



# NLP Working Group

**Chunnan Hsu**

Biomedical Informatics

UC San Diego

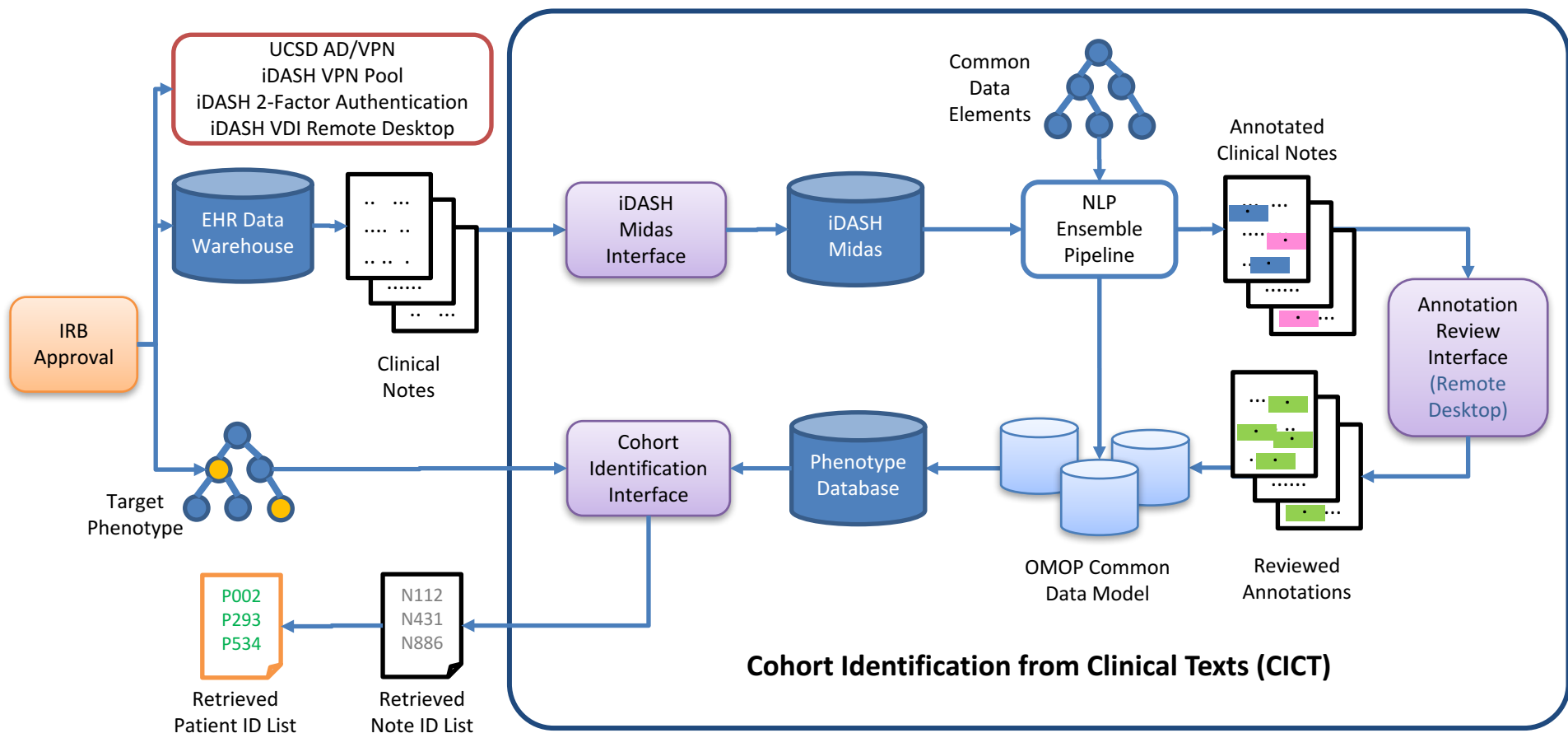
October 13, 2016

- Chunnan Hsu
- Ramana Seerapu
- Scott Duvall
- Olga Patterson
- Hua Xu
- Michael Matheny
- Glenn Gobbel
- Tsung-Ting Kuo

