The €100 Million Question: Why Is Clinical Data Still Designed Last?

-

If the protocol is the story – data is its language

Sverre Bengtsson

CEO, Digital Trial Consultants

The bio

- Started in clinical research in the early 1990's at MiniDoc, the world's first electronic patient diary
- Worked in **CRO** as Statistician, Statistical Programmer and Clinical Data Manager
- Co-founded a **CRO** which morphed into the eClinical tech company Viedoc Technologies, exited.
- Vice chair of **ACDM**
- Currently active investor and board professional in the life science sector, often focused on clinical trial technologies.
- Proud of my family and Molly, my mini schnauzer

A look at the clinical trials of today*

Data management is ~15% of trial costs but underpins 100% of trial credibility.

Technologies is approx. 5% of costs in clinical trials.

Clinical operations drives study setup, but data management often arrives too late to the table.

Leads to inefficiencies, protocol amendments, and costly rework.

Why do we still treat data as an afterthought in a €100M trial?

- Phase I 4 – 5.26 MUSD
- Phase II 7 – 20 MUSD
- Phase III 20 – 100+ MUSD

- Site costs 30 %
- Recruitment 20 %
- Data Management 15 %
- Regulatory compliance 10 %
- Other 25 % (e.g. manufacturing, OH)

* https://www.abacum.ai/blog/clinical-trial-costing

The “Data Last” Practice

- *Organisational silos:* Protocol, medical, and data functions still operate on asynchronous timelines and with different objectives. Why does **CDM** come in late to the party?
- *Procurement bias:* Contracts reward speed to study start, not quality of data design. **FPI**, then 1st amendment to fix what needs to be fixed.
- *Standard fatigue:* Over-complexity and multiple competing standards discourage early alignment.
- *Cultural hierarchy:* Data roles remain undervalued in protocol authoring and scientific design. This needs to change!

The quiet powerhouse

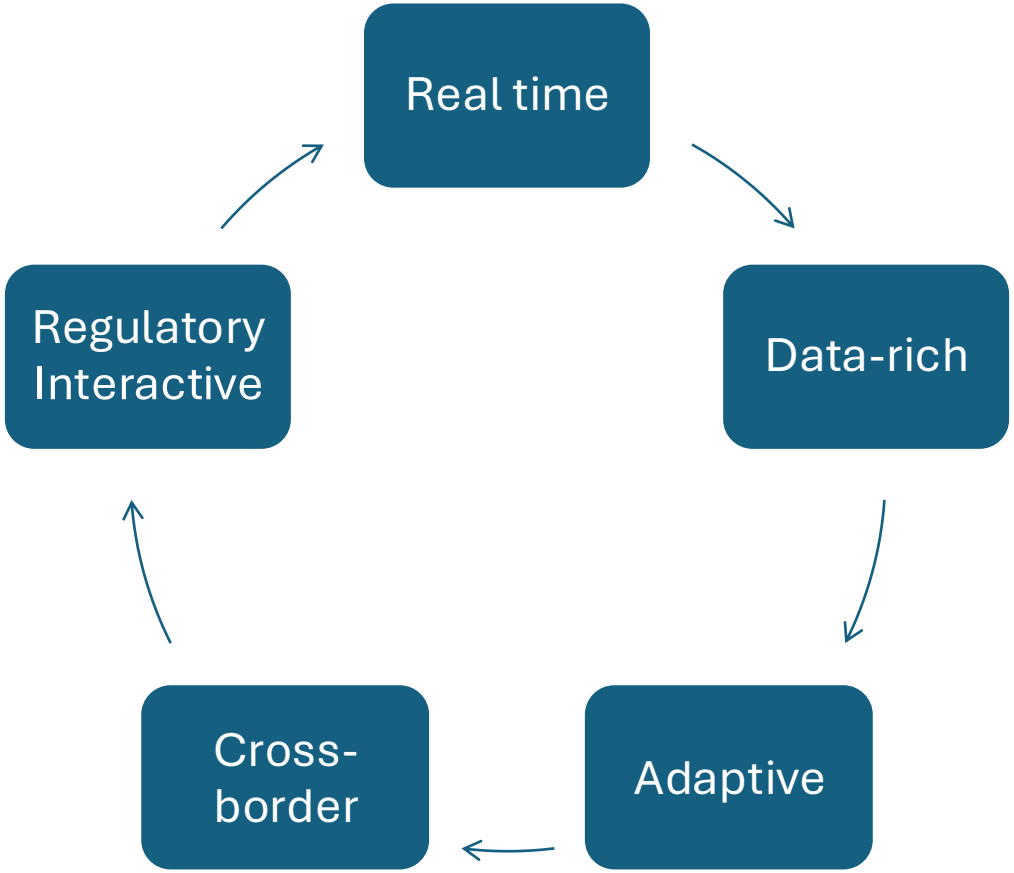
- No database is built without **CDM**
- No data is submitted without **CDM**
- No inspection passes without **CDM** traceability

