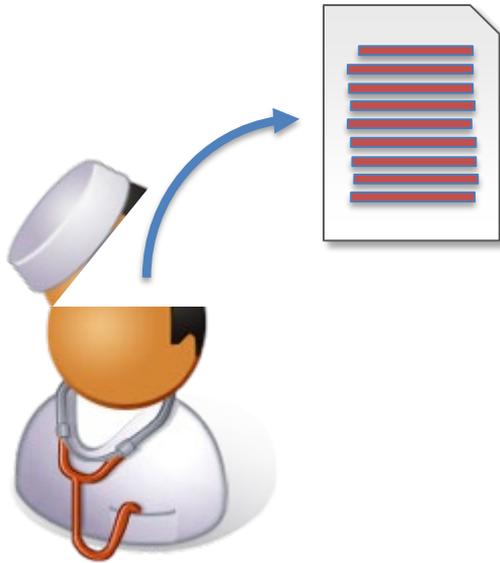


What is wrong with my patient?



How is EMR Created?

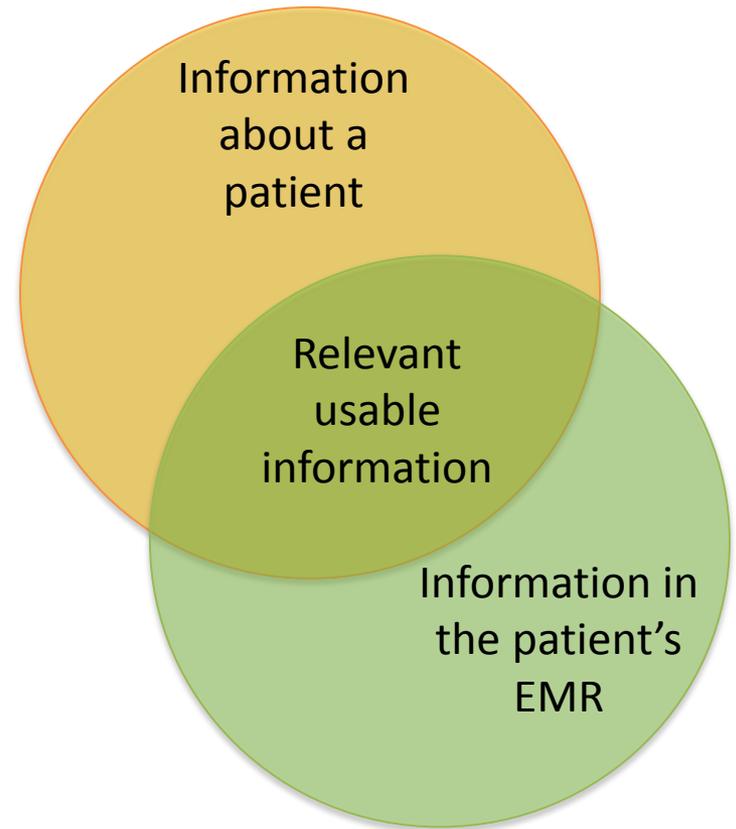
Creating patient documentation



How is EMR Created?

EMR is written by providers for other providers

- Limitations:
 - Information omission
 - verbal communication
 - shared environment
 - Information excess
 - not related to the patient
 - historical data for comparison



Data Types in the EMR

Structured and coded data



Fasting Glucose

135 mg/dL

Pulse

60 bpm



← Unstructured narrative data

synthetic medical record data

DISCHARGE SUMMARY

ALLERGIES: Tylenol, Advil, Nafcillin.

MEDICATIONS ON ADMISSION:
diazepam 25 mg daily , aspirin 325 mg
daily , Lasix 40 mg daily , Clariton
extended release 1 pill po prn.

HOSPITAL COURSE: The patient was
admitted for a brain hemorrhage and was
placed on Lasix. The patient became
stable and was transferred to ICU for
continued monitoring. While in the
hospital, she acquired MRSA and was
placed on antibiotics including
vancomycin and penicillin. Patient should
continue aspirin daily and an oral dose of
diazepam 25 mg for a total of 10 days.

Motivation: Why text?

- The patient experience

“Patient reports his knee hurts so bad he cannot sleep. He is also at risk of losing his job because he cannot work without sitting down”

- The type of illness, symptoms, and severity

“Diagnosed with relapsing remitting MS, currently mild tingling and weak grip.”

- The timing of the episode of care, disease, and treatment

“The patient saw ENT last week and surg was scheduled. She was cleared by cardiology last Monday, labs yesterday were normal. Pt taken to OR at 3:00 PM for tonsillectomy, she was taken to PACU in good condition, returned to same day, and discharged at 8 pm.”

“The chest pain started at while at the gym running, but resolved with rest. The following day the patient again had CP while walking. Today the pain was constant and he presented to the ER. The pain resolved with aspirin, oxygen, nitro, MS.”

“The patient was started on albuterol, then changed to xop and spiriva after pulm consult and cardiology saw her.”

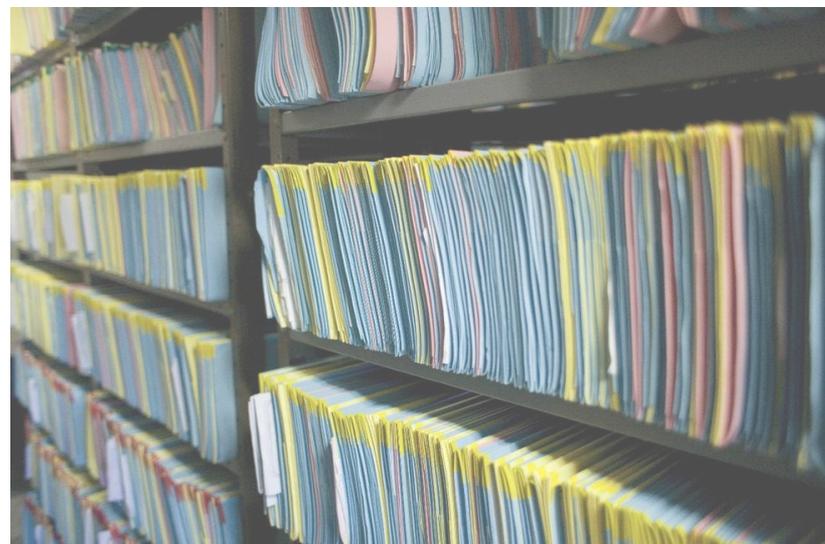
Motivation: Why text?

- The outcomes are in text.

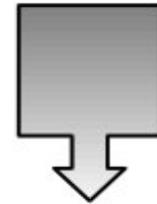
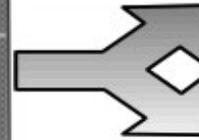
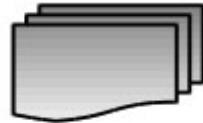
“The patient was started on 48 weeks of peg interferon and ribavirin, but tx was stopped due to constant fatigue, anxiety, and concerns of his wife that he may harm himself.”
- Even structured elements that are missing from the database are in text.

