National Institute of Diabetes and Digestive and Kidney Diseases

Principles of Data Collection and Management

Matt Breymaier, Sai Theja, & Ken Wilkins

2025-07-24

NIDDK Biostats Webinar Series,

Second Session

National Institute of Diabetes and Digestive and Kidney Diseases

REDCap Research Electronic Data Capture

Matt Breymaier, Sai Theja NIDDK; NCI & other ICOs ->

OpenClinica

MEDIDATA

From Research Study Design to Collecting, Managing, and Analyzing Data

Learning Objectives:

1. To delineate features of REDCap to support project management for research studies (e.g., how different types of studies (longitudinal vs cross-sectional etc) can be designed).*

2. To outline steps to create detailed data collection plans which fulfill regulatory requirements.

3. To identify principled approaches to data collection and management.

4. To explain the connections between research rigor and reproducibility.

Other helpful resources /links within NIH SharePoint:

** some webinar participants may well use other electronic data capture (EDC) systems, such as those from* CTSU

National Institute of Diabetes and Digestive and Kidney Diseases

Principles of Data Collection & Management Part 3 topics

- **Document organization and access as part of study planning:**

regulatory, clinical,* and *case report forms

- **Data Management and Sharing Plans**
- **Data Management for Reproducibility**
  - ***Take Home Points to follow Guiding Principles***

NIDDK Biostats Seminar Series Speaker

Today, 2025-07-24

Ken Wilkins: Mathematical Statistician, Biostatistics Program Office

Office of the Director, NIDDK

***see also the earlier webinar "Initiation, Regulatory Requirements, & Statistical Design for Research Studies conducted at NIH"**

Principles of Data Collection & Management: basis in *clinical* data

- CLINICAL CARE INFORMATION *IS* DATA (SUBJECT TO HIPAA):
  ADAPT *YOUR* USE OF IT IN *RESEARCH (PER PROTOCOLS AS IN NIH CC)*
- **as stated in a longstanding NIH course* on clinical research:**

"DATA MANAGEMENT IS THE OPERATIONALIZATION OF GOOD CLINICAL PRACTICE (GCP)"

National Institute of Diabetes and Digestive and Kidney Diseases

Principles of Data Collection & Management: basis in *clinical* data

- **CLINICAL CARE INFORMATION *IS* DATA (SUBJECT TO HIPAA):**

ADAPT *YOUR* USE OF IT IN *RESEARCH (PER PROTOCOLS AS IN NIH CC)*

- **as stated in a longstanding NIH course* on clinical research:**

"DATA MANAGEMENT IS THE OPERATIONALIZATION OF GOOD CLINICAL PRACTICE (GCP)"

CITI PROGRAM

National Institute of Diabetes and Digestive and Kidney Diseases

Principles of Data Collection & Management: basis in *clinical* data

- CLINICAL CARE INFORMATION *IS* DATA (SUBJECT TO HIPAA): ADAPT *YOUR* USE OF IT IN *RESEARCH* (*PER* PROTOCOLS *AS IN NIH CC*)
- **as stated in a longstanding NIH course* on clinical research:**

"DATA MANAGEMENT IS THE OPERATIONALIZATION OF GOOD CLINICAL PRACTICE (GCP)"

CITI PROGRAM

For more details, take a NIH-sponsored CITI course on GCP; also **try the**

***NIH Course: Intro to Principles & Practices of Clinical Research** OR *IPPCR*

Principles of Data Collection & Management Part 3 Topics

- **Document organization and access as part of study planning: *regulatory, clinical,* and *case report forms***
- Data Management for Reproducibility
- Data Management and Sharing Plans
  - *Take Home Points to follow Guiding Principles*

"All investigators are expected to conduct themselves according to the highest standards of professional conduct and integrity and to adhere to the ethical principles that address the protection of human subjects in research"

-3014-300 Investigator Responsibilities, NIH Policy Manual

Principles of Data Collection & Management Part 3 Topics

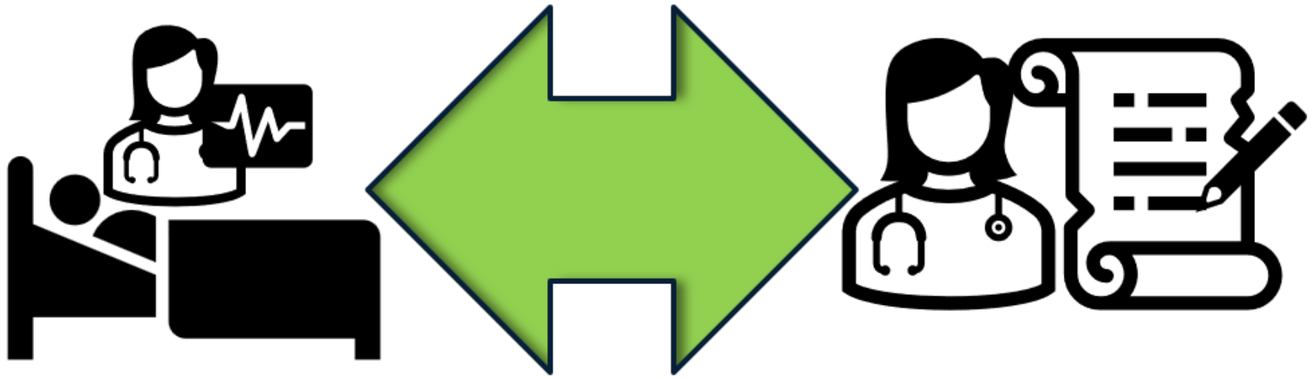
- **Document organization and access as part of study planning: *regulatory, clinical,* and *case report forms***
- Data Management for Reproducibility
- Data Management and Sharing Plans
  - *Take Home Points to follow Guiding Principles*

"All investigators are expected to conduct themselves according to the highest standards of professional conduct and integrity and to adhere to the ethical principles that address the protection of human subjects in research"
-3014-300 Investigator Responsibilities, NIH Policy Manual

Data Integrity

Data Management Principles= Ethical Principles

Document organization and access as part of study planning: *regulatory, clinical,* and *case report forms (CRFs) – defining each type*



Clinical interaction & assessment with study participants

Clinical records of interaction & assessment with study participants

National Institute of Diabetes and Digestive and Kidney Diseases

Document organization and access as part of study planning: *regulatory, clinical,* and *case report forms (CRFs) – defining each type*

