A New Breed of Digitisation:
The Path to Congruency and Faster Study Start-up

Most clinical research institutions believe their study start-up delays are a technology problem. They aren't. It's a fragmentation problem, and many technologies have only made it worse. This article examines how AI-enabled digitisation creates a single authoritative source from protocol through activation, what that means for the over a dozen internal groups who depend on that information and why speed and congruency together, not either alone, are the real prize.

Most institutions think they have a technology problem when really, they have a fragmentation problem. Interestingly, it's the technology they've deployed over the past decade that has quietly made it worse.

Here's how it happens. A new study arrives and the protocol lands in the hands of a contracts team, a budgeting team, a coverage analysis team, an IRB coordinator, a feasibility analyst, a study coordinator, a billing specialist, a lab team, a radiology team and that's just a partial list. Over a dozen internal groups, each receiving the same source document, reading it through their own lens and building their own version of the truth.

This unintentional fragmentation isn't a failure of effort or intelligence. It's a structural outcome of how clinical research institutions are built – by function, not by flow. And when a dozen versions of the truth collide downstream, you get what every site and sponsor already knows: miscommunications that require resolution, budget assumptions that don't match protocols and activity schedules, coverage analyses that conflict with billing expectations and amendments that could have been avoided if everyone had been reading from the same page.

But they weren't and therein lies the problem.

What Digitisation Has and Hasn't Solved

The industry has been digitising for years through electronic trial master files, CTMS platforms, eRegulatory systems and eSourcing tools. Each solved a piece of the puzzle, but none solved the underlying architecture.

The reason is straightforward – most digitisation efforts were built to replace paper within a function, not to connect multiple functions to each other. You get an eTMF that stores documents more reliably than a filing cabinet. You get a CTMS that tracks enrolment more accurately than a spreadsheet. What you don't get is a shared foundation – one place where the study's core parameters live and from which every downstream function draws.

So, while processes became faster and documents became more searchable and better organised, they remained siloed, just digitally.

This is the distinction that matters. Digitisation and integration are not the same thing. Most institutions have achieved the former with very few having achieved the latter. And without integration, fragmentation persists.

The Protocol as Source of Truth

Every study start-up inefficiency traces back to a single root: the protocol is read many times, by many people, for many purposes and each reading produces a separate artifact that is never reconciled against the others.

The contracts team extracts payment terms and deliverables. The budget team extracts procedures and visit schedules. The coverage analysis team extracts standard-of-care versus research-driven activities. The IRB coordinator extracts consent requirements and subject populations. These extractions happen in parallel, in isolation and with no mechanism to flag when one team's interpretation diverges from another's.

The amendment is often the first signal that something went wrong. By then, the damage is done: weeks of rework, re-signatures, resubmissions. The root cause isn't the amendment; it is that multiple teams read the same protocol and came away with different understandings and nobody found out until it was too late.

AI-enabled digitisation changes the starting point. Not by replacing human judgment on any of these questions, but by parsing the protocol once, extracting the structured data that every downstream function needs and distributing that single authoritative source before anyone builds their version of the truth.

Parse once. Distribute one foundation. Let every team work from the same ground.

What 'AI-Enabled' Actually Means Here

It's worth being specific, because the term gets stretched past usefulness.

AI-enabled digitisation in this context is not a chatbot that answers questions about your protocol. It's not a summarisation tool that produces a two-paragraph overview for busy executives. Those are useful tools, but they're not what we're talking about.

What we're talking about is structured extraction – the ability to read a protocol and produce a machine-readable data layer that downstream systems can consume. Procedures mapped to visit schedules. Visit schedules mapped to budget line items. Budget line items mapped to coverage analysis categories. Coverage analysis categories mapped to billing codes. Each connection is explicit, traceable and derived from a single source.

The AI does the extraction work that currently happens separately, in multiple departments, with multiple divergent outputs. It doesn't eliminate human review – every function still applies its expertise to the structured output. What it eliminates is the re-reading, the re-interpreting and the silent divergence that happens when every team starts from scratch.

The difference is recoverable errors versus unrecoverable ones. When every team reads from one source, disagreements surface early, cheaply and are typically fixable before they metastasise into amendments. When every team reads separately, disagreements

surface late, in the form of delays and rework that can set a study back by months.

Three things change immediately when you move to a structured, shared data layer. Speed to access: every team gets what they need from the same source, at the same time, without waiting for another function to finish their independent read. Uniformity of data: there is no longer a budget version, a contracts version and a coverage analysis version, there is one version. And when something changes, such as a protocol amendment, a scope adjustment, or a sponsor revision, the update propagates across all teams simultaneously. No more email chains, version control nightmares, or teams operating on last month's assumptions.

That last point deserves emphasis. In a fragmented system, updates are the most dangerous moment. One team gets the amendment, but another doesn't and the study moves forward on misaligned assumptions. In a shared data layer, an update for one is an update for everyone.

The Value of AI-Enhanced Conversation

Structured data also unlocks something that pure document management never could: genuine, context-aware conversation about the study.

A well-designed AI interface built on top of structured protocol data doesn't just retrieve documents, it answers questions. A coordinator can ask why a specific procedure is classified as research-driven rather than standard of care and the system can explain, traceable back to the protocol language that drove the classification. A budget analyst can ask how a scope change affects the visit schedule, and get an answer grounded in the actual data layer, not a guess.

