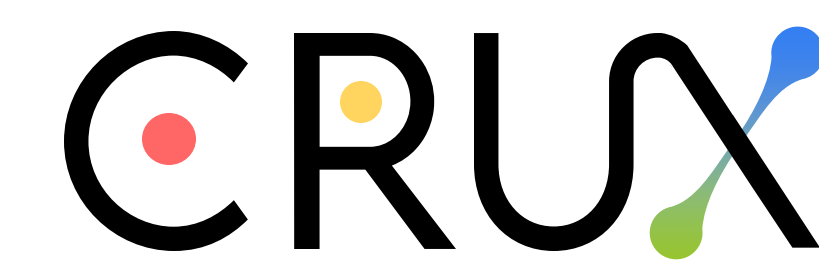


CRUX: Crowdsourced Materials Science ResoUrce and Workflow EXploration



Mengying Wang*, Hanchao Ma*, Abhishek Daundkar⁺, Sheng Guan*, Yiyang Bian*, Alp Sehirlioglu⁺, Yinghui Wu*
 *Department of Computer and Data Science, ⁺Department of Materials Science and Engineering, CWRU

OVERVIEW

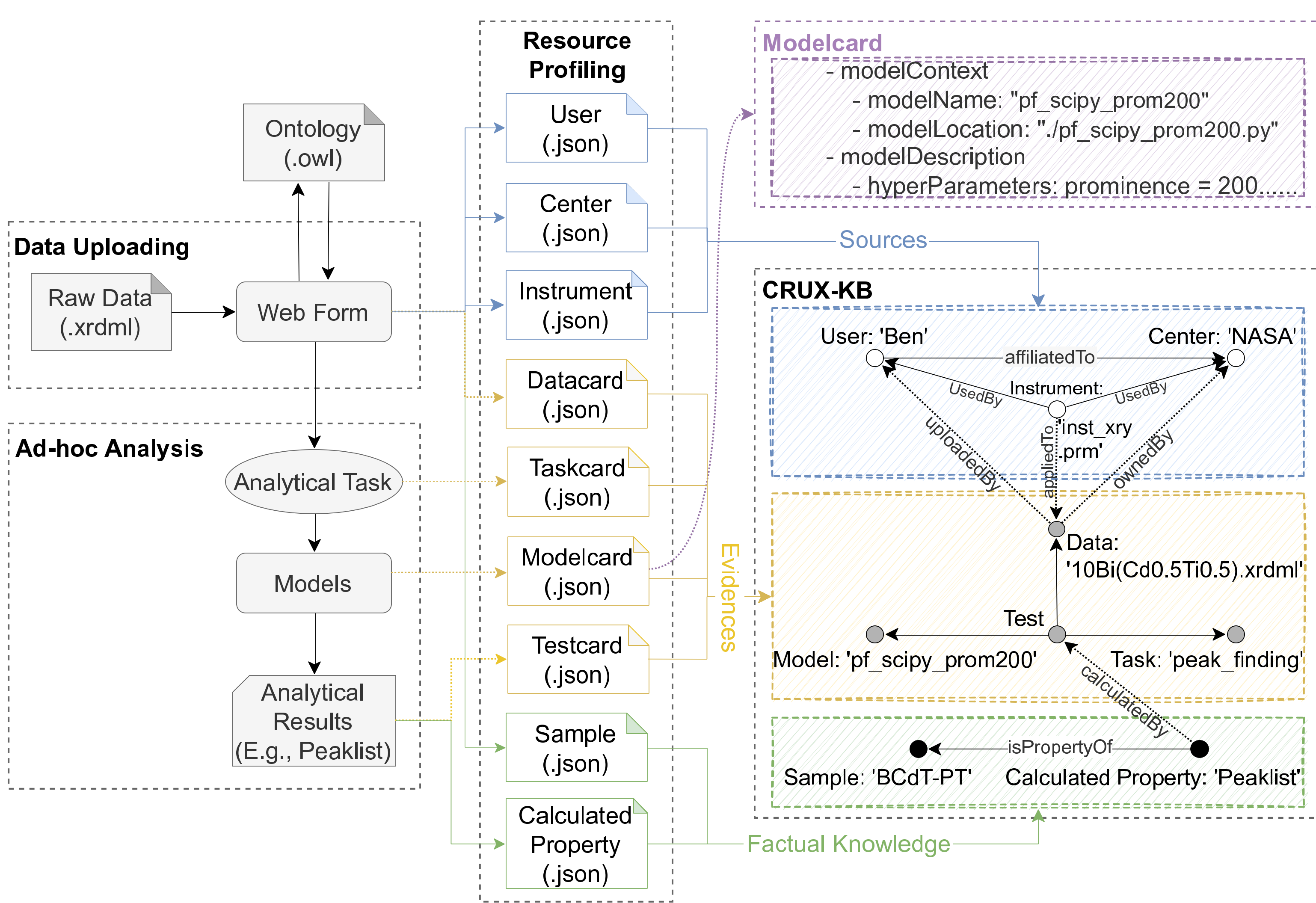
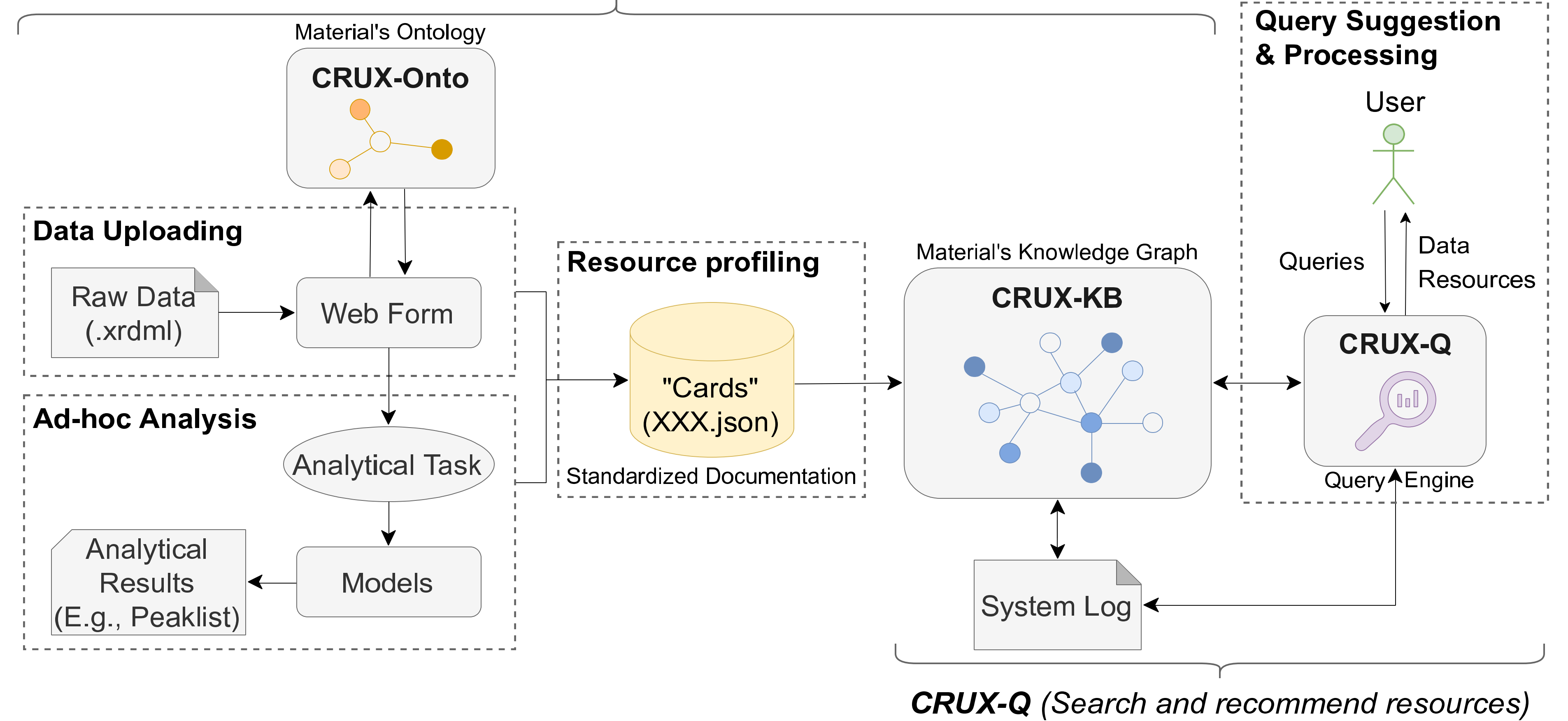
Many materials data and scripts/models are *unpublished* or *underutilized*.

