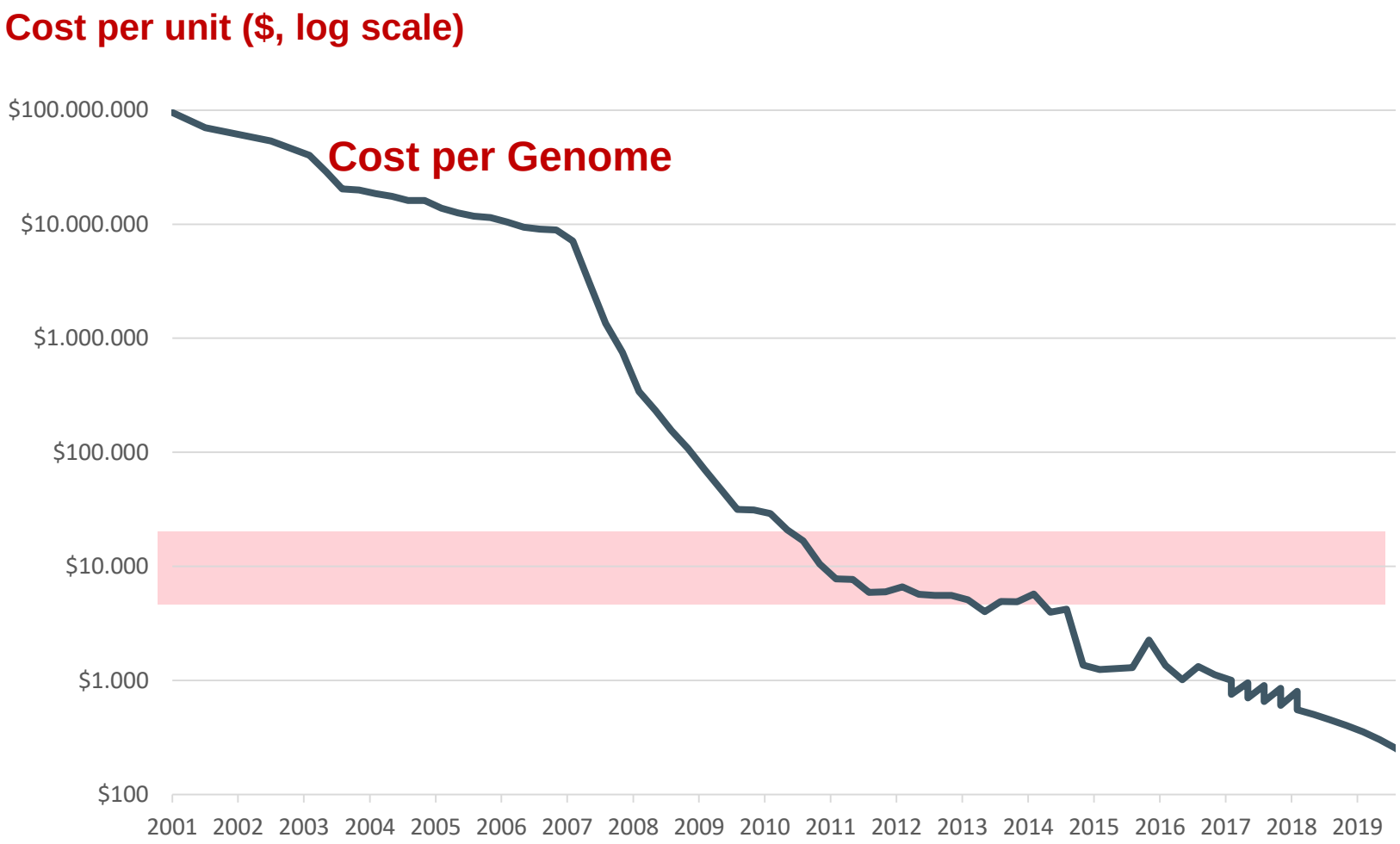
Update from Clinical Informatics Interoperability Working Group & DIGIONE pilot

Dr Piers Mahon & Prof. Giovanni Tonon

Clinical informatics is key to unlocking a research revolution as it will transform the cost of clinical and outcome research



