
Research and Applications

ATRI EDC: a novel cloud-native remote data capture system for large multicenter Alzheimer’s disease and Alzheimer’s disease-related dementias clinical trails

Gustavo A. Jimenez-Maggiora , Stefania Bruschi, Hongmei Qiu, Jia-Shing So, and Paul S. Aisen

Alzheimer’s Therapeutic Research Institute, Department of Neurology, Keck School of Medicine of USC, Los Angeles, California, USA

Corresponding Author: Gustavo A. Jimenez-Maggiora, MBA, Alzheimer’s Therapeutic Research Institute, Department of Neurology, Keck School of Medicine of USC, 9860 Mesa Rim Road, San Diego, CA 92121, USA; gustavoj@usc.edu

Received 17 August 2021; Revised 5 November 2021; Editorial Decision 24 December 2021; Accepted 31 December 2021

ABSTRACT

Objective: The Alzheimer’s Therapeutic Research Institute (ATRI) developed a novel clinical data management system, the ATRI electronic data capture system (ATRI EDC), to address the complex regulatory, operational, and data requirements that arise in the conduct of multicenter Alzheimer’s disease and Alzheimer’s disease-related dementias (AD/ADRDs) clinical trials. We describe the system, its utility, and the broader implications for the field of clinical trials and clinical research informatics.

Materials and Methods: The ATRI EDC system was developed, tested, and validated using community-based agile software development methods and cloud-native single-page application design principles. It offers an increasing number of application modules, supports a high degree of study-specific configuration, and empowers study teams to effectively communicate and collaborate on the accurate and timely completion of study activities.

Results: To date, the ATRI EDC system supports 10 clinical studies, collecting study data for 4596 participants. Three case descriptions further illustrate how the system’s capabilities support diverse study-specific requirements.

Discussion: The ATRI EDC system has several advantages: its modular capabilities can accommodate rapidly evolving research designs and technologies; its community-based agile development approach and community-friendly licensing model encourage collaboration per the principles of open science; finally, with continued development and community building efforts, the system has the potential to facilitate the effective conduct of clinical studies beyond the field of AD/ADRD.

Conclusion: By effectively addressing the requirements of multicenter AD/ADRD studies, the ATRI EDC system supports ATRI’s scientific mission of rigorously testing new AD/ADRD therapies and facilitating the effective conduct of multicenter clinical studies.

Key words: data management, clinical trials, Alzheimer disease, clinical research informatics, remote electronic data capture

Lay Summary

The Alzheimer's Therapeutic Research Institute (ATRI) developed a novel clinical data management system, the ATRI electronic data capture system, to address the complex regulatory, operational, and data requirements that arise in the conduct of multicenter Alzheimer's disease and Alzheimer's disease-related dementias clinical trials. We describe the system, its utility, and the broader implications for the field of clinical trials and clinical research informatics.

INTRODUCTION

Alzheimer's disease (AD) and Alzheimer's disease-related dementias (ADRDs) represent a serious public health crisis for healthcare systems around the world.¹ Despite significant challenges in the development of new disease-modifying therapies (DMTs) for AD/ADRD, the pipeline of candidate therapeutics remains substantial.² Considering these development efforts, the need for dedicated infrastructure and expertise to perform rigorous and efficient testing of candidate therapies became evident. The Alzheimer's Therapeutic Research Institute (ATRI) (<https://keck.usc.edu/atri/>) was founded in 2015 to address this need. As part of the Keck School of Medicine of USC, ATRI has grown into the largest NIH-funded AD/ADRD clinical trial coordinating center (CTCC) in the United States, designing, conducting, and analyzing over 20 multicenter clinical trials since its founding. As a comprehensive CTCC, ATRI provides leadership and expertise in multiple clinical research specialties and robust infrastructure. A core component of this infrastructure is an effective research data management platform. This platform, which must adapt to study-specific requirements, is anchored by a web-based clinical data management system (CDMS) developed by ATRI: the ATRI electronic remote data capture system (ATRI EDC). The purpose of this article is to describe the ATRI EDC system, its utility, and the broader implications for the field of clinical trials and clinical research informatics.

AD/ADRD clinical trial requirements**Operational complexity**

Multicenter AD/ADRD clinical trials are large, complex research enterprises that require coordination across multiple stakeholders—public and private sponsors, regulators, study teams, vendors, and trial sites. In this setting, timely communication and collaboration are critical to achieving study objectives and milestones effectively. Gaps can introduce compliance issues, reduce data quality, and compromise participant safety.

Geographic complexity

Multicenter AD/ADRD clinical trials are increasingly global in scope. Several clinical trials managed by ATRI, such as the Anti-Amyloid Treatment in Asymptomatic Alzheimer's study (A4 study),³ are conducted at trial sites across Asia, Australia, Europe, and North America. Multinational settings introduce a complex patchwork of regulatory, operational, data security, and data privacy requirements.

Complex data types, data flows, and study designs

AD/ADRD clinical trials require the management of large, heterogeneous databases composed of structured, unstructured, binary, and complex data types. These data types include clinical, biosamples (fluid, tissue), neuroimaging, genetic, multimedia (audio, video), and study documentation. The sources of data are increas-

ingly specialized and disparate, requiring ongoing monitoring of data quality and integrity. Mappings across databases must be established and maintained. Novel data types, such as sensor (IoT), wearable, administrative, log, and geospatial data, must also be managed and mapped with existing databases.⁴ The increased adoption of adaptive methods and platform designs in AD/ADRD trials introduces more demanding operational and data quality requirements.^{5,6} Adaptive designs require systems that are highly available and flexible to effectively assess data quality and maintain timely data flows to support interim analyses yielding adjustments to participant allocation and the number of treatment arms.

Extended study follow-up periods

Due to uncertain rates of disease progression, AD/ADRD clinical trial designs typically require extended follow-up periods of 18 months or more. Add to this protracted enrollment periods that span several years, and clinical trial durations frequently extend 5 years or more; instances of almost decade-long AD/ADRD clinical trials are not uncommon. Long-duration study designs present unique requirements that necessitate sophisticated forecasting and risk-based planning. In this context, sustainability, scalability, and flexibility in the face of unforeseen contingencies are critical.

Challenges to data quality

AD/ADRD trials are notoriously labor-intensive and expensive to conduct due to complex procedures, high screen failure rates, and prolonged follow-up periods.⁷ The shift from intervention to prevention studies, which rely on the timely assessment of genetic, neuroimaging, and biofluid biomarkers, further exacerbate these challenges.^{8,9} Consequently, it is common for AD/ADRD trials to experience high site staff turnover and low-enrolling sites, potentially leading to data quality issues.

Assessment of existing clinical data management systems

In late 2015, ATRI conducted an initial assessment of several open-source and commercial CMDMSs against the previously stated requirements.^{10–14} In each case, the systems offered limited capabilities, inflexible licensing models, prohibitive costs, or were no longer under active development. Therefore, ATRI decided to develop a novel CDMS (ATRI EDC) using cloud-native architecture and modern application development tools to address these requirements.

MATERIALS AND METHODS**ATRI EDC overview**

The ATRI EDC system was designed to facilitate the conduct of multicenter AD/ADRD clinical studies by providing a scalable, extensible, cost-effective, and regulatory-compliant electronic remote

Table 1. Technologies used to develop and host the ATRI EDC system

Domain	Subdomain	Tool/service name	
Applications/tools	Source code management	GitHub	
	Client-side web application development	AngularJS	
		ReactJS	
	Project management	Atlassian Suite—JIRA, Confluence	
	Continuous integration (CI), continuous deployment (CD)	Travis CI	
	Code test coverage analysis	Coveralls	
	Server-side web application development	Django—Python Web Framework	
	Application programming interface (API) development	Django REST Framework	
	Web application testing	Selenium	
	Identity management framework	SAML SSO	
	Cloud-based IT infrastructure—Amazon Web Services (AWS)	Compute	EC2
		Batch compute	AWS Batch
		Serverless compute	AWS Lambda
Web application framework		Amazon Elastic Beanstalk (EB)	
Containers		Amazon Elastic Container Service (ECS)	
Email		Amazon Simple Email Service (SES)	
Database		DynamoDB	
		Amazon RDS	
Machine learning		Amazon Sagemaker	
Machine learning-based code analysis		Amazon CodeGuru	
Developer tools		AWS CodePipeline	
		AWS CodeBuild	
Governance and management		Amazon CloudWatch	
		AWS CloudTrail	
		AWS Config	
		AWS CloudFormation	
		Amazon Macie	
		AWS System Manager	
		AWS Trusted Advisor	
	Amazon VPC		
	Amazon CloudFront		
	AWS API Gateway		
Networking and content delivery	Amazon Route 53		
	AWS Certificate Manager		
	AWS IAM		
	AWS GuardDuty		
	AWS Secrets Manager		
	AWS Key Management Service (KMS)		
	Amazon S3		
	Amazon Step Functions		
	Amazon Simple Queue Service (SQS)		
	Amazon Simple Notification Service (SNS)		
Security, identity, and compliance	Python		
	R		
Storage	Google Analytics		
	Application integration	Teamwork Desk	
Analytics and statistical reporting	Website analytics	FreshService	
	Technical support		