“Patient is transferring care from university hospital.
He is genotype 1, VI 391,000, hep B immune, HIV negative.”

Text is where the majority of clinical information is stored!



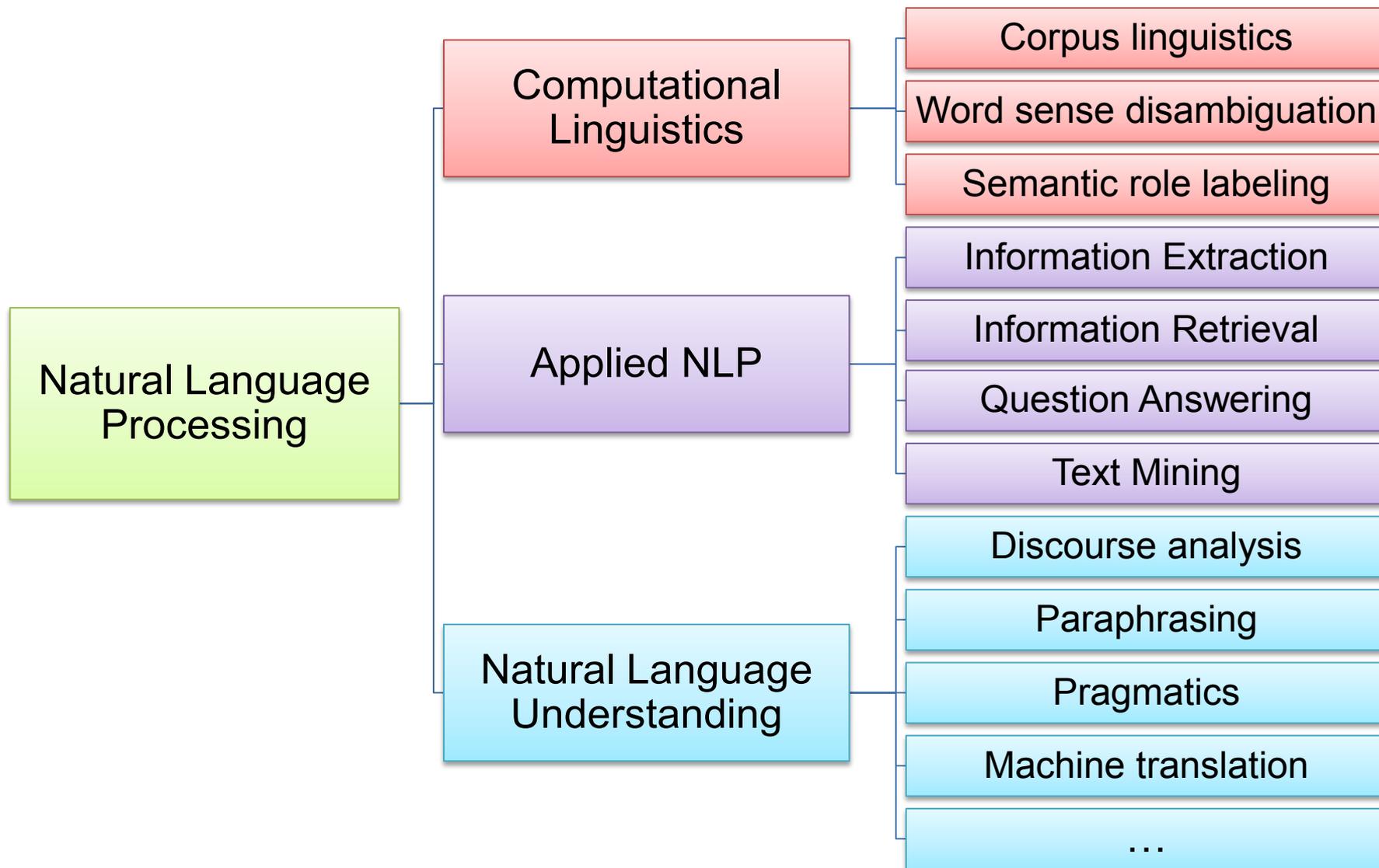
Natural Language Processing



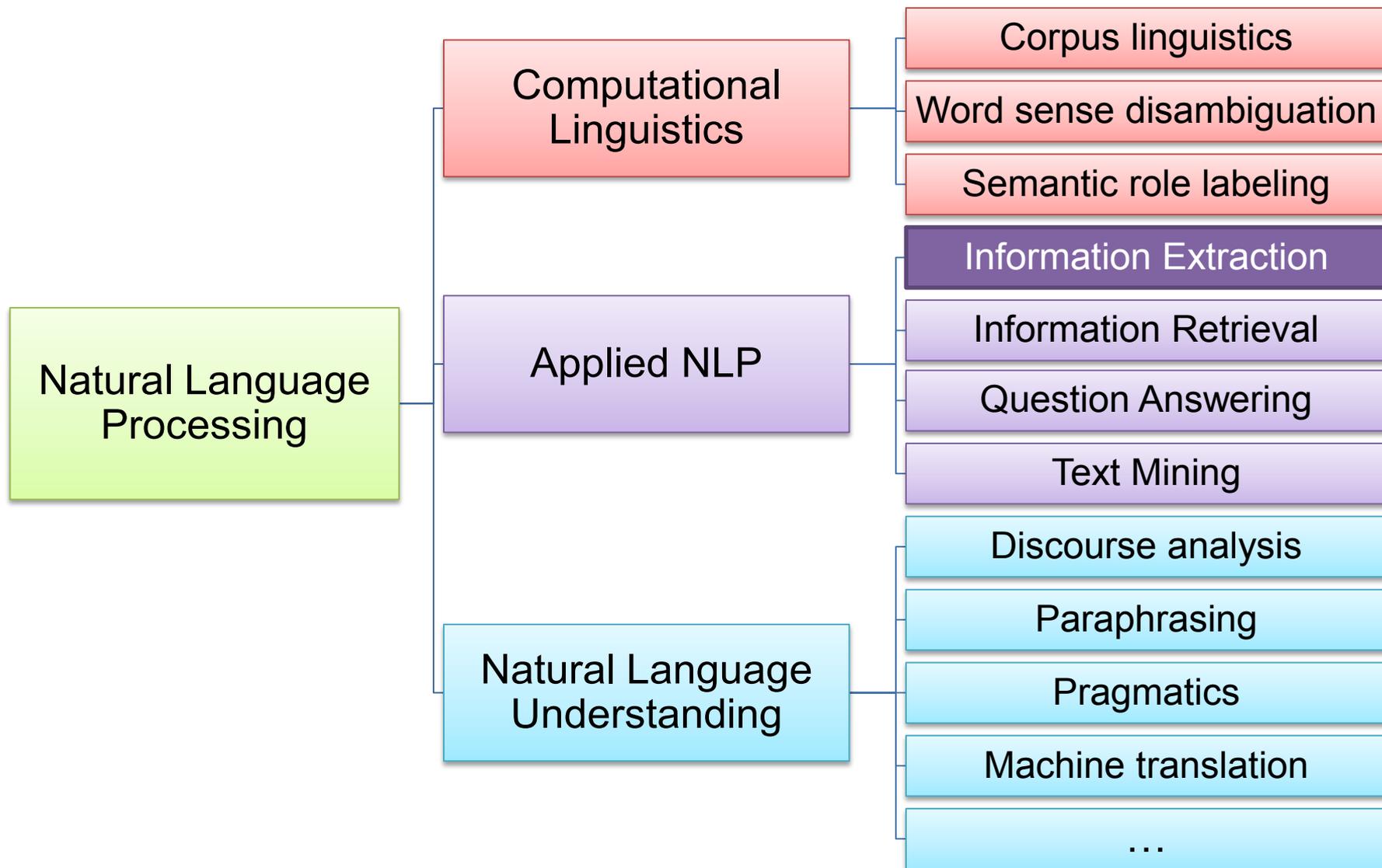
Natural language processing (NLP)

