

Digital Data Flow (DDF) Solution Showcase

July 10, 2025

Presenting Organizations:
CRScube & studyOS

Agenda

Topic

Welcome, Background, Webinar Logistics & Ground Rules

Presenting Company 1: CRScube – 30 mins

Presenting Company 2: studyOS – 30 mins

Q & A with Panelists

Closing

Today's Presenters

CRScube & studyOS



**Sylvain
Berthelot**

VP, Strategy and
Marketing

CRScube



Robert Hay

Marketing and
Solutions Consultant

CRScube



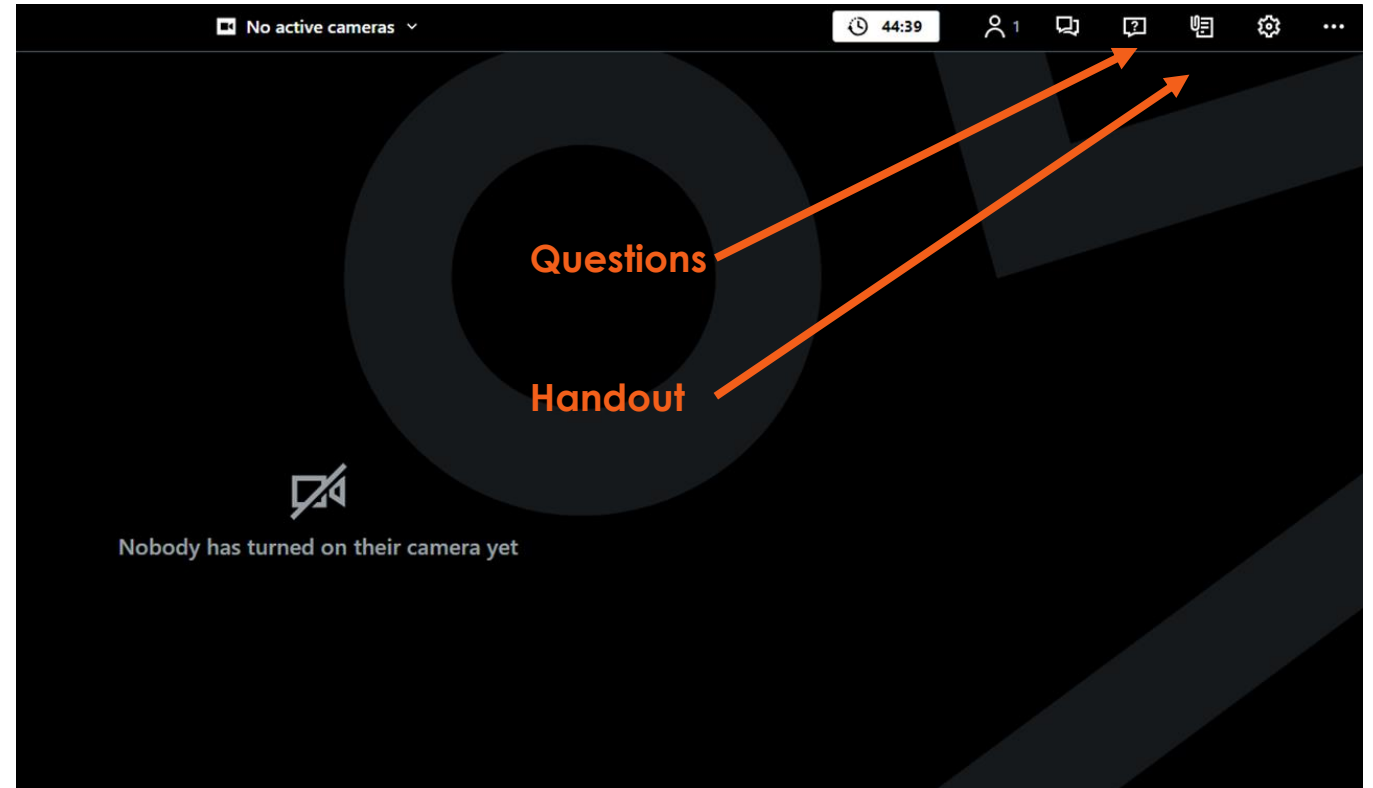
**Ellis Hiroki
Butterfield**

Chief Executive
Officer

studyOS

Logistics for the Webinar

- **All participants will be muted for this call.**
- **For audio:** Connect to audio to listen to presentations via your computer or phone
- **To submit a question to the presenters:**
 - Type your question in the Questions panel and click Send.



Reminder: This webinar may be recorded in whole or in part.

Ground Rules

- **We want to make this discussion helpful and answer as many of your questions as we can, so here are some quick ground rules:**
 - Participation is voluntary, as is using TransCelerate assets/tools
 - The responsibility for compliance with laws and regulations is owned by the solution adopter
 - You don't have to identify what company you work for
- **Things we would ask you not to post questions on:**
 - For clinical trial sponsors, what vendors/sites/CROs a company is working with or not working with
 - For tech companies, vendors, CROs, & others, what pharma companies you work with or don't work with
 - Any issues/criticisms companies have with any vendors, tech company, sites, CROs, or sponsors
 - Future and long-term development plans
 - Anything related to pricing or costs -- what you pay for the purchase of or receive for the sale of any goods or services
- **We can't answer questions about:**
 - Specific vendors or other business partners with whom member companies are working
 - Costs of using/implementing TransCelerate assets/tools
 - Which member companies are using or going to use any TransCelerate solution or any commercial product or service

TransCelerate is a Not-for-Profit Entity Created to Foster Collaboration

Our mission is to collaborate across the global biopharmaceutical R&D community to identify, prioritize, design, and facilitate the implementation of solutions designed to drive the efficient, effective, and high-quality delivery of new medicines.



About This Webinar Series

TransCelerate and CDISC are co-sponsors of this webinar series:

- TransCelerate leads the Digital Data Flow (DDF) initiative
- CDISC develops the USDM data standard for digitized protocols



Objective(s)

- Bring together DDF solution providers, sponsors, and industry stakeholders to witness innovative solutions
- Provide a platform to showcase different approaches to protocol digitalization (utilizing the USDM standard)
- Foster knowledge sharing relative to protocol digitalization



CDISC Standards

By bringing together a global community of experts to develop and advance data standards of the highest quality, CDISC creates clarity in clinical research.

Together, we enable the accessibility, interoperability, and reusability of data for more meaningful and efficient research that has greater impact on global health.

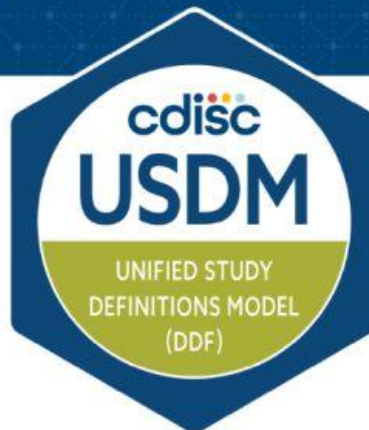


- Consensus-based standards development
- Standards for clinical and translational research
- Standards are freely available at www.cdisc.org
- IP Policy ensures open standards
- Ongoing global research support in the Americas, Europe, Japan, China, India, Korea and other regions
- Standards downloaded in 90+ countries



NEW!

USDM v4.0 Now Available



Includes:

- USDM Logical Model
- USDM Controlled Terminology
- USDM API
- USDM Conformance Rules
- USDM Implementation Guide



Alignment with ICH M11



Support for complex studies, interventional & observational studies, and medical devices



Maximise content re-use and support for multiple document templates



Model Extension mechanism to provide flexibility

Digital Data Flow

[Overview](#)[What is the USDM](#)[Participate](#)[Webinars](#)[Versions](#)[FAQ](#)[Contact Us](#)

Versions	Release Date	Information
USDM v4.0	3 June 2025	<p>USDM Implementation Guide - An implementation guide to support usage of the USDM</p> <ul style="list-style-type: none">• USDMIG v4.0 <p>USDM Logical Model</p> <ul style="list-style-type: none">• in .png (picture) format• If you use Enterprise Architect, there is also a .qea file available as well as a .xmi version for other UML software <p>Conformance Rule Specifications to support USDM v4.0</p> <ul style="list-style-type: none">• USDM v4.0 Conformance Rule Specifications <p>API Specifications</p> <ul style="list-style-type: none">• json format• yaml format <p>Controlled Terminology - supporting the UML/API</p> <ul style="list-style-type: none">• Controlled Terminology - supporting the UML/API <p>Comments from the Public Review</p> <ul style="list-style-type: none">• USDM v4.0 Public Review Comments



https://www.cdisc.org/ddf#standard_Versions

<https://github.com/cdisc-org/DDF-RA/tree/v4.0.0>

CDISC USDM Training: In-Person & Virtual Options



- Full-day USDM Training course
- Foundational understanding of USDM standards
- Real-world practical applications
- Covers key areas
 - study setup
 - activity scheduling
 - Intervention planning
 - Integration with downstream systems
- Certificate of Achievement and digital badge



Registrations OPEN!