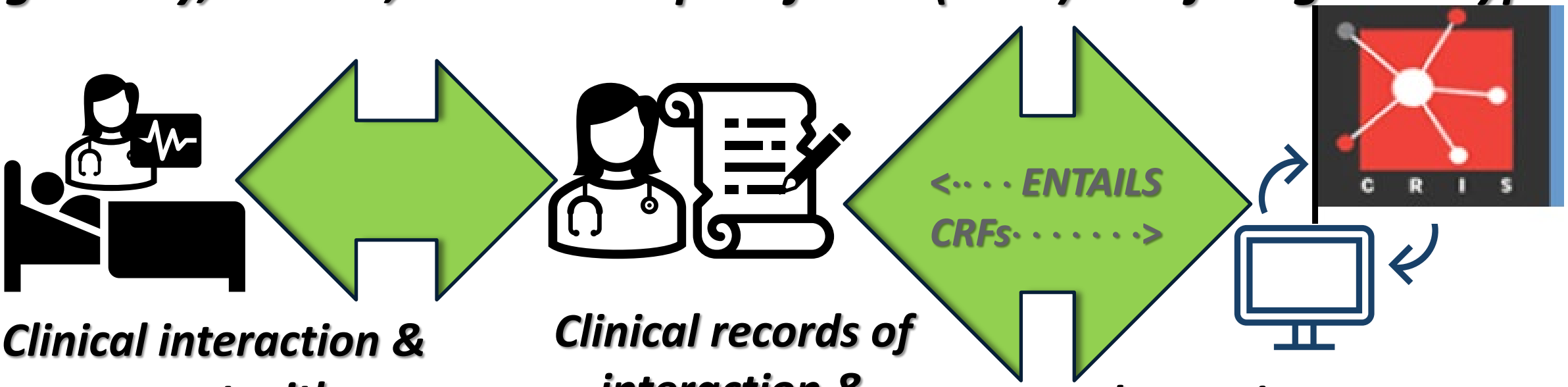
Clinical interaction & assessment with study participants

Clinical records of interaction & assessment with study participants

Electronic-system-entered records of interaction & assessment with study participants

National Institute of Diabetes and Digestive and Kidney Diseases

Document organization and access as part of study planning: *regulatory, clinical,* and *case report forms (CRFs) – defining each type*



Clinical interaction & assessment with study participants

Clinical records of interaction & assessment with study participants

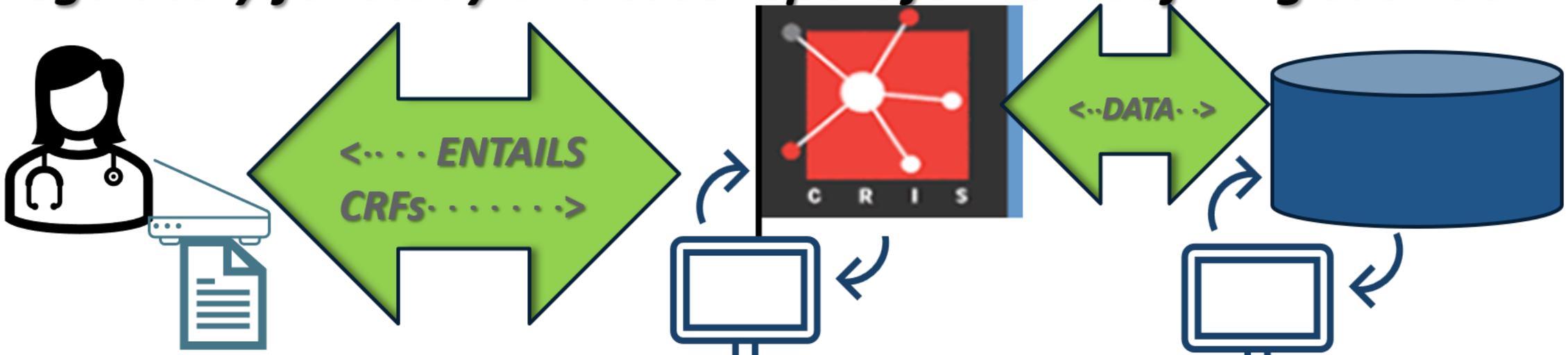
Electronic-system-entered records of interaction & assessment with study participants

CLINICAL

REGULATORY

National Institute of Diabetes and Digestive and Kidney Diseases

Document organization and access *within scope* of study planning: *regulatory for study* and *case report forms – defining such SUB-types*



Clinical source document records of interaction & assessment with study participants

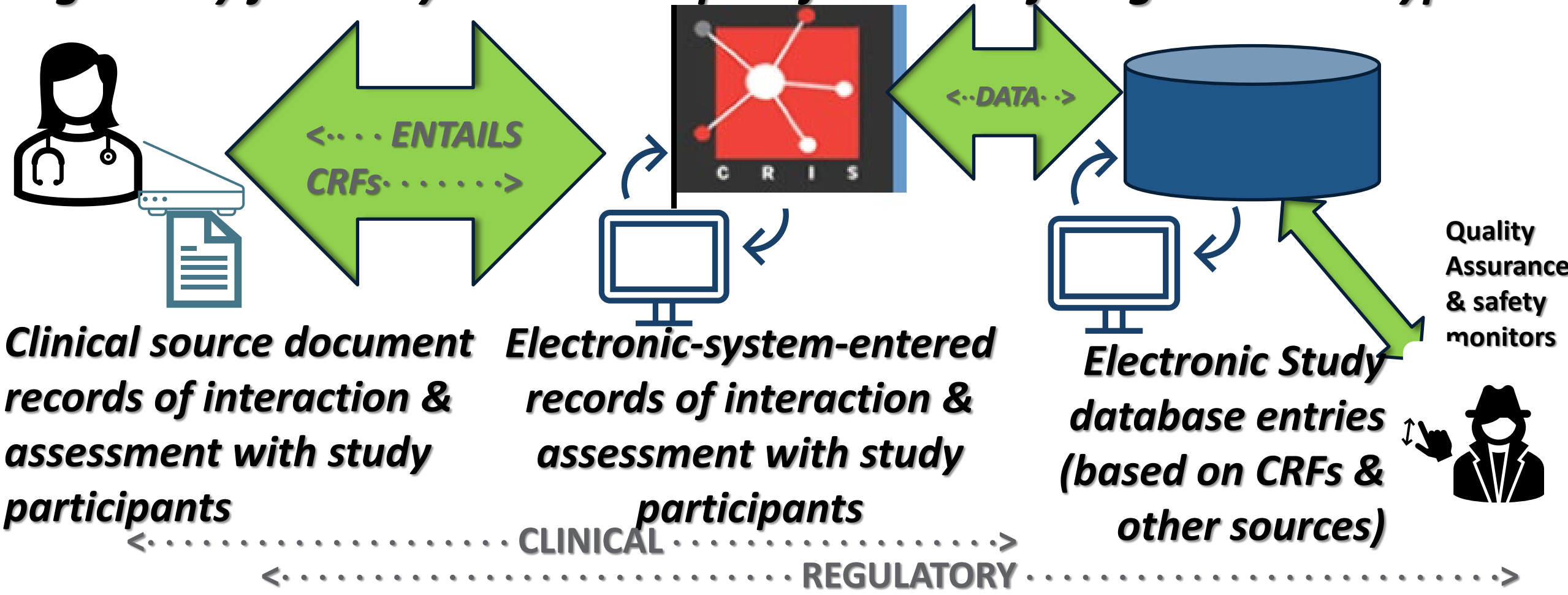
Electronic-system-entered records of interaction & assessment with study participants

Electronic Study database entries (based on CRFs & other sources)