Challenges:

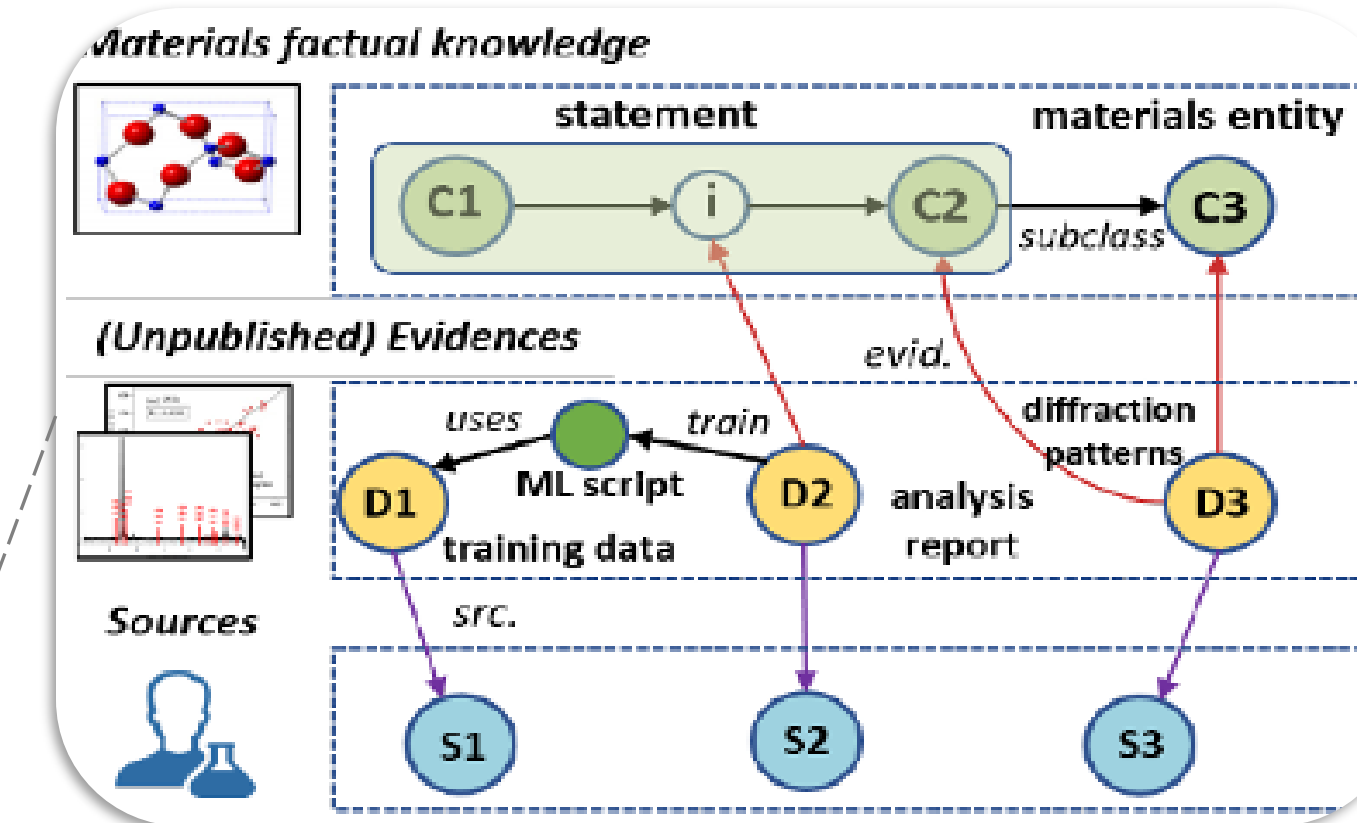
- Modeling of raw, isolated XRD data
- Make data resources discoverable for community.
- Recommend data/models for scientific workflows.

