



WHITE PAPER

**Clinical Trial Coding:
Overcoming the Challenges
through Automation within
Electronic Data Capture
Applications**

PHASE•FORWARD™

Abstract:

This paper reviews the issues surrounding the clinical trial coding process and identifies specific needs that have driven the adoption of automated solutions within electronic data capture (EDC) applications. Most notably, pharmaceutical and biotech companies have placed a premium on pharmacovigilance, process efficiencies, improved data quality, data standardization, global access to data, and fully documented regulatory compliance. This paper defines these needs, explores workflow challenges, and recommends a proven approach for companies considering conversion to an automated coding solution.

Part 1: Requirements Driving Automated Coding Adoption in EDC

Financial and competitive pressures in the pharmaceutical and biotech industry continue to motivate sponsors and Contract Research Organizations (CROs) to find new ways to efficiently conduct clinical trials. Coding is one area where automated solutions can be deployed - in combination with an EDC application - to facilitate data classification and management, streamline clinical trial processes, dramatically reduce errors and cut operational costs. The adoption of automated coding solutions is driven by a set of common sponsor requirements that include:

Safeguarding patient privacy and safety. Proper management of patient data has become an increasingly sensitive issue with the advent of Good Clinical Practice (GCP), 21 CFR Part 11, and HIPAA. In addition, the Food and Drug Administration (FDA) has raised the standard for what constitutes satisfactory pharmacovigilance. Drug safety has consequently become an ongoing process which begins earlier and ends later in the drug development life cycle. Automated coding enables accurate classification and secure management of patient information, and aids in the quick identification of adverse events, critical to meeting the goal of optimal patient safety.

Driving down costs and accelerating development efforts. In seeking to stay competitive, manufacturers are focused on achieving new process efficiencies that will reduce costs and hasten the product development cycle. Automated coding promotes consistent classification and management of a patient's medical history, including medications, and increases process efficiency.

Addressing coding complexities. The industry adoption of the Medical Dictionary for Regulatory Activities (MedDRA) and the WHO Drug Dictionary (WHO-DD) championed the need for advanced dictionary management functionality within EDC systems. However, EDC systems were not built to support the dictionary complexities of MedDRA or WHO-DD. Dictionary Management is a significant challenge since most dictionaries are dynamic and updates are routinely released which may have clinical impacts for terms previously coded. For example, MedDRA versions are released semi-annually. However, clinical organizations may elect not to update to the latest dictionary versions as often as they are released. As a result, clinical organizations may have studies utilizing different versions of the same dictionary across multiple trials. Terms can either become obsolete or modified, making clinical surveillance a challenge as terms may be coded inconsistently. Automated coding creates a consistent classification scheme and incorporates processes to routinely update the studies so that variability between coders and across studies is dramatically reduced.

Improving data quality and data standardization. The collection and management of clinical trial data can come from a variety of sources including sponsors, CROs, investigators and directly from the patient. Adding to the complexity is the interaction and communication of data between entities. The more entities responsible for the data collection, the greater the likelihood for

clinical errors, inconsistent coding assignments and more laborious review processes. The goal is always to provide error-free information, at a moment's notice, and with relative ease. Automated coding reduces error rates and greatly eases a sponsor's challenge in managing the complexity of coding medical terms across all studies and locations. Data is consistently coded through the use of multiple dictionaries (MedDRA, WHO-DD, and custom), synonym and stopword lists and other auto-coding tools. The increased consistency of coding across the organization helps reduce reconciliation at later points in the study process, enabling a far more efficient review by sponsors.

Providing global accessibility. Trials are increasingly being performed on a global scale, with the potential for variability in both nomenclature and workflow processes. Automated coding is critical to effectively work cross-organizationally, over a large geography. Automated look-ups act as a translator in standardizing key information. Most importantly, because the solution is web-enabled, information is easily accessible from anywhere.

Meeting or exceeding regulatory compliance. Sponsors today are particularly sensitive to ensuring that all regulatory requirements are met and documented for clinical trials. Automated coding provides an audit trail that sponsors may depend upon as one of the many tools to ensure their compliance with 21 CFR Part 11 and consistency for regulatory review and data validation.