<· CLINICAL ·>

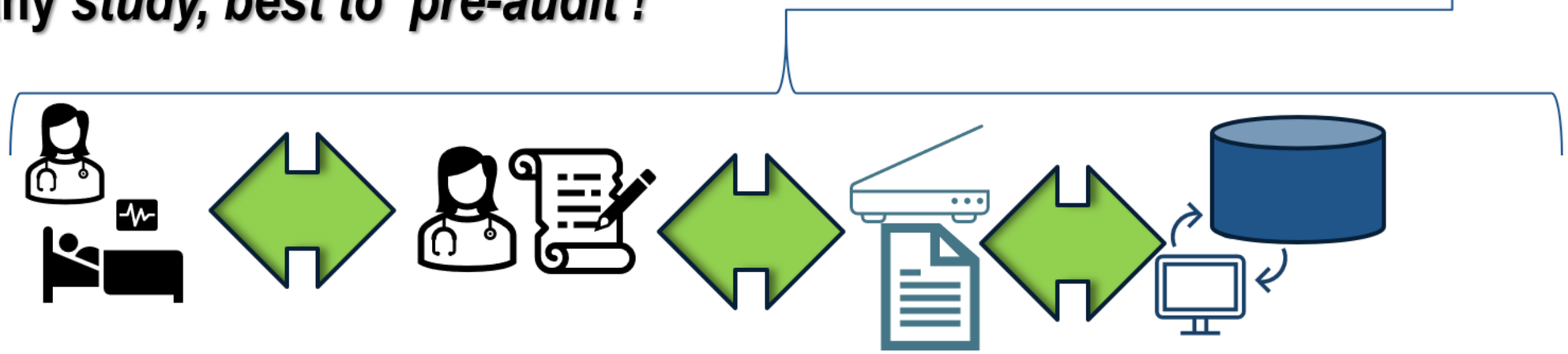
<· REGULATORY ·>

Document organization and access *within scope* of study planning: *regulatory for study* and *case report forms – defining such SUB-types*



Document organization & access as part of study planning (*with* help!): *regulatory, clinical,* and *case report forms:* data *tracked* throughout

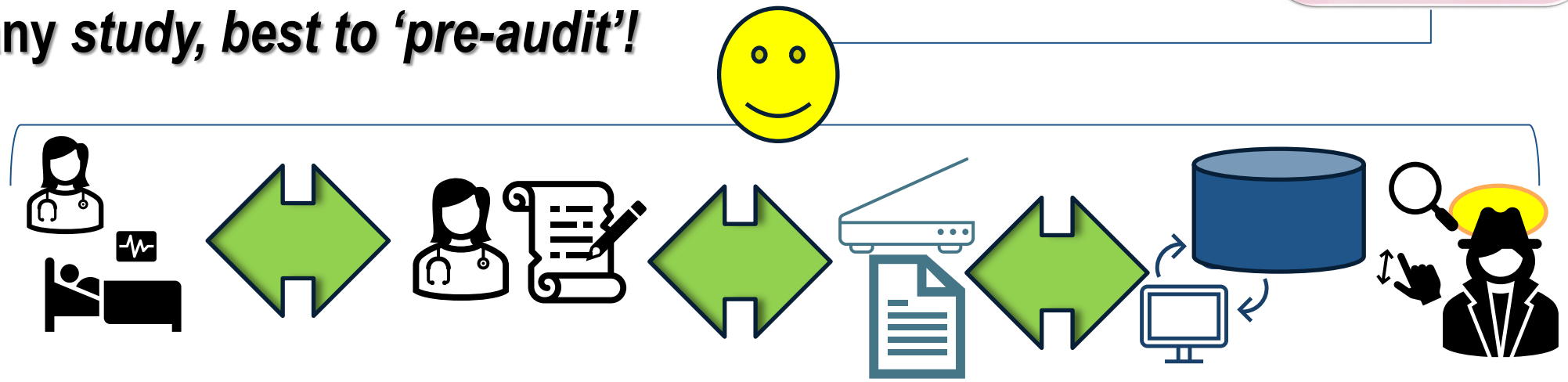
- Protocol navigation staff (from IRBO to ICO-specific, e.g., Dr. Studlack)
- Clinical Directors' Office staff, especially *Data Mgmt/Informatics!*
- *Study team members who are* closest *to data-generation*
- *For trials, also Quality Assurance(QA)/Safety monitors*
  - *For* any *study, best to 'pre-audit'!*



Document organization & access as part of study planning (*with* help!): *regulatory, clinical,* and *case report forms:* data *tracked* throughout

- Protocol navigation staff (from IRBO to ICO-specific, e.g., Dr. Studlack)
- Clinical Directors' Office staff, especially *Data Mgmt/Informatics!*
- *Study team members who are closest to data-generation*
- *For trials, also Quality Assurance(QA)/Safety monitors*
  - *For any study, best to 'pre-audit'!*

Team Members to keep in mind for later
Investigator
Clinician Co-Investigator/Study Coordinator
Data Manager/ Informaticist
Data Analysts/ Biostats



From Study Design to Collecting, Managing, & Analyzing Data: *KEY STUDY DOCS*

BEGINNING WITH THE END IN MIND: *protocol will be posted publicly*

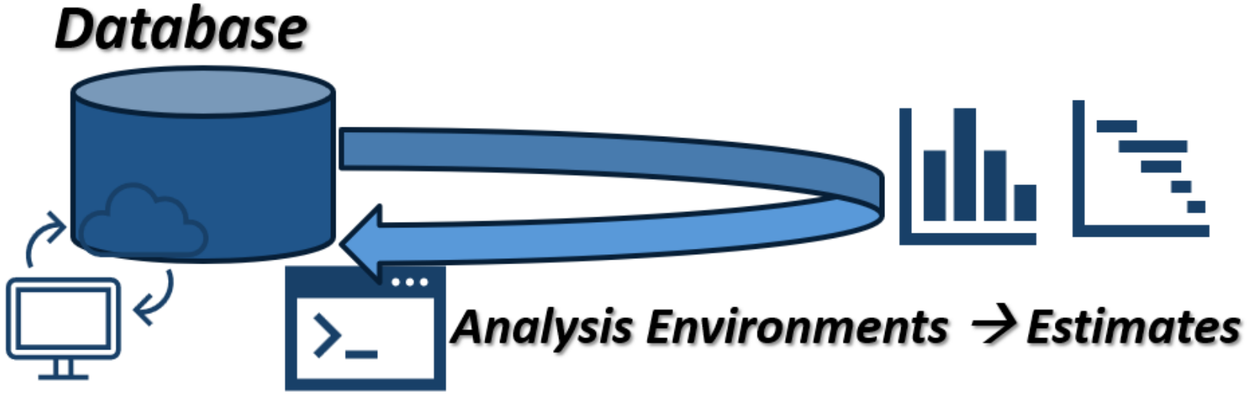
NOTE: journals *require* rigor, prior-specified data analyses *documented in:*

- **FULL & FINAL protocol: use templates adhere to type (e.g., clinical trial)**
  - **See walk-through of protocol elements during earlier webinar in BTEP series**
  - **Templates for varied types of studies (trials v. observational v. 'secondary')**
    - **Among trials, distinctions by 'intervention' type: drugs/biologics v. device v. behavioral**
    - **Protocol Section 9 may well define distinct 'analysis populations' per distinct study aims**
  - **For reproducibility (covered in depth later) *much additional metadata needed***
    - ***High-level description of approach in protocol, details (metadata) can be triaged to SAP***
    - ***Additional structural details (metadata) would fall within the Data Management (DM) plan***

