CIPHER-OHDSI Phenotype Library Integration

OHDSI Phenotype Development and Evaluation Workgroup
6/28/2024

- Ashley Galloway, Associate Director, Strategic Partnerships and Outreach
- Anne Ho, Director for Data Operations
- Jackie Honerlaw, Deputy Director
- Michael Murray, Lead Technical Architect
- Ed Zielinski, Database Manager

CIPHER
CENTRALIZED INTERACTIVE PHENOMICS RESOURCE

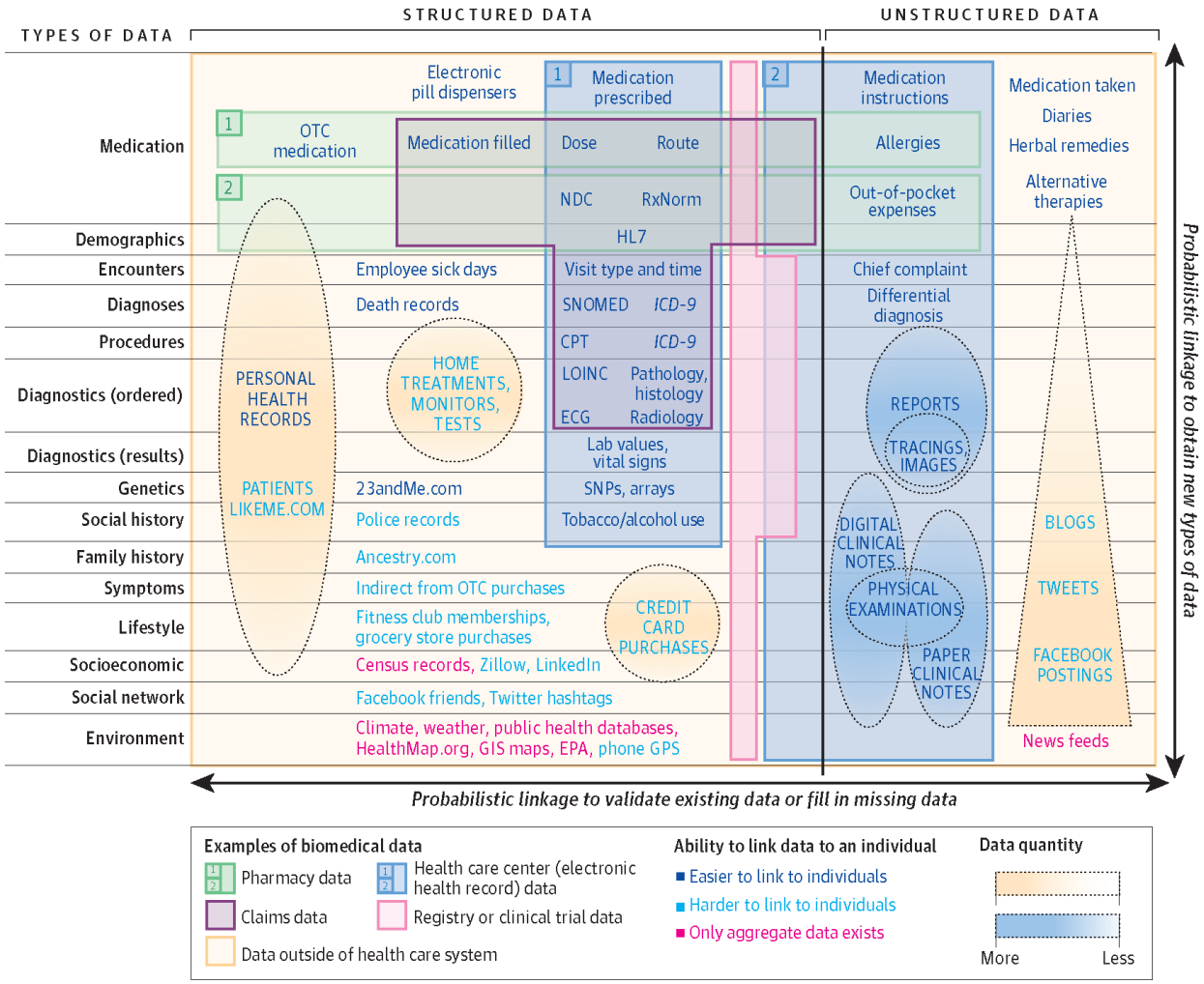
INNOVATION INTEROPERABILITY COLLABORATION

VA | U.S. Department of Veterans Affairs

Objectives

1. Overview of CIPHER phenotype library (5 min)
2. Review proposed pilot for integration of OHDSI Phenotype Library definitions in CIPHER (5 min)
3. Discuss proposed integration (5 min)

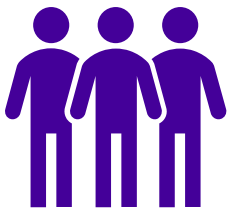
EHRs are a rich resource for clinical research and healthcare operations