- **computational processing**
- **unstructured text**
- **broad range of purposes**

NLP subfields and tasks

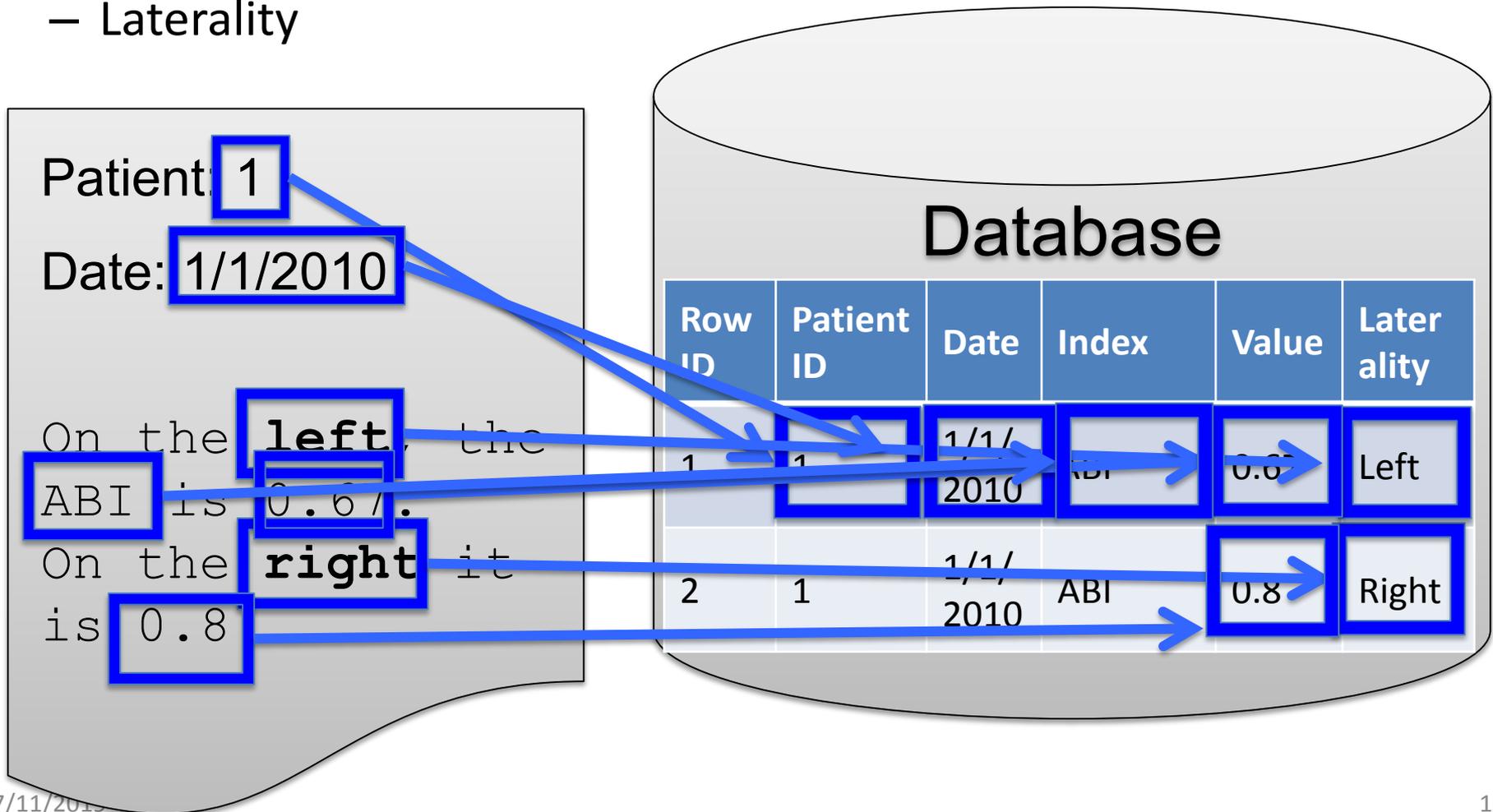


NLP subfields and tasks

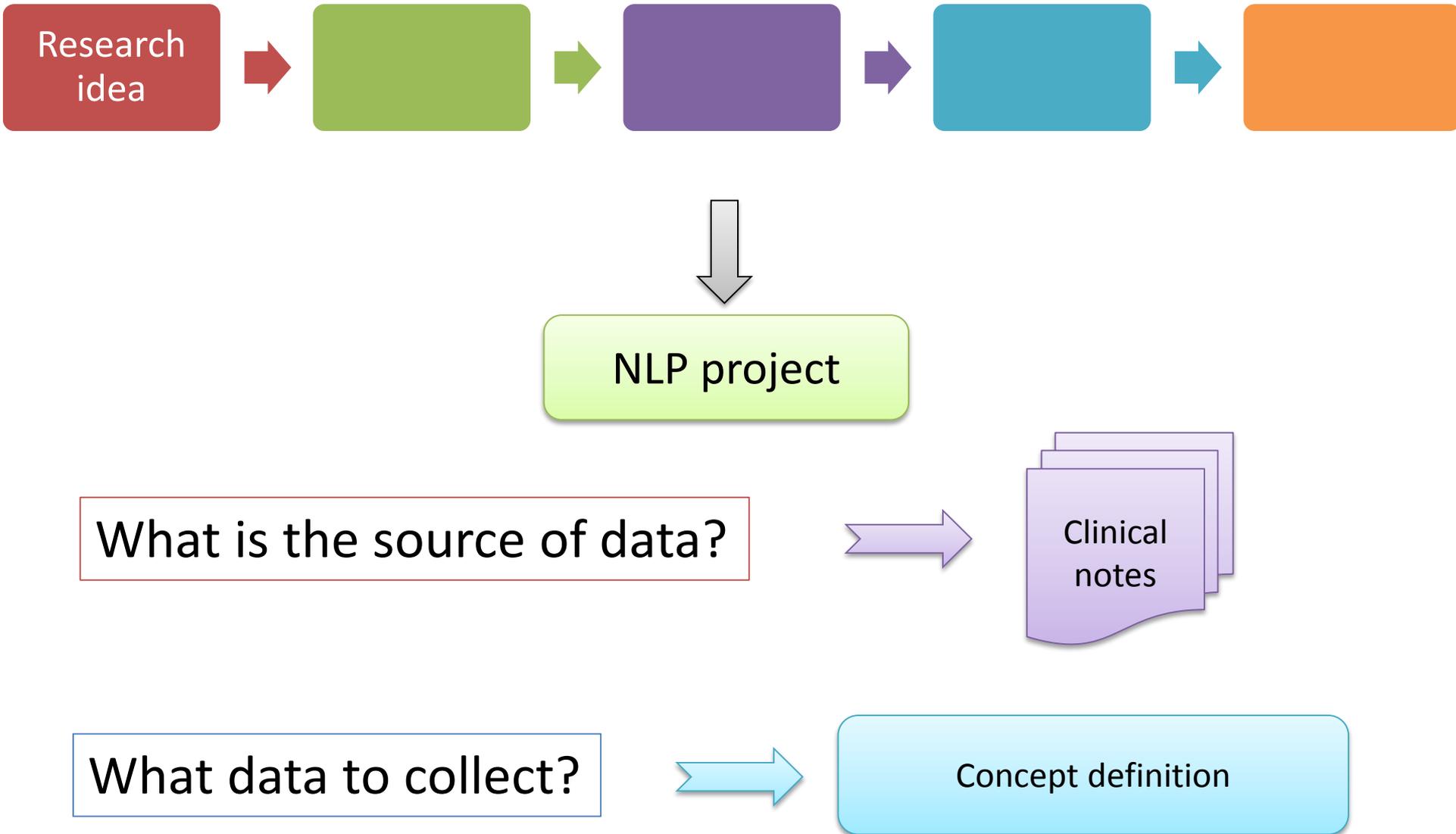


Clinical NLP: Information Extraction

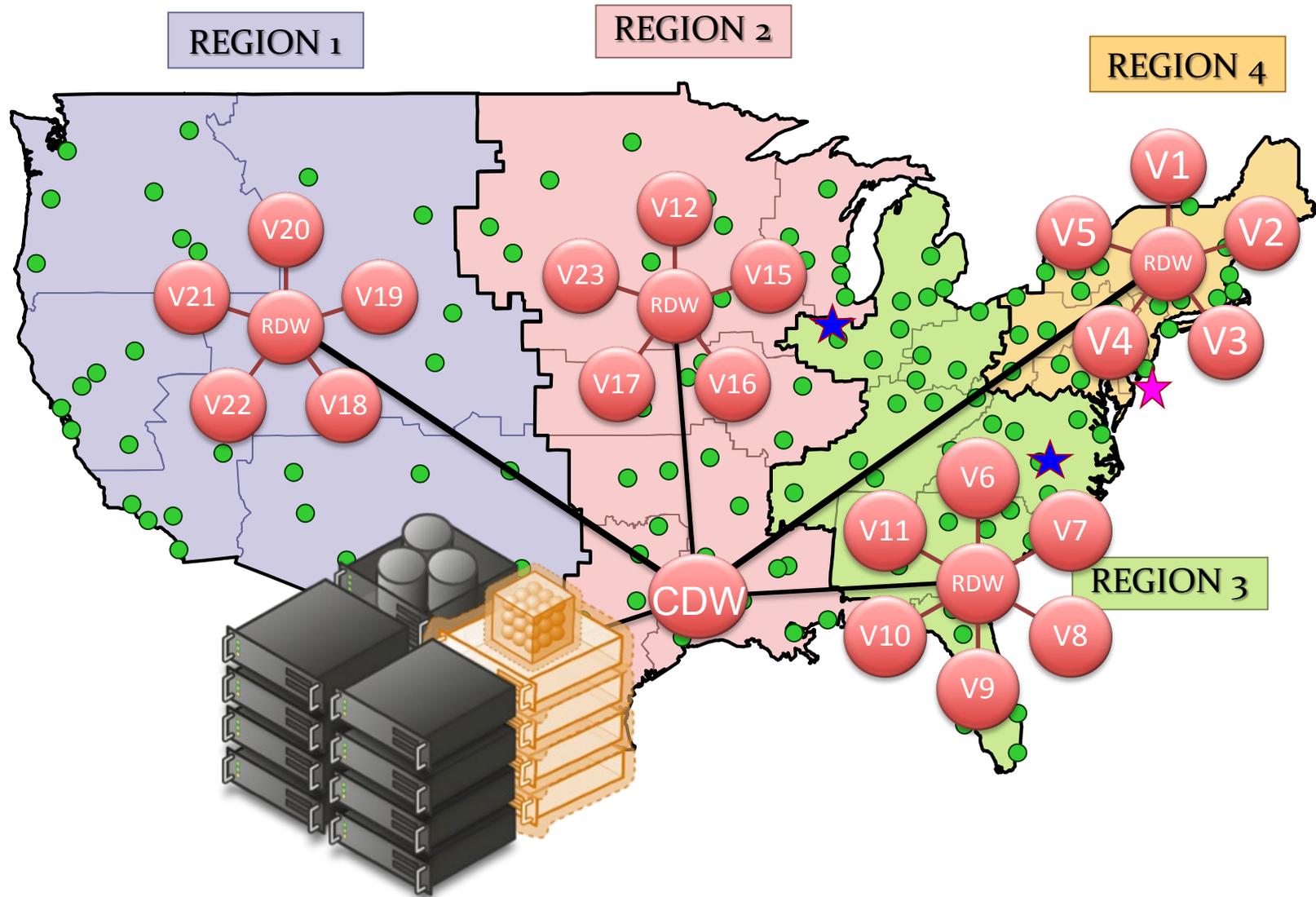
- Ankle Brachial Index
 - Value
 - Laterality



Clinical research project workflow



Corporate Data Warehouse (CDW)



Brief intro to VA clinical notes

- CPRS
 - Computerized Patient Record System
 - Packages: Pharmacy, Lab, Radiology
 - ... Text Integration Utility (TIU)
- CDW
 - Corporate Data Warehouse
 - SQL Server databases