CRUX: A Crowdsourced data resources platform empowered by *Knowledge Graphs*, *Deep Graph Learning*, and *Exploratory Graph Search*.

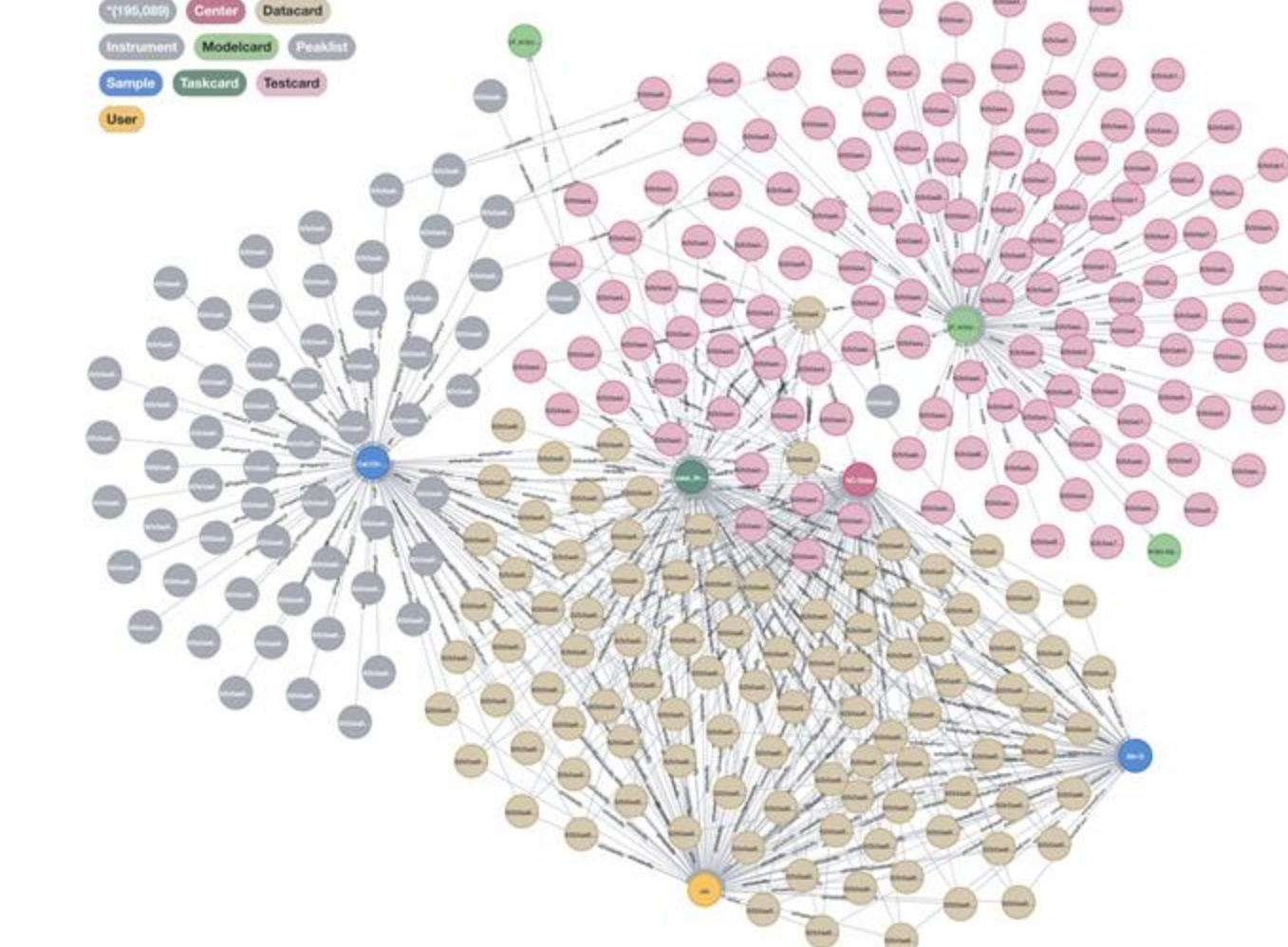
CRUX-IE (Transform raw data to knowledge graph)



"Three-tier" Knowledge Graph



Knowledge Graph (fraction)



CRUX-IE

- Information extraction engine:**
- Profile and extract facts from the shared raw data resources and metadata into a set of standardized documentation in JSON format, called "Cards",
 - Validate the JSON objects with a material ontology **CRUX-Onto**.
 - Recognize triples from JSON objects to enrich **CRUX-KB**, a multi-layered materials knowledge graph.

