

Electronic Health Record Data Extraction: Current Status and Continuing Challenges

2023 Childhood Cancer Data Initiative Annual Symposium

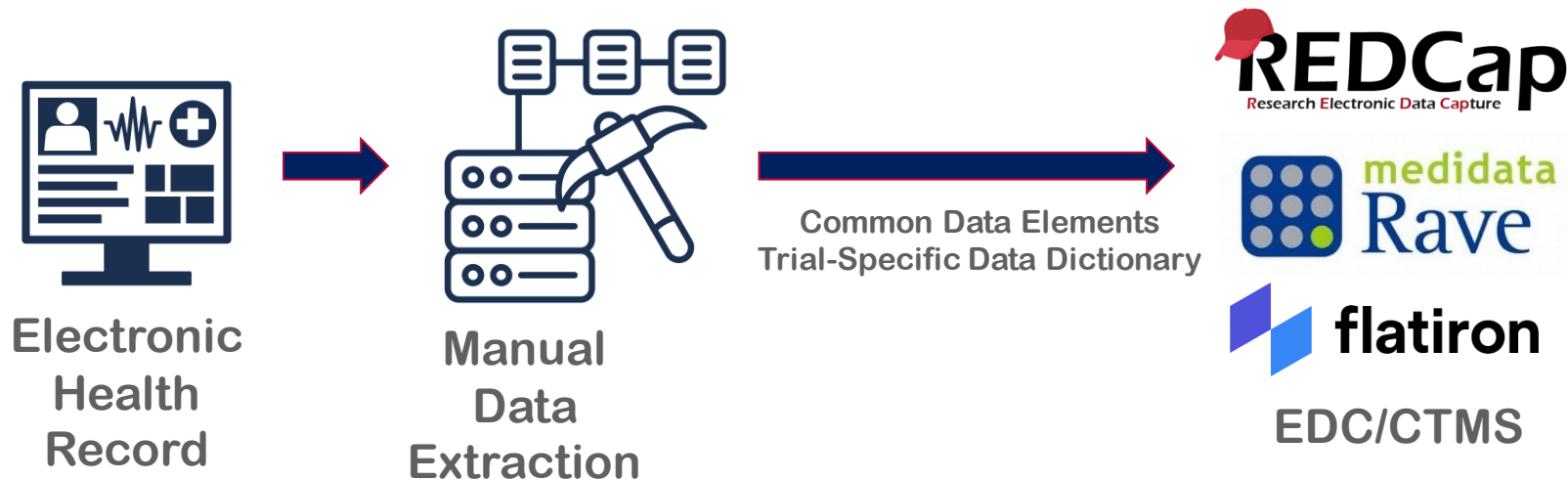
Breakout Session #3

Discussion Questions

- What are the potential challenges with EHR data extraction at a broader range of institutions?
- What are the limitations of extracting free text EHR data that we need to find ways to approach?
- How do we approach data in PDFs?
- What other data modalities should be extracted?
- What is the promise of NLP?
- What is the potential for data standardization within the EHR pre-extraction?
- Should there be an exchange standard that will facilitate exchange of data across sites?
- What are concerns around consent and deidentification for data sharing with CCDI?

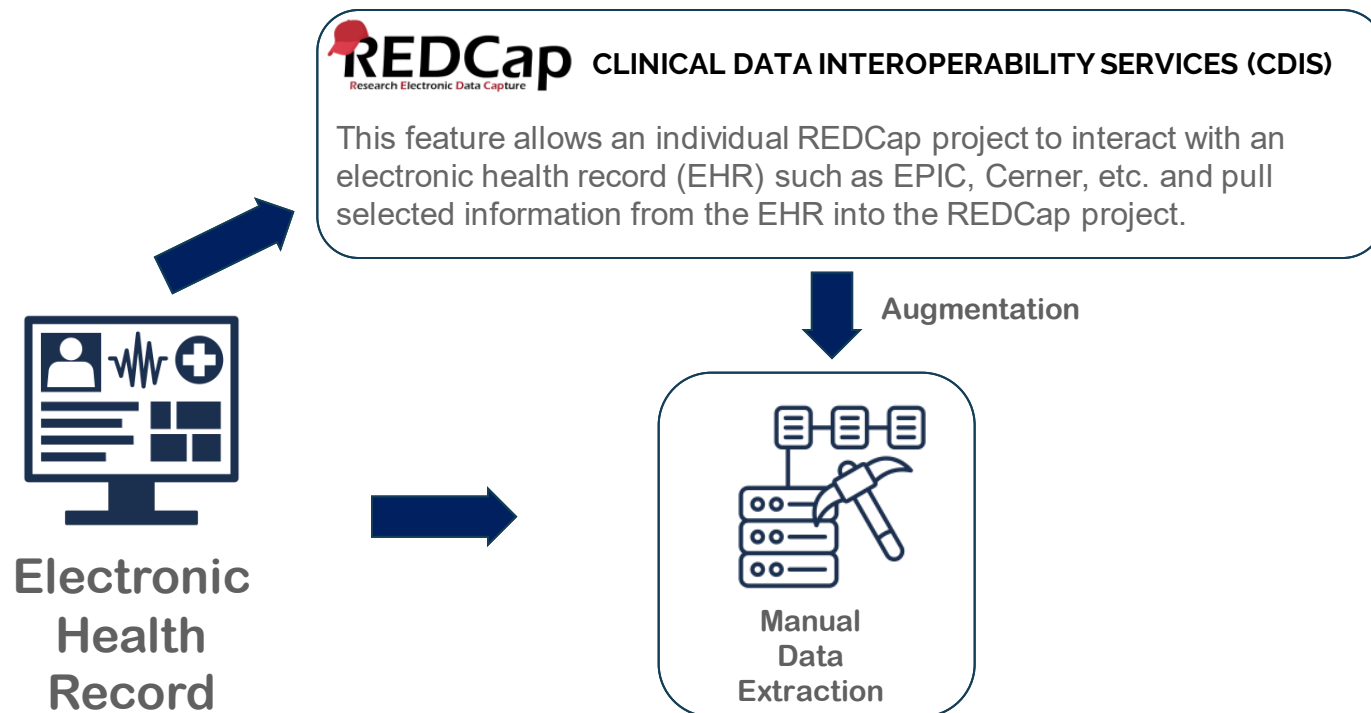
Current State of Clinical Data Collection