Registration is now open for the DDF In-person Event for 2025 (2 Days, 2 Locations)

“DDF: MISSION POSSIBLE”

Dates: September 24th and 25th of 2025

Locations: New Jersey, USA and Basel, Switzerland

(Registration link posted in this meeting's chat window)

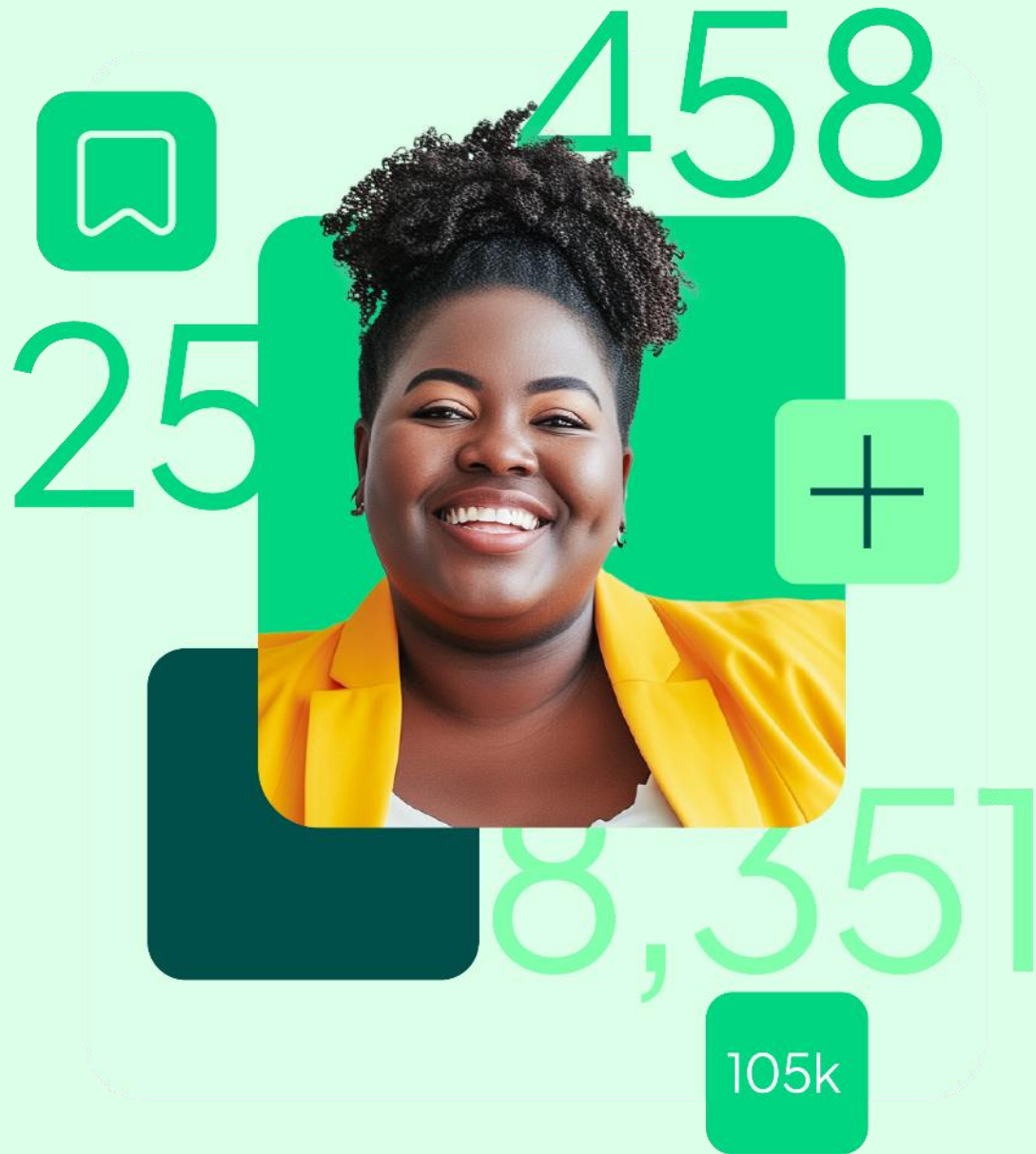
Solution Providers are also invited to present a poster during this event!

If you are interested in submitting a poster, please email us at:

ddf@transceleratebiopharmainc.com

Solution Showcase Presentations

CRScube Presentation



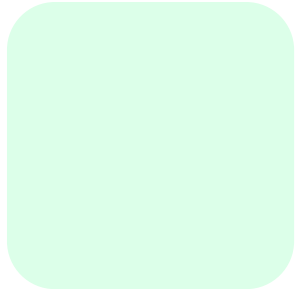
Embracing USDM standards to accelerate eClinical technology setup

CRScube's eClinical platform

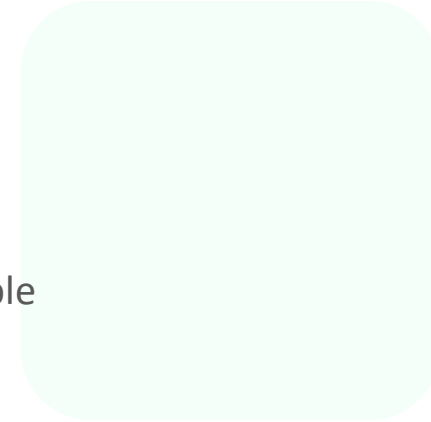
Team introduction



Sylvain Berthelot
VP, Strategy & Marketing
20 years industry experience
15 years at eClinical technology vendors
Advocates for making clinical trials more accessible
Patient advocacy podcast *On One Condition*



Robert Hay
Marketing & Solutions Consultant
5+ years working in clinical trials and pharma
Believes in leveraging AI and automation
to accelerate access to new treatments



Agenda

1

Company introduction

2

DDF implementation goals

3

Development to date

4

Future outlook



01

Company introduction

Who we are, what we stand for



CRScube introduction

Global eClinical solution provider



110 employees

Offices in Korea, Japan, China, UK and US

In numbers

Established in 2011

Over 5,800 trials supported

#1 eClinical vendor in Korea

#3 eClinical vendor in Japan

Clinical data, simply

Our mission



Make clinical trials more accessible

Not that simple!

We welcome the increased focus on standardization and knowledge sharing.

We would benefit from better global awareness, especially in Asia.

The steps of standardization

Moving from homemade to homemade+



Homemade

We implemented ways to reuse functionalities between studies:

- Pre-validated library
- Build-your-own library
- Study to study copy

Standards are confined within each client's own definitions and usage.



Homemade+

We implemented CDISC and SDTM standards in our platform.

We are now in the process of developing cubeDDF, a solution leveraging the USDM.

We also wait eagerly for the release of ICH M11 guidelines in the fall.



Global awareness

Our existing clients in Asia are not aware of the DDF initiative and its benefits.

It is challenging for us:

- Lack of buyer interest
- No concrete use case

It reduces our ability to provide innovative solutions based on DDF.

