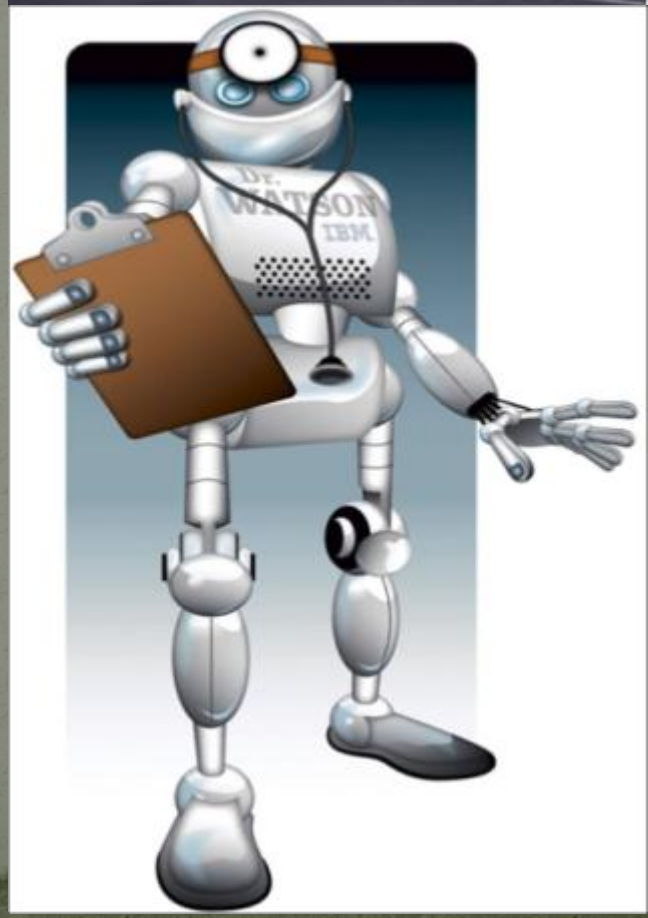
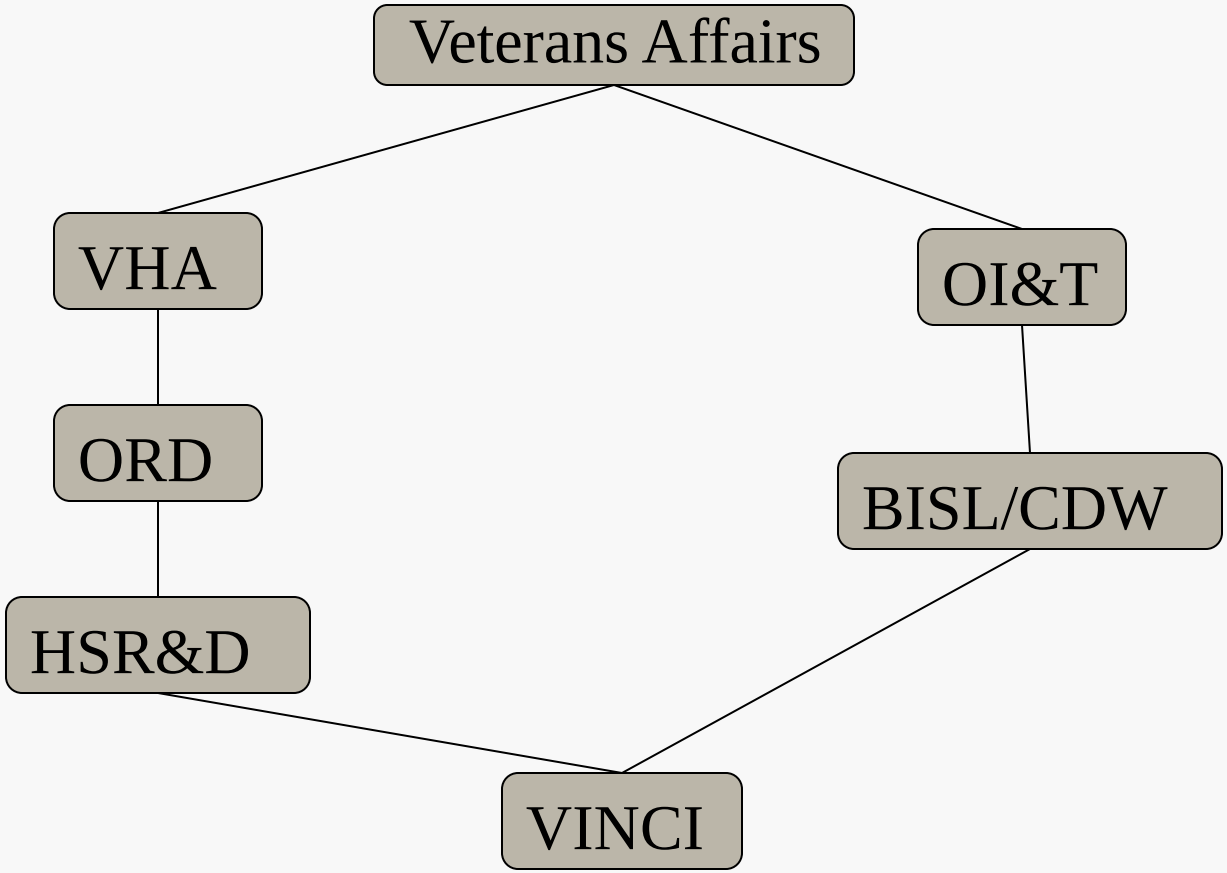
VINCI and the Center for Computational Intelligence at University of Maryland



VA Informatics and Computing Infrastructure
VINCI Organization

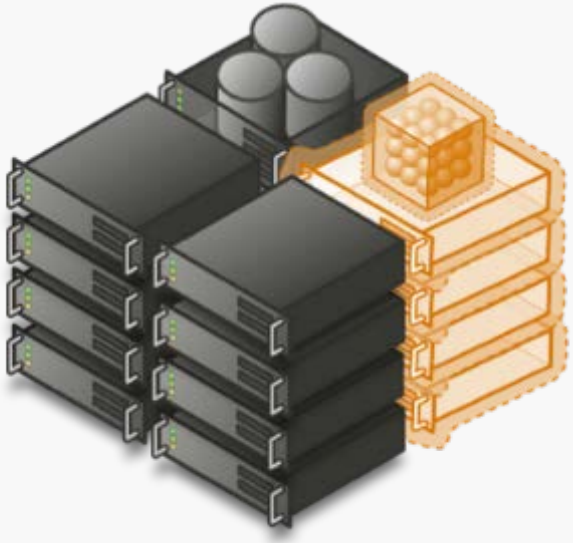


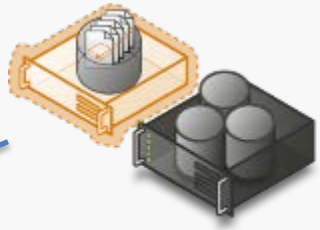
A Research / IT Partnership

- Establish a National Research Infrastructure
- Develop New Data Privacy & Security Capabilities
- Develop Remote Analysis Capability
- Prepare for Future Data Capabilities
- Partner With Other VA and Non-VA Entities