- Most clinical data extraction for the purposes of IRB-approved research and observational or clinical trials is performed by human beings who are trained and have some level of domain expertise.



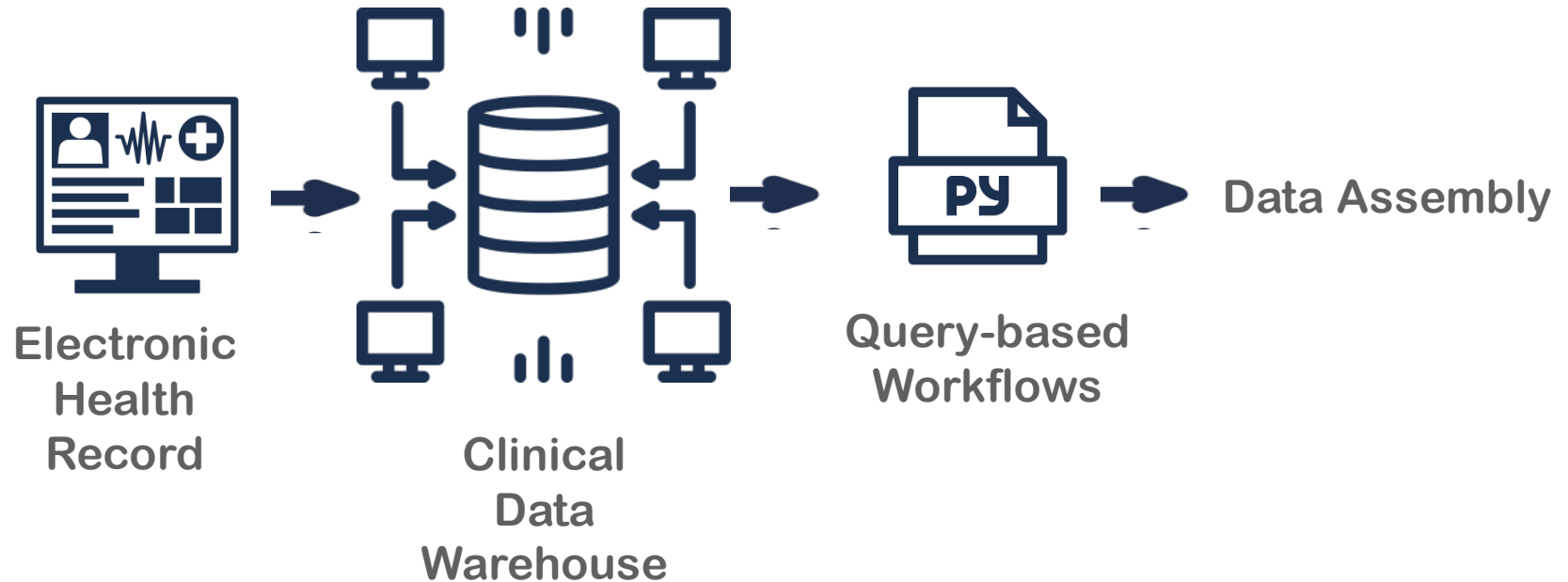
Emerging State of Clinical Data Collection

- Augmentation of human curation by automated workflows (API-Based)

