Elements: Crowdsourced Materials Data Engine for Unpublished XRD Results

CASE WESTERN RESERVE UNIVERSITY EST. 1826

CRUX: CRowdsourced data infrastructure to curate, discover, and recommend Unpublished XRD data and analytical results empowered by Knowledge Graphs, Deep Graph Learning, and Exploratory Graph Search.

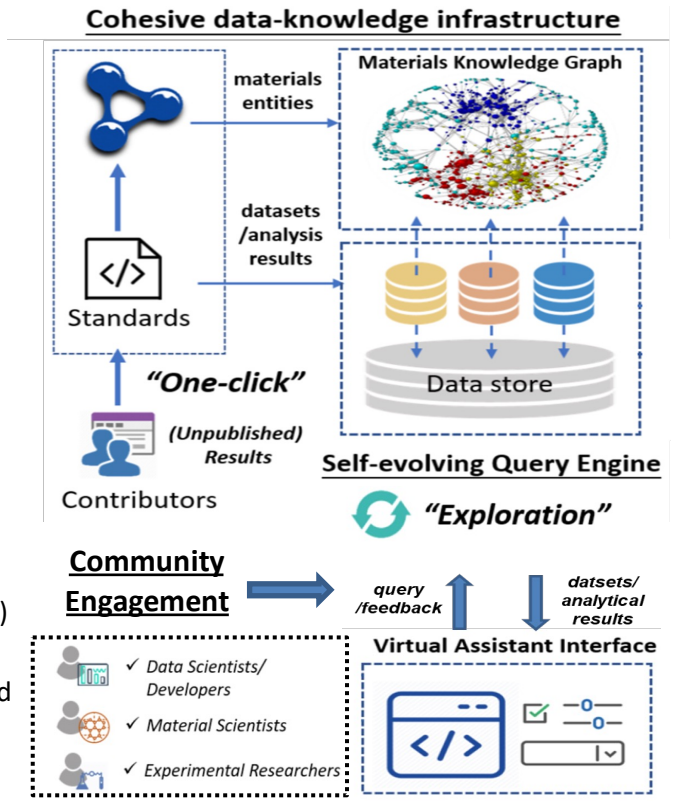
High-value materials data resources such as multiphase X-Ray Diffraction (XRD) remain *unpublished* and never made available for public.

Challenge:

- Modeling of raw, isolated XRD data
- Search and access high-value XRD data and the corresponding processing metadata
- Recommend XRD data and resources for scientific workflows

Research:

- Three-tier Knowledge Graph Model (CRUX-KB)
- "One-click" Data Sharing & Data Fusion
- Exploratory Query Infrastructure to search and explore with "What", "Why", "How" and "Who" queries



Highlight:

- **CRUX-KB**: 65K in 8 types (e.g., center, user, tests, sample, data, models, tasks); 167K triples in 12 types (e.g.,ownedBy, testwith, caculatedBy); 2121 datasets; 31826 tests; 8 contributors.
- Community: ICDD, MDI, UIUC, NC State, NASA, CWRU, UNSW, etc.
- Novel techniques developed to explain query answers (**SIGMOD** '21, **ICDE** '21) and suggest queries in knowledge graphs (**WSDM** '22, **ICDE** '22)

Scientic Impact:

- Query suggestion and knowledge search to explore materials data
- Data resource recommendation for materials scientific workflows
- Open and self-evolving materials knowledge repository towards unlocking new research

Broader Impact :

- Promote collaborative Research
- Pathways to large ML/AI driven tasks e.g., prediction of crystal structures or solubility limits of different compounds
- Provide open data vault of processing metadata tied to XRD data
- Outreach and Promote underrepresented groups
- Curriculum & Tools: CRUX-Q-based Why & What-if analysis for materials science query and analysis pipelines

CRUX team:


Yinghui Wu (Dept. of Comp. & Data Sciences, CWRU)


Alp Sehirlioglu (Dpt. of Mat. Sci. & Eng., CWRU)


Mengying Wang (PhD, CDS)


Hanchao Ma (Ph.D, CDS)


Abhishek Daundkar (Ph.D, MSE)


Sheng Guan (PhD, CDS)


Yiyang Bian (MS, CDS)

Project Site: crux-project.github.io

NSF CSSI OAC-2104007

PI: Yinghui Wu, Co-PI: Alp Sehirlioglu
Case Western Reserve University
yxw1650@case.edu, axs461@case.edu