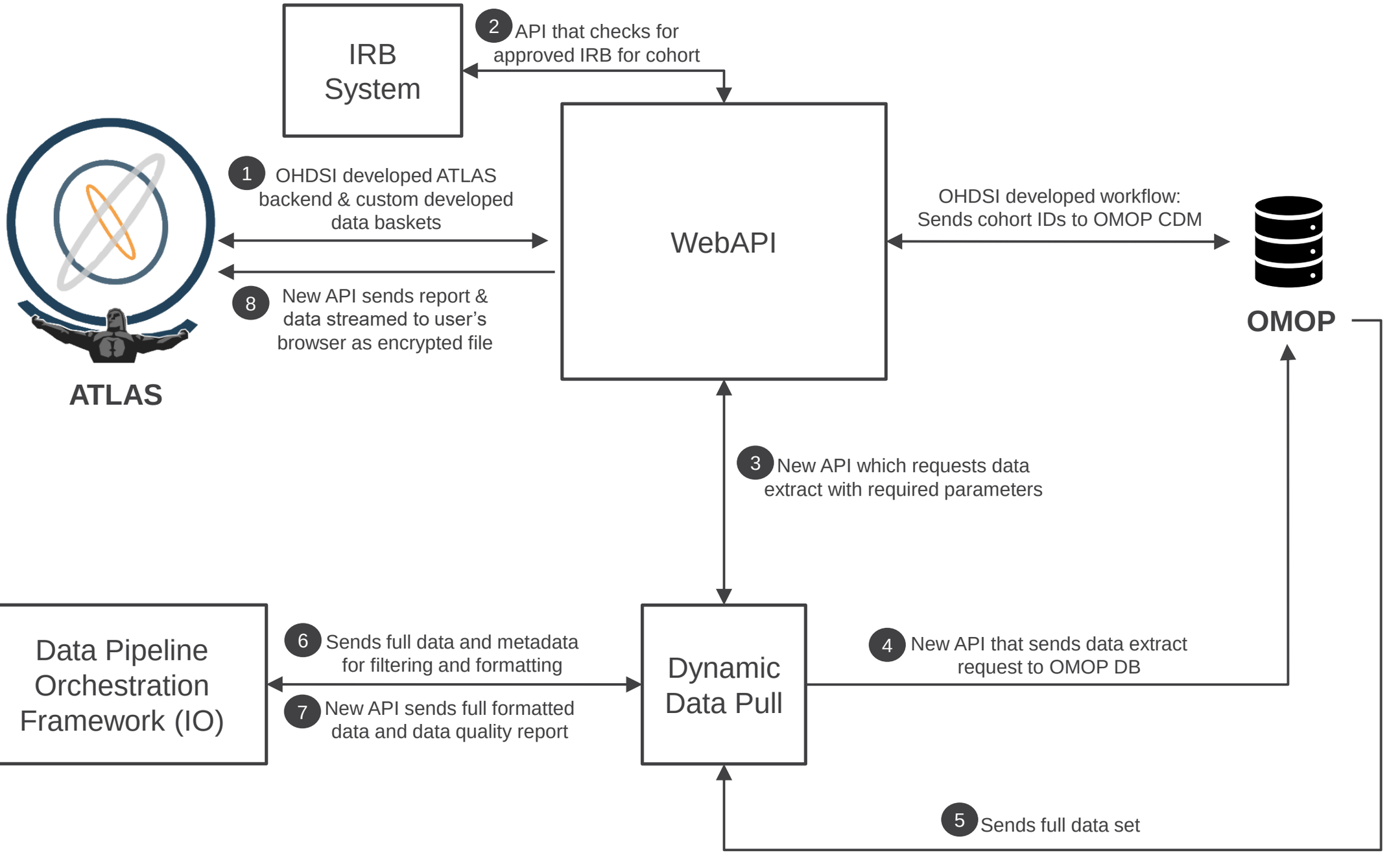
Data Extraction Workflow in ATLAS



Einstein-ATLAS: Leveraging OHDSI/ATLAS and Open-Source Development to Support Translational Research, Data Science, and Regulatory Compliance

Parsa Mirhaji, Selvin Soby, Erin Henninger, Chandra Nelapatla, Manuel Wahle, Boudewijn Aasman, Eran Belin

Background

OHDSI ATLAS is an open-source framework to enable self-service cohort identification, and some basic cohort-based analytics studies, using real-world observational health data (RWD), without exposing users to the underlying health records. ATLAS provides a framework for distributed analytics and collaboration by leveraging OHDSI/ATHENA terminology system, OHDSI/OMOP common data model, and the query formulation and sharing facilities. Although this provides a secured and collaborative framework to use RWD for research and discovery, it does not scale to unleash the full potential of RWD for diverse scenarios that require direct access to the underlying patient level records, record linkage, or identifiable health information (e.g., identifying patients for clinical trial matching and recruitment, pragmatic clinical trials, and personalized medicine).