Weber JAMA 2014

Phenotyping

- Code curation
- Machine learning
- Other approaches



Heart failure

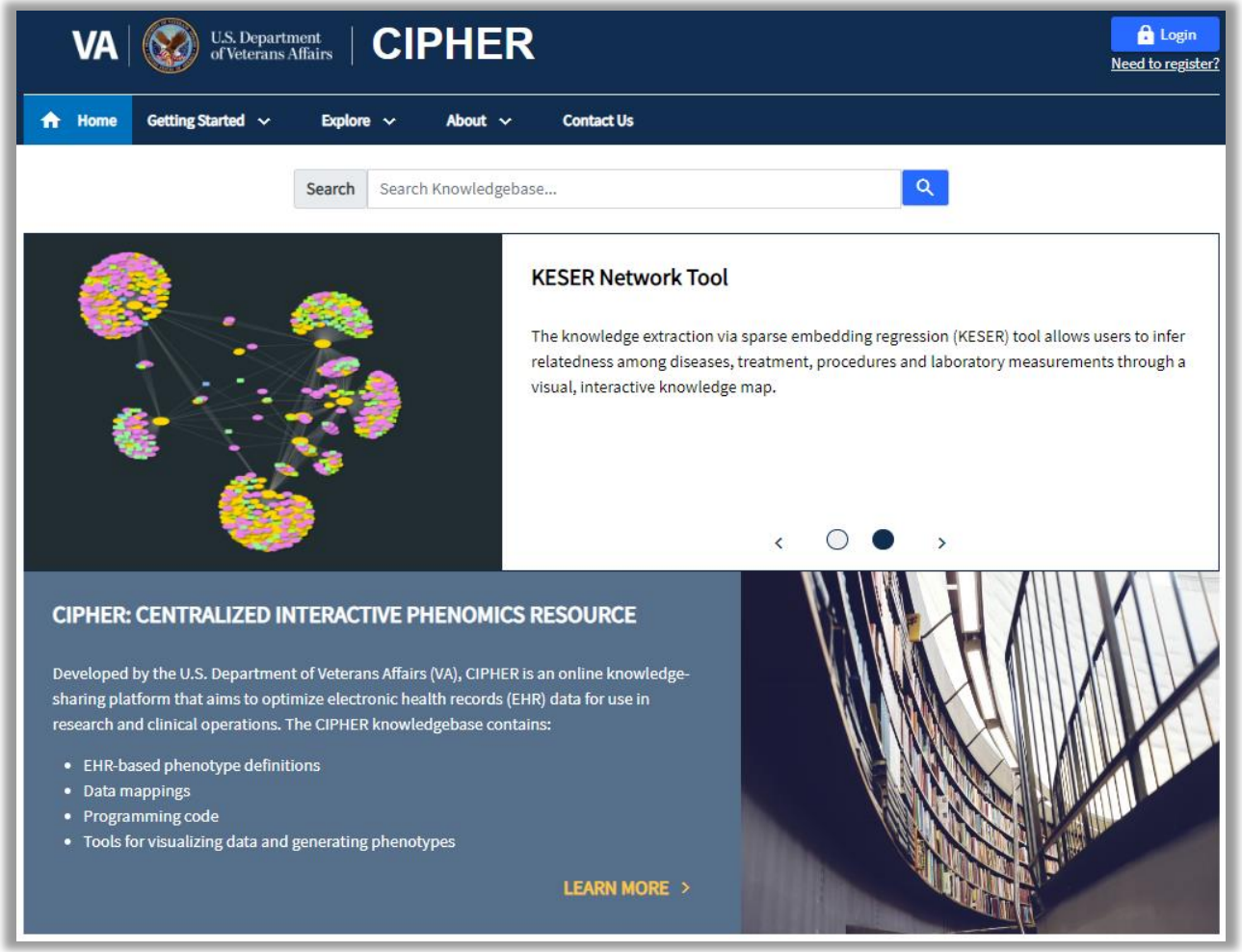
Diabetes

PTSD

Centralized Interactive Phenomics Resource (CIPHER) Overview

A publicly accessible platform centralizing phenotype definitions and phenotyping resources

Mission: Accelerate health data innovation by providing an integrated and interactive knowledge sharing platform