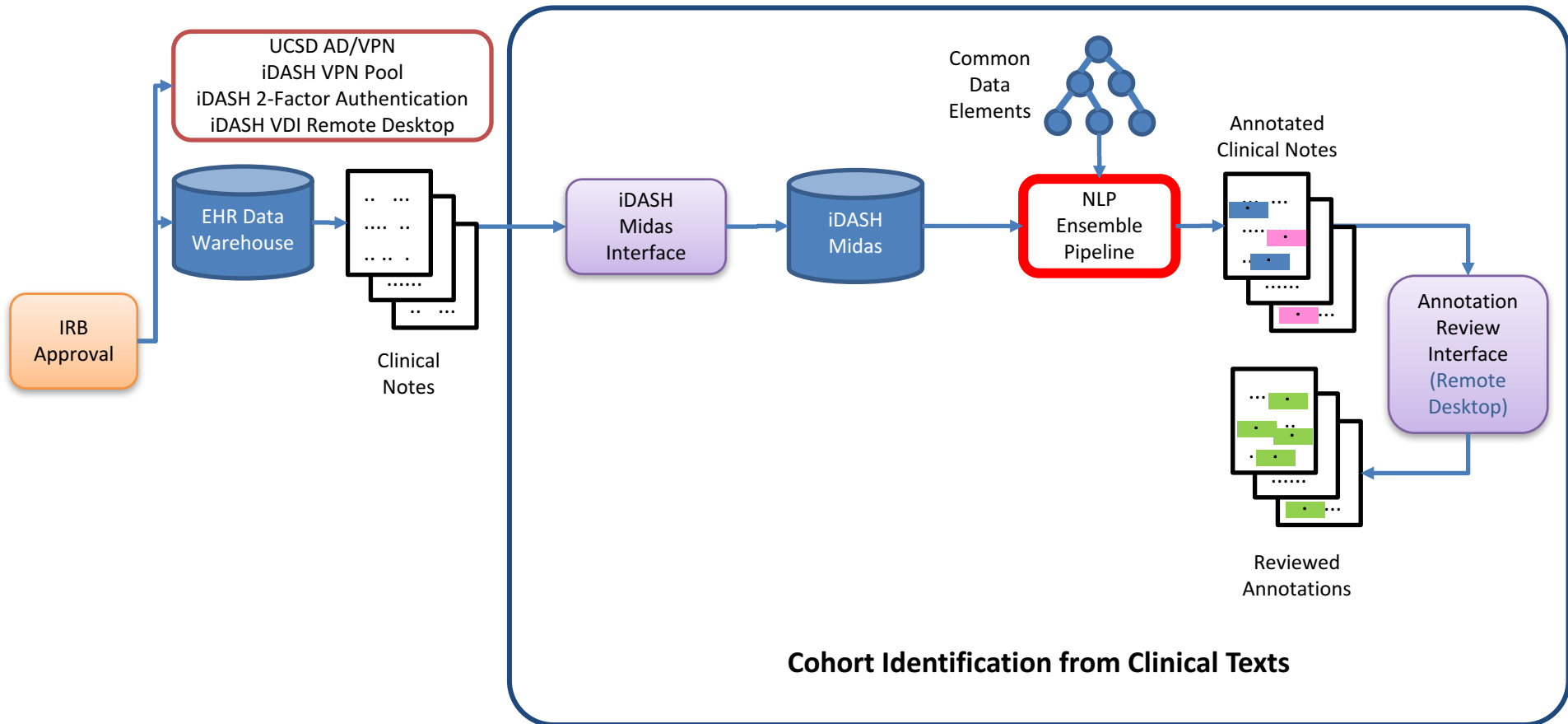
- The NLP working group is tasked to accurately extract phenotypes for three clinical conditions: Kawasaki Disease (KD), Weight Management / Obesity (WM/O), and Congestive Heart Failure (CHF), from tens of millions of clinical notes shared by participating institutes in pSCANNER, and seamlessly integrate with the shared structured data.