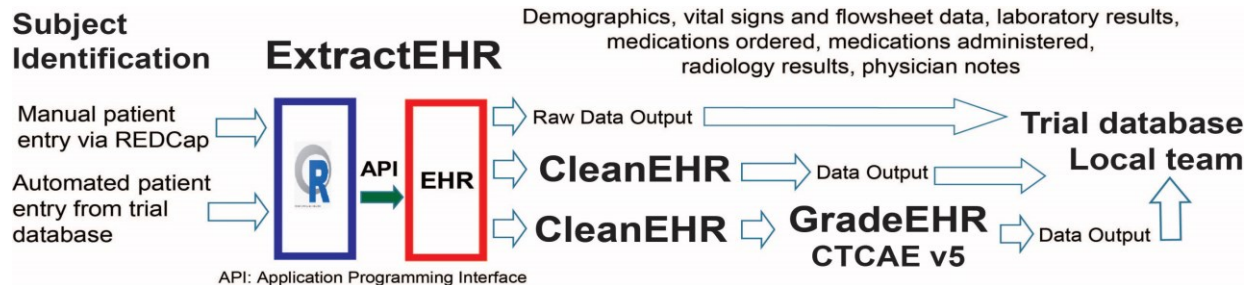


Emerging State of Clinical Data Collection

- Augmentation of human curation by automated workflows (query-based)



ExtractEHR Pilot



- R package that extracts data from EHR data warehouse
 - Implemented at multiple hospitals to extract range of data elements with high accuracy
 - Includes range of data types: structured (e.g. labs), semi-structured (e.g. radiology reports), unstructured (e.g. notes)
- Extracted data needs processing for use (CleanEHR, GradeEHR)
 - Key to focus on data quality, formatting to match needs of use case
 - Use to create granular analytic grade datasets (e.g. accurate rates of laboratory adverse events during therapy)
- CCDI pilots ongoing to test use for clinical trials (NCT05020951, NCT03126916) and SEER registry

EHR Extraction Use-Case Opportunities/Challenges/Requirements

- Enables scaled longitudinally, increased accuracy, and automation

EHR Extraction Use-Case Opportunities/Challenges/Requirements

- Enables scaled longitudinally, increased accuracy, and automation
- Data must still be converted to Information through interpretation: a human element will always be needed but that some elements of data extraction can be optimized with automation

EHR Extraction Use-Case Opportunities/Challenges/Requirements

- Enables scaled longitudinally, increased accuracy, and automation
- Data must still be converted to Information through interpretation: a human element will always be needed but that some elements of data extraction can be optimized with automation
- EHR extraction prevents some of the “pdf-ing” challenge. However, important data resides in clinical notes and non-standard texts. Opportunities for downstream optimization of NLP or other workflows.

EHR Extraction Use-Case Opportunities/Challenges/Requirements

- Enables scaled longitudinally, increased accuracy, and automation
- Data must still be converted to Information through interpretation: a human element will always be needed but that some elements of data extraction can be optimized with automation
- EHR extraction prevents some of the “pdf-ing” challenge. However, important data resides in clinical notes and non-standard texts. Opportunities for downstream optimization of NLP or other workflows.
- Opportunities to optimize not only on data coming out, but also on data going into the EHR. Pilot projects could explore standardizing clinical data entry in ways that could be expanded across a national network

EHR Extraction Use-Case Opportunities/Challenges/Requirements



Electronic
Health
Record



Data Exchange Standards



mCODE™

PCDC-INPSIRE

EHR Extraction Use-Case Opportunities/Challenges/Requirements



Electronic
Health
Record



Need to integrate non-EHR
resident data



Radiation Treatment



Pathology Slides



Imaging Data



Research/Clinical Trial
Data

EHR Extraction Discussion Summary

- Pediatric Cancer Research is a good environment to lead these efforts due to existing networks, tighter scope of focus areas, and the higher likelihood that pediatric patients stay within a limited medical network for their care
- CCDI provides unique opportunities for EHR-based tactics that can enhance observational and natural history components of clinical trials
- Current COG or CCDI pilot projects can be connected to the EHR extraction efforts
- Initial pilot projects need to pair the technical compatibility and feasibility of implementation at a particular site with impactful use cases to demonstrate the potential and assess extensibility

EHR Extraction Discussion Summary – Next Steps

- Pilot efforts should be implemented presently with short (1-2 year) intervals
- Align efforts with broad / long-term CCDI goals including both:
 - (1) supporting scientific discovery in pediatric cancer biology and in clinical trials and
 - (2) mechanisms for real-time decision support tools for a national tumor board

Discussion Questions

- What are the potential challenges with EHR data extraction at a broader range of institutions?
- What are the limitations of extracting free text EHR data that we need to find ways to approach?
- How do we approach data in PDFs?
- What other data modalities should be extracted?
- What is the promise of NLP?
- What is the potential for data standardization within the EHR pre-extraction?
- Should there be an exchange standard that will facilitate exchange of data across sites?
- What are concerns around consent and deidentification for data sharing with CCDI?

U.S. Department of Health & Human Services
National Institutes of Health | National Cancer Institute

cancer.gov

1-800-4-CANCER