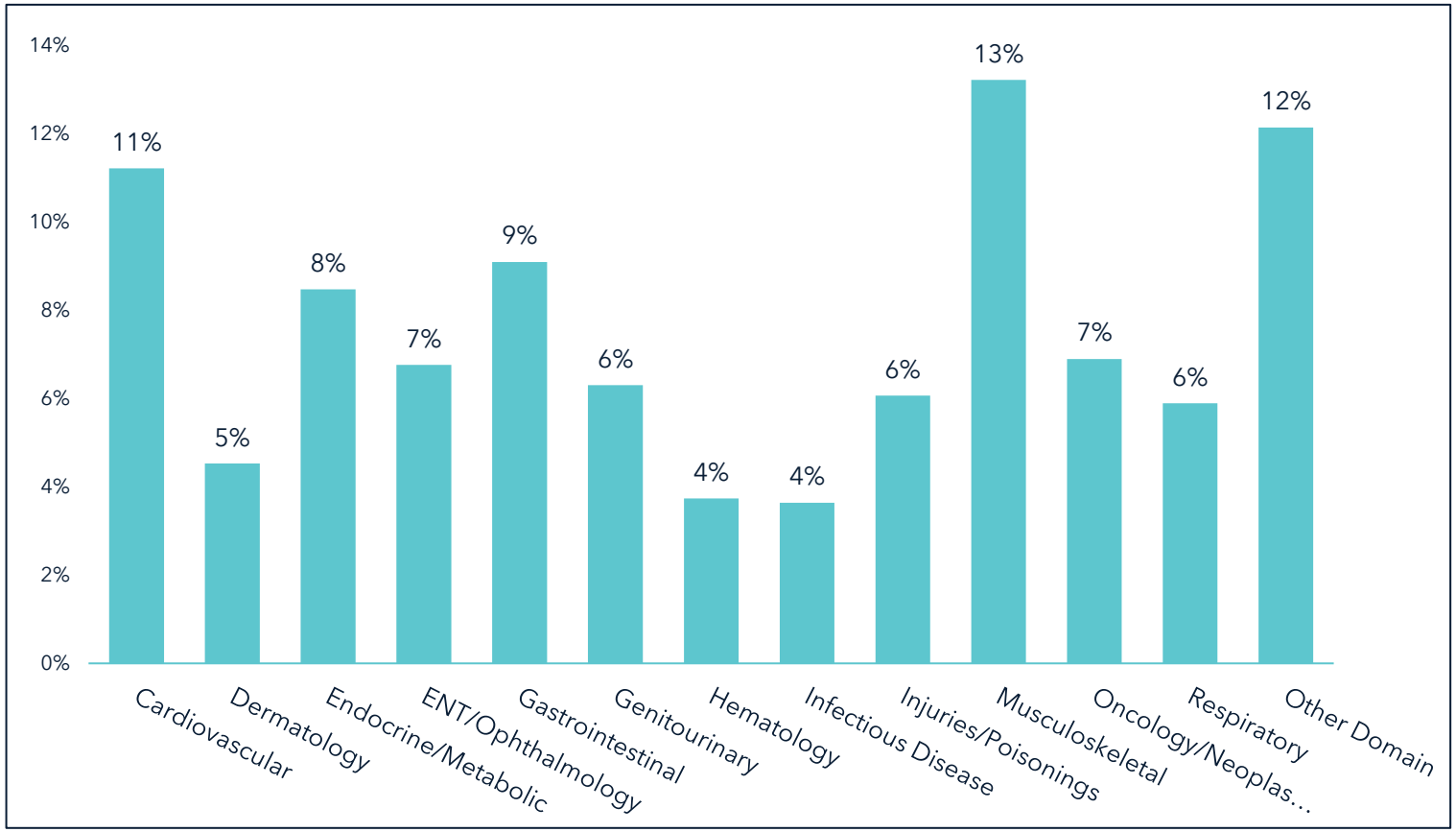


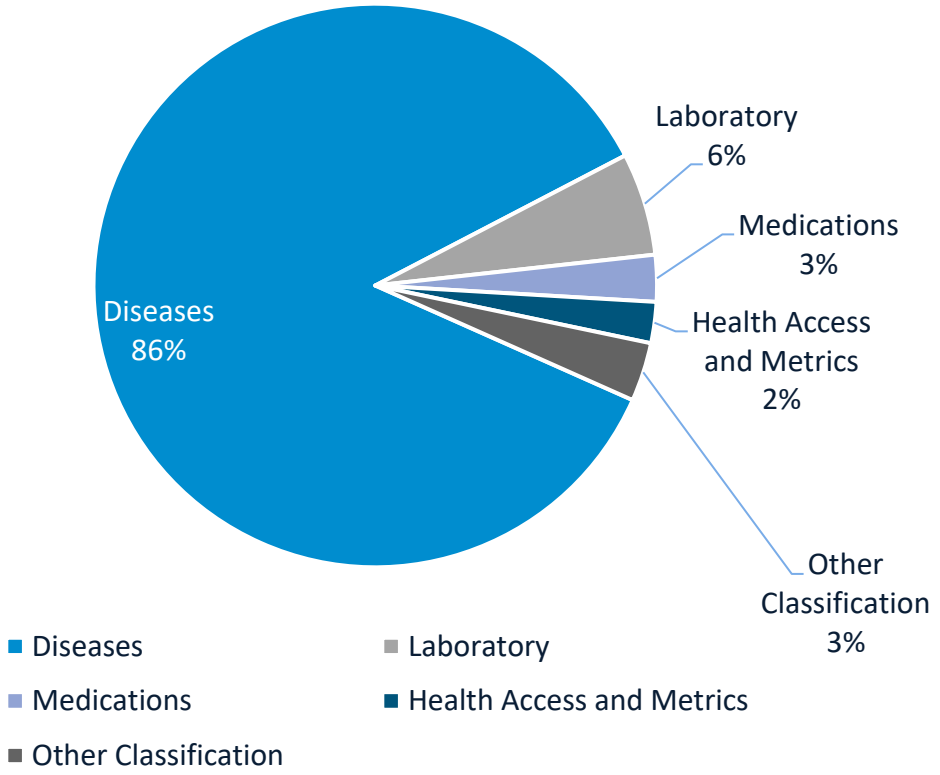
Funded by the US Department of Veterans Affairs, Office of Research and Development

Expanding CIPHER Knowledgebase Content

CIPHER

6,000 + Phenotypes





VA | U.S. Department of Veterans Affairs

CIPHER's Userbase Continues to Grow

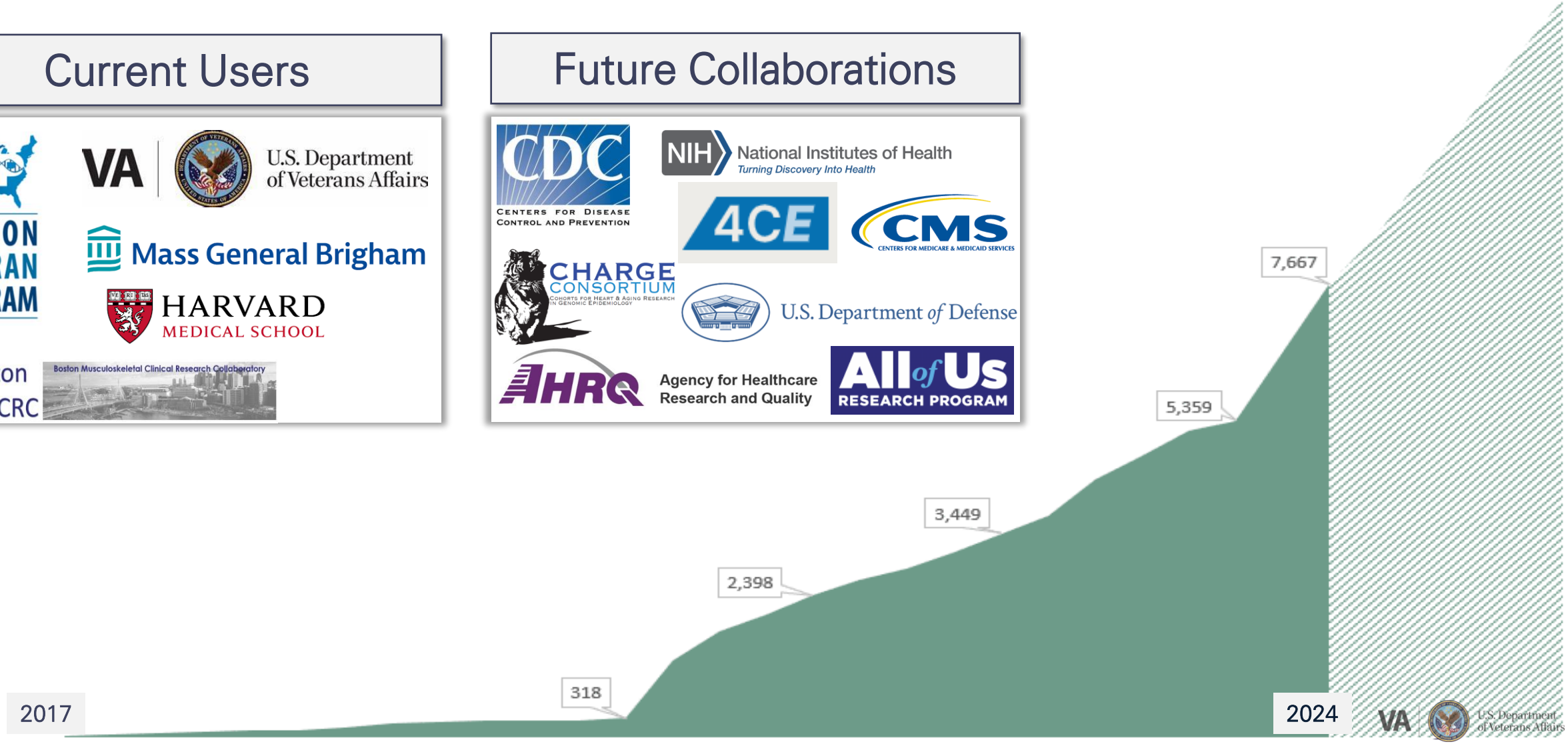
CIPHER

Current Users

- Million Veteran Program
- VA | U.S. Department of Veterans Affairs
- Mass General Brigham
- Harvard Medical School
- Boston MCRC — Boston Musculoskeletal Clinical Research Collaboratory