Providing direct access to the underlying patient level health records can provide vast opportunities for investigators, data scientists, application developers, and informatics analysts to unleash power of RWD research, but puts OHDSI/ATLAS at odds with its original design principles for supporting RWD research and collaboration in a secured and confidential environment using deidentified observational health datasets. In this workshop we will introduce a modernized version of ATLAS extended by Einstein-Montefiore Center for Health Data Innovations to convert OHDSI/ATLAS to an enterprise self-service tool for RWD, supporting full spectrum of use-cases and data needs in an everchanging and expanding landscape.

Methods

We have established a dedicated RWD based on OHDSI/OMOP-CDM v5.3.1 to support the research community with high-quality standards-based analysis-ready datasets and participate in open science through multi-institutional collaborations and multi-site data consortiums. In order to support a wide range of users and use-cases, we have deployed extensive software development, interaction design, and data management resources to adopt and extensively improve upon the open-source OHDSI/ATLAS analytics and collaboration platform. Our extensions and improvements include, but is not limited to the following functionalities:

1) HIPAA compliant user level authorization security based on a *least-privilege* permission model to allow a private space to access and retrieve RWD.

2) A versatile interactive framework to design data baskets and index events linked to ATLAS cohorts, which then can be used to retrieve patient-level data and extract analysis-ready clinical datasets (in identified, limited-identified, and linkable de-identified, and fully de-identified formats).

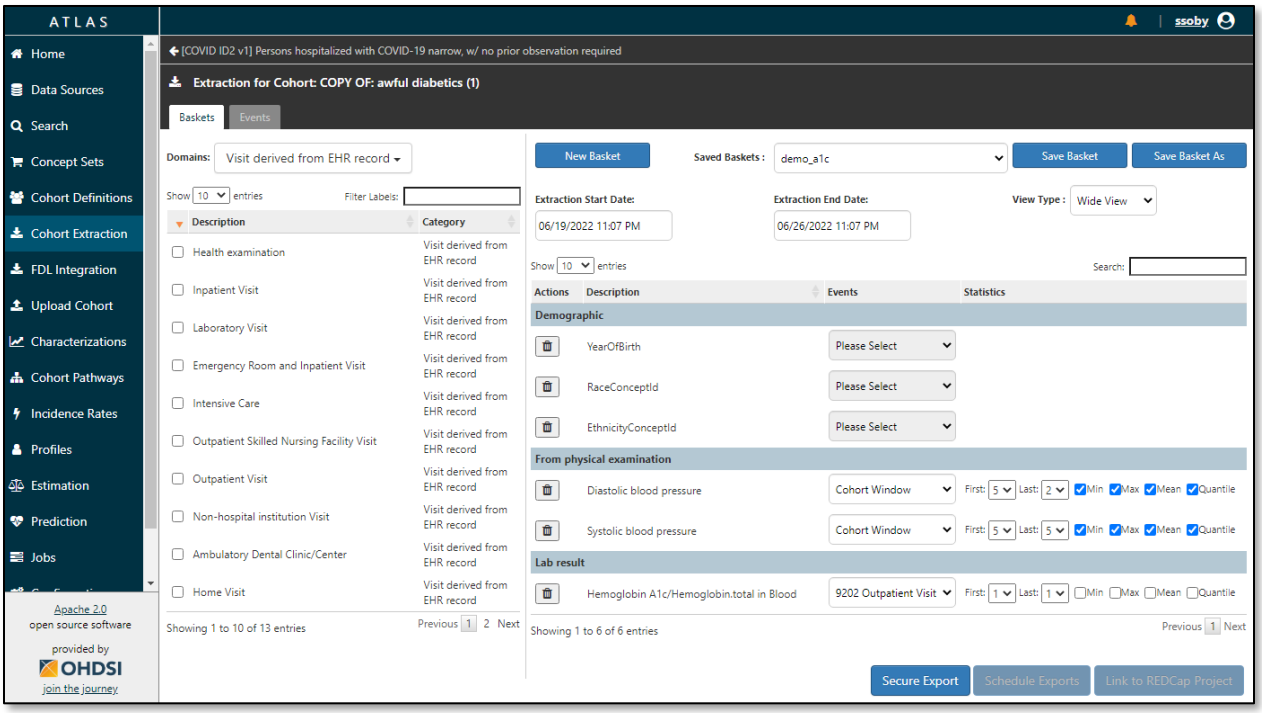


Figure 1: Cohort extraction workflow with data baskets

3) A comprehensive data quality characterization and reporting service linked to cohorts and data baskets to provide granular data quality reports

4) A workflow to upload user-defined cohorts into ATLAS for analysis, data extraction, and subsequent collaboration. This feature will be linked to REDCap to provide a two-way data flow between cohorts defined within REDCap projects, and Cohorts defined by ATLAS.