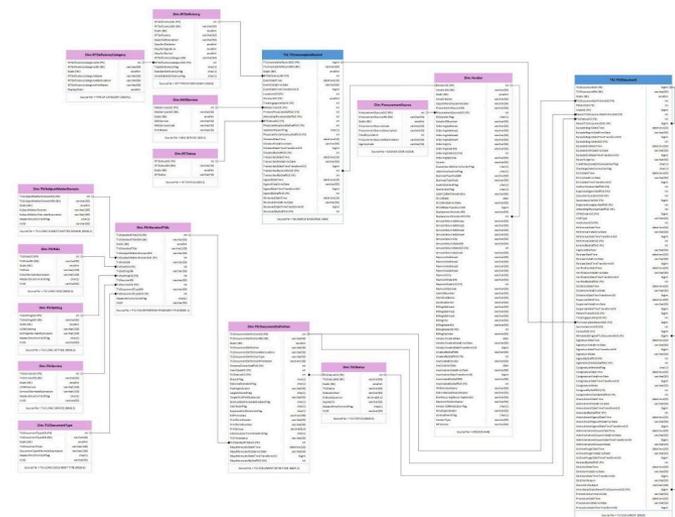
CDW Metadata	
Contains a grouped list of available CDW ER Diagrams and members.	
ERDiagramTitle	
Active Directory 1.0	ER Diagram
Allergy 1.0	ER Diagram
Appointment 2.2	ER Diagram
BeneTravel	ER Diagram
Consult 2.1	ER Diagram
ConsultFactor	ER Diagram
CPRSOrder 1.1	ER Diagram
CPT 2.0	ER Diagram
Dental 1.0 Diagram 1 of 2	ER Diagram
Radiology 1.0 Dimensions	ER Diagram
Radiology 1.0 Fact	ER Diagram
Radiology 1.0 Fact 2	ER Diagram
Radiology 1.0 Patient	ER Diagram
SurgeryINTRA 1.0	ER Diagram
SurgeryPOST 1.1	ER Diagram
SurgeryPRE 1.0	ER Diagram
Text Document Titles 1.0	ER Diagram
Vista Compensation and Pension 1.5	ER Diagram
Vista Data Dictionary 1.0	ER Diagram
Vista Permissions 1.0	ER Diagram
Vista WaitList 1.0	ER Diagram
Vital Sign 1.1	ER Diagram
Women's Health v1.0	ER Diagram

Text Integration Utility (TIU)

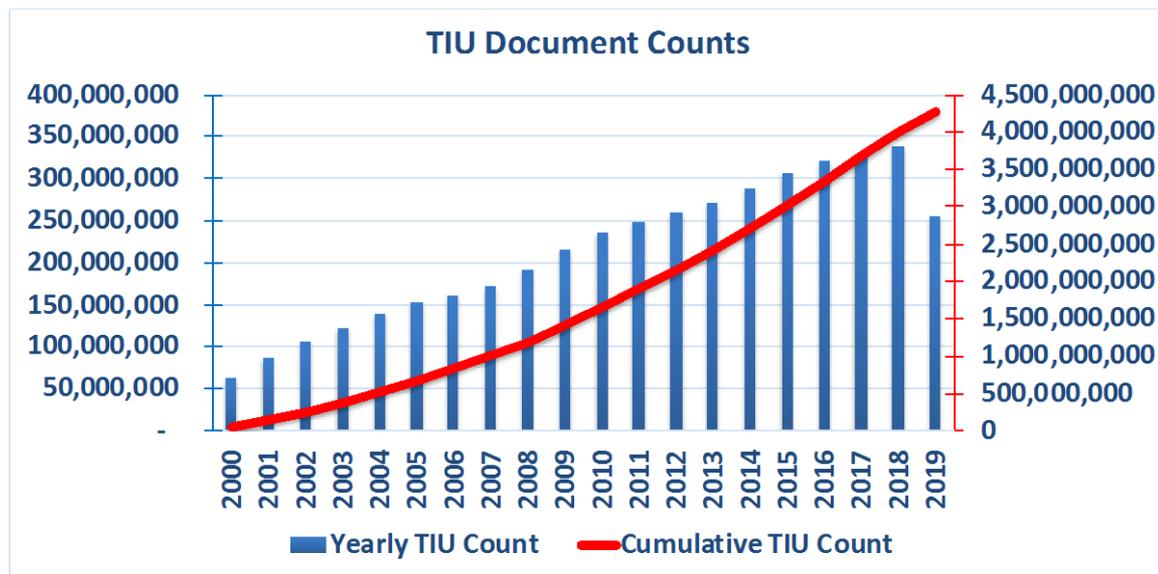
- TIU structured data:
 - TIUDocumentDefinition - document type
 - PatientSID - link to patient data
 - ReferenceDateTime - date of the document

https://vaww.cdw.va.gov/metadata/Reports/ERDiagramsOfViews/TIU_5582.jpg

Text Document Files (TIU) 1.0



- TIU unstructured data:
 - TIU.TIUDocument_8925
 - TIUDocumentSID - bigint
 - Sta3n - smallint
 - ReportText - varchar(max)
 - Raw data, full text index
 - Total count in July 2019
4,265,850,943