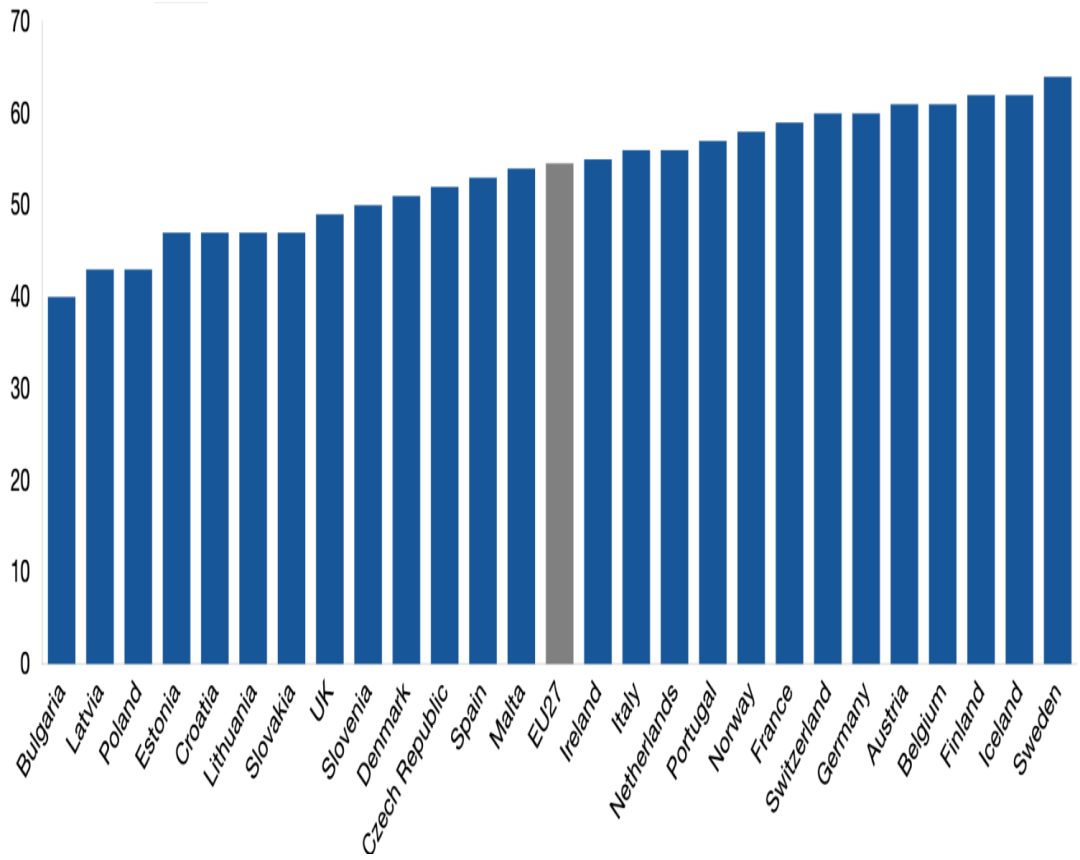
Cost per Clinical Record by eCRF

Typical cost in a large scale post-approval safety study or national registry **via manual retype methods**

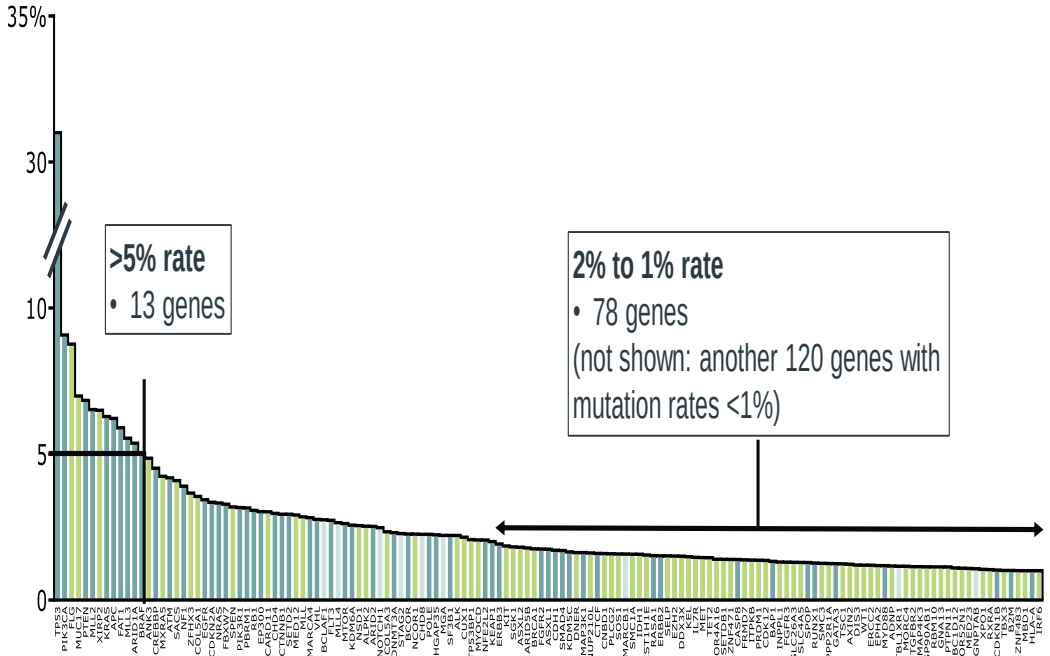
Source: Genome costs: National Institutes of Health, US; study costs: IQVIA internal data, IQVIA analysis of 2 national cancer registry budgets

With lower cost we can get the scale to tackle both care quality and precision oncology research

5 year age standardised survival (%)



Pan-cancer non-silent mutation frequency (%)



Hofmarcher, T et al. (2019) Comparator Report on Cancer in Europe 2019 - Disease Burden, Costs and Access to Medicines. IHE Report 2019:7
Mahon & Tenenbaum, 2015

DigiCore

Milan 2022: we discussed the technical challenges and key principals for an open standard based, multi-vendor networked solution