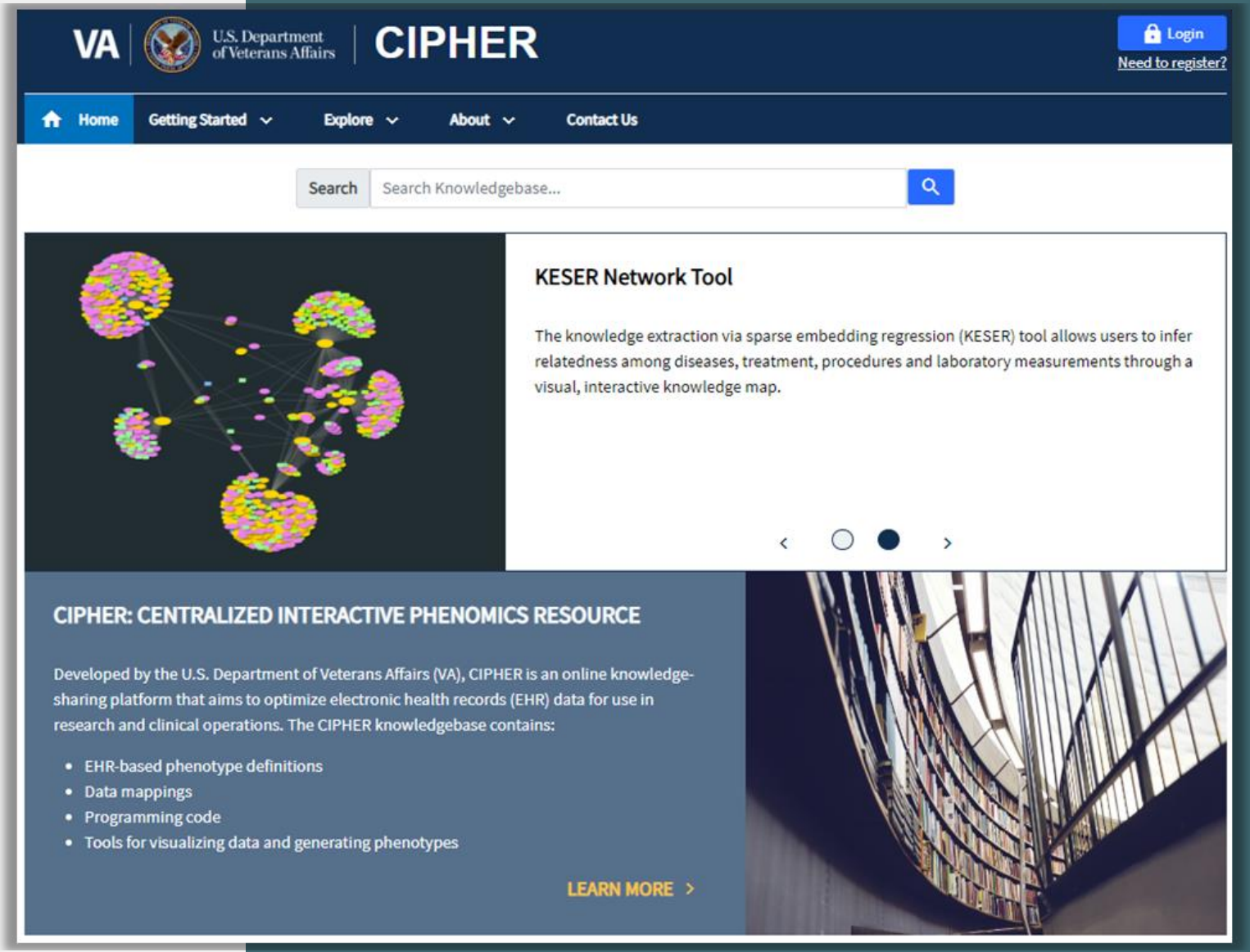
Future Collaborations

- CDC — Centers for Disease Control and Prevention
- NIH — National Institutes of Health, Turning Discovery Into Health
- 4CE
- CMS — Centers for Medicare & Medicaid Services
- CHARGE Consortium — Cohorts for Heart & Aging Research in Genomic Epidemiology
- U.S. Department of Defense
- AHRQ — Agency for Healthcare Research and Quality
- All of Us Research Program



CIPHER Website

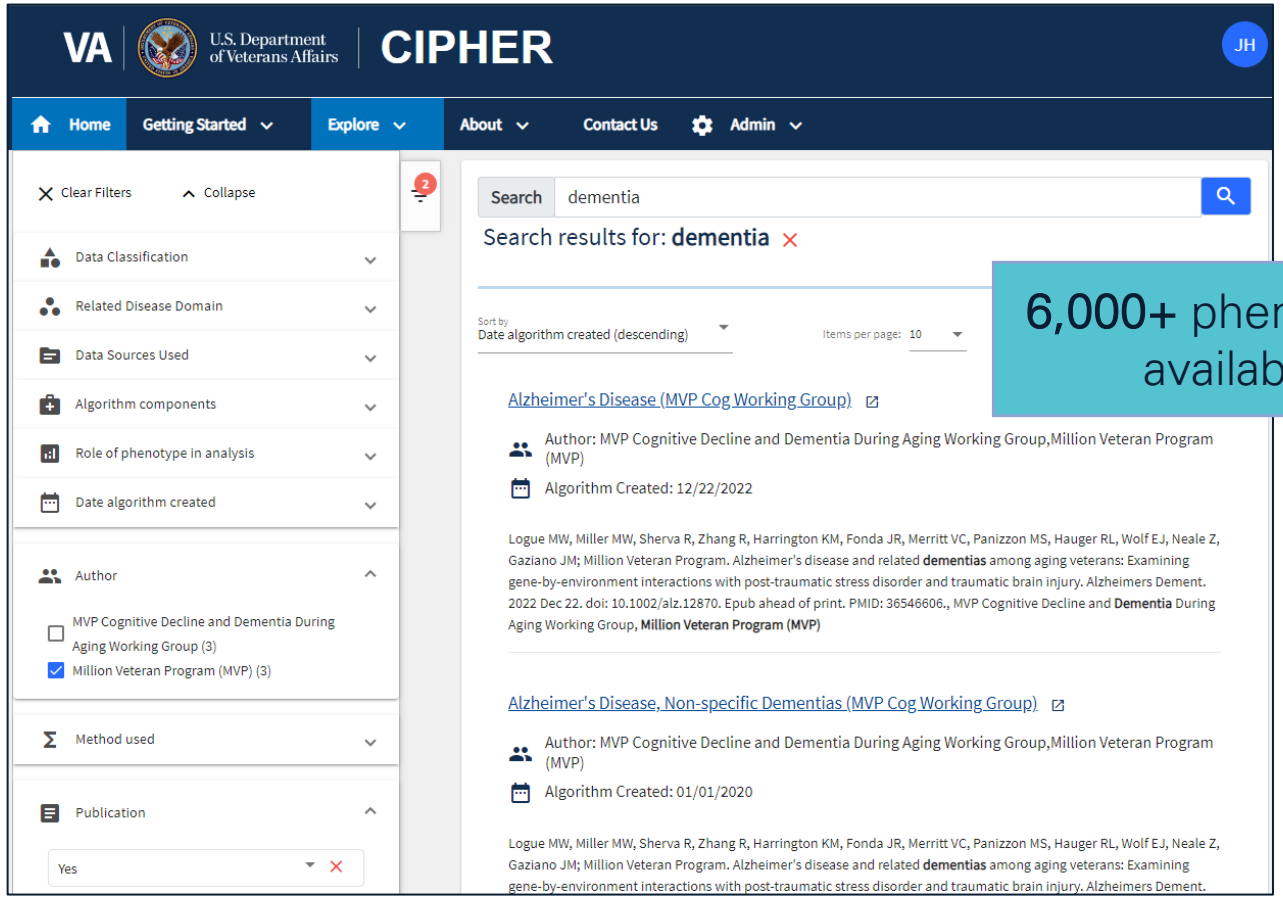
A. Phenotype Knowledgebase
B. Phenotype Collection Workflow
C. Data Visualization Tools