Note: Some tools and services are listed more than once due to their application in multiple domains.

data capture (eRDC) platform to manage large, heterogeneous data types and model complex dataflows and workflows. The system provides study teams with real-time visibility into a study's operations and facilitates communication and collaboration between the CTCC, sponsors, vendors, and sites. It accomplishes this by providing user interfaces that leverage contemporary web application design practices and methods (Table 1). At its core, the system uses metadata derived from a study's protocol (schedule of events, case report forms, data types, and key outcomes) to present a structured, study-specific, web-based clinical data management environment that guides users through the accurate and timely completion of study activities.

System development

Community-based agile development process

The initial version of the ATRI EDC system was released in September 2016, after 10 months of development. The development process adhered to the system development life cycle (SDLC) model, adapted to incorporate rapid prototyping and iterative stakeholder assessment consistent with Agile software development principles and methods.^{15,16} During this process, the development team met with multidisciplinary stakeholders frequently to present system prototypes and elicit feedback. The SDLC incorporated test-driven development, behavior-driven development and testing, continuous integration, and continuous deployment

methods to build, test, and validate the system continuously (Table 1).

Web application design principles

The ATRI EDC system was built using open-source web application frameworks and tools (Table 1) following the single-page application design pattern.¹⁷ This approach aims to provide users with a responsive, fluid experience reminiscent of a desktop application via a browser-based application. A complete description of the single-page application design pattern is beyond the scope of this article.

Partnership with cloud IT provider

During the development process, Amazon Web Services (AWS)¹⁸ was selected as the preferred cloud platform for the ATRI EDC project. This choice led to the establishment of an extensive ongoing partnership between ATRI and AWS that provides access to expertise, training, and support in cloud-based architecture and best practices. While designed to be platform-agnostic, the ATRI EDC system's performance characteristics can be optimized when running on the AWS cloud (Table 1). This optimized configuration leverages specific AWS services and features to achieve increased performance, security, scalability, and resilience.

Regulatory compliance

In the United States, computer systems used to manage data for submissions to regulatory agencies must comply with 21 CFR Part 11 (Part 11).¹⁹ This requirement applies to the ATRI EDC system. Traditional methods to demonstrate compliance require protracted, labor-intensive manual testing and validation procedures. These methods are costly, error-prone, lack reproducibility, and, ultimately, slow system innovation. The ATRI Informatics team developed novel automated testing and validation methods to address these challenges.²⁰ Integrating these methods into the SDLC process generated efficiencies and ensured Part 11 compliance on an ongoing basis. An independent assessment of the ATRI EDC system completed in January 2018 provided further evidence of Part 11 compliance.

Focus on data security and data privacy

Data security and data privacy are critical requirements for CDMSs. In the United States, the Health Insurance Portability and Accountability Act²¹ mandates these requirements and defines penalties for non-compliance. More stringent data regulations are in force in jurisdictions outside the United States. To address these requirements, ATRI partnered with Amazon Web Services Professional Services (AWS ProServ) to establish a HIPAA-compliant IT infrastructure to host the ATRI EDC system (Figure 1). This environment implements automated compliance and security policies using DevSecOps methods such as automated data classification, configuration, threat detection, and web application firewalls.^{22,23} As of August 2021, further work with AWS ProServ focused on compliance with HITRUST Common Security Framework²⁴ and the European Union General Data Protection Regulation²⁵ was underway.

System design

Clinical data management system and applications

The ATRI EDC system is a general CDMS package that serves as a framework for managing independent study-specific CDMS instances (study-CDMS) (Figure 2). In this approach, each study-CDMS is customized using study-specific configurations, data types, workflows, and functionality. This approach ensures strict isolation be-

tween study-CDMSs while still allowing the use of shared IT infrastructure. Furthermore, this approach supports the implementation of good clinical data management practices, such as the use of distinct study-specific production and development study-CDMSs to support well-controlled data management processes.

Study team-centered management

The ATRI EDC system supports a study team-centered approach to the configuration and administration of study-CDMSs. In this approach, study teams use web-based administrative user interfaces to manage the system's features and functionality. This approach empowers study teams and minimizes the need for IT team involvement. The use of these interfaces requires no programming skills. Configuration metadata may also be imported using spreadsheets. These spreadsheets support collaborative configuration efforts.

Study portal and modules

A study-CDMS's functionality, features, and data are available via a secure study-specific web portal. Portals are assigned study-specific web addresses (URLs) to facilitate user access. Within the portal, users can easily find modules that support various study activities such as participant registration, eCRF capture, query resolution, study document retrieval, and neuroimaging study upload (Figure 2). Access rights to each module are governed by a user's role in the study.

Data model and objects

The ATRI EDC system's data model combines a data ledger and a metadata database that provide a comprehensive, immutable, and cryptographically verifiable audit trail (Figure 3). Layered atop the data ledger, the metadata model defines an elementary set of primary data types: study, participant, event, field, form, file. The primary data types may be combined or extended to define derived composite data types. Data type definitions offer a mechanism to describe data in rich detail and may include validation rules, mappings to external ontologies, and relationships with other data types. The system uses these data type definitions during data collection activities to create, correlate, and validate data object instances (DOIs) in the data ledger. Each DOI represents a unique, fully characterized, version-controlled, cryptographically signed data value. A data record, which itself is a DOI, is composed of a prespecified set of data item DOIs. The DOI construct serves multiple purposes: identification, attribution, authentication, correlation, and verification. At any point in time, the authenticity and integrity of a DOI can be verified by reviewing the ATRI EDC system's audit trail. This data model design offers the necessary rigor and flexibility to manage the large heterogeneous data types common in AD/ADRD clinical research in accordance with regulatory requirements.

Data workflows

DOIs are processed using prespecified workflows. Workflows are defined using an ordered set of elementary actions: insert, edit, delete, annotate, link, copy, and execute. More complex workflows support conditional logic and parallel or asynchronous actions. The state of each DOI is updated and logged as it makes its way through a given workflow. The execute action merits special mention, as it may be used to define workflows that incorporate functionality conducted by external systems. This capability allows the ATRI EDC system to coordinate complex study-specific activities across multiple systems.