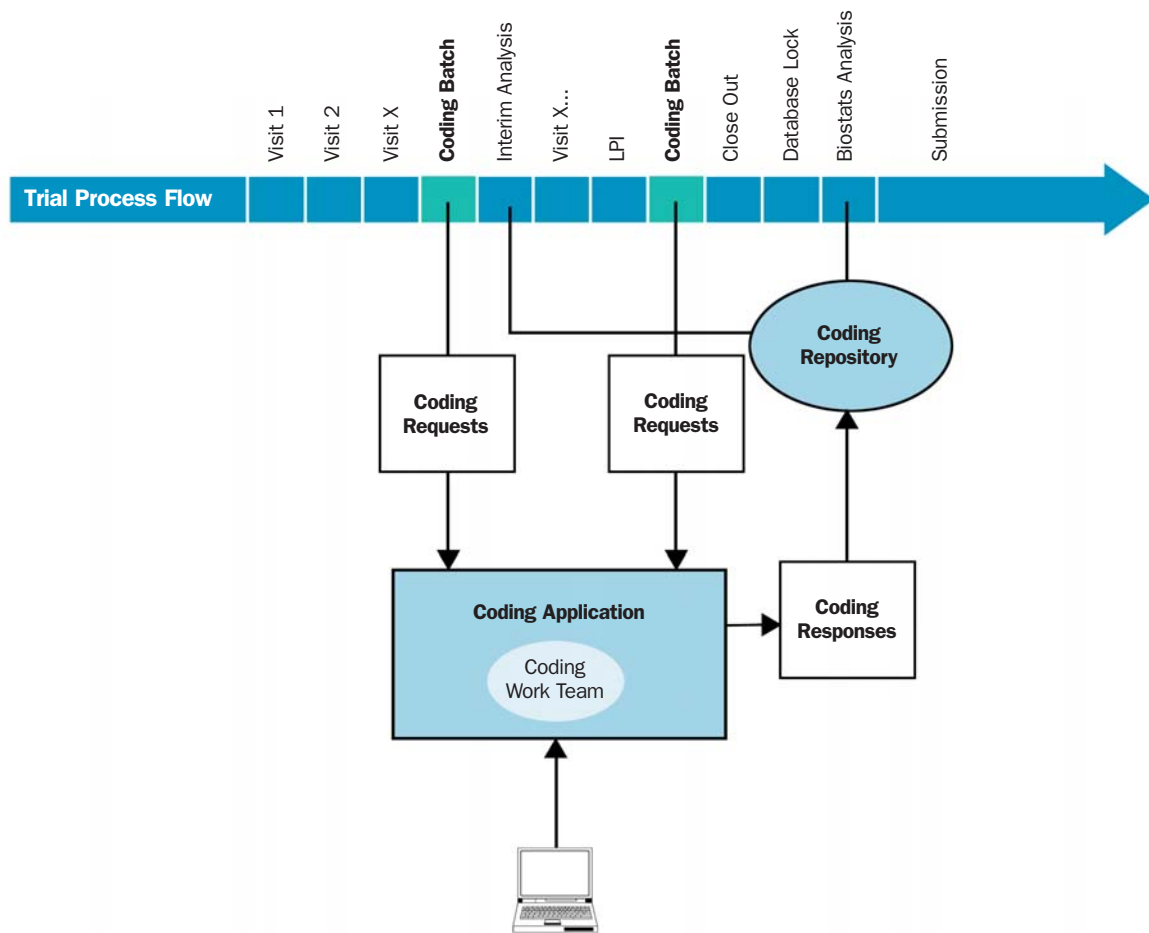
Part 2: Traditional Coding Workflow

Stand-alone Coding + Electronic Data Capture (EDC). While there are variations in the processes used to conduct clinical trials, in the area of coding many standard practices exist. As the process flowchart illustrates (*Figure 1*), coding is most often performed as a stand-alone operation, separate from electronic data capture. This separation exists because there is a specialized team of professionals who exclusively perform coding activities. These "medical coders" are not EDC users, and they remain focused on ensuring that medical history, medication and any related events are correctly classified, using standard classification schemes.