Technical challenge – hospital data is “ugly data”

The Tower of Babel

Pieter Bruegel the Elder

× We speak **multiple languages**

The Digital Institute for Cancer Outcomes Research

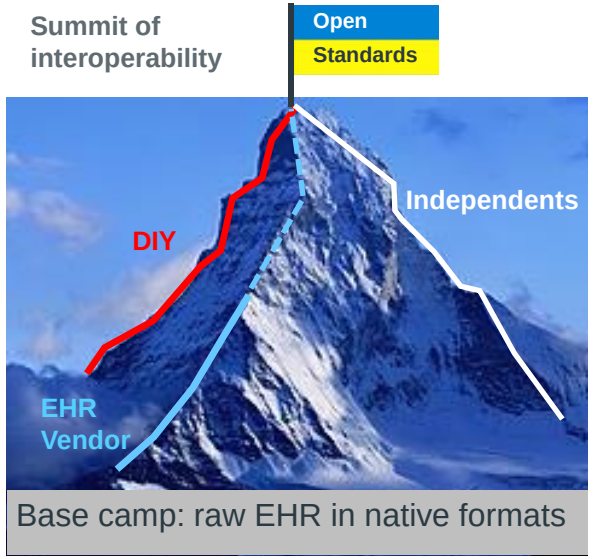
Common data models in research data repositories

Data item	Data model	Extraction “Tooling”	Conformed research data repository
• Specific medical concept that can be measured in data, a “protocol element”	• A conceptual schema for storing data elements in standardised ways in standardised uni for reliable analysis	• Software to “pull”	• The result: clean data

Cars	Empty Car Park
	

The Digital Institute for Cancer Outcomes Research

Open standards and a multi-vendor market



- DIY: Do-it-yourself using Open Source tools (The IT version of climbing with no guide)
- EHR vendor supported (e.g. EPIC, Varian, Daedelus, Cerner etc)
- But will they get beyond a Clinical Data Warehouse in a proprietary data model?

Independent specialist systems integrators (IQVIA, EHDEN accredited SI vendors)

DigiCore

The Digital Institute for Cancer Outcomes Research

We also reviewed the international consensus work we've done to define a Minimal Essential Description of Cancer (MEDOC) that most can deliver

We have built international consensus across 16 hospitals in 13 countries to define a minimum data model for cancer: MEDOC

Step 1: 5 experts, 4 countries → **Step 2: 16 Hospitals, 13 countries**

1. National cancer datasets (# items)

- UK COSD (~1200)
- German ADT (~300)
- French OSIRIS (65)
  (and chosen as start point)

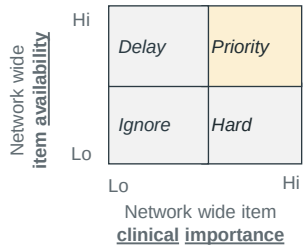
2. International outcomes research experience

- Data availability in EHR
- Importance of item in outcome research

3. Expert hypothesis modified from OSIRIS

- Senior experts in France, Italy, Germany, UK to "slim down" OSIRIS where possible, identify gaps (e.g. weight & height for cachexia or chemotherapy dosing)

4. Clinical priority / feasibility trade-offs by e–survey



5. Traditi… proc… "conte…

- Item by i… the "Har… pragmat…
- For exar… CCI co-r… co-morb… NLP imp…

MEDOC defines a minimum data standard most Cancer Centres can achieve

1. Demographics (=6)	2. Clinical Phenotype (=7)	3. Biomarkers (=3)	4. Treatment (=14*)	5. Outcomes (=6)
1.1.Date of birth (month)	**2.1 ICD10** for primary diagnosis and comorbidities (*& timestamp*)	3.1 Biomarker name *& time stamp*	**4.1 Line of therapy**	5.1 Date of death (any location, in-hospital or from national deaths)
1.2 Sex	*Note: comorbidities often will need NLP and will be optimised for only the 17 CCI co-morbidities*	3.2 Biomarker measure *& time stamp*	4.2 Anti-cancer treatment name	**5.2 Time to next treatment** (*derived*)
1.3 Weight *& timestamp*	*By implication, often not complete*	3.3 Biological sample identifier *& timestamp*	4.3 Molecule generic name	5.3 Metastasis presence / absence
1.4 Height *& timestamp*	**2.2 Charlson comorbidity index** (CCI – *timestamped and derived*)	***Notes on biomarkers:***	4.4 Start date for drug treatment	5.4 Metastasis location
1.5 Healthcare ID	2.3 Date of primary diagnosis	*We will aim to get to the same level of detail as in OSIRIS –omics for biomarkers anticipated in the guidelines in 2024 from the drug pipeline (even if from NLP / OCR)*	4.5 Treatment dose	5.5 Date of last visit/follow-up
1.6 Legal basis for data processing, *e.g.* consent or non-opposition	2.4 Method of primary diagnosis	*Tests formats will cover:*	4.6 End date for drug treatment	5.6 Vital status (derived)
Notes: legal basis and a healthcare ID are likely to be in national schema, and may be multi-concept in some counties or settings	2.5 Performance status (e.g. ECOG, Karnofsky) *& timestamp*	- *Core routine Blood biochemistry commonly used in cancer* - *IHC – including HER2+ low* - *FISH* - *Somatic mutations, likely as amino acid change or similar* - *Germline, e.g. BRCA1*	4.7 Radiotherapy type (e.g. procedure code of treatment)	*Note: routine death registry linkage is not allowed in some European countries, and will require careful design of delivery of 5.1 and 5.6*
Key: Yellow = item must follow local / national rules or norms Red = Item not in original OSIRIS starting 65 concepts *Italics* = implementation notes	2.6 Disease stage *& timestamp* (e.g. TNM, size, node and metastasis)		4.8 Radiotherapy Start date	*** Notes on Treatment** In some countries we anticipate that claims data is not accessible, only the core EHR which may need NLP routines to extract dates. Where claims data accessible, dates may be derived via timestamps
	2.7 Histological cell type & *timestamp* (e.g. ICD-O-3)		4.9 Radiotherapy dose	
	Note: we anticipate multiple cancer specific schema for stage and cell type and will phase implementation		4.10 Radiotherapy end date	
			4.11 Surgery type (e.g. procedure)	
			4.12 Surgery date	
			4.13 Participation in clinical trial	
			4.14 Date of trial consent	