02

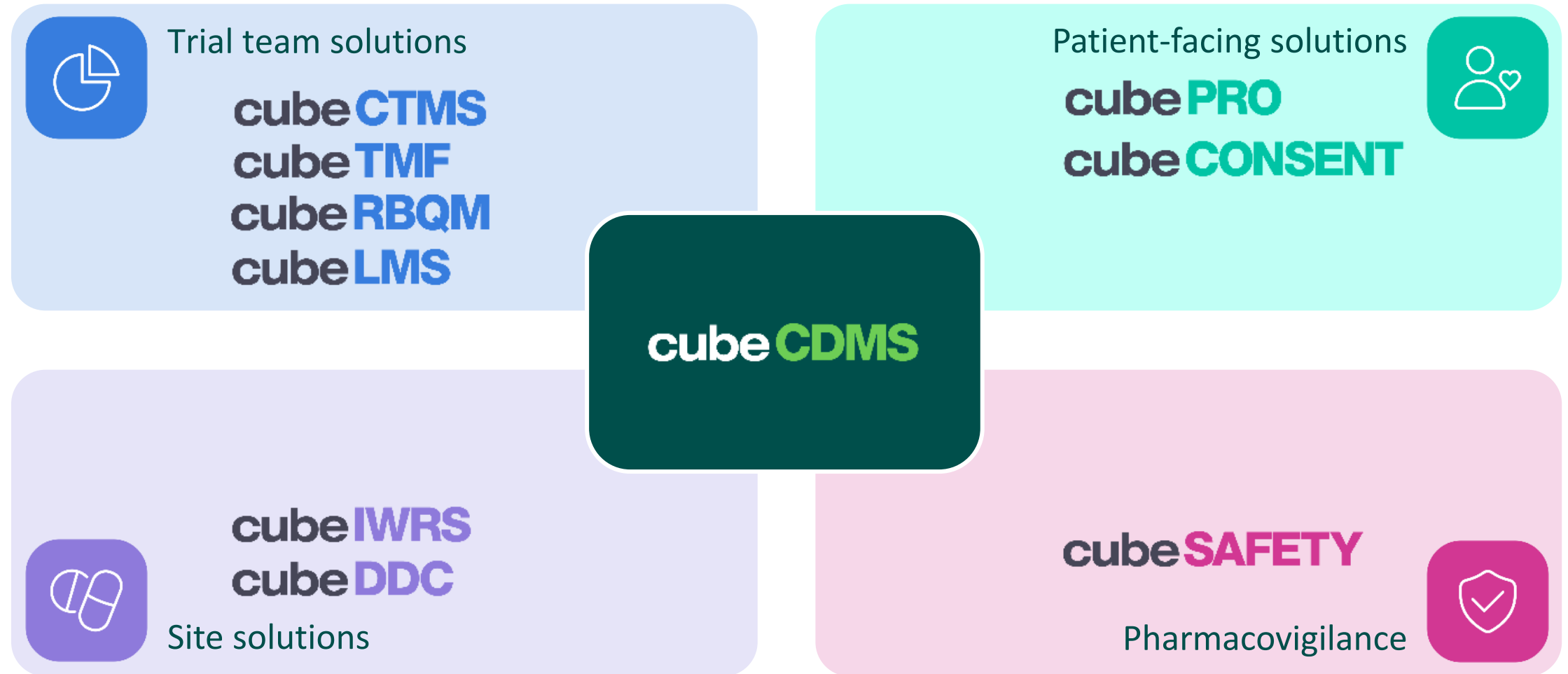
DDF implementation goals

What we aim to achieve



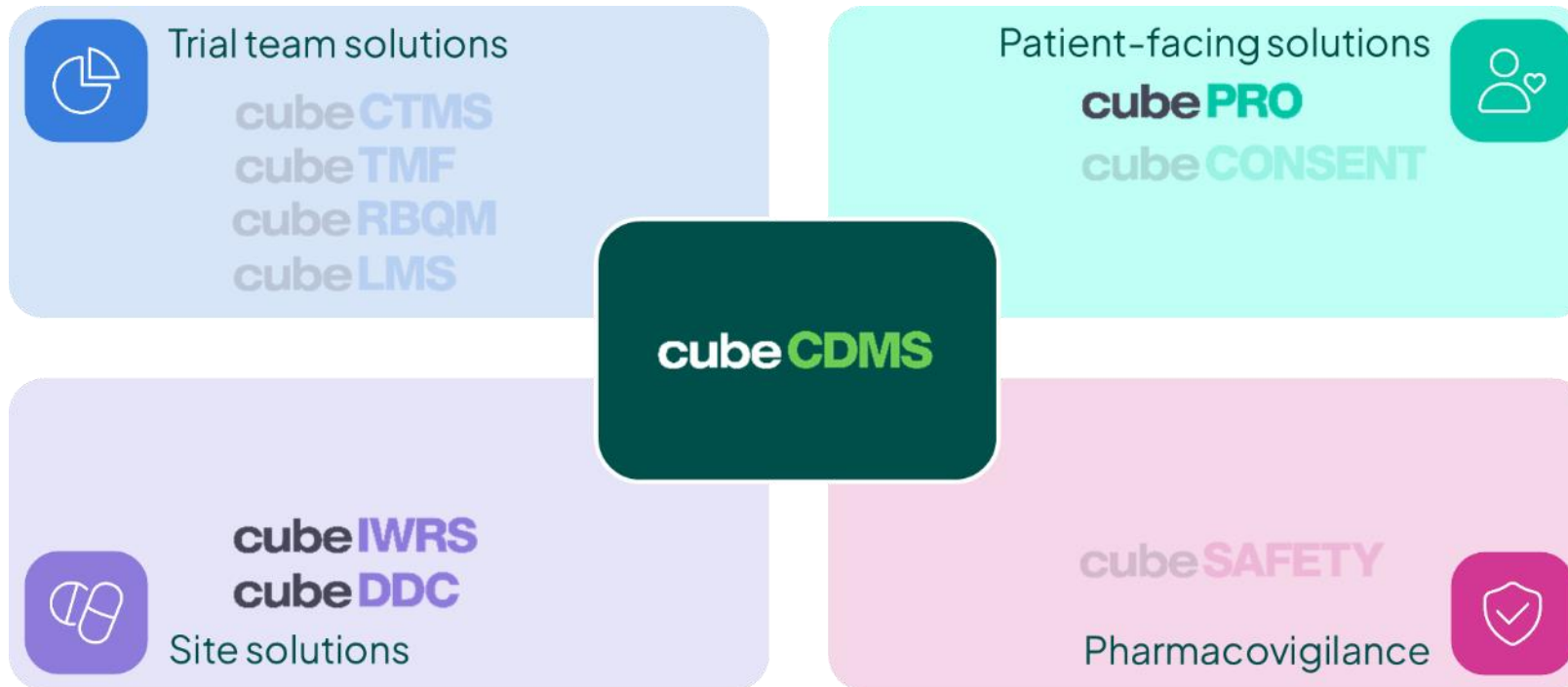
CRScube platform

Unified set of solutions for various users



CRScube platform

Impact of our developments with USDM



?

Why those solutions?

Focused on data capture from sites and patients.

Single data structure.

Configured based on protocol design.

Share the same tool for study-level configuration, cubeBUILDER.

Objectives of cubeDDF

Increase automation, platform for future innovation



Reduce effort

The setup from a digital protocol will reduce human-led configuration and validation.

Human design will focus on edit checks, items not in the protocol (supply chain etc.).

SDTM conversion will be semi-automated.



Increase quality

Less room for errors with eClinical solutions built from a digital protocol.

Protocol amendments will be implemented automatically.

Using a separate module (cubeDDF) reduces the impact to existing configuration.



Accelerate innovation

As the scope of the DDF initiative expands, our clients will benefit from additional automation and standardization.

cubeDDF will leverage future standards and guidelines updates.

cubeDDF principles

Context



Digitized Protocol



Study Definitions
Repository



Clinical Operations
Systems

cubeDDF principles

Context

Other solution, not built
by CRScube



Digitized Protocol



cubeDDF



Study Definitions
Repository



cubeCDMS
cubeIWR
cubeDDC
cubePRO



Clinical Operations
Systems

cubeDDF principles

Two initial use cases

Use case 1: study configuration



Use case 2: EDC to SDTM conversion



03

Development to date

What we have achieved



Study configuration use case

Placeholder

New screen recording will be provided shortly.



The screenshot shows a web form for signing in to cubeDDF. At the top is the cubeDDF logo, which consists of a purple square icon with a white cross-like pattern followed by the text 'cubeDDF' in a bold, sans-serif font. Below the logo is the heading 'Sign In'. There are two input fields: the first is labeled 'Email' and contains the text 'shyu@cs.cube.io'; the second is labeled 'Password' and contains a series of dots, indicating a masked password. To the right of the password field is a small icon of an eye, typically used to toggle password visibility. Below these fields is a wide, solid purple button with the word 'Login' in white text. A mouse cursor is visible at the bottom center of the slide, pointing towards the bottom of the form area.