As *Figure 1* indicates, the EDC system acts as the initial collection point for terms that will need to be coded. Terms are periodically extracted to a coding application where coding can be performed, then coded data is typically fed into a data management system or coding repository. Retrieval of coded data by the EDC system requires that the EDC system tie into the data management system or coding repository.

There are generally three distinct coding roles involved in processing a request: the coder who initiates the response for coding assistance by determining the classification code, the approver who must validate and assess that the coding is correct, and the administrator who oversees the process, assigns the coders their work, and triages where necessary. It is not unusual

Figure 1: Traditional Batch Coding Workflow



for an individual to be responsible for coding, approving and administering the application. Typically, a clinical data manager would oversee coding assignments, synonym table management and workflow distribution, and the systems administrator would be responsible for dictionary load and access rights.

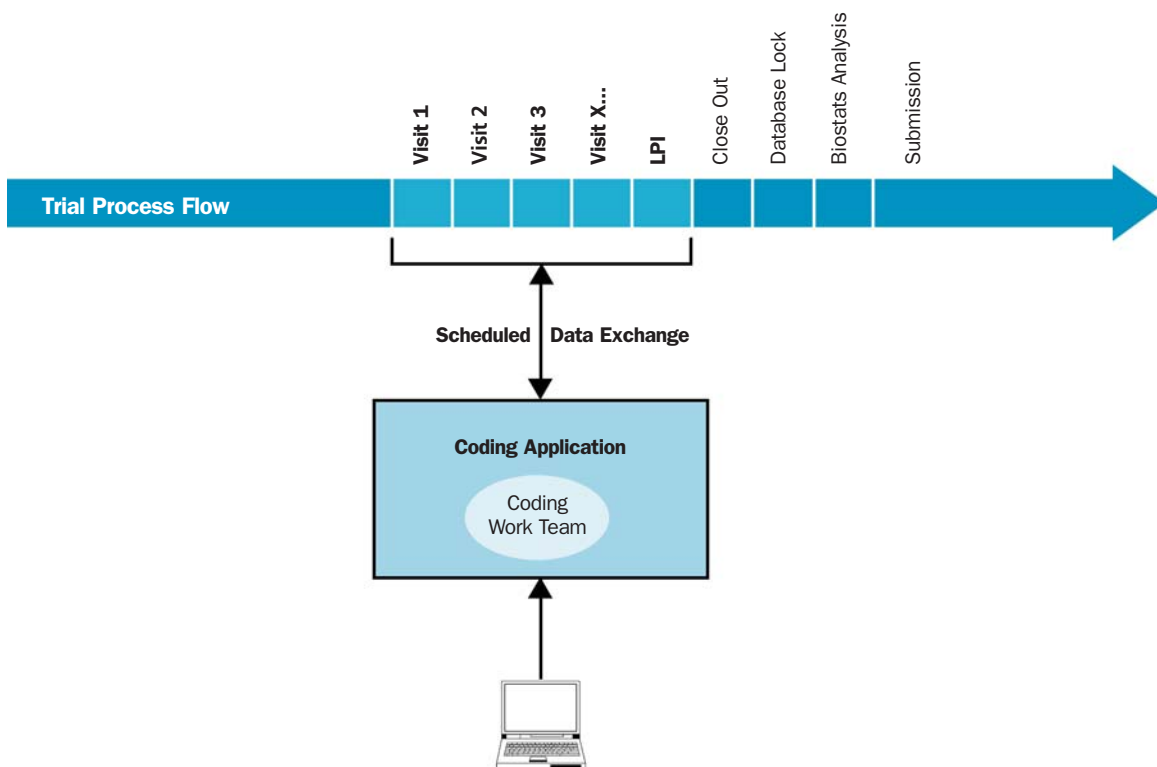
Once a code has been selected and approved, the coding team transmits the designated code to either a coding repository or back to the EDC system. Verbatim terms which are misleading or need further contextual information can be queried and sent back to the investigative site for clarification. Since most EDC systems are not integrated with a coding application, sponsors must develop a query process outside of the EDC system. Email is typically used as the mode of communication to answer any questions they may have. In general, the process is decentralized, and does not capture the audit trail, nor trace the historical context of the coding assignment.

