



Society for Clinical Data Management
DATA DRIVEN

A Topic Brief Series on eSource

Society for Clinical Data Management (SCDM)

eSource Implementation Consortium



Playbook 3: Practical advice for pharma

Version 1 (September 2023)

Authors

- Linda S. King, SCDM, Astellas
- MD Naqib Alam Ansari, AbbVie
- Magda Jaskowska, PhD, GSK
- Lauren McCabe, Pfizer
- Muzafar Mirza, Pfizer
- Joseph Angiolelli, Pfizer
- Peter Casteleyn, Janssen R&D
- Rakesh Maniar, Merck & Co., Inc.
- Nadir Ammour, DDS, Sanofi

Reviewer

- Patrick Nadolny, SCDM, Sanofi

Author notes:

SCDM eSource Implementation Consortium, Linda S. King, SCDM
Facilitator

No conflicts of interest

Correspondence relating to this article: Linda S. King
(linda.king@astellas.com)

Disclaimer: Not all the views expressed in this playbook may be those of the individual companies or entities for which the authors are employed or affiliated.

Acknowledgement: SCDM would like to thank Amanda Leweson from Discovery PR for her technical writing assistance in this manuscript.

Abstract

Clinical research is in the midst of a digital transformation, with the emergence of eSource data promising to accelerate drug development timelines, enhance patient centricity, and unlock previously unseen insights. While much has been written on the rationale for eSource approaches, practical advice on their implementation has been less widely available. As the world's leading advocate for the discipline of clinical data management, the Society for Clinical Data Management (SCDM) is in a unique position to fill this knowledge gap. To achieve this aim, the group has produced a series of podcasts in which leading experts from across the clinical research ecosystem share their case studies and practical advice on moving eSource from theory into practice. We then distilled their learnings into four playbooks, each from the standpoint of one of the main stakeholder groups: CROs and vendors, pharma, regulators, and academia/sites. This paper focuses on the pharma perspective.

Methodology

The eSource Implementation Consortium is publishing an eSource topic briefs series intended to serve as orientation guides on eSource which are contributing directly or indirectly to the evolution of Clinical Data Management (CDM) into Clinical Data Science (CDS). Due to the absence of a comprehensive and authoritative literature base regarding the wide implementation of eSource within the Drug Development industry, this content was gathered from industry leaders through an opinion-based methodology. As eSource implementation matures, and technology evolves, the Consortium anticipates that literature on this topic will blossom.

Podcast interviewees were selected for their eSource expertise according to SCDM Board recommendations and/or were members of the SCDM eSource Implementation Consortium. Efforts to reduce bias included using a standard set of questions, based on input from the SCDM eSource Implementation Consortium and conducting interviews with 17 contributors from four different perspectives. Contributors were asked to share their thoughts on barriers to eSource adoption and implementation from their personal experiences of the approach, and to provide case studies.

Post-podcast recording, the recordings were grouped into four perspectives: CROs and vendors, pharma, regulators, and academia/sites. The transcripts were reviewed to identify key themes, which were then summarized to form a narrative, playbook-style report. Podcast contributors were asked to review the drafted content to ensure their viewpoints had been represented faithfully.

Interviewees: *

Name	Job title / Organization	Sector
Jonathan Andrus	President and COO at CRIO, and SCDM Treasurer	Vendor/CRO
Alex Crawford	Director of Decentralized Clinical Trial Products, DCT Operations, ICON	Vendor/ CRO
Kristen Harnack	Director of Solutions Consulting, Castor	Vendor/CRO
MD Naqib Alam Ansari	Senior Manager, Clinical Data Strategy and Operations, AbbVie R&D	Pharma
Magda Jaskowska, PhD	Global Director/Leader Oncology, Data Strategy and Management, GSK	Pharma