- Refine the problem definitions of the extraction of common data elements to guide the finalization of a common evaluation guideline, mapping to a common output data model, and design, reuse, and sharing of NLP tools
- Finalize output format definitions with OHDSI
- Create secure, privacy-preserving, cross-institution clinical NLP infrastructure where tens of millions of clinical notes can be processed and the quality of processing can be assessed semi-automatically.
- With the infrastructure, create large data warehouse of NLP extracted data from clinical notes to support phenotyping of the three pSCANNER use case conditions.

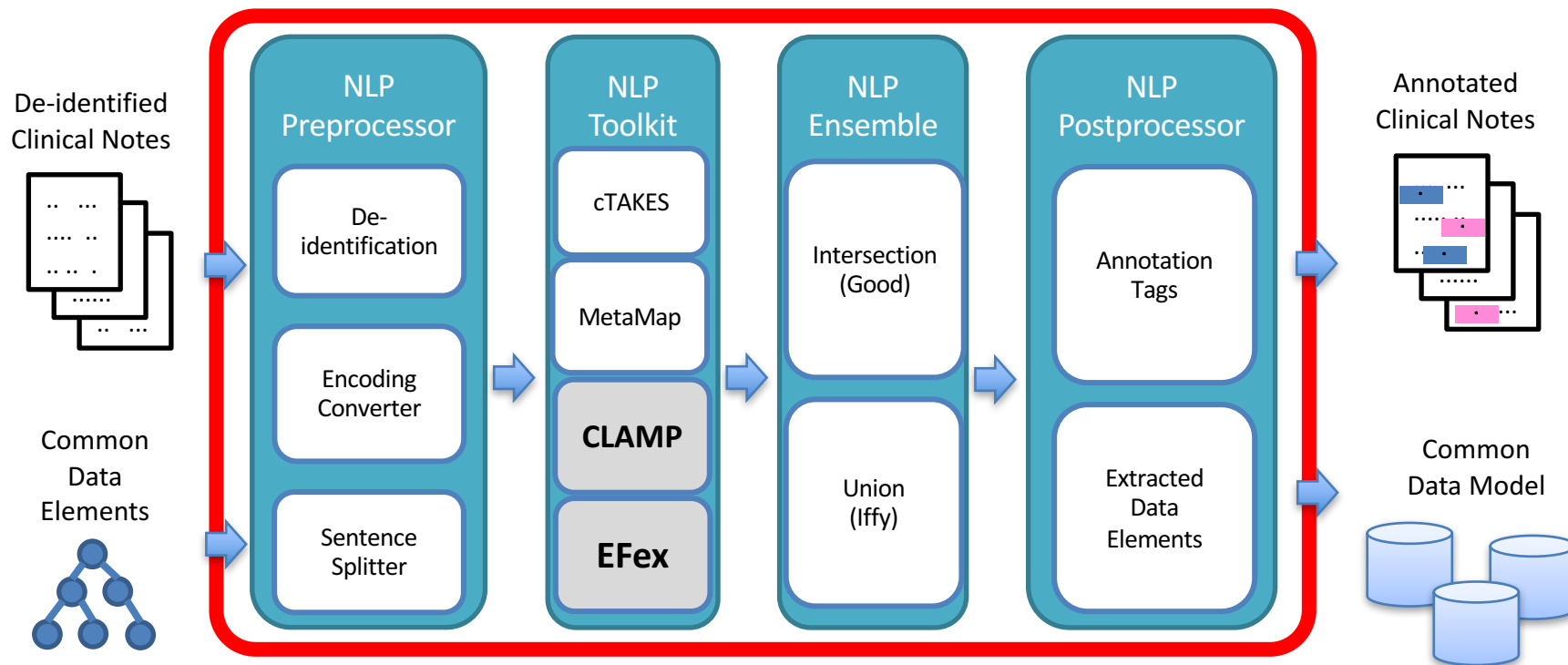
# Workflow



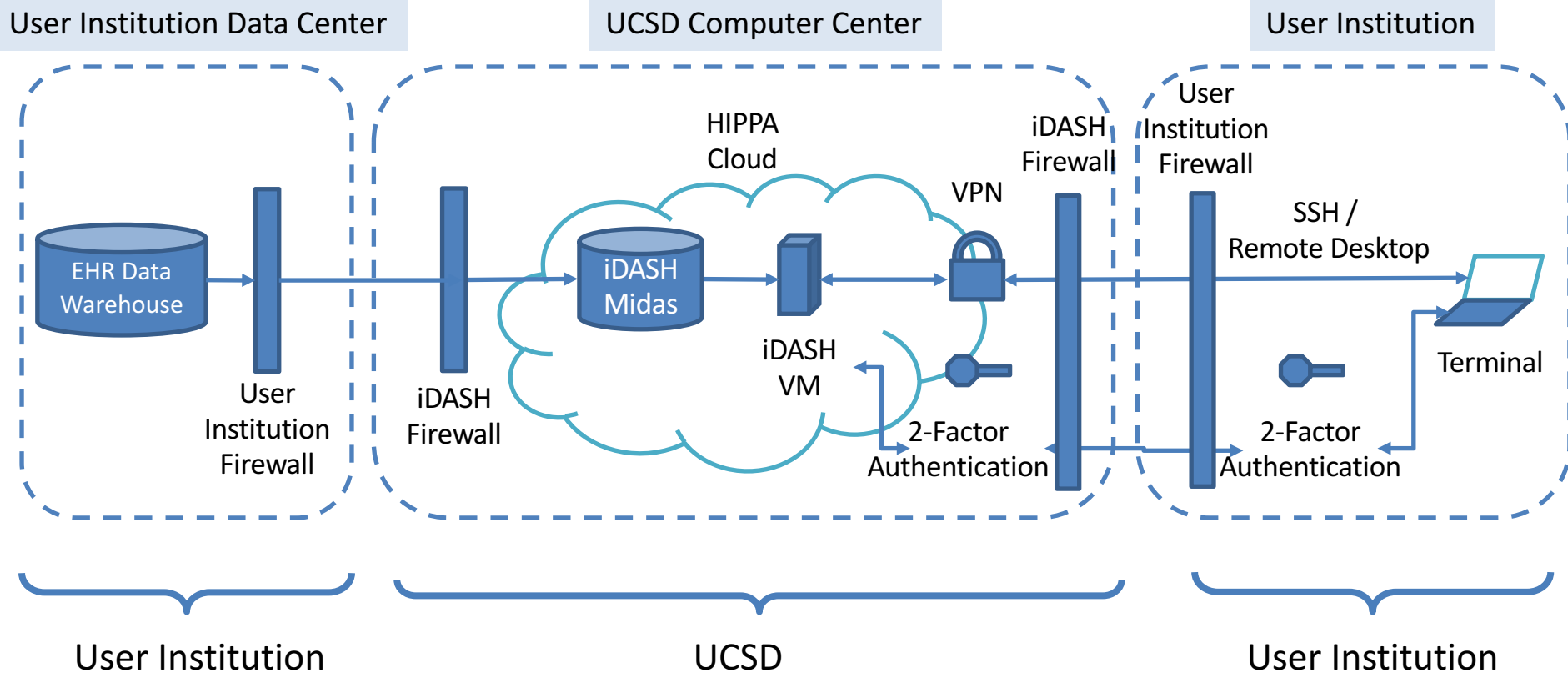
# Current Progress



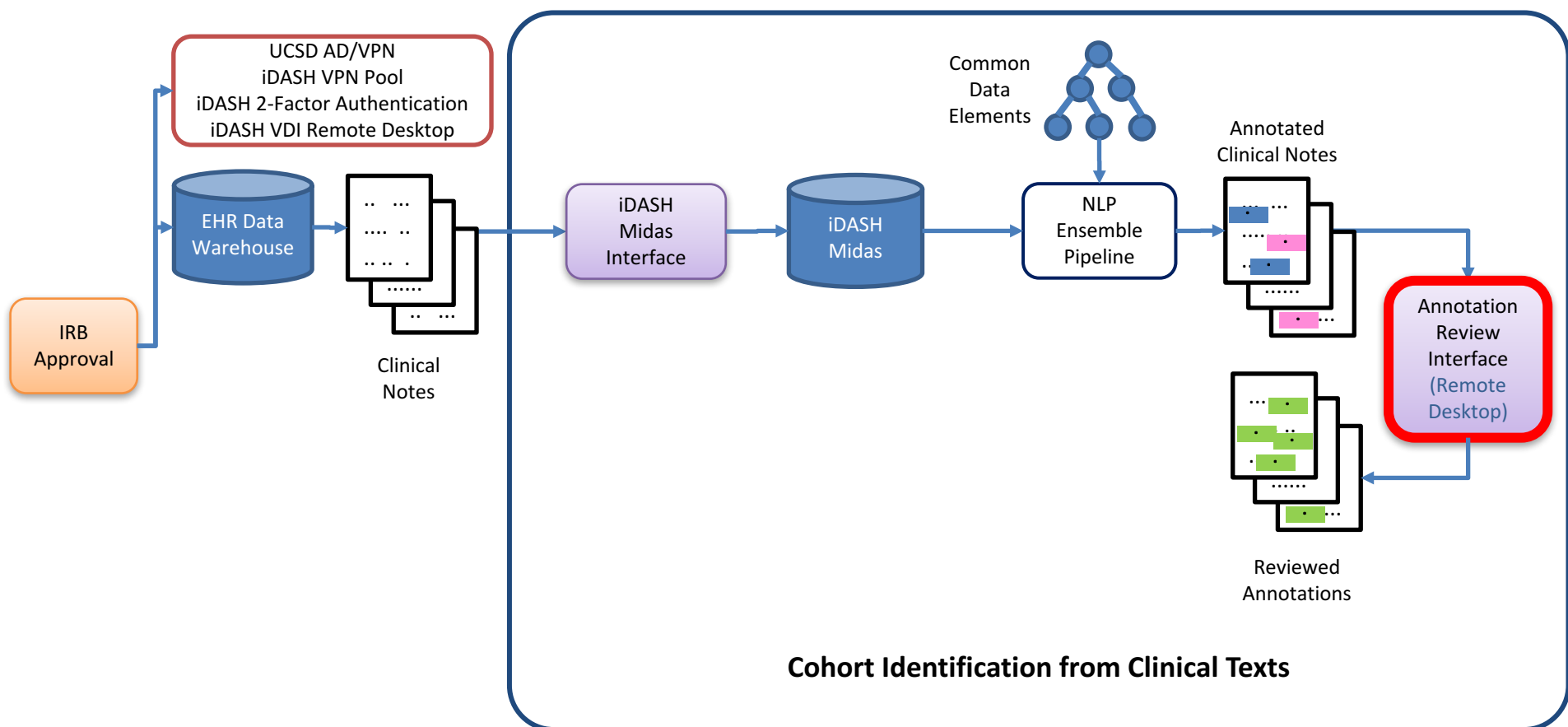
# CICT: NLP Ensemble Pipeline



# CICT: Secured Environment







<http://textmining.ucsd.edu:5005>