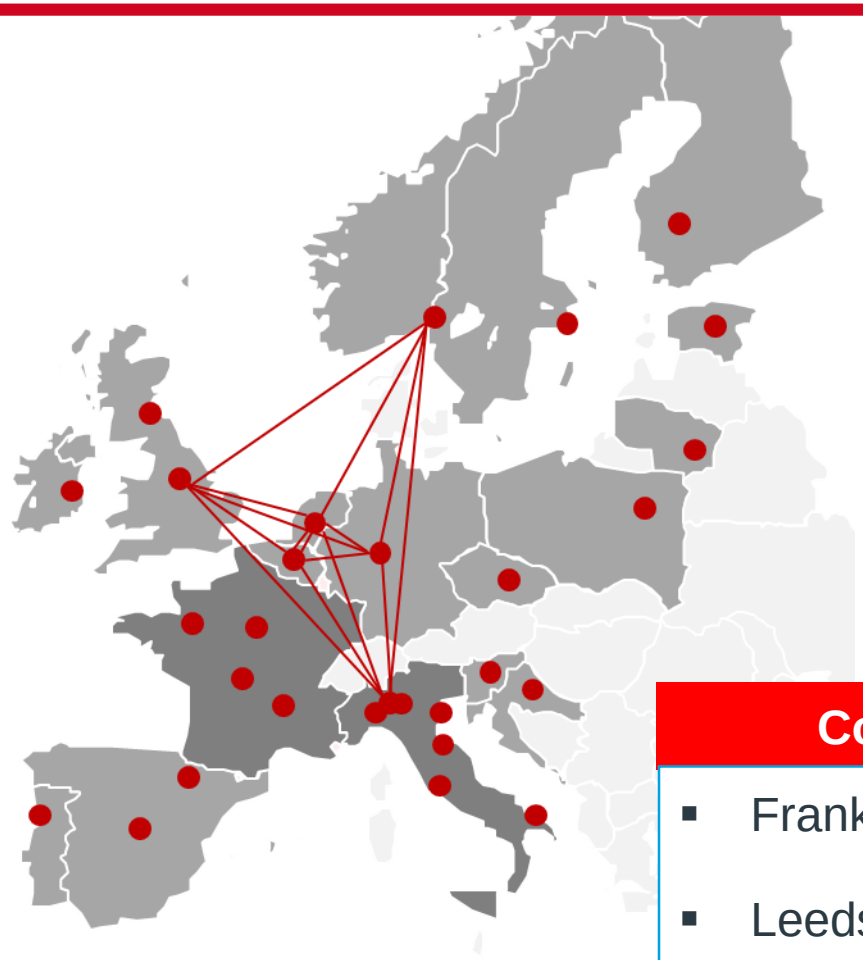
DigiCore 5

The Digital Institute for Cancer Outcomes Research

DigiCore

2022 we also announced who secured seed funding to build a prototype for the Digital Oncology Network for Europe (DigiONE – Pilot)