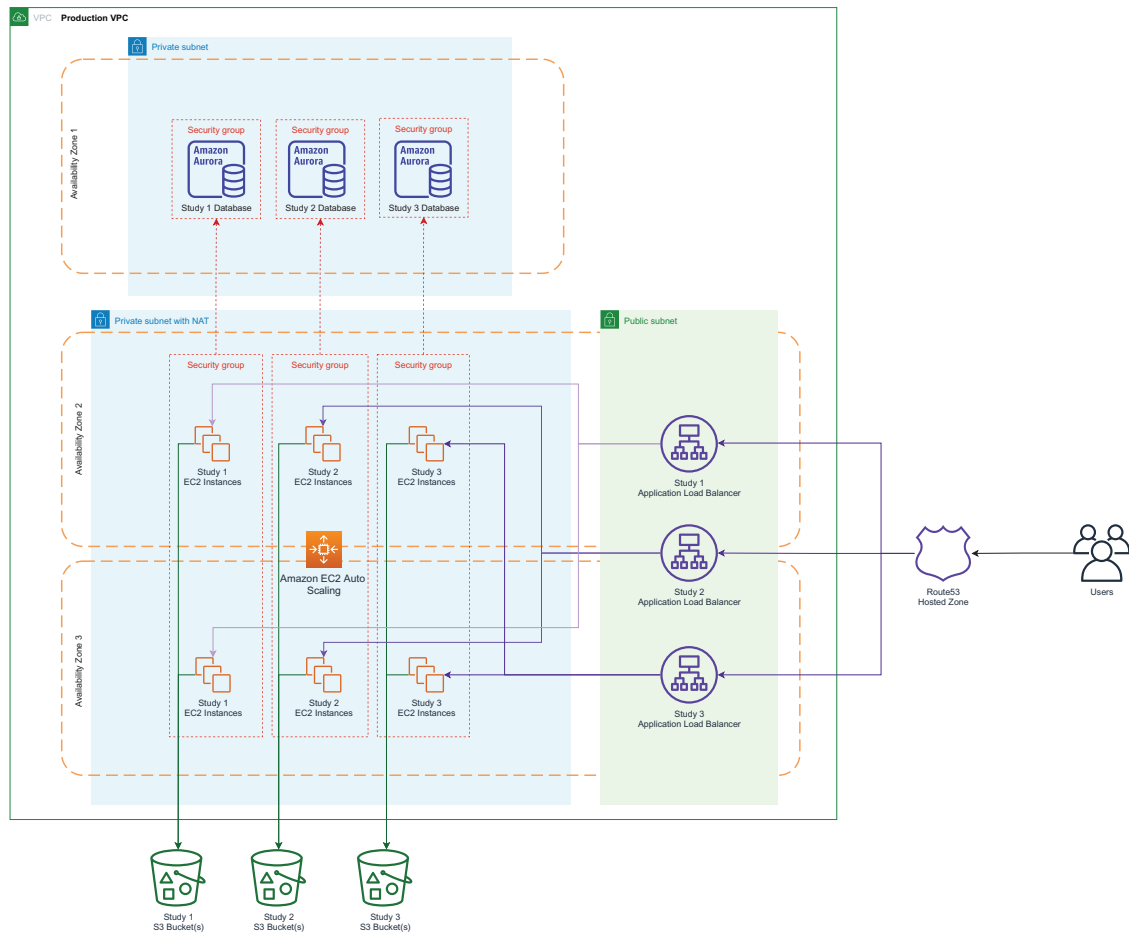


Figure 1. ATRI EDC AWS architecture. This cloud-based IT architecture assures strict isolation of configurations, data, and study-specific code between studies. It is implemented using a multi-account strategy that compartmentalizes core functions (production, staging and development workloads, audit, analytics, and logs) to improve security by minimizing the impact of a potential breach. Security groups and policies enforce network security, strong encryption at rest and in transit, and multifactor authentication for administrative accounts. High durability and availability are ensured via multi-region data replication and auto-scaling application groups. IT infrastructure is managed programmatically to minimize the possibility of human error in configuration management. VPC: virtual private network

Core modules

Electronic remote data capture of eCRFs

The EDC module serves as the primary interface for remote study data collection and implements user role-, eCRF-, and site-based permissions. A novel user interface design supports fluid and efficient navigation, review, and input of participant records by seamlessly combining eCRF, query, and audit trail information.

EDC use by trial sites

Site personnel use the EDC module to collect participant eCRF data, upload source documents, and submit neuroimaging studies at predefined study event timepoints. The Participant Event Matrix interface displays the status of all captured, queried, and missing data for participants at their site (Figure 4). Newly captured data are validated against a battery of edit checks and quality control rules; errors are immediately reported to the user for correction (Figure 5). Unresolved errors automatically lead to the generation of queries to facilitate tracking and resolution. All transactions on study data are recorded in the audit trail and may also trigger notifications to the study team or asynchronous data processing pipelines.

EDC use by study teams

Study teams use the EDC module to conduct data management and monitoring activities across all study sites. The EDC module supports the creation and resolution of item- and eCRF-level data queries, data and audit trail review, data locking, and source document verification.

Shared use of the EDC module facilitates communication and collaboration

The shared view of a participant's comprehensive data record, which integrates eCRF, query, and audit trail information, facilitates efficient and effective communication and collaboration between the study team and trial sites.

Metadata-driven EDC and eCRF configuration

Metadata drives the EDC module's functionality and features. In the case of an eCRF, this includes item names, prompts, edit checks, quality rules, and skip patterns (Table 2). New metadata is validated using a comprehensive set of quality and consistency checks. Once validated, the new metadata updates the EDC module's web-based views, functionality, and eCRF configurations. These configurations

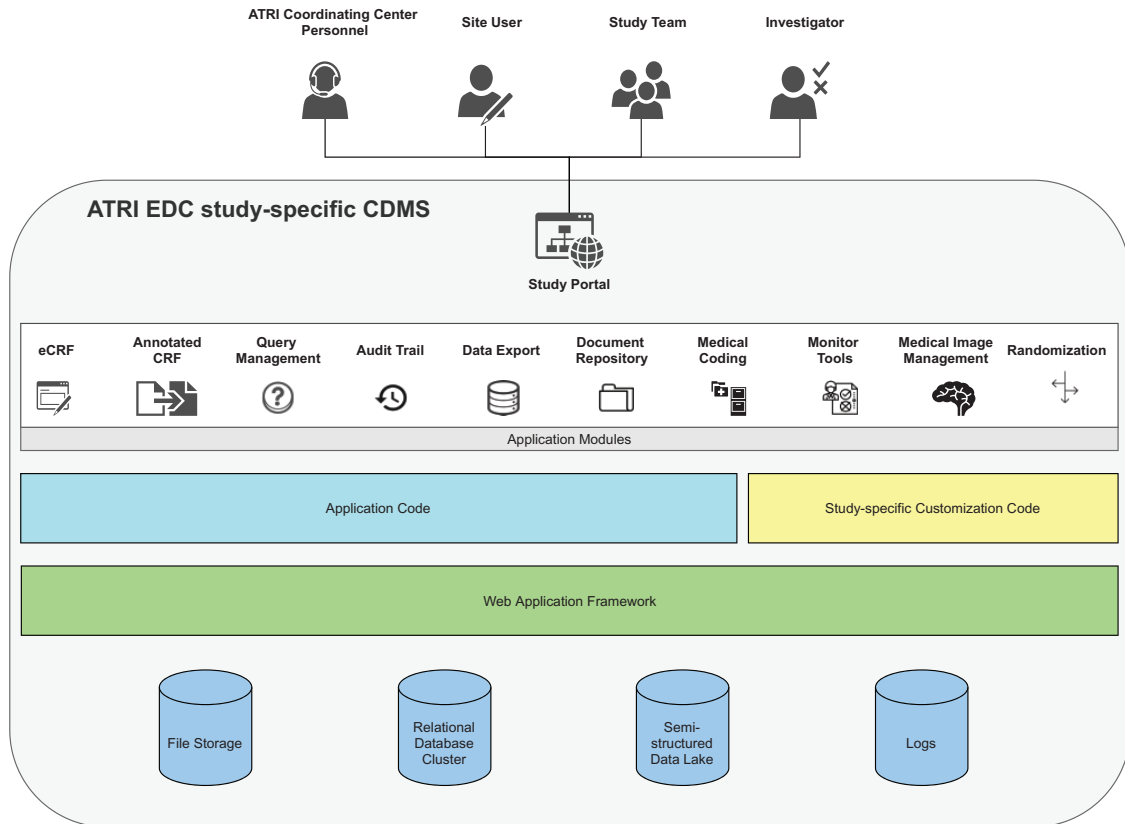


Figure 2. ATRI EDC study-CDMS architecture. Each study-specific CDMS is built using a dedicated set of application components hosted on shared cloud-based IT infrastructure. This design guarantees strict isolation of configurations, data, and study-specific code between studies. It also allows IT resources (compute, service, storage) to be optimized to address study-specific regulatory, operational, data, and fiscal requirements. From the user perspective, this design offers authorized users a single point of access to multiple study-specific application modules. Access to these modules is governed by role-based permissions. This approach facilitates communication and collaboration in the accurate and timely completion of study activities.