BEGINNING WITH THE END IN MIND: smooth Data Management(DM)→Analysis

NOTE: journals *require* rigor, prior-specified data analyses *documented in:*

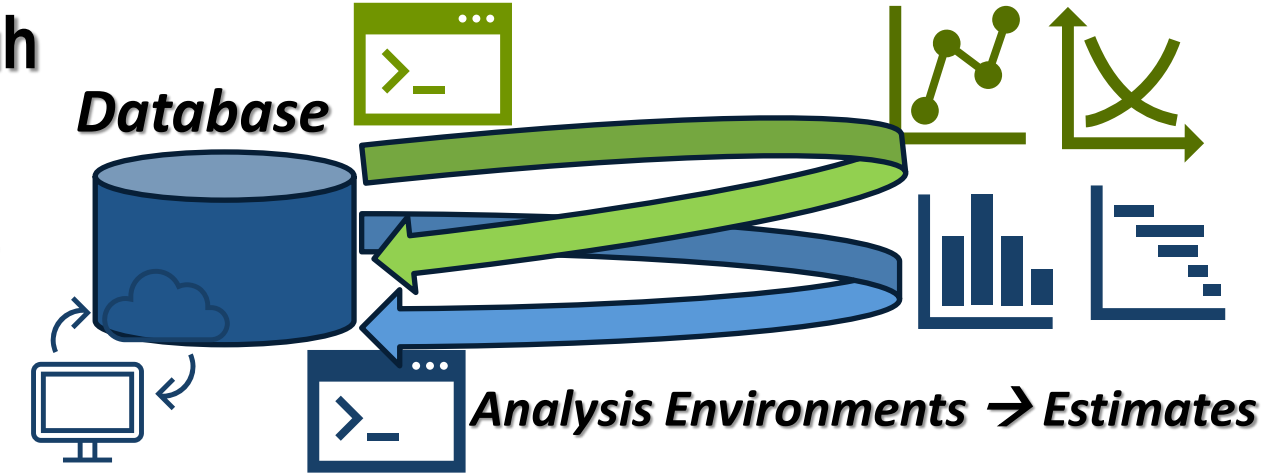
- **Statistical Analysis Plan (e.g., extramural templates help; *more to come!*)**
  - **THEN can go into level of detail needed to REPRODUCE a study-specified analysis**
  - **As covered in *prior* webinar by Dr. Auh**
  - **MORE TIME UP-FRONT YET:**
  - ***ANALYSIS SETUP DONE WITH DM SPEC'S:***
    - ***CLEAR EXPECTATIONS AMONG TEAM***



BEGINNING WITH THE END IN MIND: smooth Data Management(DM)→Analysis

NOTE: journals *require* rigor, prior-specified data analyses *documented in:*

- **Statistical Analysis Plan (e.g., extramural templates help; *more to come!*)**
  - **THEN can go into level of detail needed to REPRODUCE a study-specified analysis**
  - **As covered in *prior* webinar by Dr. Auh**
  - **MORE TIME UP-FRONT YET:**
  - ***ANALYSIS SETUP DONE WITH DM SPEC'S:***
    - ***CLEAR EXPECTATIONS AMONG TEAM***
    - ***CLEAR DIVISION OF DATA TASKS***
    - ***GOALS:* VERSIONED *REPRODUCIBILITY* AND *MITIGATING FALSE POSITIVE FINDINGS***



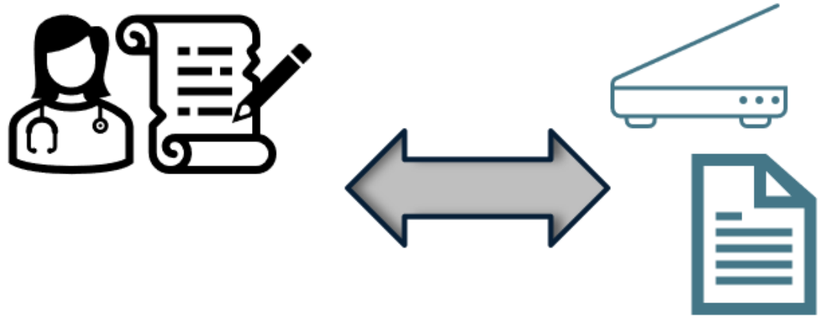
From Research Study Design to Collecting, Managing, & Analyzing Data: *clinical* context

Notably, *all* clinical care (done for protocol enrollees) requires documentation

NIH Clinical Center
America's Research Hospital

As such documentation is *required* for rigor, need to *organize* (as in PQS*)

- **Document organization & access as part of study planning & conduct: *regulatory, clinical, & case report forms (CRFs)***



Source documents ***CRF entries***

*PROTRAK Query System, login with NIH auth. here

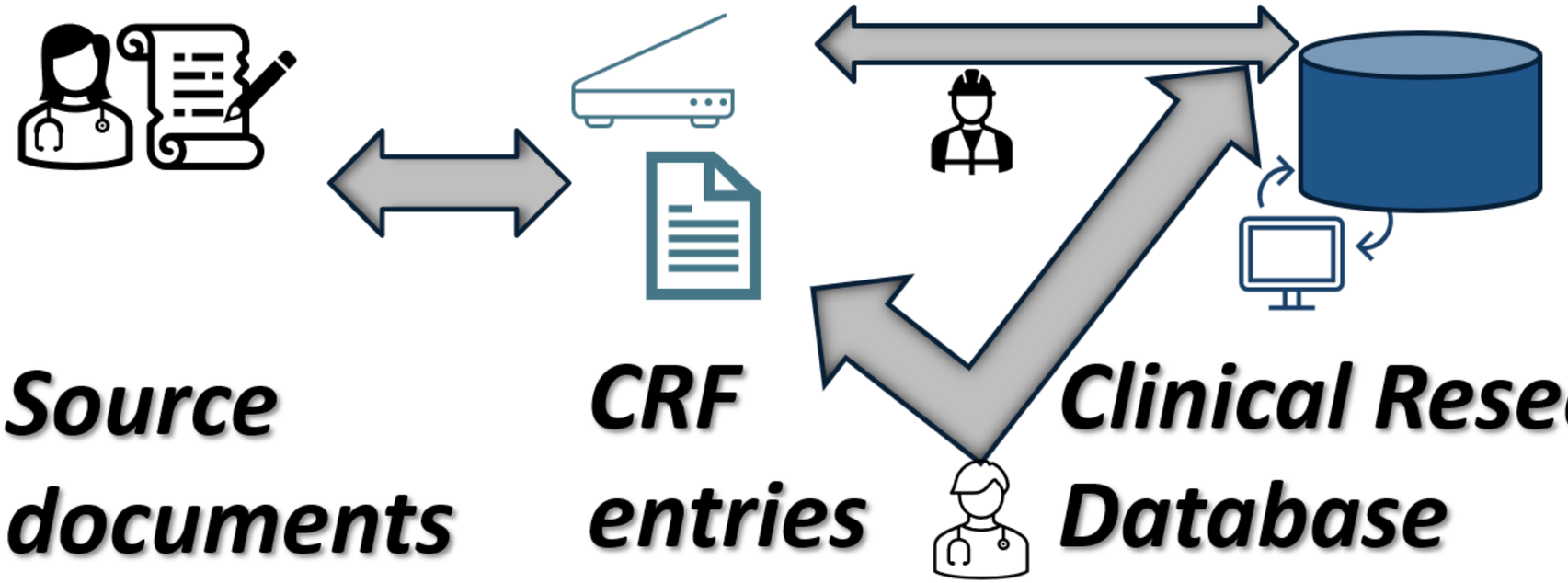
From Research Study Design to Collecting, Managing, & Analyzing Data: *clinical* context