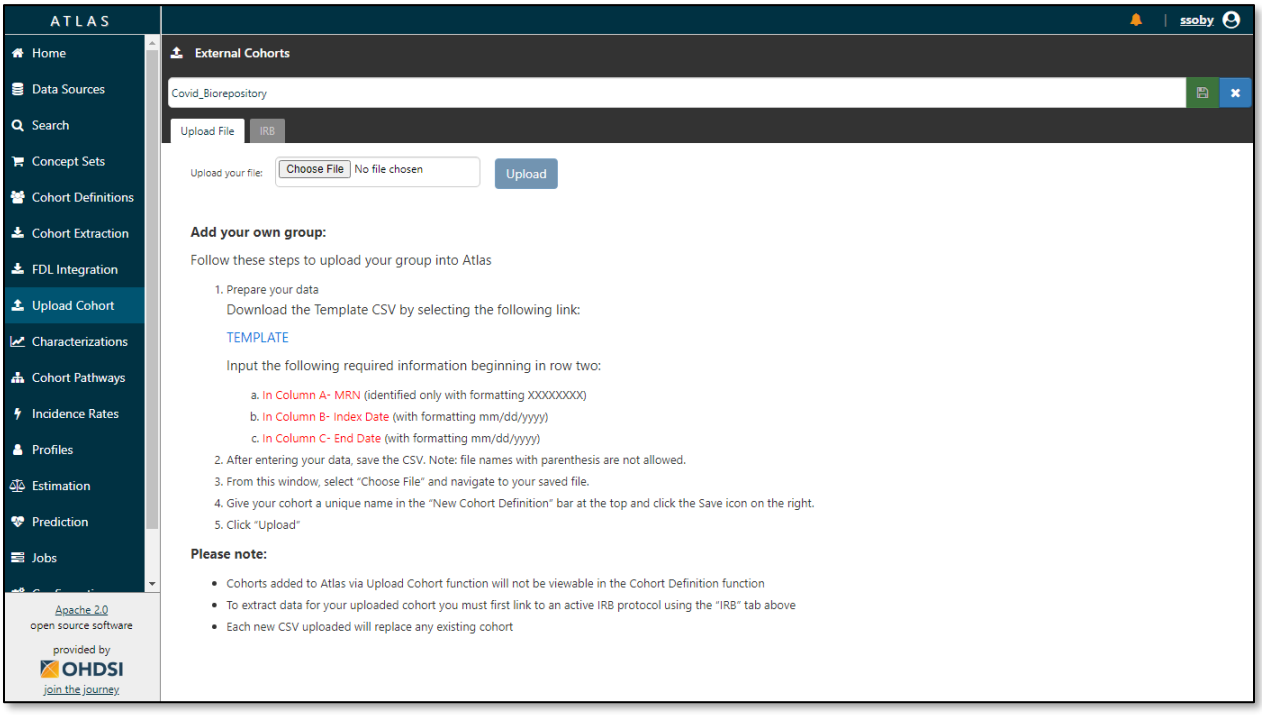


Figure 2: Cohort upload page

5) Using cryptography and a Master Entity Indexing (MEI) service to generate identifiers for Privacy Preserving Record Linkage (PPRLs) that support the linking of OMOP data to other modalities of data

6) A novel IRB Web-API and services-based integration framework to enable linking users, cohorts, data baskets, and analytics to specific IRB protocols which enforce regulatory monitoring, compliance and reporting requirements, at the user and protocol levels in near real-time