Lauren McCabe	Associate Director, Clinical Data Science, Pfizer	Pharma
Muzafar Mirza	Senior Group Lead, Clinical Data Sciences, Pfizer	Pharma
Joseph Angiolelli	Director, Information Management/Clinical Trial Solutions, Pfizer	Pharma
Peter Casteleyn	Director, Data Collection Solutions-EHR, Janssen R&D	Pharma
Rakesh Maniar	Executive Director and Head of eClinical Technologies, Global Clinical Trials Operations– Global Data Management and Standards, Merck & Co., Inc.	Pharma
Nadir Ammour, DDS	Global Lead for external engagement, Transformation & Performance Office, Clinical Science & Operations/Development, Sanofi R&D	Pharma
Mitra Rocca	Senior Medical Informatician, FDA	Regulatory
Jeff Stein	President of Stamford Therapeutics Consortium	Sites/academia
Michael Buckley	Associate Director of Product Management, Clinical Research Informatics and Technology, Memorial Sloan Kettering Cancer Center	Sites/academia
Elena Christofides, MD	Owner, Endocrinology Research Associates	Sites/academia
Cory Ennis	Director of Information Technology- Engagement and Assistant Dean for Research Systems, Duke University School of Medicine’s Office of Academic Solutions and Information Systems	Sites/academia
Denise Snyder	Associate Dean for Clinical Research, Duke University School of Medicine, Office of Clinical Research	Sites/academia

* All interviewees consented to use of their quotes. All information included in this report has been reproduced with the permission of the interviewees and the SCDM.

Introduction

Clinical research is in the midst of a digital transformation, with the emergence of eSource data promising to accelerate drug development timelines, enhance patient centricity, enhance sponsor, and site efficiencies, and unlock previously unseen insights. eSource refers to the direct collection (entry or acquisition) of clinical data into an eSource system from site staff, clinical trial participants, or care givers. It can include direct from device, such as wearables or sensors, direct from clinical trial participants or clinician/site staff, such as eCOA, or direct from an electronic health record (EHR).¹ The approach reduces the need for source data verification (SDV), minimizing the need for transcription and providing real-time guidance on illogical or inconsistent data at the point of collection. If implemented correctly and in compliance with ICH-GCP, it can reduce site burden, boost patient centricity, and improve data quality.²

As the industry moves from the “why” to the “how” of eSource, however, it is clear that adoption can sometimes present just as many challenges as it does opportunities. The new paradigm often requires the integration of disparate data sets, using multiple technologies, and redesigning existing work and data flows, for example. While much has been written on the rationale for eSource approaches, practical advice on their implementation has been less widely

available.³ As the world’s leading advocate for the discipline of clinical data management, the Society for Clinical Data Management (SCDM) is in a unique position to fill this knowledge gap.

SCDM eSource Implementation Consortium

SCDM is one of several industry bodies backing the use of eSource, which offers a wide range of benefits. The consensus is that it can “improve protocol design and clinical trial participant recruitment, modernize, and streamline data collection, monitoring and reporting”², thereby improving healthcare and outcomes. It can enhance “site and participant experience, reduce data entry errors, minimize the ‘burden of source data verification’, and ‘facilitate’ the use of ‘risk-based monitoring (RBM)’”, as well as enable real-time data review and generate the outcomes-based evidence sponsors need to demonstrate the value of their products”².

Despite the well-documented advantages and wide availability of eSource tools, challenges around implementation mean adoption has been slow. SCDM eSource Implementation Consortium, which includes representatives of leading biopharmaceutical companies, academic medical centers, regulatory bodies, and healthcare technology providers, was established in 2017 to further the adoption of eSource approaches.³ As part of that work, the group has produced a series of podcasts in which leading experts from across the clinical research ecosystem share their practical advice on moving eSource from theory into practice. We have also distilled their learnings into an eSource Topic Brief series of four playbooks, each from the standpoint of one of the main stakeholder groups: CROs and vendors, pharma, regulators, and academia/sites.

Playbook 3: Practical advice for pharma

As part of the Playbook project, representatives from pharmaceutical organizations shared their experiences:

- MD Naqib Alam Ansari, Senior Manager, Clinical Data Strategy and Operations, AbbVie R&D
- Magda Jaskowska, PhD, Global Director/Leader of TA for Oncology, Data Strategy and Management, GSK
- Lauren McCabe, senior clinical data scientist, Clinical Data Science, Pfizer
- Muzafar Mirza, Senior Group Lead, Clinical Data Sciences, Pfizer
- Joseph Angiolielli, director, Information Management/Clinical Trial Solutions, Pfizer
- Peter Casteleyn, Director, Data Collection Solutions, EHR, Janssen R&D
- Rakesh Maniar, Global Clinical Data Sciences Leader, Executive Director and Head of eClinical Technologies, Global Clinical Trial Operations – Global Data Management and Standards, Merck & Co., Inc., Rahway, NJ, USA, and founding original co-chair of SCDM eSource Implementation Consortium
- Nadir Ammour, DDS, Global Lead, Clinical Innovation and External Partnerships, Sanofi

Typical challenges to adoption

eSource adoption must be a joint effort, and sponsors need both top-down management support and site engagement for successful implementations. Change management can be challenging and most contributors said their organizations had launched eSource initiatives as proof-of-concept (POC) projects.