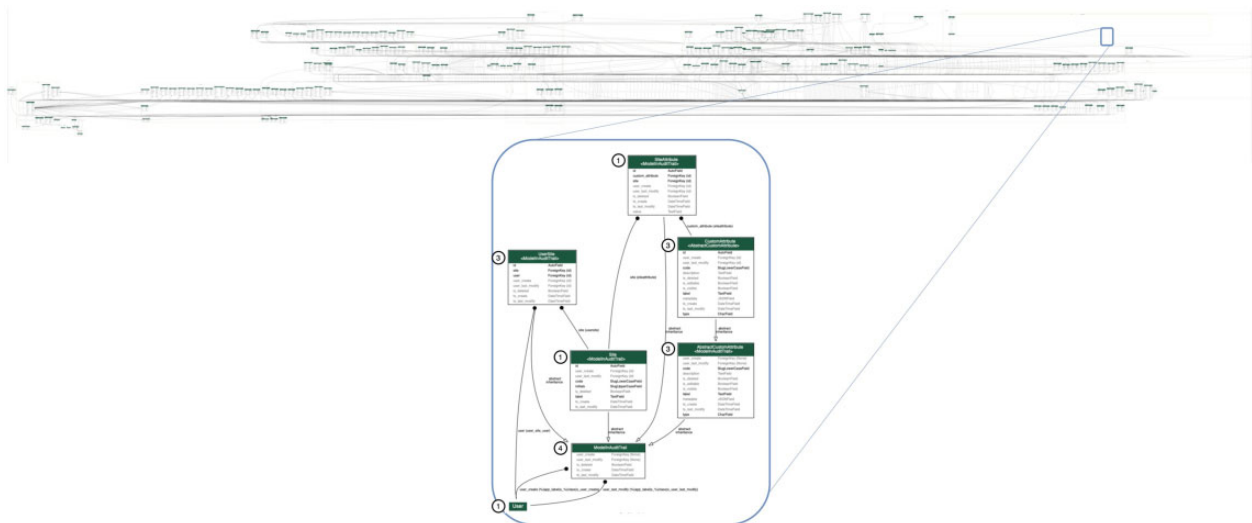


Figure 3. ATRI EDC entity relationship diagram (ERD). The ATRI EDC ERD visualizes the system's data model. The data model contains over 100 entities that serve as the foundation for the system's database. Each entity may be classified into 1 of 4 types: (1) data ledger, (2) data, (3) metadata, and (4) audit trail or history. Entities may also be divided into core and application-specific entities. To further illustrate this design, the inset diagram presents the ERD for a specific application. This application-level ERD contains 3 entity types and demonstrates the relationships between them.

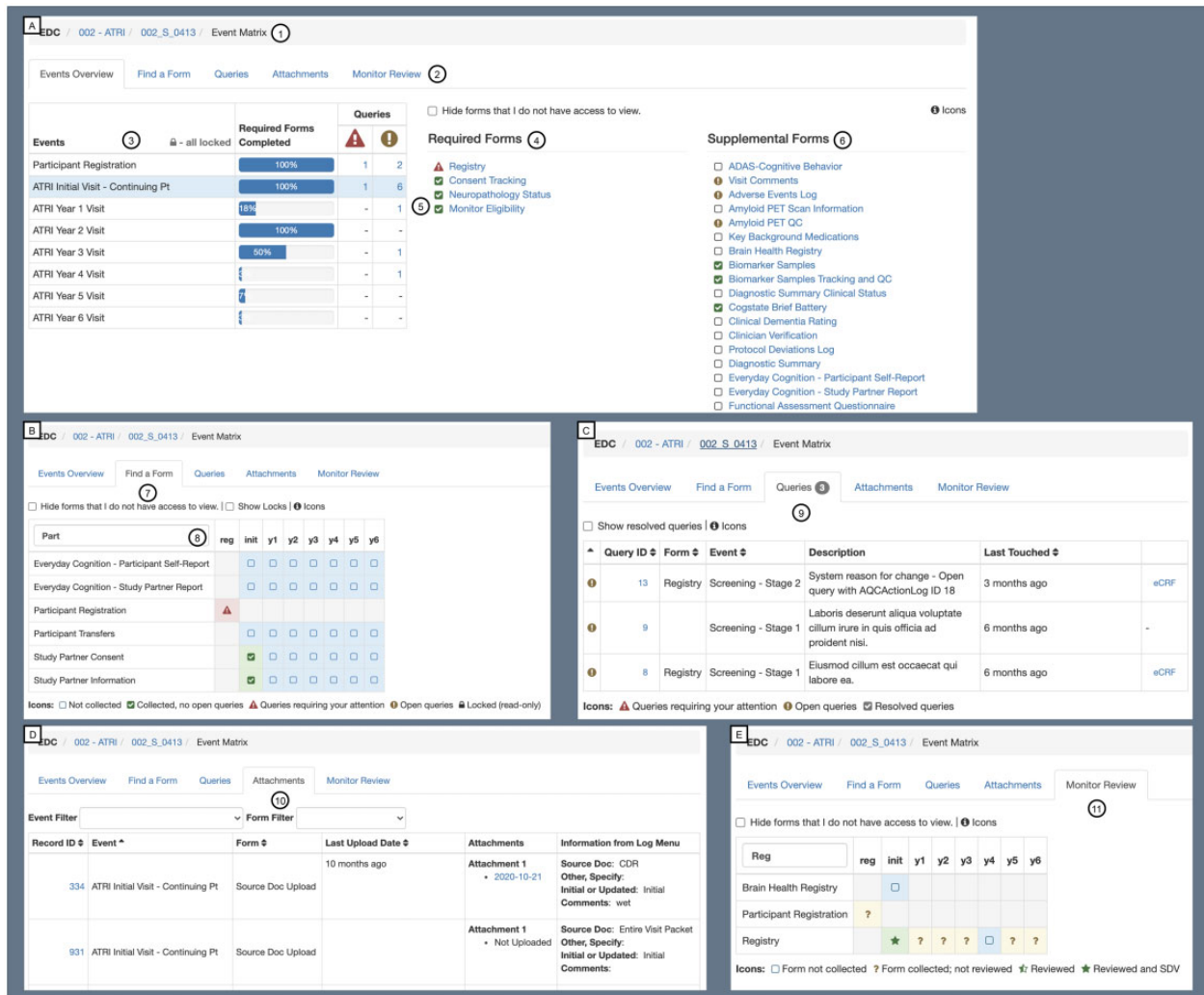


Figure 4. ATRI EDC Participant Event Matrix interface. This interface offers a unified overview of a participant’s study record by integrating information from multiple sources—eCRFs, queries, file uploads, data locking, clinical monitor review, and source document verification (SDV). Panel A demonstrates the “Events Overview” tab view, which includes the following features: (1) case navigation and orientation, (2) interface tab selector, (3) event dashboard with study event lock status, eCRF completion progress meter, and unresolved query counts, (4) list of required eCRFs, (5) eCRF completion status, and (6) list of supplemental eCRFs and completion status. Panel B demonstrates the “Find a Form” tab view, which includes the following features: (7) eCRF completion status per study event dashboard, and (8) eCRF name search. Panel C demonstrates the “Queries” tab view, which features the unresolved query dashboard (9). Panel D demonstrates the “Attachments” tab view, which features the file uploads dashboard (10). Panel E demonstrates the “Monitor Review” tab view, which features the eCRF review status per study event dashboard (11).

are managed by the system’s metadata model, using a version and revision control scheme, and applied prospectively, maintaining the integrity of previously captured data. This metadata-driven approach allows study teams to implement intrastudy protocol amendments in a well-controlled manner.

Annotated eCRFs

The annotated eCRFs module offers an additional mechanism to support collaborative study design and development. This module allows study teams to generate a highly detailed, printable representation of each version and revision of an eCRF annotated with configuration information such as item name, prompt, coding, edit checks, and skip patterns (Table 2). These documents facilitate effective and efficient multistakeholder eCRF design, review, and validation processes.

are managed by the system’s metadata model, using a version and revision control scheme, and applied prospectively, maintaining the integrity of previously captured data. This metadata-driven approach allows study teams to implement intrastudy protocol amendments in a well-controlled manner.