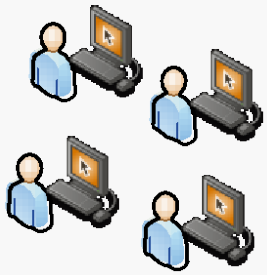
VINCI

Corporate Data Warehouse



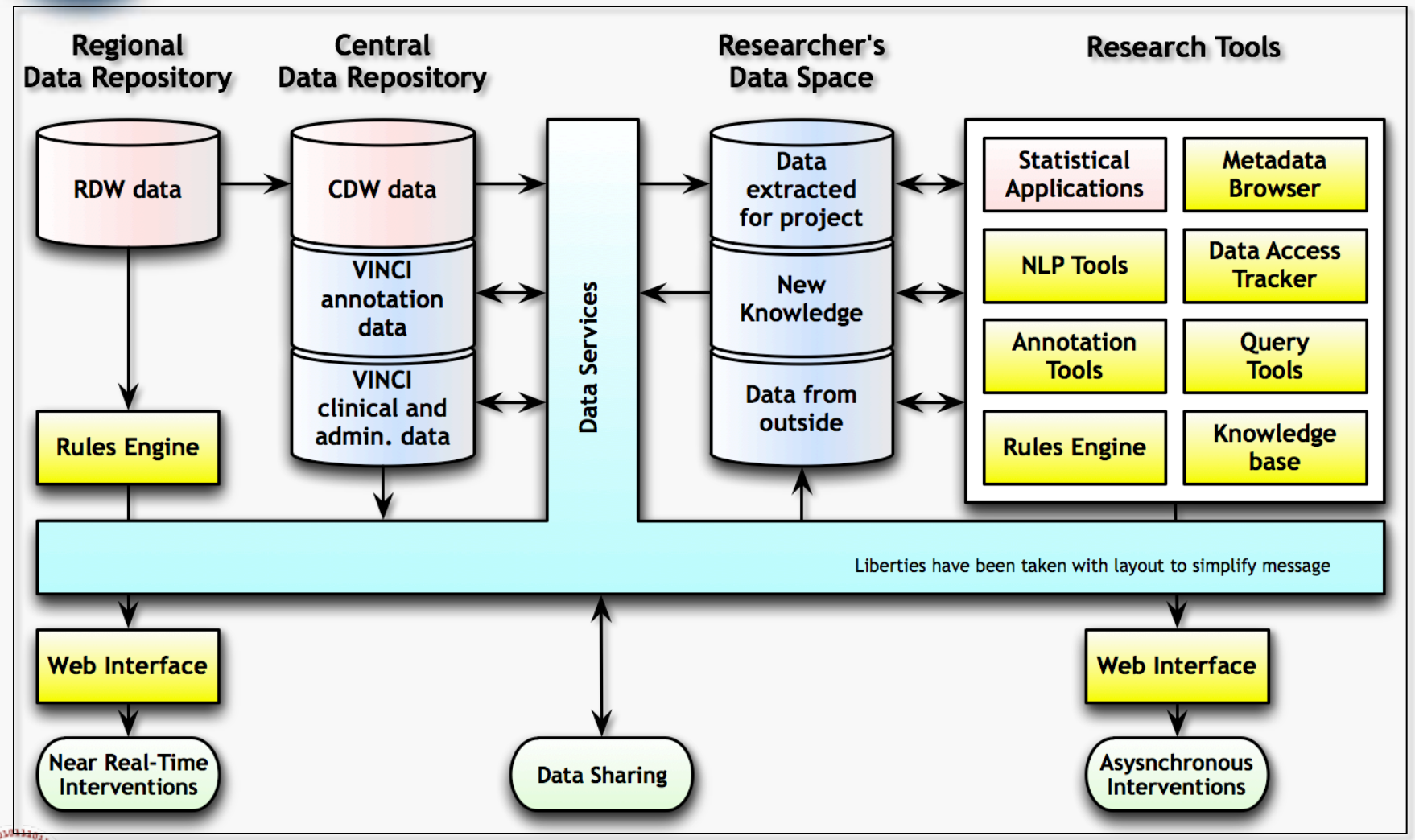


Specialty VA data sources

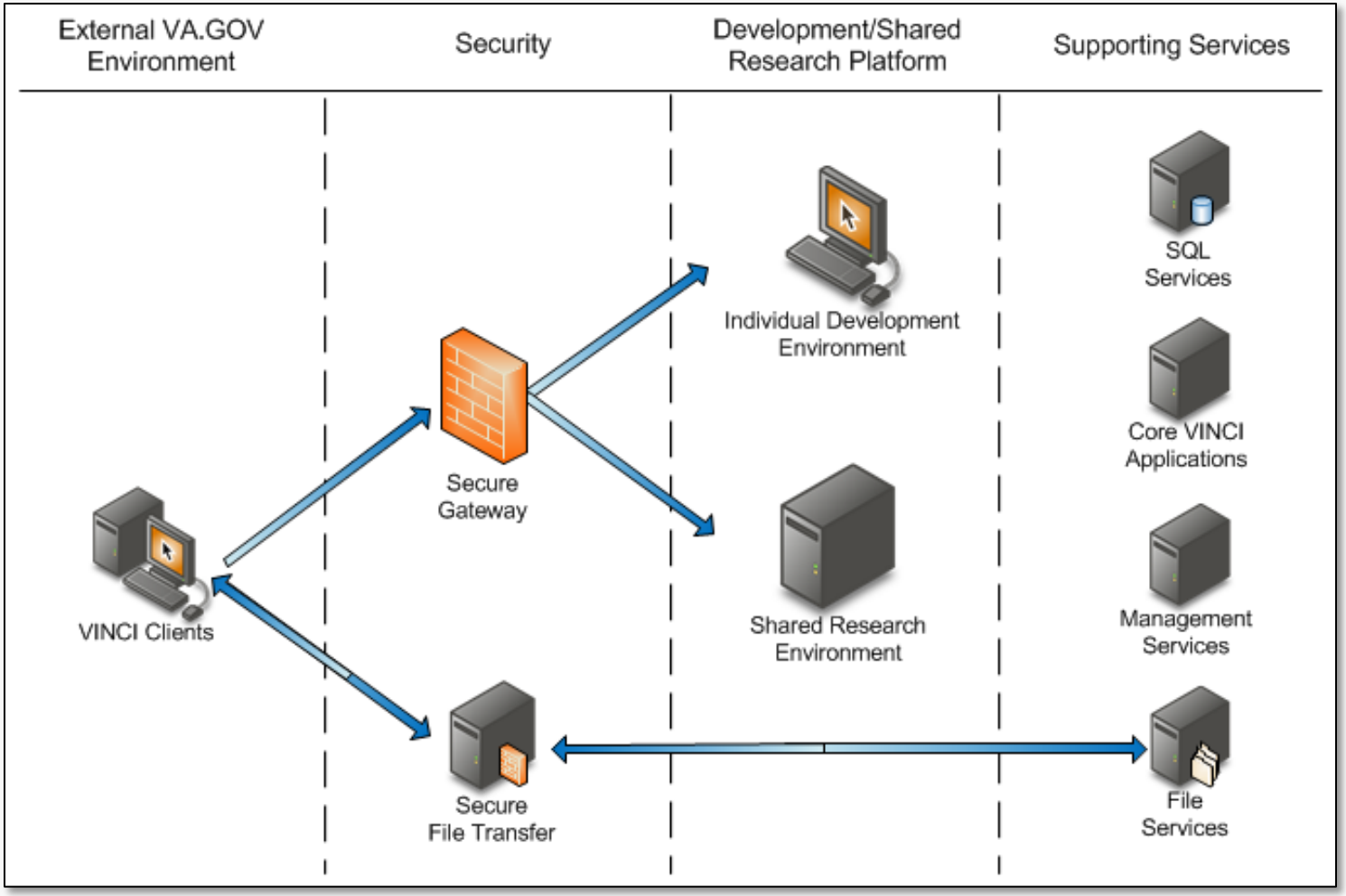


Local VistA implementations

A Research / IT Partnership



VINCI Architecture



Data Extraction

- Research Study Team uploads cohort to a secure site or VINCI creates cohort per requirements
- Research Study Team completes the Data Selection Forms on correspondence site
- NDS approved data domains are extracted and provided
- Analysis performed by research project team staff

External Data

- Research Study Team may upload other data for analysis into project database or workspace
- Secure data upload process
- Optional direct database upload
- VINCI data managers work on behalf of research team to upload data from other data providers