Objectives for DigiONE Pilot



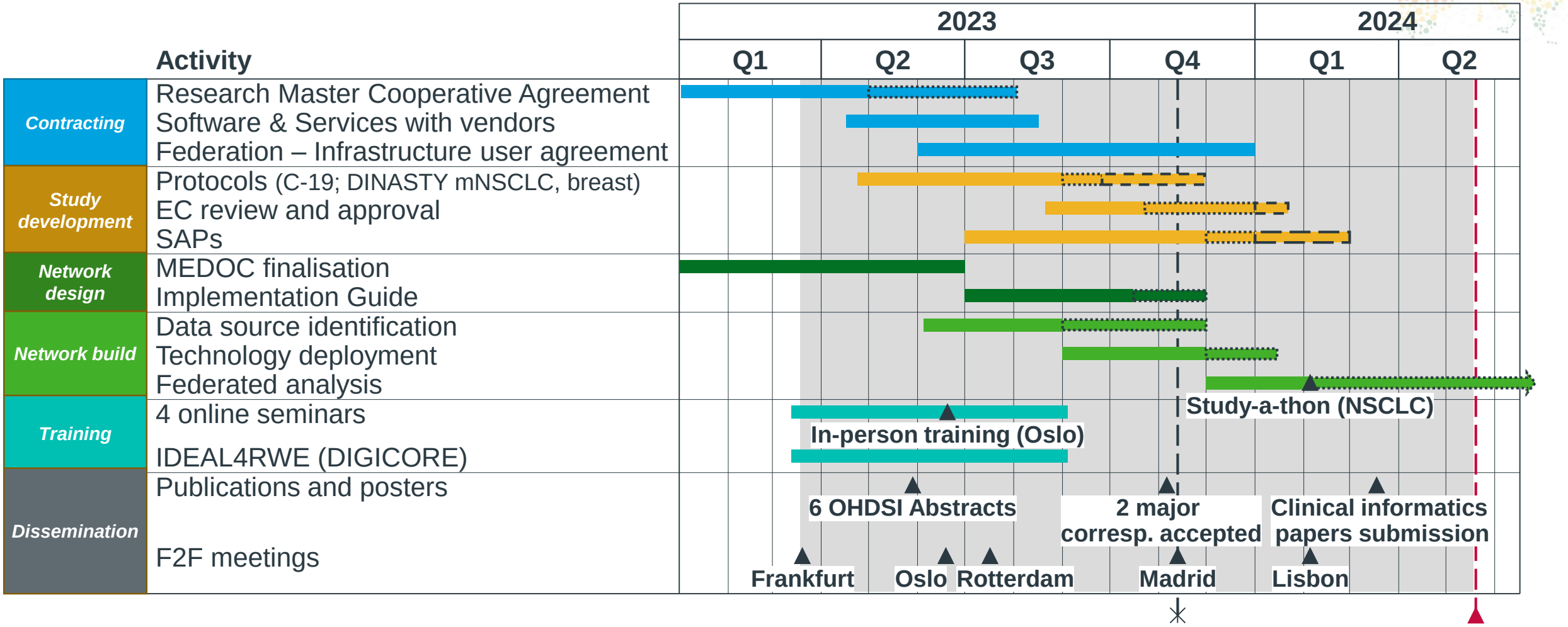
1. Define a **scalable common international minimum dataset for cancer outcome research** in precision oncology / OMOP
2. **Achieve interoperability and high data quality** on that dataset between 6 centres across Europe under GDPR
3. **Federate those centres** to allow complex protocolized research, such as disease natural history + outcome studies
4. **Demonstrate "fully digital" real world evidence possible** in a broader range of European countries to attract funding

Competitive process: 16 sites applied => 6 centres selected

- Frankfurt (Janne Vehreschild)
- Leeds (Geoff Hall)
- Maastricht (Andre Dekker)
- Oslo (Åslaug Helland / Sissel Jor)
- St Luc (Cédric Van Marcke)
- San Raffaele (Giovanni Tonon)

**with funding from Illumina and IQVIA*

This is where we are in the pilot



Study key

Extension for mNSCLC

Extension for breast

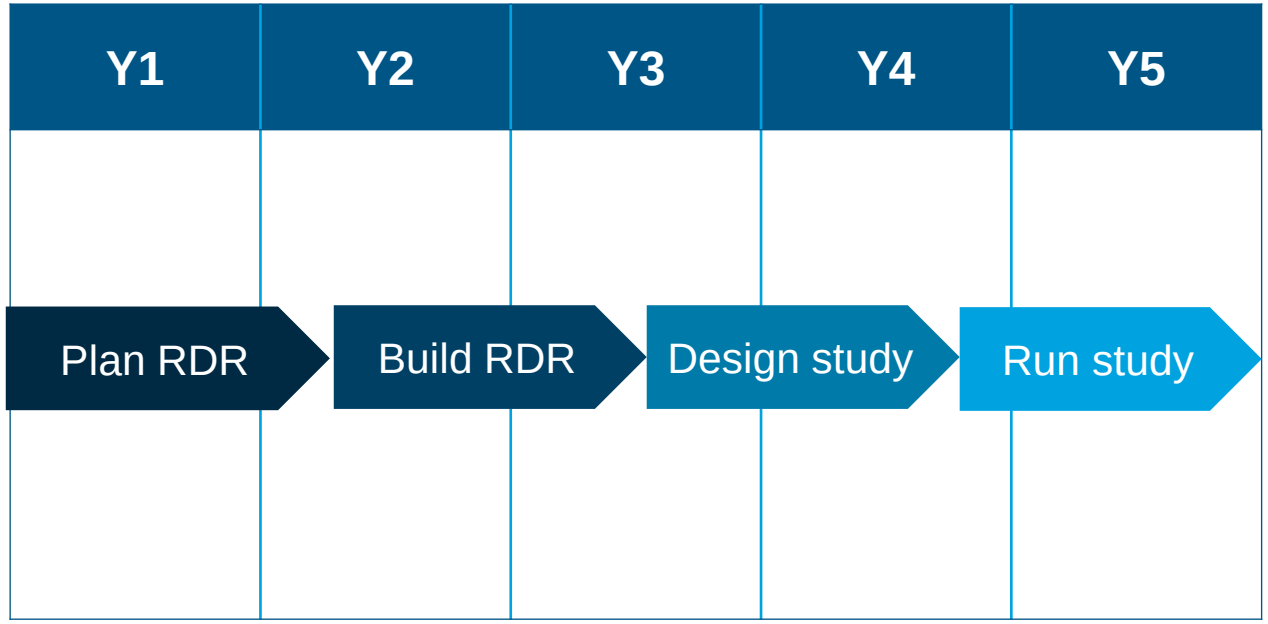
Core project duration from start to study output ~14 months