Query management

The Query module is designed to promote transparency and shared accountability between study teams and sites for the timely resolution of data quality issues. It is capable of functioning as a stand-alone module or embedded within other modules. It implements a novel conversational user interface that supports query-related communication and collaboration. These interactions are recorded in the system’s audit trail and can be easily referenced and analyzed to identify personnel training issues or eCRF improvements. Query status changes trigger role-based alerts and notifications. Queries can

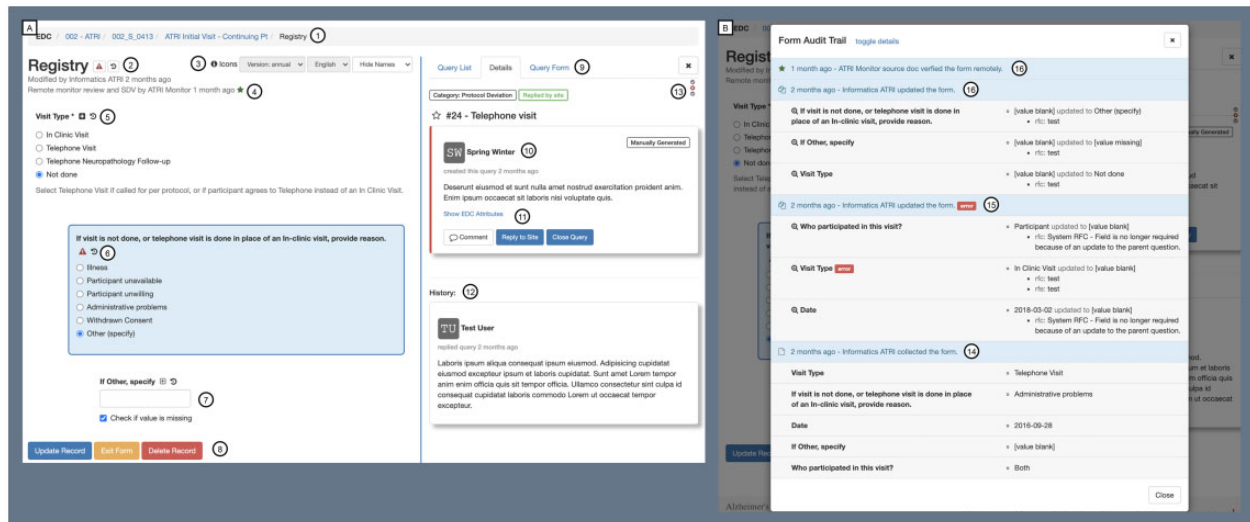


Figure 5. ATRI EDC integrated eCRF-Query-Audit Trail interface. This novel interface integrates multiple data sources—eCRF, source document verification (SDV), recording locking, eCRF- and item-level queries, eCRF- and item-level audit trail, and eCRF version and revision information. It supports accurate and timely communication and collaboration between study teams and sites by offering a shared view of these data sources and a conversational interface for data query creation and resolution. Panel A demonstrates the primary eCRF interface, which includes the following features: (1) case navigation and orientation, (2) 1-click eCRF-level query and audit trail review, (3) eCRF version and language settings, (4) dates of last modification and SDV, (5) various input types with accompanying prompt, units, notes, and 1-click item-level query and audit trail review, (6) real-time item-level error feedback with query integration, (7) item indenting, missing data check box, and skip patterns, (8) color-coded action buttons, (9) integrated query listing, details, and creation, (10) conversational query interface, (11) query comments, response, and closure, (12) query conversation history, and (13) field navigation, orientation, and query status. Panel B demonstrates the integrated Audit Trail interface, which includes the following features: (14) summary and detailed view of eCRF insert, (15) summary and detailed view of eCRF update with item-level change summary and query status, (16) eCRF update with resolved query status, and (17) SDV summary.

also be annotated and categorized to support further analysis and reporting. A suite of reports provides query performance metrics that study teams can use to manage data quality throughout a study.

Audit trail

Consistent with regulatory requirements, every transaction (insert, update, delete, and upload) performed on a study-CDMS's study metadata and data is recorded in the system's audit trail. The Audit trail module integrates information from multiple sources (metadata, item data, reason for change [RFC], file version, query data, data locks, source document verification) to present a comprehensive view of the factors associated with study data changes over time.

Document management

The Document Repository module provides a scalable, version-controlled mechanism to manage file-based data such as documents, videos, spreadsheets. It allows users to access, preview, upload, download, and annotate file-based data. The module implements a file-system interface and supports the organization of files by topic, a construct like a single-level file-system folder. Users may subscribe to receive notifications when new files are uploaded to or deleted from a specific topic. User preview and download activity in this module is recorded in the module's usage log.

Topic types

The Document Repository module supports multiple topic types, each with its associated metadata, configurations, and workflows. Standard topics offer file version control, annotations, and RFC features. Site topics have identical functionality, but automatically create site-specific sub-topics as new trial sites to the study-CDMS. In addition to

role- and topic-based permissions, these site-specific sub-topics implement site-based permissions. Additional topic types are available to support data archival (read-only) and large data transfers (read-write).

Data transfers

The Document Repository module's functionality—access control, version control, file retention, annotations, RFC, notifications, and logging—makes it a robust mechanism to facilitate inbound and outbound data transfers. To further improve support for this use case, file naming and content validation rules can be configured and enforced at the point of upload.

Data export

The Data Export module provides a standardized mechanism to export study metadata and data. The module supports export to multiple formats: comma-separated (2-dimensional) and JSON (multi-dimensional). As of August 2021, export to additional formats, such as PDF, R, and CDISC ODM, is under development.

Data lake

As previously noted, AD/ARD studies collect large heterogeneous databases from multiple sources. Loading external data into the study-CDMS using traditional ETL methods is labor-intensive and fragile. At scale, providing users with real-time exports of study data from a study-CDMS is inefficient. One approach to address these issues involves importing all study data regularly to a semi-structured data repository, also referred to as a "data lake."²⁶ This approach bypasses the need to upload data into a structured data store, such as a relational database or data warehouse, while still providing indexing and standardization to improve discoverability and usability. The Docu-

Table 2. ATRI EDC eCRF configuration metadata dictionary

Class label	Attribute label	MLS	Description
CRF metadata	CRF_NAME	—	Short name of the eCRF.
	CRF_LABEL	—	Long name of the eCRF.
	ALLOW_MULTIPLE_ENTRY	—	Allow multiple records of an eCRF to be collected at a single study visit.
	IS_CROSS_EVENTS	—	Allow multi-entry eCRFs to span multiple study visits; used for concurrent medications and adverse event logs.
	ALLOW_CHANGE_VERSION	—	Allow users to choose which version of an eCRF they will collect, otherwise, eCRF versions are selected by the system.
Version metadata	LOG_MENU_FIELD	—	Specifies the set of fields and labels that will be presented in the multi-entry eCRFs log view.
	V_VERSION	—	eCRF version name; multiple versions of the same eCRF can be defined.
Revision metadata	FORM_SUBTITLE	Yes	eCRF subtitle.
	FORM_INSTRUCTION	Yes	eCRF instructions, below subtitle.
Fieldset metadata	TITLE	Yes	Field section title.
	DESCRIPTION	Yes	Field section description.
Field metadata	FIELD_NAME	—	Field short name.
	QUESTION	Yes	Field prompt.
	ORDER	—	Field order number.
	UNITS	Yes	Field units (not required).
	NOTES	Yes	Filed notes (not required); supports Markdown.
	WIDGET	—	Field input type: select; multiCheckbox; input; textarea; radio; radioHorizontal; pin; datePickerStrict; datePicker; timepicker24. System allows key/label pairs to be specified for input types that support multiple choice options.
	TYPE	—	Field type: (T) Text; (N) Numeric; (D) Date.
	VALUE	—	Range for a numeric field: <i>[min value]..[max value]:[step]</i>
	REQUIRED	—	Indicate if the field is required.
	ALLOW_MISSING	—	Indicate if the field is allowed to be collected as “missing”; a “missing” checkbox input will be appended to this field in the eCRF.
	LENGTH	—	Length of the field: <i>[number, decimal]</i>
	PHI	—	Indicate if the field may contain PHI/PII; will be masked or removed from data exports by default.
	INDENT	—	Field indentation from the left margin of the eCRF.
	SUBQUESTION	—	Specify conditional field visibility rules and skip patterns.
DEFAULT_VALUE	—	Specify a default value for the field.	
DOUBLE_ENTRY	—	Activate field-level double entry.	
File metadata	NAME	—	File attachment: File name.
	LABEL	Yes	File attachment: File label.
	RESTRICTIONS	—	File attachment: File extension restriction. By default, no restrictions are applied.