“Depending on the eSource modality, multinational studies with large recruitment goals (hundreds and thousands of participants) or studies with complex adaptive designs can be challenging projects to begin eSource implementation. Smaller studies, which are typically in Phase I or II may be better suited for this work.” McCabe.

Compared to around a decade ago, the regulatory landscape is no longer viewed as such a challenge to adoption. Regulatory bodies, including the FDA and EMA, have, or are in the process of, publishing guidelines on eSource. Such documents are expected to evolve in the coming years, and cross-sector working is recommended.

“You should never assume what is OK in one geography is OK for the rest of the world. Whenever you are working with something new, you should have early engagement with health authorities and regulators in the key markets where you intend to file. It can help avoid surprises during review with regulators.” Maniar.

In terms of EHR to EDC extraction, there are challenges around legal and compliance, such as anonymization and data quality and integrity considerations, including accounting for errors within EHR data. There are also technical and operational challenges, with a lack of standardization hindering implementation, which must be carefully mitigated as part of the proof of concept (POC).

“One of the top challenges to bringing EHR data into the EDC is anonymization. We always have to carefully consider what we are bringing in and ensure we have a full audit of what we are sourcing, and where we are sourcing it from.” Ansari.

Mapping EHR data to EDC is often a challenge, as health records may contain inaccurate or irrelevant information (e.g., a ‘rule out’ diagnosis), and codes do not always align. Some teams recommended speaking with sites to select the right data and to map data fields manually, though they also acknowledged that this was not always possible, and that site-to-site differences in security protocols could be difficult to navigate.

“There’s huge diversity in how data is being captured by different research sites. Bringing that all together consistently into a clinical trial that crosses different sites is not straightforward.” Casteleyn.

There are similar challenges with eCOA that present difficulties with relation to cleaning, integration and real-time data access. However, solutions are being developed and deployed in this area.

Case study: Selecting core data sets.

The proposition: To demonstrate the feasibility of collecting clinical trial data from the EHRs of four separate hospitals with different protocols.

The challenges: The heterogeneity of data being collected in local hospitals can challenge efforts to aggregate the information into global clinical trial databases. Questions include:

- which data is common across trials?

- how can we assess safety?
- how can we assure return on investment?

The solution: An analysis of multiple protocols across multiple indications revealed a common subset of core datasets. The findings were confirmed by a group of data manager experts. After deciding to focus on structured data, the selected data sets were narrowed down to laboratory data, vital signs, concomitant medications, and demographics.

“With these data types, we can collect a good proportion of the data we need, while generating sufficient value for the sites as well as us sponsors.” Ammour.

Data integration

Pharma organizations are still working to find the best integration approaches. For EHR to EDC, teams are making “use of Application Programming Interfaces (APIs) and Fast Healthcare Interoperability resources (FHIR) standards”.^{4,5}

Methods will often depend on the EDC system and the eSource being used. Some EDCs, for example, will allow direct, real time, or near real time integration with eSources such as eCOA, though this is not always possible or, indeed, necessary. Sponsors can integrate third party data from digital medical devices into the backend of the EDC database, or through a data workbench or data repository. Site preferences are also important: considering each site’s requirements and needs can help secure their buy-in and engagement in the process. That is why Casteleyn believes in working with multiple data brokers to collaborate with different sites to rely on a variety of integration solutions. His approach is to create a “light menu card” of options that allows sites to consider the one that best fits their environment.

At Sanofi, the team were keen to ensure that any data integration method would be hybrid. In a request for proposal (RFP) process, the sponsor asked vendors to explain how they would account for sites that still worked manually. From a sponsor perspective, then, it was a question of evaluating and quantifying to ensure the proper required forms were populated in the EDC.

Our speakers saw one of the next big challenges to be processing high density, high dimensional data, which can be unstructured and terabytes in size, from digital medical devices and sensors into clinical trials.

Case study: Working with high-dimensional data.

The proposition: To utilize wearable sensors during a large clinical trial for direct from patient data collection.