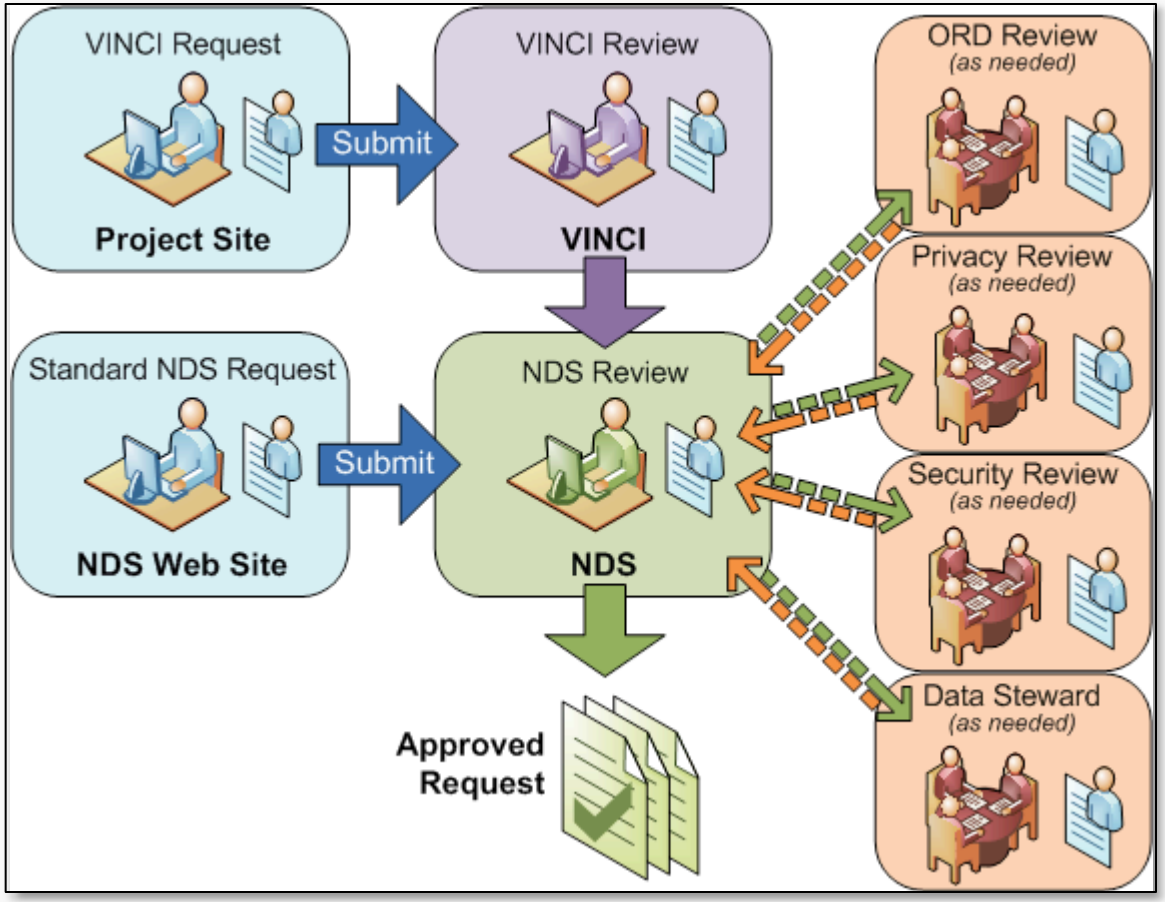
Data Processing/Analysis tools

- SQL Server as primary data store
- Multiple high performance servers
- Most data queries performed in SQL Server
- Accessible by all analysis software
- SSIS, SSRS, SSAS (SQL server integration, reporting, and analysis)
- High speed intra-server network will allow distributed queries

SAS

- High performance SAS server
  - 2 TB of RAM, 64 cores, 2 TB SSD
  - Launch grid jobs
- SAS grid – very large data analysis work
  - 10 high performance servers
  - Most advanced SAS implementation in VA
  - Dedicated SAN
  - Additional SAS modules
  - SAS knowledge base SharePoint site
- Dedicated SAS administrator

Data Approvals

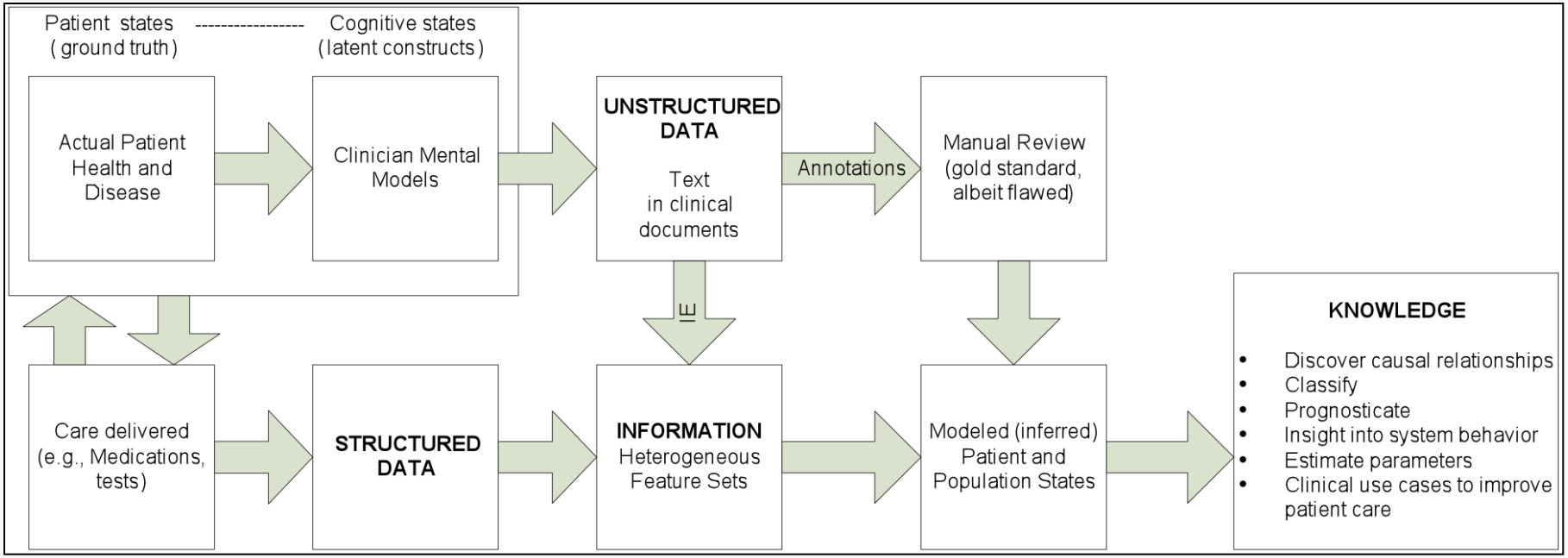


Data Access & Security

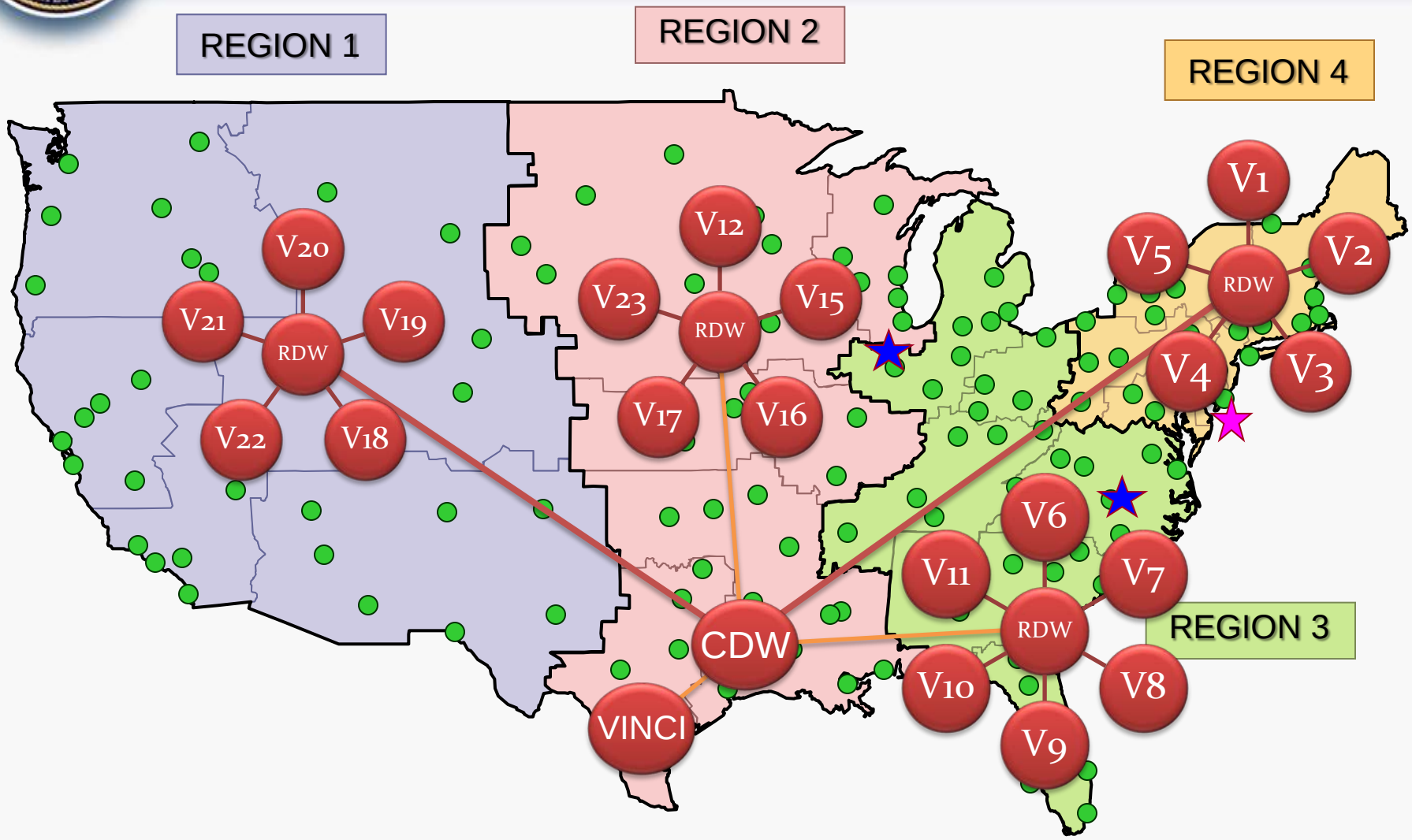
- Access groups created based on IRB and NDS approved research team
- Only research team members have access to the data
- Data stored on secure VINCI servers
- Regular data backup and archiving
- Workspace vs. collaboration site
- Project work can be performed in VINCI
- Export final result & publication