Note: Attribute templates are indicated in italics.

Abbreviation: MLS: multi-language support.

ment Repository module can be configured to serve as an interface to a study-specific data lake to facilitate user access.

Standard reports

The Reports module provides a standard suite of reports that are available on all CDMS instances. These reports summarize study operations, data inventories, data quality, and data flows.

Additional modules

The ATRI EDC system provides several additional modules that support the effective and efficient conduct of AD/ADR clinical trials (Figure 6). These modules include randomization, medical image management, adverse event coding, safety reporting, and clinical monitor visit reporting. A complete description of these modules is

beyond the scope of this article; detailed descriptions are planned for future articles.

Integration with external systems

The ATRI EDC system provides a comprehensive set of application programming interfaces (APIs) to support integration with external systems. Examples of these integrations include automating inbound and outbound data transfers with sponsors and vendors, ordering the shipment of investigational product at the point of participant randomization, and registering study participants in an external image management system.

RESULTS

As of August 2021, the ATRI EDC system has supported the conduct of 10 AD/ADR multicenter clinical studies (Table 3).

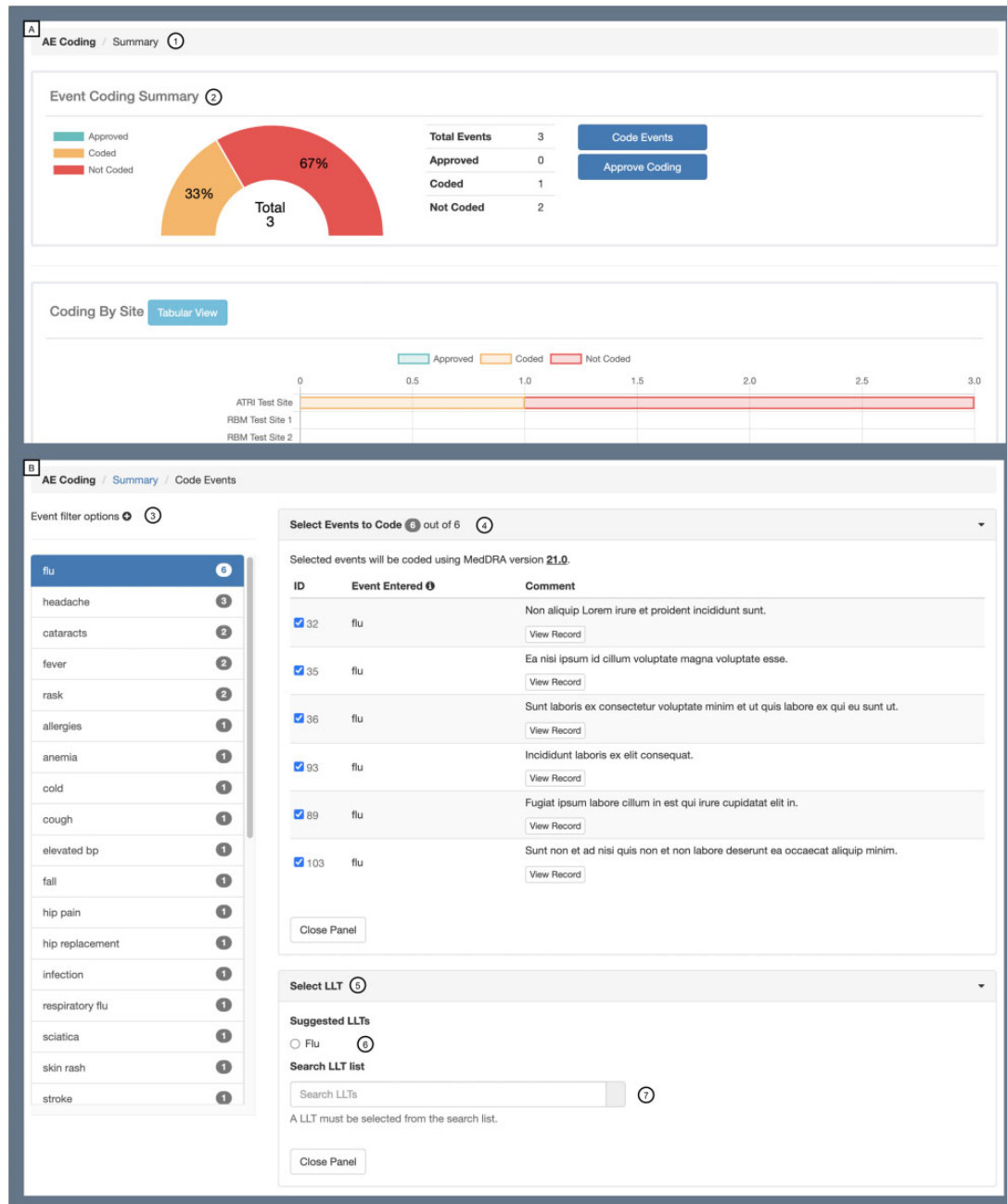


Figure 6. ATRI EDC additional modules example: Adverse Event Coding module. This module supports the coding of adverse event (AE) data. It leverages a code suggestion feature powered by a text classification API that combines natural language processing (NLP) and machine learning methods.³⁰ Panel A demonstrates the Event Coding Summary view, which includes the following features: (1) navigation and orientation and (2) event coding summary dashboard. Panel B demonstrates the Event Coding interface, which includes the following features: (3) event filtering and grouping, (4) event coding selection, and (5) event code suggestion and search. LLT: MedDRA lowest level term.

Each study database was developed collaboratively by ATRI study teams in 3–6 months. Table 4 summarizes the modules and integrations with external systems that support each study. To date, the ATRI EDC has been used to capture data for 4596 study participants from trial sites in Asia, Australia, Europe, and North America. For these participants, more than 330 000 eCRFs have been collected. Among these studies, the 3 cases described below illustrate the ATRI EDC system's flexibility and range of capabilities.

Case: ADNI 3 study

The Alzheimer's Disease Neuroimaging Initiative (ClinicalTrials.gov ID: NCT02854033) (<http://adni.loni.usc.edu>) (Table 3),²⁷ a large multicenter prospective observational cohort study that seeks to enroll 2000 participants across various AD diagnostic groups (Cognitively Normal, Mild Cognitive Impairment, and Mild AD Dementia). The study is funded by a public–private partnership between the NIH and several private companies. In this study, the ATRI EDC system serves several purposes: (1) the EDC, Query, and

Table 3. Multicenter AD/ADRD clinical studies supported by the ATRI EDC system (as of August 3, 2021)

Study	Study type	Start year	Users	Sites	Participants screened	eCRFs		Queries		Document repository			Image studies
						N	Transactions	N	Rate (%)	Files	File versions		
ABC-DS	OC	2021	22	8	—	—	—	—	NAN	19	19	—	
ADNI3	OC	2016	836	63	1802	136 876	152 773	11 644	8.5	1217	135 797	—	
AHEAD3-45	RCT	2020	1065	108	1147	63 200	69 047	9275	14.7	4161	14 333	449	
COGRX	RCT	2022	—	—	—	—	—	—	—	—	—	—	
LEADS	OC	2018	311	19	415	28 843	34 497	4919	17.1	539	11 721	—	
LIBBY	RCT	2022	12	—	—	—	—	—	—	—	—	—	
MIND	RCT	2017	443	51	623	61 656	72 588	7361	11.9	446	34 040	—	
NiAD	OC	2016	102	5	248	20 623	24 391	2601	12.6	11	4186	—	
TRC-PAD	OC	2019	598	51	360	26 592	28 545	1086	4.1	486	13 740	252	
TRC-DS	OC	2021	149	17	1	26	27	1	3.8	410	419	2	
Total			3526	322	4596	337 816	381 868	36 887		7289	214 255	703	

Note: Some sites participate in multiple studies. AD/ADRD RCTs typically experience 50%-80% screen failure rates; participant enrollment is ongoing in all studies. Two new RCTs, COGRX and LIBBY, are currently under development and scheduled to start in 2022.