The challenges: Smart watches were transferring three terabytes of data to the sponsor at regular intervals, presenting challenges to data acquisition, data cleaning, and data processing. The high-dimensional, unstructured nature of the data rendered traditional data management approaches and regular data formatting and transformations unfeasible.

The solution: Sponsors need to move from traditional data management to clinical data science by building data acquisition, cleaning, and review processes and new tools that are appropriate for use with big data. They also need to upskill staff accordingly.

Data cleaning

Some contributors said that eSource was fundamentally changing the practice of data cleaning, transforming it into more of a data validation process. The Pfizer team, for example, said

they validated the EHR data mappings in a test environment before pulling data into production. Therefore, data mapped from an EHR in an EDC production environment would not require traditional cleaning, such as confirmatory queries or source data verification (SDV). The EHR data still would require source document review (SDR) per the monitoring plan, as well as manual or system queries (i.e., edit checks), as applicable.

However, removing manual transcribed data entry will not remove all data discrepancies and will require data reconciliation/mapping validations, and this needs to be accounted for.

“We will never eliminate queries 100% because some will have nothing to do with the data sourced from the EHR, they are protocol deviations. There may, for example, be out of range data that entirely justifies a query and a response.” Ammour.

In addition, due to their extremely large volume and complex raw data output, eSources such as wearables are simply not suited to traditional data cleaning approaches.

“The term data cleaning is connected to activities that are almost impossible to carry out on the terabytes of data that come from wearables. We should be thinking about processing and pre-processing, focusing on verification of completeness, verification that the data is acquired as per the transfer agreement, and verification of the quality and content of the file as per the written specification.” Jaskowska.

Standardization and terminology

Most contributors agreed that there is a standards gap between the healthcare and clinical research spaces. While research tends to use CDASH and SDTM (CDISC standards), a majority of healthcare networks, at least in the USA, will use Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR). The HL7’s Vulcan accelerator project aims to bridge that divide by designing and implementing HL7 FHIR data exchange standards, and work is ongoing to address this domain by domain.

“If EHR had all the fields needed for research, the data could be entered once for the dual purposes of clinical care and clinical research. It would enable us to bring the protocol-relevant data from the EHR to the sponsor system with minimal transformations.” Maniar.

In addition, while FHIR standards have evolved in recent years, they are still not universally adopted on the healthcare side and many clinical research focused FHIR standards are still in progress. Sites with more legacy EHRs, or those built by smaller vendors, will often not be working with industry standardized data. This hinders the speed and efficiency with which sponsors can extract EHR data into EDC systems, and will continue to need custom solutions.

“There are so many EHR systems out there, and each has been made in a proprietary way. It makes it difficult for hospitals, sites, or clinics to share that data with sponsors... there is still a long way to go, and sponsors need to play a big role in helping sites to evolve.” Ansari.

As the FDA will only accept data in CDISC standards, sponsors still need to map EHR FHIR data and/or non-FHIR format data to CDISC. In the context of FHIR, this is not always straightforward since FHIR and CDISC have different structures and are used for different purposes. For example, the Concomitant Medication CDISC domain can be difficult to populate from FHIR data because a single medication may be represented across all four FHIR medication related resources (Request, Dispense, Administration, Statement) or the FHIR data are unclear about whether the medication was prescribed versus administered.

“As more sites adopt FHIR, the industry needs to evaluate whether it still makes sense to map from FHIR to CDISC.” McCabe.

Existing standards, our contributors said, were unfeasible for high-dimensional, unstructured wearables data. They called for expert discussion on how this gap could be best filled.

“There is a need for the industry and the healthcare sector to come together to develop mapping libraries that would become publicly available. Developing the concept of core data sets could become the backbone to help with generalizations and uptake.” Ammour.

Case study: Multi-EHR data integration

The proposition: To integrate data from two EHRs, each developed by different technology vendors, for use in a non-interventional study of around 50 participants.

The challenges: Each EHR collected the data in their vendor’s proprietary format, making comparison and analysis extremely challenging. It meant the sponsor, AbbVie, was faced with conducting time- and resource-consuming transformations in order to integrate the data.

“Even though integration was achievable through transformations and programming on our end, this is not efficient EHR to EDC extraction.” Ansari.

The solution: Greater cross-vendor standardization of the types of data being collected; data anonymization and storage formats could significantly streamline EHR to EDC extraction when working with multiple systems.