Achievements

Study configuration use case

Use case 1: study configuration



What we have done

- JSON & Excel file uploader & converter
- USDM-compatible data structure
- Visualisation in cubeDDF
- Version controller

Challenges

- Studying the details relating to the USDM and the DDF initiative
- Interpreting the equivalent of Biomedical Concepts in our EDC

Requirements

SDTM conversion use case

Use case 2: EDC to SDTM conversion



What we need to do

- Data replication from cubeCDMS
- Version controller
- EDC to SDTM conversion
- JSON file export

Challenge

- Until we have completed the study configuration use case, we are not able to proceed with the SDTM conversion

04

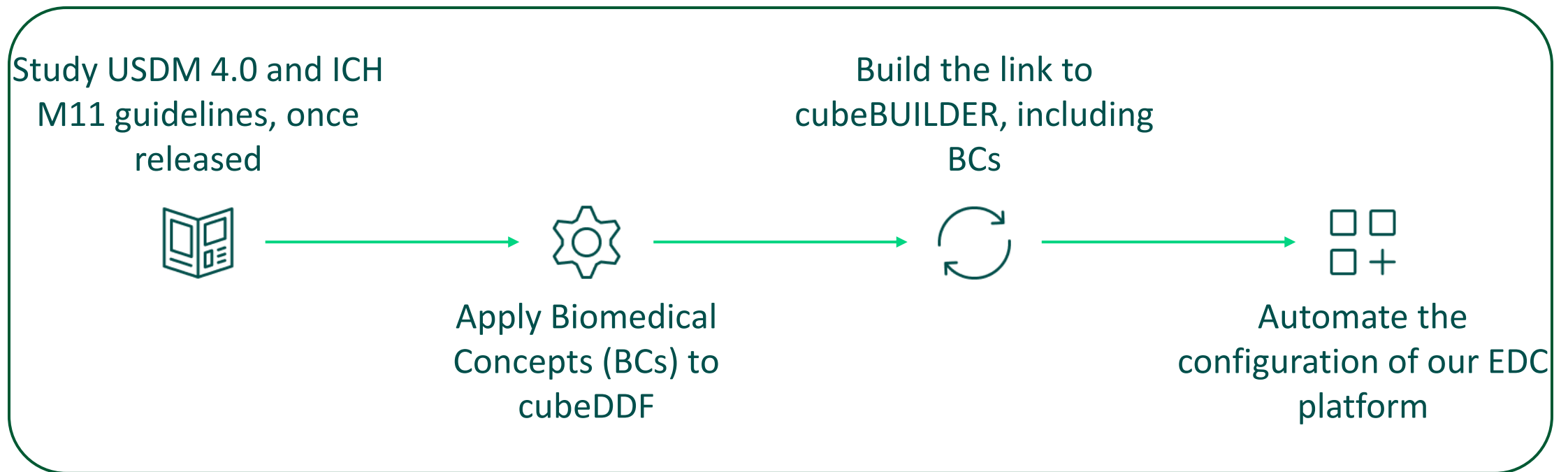
Future outlook

What is left to do



Next steps

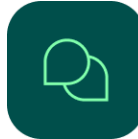
Build up from what we have achieved



Promoting in Asia



How can we increase awareness?



We promote the benefits of the DDF initiative and the USDM.



We are looking for other ways and partners to share about these initiatives.



Our goal is to move from ignorance to awareness to demand.





Thank you!