Notably, *all* clinical care (done for protocol enrollees) requires documentation

NIH Clinical Center
America's Research Hospital

As such documentation is *required* for rigor, need to *organize* (as in PQS*)

- **Document organization & access as part of study planning & conduct:** ***regulatory, clinical, & case report forms (CRFs)***



*PROTRAK Query System, login with NIH auth. here

National Institute of Diabetes and Digestive and Kidney Diseases

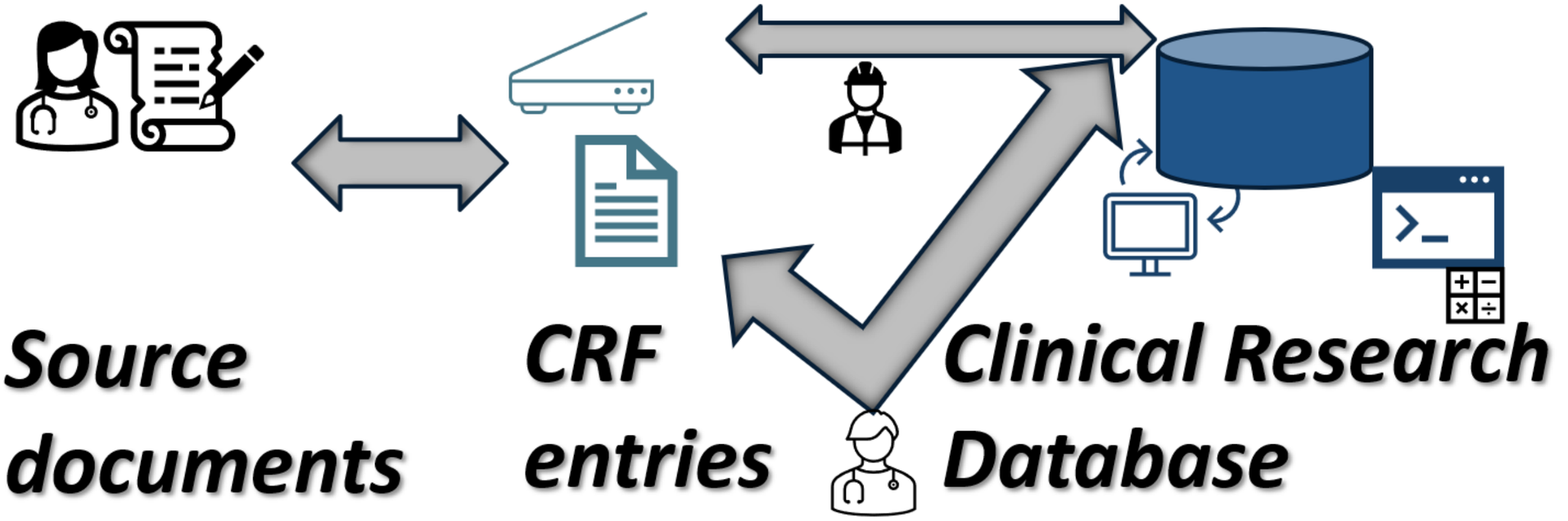
From Research Study Design to Collecting, Managing, & Analyzing Data: *clinical* context

Notably, *all* clinical care (done for protocol enrollees) requires documentation

NIH Clinical Center
America's Research Hospital

As such documentation is *required* for rigor, need to *organize* (as in PQS*)

- **Document organization & access as part of study planning & conduct: *regulatory, clinical, & case report forms (CRFs)***



*PROTRAK Query System, login with NIH auth. here

National Institute of Diabetes and Digestive and Kidney Diseases

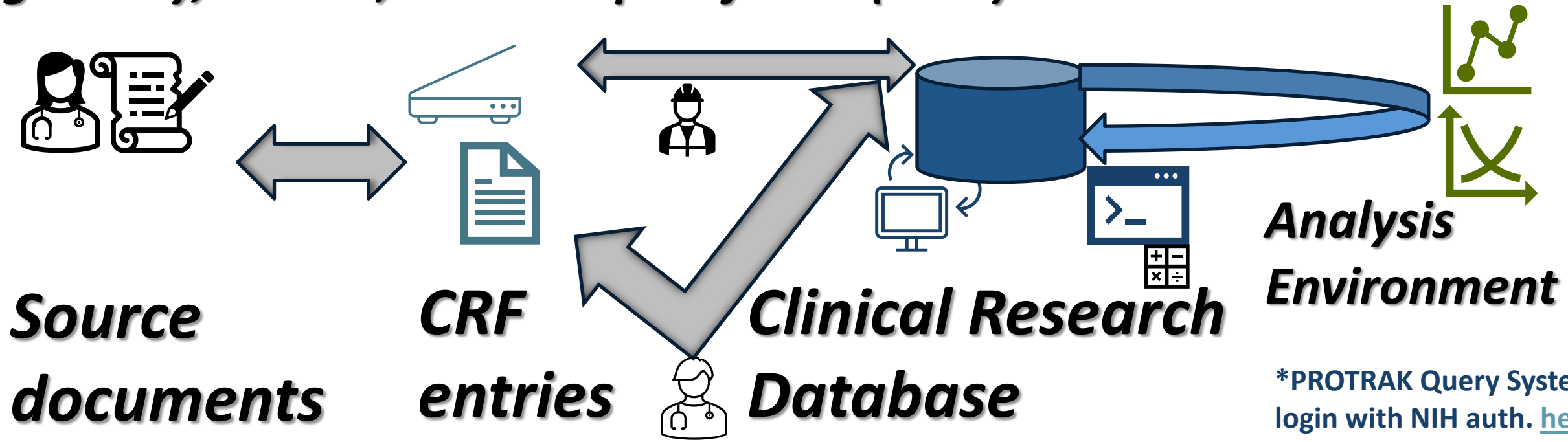
From Research Study Design to Collecting, Managing, & Analyzing Data: *clinical* context

Notably, *all* clinical care (done for protocol enrollees) requires documentation

NIH Clinical Center
America's Research Hospital

As such documentation is *required* for rigor, need to *organize* (as in PQS*)

- **Document organization & access as part of study planning & conduct:** ***regulatory, clinical, & case report forms (CRFs)***