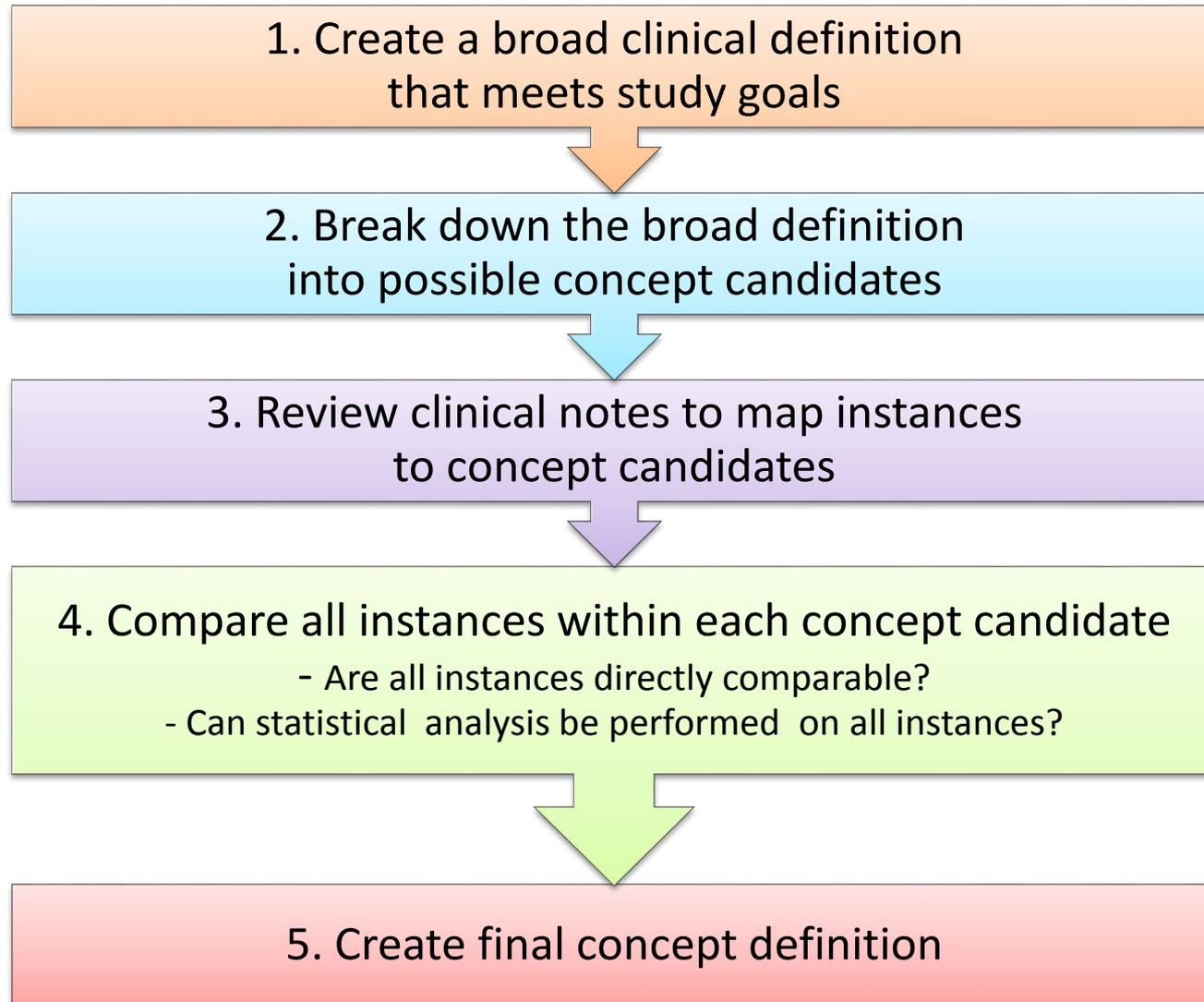
Other sources of narrative data

- Laboratory results
 - Chem.LabPanel.LabPanelComment
 - narrative description of the laboratory results
- Radiology
 - SPatientText.RadiologyReportText.ReportText
 - SPatientText.RadiologyImpressions.ImpressionText
- Surgery
 - Surg.SurgeryPRE.CancelationComments
-

What is a clinical concept?

- Concept = entity, idea, thought, meaning
- Clinical concept - entity targeted by NLP
 - Examples:
 - Diagnosis, symptom, finding
 - Lab value
 - Vital sign measurements
- Characteristics of a clinical concept for NLP
 - single, **unified** meaning across all instances of the concept
 - instances of the concept are directly comparable to each other
 - **project specific** definition
 - documented in electronic medical record (EMR)

Data driven concept definition process

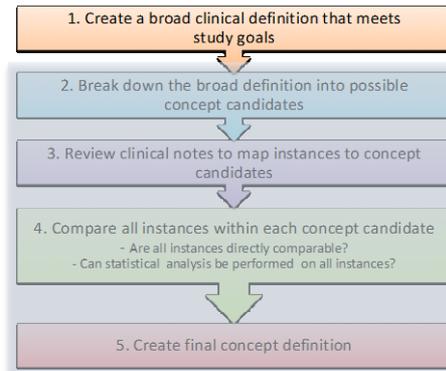


Broad clinical definition

Use case:

find patients with Peripheral Arterial Disease
(PAD)

Peripheral Arterial Disease



Concept candidates

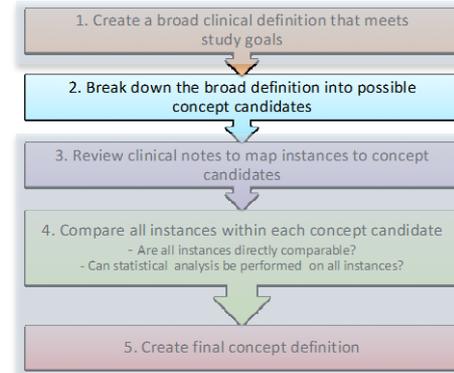
Use case:

find patients with Peripheral Arterial Disease (PAD)



Mentions of PAD in clinical notes

Ankle Brachial Index



Review clinical notes

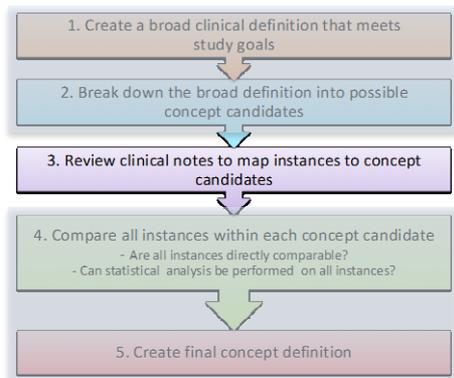
Use case:

find patients with Peripheral Arterial Disease (PAD)

Peripheral Arterial Disease

Mentions of PAD in clinical notes

Ankle Brachial Index



Patient diagnosed with **Peripheral Arterial Disease** ...

PROBLEMS: **Peripheral Vascular Disease**

... suggests left subclavian **arterial occlusive disease**...

... is a 85-year-old male with CAD s/p PTCA LCx, **PAD**, COPD...

... History of Present Illness:
56yo male with a history of **peripheral vascular disease** ...

Review clinical notes

Use case:

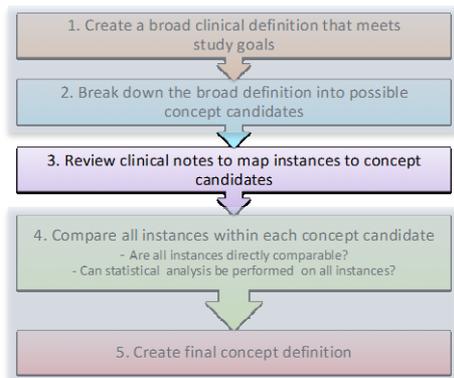
find patients with Peripheral Arterial Disease (PAD)

Peripheral Arterial Disease

Mentions of PAD in clinical notes

Ankle Brachial Index

- ... relief with ice packs or heating **pad** ...
- ... wait for arterial doppler to **rule out** PAD
- ... **no evidence** to support a diagnosis of peripheral vascular disease...
- ... **no significant** hemodynamic peripheral arterial disease ...



Review clinical notes

Use case:

find patients with Peripheral Arterial Disease (PAD)

Peripheral Arterial Disease

Mentions of PAD in clinical notes

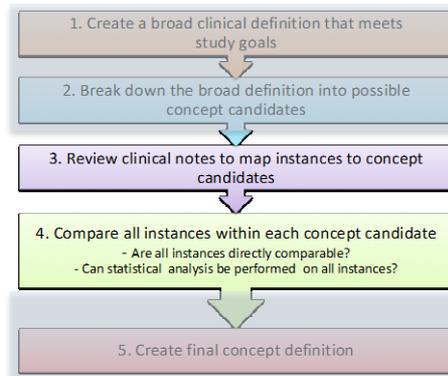
Ankle Brachial Index

On the **right**, the ABI is 0.6.
On the **left** it is 0.44.