Searchable database of phenotype articles

A. Phenotype knowledgebase
B. Phenotype collection workflow
C. Data visualization tools

- ✓ CIPHER standard
- ✓ Unique phenotype article identifier
- ✓ Smart search
- ✓ Version control



6,000+ phenotypes available

 https://phenomics.va.ornl.gov/web/cipher/search

CIPHER Metadata Standard

Phenotype Name

Phenotype Description
Phenotype purpose and aims (E.g., to maximize sensitivity)

Publication / Acknowledgment

Algorithm Description
Logic applied to components to generate algorithm

Code Samples
Programming code or code repository link

Population
Population used to develop algorithm

Authors / Contacts

Data Sources
Data source used to generate the phenotype

Data Classifications
Disease, demographics, laboratory test, etc.

Role in Analysis
Primary outcome/exposure, inclusion requirement, etc.

Related Disease Domain
Cardiovascular, dermatology, genitourinary, etc.

Unique Identifier

Keywords
MeSH Terms

Methods Used
Rules based, machine learning, etc.

Context of Development
Research, healthcare operations, etc.

Algorithm Components
Medical Vocabularies
(ICD, LOINC, etc.)

Algorithm Creation Date

Lab Units / Values

Algorithm Validation / Performance Metrics
Sensitivity, specificity, etc.

Attachments
Tables, figures, slides, etc.

Legend
- Free Text
- Semi-controlled Authority
- Standard Vocabulary
- Empirical Value
- Other

Honerlaw JAMIA 2023

Example Phenotype Content: Alzheimer's Disease and Related Dementias (MVP Cognitive WG)

1

Basic Information and Contact

Abbreviations and Keywords

ADRD | Cognitive Decline

Publication

Alzheimer's disease
disorder and trauma

Data Classification(

Diseases

Related Disease

Mental/ Behavioral H

Author(s)

MVP Cognitive D

Contact

MVP Cognitive Decli

Acknowledgement

Mark W. Logue, Richa
During Aging Workin

2

Algorithm Ov

Source of Phenotyp

CDW (Corporate

Context of Develop

Research

Role of Phenotype i

Primary Outcome/Ex

Phenotype Descript

The Million Veteran
dementia phenotype
developed for use in

Population

Million Veteran Prog

Date Algorithm Cre

06/01/2023

3

Algorithm Components

Method Used

Rules-Based (i.e., only structured data were used)

Algorithm Description

To quality as a case requires the presence of two or m

Algorithm Components

ICD-9 Diagnostic Codes (2

ICD-10 Diagnostic Codes (1

(1) Programming Code SQL

```
/*******************    MVP Phenotype Code
Project Name: Alzheimer's Phenotyping in
Purpose: To create a single program that
*****************************************

/*****Step 1. Selecting persons with any
 DROP table if exists ##Dementia_InpatOut
 SELECT Distinct
        [VINCI_ID]
       ,CAST([DischargeDate] AS DATE) AS
       ,[ICDCode]
       ,[ICDDescription]
              /*AD 1/30/2023: Creating
```

4

Validation

Validated

Yes

Description of Validation

We have performed a chart review of n=39 "difficult cases". That is, MVP participants with either a) 1 ICD code for either MCI, AD, or dementia or b) subjects with low to moderate AD case probability according to the Multimodal Automated Phenotype (MAP) algorithm (Liao et al. Am Med Inform Assoc, 2019). These were reviewed in tandem by Drs. Hauger and Meritt of the AD/MCI working group and classified as "Likely not", "Possible", or "Likely" cases of MCI, AD, and dementia. Chart review results were compared to the classifications according to the MAP algorithm and according to our ICD-code based definitions of MCI, AD, ADRD, and dementia (Table 2). The performance of the ICD code set in shown in "Algorithm Performance Measures". The authors also evaluated the suitability of our ICD code based phenotypes in genetic studies by testing their association with the APOE E4 isoform (the strongest AD genetic risk factor) in European-descent MVP subjects (Table 3).

Performance Metric	Value
Sensitivity	1.00
Specificity	0.35
Negative Predictive Value (NPV)	1.00
Positive Predictive Value (PPV)	0.67
Area Under the ROC Curve (AUC)	N/A
Kappa	N/A