Contact us
sylvain@crscube.io

 **CRScube**

Clinical data, simply

studyOS Content



Ellis Hiroki Butterfield
CEO @ studyOS
hiroki@studynos.co

How USDM powers effective AI use in Clinical Trials

Agenda

Introduction

Problem

Naïve solutions

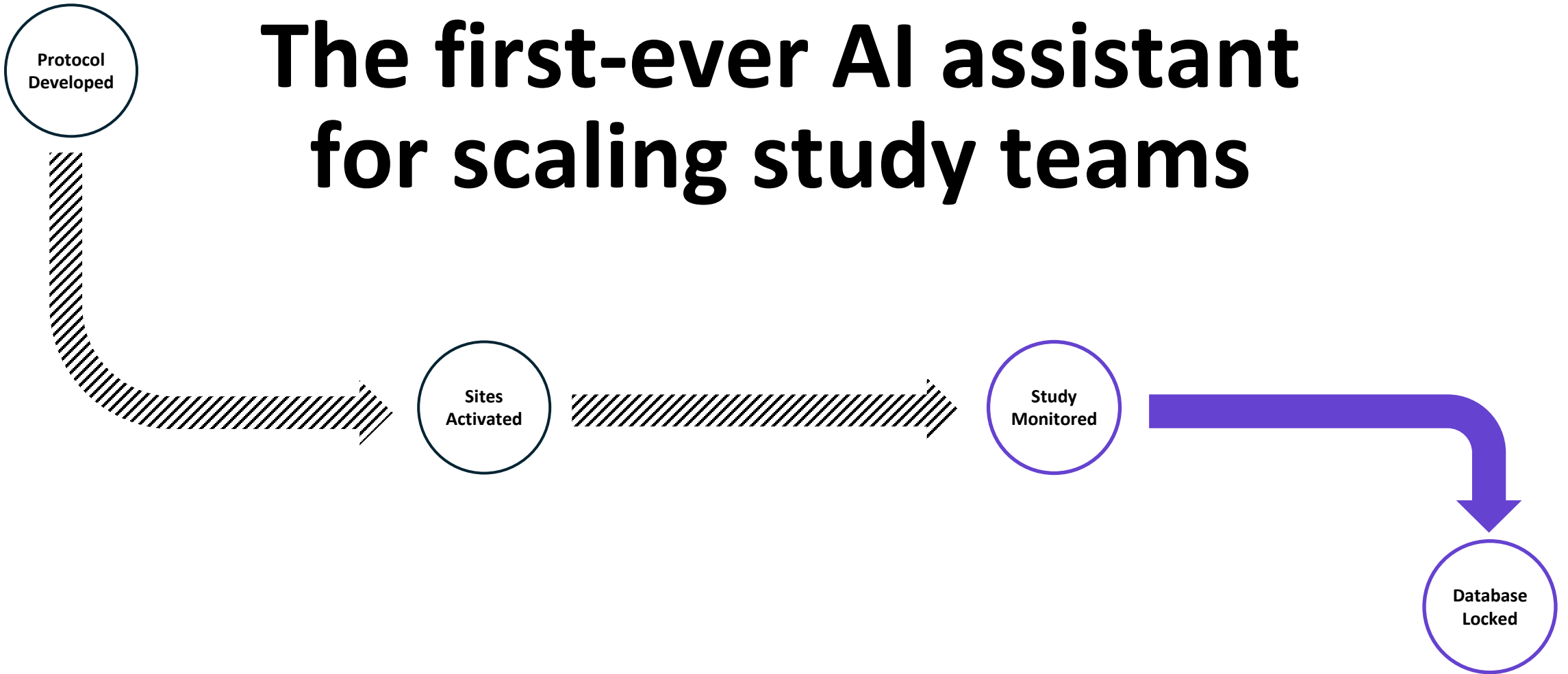
USDM

Gaps

Adoption

Meet Ash

The first-ever AI assistant for scaling study teams



And no, it's not
named after
this guy



<https://www.denofgeek.com/movies/ian-holm-ash-scariest-monster-in-alien/>

But actually,

Ashish Vaswani

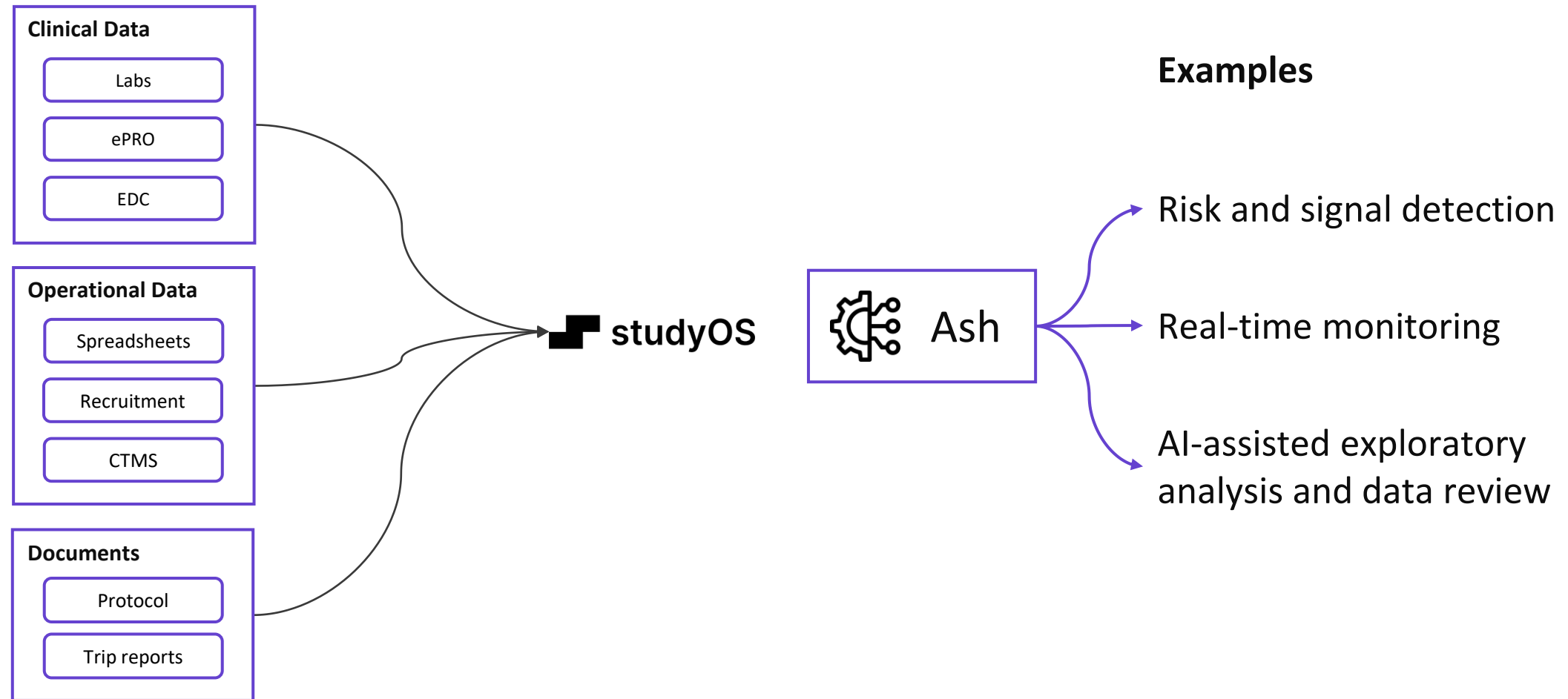
Lead author of
“Attention Is All
You Need”



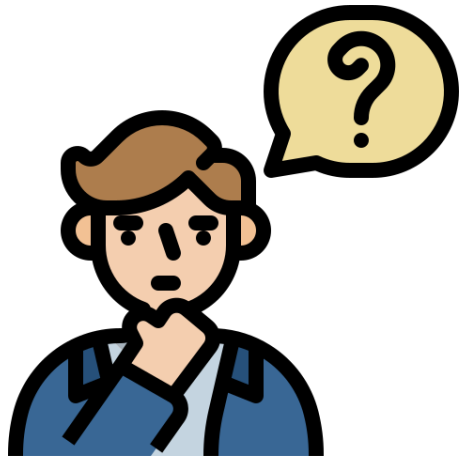
<https://montgomerysummit.com/wp-content/uploads/Ashish-2.png>

31st Conference on Neural Information Processing
Systems (NIPS 2017)

Ash powers studyOS



Narrowing the gap



Intent

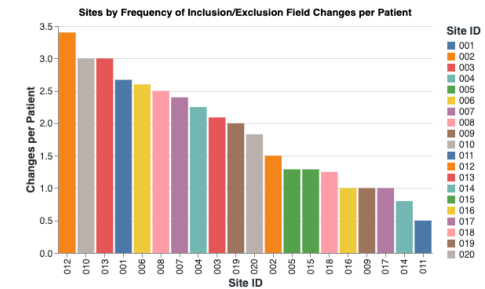


Outcome

Working with Ash today



Which sites have changed inclusion/exclusion fields most frequently, normalized on enrolled patients?

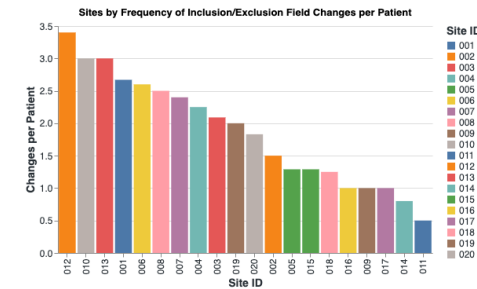


- “ The top 5 sites with the most frequent changes to I/E fields (normalized by patient count) are:
1. Site 012: **3.40** changes per patient
 2. Site 010: **3.00** changes per patient
 3. Site 013: **3.00** changes per patient
 4. Site 001: **2.67** changes per patient
 5. Site 006: **2.60** changes per patient

Working with Ash today

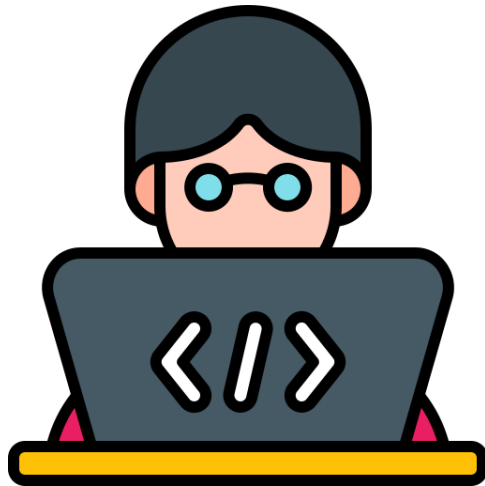


Which sites have changed inclusion/exclusion fields most frequently, normalized on enrolled patients?



- “ The top 5 sites with the most frequent changes to I/E fields (normalized by patient count) are:
1. Site 012: **3.40** changes per patient
 2. Site 010: **3.00** changes per patient
 3. Site 013: **3.00** changes per patient
 4. Site 001: **2.67** changes per patient
 5. Site 006: **2.60** changes per patient

Problem – Methods today



Programming



Manual

In an ideal world

Operational

Alert the CRA when a lab shows as collected but there is no corresponding lab report within 2 days of the visit.

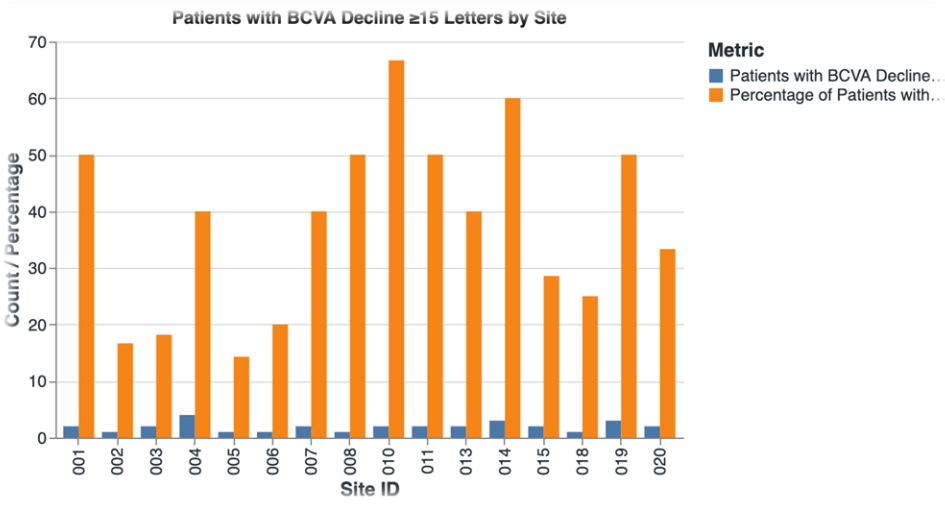
Clinical

Are there subgroups of patients who appear to be experiencing adverse events at higher-than-expected rates?

Forensic

Does the audit log show any sites updating data between two visits after initial data entry?