As a way of explanation, it is important to understand why coding would not usually be based inside the EDC operation. In a stand-alone operation, the coder can quickly deploy his skills and process each request in a batch workflow environment. When coding is performed within EDC, the coder needs to open each electronic case report form (eCRF), find and respond to the coding request, exit the eCRF, and then enter the eCRF for the next coding request. This process would be extremely cumbersome and inefficient. For this reason, coding is most effectively performed as a stand-alone operation, and then tied into the EDC data repository.

Part 3: Centralizing Coding Workflow Adds New Promise

Managing Disparate Applications. Each clinical organization typically has dictionary management functionality either built into their clinical applications or integrated with a third-party coding application. Each coding application has its own search engine, dictionary load and application interface. The more clinical coding applications needed to perform coding, the greater probability for clinical error. Disparate clinical data management application issues are more complicated when third-party coding applications are required to perform adverse event and medical drug coding. Since the coding application and the clinical data management application are not tightly coupled, process delays are incurred due to query management and workflow communication constraints. Furthermore, the maintenance and product release for each application are not

Figure 2: Automated Coding Workflow



synchronous thus causing additional cost for validation and support from multiple vendors.

Opportunities for Process Improvements Using Centralized Coding and EDC. There is a significant opportunity to improve the clinical trial process by centralizing the automated coding workflow. Centralized coding solutions yield new efficiencies as the workflow process switches from batch to real-time. The result: continuous data exchange between the coding and source systems provides real-time feedback to sponsors.

There are many advantages to this centralized approach, in that it:

- **Extends the value of the automated look-up function.** The goal in any technology-based coding solution is to automatically assign as many codes as possible, thus minimizing the time-consuming task of manual look-up. With centralized coding solutions, users can centrally create and store definitions - automated algorithms used for different trials and different therapeutic areas. This provides medical coding administrators the level of flexibility and automation needed, all in one application.
- **Automates code propagation across all trials.** As coders encounter "like verbatims," the code selected for one is automatically propagated across multiple trials, saving redundant manual coding steps for the coder.
- **Allows information to be exchanged more frequently with the EDC system.** While traditional automated coding can sync up with EDC, it requires more effort because the applications are not as tightly integrated. Communication is often disjointed, causing the need for further clarification before work can proceed. Centralized coding, however, integrates functionality at the application level, which makes workflow, data queries, and data exchange all much easier. These benefits allow eCRFs to be closed out more quickly. Assignments can be better routed within the team because of real-time volume information.
- **Provides a seamless interface with EDC systems, reducing the need for costly integration services.** Facile exchange of information between centralized coding and EDC systems is critical for a well-designed clinical trial operation. Depending on the environment, application program interfaces (APIs) may be needed for connectivity. Sponsors, in general, prefer to minimize their investment in customized integration services.
- **Configures the workflow so that automated and manual coding processes work intelligently together.** A central coding solution should support both automated and manual processes, using a methodology that intelligently assigns verbatim terms to a designated coder. Based on Standard Operating Procedure (SOP) coding guidelines and workflow review steps, all approved terms may automatically be re-routed back to the EDC system.

- **Accelerates data visibility for more informed decision-making.** When central coding is set up to work independently from the EDC solution, parallel work paths can be established that allow clinical study teams and clinical coding teams to work simultaneously. One significant advantage is that coding can happen earlier in the trial cycle, providing valuable data visibility for pharmacovigilance managers reviewing safety data. The improvements are due to higher quality data, easier data access and an automated and more streamlined workflow. While no single benefit may seem extraordinary, taken in total, these differences offer significant competitive advantages for pharmaceutical and biotechnology companies seeking to reduce costs and speed up product development and rollout.