- Patient demographic information
- Vital signs
- Diagnoses and procedures from patient encounters
- Outpatient pharmacy data
- Laboratory values
- Immunizations
- Microbiology reports
- Text notes (including progress notes, discharge summaries, radiology reports)
- VA Decision Support System (DSS) in database tables linked with other VINCI data or as the original SAS files
- VHA Medical SAS datasets

Using All You Have



VINCI


REGION 1
REGION 2
REGION 4
REGION 3
V20
V21
RDW
V19
V22
V18
V12
V23
RDW
V15
V17
V16
V1
V5
RDW
V2
V4
V3
V6
V11
V7
RDW
V10
V8
V9
CDW
VINCI

Data Standardization

MODERATE STAPHYLOCOCCUS AUREUS
E. COLI
> 100 CFU STAPHYLOCOCCUS AUREUS
16 COLONIES SERRATIA SP

GRAM NEG ORGANISM
ANAEROBIC GRAM NEGATIVE RODS
GRAM NEG BACILLI

ENTEROCOCCUS FAECALIS - VANCOMYCIN RESISTANT
ENTEROCOCCUS FAECIUM
ENTEROCOCCUS SPP.

INFLIXIMAB 100MG/VIL (PF) INJ

INFLIXIMAB 100MG/VIL PF INJ

INFLIXIMAB 100MG INJ

INFLIXIMAB 100MG VIAL

INFLIXIMAB 100MG

INFLIXIMAB 100MG/20ML INJ

INFLIXIMAB 100MG/VIAL INJ

INFLIXIMAB 100MG/20ML VIAL

INFLIXIMAB 100MG IV

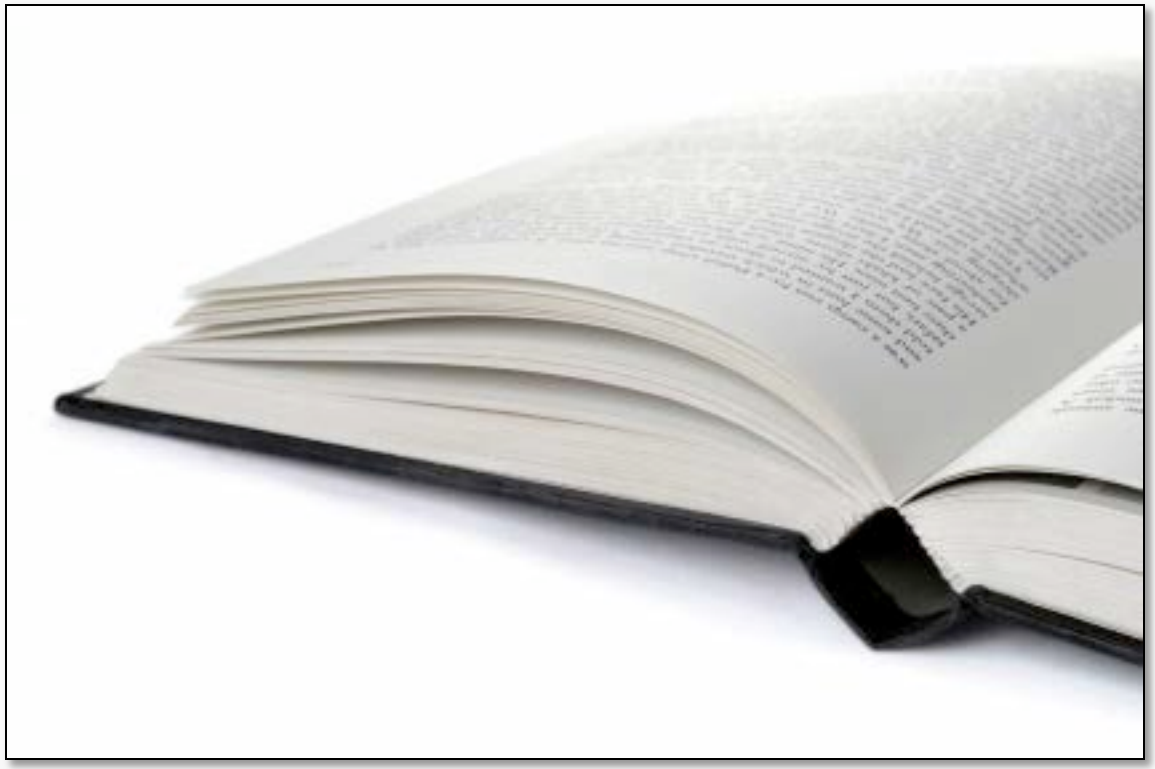
ZZINFLIXIMAB 100MG INJ

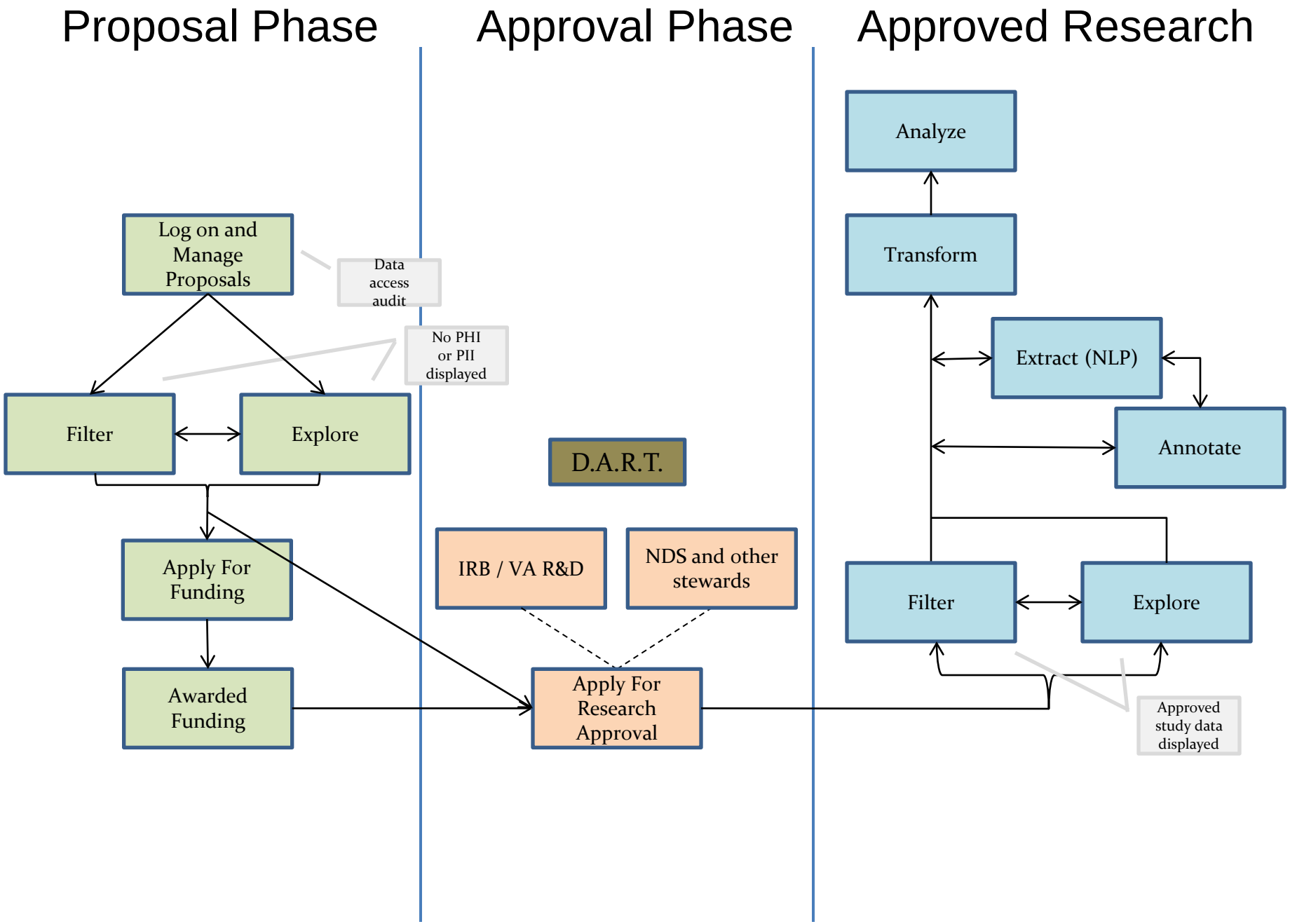
ZZINFLIXIMAB 100MG/20ML

Understand the Target Study Population

Understand the Care Workflow

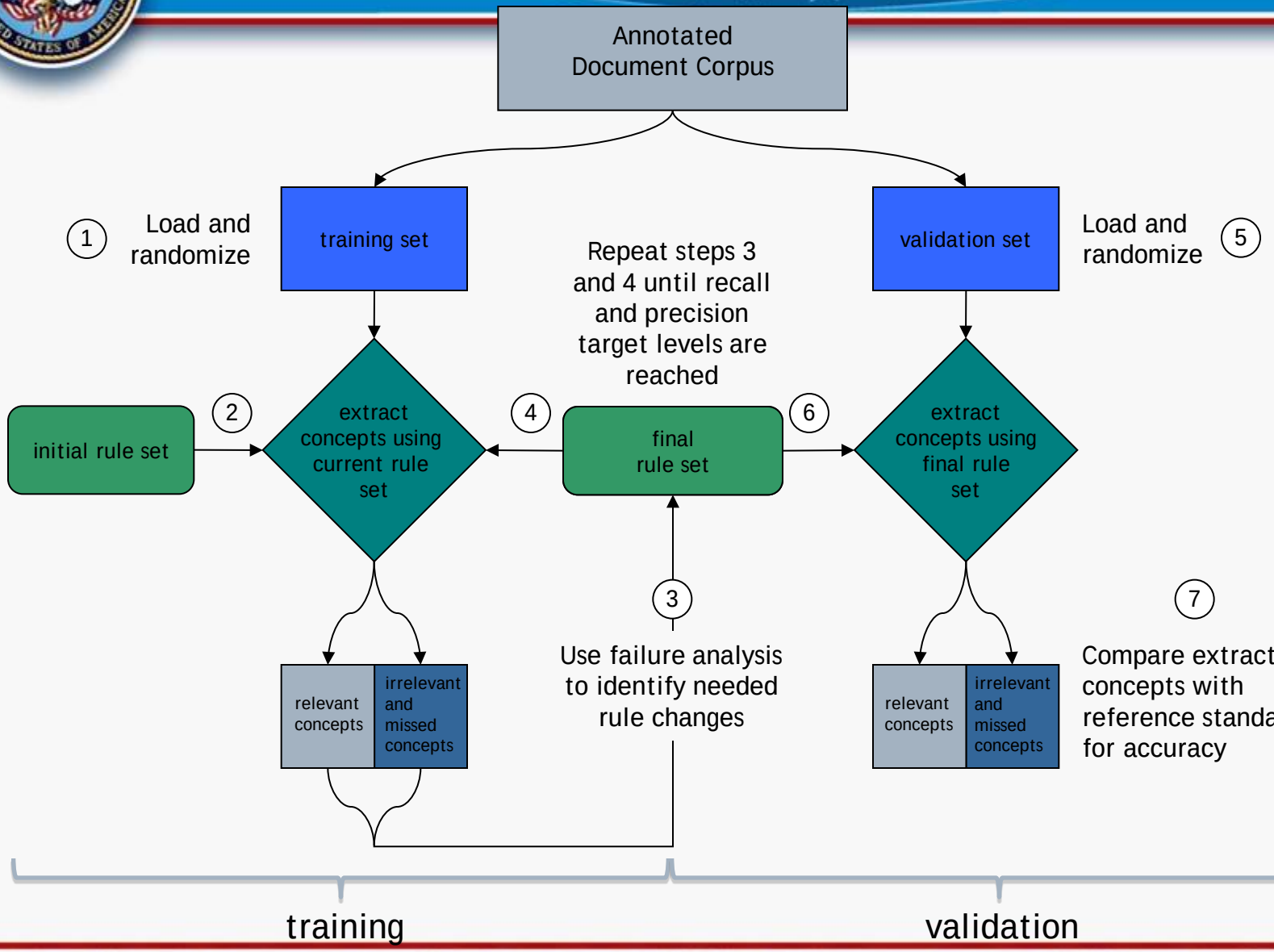
Understand the Underlying Data



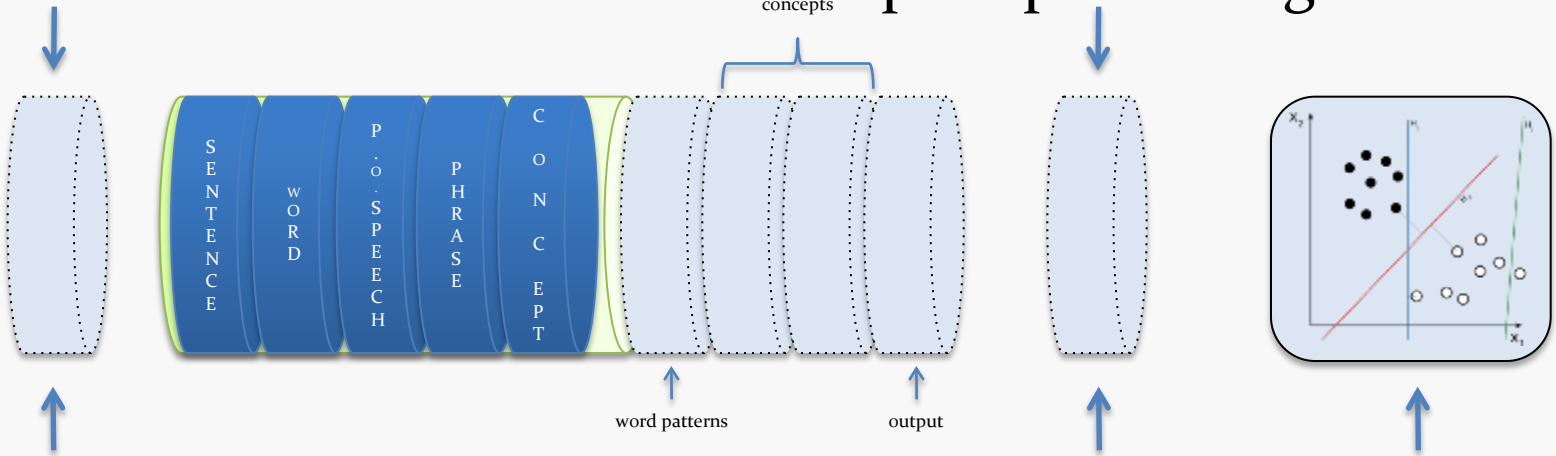