Data Management may only account for 15% of the budget — but it controls 100% of the credibility.

The legacy linear trial

Outdated for most modern clinical trials

The modern model



Clinical trials of today/tomorrow cannot function if data is designed last

Clinical Operations can’t do it alone

Clinical Operations is critical for sites, logistics, timelines, vendor relationships. But when Clin Ops unilaterally decides:

- What data will be captured
- How it will be captured
- Which systems will do the capturing
- And when **CDM** is allowed into the conversation…

The cost of sidelining data management

The cost is real:

- Data chaos
- Protocol rework
- Ballooning queries
- Overburdening study sites with different technologies
- Different vendors with different set ups
- Helpdesks that are only partially involved in the whole trial
- Database builds that break reality
- Regulatory non-compliance
- …

CDM is essential from study defintion

Modern data management is about system design, data architecture, risk mitigation, and regulatory foresight. It’s the spine of the entire trial.

When **CDM** is involved from day one, trials:

- Avoid ambiguous endpoints
- Capture only what’s necessary, in the right way
- Map cleanly to submission standards like **SDTM**
- Build smarter edit checks and reduce site burden
- Enable real-time dashboards that actually mean something
- Ensure compliance from the beginning, not during a last-minute audit fire drill

Start with data and science

The false dichotomy: the protocol is a scientific document, and **CDM** is just technical implementation.

A data-informed protocol process would ask:

- How will this data point be collected?
- How will it be verified?
- Who will monitor it, and when?
- Can it be harmonized with other endpoints?

The Strategic Shift

The bigger vision. **CDM** shouldn't just be "brought in earlier." It should be repositioned entirely, from operational support to strategic architect.

- Designs data flows, not just **CRFs**
- Anticipates regulatory and technical risk
- Leads cross-functional data governance
- Collaborates with biostats, safety, and medical on real-time insights
- Champions patient-centricity through smarter data collection

Let's call this what it is: Clinical Data Architecture.

How do we do this?