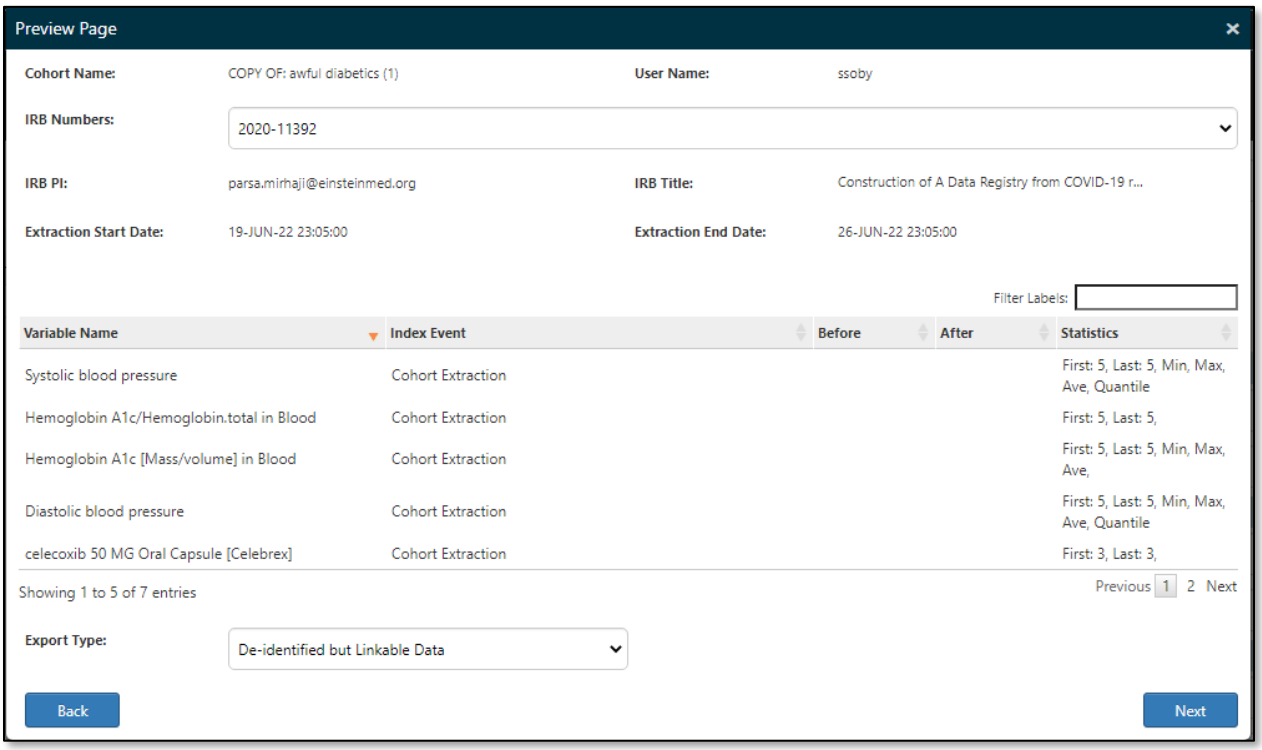


Figure 3: Cohort extraction overview page with data basket and linked IRB information

7) Honest brokering/re-identification services controlled by IRB interoperability to support the entire life cycle of clinical research and trials

8) Further extending OMOP-CDM and OHDSI terminology system to allow more granular site-specific names of facilities, care providers, and care locations, for cohort creation, and data extraction queries

9) A versatile model to expand ATLAS cohort definition framework to allow linked federated data (natural language text processing, genetic information) for cohort identification.

10) An innovative web-API extension that provides real-time interoperability with our two IRB systems at Einstein and Montefiore: Einstein IRB and BRANY Central IRB. This provides just-in-time access to updated IRB protocols (including information about status, study team, PHI authorization levels, and other important protocol-level metadata). The IRB web-API is leveraged by the Einstein/ATLAS Self-Service tool to implement a virtual honest-broker service that consistently enforces and protects the confidentiality and privacy of human subjects in all data and research pipelines supported by OHDSI ATLAS and OMOP platforms.

Results

The Einstein-Montefiore OMOP/ATLAS implementation supported 110 research protocols, nine national data consortiums, and seven network trials in 2021. The RDW is updated daily with respect to clinical and administrative data domains, clinical notes (clinical, radiology, pathology reports), and patient

demographics (> 3.8 million unique patients). The novel Einstein/ATLAS Self-Service Analytics platform, which currently has 358 registered users, enables self-service computable phenotyping and querying tools for defining cohorts, eligibility screening, case finding, real-world-data collaborations, and secured data extractions.

Conclusion

OHDSI open source platforms and community resources have enabled Albert Einstein college of medicine and Montefiore Health System to extend and leverage innovative informatics, software engineering, and data science over the past several years to develop robust and scalable infrastructure to support our goals of accelerating the integration and use of multimodal, heterogeneous datasets, enhancing the health informatics and analytics services, democratizing and scaling access to observational health data for research and discovery, diversifying clinical trials recruitment to promote health equity, and to provide informatics and data science education and training to our community.

We will contribute Einstein/ATLAS platforms for free and open-source use through OHDSI and CD2H GitHub repositories.

References

1. N/A