*PROTRAK Query System, login with NIH auth. here

BEGINNING WITH THE END IN MIND: external publication expectations

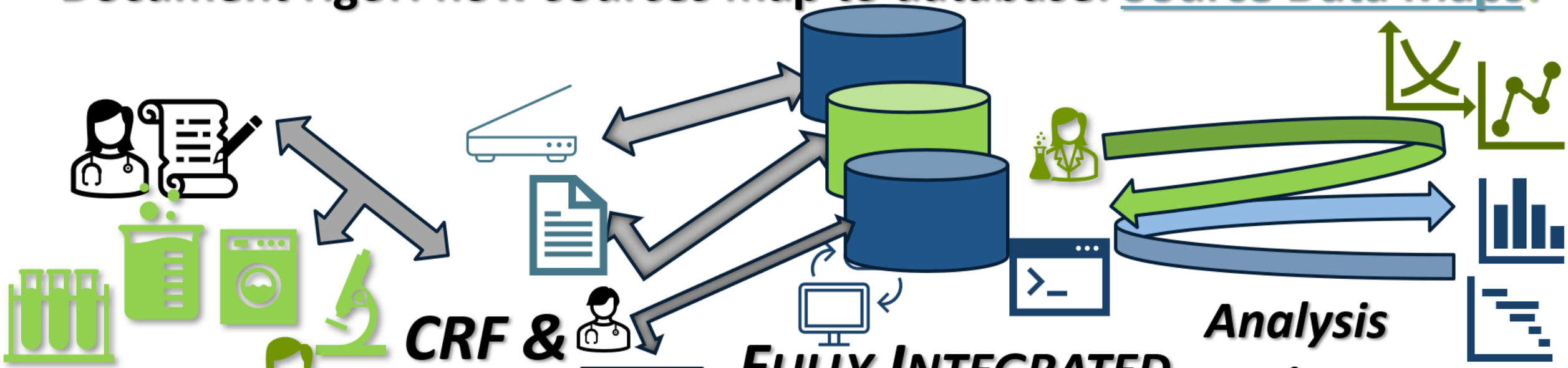
NOTE: *most* journals *<u>require</u>* rigor, docs with prior-specified data analyses

- PROTOCOL-specified analysis plans
- Statistical Analysis Plan (SAP)
- **Occasionally, depending on the journal-meets-editorial discretion:**
  - **Data collection instrumentation (see Best Practices in Resources Doc)**
  - **Documentation of data workflows (see Data Mgmt/Sharing Plan [DSMP], below)**
  - **Documented oversight by independent agents (IRB, Safety Monitoring bodies)**

From Research Study Design to Collecting, Managing, & Analyzing Data: **Source Maps**

To meet this, document <u>all</u> *as you organize* (and *map out* data integration)

- **Document rigor: how sources map to database: <u>Source Data Maps</u>!**



Source documents/ readouts*

CRF & inter-mediate file entries

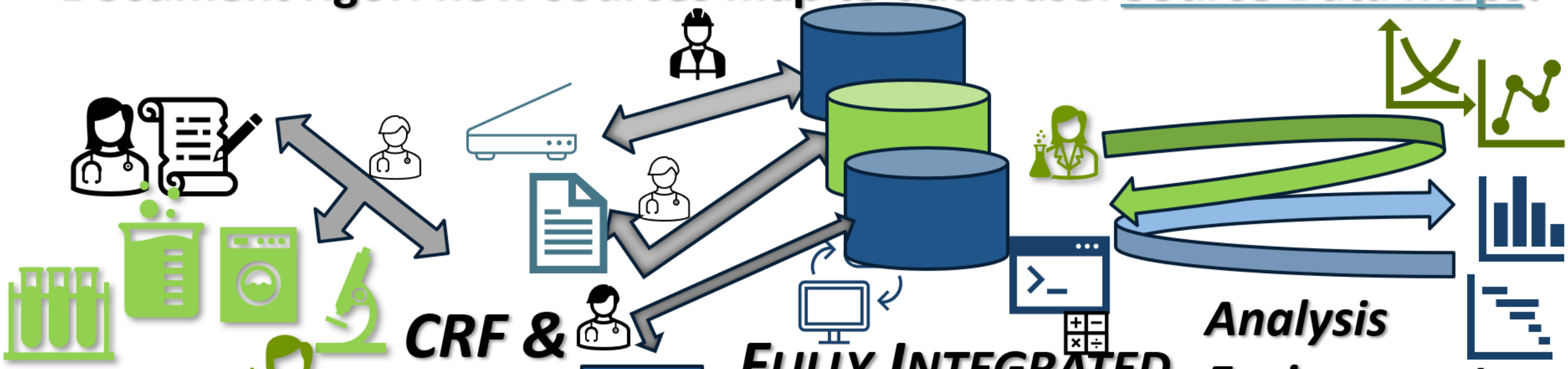
FULLY INTEGRATED Clinical Research Database

Analysis Environment… integration key to **<u>monitoring</u>**

From Research Study Design to Collecting, Managing, & Analyzing Data: **Source Maps**

To meet this, document <u>all</u> *as you organize* (and *map out* data integration)

- **Document rigor: how sources map to database: <u>Source Data Maps</u>!**



Source documents/ <u>readouts*</u>

CRF & inter-mediate file entries

FULLY INTEGRATED Clinical Research Database

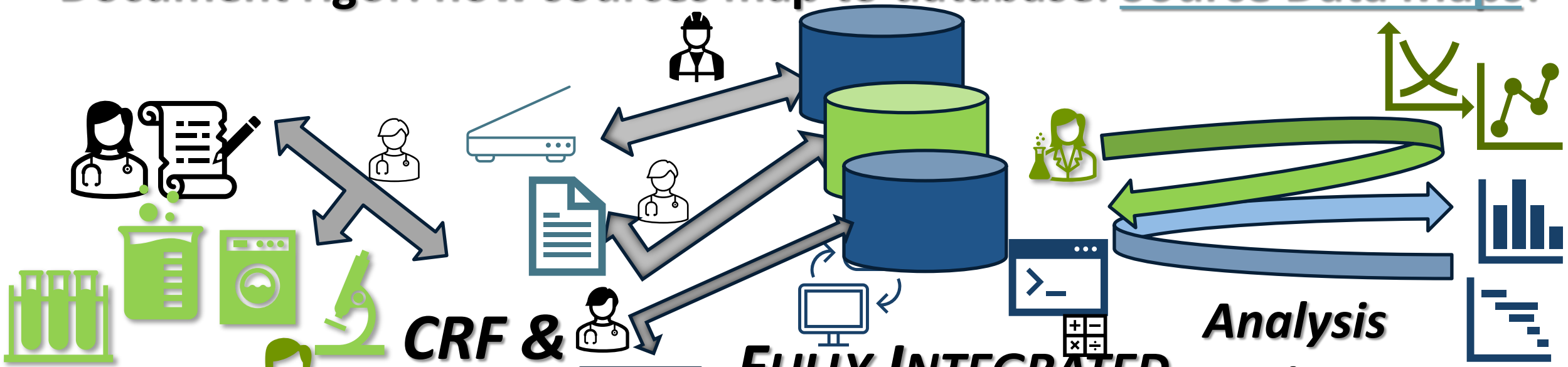
Analysis Environment… integration key to **<u>monitoring</u>**

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to Collecting, Managing, & Analyzing Data: **Source Maps**

To meet this, document <u>all</u> *as you organize* (and *map out* data integration)

- **Document rigor: how sources map to database: Source Data Maps!**



* On note of "source documents" it's noteworthy that any document created within a 21CFR-part-11-compliant system will *itself* be considered a ***source document*** – e.g., REDCap CRFs, 21CFR11-compliant device's exported files like those in some CLIA-certified equipment for lab assays