CRUX-Q

Query engine with GNN-based inductive recommendation (Oracle):

Resource Search:

Input: keywords.

Output: suggested queries/resources.

"Workflow-centric" Search:

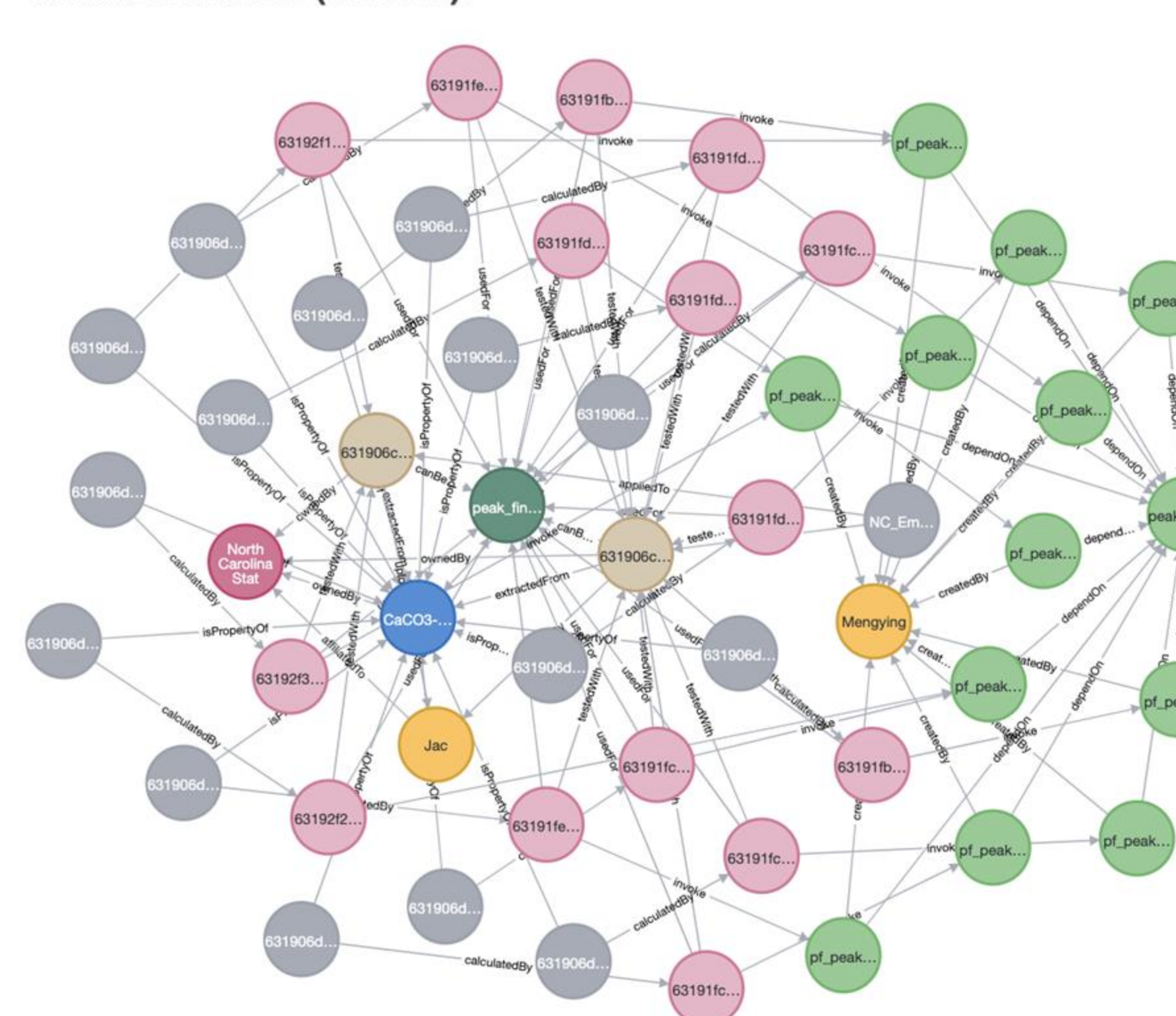
Input: keywords or pre-defined workflow by YAML.

