

Ancillary Study Data Analysis Center (DAC) Support Budgets

Instructions

Please fill in your contact information and the estimated DAC scope of work for your ancillary study in this form, then submit to ECHO-Help@rti.org with subject "Ancillary Study DAC Cost inputs" to receive your DAC ballpark cost estimate.

Contact Information

Name: _____

Institution: _____

Email: _____

Phone Number: _____

Scope of Work

Depending on the needs of the ancillary study, the services that may be provided by the ECHO Data Analysis Center (DAC) are shown below. Please indicate what services may be needed so the DAC can provide a cost estimate for your project.

1. **Study Design and Orchestration** – Leverage the administrative, data management, and infrastructure systems and procedures developed for ECHO that support rapid implementation of the ancillary study protocol. Assist in orchestrating a workflow that integrates with the ancillary study protocol requirements and timeline for ECHO.
 - Generate specimen pull list based on participants inclusion criteria (e.g., exposure, covariates, and outcomes, repeated measurement requirements), assay requirements (e.g., volume, metal free tube, freeze/thaw status), reserve policy.
2. **Data Management** – Expand ECHO DAC systems to receive and perform data quality checks of incoming data and maintain documentation on ECHOPortal, a secure web portal for the submission, processing, storage, harmonization, and analysis of ECHO Cohort data and for collaboration of researchers using these data. Specify data types:
 - Laboratory assay results
 - Geospatial data
 - Other, specify _____
3. **Systems Pipelines** - Ensure the scientific validity of the study by implementing systems pipelines that apply validation algorithms for preventing, detecting and reporting on problems related to data integrity when ingesting data from various sources into the database.

- Develop new pipeline to ingest data type not previously collected in ECHO including setting up validation rules.
 - Update existing pipeline to ingest data with structure and validation rules similar to other data already being collected for ECHO.
 - Update assay pipeline to ingest new assay results from the Lab Core or another lab and perform validations.
4. **Infrastructure/Security** –Update the ECHO cloud-based computing infrastructure (ECHO Analysis Workbench, EAW) to accommodate new data or analytical tools; support new software or external data needed to conduct the research; maintain data use agreements to allow access to the EAW that describe the terms and conditions of use, and monitor activities conducted on the computer system to facilitate protection against unauthorized access, and to verify security procedures, survivability, and operational security.
- Store large files in high performance computing environment (e.g., genetic, epigenetic, metabolomics, microbiome, expression, imaging)
 - Generate analyzable data sets from genome wide SNP data: genotype calling, marker and sample QC, imputation of genotypes, merging with ECHO samples, harmonization of datasets, and annotation of SNPs.
 - Whole genome methylation data: calling of feature level methylation data, sample and probe level QC, combine with ECHO data, generation of beta matrix files, and annotation of probe files.
 - Metabolomic data: quantification of peaks, identification of metabolites associated with peaks, identification of alternative peak identifications, adjust quantification of batch and sites effects, bioinformatic annotation of metabolites and linkage to pathways, imputation/windsorization of missing data, generation of final analyzable datasets.
 - Microbiome data: QC of sequencing reads, alignment of reads per samples, harmonize data across samples, batch and sites, taxonomic annotation.
 - Monitor cluster performance in high performance computing environment for computationally intensive analyses.
 - Install specialized software for analysis in the EAW (e.g., image viewer)
5. **Data Analysis and Reporting** – Ensure success of the study by monitoring timelines and data quality through regular reporting and dissemination of QC/QA queries. Create de-identified version of the data and harmonize to data standards for submission to repositories to maximize the value of data to the scientific and general community.
- Provide training/orientation for data analysts that covers the ECHO Analysis Workbench (EAW) Setup, ECHO Portal Resources, overview of the ECHO Cohort Data Platform, SQL Table Names and Conventions, Data Warehouse Snapshots, EAW Shared Resources, Best Practices Tips & Tricks.
 - Provide more intensive 1-1 support for analyst in developing their analysis dataset.
 - Develop reports to track study status (e.g., transfer of data from the Lab Core) and data quality.

- Develop best practices guidelines for analysis for new data being added to the ECHO Cohort Data Platform.
- Submit de-identified data to the relevant repositories.

6. **Communications and ECHO Cohort Site Support** - Respond to inquiries to the Help Desk and maintain a Knowledge Base of FAQs, and conduct training for site staff on the requirements, methods and functions of the various systems to support.

- Provide training of study coordinators and data managers on data systems
- Addressing High Performance Computing (HPC)/analysis questions, deal with data issues that arise, address review of data for removal from HPC, limited writing of methods for publications, and review of publications.

Computer and Storage Other Direct Costs

____ Number of staff needing access to the ECHO Analysis Workbench

High-performance computing (HPC) environment needed.

____ Number of staff needing access to the HPC

____ Storage, specify units: Gb or Tb

____ CPU years of compute time per year

Example DAC costs (not including Other Direct Costs)

Scenario 1: Plans to use existing data on the ECHO Cohort Data Platform

Data Analysis and Reporting

- Provide training/orientation for data analysts that covers the ECHO Analysis Workbench (EAW) Setup, ECHO Portal Resources, overview of the ECHO Cohort Data Platform, SQL Table Names and Conventions, Data Warehouse Snapshots, EAW Shared Resources, Best Practices Tips & Tricks.

Cost: \$5,000

Scenario 2: Analysis of biospecimens to look at inflammatory markers in pregnancy related to a child outcome.

Study Design and Orchestration

- Identify specimen pull list based on participants inclusion criteria (e.g., exposure, covariates, and outcomes, repeated measurement requirements), assay requirements (e.g., volume, metal free tube, freeze/thaw status), reserve policy.

Data Management

- Laboratory assays results

Systems Pipelines

- Update assay pipeline to ingest new assay results from the laboratory (Lab Core or external lab) and perform validations.

Data Analysis and Reporting

- Provide training/orientation for data analysts that covers the ECHO Analysis Workbench (EAW) Setup, ECHO Portal Resources, overview of the ECHO Cohort Data Platform, SQL Table Names and Conventions, Data Warehouse Snapshots, EAW Shared Resources, Best Practices Tips & Tricks.

Cost: \$15,000