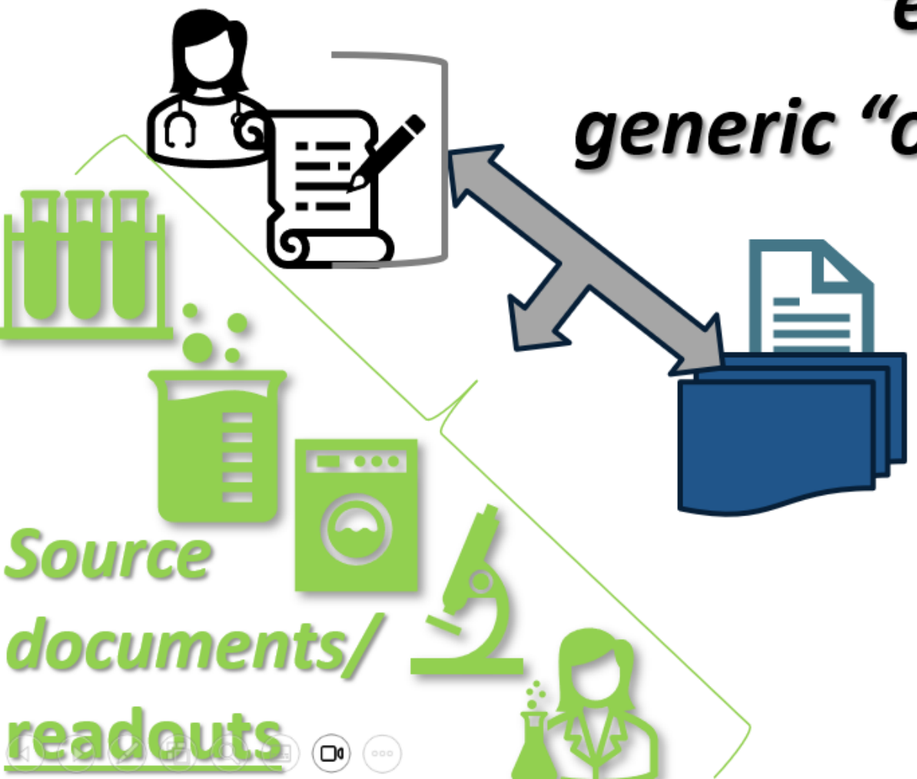
From Research Study Design to Collecting, Managing, & Analyzing Data: **Provenance**

So, just document <u>all</u> *as you organize* (and *map out* data integration)

- **Document rigor: how sources map to database: <u>Source Data Maps</u>**

Example 1-arm trial: note distinct yet equivalent diction for key data

'endpoint' per Protocol Template* specific use *of generic "outcome measure" (term in CTG) CTG=ClinicalTrials.gov



<u>PROVENANCE</u>: 'bigger tent' for 'trace-ability', i.e., audit-ready capacity to trace <u>each</u> quantity back to source information; illustrated below as key part of REPRODUCIBILITY

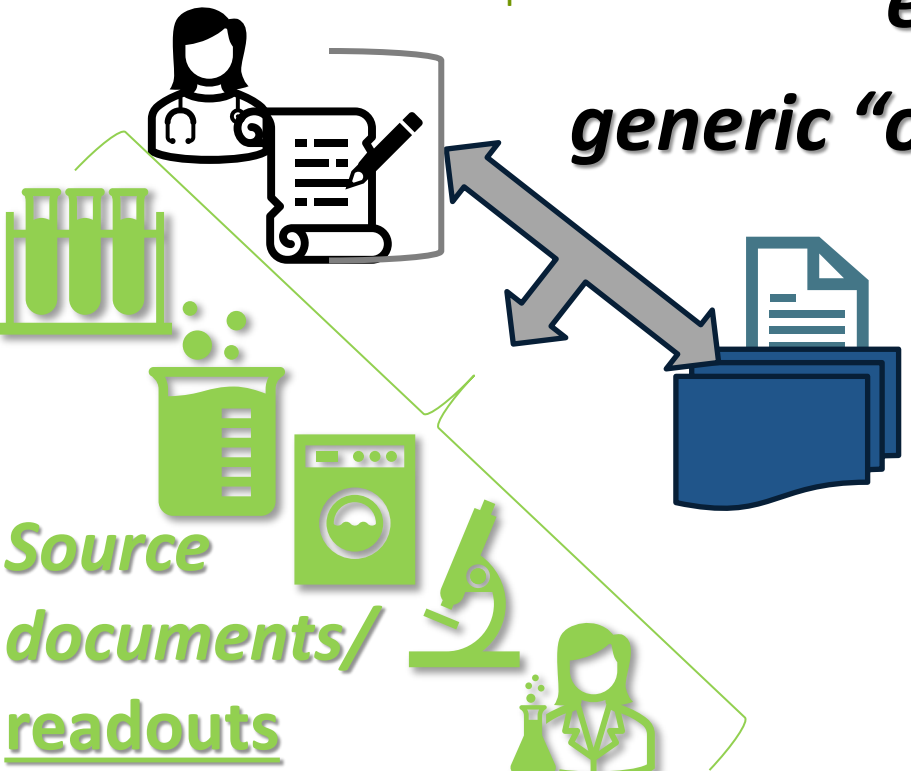
From Research Study Design to Collecting, Managing, & Analyzing Data: **Provenance**

So, just document <u>all</u> *as you organize* (and *map out* data integration)

- **Document rigor: how sources map to database: <u>Source Data Maps</u>**

Example 1-arm trial: note distinct yet equivalent diction for key data

"Model-estimated fold-change in mean urine Protein to Creatine ratio 3 months post-randomization" ***'endpoint' per Protocol Template* specific use *of generic "outcome measure" (term in CTG)*** CTG=ClinicalTrials.gov "Change in proteinuria"



<u>PROVENANCE</u>: 'bigger tent' for 'trace-ability', i.e., audit-ready capacity to trace <u>each</u> quantity back to source information; illustrated below as key part of REPRODUCIBILITY

Principles of Data Collection & Management Part 3 Topics

- Document organization and access as part of study planning: *regulatory, clinical,* and *case report forms*
- **Data Management for Reproducibility: *rigor-anchored transparency***
- Data Management and Sharing Plans
  - *Take Home Points to follow Guiding Principles*

"Data Management spans protocol conception to completion"

-Matt Breymaier, 1st part of this webinar

From Study Design to Collecting, Managing, Analyzing Data: **Defining *Provenance***

BEST PRACTICE: document all *as you organize* (& *map out* how data 'connect')

Source Data Map *key to provenance for each *given product* of a study**

- ***PROVENANCE: audit-ready capacity to trace* each & every *quantity back to source info; key part of REPRODUCIBILITY***