Site	Total Patients	Percentage of Patients with Decline	Patients with BCVA Decline ≥15 letters
001	4	50.00	2
002	6	16.67	1
003	11	18.18	2
004	10	40.00	4
005	7	14.29	1



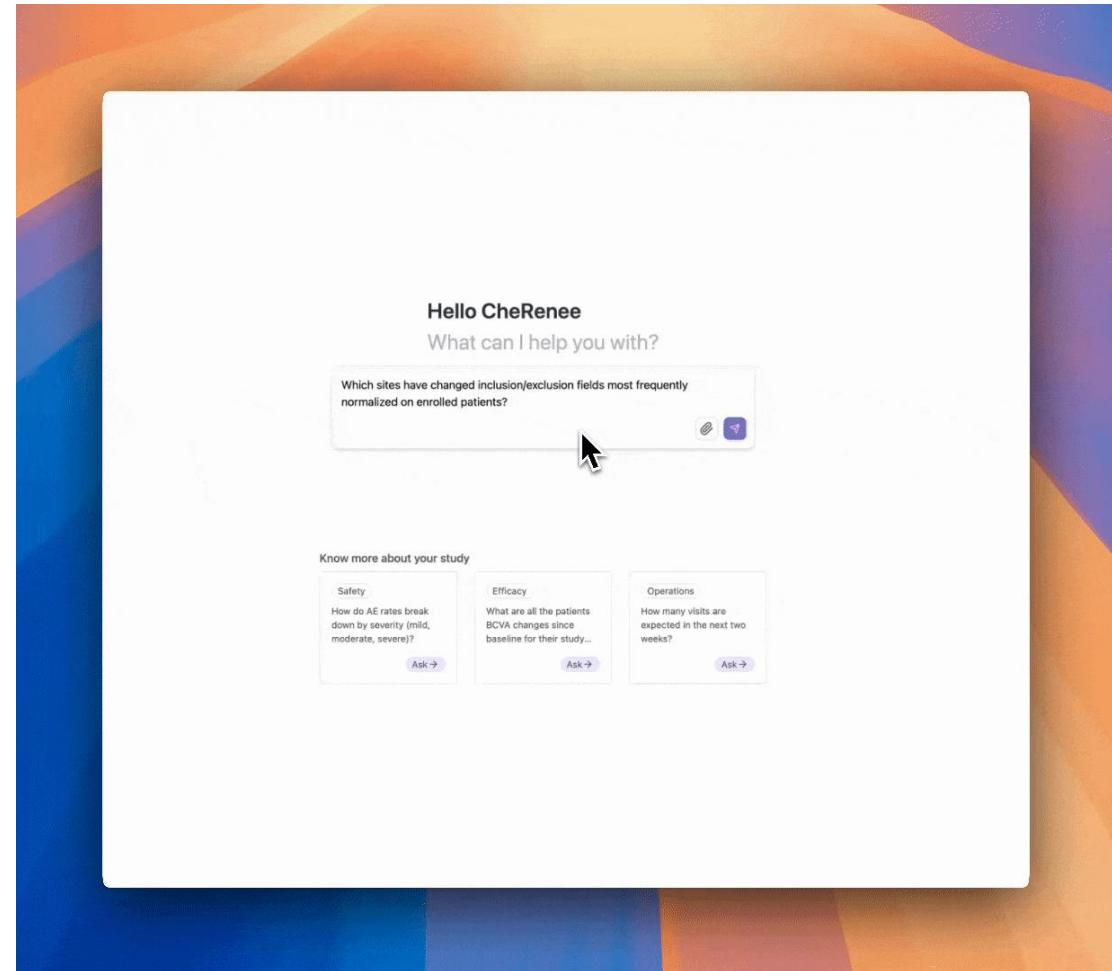
Ash

How can we allow users to effectively and accurately interrogate data and set controls on studies?

How can we reduce the time from user intent to automated analysis?

Applicable across

- Data Review
- Central Monitoring
- RBQM
- Audit Trail Analytics



Problem – Clinical trials present a unique analytics challenge

The problem of NLP for analytics is well researched.

However, general-purpose solutions are ineffective at answering questions in clinical trials

Products are generally built for marketing and sales -> rollups, aggregation, transaction data

Problem – Clinical trials present a unique analytics challenge

The problem of NLP for analytics is well researched.

However, general-purpose solutions are ineffective at answering questions in clinical trials

Products are generally built for marketing and sales -> rollups, aggregation, transaction data

1. Subgroups
2. Subperiods
3. Study-specific definitions
4. Branching conditional logic
5. Large amounts of text data
6. Need for reproducibility
7. Schedule-based rather than calendar-based temporal logic
8. Heterogenous data types
9. ...

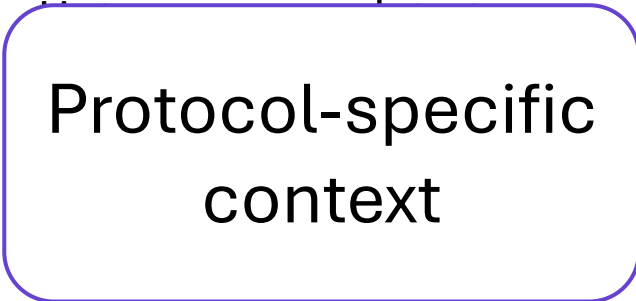
Problem – Clinical trials present a unique analytics challenge

The problem of NLP for analytics is well researched.

However, general-purpose solutions are ineffective at answering questions in clinical trials

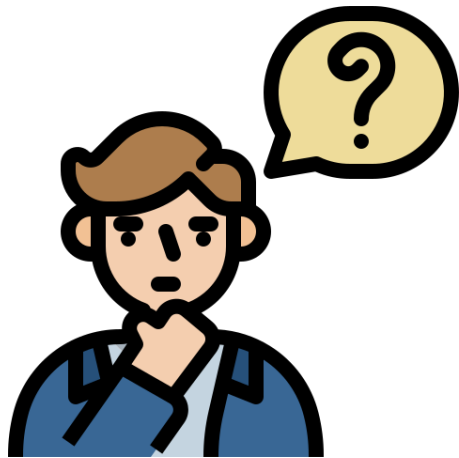
Products are generally built for marketing and sales -> rollups, aggregation, transaction data

1. Subgroups
2. Subperiods
3. Study-specific definitions
4. Branching conditional logic
5. Large amounts of text data
6. Need for reproducibility
7. Schedule-based rather than calendar-based temporal logic

8. 
- 9.

Problem

How do we provide relevant protocol data to the model accurately?



Which site has had the most changes to **primary endpoint data** after **baseline**, normalized on enrolled patients?

Uh-oh

Without protocol context, many prompts are unanswerable!


Which site has had the most changes to primary endpoint data after baseline, normalized on enrolled patients?



Answer Preparation Steps

(Click to expand/collapse)

Step 1  Processing question

Step 2  I can help analyze changes to data after baseline, but I need to clarify a few things about your question. The database contains an audit log that tracks changes to data, which would show modifications made after initial entry. However, I don't have explicit information about which measurements are considered "primary endpoints" for this clinical trial.

I can see that the audit_log table tracks all data changes including who made them, when they were made, and what values were changed. I can also see patient enrollment information across different sites. However, without knowing which specific measurements constitute the primary endpoints for this study, I can only provide partial information about data changes by site without the specific focus on primary endpoints.


If you'd like, I can help you explore general data modification patterns by site based on the audit log information available in the database.

Step 3  Done! Let me know if you need anything else!


Uh-oh


Context is everything!


Which site has had the most changes to primary endpoint data after baseline, normalized on enrolled patients?



Answer Preparation Steps (Click to expand/collapse)

Step 1  Processing question

Step 2  However, I don't have explicit information about which measurements are considered "primary endpoints" for this clinical trial

Step 3  Done! Let me know if you need anything else!

ify a few things
s changes to data,
don't have explicit
ndpoints" for this
who made them,
ee patient
wing which specific
n only provide
mic focus on primary
endpoints.
If you'd like, I can help you explore general data modification patterns by site based on the audit log information available in the database.

Naïve solution – Infer from EDC or similar data structures

EDCs capture an increasingly smaller percentage of total study data

Event (schedule) data only contain enough information to fulfill statistical analysis needs

Structures are different per study – not scalable

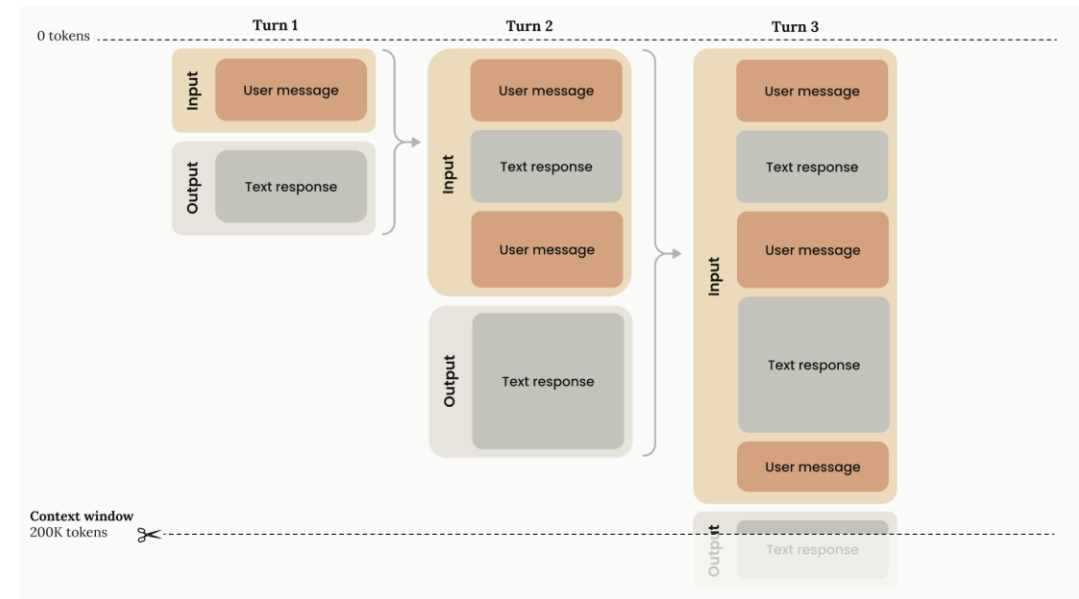
Naïve solution – Embed protocol directly into context window