This is materially different from a general-purpose chatbot. Intelligence is only as good as the data underneath it. When the data is structured, consistent and derived from a single source, the conversation becomes a genuine tool for the people doing the work.

Speed and Congruency: Two Outcomes, One Architecture

Everyone wants faster study activation with shorter start-up timelines, days-to-first-patient, sites opening and enrolling. These are the right goals, but speed without congruency is a liability, not an achievement. A study that activates fast on misaligned assumptions costs more to correct than one that took two extra weeks to get aligned at the start.

The institutions that activate fastest and run cleanest aren't trading one for the other. They've built the architecture that produces both. Congruency, the state where every function's understanding of a study is aligned with every other function's understanding, is what makes speed durable. Fewer amendments, fewer billing errors, fewer coverage analysis exceptions that require IRB re-review, and fewer budget reconciliation conversations between the site and sponsor that consume weeks of calendar time.

An institution with congruency doesn't just activate faster – it maintains that speed. The speed doesn't erode into rework and correction cycles three months into the study. That's the difference between a fast start and a clean run. And that's where AI-enabled digitisation is the architecture that produces both.

Good Governance Is Not Optional

Deploying AI-enabled digitisation responsibly requires more than good technology. It requires a governance framework that keeps humans genuinely in the loop, not ceremonially, but operationally.

The most important governance mechanism is bidirectional visibility. Any downstream team that identifies a problem like a misclassified procedure, a scope discrepancy, or a data anomaly must have a clear, fast path to flag that issue back upstream, where it

becomes visible to all stakeholders simultaneously. Not an email to a supervisor or a ticket in a queue. A live signal surfaced to the shared data layer, which every affected function can see and act on.

This is what separates AI-enabled digitisation from AI-enabled automation. Automation removes humans from the loop. Properly executed digitisation puts humans in a better position in the loop. AI extracts and structures, while humans validate, question and correct. When a downstream team catches something that the AI missed, that catch makes the shared source more accurate for everyone. The system improves because humans are genuinely engaged with it, not just downstream consumers.

Human checkpoints are a feature, not a workaround. When you build them in explicitly, make them visible, and measure how often they catch something, that measurement will tell you where the AI needs to improve and where human expertise is irreplaceable.

What Implementation Actually Looks Like

None of this is theoretical. Institutions that have already moved in this direction share a few common patterns.

They start with the protocol since it is the source document from which everything else derives. If you can extract the structured data layer from the protocol accurately, every downstream function benefits.

They build for consumption, not for storage. The goal isn't a better filing system for protocol documents; it is a structured output that every downstream system, including CTMS, budget tool, coverage analysis platform and eRegulatory system, can read and use. It is designed for the consumer, not the archivist.

They measure both speed and congruency. Activation timelines matter, as do amendment rates, budget variance at closeout and coverage analysis exception rates. If you're only measuring how fast you activated and not how cleanly you ran, you're missing half the picture.

The Site-Sponsor Dimension

Everything above applies within an institution. The fragmentation problem doesn't stop at the site's walls.

Sites and sponsors often begin study start-up from the same protocol and arrive at materially different understandings of scope, procedures and expectations. Those differing understandings only collide when the contract is being negotiated, or the budget is being reconciled. By then, both parties have invested weeks in versions of the truth that don't match each other.

AI-enabled digitisation creates the possibility of a shared foundation between the site and sponsor – not just within each organisation, but across the table. When both parties work from the same structured protocol data layer, the negotiation starts from alignment rather than from divergence. That changes the nature of the conversation. It doesn't eliminate disagreement, but it moves the disagreement upstream, where it's cheap to resolve, rather than downstream, where it's expensive.

This is the longer arc of where the industry is heading. Not just faster start-up within institutions, but a congruent start-up across them, and a shared data foundation that makes both possible.

A Framework for Getting Started

Assessing your fragmentation points is the right first step. Ask these questions honestly:

- How many times is your protocol read before the study activates?
- By how many different teams?
- What does each team produce from that reading?
- When does anyone compare those outputs? Additionally, consider where your amendments cluster.
- If you map your last amendments against the workflow stage where the root cause originated, a pattern will emerge. Most institutions find the same two or three fragmentation points generating much of their amendment volume.

What would it take to parse the protocol once and distribute the output? This is a technology question, but it's also a workflow question and a governance question. The technology exists, but the harder work is redesigning the workflow around a shared source rather than a siloed one.

The institutions that answer these questions honestly, and act on the answers, are the ones that will look back in five years and wonder how they ever did it the other way.

Closing

The fragmentation problem in clinical research is not new. What's new is that we now have the tools to solve it at the source rather than manage it downstream.

AI-enabled digitisation isn't a faster version of what we already have. It's a different architecture. It is developed with one source and foundation. Every function is working from the same grounding level and every function is empowered to improve that ground when they find something wrong.

The studies that activate in weeks instead of months and stay on track after they open aren't doing more. They're fragmenting less, and that is the new path forward.

Jim Wagner

Jim Wagner is senior vice president of AI and Product Innovation – Site Enablement at WCG, where he leads the expansion of AI-driven capabilities across study start-up and site enablement following WCG's acquisition of The Contract Network. Previously, Jim was co-founder and CEO of The Contract Network, an AI platform designed to accelerate clinical trial contracting and improve alignment between sponsors, sites and CROs. Earlier in his career, he served as vice president of Agreement Cloud Strategy at DocuSign and as president of Seal Software. Jim is a Duke Law School graduate and a member of the Duke Law Tech Lab advisory board.