Abbreviations: ABC-DS: Alzheimer Biomarker Consortium—Down Syndrome; ADNI3: Alzheimer’s Disease Neuroimaging Initiative 3 (ADNI3) Protocol; AHEAD3-45: AHEAD 3-45 Study: A Study to Evaluate Efficacy and Safety of Treatment With Lecanemab in Participants With Preclinical Alzheimer’s Disease and Elevated Amyloid and Also in Participants With Early Preclinical Alzheimer’s Disease and Intermediate Amyloid; COGRX: Randomized Double Blind, Placebo Controlled, Parallel Group Trial to Evaluate the Safety and Efficacy of CTT1812 in Early Alzheimer’s Disease over 18 Months; LEADS: Longitudinal Early-onset Alzheimer’s Disease Study Protocol; LIBBY: Life’s end Benefits of Cannabidiol and Tetrahydrocannabinol (LIBBY) Trial; MIND: Memory Improvement Through Nicotine Dosing (MIND) Study; NIAD: Neurodegeneration in Aging Down Syndrome (NIAD): A Longitudinal Study of Cognition and Biomarkers of Alzheimer’s Disease; TRC-PAD: TRC-PAD Program: In-Clinic Trial-Ready Cohort; TRC-DS: Trial-Ready Cohort-Down Syndrome; OC: observational cohort; RCT: randomized clinical trial.

Table 4. ATRI EDC modules and external integrations supporting AD/ADRD studies (as of August 3, 2021)

Study	Study type	Standard modules	Clinical monitor visit reporting	AE coding	Image management	Safety reporting	External integrations
ABC-DS	OC	X					Imaging
ADNI3	OC	X		X			Imaging, RBM, Labs
AHEAD3-45	RCT	X	P		X		IRT, ECOA, EPRO, RBM, Labs, ECG
COGRX	RCT	X	P	X	X	P	IRT, RBM, Labs
LEADS	OC	X		X			
LIBBY	RCT	X	P	X		P	IRT, RBM, Labs
MIND	RCT	X	X	X		X	IRT, RBM, Labs
NiAD	OC	X					
TRC-PAD	OC	X	P	X	X	P	RBM, Labs
TRC-DS	OC	X	P	X	X	P	RBM, Labs

Note: Standard modules are used in all studies and include eCRFs, Annotated CRFs, Query Management, Audit Trail, Document Management, Data Export, Standard Reporting, and Data Lake.

Abbreviations: ABC-DS: Alzheimer Biomarker Consortium—Down Syndrome; ADNI3: Alzheimer’s Disease Neuroimaging Initiative 3 (ADNI3) Protocol; AHEAD3-45: AHEAD 3-45 Study: A Study to Evaluate Efficacy and Safety of Treatment With Lecanemab in Participants With Preclinical Alzheimer’s Disease and Elevated Amyloid and Also in Participants With Early Preclinical Alzheimer’s Disease and Intermediate Amyloid; COGRX: Randomized Double Blind, Placebo Controlled, Parallel Group Trial to Evaluate the Safety and Efficacy of CT1812 in Early Alzheimer’s Disease over 18 Months; LEADS: Longitudinal Early-onset Alzheimer’s Disease Study Protocol; LIBBY: Life’s end Benefits of Cannabidiol and Tetrahydrocannabinol (LiBBY) Trial; MIND: Memory Improvement Through Nicotine Dosing (MIND) Study; NIAD: Neurodegeneration in Aging Down Syndrome (NiAD): A Longitudinal Study of Cognition and Biomarkers of Alzheimer’s Disease; TRC-PAD: TRC-PAD Program: In-Clinic Trial-Ready Cohort; TRC-DS: Trial-Ready Cohort-Down Syndrome; OC: observational cohort; RCT: randomized clinical trial; P: module activation planned; X: module active; ECG: electrocardiogram; ECOA: electronic clinical outcome assessment; EPRO: electronic patient-reported outcomes; IRT: interactive response technology for randomization and study supplies; RBM: risk-based monitoring.

Audit Trail modules are used to collect and manage the clinical database; (2) the Document Repository module supports study document management, inbound data transfers from multiple external biomarker labs, and large data exports to support analysis and statistical reporting; (3) the AE Coding module supports coding of adverse events; and (4) the system’s APIs integrate the study-CDMS with a data-sharing platform hosted by the USC Laboratory of Neuro Imaging (<https://www.loni.usc.edu>).

Case: TRC-PAD study

The Trial Ready Cohort for the Prevention of Alzheimer’s Dementia (TRC-PAD) study (ClinicalTrials.gov ID: NCT04004767) (<https://trcpad.org>) (Table 3),⁹ a large multicenter prospective observational cohort study that seeks to enroll 2000 participants with increased risk of memory loss caused by AD. The study is funded by the NIH. In this study, the ATRI EDC system serves several purposes: (1) the EDC, Query, and Audit Trail modules are used to collect and manage the clinical database; (2) the Document Repository module supports study document management, inbound data transfers from multiple external biomarker labs, and large data exports to support analysis and statistical reporting; (3) the Imaging module supports the acquisition and management of PET neuroimaging studies; (4) the AE Coding module supports coding of adverse events; and (5) the system’s APIs integrate the study-CDMS with the TRC-PAD Informatics Platform and the APT Webstudy (<https://www.aptwebstudy.org>).²⁸

Case: AHEAD3-45 study

The AHEAD3-45 study (ClinicalTrials.gov ID: NCT04468659) (<https://www.aheadstudy.org>) (Table 3), a large multicenter double-blinded randomized clinical trial (RCT) that seeks to enroll 1400 participants to determine the efficacy of treatment with Lecanemab (BAN2401) relative to placebo in 2 diagnostic groups (Preclinical AD and Early Preclinical AD). The study is funded by a public-private

partnership between the NIH and Eisai, Inc. The study is conducted via a global network of 108 sites in Asia, Australia, Europe, and North America. In this study, the ATRI EDC system serves several purposes: (1) the EDC, Query, and Audit Trail modules are used to collect and manage the clinical database; (2) the Document Repository module supports study document management, inbound and outbound data transfers with multiple external biomarker labs and vendors (ePRO, eCOA, IRT, RBM), and large data exports to support analysis and statistical reporting; (3) the Imaging module supports the acquisition and management of MRI neuroimaging studies; and (4) the system’s APIs integrate the study-CDMS with external pharmacovigilance, safety monitoring, and reporting systems.

DISCUSSION

Evolving requirements and technological innovation

The ATRI EDC system was developed to address the complex and demanding set of requirements of AD/ADRD clinical trials. As described in previous sections, the system’s flexibility and capabilities in supporting a wide array of complex study designs and operational implementations confirm that these requirements have been met. However, the pace of innovation in AD/ADRD clinical research continues to accelerate, requiring further development of the system. In parallel, rapid shifts in technology and computing offer opportunities to address existing and emerging study requirements using novel methods. Incorporating capabilities based on machine learning, microservices, and serverless computing into the ATRI EDC system that will increase data quality, improve productivity, and reduce the burden on study teams and sites are now being planned.

Open science

Consistent with the principles of open science, the ATRI EDC system is intended to be shared with the scientific community. Inter-

ested investigators can obtain system source code and documentation for evaluation and use via a technology transfer agreement by submitting a request via email to: softwarelicensing@atrihub.io. The ATRI Informatics team hosts regular community-based webinars with investigators and their study teams to discuss the system roadmap, preview updates, and gather feedback. As an example, investigators from the University of Tokyo are using the ATRI EDC to support the J-TRC study,²⁹ a multicenter prospective observational cohort study in preclinical and prodromal AD.

Looking beyond AD/ADRD clinical studies

More generally, the ATRI EDC system has several advantages to offer other groups that are coordinating trials. Its flexibility and range of capabilities make it a suitable option for studies in other fields of medical research.

Limitations

The ATRI EDC system has several limitations that must be acknowledged. First, the system was designed to address the requirements of large multicenter clinical studies, where the costs of commercial options are substantial and the capabilities of academic or community-based systems are incomplete. The advantages of this design, however, are not generalizable to all study types. A small single-center study, for example, may be better served by an academic or community-based EDC solution, which is simpler to use and more cost-effective. To address these limitations, we plan to develop a new “quick start” module with a simplified user interface that will guide study teams through the study configuration process in a stepwise manner, using a series of prompts to select appropriate system settings. We also plan to enhance our user and system documentation and training materials.