Right ABI was 0.4, **resting**.

... ABI was 0.88 on the right which decreased to 0.46 **after exercise**. ...

... The ABI on the right based on the **DP artery** is 0.55 and on the left based on the **PT artery** is 0.53...



Review clinical notes

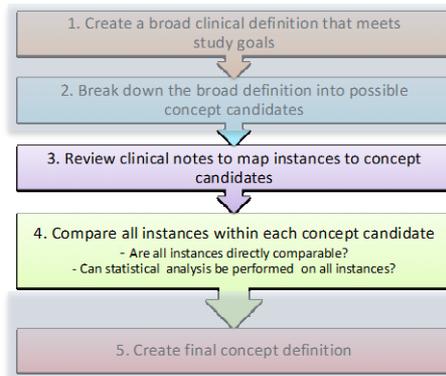
Use case:

find patients with Peripheral Arterial Disease (PAD)

Peripheral Arterial Disease

Mentions of PAD in clinical notes

Ankle Brachial Index



Ankle brachial index was performed:
r. arm systolic - 178
r. ankle - 90

Findings: R ABI = 0.41
(previously 0.49)

Comparison: R ABI is 0.41, but in 2010 ABI was 0.49

Final concept definitions

Use case:

find patients with Peripheral Arterial Disease (PAD)

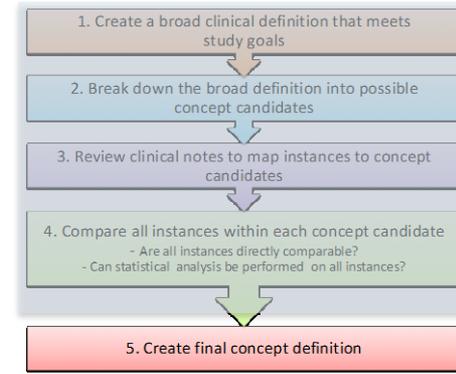
Peripheral Arterial Disease

Mentions of PAD in clinical notes

Ankle Brachial Index

Concept: Peripheral Arterial Disease
Definition: **Explicit historical or current affirmed** mentions of PAD **including** peripheral vascular disease, and vascular occlusions.

Concept: ABI
Definition: **Explicit historical or current affirmed** mentions of ABI **numeric values, regardless of** resting/after exercise, **regardless of** specific arteries
Attributes: Laterality (left, right, bilateral, not specified)



Concept sheets

- To formalize the definitions and as a communication tool
- Elements
 - Concept name
 - Detailed definition
 - Attributes
 - Examples: What it is and what it is not

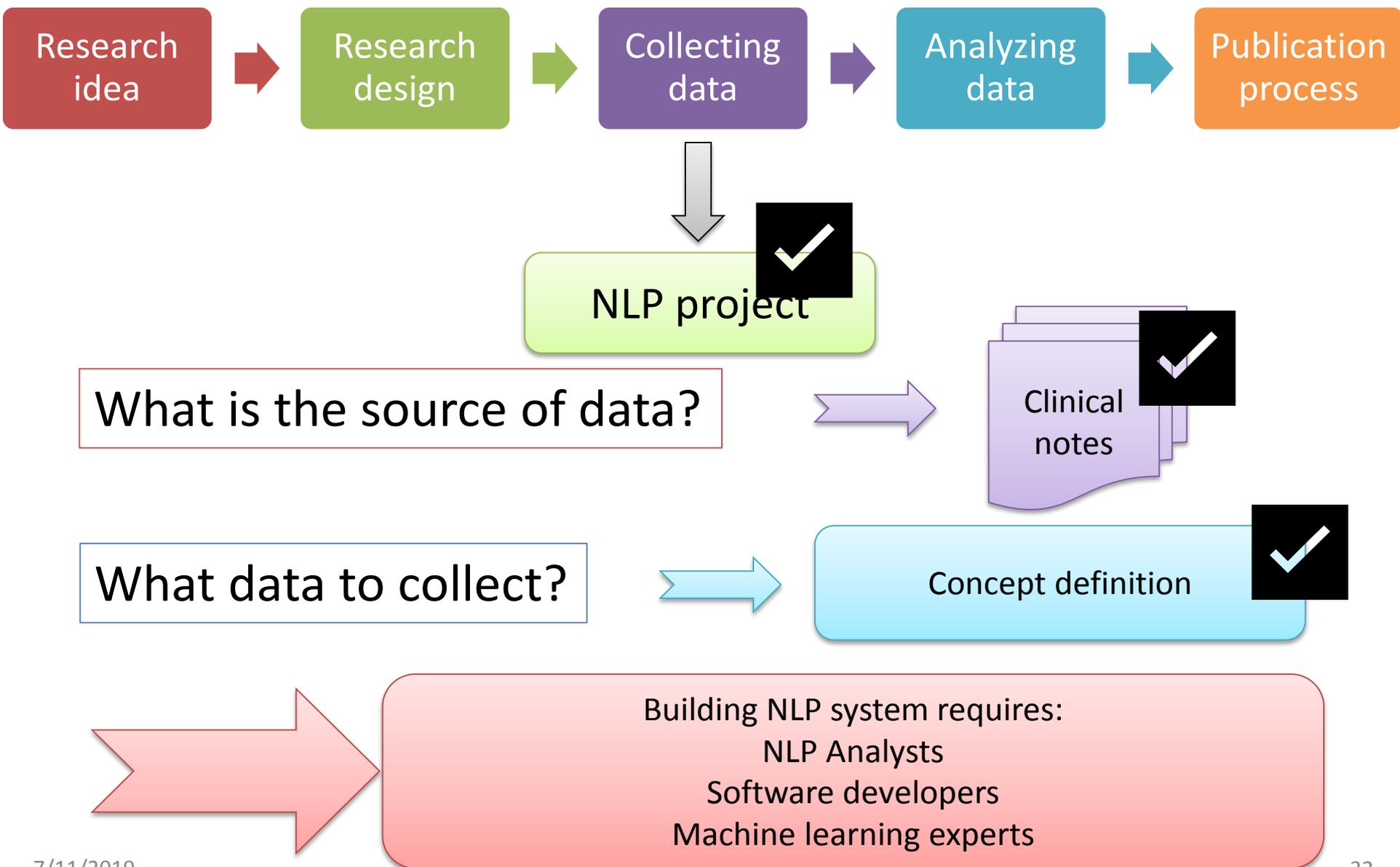
Concept: Ankle Brachial Index

Definition: Explicit **historical or current affirmed** mentions of ABI **numerical values**, **regardless** of resting/after exercise, **regardless** of specific arteries

Attributes: Laterality (left, right, bilateral, not specified)

Exclude: pressure values

Clinical research project workflow



Example: NLP system design

Goal:

Extract Ankle Brachial Index

Steps:

1. Find all ABI and attribute terms

FINAL REPORT

ARTERIAL DOPPLER LOWER EXTREMITY

REASON: Diminished pulses and pain.