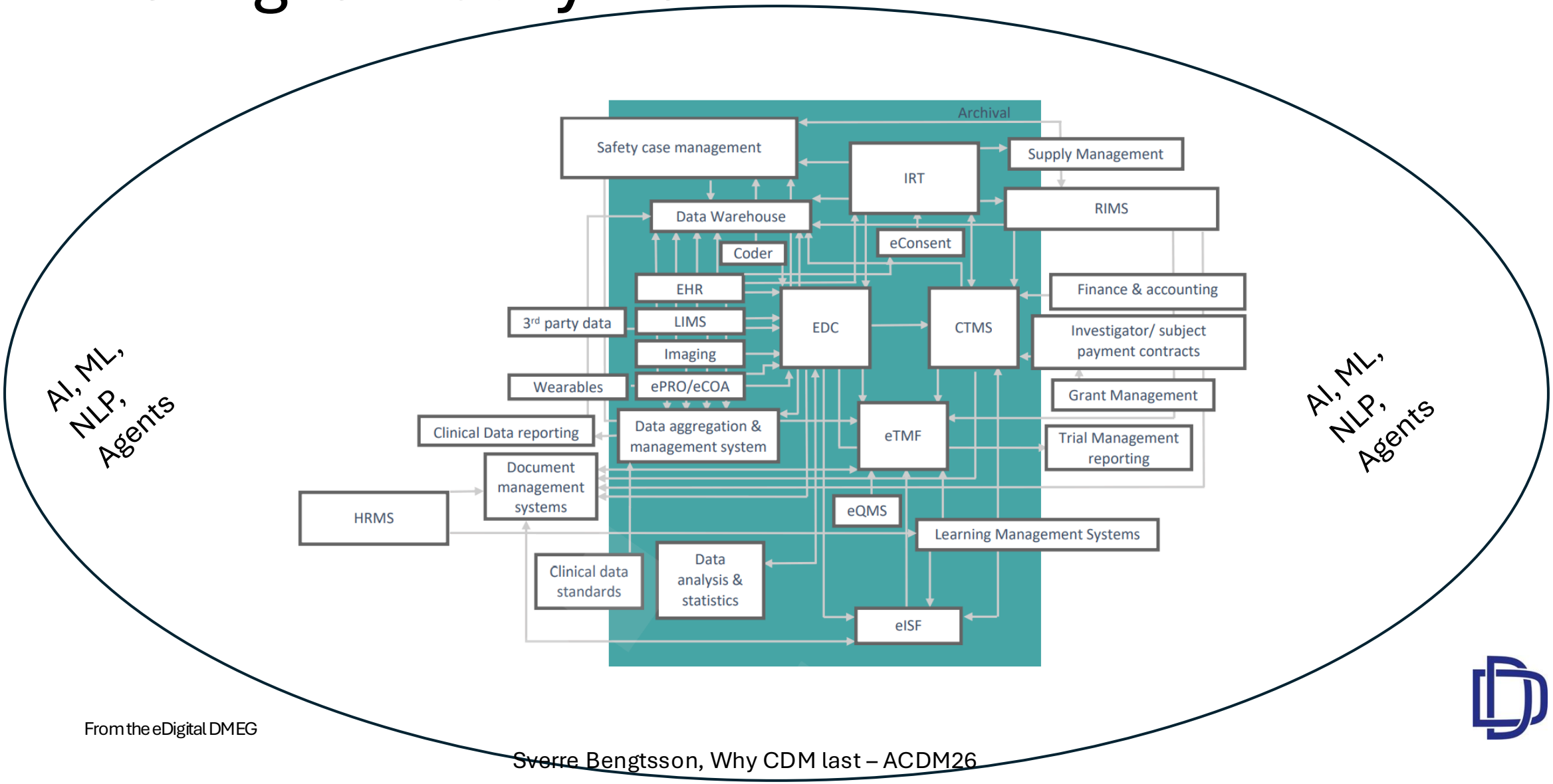
1. *Mandate **CDM** in protocol design*
   - **CDM** must have a seat at the table, not as a courtesy, but as a requirement. Every protocol team should include a data architect.
2. *Adopt Structured Protocol Tools*
   - Embrace platforms that allow protocol authoring using **USDM** or similar models — ensuring seamless translation into **EDC** and analytics systems.
3. *Shift Budget and Recognition*
   - Stop treating **CDM** as a cost centre. It is the asset manager of your clinical trial. Budget and empower accordingly.
4. *Train Ops to think in data*
   - Help other functions understand the downstream impact of their decisions — and how to collaborate with **CDM** proactively.
5. *Break the org chart*
   - **CDM** is an equal partner in trial design, strategy, and delivery, as a co-equal steward of study success

Clinical Trials are now data ecosystems

Modern studies include:

- **EDC**
- **ePRO**
- **IRT/RTSM**
- Wearables
- Imaging
- Genomics
- **EHR** data
- **DCT** platforms (e.g. eConsent, remote trial components)

The Digital Ecosystem



From the eDigital DMEG

The cost of poor data design

As established, phase III trial costs are very costly

Often in the range of 50 – 100 M€

Yet the data architecture is often finalised late in study startup

- How sites enter data
- How systems integrates
- How quickly we can clean data
- How easily we can generate submission datasets

Database delays

- Many setups are released after first patient first visit
- Or they are rushed into release and then there's a protocol amendment, and another one…
- Late releases increase likelihood of data entry delay / integrations with external data not functioning properly
- Locking timelines increase

Clean data is a design problem

We don’t need better data cleaning. We need better data thinking.

If we keep building trials backwards, treating data as an output, not an input, eventually we will keep drowning in queries, amendments, delays, and compliance nightmares.

But if we redesign the process and we put **CDM** where it belongs, we can build faster, smarter, more compliant trials that work not just for sponsors, but for patients.

So next time you hear “we’ll figure out the data later,” stop the meeting.

Because that’s exactly when the real conversation should begin.

And it is a €100 Million question after all…

Why this matters

Four major forces are reshaping clinical trials

- **ICH E6 (R3)**
- **Unified Study Definition Model (USDM)**
- Explosion of complex data sources
- The new tools for analysing and managing data: AI & Agents

ICH E6(R3): The Tectonic Shift

ICH E6 (R3) pushes the industry toward:

- Proactive quality-by-design
- Cross-functional risk management from protocol forward
- Better stakeholder collaboration
- A data-centric mindset at all stages

ICH E6 (R3) explicitly encourages early integration of quality, risk, and data considerations into protocol development, moving upstream what was previously treated as downstream.

How you define your data at protocol stage defines your trial's quality, cost, speed, and credibility.

Unified Study Definitions Model: **USDM**

- Real-time collaboration between functions
- Direct export to downstream systems (**EDC, CTMS, IRT**)
- Fewer translation errors between protocol and database
- Automated impact analysis when endpoints change
- Better alignment with **SDTM** and regulatory submission
- Faster study startup

Requires that **CDM** is involved from protocol design and onwards.

Implications for Clinical Data Management

CDMs should be involved in :

- Defining critical data
- Eliminate unnecessary data
- Designing data collection workflows; integrations, sites, patients…

The structural gap

The reality too often today:

Protocol finalised ⇒ Operational planning ⇒ Database build

From chaos to architecture

Protocol design

⇓

Unified Study Definitions Model

⇓

Automated generation of

- EDC
- CRF templates
- Visit schedules
- Integration mappings
- SDTM metadata
- Data pipelines

⇓

- Unified Clinical Data Platforms

The strategic shift

Trials are now moving form

- Document driven ⇒ Data-driven
- Data management needs to become data architects

What should CDM push for

- Data architecture review during protocol design, at latest
  - Preferably earlier, at program/company level
- Early CRF prototyping
- Adoption of structured protocols
  - Allows time savings in the whole process, increase quality

Leadership perspective

Data architecture affects:

- Timelines
- Costs
- Regulatory readiness
- Quality

Let’s avoid this

Brad Hightower LinkedIn post

Redesigning the Future, Backwards

We often build our most complex systems backwards, narrative ⇒ data.

In a world defined by complexity, compliance, and speed, the future belongs to those who can align science, operations, and data from the very first sentence of a trial design.

USDM and ICH E6(R3) is a huge help on the way.

But only a cultural shift, and a redefinition of roles, will make it real.
Let's stop asking if CDM should be brought in earlier.

Let's start asking why they're not already leading.

It is a €100 Million question after all…

Thanks!

sverre@digitaltrialconsultants.com

LinkedIn: https://www.linkedin.com/in/sverrebengtsson/