Second, despite efforts toward simplification, the technical expertise and training necessary to effectively administrate the system are significant and may represent a barrier to further adoption by other research organizations and teams. This may slow community-building efforts and innovation. To address this limitation, we plan to offer a containerized version of the system that will reduce the administrative requirements and facilitate hosting in a wider array of computational environments.

Finally, maintaining strict isolation of the system’s application code and data is achieved at the expense of higher resource utilization and storage density. This approach is less cost-effective and has a higher environmental impact.

CONCLUSION

In conclusion, AD/ADRD clinical trials present a complex set of scientific, operational, and regulatory challenges that require innovative technology and team-based science. ATRI designs and conducts large multicenter AD/ADRD clinical studies across a global network of sites by building multidisciplinary study teams of clinical trialists and leveraging the increasing functionality and capabilities of the ATRI EDC remote data capture system. The combination of human expertise and novel technologies facilitates ATRI’s mission to conduct rigorous and efficient therapeutic AD/ADRD trials.

AUTHOR CONTRIBUTIONS

GJM conceived and led the project. GJM, SB, HQ, and JS designed and implemented the project. GJM, SB, HQ, and JS participated

equally in the writing of the manuscript. PSA provided scientific and medical expertise. All authors provided feedback and reviewed and approved the final version of the manuscript.

ACKNOWLEDGMENTS

Amazon Web Services, the “Powered by AWS” logo, EC2, AWS Batch, AWS Lambda, Amazon Elastic Beanstalk, AWS Elastic Container Service, Amazon Simple Email Service, DynamoDB, Amazon RDS, Amazon Sagemaker, Amazon CodeGuru, AWS CodePipeline, AWS CodeBuild, Amazon CloudWatch, AWS CloudTrail, AWS Config, AWS CloudFormation, Amazon Macie, AWS System Manager, AWS Trusted Advisor, Amazon VPC, Amazon CloudFront, AWS API Gateway, Amazon Route 53, AWS Certificate Manager, AWS IAM, AWS GuardDuty, AWS Secrets Manager, Amazon S3, AWS Step Functions, Amazon Simple Queue Service (SQS), AWS Key Management Service, and Amazon Simple Notification Service are trademarks of Amazon.com, Inc. or its affiliates in the United States and/or other countries. AWS architecture icons (<https://aws.amazon.com/architecture/icons/>) used with permission from Amazon Web Services.

FUNDING

The development of this project was supported by infrastructure funding received from National Institute on Aging (NIA) grant # 3U24AG057437-03S1 (Principal Investigator: PSA).

CONFLICT OF INTEREST STATEMENT

The authors report grants from National Institute on Aging, during the development of this manuscript. None of the authors have additional financial interests, relationships, or affiliations relevant to the subject of this manuscript.

DATA AVAILABILITY

No new data were generated or analyzed in support of this research.

REFERENCES

- 2020 Alzheimer’s disease facts and figures. *Alzheimers Dement* 2020; 2020: 12068.
- Cummings J, Lee G, Ritter A, *et al.* Alzheimer’s disease drug development pipeline: 2020. *Alzheimers Dement (N Y)* 2020; 6 (1): e12050.
- Sperling RA, Rentz DM, Johnson KA, *et al.* The A4 study: stopping AD before symptoms begin? *Sci Transl Med* 2014; 6 (228): 228fs13.
- Kaye J, Aisen P, Amariglio R, *et al.* Using digital tools to advance Alzheimer’s drug trials during a pandemic: the EU/US CTAD Task Force. *J Prev Alz Dis* 2021; 2021: 1–7.
- The Adaptive Platform Trials Coalition. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat Rev Drug Discov* 2019; 18: 797–807.
- Aisen PS, Bateman RJ, Carrillo M, *et al.* Platform trials to expedite drug development in Alzheimer’s disease: a report from the EU/US CTAD Task Force. *J Prev Alzheimers Dis* 2021; 8 (3): 306–12.
- Watson JL, Ryan L, Silverberg N, *et al.* Obstacles and opportunities in Alzheimer’s clinical trial recruitment. *Health Aff (Millwood)* 2014; 33 (4): 574–9.
- Aisen P, Touchon J, Andrieu S, *et al.* Registries and cohorts to accelerate early phase Alzheimer’s trials. A report from the E.U./U.S. Clinical Trials in Alzheimer’s Disease Task Force. *J Prev Alzheimers Dis* 2016; 3 (2): 68–74.

9. Aisen PS, Sperling RA, Cummings J, *et al.* The Trial-Ready Cohort for Pre-clinical/Prodromal Alzheimer's Disease (TRC-PAD) project: an overview. *J Prev Alzheimers Dis* 2020; 7 (4): 208–12.
10. Das S, Zijdenbos AP, Harlap J, *et al.* LORIS: a web-based data management system for multi-center studies. *Front Neuroinform* 2011; 5: 37.
11. Harris PA, Taylor R, Thielke R, *et al.* Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; 42 (2): 377–81.
12. Nguyen L, Shah A, Harker M, *et al.* DADOS-Prospective: an open source application for web-based prospective data collection. *Source Code Biol Med* 2006; 1: 7.
13. Löbe M, Meineke F, Winter A. Scenarios for using openclinica in academic clinical trials. *Stud Health Technol Inform* 2019; 258: 211–5.
14. Rave electronic data capture (EDC) system. Medidata.com. 2021. <https://www.medidata.com/en/clinical-trial-products/clinical-data-management/edc-systems>. Accessed October 30, 2021.
15. Beck K, Beedle M, van Bennekum A, *et al.* Manifesto for Agile Software Development. <http://agilemanifesto.org/>. Accessed May 13, 2020.
16. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. *Comput Biol Med* 2002; 32 (3): 141–9.
17. Ambler T, Cloud N. *JavaScript Frameworks for Modern Web Dev*. New York, NY: Apress LP; 2015.
18. Amazon Web Services (AWS)—cloud computing services. <https://aws.amazon.com/>. Accessed September 3, 2020.
19. U.S. Food & Drug Administration. Guidance for industry—part 11, electronic records; electronic signatures: scope and application. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application>. Accessed August 9, 2021.
20. Bruschi S, Xiao L, Kavatkar M, *et al.* Behavior driven development (BDD): a case study in healthtech. Presented at: Pacific NW Software Quality Conference; 2019; Portland, Oregon, USA. <http://uploads.pnsqlc.org/2019/papers/Xiao%20-%20Behavior%20Driven%20Development%20A%20Case%20Study%20in%20Healthtech.pdf>. Accessed August 9, 2021.
21. ASPE—Office of the Assistant Secretary for Planning and Evaluation. Health Insurance Portability and Accountability Act of 1996. 1996. <https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996>. Accessed August 9, 2021.
22. Amazon Web Services. AWS well-architected framework. 2020. <https://docs.aws.amazon.com/wellarchitected/latest/framework/welcome.html>. Accessed August 9, 2021.
23. Carter K, Francois Raynaud on DevSecOps. *IEEE Softw* 2017; 34 (5): 93–6.
24. HITRUST Alliance. HITRUST Common Security Framework. <https://hitrustalliance.net/product-tool/hitrust-csf/>. Accessed August 9, 2021.
25. European Commission. Data protection. https://ec.europa.eu/info/law/law-topic/data-protection_en. Accessed August 9, 2021.
26. Sawadogo P, Darmont J. On data lake architectures and metadata management. *J Intell Inf Syst* 2021; 56 (1): 97–120.
27. Weiner MW, Veitch DP, Aisen PS, *et al.* The Alzheimer's Disease Neuroimaging Initiative 3: continued innovation for clinical trial improvement. *Alzheimers Dement* 2017; 13 (5): 561–71.
28. Jimenez-Maggiola GA, Bruschi S, Raman R, *et al.* TRC-PAD: accelerating recruitment of AD clinical trials through innovative information technology. *J Prev Alzheimers Dis* 2020; 7 (4): 226–33.
29. Sato K, Niimi Y, Ihara R, *et al.* Efficacy and cost-effectiveness of promotion methods to recruit participants to an online screening registry for Alzheimer disease prevention trials: observational study. *J Med Internet Res* 2021; 23 (7): e26284.
30. Ravindranath PA, Bruschi S, Ernstom K, *et al.* Machine learning in automated classification of adverse events in clinical studies of Alzheimer's disease. *Alzheimers Dement* 2017; 13 (7): P1256.