Doppler evaluation was performed of both lower extremity arterial systems at rest. On the right, all waveforms are monophasic. On the left, waveforms are triphasic at the femoral and popliteal levels. On the right, no ankle brachial index was calculated. Pulse volume recordings are essentially flat from the low thigh down. On the left, the ABI is 0.54 with a significant decrease at the calf at the metatarsals.

IMPRESSION: Significant aortoiliac and likely multisegmental occlusive disease with severe flow deficit to the whole right lower extremity. On the left, there is significant SFA/tibial artery occlusive disease.

Example: NLP system design

Goal:

Extract Ankle Brachial Index

Steps:

1. Find all **ABI** and **attribute** terms

FINAL REPORT

ARTERIAL DOPPLER LOWER EXTREMITY

REASON: Diminished pulses and pain.

Doppler evaluation was performed of both lower extremity arterial systems at rest. On the **right**, all waveforms are monophasic. On the **left**, waveforms are triphasic at the femoral and popliteal levels. On the **right**, no **ankle brachial index** was calculated. Pulse volume recordings are essentially flat from the low thigh down. On the **left**, the **ABI** is 0.54 with a significant decrease at the calf at the metatarsals.

IMPRESSION: Significant aortoiliac and likely multisegmental occlusive disease with severe flow deficit to the whole **right** lower extremity. On the **left**, there is significant SFA/tibial artery occlusive disease.

Example: NLP system design

Goal:

Extract Ankle Brachial Index

Steps:

1. Find all **ABI** and **attribute** terms
2. Find all **numeric** values

FINAL REPORT

ARTERIAL DOPPLER LOWER EXTREMITY

REASON: Diminished pulses and pain.

Doppler evaluation was performed of both lower extremity arterial systems at rest. On the **right**, all waveforms are monophasic. On the **left**, waveforms are triphasic at the femoral and popliteal levels. On the **right**, no **ankle brachial index** was calculated. Pulse volume recordings are essentially flat from the low thigh down. On the **left**, the **ABI** is **0.54** with a significant decrease at the calf at the metatarsals.

IMPRESSION: Significant aortoiliac and likely multisegmental occlusive disease with severe flow deficit to the whole **right** lower extremity. On the **left**, there is significant SFA/tibial artery occlusive disease.

Example: NLP system design

Goal:

Extract Ankle Brachial Index

Steps:

1. Find all **ABI** and **attribute** terms
2. Find all **numeric** values
3. Build candidate links

FINAL REPORT

ARTERIAL DOPPLER LOWER EXTREMITY

REASON: Diminished pulses and pain.

Doppler evaluation was performed of both lower extremity arterial systems at rest. On the **right**, all waveforms are monophasic. On the **left**, waveforms are triphasic at the femoral and popliteal levels. On the **right**, no **ankle brachial index** was calculated. Pulse volume recordings are essentially flat from the low thigh down. On the **left**, the **ABI** is **0.54** with a significant decrease at the calf at the metatarsals.

IMPRESSION: Significant aortoiliac and likely multisegmental occlusive disease with severe flow deficit to the whole **right** lower extremity. On the **left**, there is significant SFA/tibial artery occlusive disease.

Example: NLP system design

Goal:

Extract Ankle Brachial Index

Steps:

1. Find all **ABI** and **attribute** terms
2. Find all **numeric** values
3. Build candidate links
4. Evaluate relationships

FINAL REPORT

ARTERIAL DOPPLER LOWER EXTREMITY

REASON: Diminished pulses and pain.

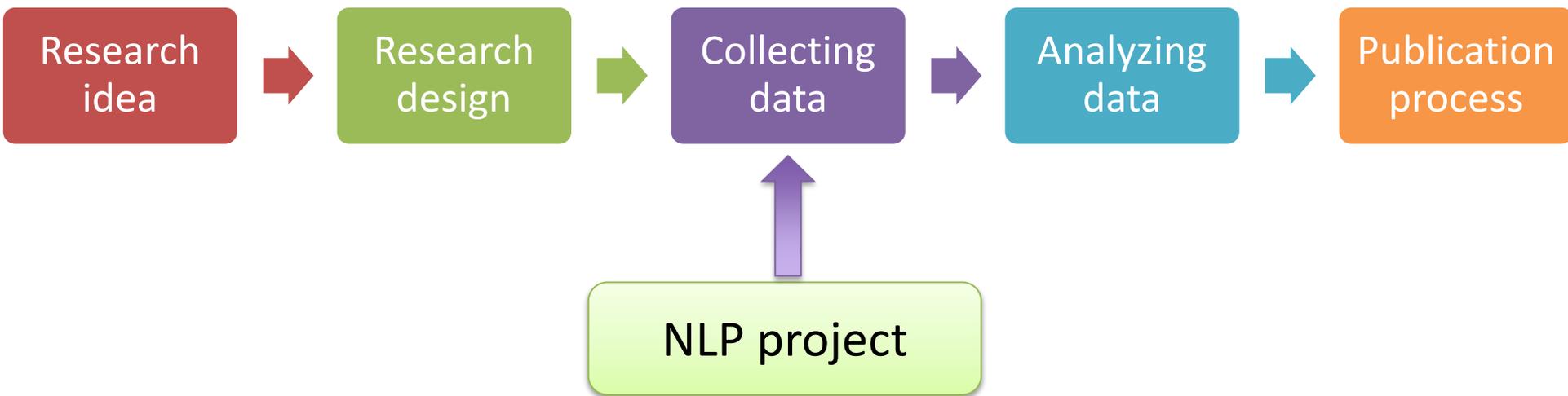
Doppler evaluation was performed of both lower extremity arterial systems at rest. On the **right**, all waveforms are monophasic. On the **left**, waveforms are triphasic at the femoral and popliteal levels. On the **right**, no **ankle brachial index** was calculated. Pulse volume recordings are essentially flat from the low thigh down. On the **left**, the **ABI** is **0.54** with a significant decrease at the calf at the metatarsals.

IMPRESSION: Significant aortoiliac and likely multisegmental occlusive disease with severe flow deficit to the whole **right** lower extremity. On the **left**, there is significant SFA/tibial artery occlusive disease.

Result of Information Extraction

Row ID	Patient ID	Date	Index	Value	Laterality
1	1	1/1/2010	ABI	0.67	Left
2	1	1/1/2010	ABI	0.8	Right
3	1	1/3/2010	ABI	0.67	Left
4	1	6/1/2010	ABI	<0.8	
5	1	1/1/2015	ABI	0.67	Left
6	1	1/1/2015	ABI	0.58	Left

Clinical research project workflow



Questions about VA data and text processing in the VA?

Email: vinci@va.gov