Output: suggested workflow queries and instantiated workflow with recommended resources.

Query Example:

"Find the *datasets* for sample 'CaCO3-TiO2' provided by 'NC-State', a *task* with the name 'peak_finding', and the *models* implemented with the library 'peakutils'."

Visualized Result (fraction):



Overview

Node labels

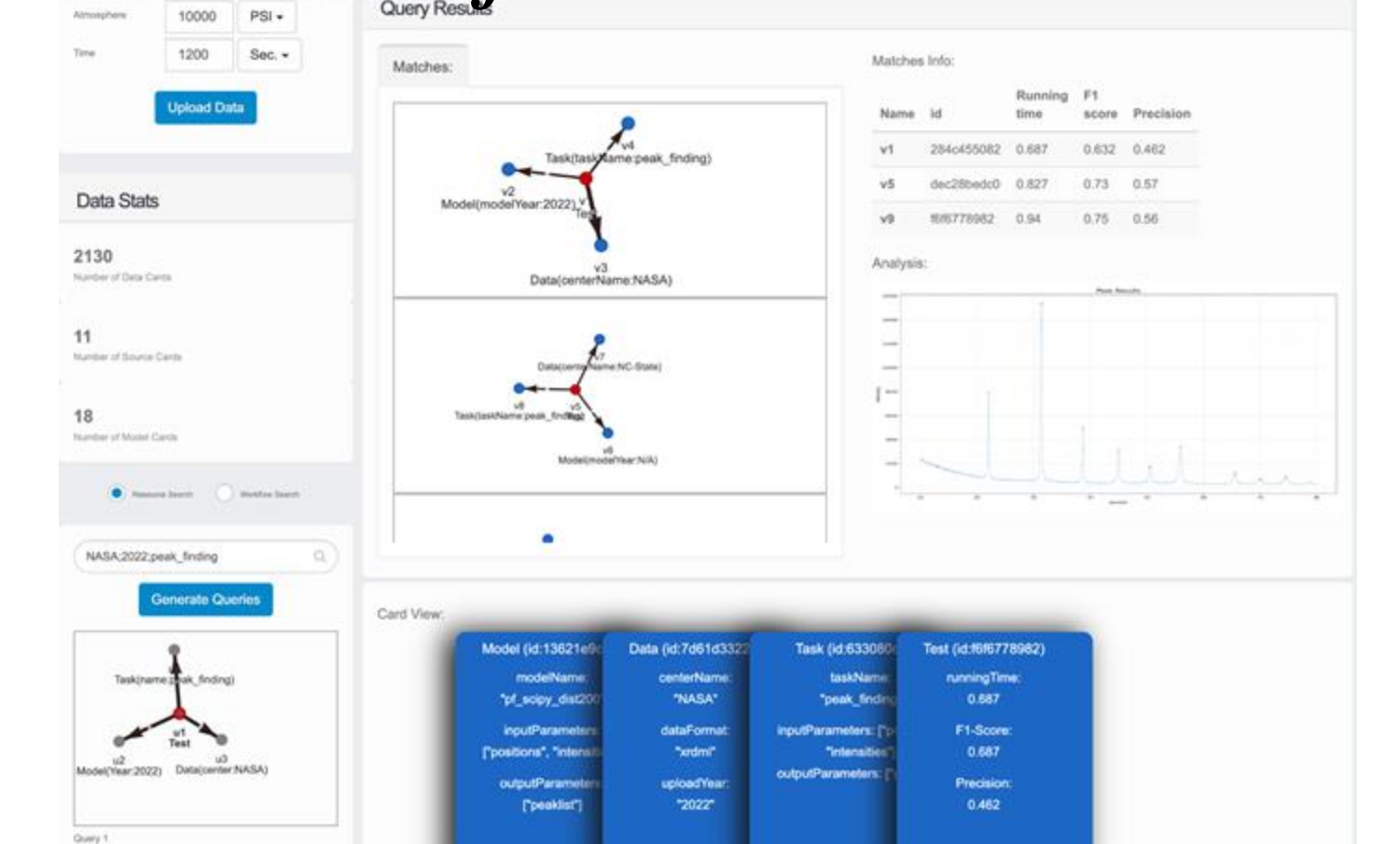
- * (51)
- Sample (1)
- Center (1)
- Datacard (2)
- Modelcard (13)
- Taskcard (1)
- Testcard (15)
- User (2)
- Peaklist (15)
- Instrument (1)