Information is scattered throughout protocol

Prone to hallucinations

Not reproducible

Expensive! May not be technically feasible



<https://docs.anthropic.com/en/docs/build-with-claude/context-windows>

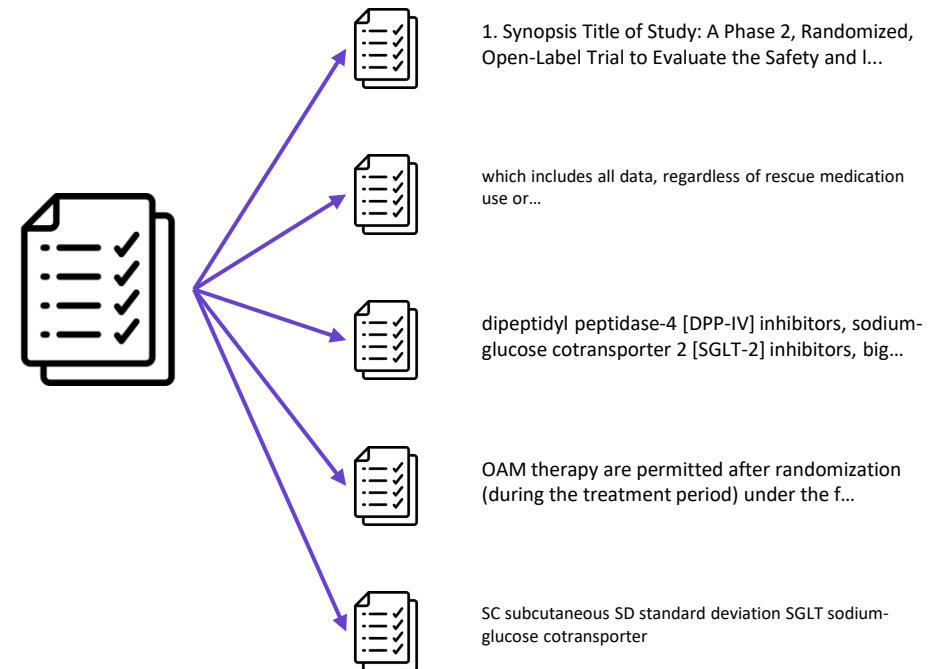
Naïve solution – Retrieval Augmented Generation

Protocol documents do not share the same layout

The layout contains meaning – text chunking is insufficient

Difficult to know how many top-k results to include for a given question

Largely incompatible with tables – study schedule



What is needed

Reproducible/deterministic

Scalable

Context rich



What is needed

Reproducible/deterministic

Scalable

Context rich

- Protocol-as-data allows the same code to produce the same results
- Can use narrative elements directly as embeddings rather than arbitrary text chunks

What is needed

Reproducible/deterministic

Scalable

Context rich

- Consistent interface unlocks downstream automation – write once, use many!
- Part of larger industry standardization effort
- Covers **almost** all scenarios

What is needed

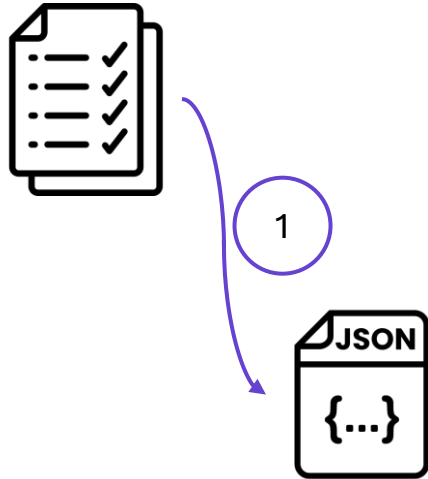
Reproducible/deterministic

Scalable

Context rich

- Related pieces of content link to one another
- Contains biomedical concepts allowing for more granular standardization and association with data
- Combines semi-structured narrative

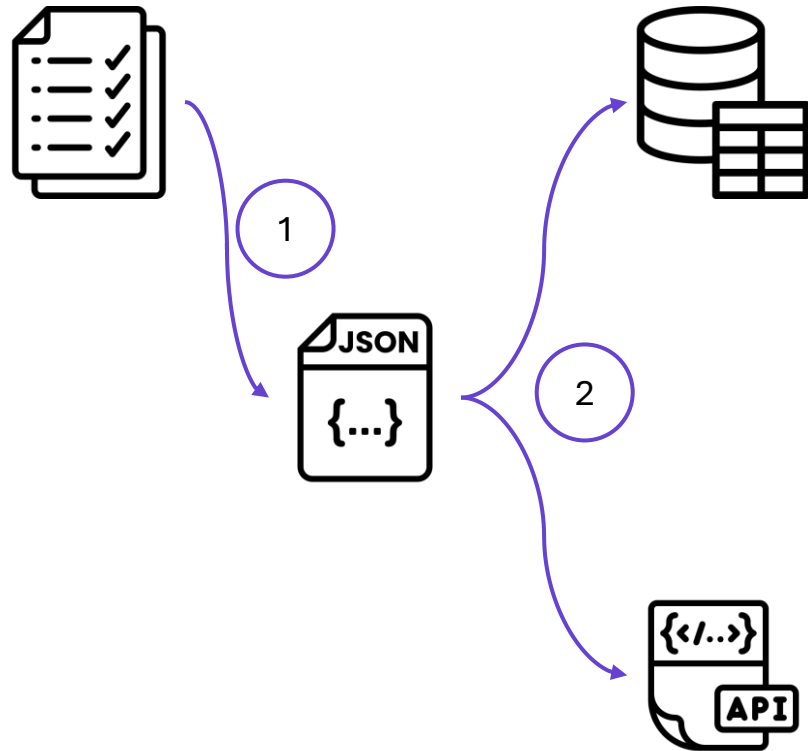
USDM use at studyOS



1

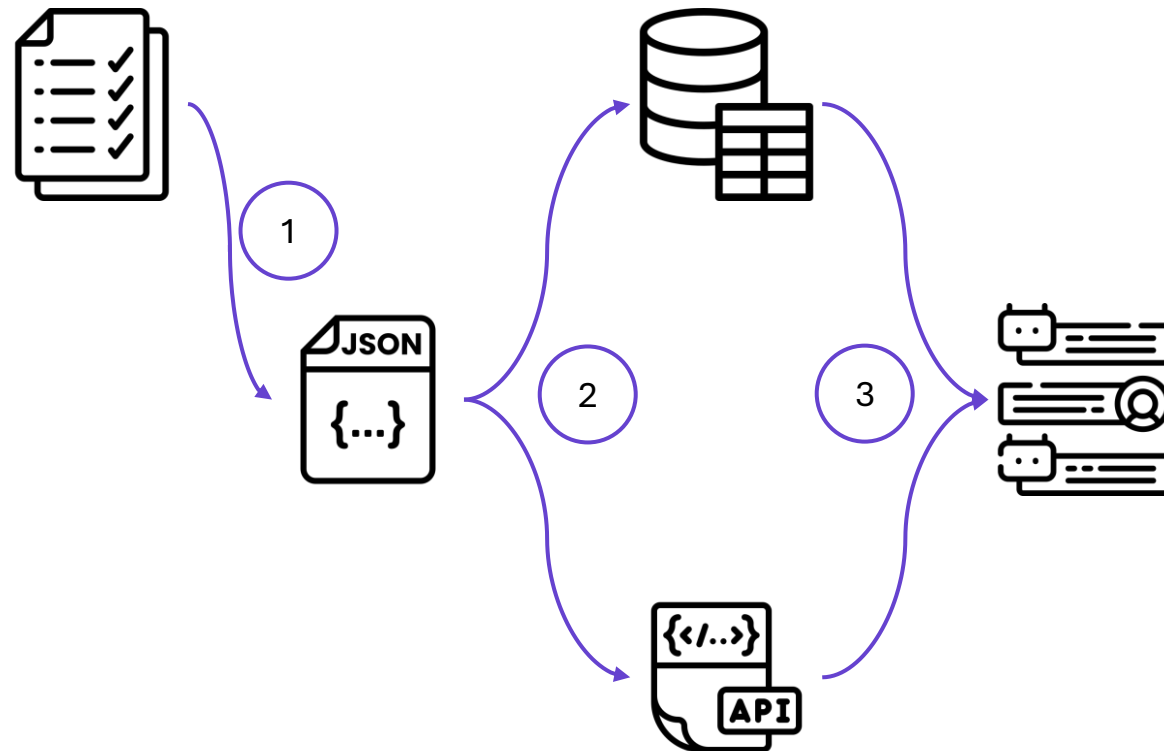
Use Ash to bootstrap the USDM format from the protocol PDF