Additional Information

Attachments

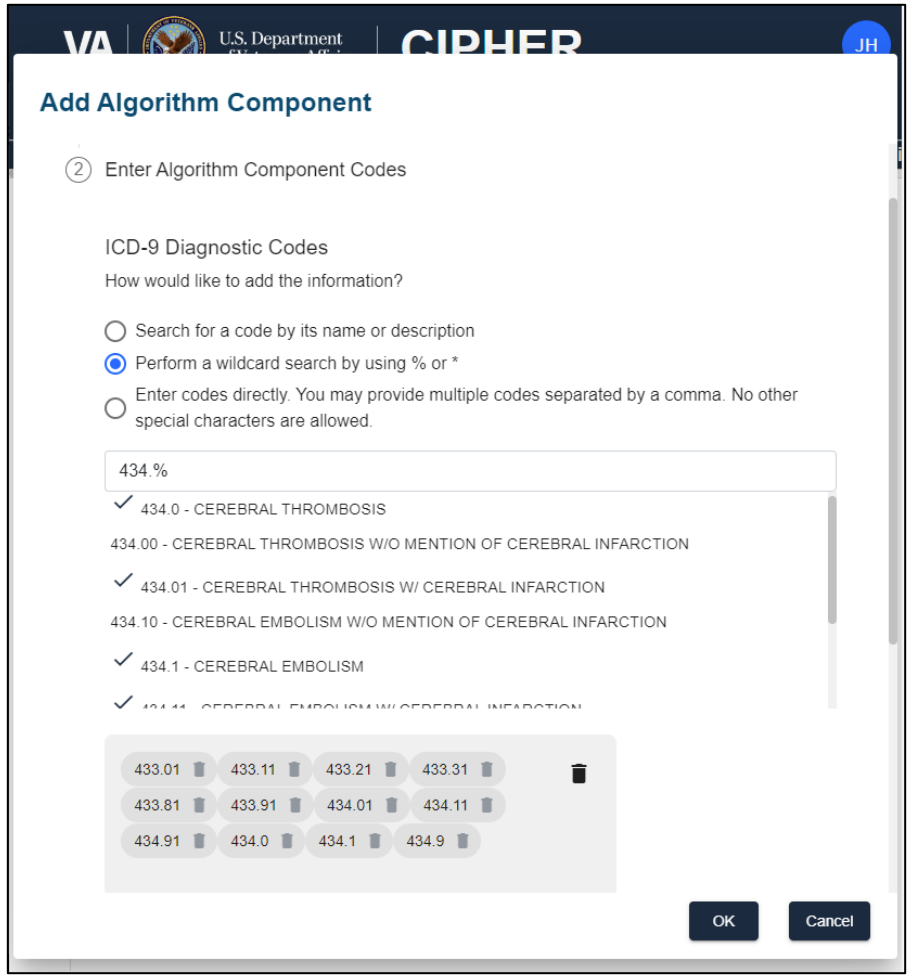
Type	Description
Additional Information	Algorithm components/ performance metrics

Alzheimer's Disease and Related Dementias

Standardized collection of phenotype metadata

A. Phenotype knowledgebase
B. Phenotype collection workflow
C. Data visualization tools

- ✓ Validation against standard vocabularies
- ✓ CIPHER review
- ✓ Populate into knowledgebase

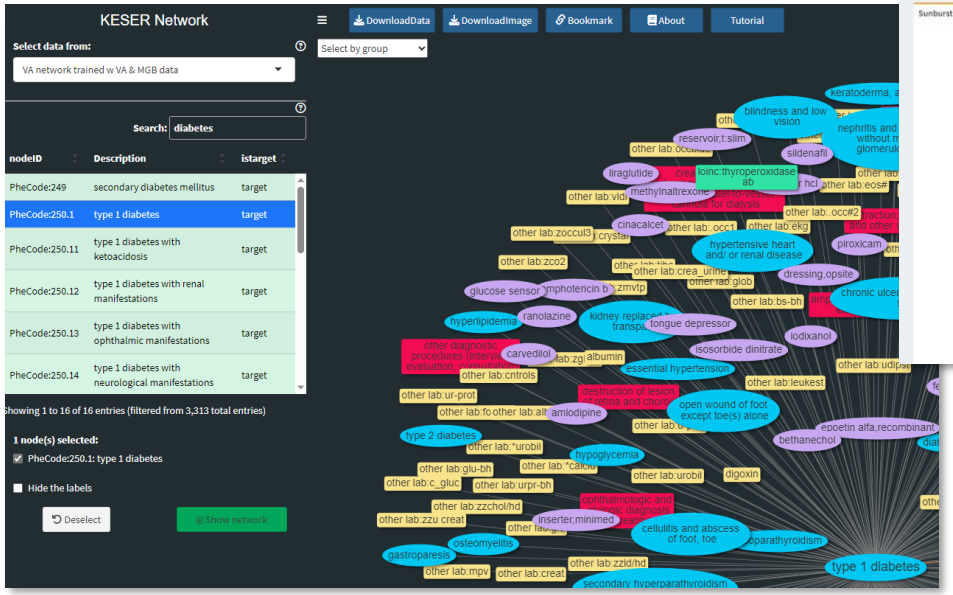


https://phenomics.va.ornl.gov/web/cipher/contribute

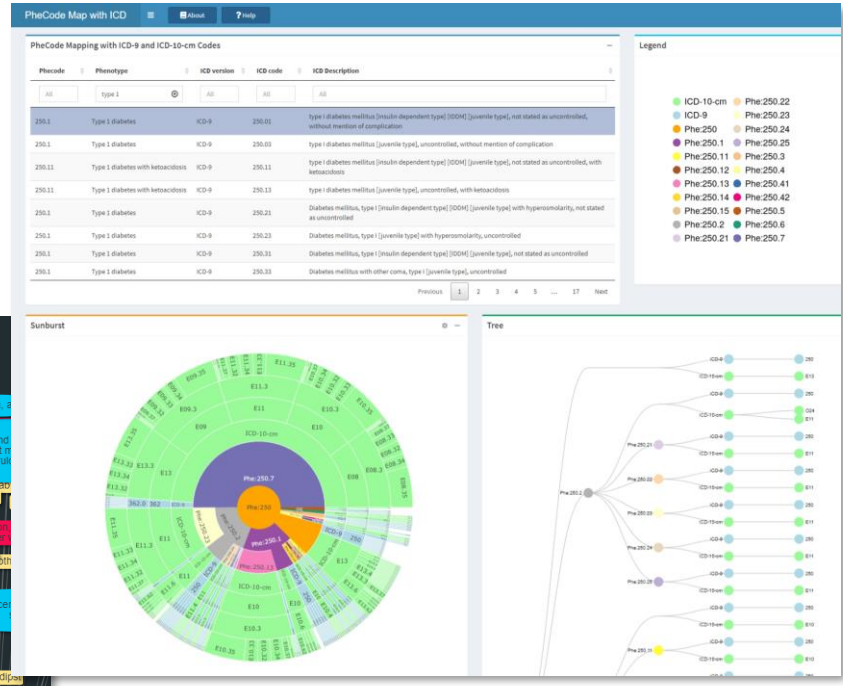
Tools connected to phenotype definition knowledgebase

A. Phenotype knowledgebase
B. Phenotype collection workflow
C. Data visualization tools

- ✓ Provides interactive approach to exploring metadata
- ✓ Aids in phenotype development
- ✓ Linkage to knowledgebase



KESER creates a knowledge map to allow users to visualize relatedness among diseases, treatment, procedures, and laboratory measurements.



Phecode to ICD Map assists users in developing clinically meaningful disease phenotypes by enabling the user to visualize the relationships between various ICD codes and Phecodes.

https://phenomics.va.ornl.gov/web/cipher/vistools

Learn More About CIPHER

CIPHER

CIPHER Metadata Standard

Journal of the American Medical Informatics Association, 30(5), 2023, 958–964
https://doi.org/10.1093/jamia/ocad030
Advance Access Publication Date: 7 March 2023
Brief Communication

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.