Relationship types

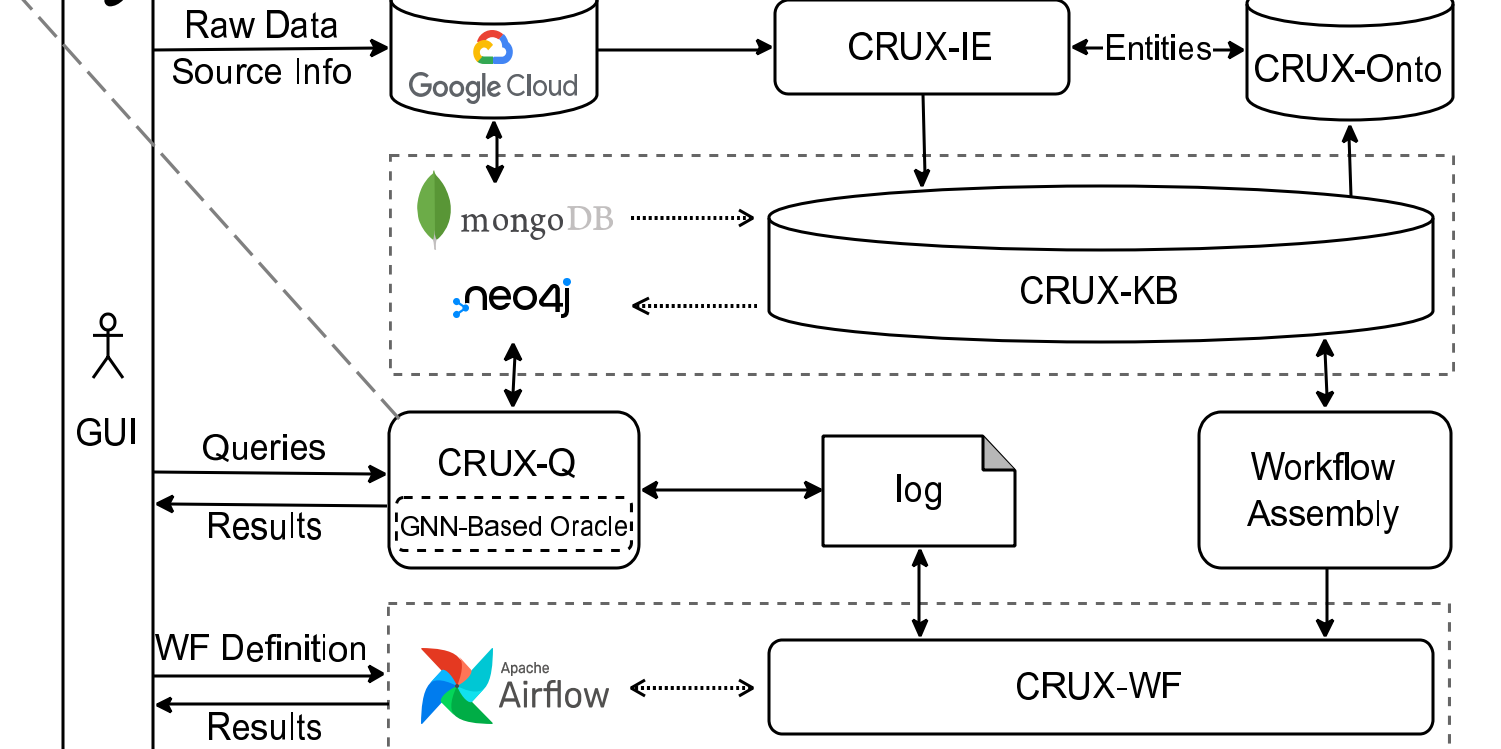
- * (111)
- ownedBy (3)
- canBeUsedTo (2)
- extractedFrom (2)
- uploadedBy (2)
- createdBy (12)
- dependOn (12)
- testedWith (15)
- usedFor (15)
- invoke (15)
- affiliatedTo (1)
- isPropertyOf (15)
- calculatedBy (15)
- appliedTo (2)

Displaying 51 nodes, 111 relationships.