So what have we achieved in 2023?

- ❑ MEDOC, our minimal dataset, has been defined in Cancer OMOP
- ❑ 58 page MEDOC implementation guide developed so any hospital can copy us (in beta)
- ❑ 18 contracts signed to allow build (with another 3 pending for federation)
- ❑ 6 hospitals well into build, with multiple vendors and approaches
- ❑ 4 OMOP studies (C19, mBC, mNSCLC, EOC) developed to test MEDOC
- ❑ Data ready for first study in >3 centres now (C19), will be ready for mNSCLC mid-January
- ❑ 6 abstracts accepted to OHDSI Europe summer 2023 (the main OMOP conference)
- ❑ 2 major correspondence pieces in press (but under embargo)
- ❑ Funding from ESMEIA / I3 scheme for another 15 hospitals to achieve “OMOP + NLP”

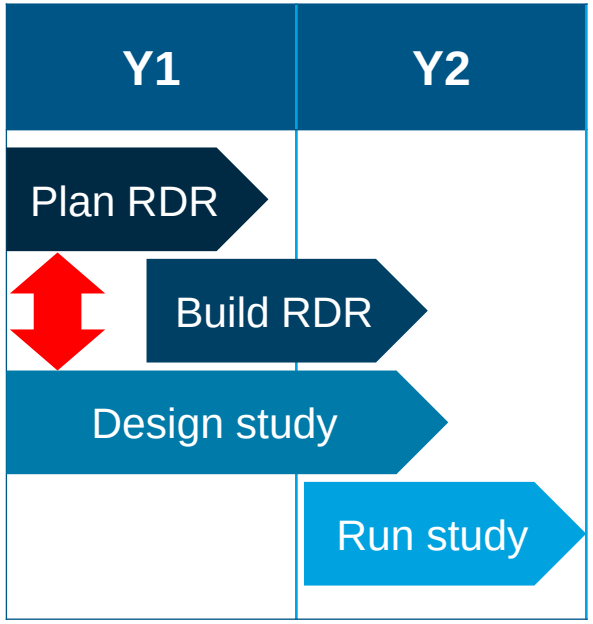
Lesson 1: DigiONE's minimalist, study focused, and pragmatic approach works and is faster than traditional approaches to Cancer OMOP

Early Cancer OMOP pilots ran in series
3-5 years



VS.

DigiONE runs in parallel
1.5-2 years



2-3x faster by tightly coupling study design and RDR specification

RDR, research data repository

Lesson 2: we can get interoperability with multi-vendor approaches (just look at the diversity of approaches the hospitals took)

Hospital	Unstructured data approach	Structured data approach	Approach to build
Frankfurt	• Averbis NLP in German	• Kairos ETL OMOP conversion	• Company
Leeds	• *May not need – considering NLP options for biomarker**	• Hospital built ETL, repurposing a non-Cancer OMOP version	• Hospital / open source
Maastricht	• CTcue NLP in Dutch	• OMOP conversion from CTcue data model	• Company + open source
Oslo	• Simple text mining for semi-structured data (no Norwegian NLP)	• Hospital built ETL, with advice from EdenceHealth	• Hospital / open source
Saint-Luc	• EarlyTracks/ manual (in French)	• EdenceHealth / hospital ETL OMOP conversion	• Company / Hospital
San Raffaele	• CGP – home developed eCRF solution (manual retype)	• Hybrid - Hospital eCRF tools integrated with IQVIA Health Data Research Platform	• Hybrid

*Leeds' EHR is very structured, and so for most MEDOC NLP is not required

DigiCore

Lesson 3: OMOP is so flexible we need a detailed implementation guide to ensure hospitals and vendors can “build the same thing”