Proposal Phase
Approval Phase
Approved Research
Log on and Manage Proposals
Data access audit
No PHI or PII displayed
Filter
Explore
Apply For Funding
Awarded Funding
D.A.R.T.
IRB / VA R&D
NDS and other stewards
Apply For Research Approval
Analyze
Transform
Extract (NLP)
Annotate
Filter
Explore
Approved study data displayed

Why natural language processing?

- Increasing demands for more detailed clinical data
  - Quality Measures
  - Evidence-based medicine
  - Phenotyping for genomic-related analysis
  - Biosurveillance
- The majority of EMR data is free text


Annotated Document Corpus
1
Load and randomize
training set
validation set
Load and randomize
5
Repeat steps 3 and 4 until recall and precision target levels are reached
initial rule set
2
extract concepts using current rule set
4
final rule set
6
extract concepts using final rule set
3
Use failure analysis to identify needed rule changes
7
Compare extracted concepts with reference standard for accuracy
relevant concepts
irrelevant and missed concepts
relevant concepts
irrelevant and missed concepts
training
validation

An NLP Pipeline



Pipeline Issues

Base modules have trouble with:

- Templates
- Incomplete Sentences
- Jargon

Pipeline Issues

Project-specific modules are:

- Limited by base module output
- Affected by propagation of errors from previous modules