Demo Interface



System Architecture



ACKNOWLEDGEMENT

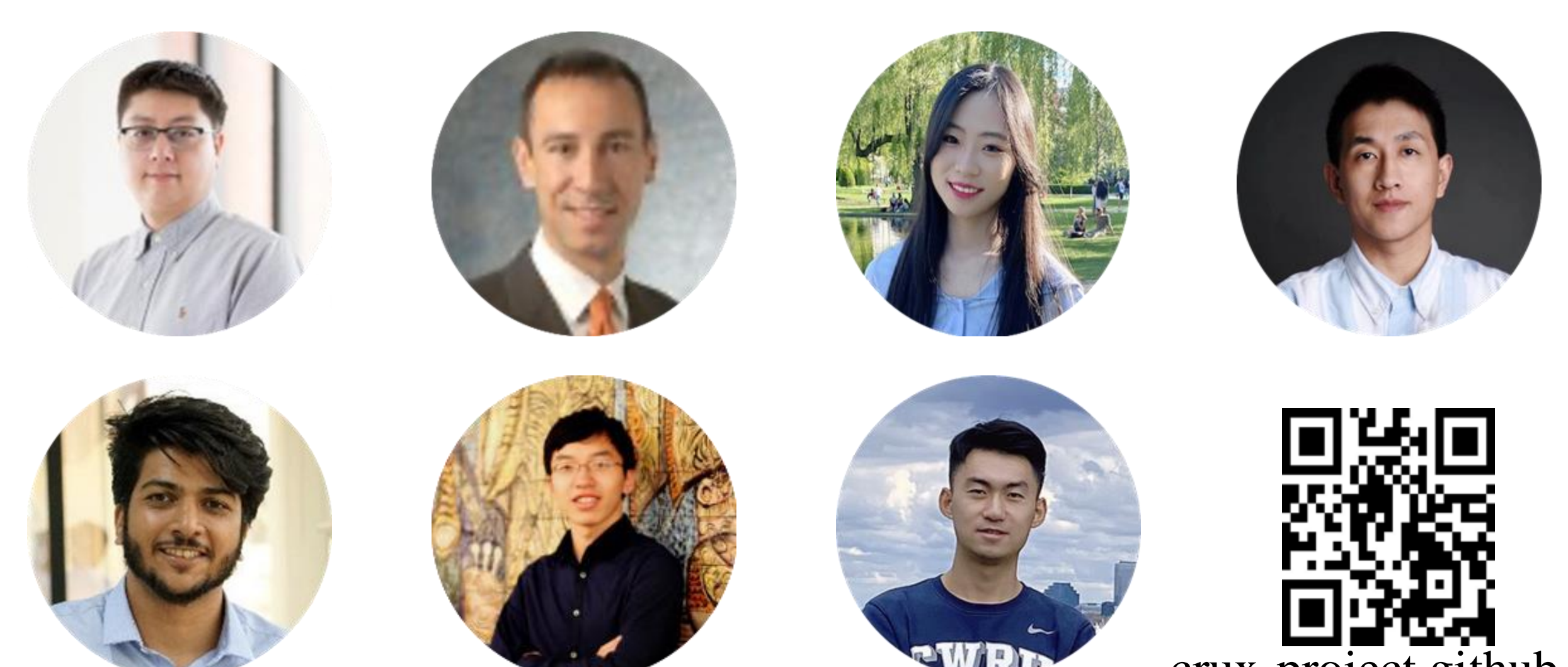
CRUX Community (Growing):



Funded by NSF:
(CSSI OAC-2104007)



CRUX TEAM



crux-project.github.io