Brief Communication

Framework of the Centralized Interactive Phenomics Resource (CIPHER) standard for electronic health data-based phenomics knowledgebase

Jacqueline Honerlaw[1], Yuk-Lam Ho[1], Francesca Fontin[1], Jeffrey Gosian[1], Monika Maripuri[1], Michael Murray[1], Rahul Sangar[1], Ashley Galloway[1], Andrew J. Zimolzak[2,3], Stacey B. Whitbourne[1,4,5], Juan P. Casas[1,4,5], Rachel B. Ramoni[6], David R. Gagnon[1,7], Tianxi Cai[1,8,9], Katherine P. Liao[1,4,9,10], J. Michael Gaziano[1,4,5], Sumitra Muralidhar[6], and Kelly Cho[1,4,5]

[1]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, Massachusetts, USA, [2]Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey VA Medical Center, Houston, Texas, USA, [3]Department of Medicine, Baylor College of Medicine, Houston, Texas, USA, [4]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA, [5]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA, [6]Office of Research and Development, Veterans Health Administration, Washington, District of Columbia, USA, [7]Department of Biostatistics, School of Public Health, Boston University, Boston, Massachusetts, USA, [8]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA, [9]Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA, and [10]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Boston, Massachusetts, USA

Corresponding Author: Jacqueline Honerlaw, Massachusetts Veterans Epidemiology Research and Information Center

Honerlaw J, Ho YL, Fontin F, et al. Framework of the Centralized Interactive Phenomics Resource (CIPHER) standard for electronic health data-based phenomics knowledgebase. J Am Med Inform Assoc. 2023; ocad030.

CIPHER Design and Resources

Journal of the American Medical Informatics Association, 2024, 1–9
https://doi.org/10.1093/jamia/ocae042
Research and Applications

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY. OXFORD

Research and Applications

Centralized Interactive Phenomics Resource: an integrated online phenomics knowledgebase for health data users

Jacqueline Honerlaw, RN, MPH*[,1,2], Yuk-Lam Ho, MPH[1,2], Francesca Fontin, MPH[1,2], Michael Murray, MS[1,2], Ashley Galloway, MPH[1,2], David Heise, MS[3], Keith Connatser, BS[3], Laura Davies, PMP[3], Jeffrey Gosian, BS[1,2], Monika Maripuri, MBBS, MPH[1,2], John Russo, MS[1,2,4], Rahul Sangar, MPH[1,2], Vidisha Tanukonda, MD[1,5], Edward Zielinski, ALM[1,2], Maureen Dubreuil, MD, MSc[2,6], Andrew J. Zimolzak, MD, MMSc[7,8], Vidul A. Panickan, MS[2,9], Su-Chun Cheng, ScD[2,9], Stacey B. Whitbourne, PhD[2,10,11,12], David R. Gagnon, MD, PhD[2,13], Tianxi Cai, ScD[2,9,14], Katherine P. Liao, MD, MPH[2,12,15,16], Rachel B. Ramoni, DMD, ScD[17], J. Michael Gaziano, MD, MPH[2,10,11,12], Sumitra Muralidhar, PhD[17], Kelly Cho, PhD, MPH[1,2,10,11,12]

[1]Centralized Interactive Phenomics Resource (CIPHER), Office of Research and Development, Veterans Health Administration, Washington, DC 20002, United States, [2]VA Boston Healthcare System, Boston, MA 02111, United States, [3]Oak Ridge National Laboratory (ORNL), Oak Ridge, TN 37830, United States, [4]Department of Computer Science, Landmark College, Putney, VT 05346, United States, [5]VA Atlanta Healthcare System, Decatur, GA 30033, United States, [6]Section of Rheumatology, Boston University Chobanian and Avedisian School of Medicine, Boston, MA 02118, United States, [7]Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey VA Medical Center, Houston, TX 77030, United States, [8]Department of Medicine, Baylor College of Medicine, Houston, TX 77030, United States, [9]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, United States, [10]Million Veteran Program (MVP) Coordinating Center, VA Boston, Boston, MA 02111, United States, [11]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, United States, [12]Department of Medicine, Harvard Medical School, Boston, MA 02115, United States, [13]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, United States, [14]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA 02115, United States, [15]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Boston, MA 02115, United States, [16]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, United States, [17]Office of Research and Development, Veterans Health Administration, Washington, DC 20002, United States

*Corresponding author: Jacqueline Honerlaw, RN, MPH, VA Boston Healthcare System, 2 Avenue De Lafayette, Boston, MA 02111, United States (Jacqueline.Honerlaw@va.gov)

J. Honerlaw and Y.-L. Ho authors contributed equally.

Abstract

Honerlaw, J, Ho YL, Fontin F, et al. Centralized Interactive Phenomics Resource: an integrated online phenomics knowledgebase for health data users. J Am Med Inform Assoc. 2024 Mar 13:ocae042. doi:10.1093/jamia/ocae042. Epub ahead of print. PMID: 38481028.

VA U.S. Department of Veterans Affairs

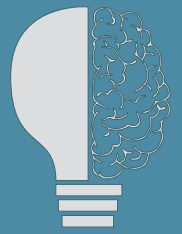
Future directions

CIPHER

Enhance Site Features

- Comparison of phenotype definitions and versions
- Phenotype discovery

Expand Knowledgebase



- Expand phenotypes
- Capture information on phenotype reuse
- Data dictionaries

Integrate Additional Tools

- Interactive map of phenotype prevalence
- Tools for phenotype development

Expand Partners and Contributors

Including:

- Health systems: QI/ CDS
- Cohort studies
- Federal partners
- Research funders

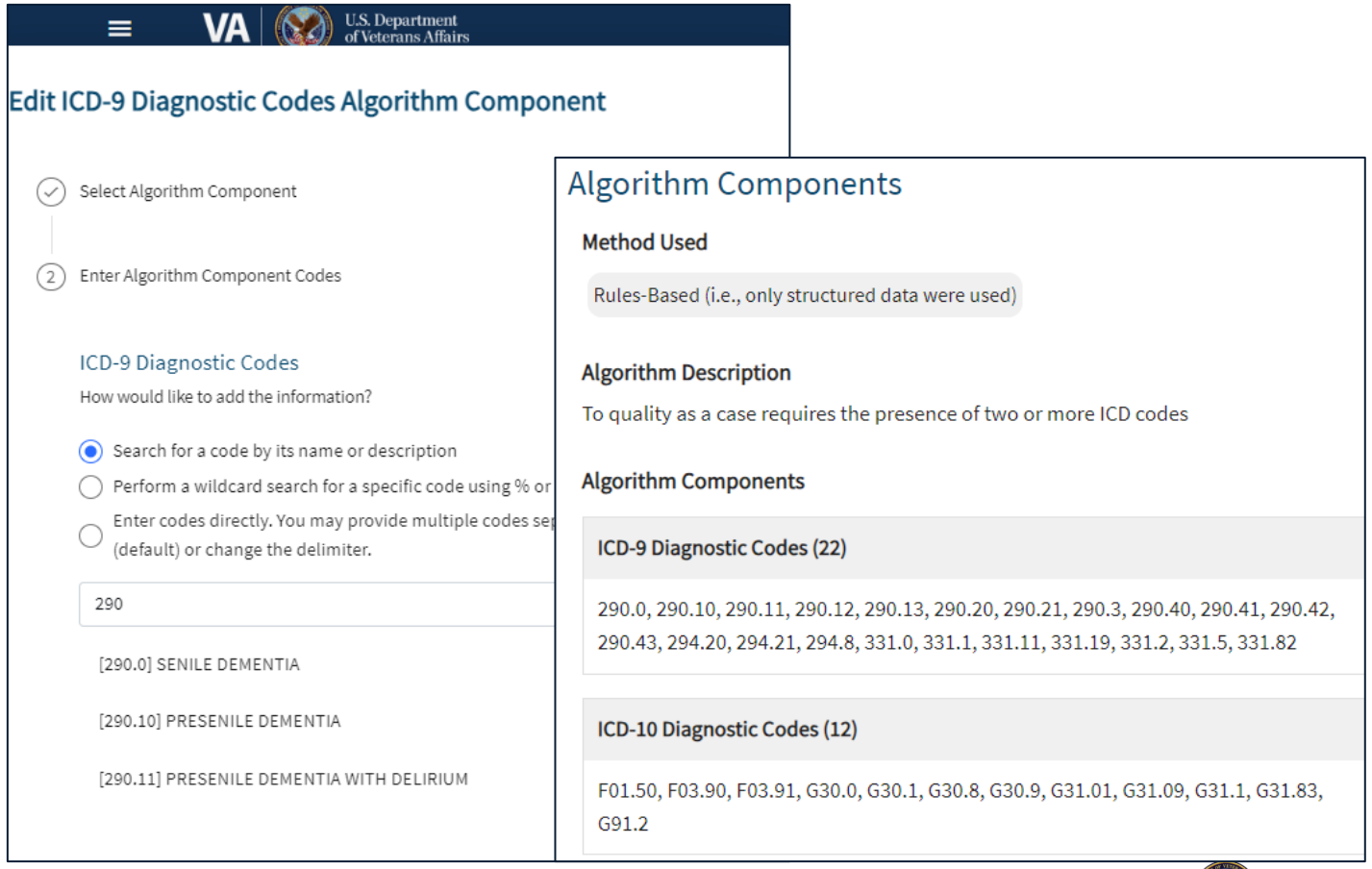
QI: Quality improvement; CDS: Clinical Decision Support

OMOP on CIPHER

Goal: Ensure OMOP users can search and share phenotype definitions in CIPHER

Planned approach

- Integrate OMOP concepts into CIPHER
- Pilot collection of OMOP based phenotypes with All of Us (AoU)
- Connect to phenotypes developed by OHDSI experts



OHDSI-CIPHER Library Integration

CIPHER

Goals:

- Promote OMOP use and OHDSI Phenotype Library to CIPHER community
- Facilitate search and reuse of OHDSI phenotypes

Proposed approach

1. Identify 5 phenotypes to pilot
2. Collect phenotype metadata required per the CIPHER standard
3. Enter phenotypes in knowledgebase, reference OHDSI Phenotype Library for select fields (programming code, etc)

Pilot phenotypes

- ✓ Shared in manuscripts/ conference proceedings
- ✓ Additional metadata available
- ✓ Author collaboration

Cohort Id	Cohort Name
33	Dementia
255	Alzheimer's disease (based on Imfeld, 2013)
863	Cognitive impairment, incident
249	Ischemic (Non-hemorrhagic) Stroke In Inpatient
892	Stroke (ischemic or hemorrhagic) events
76	Transient ischemic attack events

Q&A and Next Steps

CIPHER

- Discussion of proposal
- Follow-up
  - Demo of CIPHER library
  - Questions for OHDSI

1. OHDSI Phenotype Library
   a. How are phenotypes entered and what is the review process?
   b. How often are phenotypes updated?
2. OMOP Concepts
   a. How useful are OMOP based phenotypes to researchers not using OMOP?
   b. How are non-standard concepts maintained over time? Are there concepts that expire?

CIPHER Team

- **Sumitra Muralidhar** – VACO Lead
- **Kelly Cho** – Director
- **Jackie Honerlaw** – Deputy Director
- **Anne Ho** – Director for Data Operations
- **Francesca Fontin** – Project Manager, VA Partnerships and User Experience
- **Ashley Galloway** – Associate Director, Strategic Partnerships and Outreach
- **Jeff Gosian** – Systems Support Librarian
- **Monika Maripuri** – Project Manager, Clinical Phenotype Validation
- **Michael Murray** – Lead Technical Architect
- **Rahul Sangar** – Data Services Specialist
- **Tiffany Sim** – Project Manager, CIPHER Online
- **Joanne Sordillo** – SME for Environmental Exposure Data Domain
- **Vidisha Tanukonda** – Project Manager, Clinical Adjudication
- **Edward Zielinski** – Database Manager
- **CIPHER Online ORNL Team** – David Heise, Laura Davies, Keith Connatser, Adrian Degraffenreidt
- **Environment and Infrastructure Partners** – VINCI (Scott DuVall); MVP (Mike Gaziano); ORNL-DOE (David Heise)
- **CIPHER Partners and Contributors**

https://phenomics.va.ornl.gov/

Contact Us
CIPHER@va.gov

Relevant Links

- CIPHER Online
- CIPHER ORD Program Page
- VIReC Database and Methods Cyberseminar

References

- Honerlaw J, Ho YL, Fontin F, et al. Framework of the Centralized Interactive Phenomics Resource (CIPHER) standard for electronic health data-based phenomics knowledgebase. J Am Med Inform Assoc. 2023;30(5):958-964. doi:10.1093/jamia/ocad030 (link)
- Honerlaw J, Ho YL, Fontin F, et al. Centralized Interactive Phenomics Resource: an integrated online phenomics knowledgebase for health data users. J Am Med Inform Assoc. Published online March 13, 2024. doi:10.1093/jamia/ocae042 (link)

Please reach out to CIPHER to schedule a demonstration, refer us to health data users who are not part of the CIPHER community, or to receive summary slides to help us spread the word at conferences or other meetings.