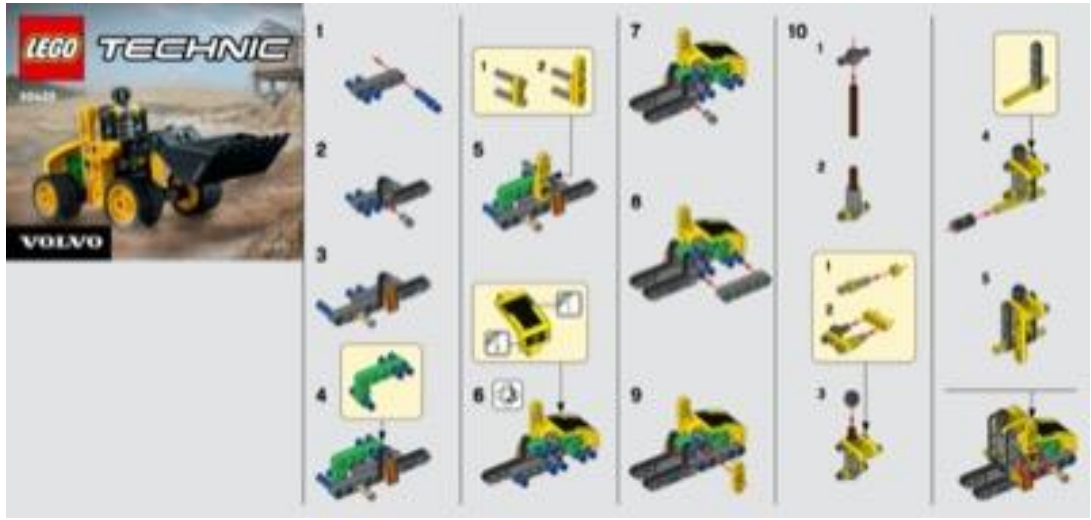
OMOP is a very flexible data model

- Designed for registry, claims or EHR data
- Lots of components, some that duplicate
- Very flexible individual tables
- Highly customisable implementation
- ETL “expert only” documentation

…Like a big pile of Technic Lego

Needs an **implementation guide**

- Assumes we start with hospital EHR data with non-expert teams on 38 MEDOC variables
- Agrees how we will build key tables and concepts so we “end up in the same place”
- 58 pages, 12 tables for local ETL planning



The Lego instructions to get the same model

Our panel today

What is it like to build a Cancer OMOP instance in a hospital?

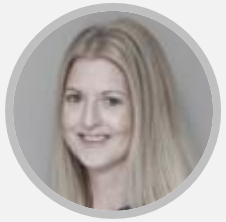
What does it allow us to do?

Leeds Teaching Hospitals NHS Trust

Stelios Theophanous

LTHT has a homebrew EMR and currently hosts a non-cancer OMOP instance they are extending for MEDOC

Oslo University Hospital



Elisabeth Ross

OUH is building its own OMOP instance following consulting advice from edenceHealth

edenceHealth

Lars Halvorsen

edenceHealth is collaborating with Saint-Luc on their OMOP implementation of MEDOC

San Raffaele Scientific Institute

Giovanni Tonon

Overseeing pilot implementation in San Raffaele, and coordinating large scale Europe initiatives that could use such an infrastructure

What are the “take homes” for senior hospital leadership?

Care quality is primary focus

- **Improving care quality** via guideline benchmarking and outcome research
- Designed to **engage front line clinicians** (not just the research focused)
- Set up as an **internal service to all** (not a restrictive research dataset)

Right hospital leadership and team key

- **Senior sponsor** to solve roadblocks like contracting (head of R&D, head of IT, CIO etc)
- **Project manager** to drive coordination day to day (across contracting to clinical engagement)
- **Data analyst** who understands “what data lives where and how to access it”
- **SQL engineer** to learn and then implement OMOP ETL tools
- **Clinician or nurse to advise on data meaning** (via study implementation in their disease)

Doing OMOP together = faster & better

- **Design for study interoperability together from the start** (Implementation guide)
- **Peer to peer advice** helps turbocharge new teams
- Often **“someone has solved that problem before”** (don’t reinvent the wheel)

Just get started and then hunt funds to upgrade

- **Any hospital can do C19 study and build a minimal OMOP instance on their own**
- **2024 OMOP ETL training programme** to help you get started
- Solving **unstructured data is slower,** harder and more expensive
- **Full MEDOC implementation needs grants** and often vendor support

After the CCI4EU sesson we split for the final session of the day

Session A

Topic: DIGICORE 2024 clinical research priorities and working groups **discussion**

With: Adriana Albini & Piers Mahon

For: more clinically focused people

Location: Stay here

Or

Session B

Topic: Clinical informatics tools in open source to help DIGICORE members digitise – selected **poster** presentation

With: Alberto Traverso & Xose Fernandez

For: more IT /data science people

Location: go outside with Alberto & Xose

BACKUP

Technical challenge – hospital data is “ugly data”

The Tower of Babel

Pieter Bruegel the Elder

- ✕ We speak **multiple languages**
- ✕ We **practice medicine differently**
- ✕ Most of the data in a hospital is **unstructured**
- ✕ **Critical data is missing**
- ✕ We have **bespoke IT systems** and vendors in every hospital with **proprietary data formats**
- ✕ We have **different clinical coding standards and claims systems** in every country
- ✕ We have different **national care quality** agendas
- ✕ We have **different national (and local) interpretations of GDPR** & privacy requirements