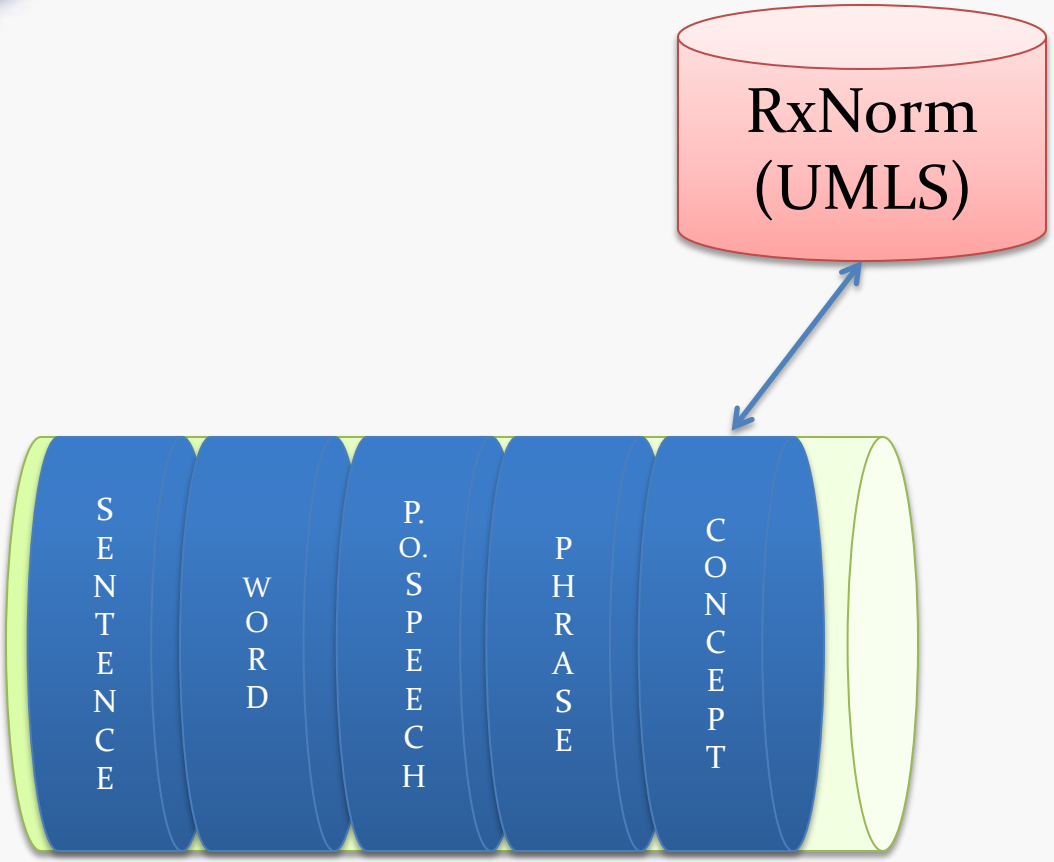
Lookup Issues

- Target does not exist in database

  Semantic bootstrapping

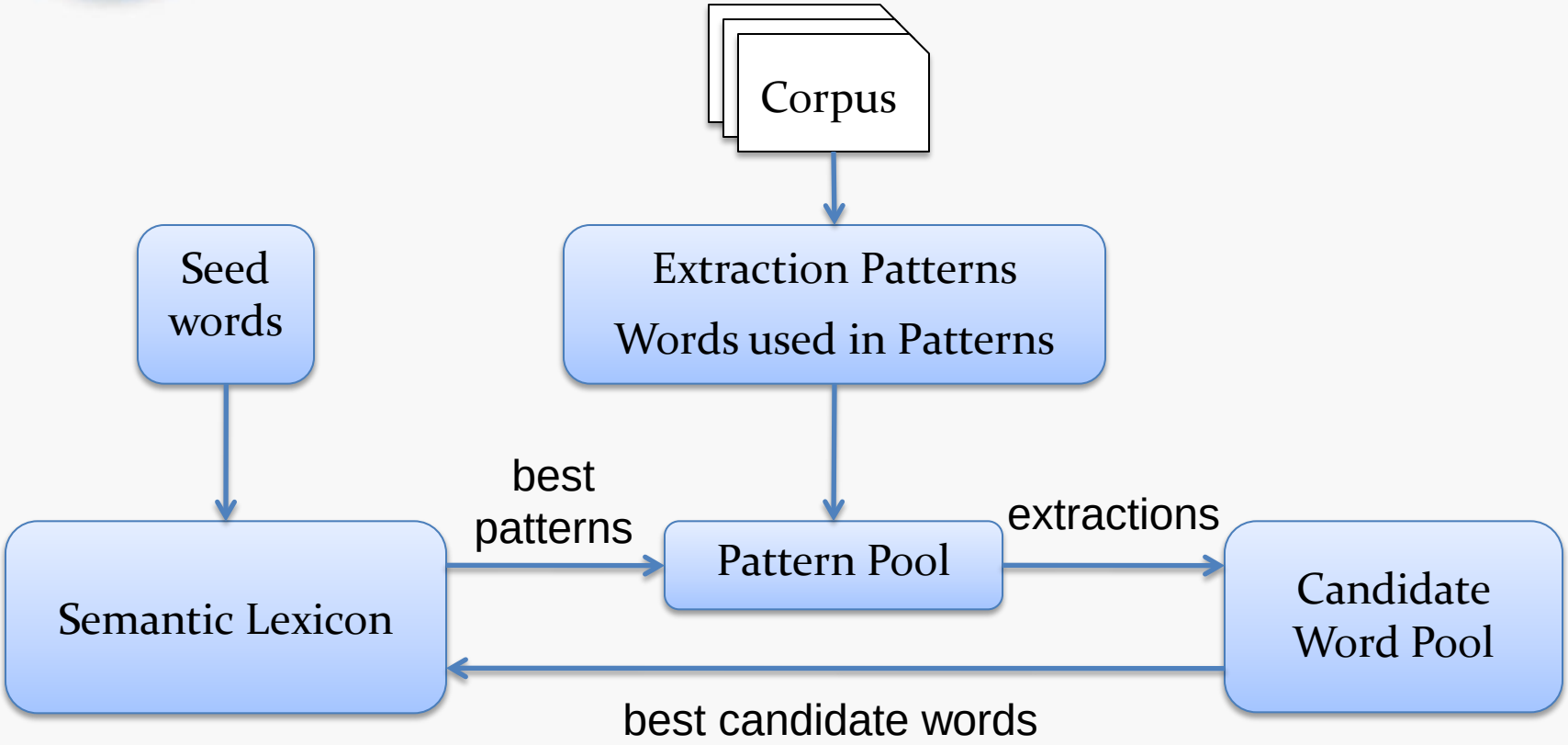
- Lookup method insufficient to find target