Part 4: Planning Ahead for Centralized Coding - Some Practical Tips

While the process for how to automate and centralize coding may vary from organization to organization, there are several practical tips that have proven to be helpful to sponsors. Here are a few recommendations to consider when evaluating or implementing a centralized coding solution:

- **Articulate the business objectives for the revised operation.** Describe whether accuracy, speed, consistency, and/or cost reduction drive the need for analysis and change. Then establish measurable goals to be achieved by the new set-up.
- **Document the current workflow including any bottlenecks and redundancies, as well as areas where the current operation runs efficiently.** This is important so that in the revised operation, those areas needing improvement can be addressed up front in the design, and those areas working well can either be moved intact or slightly modified based on adjacent activities.
- **Forecast a range of eCRF volumes from low to peak volumes and estimate the coding demands.** This not only helps with staffing concerns, but is also important as one of the variables affecting database lock time, validation and analysis. A projection of autocoding rates will be needed in order to estimate the degree of manual intervention.
- **Redesign the workflow, first, in the areas that form the biggest bottlenecks.** This mirrors what typically happens in business process optimization projects where teams focus on fixing what is "most broken" first. From there, the workflow can be revisited in adjacent areas.
- **Before finalizing any workflow changes, do preliminary testing to see if desired improvements have been achieved.** Benchmarking the performance of the revised operation under a variety of scenarios will help the sponsor fine-tune where necessary. It will also help management understand when unanticipated slow-downs occur whether it is due to new workflow or some other factor.

Part 5: Phase Forward's Approach to Automated Centralized Coding

When Phase Forward began designing a new centralized coding solution - Central Coding for InForm™ – the company paid particular attention to sponsor-driven requirements, which include:

Global accessibility to data

Increased visibility of results

Improved process efficiencies

Substantial cost reductions

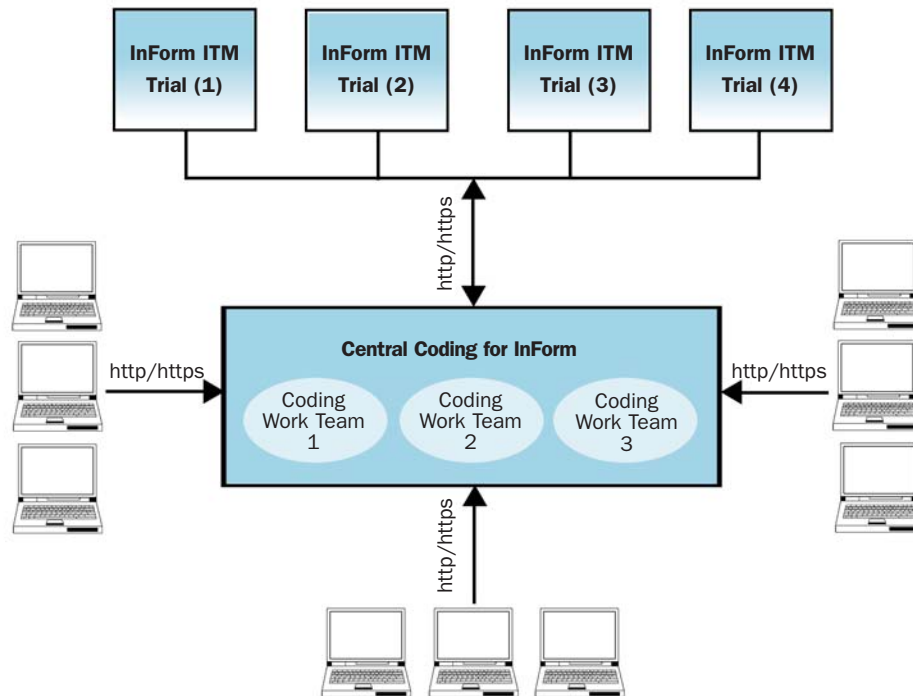
Based on these requirements, Phase Forward adopted several key design tenets:

- **Central coding should be designed as an independent module capable of communicating with its source systems.** This provides one view for all coding needs and enables consistent coding across the organization. As a stand-alone module, it also streamlines dictionary maintenance to one location.
- **Enhanced automation must significantly improve autocode hit rates.** Automation naturally reduces manual intervention and promotes data standardization. The additional use of synonym lists, stop word filtering, auto-suggestion, customizable coding algorithms and the like improve the autocode hit rate.
- **Configurable workflow is critical for process optimization.** Flexibility is needed by sponsors in order to manage multiple trials across multiple sites, with varying therapeutic areas and different geographies. Dictionaries, work assignments and process flow must all reflect this complicated environment.
- **Centralized coding must be tightly integrated with an EDC system.** Integration is essential for permitting bi-directional, real-time information access. This ability allows professionals to review source eCRFs as needed and/or initiate queries before assigning codes.
- **Central coding must be web-enabled.** The web has become a universal tool used by businesses to share and retrieve information globally – and automated coding is no exception. As a web application, central coding requires lower installation and support costs, and best achieves cross-organizational collaboration and global access.
- **Support the customer with tools that allow 21 CFR Part 11 compliance.** Audit trails, systematic tracking of data, and enhanced visibility are part of the Phase Forward solution.

Part 6: Introducing Phase Forward's Central Coding for InForm™

Phase Forward's Central Coding solution has been designed to meet sponsor needs in the broadest and most complete way possible.

Figure 3: Multi-team, Multi-trial Coding Workflow Using Central Coding for InForm



Some of the key features of Phase Forward's Central Coding solution are:

- **Automated Task Management.** Phase Forward's Central Coding product's advanced features include accurate high performance auto-coding, customizable synonym and stopword lists, propagate functionality, and configurable workflow. These capabilities provide medical coders a highly efficient way to review, resolve and approve coded terms - all of which can take place simultaneously across staff, sites, trials and locations.
- **Centrally Optimized Process Control.** Phase Forward's Central Coding product is architected to connect to all sponsor and CRO trials, serving as an independent collection and access point for all verbatims. The product further provides the workflow and management interface the coding team will use to assign, code and approve coded terms.

- **Coding-centric Workflow.** Phase Forward's Central Coding product supports both automated and manual coding processes, using an inbox "to-do" methodology. The product's automated workflow capabilities include auto-code and auto-approval of coded terms and auto-suggest and auto-review of coded terms. The product's configurable role and work management capabilities intelligently assign and rout verbatim terms to the designated coder for coding, and the designated reviewer for approval. Once terms are approved, they are automatically sent back to the originating source system.
- **InForm (EDC) as the Source System.** Designed to complement the trial management efficiencies found within the InForm ITM solution, Phase Forward's Central Coding product seamlessly integrates with InForm as the source system for verbatim terms. For sponsors and CROs, this powerful product combination significantly reduces costly integration services and time delays when integrating and deploying a coding solution.
- **Advanced Dictionary Administration.** Phase Forward's Central Coding product provides multi-level, multiaxial, and multi-version dictionary support for the industry standard WHO Drug Dictionary (WHO-DD) that defines drug terms, and the Medical Dictionary for Regulatory Affairs (MedDRA) that defines adverse events. Additionally, the Central Coding product supports one-to-many hierarchically structured custom or proprietary dictionary formats used by sponsors and CROs.
- **Powerful Performance Tools.** Phase Forward's Central Coding product contains a variety of performance tools to support sponsors and CROs, including link-through capabilities to InForm for reference and query initiation; global, multi-trial/multi-center connectivity; and ready-to-run reports for coding classification, consistency, coded items and user reports.

For more information about Phase Forward's Central Coding for InForm solution, visit our website: www.phaseforward.com/products or email us at: ROI@phaseforward.com

The information contained in this publication (i) is provided without warranty of any kind, (ii) references only laws, regulations and rules as in effect at the time of publication and (iii) is subject to change without notice. Nothing contained in this publication should be construed as the rendering of legal advice, and readers are advised to seek the appropriate advice for specific cases. The information contained herein regarding the products and services of Phase Forward Incorporated is an incomplete summary and is not a substitute for obtaining products and services from Phase Forward. Any such products and services are governed solely by the agreements between Phase Forward and its customers.

Copyright © 2006 Phase Forward Incorporated. All rights reserved. Phase Forward, InForm, Clintrial, Clintrace, WebVDME, WebSDM, CTSD and the Phase Forward logo are trademarks or registered trademarks of Phase Forward Incorporated in the U.S. Patent and Trademark Office and other jurisdictions. Other marks may be trademarks or registered trademarks of their respective owners.

PHASE•FORWARD™