USDM use at studyOS



- 1 Use Ash to bootstrap the USDM format from the protocol PDF
- 2 Generate reference tables and protocol API

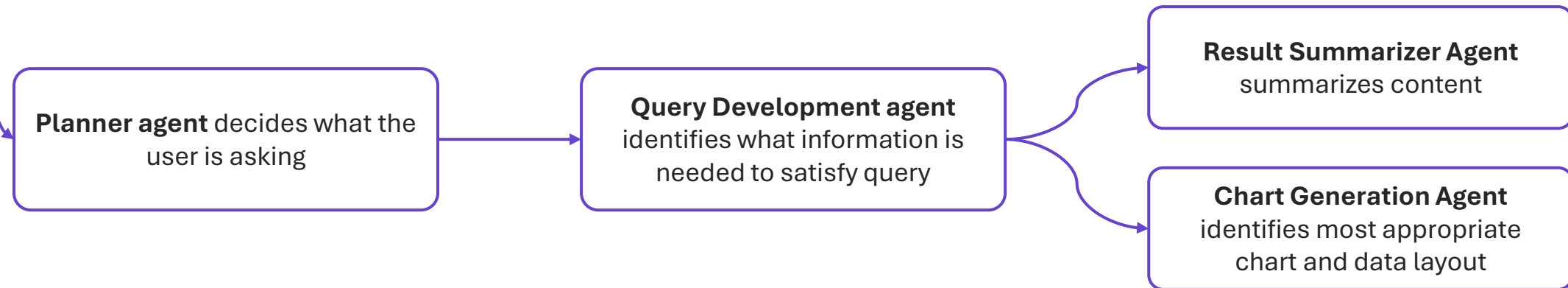
USDM use at studyOS



- 1 Use Ash to bootstrap the USDM format from the protocol PDF
- 2 Generate reference tables and protocol API
- 3 Provide protocol context to Ash during query generation

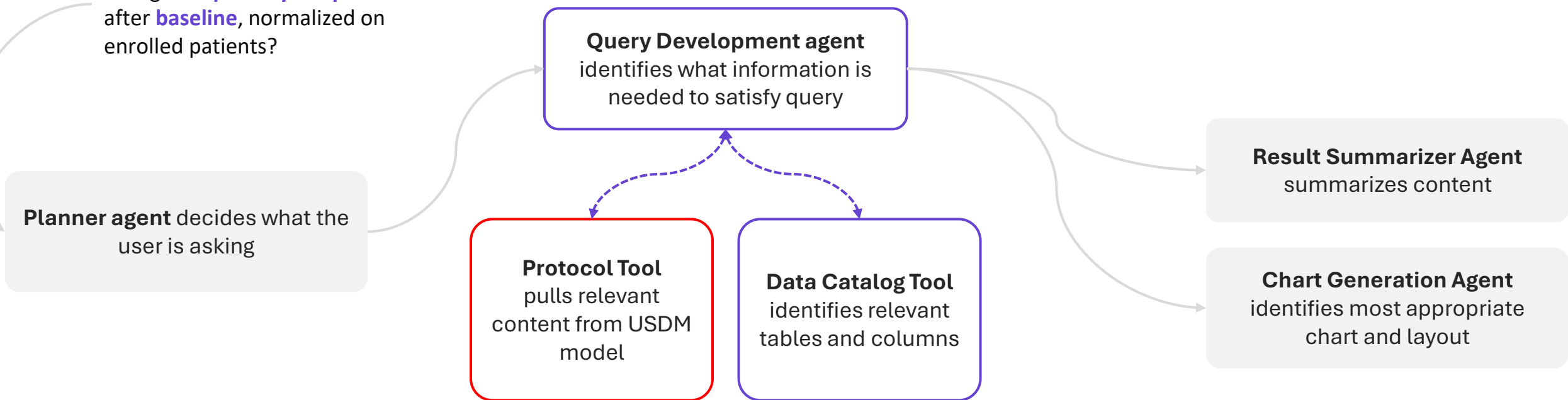
Agentic flow (simplified)

“ Which site has had the most changes to **primary endpoint data** after **baseline**, normalized on enrolled patients?



Agentic flow (simplified)

“ Which site has had the most changes to **primary endpoint data** after **baseline**, normalized on enrolled patients?



Protocol Tool (Simplified)

Term	Method	Code	Output
primary endpoint	Direct map	<code>\$.endpoints[?(@.level.decode == 'Primary Endpoint')]</code>	The proportion of patients achieving treatment success defined as either an increase in PG to >70 mg/dL or....
baseline	Embedding	<code>[0.0019510742,0.0016076918,0.0072813625,0.0...</code>	Triplicate ECGs collected 30 and 15 minutes prior to the start of insulin-induced hypoglycemia on Day 1 of each period will be used to</p>\n<p>establish a baseline. The consecutive triplicate...

USDM use at studyOS – Partial adoption still useful

What we didn't do

Skipped amendment and version handling

Skipped non-essential areas

Only used as context and not for downstream systems

And yet we still gained

Reproducible, accurate Clinical Trial Agent™

Automatic generation of reference tables

Reduced implementation process and timelines

Challenges

No wide adoption...yet

Bootstrapping PDFs to USDM format is non-trivial

Non-obvious incremental adoption strategy

Lack of standardized way to link to data

Eligibility criterion standardization

Exploration – Common Data Model

Standardize data as well according to USDM (BRIDG)

Unified ontology. Query re-use across studies

Extremely expensive and ROI lower than expected

Hard to keep synchronized

- cdm_anatomic_site_code
- cdm_anatomic_site_laterality_code
- cdm_date_range
- cdm_day_range
- cdm_defined_activity
- cdm_defined_observation_result
- cdm_medical_history
- cdm_performed_activity
- cdm_performed_adverse_event
- cdm_performed_lateral_activity
- cdm_performed_lateral_activity_base
- cdm_performed_lateral_activity_det
- cdm_performed_observation_result
- cdm_planned_activity
- cdm_planned_criterion_group_to_observation_result
- cdm_prior_concomitant_medication
- cdm_site
- cdm_subject

Internal opportunities

Better automation to handle versioning

Standardize connection to data

Expose USDM bootstrapper and protocol API externally

Thank you!



Ellis Hiroki Butterfield



hiroki@studynos.co



<https://www.studynos.co>

Presenter Q&A

Meet Our Expert Panel



Govind Subramanian
Digital Data Flow
Project Manager



**Sylvain
Berthelot**

VP, Strategy and
Marketing

CRScube



Robert Hay

Marketing and
Solutions Consultant

CRScube



**Ellis Hiroki
Butterfield**

Chief Executive
Officer

studyOS



Julie Smiley

Vice President, Data
Sciences

CDISC

If you have a question, **please denote to whom** the question is directed.

Note: depending on time, we will not be able to answer all questions

As a reminder, we can't answer questions about:

- Specific vendors with whom organizations are working
- Costs of using/implementing TransCelerate assets/tools
- Which member companies are using the assets/tools



TransCelerate
BIOPHARMA INC.

Tools & Resources



Visit us, for more information:
www.TransCelerateBioPharmaInc.com



Stay Connected:
[Sign Up](#) for our Awareness
& Implementation Community!

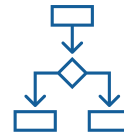


Follow us on social media:
[TransCelerate
BioPharma Inc.](#)

 [@TransCelerateBioPharma](#)



To find out more about our events, click [here](#)



Click [here](#) to learn more about our solutions



View our annual achievements archive, [here](#)



Click [here](#) to read our recent blog posts

Thank you!

Please reach out with any additional questions:
Events@transceleratebiopharmainc.com