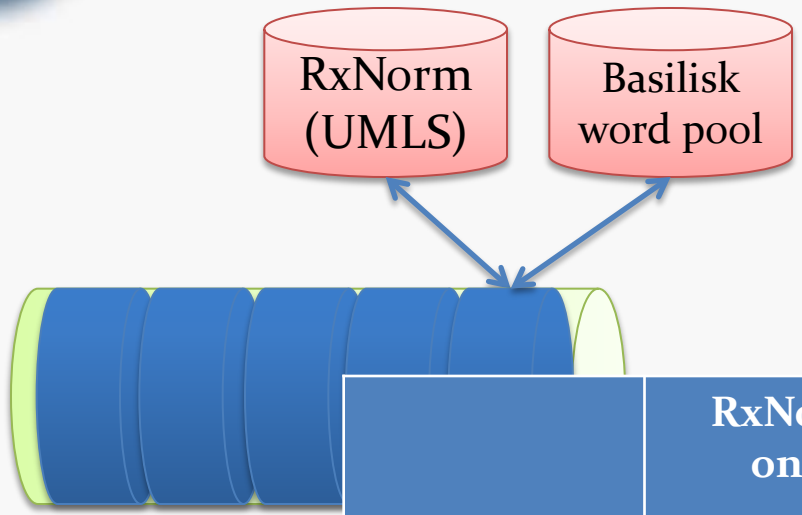
  Structural cues



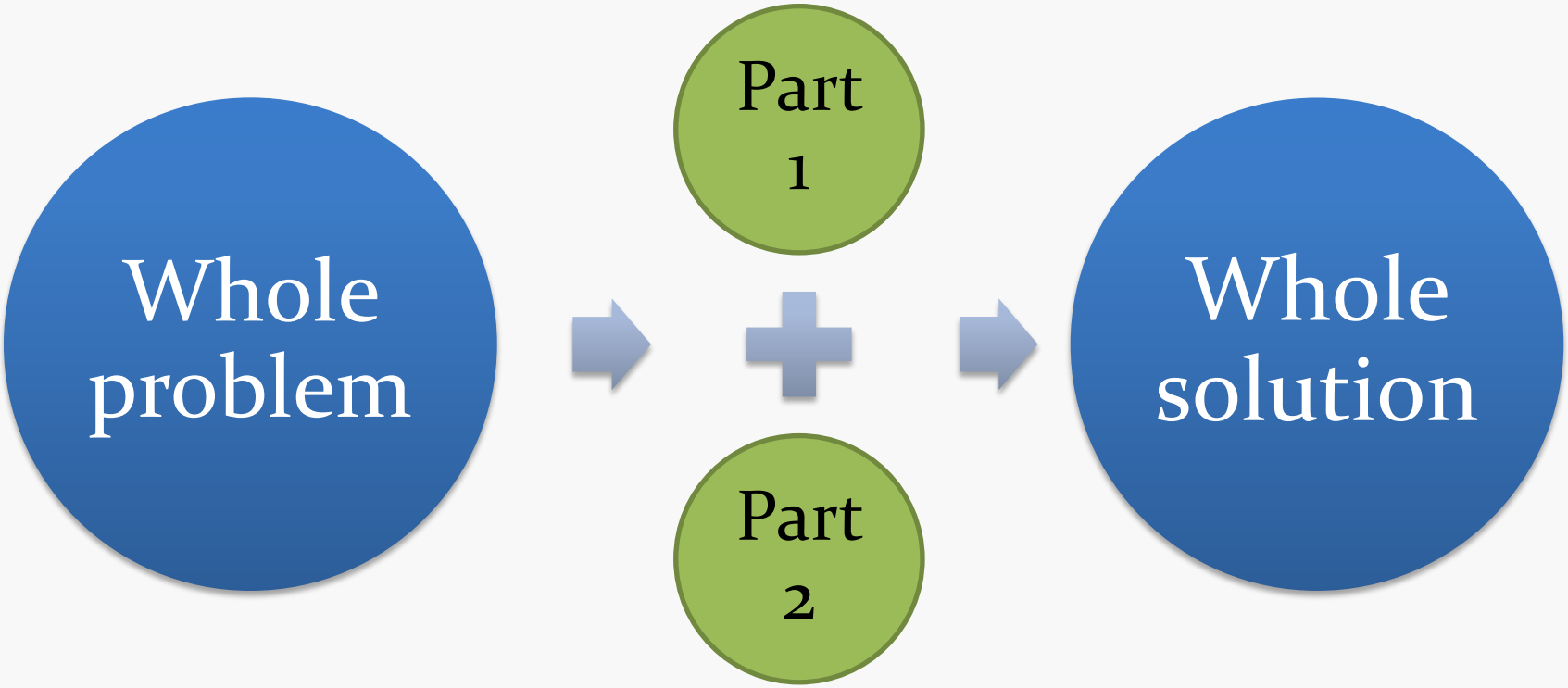
Medication Name	
Precision	61.86%
Recall	66.93%
F1	64.29%

BASILISK


Corpus
Seed words
Extraction Patterns
Words used in Patterns
best patterns
extractions
Semantic Lexicon
Pattern Pool
Candidate Word Pool
best candidate words



	RxNorm only	RxNorm + Basilisk	1-pass Filtering
Precision	61.86%	66.74%	71.95%
Recall	66.93%	82.80%	82.80%
F1	64.29%	73.91%	77.00%


Whole problem
Part 1
Part 2
Whole solution

- Craft *very specific* rules to handle specific situations
- Craft *highly precise* rules to determine which rules to apply in which situations
- Have general rules to apply elsewhere

----MICROBIOLOGY----

Accession: VA 123 Received: Oct 12, 2009
Collection sample: WOUND Collection date: Oct 12, 2009
Provider: Dr. General Hawk
Specimen: SWAB

CULTURE RESULTS: 1. MODERATE STAPHYLOCOCCUS AUREUS
Comment: CIPROFLOXACIN = S
ERYTHROMICIN = S
2. E. COLI
Comment: GENTAMICIN 500 synergy screen → RESISTANT
Confirmed sensitive to Penicillin

ANTIBIOTIC SUSCEPTIBILITY TESTS RESULTS:

	1. STAPHYLOCOCCUS AUREUS	2. ESCHERICHIA COLI
AMPICLN		S
PENICLN	S	

----MICROBIOLOGY----

Accession: VA 123
Received: Oct 12, 2009
Collection sample: WOUND
Collection date: Oct 12, 2009
Provider: Dr. General Hawk
Specimen: SWAB

CULTURE RESULTS: 1. MODERATE STAPHYLOCOCCUS AUREUS
Comment: CIPROFLOXACIN = S
ERYTHROMICIN = S
2. E. COLI
Comment: GENTAMICIN 500 synergy screen → RESISTANT
Confirmed sensitive to Penicillin

ANTIBIOTIC SUSCEPTIBILITY TESTS RESULTS:

	1. STAPHYLOCOCCUS AUREUS	2. ESCHERICHIA COLI
:		
AMPICLN		S
PENICLN	S	

----MICROBIOLOGY----

Accession: VA 123
Collection sample: WOUND
Provider: Dr. General Hawk
Specimen: SWAB

Received: Oct 12, 2009
Collection date: Oct 12, 2009

CULTURE RESULTS: 1. MODERATE STAPHYLOCOCCUS AUREUS
Comment: CIPROFLOXACIN = S
ERYTHROMICIN = S
2. E. COLI
Comment: GENTAMICIN 500 synergy screen → RESISTANT
Confirmed sensitive to Penicillin

ANTIBIOTIC SUSCEPTIBILITY TESTS RESULTS:

	1. STAPHYLOCOCCUS AUREUS	2. ESCHERICHIA COLI
AMPICLN		S
PENICLN	S	

----MICROBIOLOGY----

Accession: VA 123 Received: Oct 12, 2009
Collection sample: WOUND Collection date: Oct 12, 2009
Provider: Dr. General Hawk
Specimen: SWAB

CULTURE RESULTS: 1. MODERATE STAPHYLOCOCCUS AUREUS
Comment: CIPROFLOXACIN = S
ERYTHROMICIN = S
2. E. COLI
Comment: GENTAMICIN 500 synergy screen → RESISTANT
Confirmed sensitive to Penicillin

ANTIBIOTIC SUSCEPTIBILITY TESTS RESULTS:
1. STAPHYLOCOCCUS AUREUS
: 2. ESCHERICHIA COLI
AMPICLN S
PENICLN S

Get numbered list of organisms

Look for Comments

----MICROBIOLOGY----

Accession: VA 123 Received: Oct 12, 2009
Collection sample: WOUND Collection date: Oct 12, 2009
Provider: Dr. General Hawk
Specimen: SWAB

Pull out susceptibility formatted like DRUG = SUSCEPT.

CULTURE RESULTS: 1. MODERATE STAPHYLOCOCCUS AUREUS

Comment: CIPROFLOXACIN = S
ERYTHROMICIN = S

2. E. COLI

Comment: GENTAMICIN 500 synergy screen → RESISTANT
Confirmed sensitive to Penicillin

ANTIBIOTIC SUSCEPTIBILITY TESTS RESULTS:

	1. STAPHYLOCOCCUS AUREUS
:	2. ESCHERICHIA COLI
AMPICLN	S
PENICLN	S

Pull out susceptibility formatted in table

CULTURE RESULTS:

1. MODERATE STAPHYLOCOCCUS AUREUS

Comment: CIPROFLOXACIN = S

ERYTHROMICIN = S

2. E. COLI

Comment: GENTAMICIN 500 synergy screen → RESISTANT

Confirmed sensitive to Penicillin

ANTIBIOTIC SUSCEPTIBILITY TESTS RESULTS:

	1. STAPHYLOCOCCUS AUREUS	: 2. ESCHERICHIA COLI
AMPICLN		S
PENICLN	S	

CULTURE RESULTS:

1. <ORGANISM>

Comment: <DRUG> = <SUSCEPTIBILITY>

<DRUG> = <SUSCEPTIBILITY>

2. <ORGANISM>

Comment: <DRUG> 500 synergy screen → <SUSCEPTIBILITY>

Confirmed <SUSCEPTIBILITY> to <DRUG>

ANTIBIOTIC SUSCEPTIBILITY TESTS RESULTS:

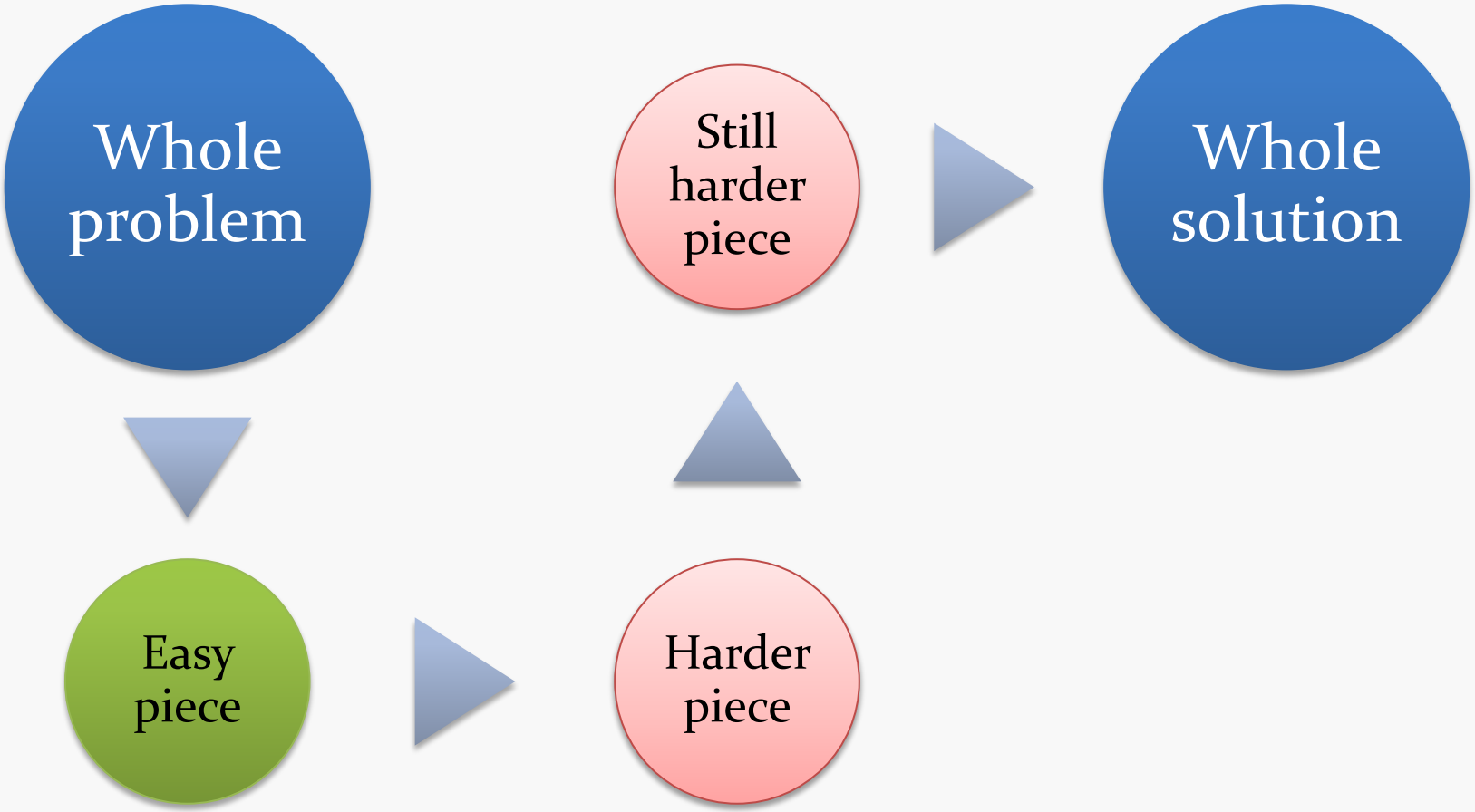
1. <ORGANISM>

: 2. <ORGANISM>

<DRUG> <SUSCEPTIBILITY>

<DRUG> <SUSCEPTIBILITY>

Divide and Construct



Divide and Construct

- Break the problem down into pieces
- Determine which pieces are easiest to solve
- Use what you know about the pieces to find solutions
- Use solution from easier pieces to help solve harder pieces

Divide and Construct

DISCHARGE SUMMARY

ALLERGIES: Tylenol, Advil, Chocolate.

MEDICATIONS ON ADMISSION: diazapam 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 apirin daily and an oral dose of diazapam 25 mg for a total of 10 days.

Divide and Construct

DISCHARGE SUMMARY

Ignore this section

ALLERGIES: Tylenol, Advil, Chocolate.

Has structure

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

No structure

HOSPITAL COURSE: The patient was admitted for a brain hemorrhage and was placed on Lasix. The patien 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 apirin daily and an oral dose of diazapam 25 mg for a total of 10 days.

Divide and Construct

DISCHARGE SUMMARY

Ignore this section

ALLERGIES: Tylenol, Advil, Chocolate.

List

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

Narrative

HOSPITAL COURSE: The patient was admitted for a brain hemorrhage and was placed on Lasix. The patien 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 apirin daily and an oral dose of diazapam 25 mg for a total of 10 days.

Divide and Construct

DISCHARGE SUMMARY

ALLERGIES: Tylenol, Advil, Chocolate.

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

What type of list?

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 apirin daily and an oral dose of diazapam 25 mg for a total of 10 days.

Divide and Construct

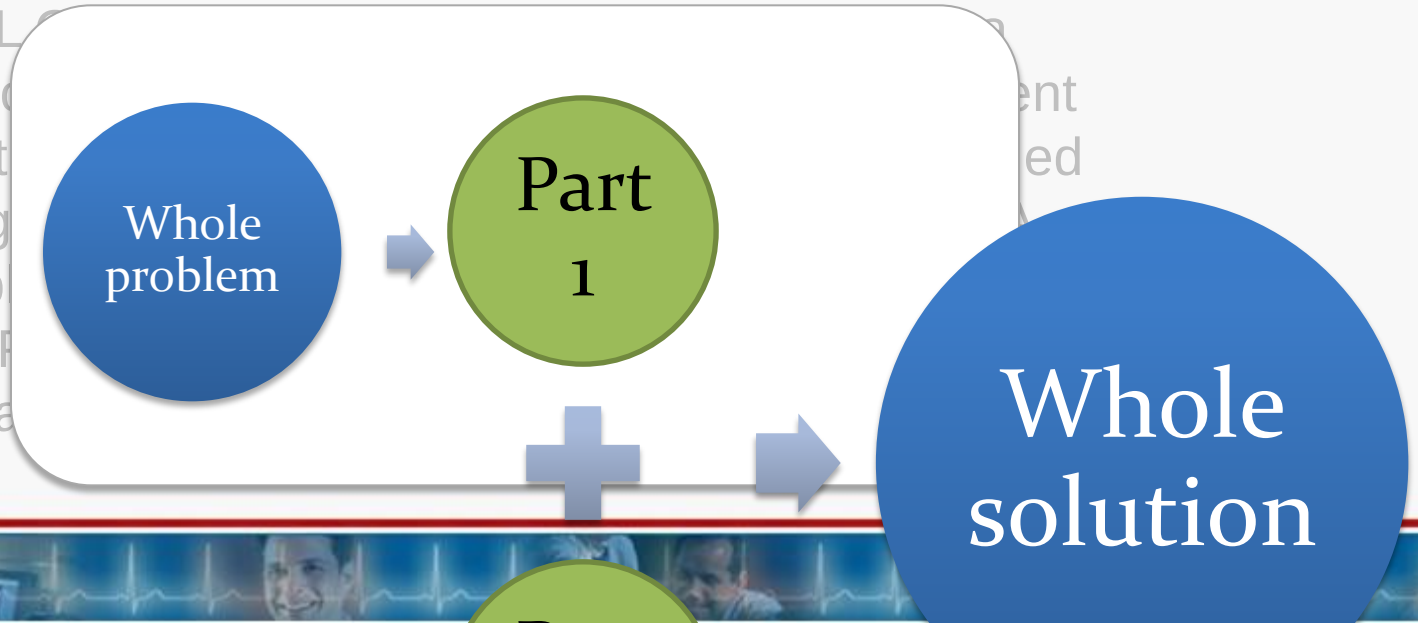
DISCHARGE SUMMARY

ALLERGIES: Tylenol, Advil, Chocolate.

List pieces

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

HOSPITAL
brain hem ent
became st ed
monitoring
and was p
penicillin.
dose of dia



Assumptions

- Some pieces must be easy
- The easy pieces have to be identified
- The easy pieces can aid in solving harder pieces

Conclusion

- NBS involved from the start and would love to re-engage NIST
- Opportunities include:
  - Help in creating bridging architecture to next generation system including migration from M origins to a more modern architecture with ability to provide support from a whole new generation of developers using modern Web technology

Opportunities

- Incredible opportunity to study and create laboratory for usability and safety
- Great opportunity to study healthcare security
- Excellent opportunity to evaluate and test standards

- Can evaluate and text logic used for business and clinical analytics and then test and release clinical software
- Superb opportunity to utilize VA's VINCI and other data as sandbox to evaluate big data using the 3.2 billion records available in the VINCI system and have access to not only data but also patient images

Acknowledgements

Dr. Robert Kolodner - VISTA

Dr. Scott DuVall – VISTA and VINCI