Data flow

Pilot/POC projects at AbbVie have included creating intermediary EDCs that pull data from site data warehouses to the sponsor EDC using CDISC ODM standards.

At Sanofi, implementation has focused on minimizing the impact on workflows. Sites enter data, either manually or electronically, into a third-party vendor system that facilitates the necessary data loading transformations. The system module or middleware ensures only the values and data required by the protocol are pulled from site to sponsor, ensuring protocol compliance. The data is then reviewed by study coordinators, who then transfer it into the EDC.

Janssen has been operating on a “push” principle that enables sites to feed the data through to the sponsor, while still maintaining full control and ownership of it per GCP-ICH expectations. Such frameworks utilize FHIR standards to facilitate interoperability, with data brokers mapping and transforming data into eCRF in CDISC ODM structure, and pushing it to the sponsor. However, new systems and pipelines, such as data exchange tools, are needed to transfer unstructured data.

“Application Programming Interface (API) layers are ideal. But when we started, we used a point-to-point data transfer method where sites would push data to secure file servers, and we would bring it into our ecosystem, and load it into our data management system.” Maniar.

Case study: Automatic transfer of EHR data to EDC

The proposition: To transfer data seamlessly and efficiently from multiple site EHR systems to a central sponsor EDC.

The challenges: EHR and EDC data fields were misaligned.

The solution: The sponsor partnered with a vendor to build middleware that correlated EHR data fields to EDC data fields. Mapping was configured according to sponsor-provided specifications. The solution, based on a FHIR API, enabled data to be transferred to the sponsor within 24 hours of site input.

The future of eSource

When asked about the future of eSource, many contributors pointed to the more seamless integration of data.

Said Maniar, “The future will see data coming from many different sources directly into a single sponsor ecosystem. The technology is out there, with cloud computing and the decreased cost of memory, to enable that.”

Ammour agreed, describing eCRF as a “transitional technology”, and explaining that future protocols would collect data from its original sources, in electronic formats, while Mirza said that getting patient treatments on the market quicker relied on “having real time data feeds for our trials”. “It is something we need to continually push for,” he added.

Casteleyn pointed to a future where patients were in control of their own data, and in which that data could be automatically used in research without the need for multiple integrations. “It will ultimately make clinical research much more patient-mediated and more efficient,” he said.

All will be facilitated by a shift from data management to clinical data science, meaning sponsors need to be ready to support skills development in this area, the speakers said. “We need data science functions within clinical operations to handle big data, and to achieve that we need the right people and the right processes,” explained Jaskowska.

Another crucial enabling factor will be greater standardization. Ansari said standards needed to continue to evolve “so that they can simplify the process of adoption, both from a legal/compliance standpoint and from a data integrity standpoint”.

Achieving this end requires cross-sector working. “The more pharma companies group together to create common ground and contribute to the process, the faster we will be able to reach our goals,” said Ansari. Casteleyn agreed: “I don’t believe this is something that one sponsor or regulator can do alone. We all must work together. As the saying goes, ‘If you want to go fast, go alone. If you want to go far, go together’.”

References

1. U.S. Department of Health and Human Services Food and Drug Administration, Electronic Source Data in Clinical Investigations: Guidance for Industry. FDA. September 2013. Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/electronic-source-data-clinical-investigations>. Accessed July 17, 2023.
2. Parab AA, Mehta P, Vattikola A, et al. Accelerating the adoption of eSource in clinical research: a TransCelerate point of view. *Ther Innov Regul Sci*. 2020;54(5):1141-1151. DOI: 10.1007/s43441-020-00138-y
3. eSource Implementation Consortium. Society of Clinical Data Management. Date unknown. Available at: <https://scdm.org/esource-implementation-consortium/>. Accessed July 17, 2023.
4. Jennings DG, Nordo A, Vattikola A, Kjaer J. Technology considerations for enabling eSource in clinical research: industry perspective. *Ther Innov Regul Sci*. 2020;54(5):1166–1174. DOI: 10.1007/s43441-020-00132-4.
5. Monica K, EHR Intelligence (ehrintelligence.com), “VA Awards Cerner \$140M Task Order for EHR Interface Support”, news article, June 07, 2019, Available at: <https://ehrintelligence.com/news/va-awards-cerner-140m-task-order-for-ehr-interface-support>. Accessed September 25, 2023.