DigiCore

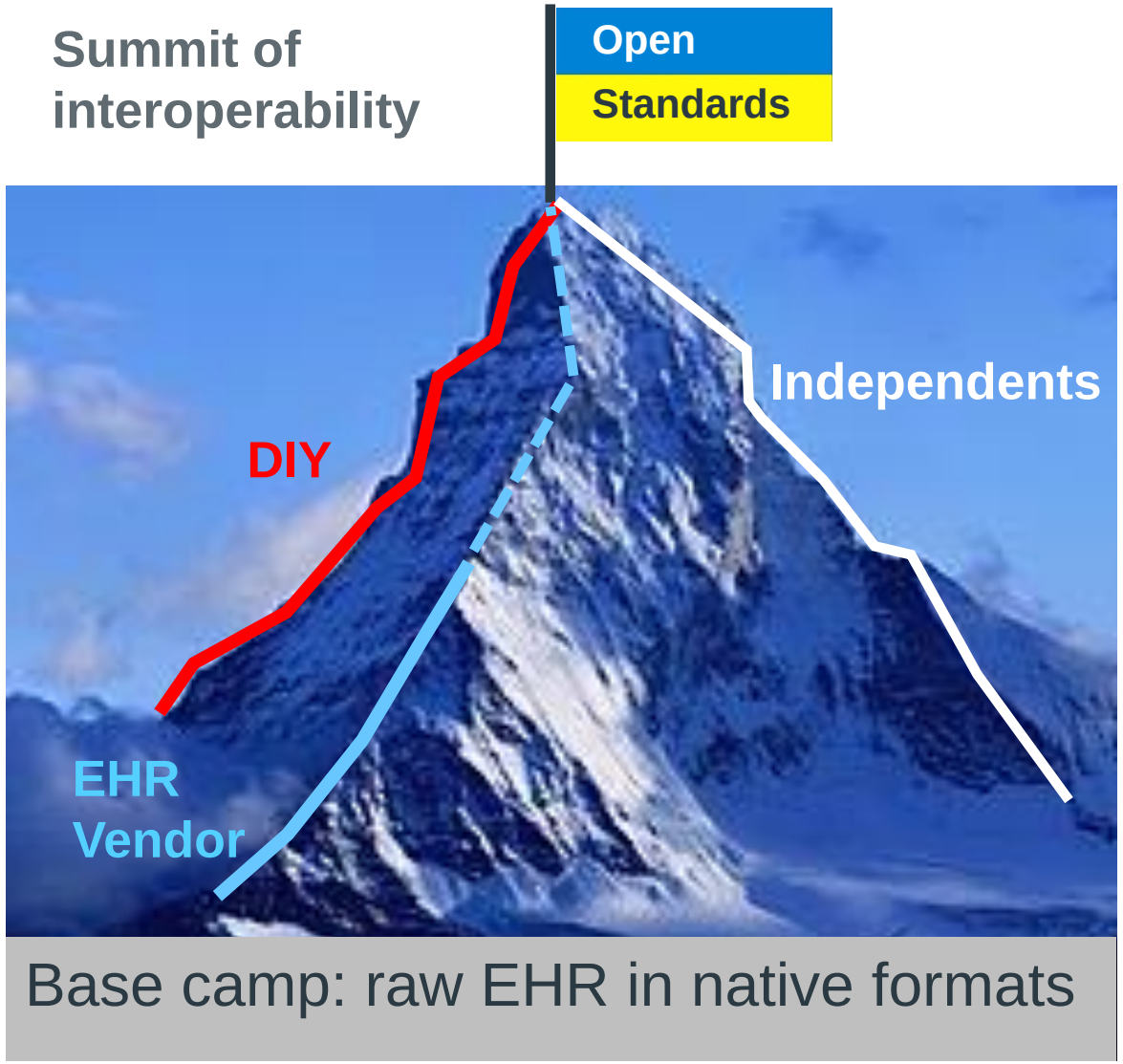
Common data models in research data repositories

Data item	Data model	Extraction "Tooling"	Conformed research data repository
• Specific medical concept that can be measured in data, a "protocol element"	• A conceptual schema for storing data elements in standardised ways, in standardised units for reliable analysis	• Software to "pull" data from existing messy storage, clean it, standardise and "push" into a data model	• The result: clean data in a standardised format in a robust data model held under hospital control for research use

An analogy…

Cars	Empty Car Park	Car Park Attendants	Filled, Neat Car Park
			

Open standards and a multi-vendor market



- DIY: Do-it-yourself using Open Source tools (The IT version of climbing with no guide)
- EHR vendor supported (e.g. Epic, Varian, Dedalus, Cerner etc)
- But will they get beyond a Clinical Data Warehouse in a proprietary data model?

Independent specialist systems integrators (IQVIA, EHDEN accredited SI vendors)

An IT deployment

Data item	Data model	Extraction "Tooling"	Conformed research data repository
• Specific medical concept that can be measured in data, a "protocol element"	• A conceptual schema for storing data elements in standardised ways, in standardised units for reliable analysis	• Software to "pull" data from existing messy storage, clean it, standardise and "push" into a data model	• The result: clean data in a standardised format in a robust data model held under hospital control for research use

An analogy…

Cars	Empty Car Park	Car Park Attendants	Filled, Neat Car Park