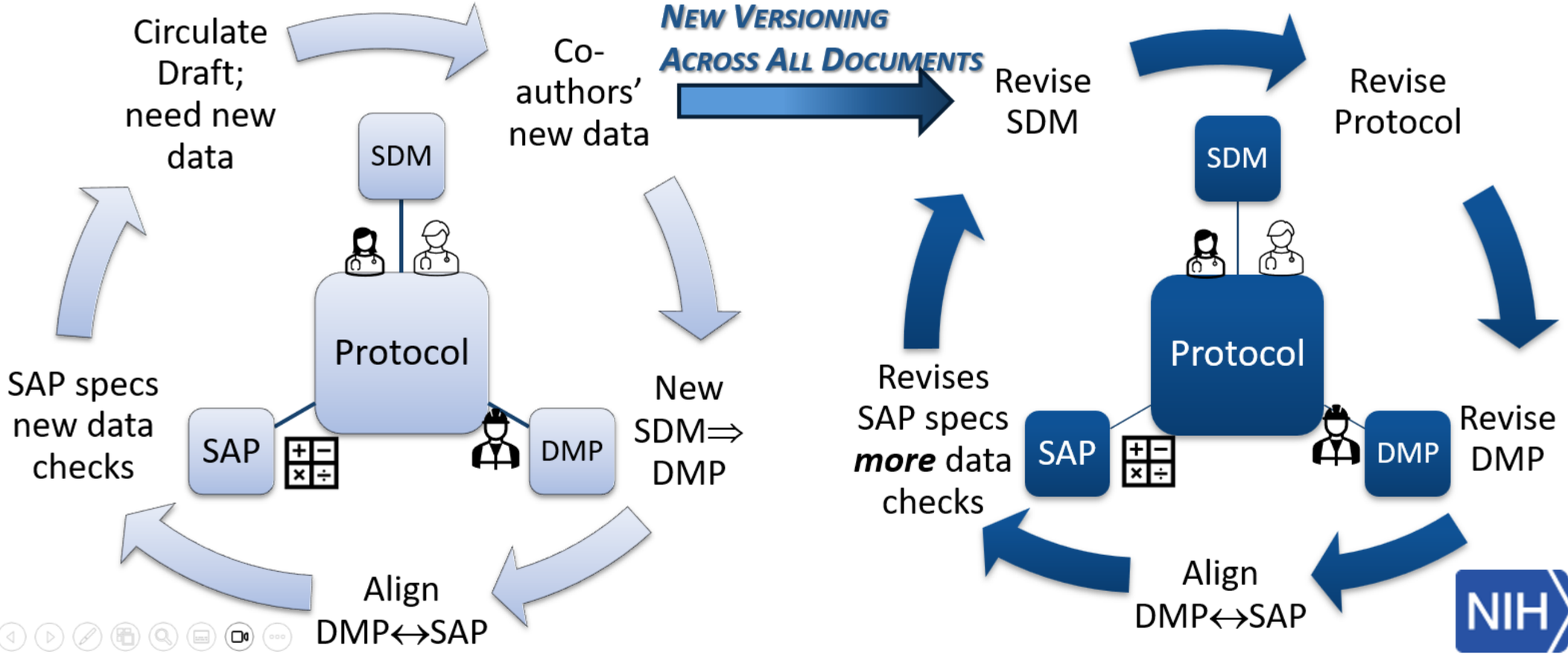
- ***VERSIONING & ACCESS CONTROL (WITH DE-IDENTIFICATION) WITHIN DM ARE KEY***
- ***LACK OF PROVENANCE (AND REPRODUCIBILITY) CAN BE AN ETHICAL ISSUE DUE TO HUMAN ERROR: SEE CANCER TRIALS' UNETHICAL TREATMENT OF >100 PATIENTS STARTED WITH DUKE-PIS' SPREADSHEET-DATA-MANAGEMENT ERROR****

** INITIALLY UNCOVERED DUE TO TRIAL-MOTIVATING STUDY'S FINDINGS BEING BASED ON PUBLICLY-ACCESSIBLE DATA*

BEST PRACTICE: version documents *as you revise* (especially *map* of data)

Source Data Map (SDM) *key to constraints on SAP & DM plan (DMP) for a* study

- ***VERSIONING OF INTERDEPENDENT DOCS & ROLE-BASED ACCESS CONTROL ARE KEY***

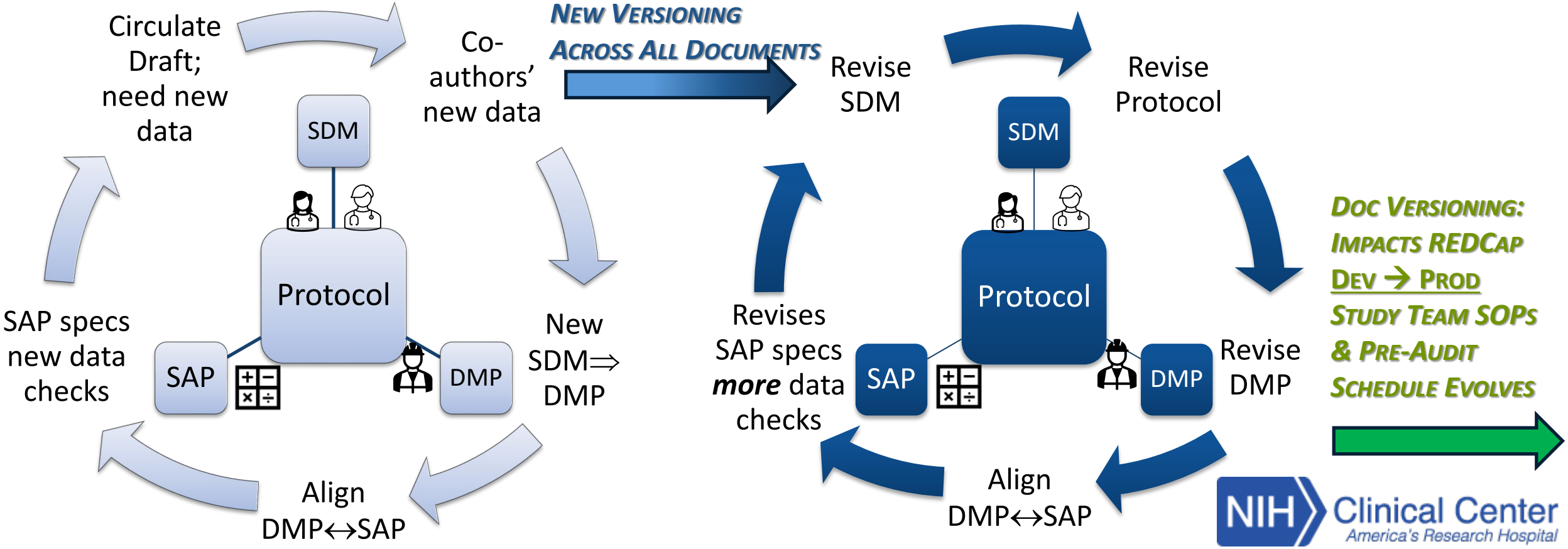


From Study Design to Collecting, Managing, Analyzing Data: **Defining *Provenance***

BEST PRACTICE: version documents *as you revise* (especially *map* of data)

Source Data Map (SDM) *key to constraints on SAP & DM plan (DMP) for a* study

- ***VERSIONING OF INTERDEPENDENT DOCS & ROLE-BASED ACCESS CONTROL ARE KEY***



National Institute of Diabetes and Digestive and Kidney Diseases

From *Prospective* Study Design to Collecting, Managing, Analyzing Data: **Reproducibility**

BEST PRACTICE: document all *as you organize* (& *map out* how data 'connect')

Responsible Team Member	Pre-collection: conceptual	Pre-collection: operational	Collection: at study kickoff	Collection: ongoing	Collection: closeout	Post-collection: conceptual	Post-collection: operational	Post-collection: closeout
Investigator	Set target for completion, Quantify Questions	Input into CRFs, SOPs, monitor data priorities for each aim	Establish plan for versioning changes as we revise protocol	Iterate with team to ensure SOPs followed & data queries resolved	Document & sign off on all remaining unresolved queries	Iterate with team on each data variable's full provenance	Iterate with team on data variables DMSP metadata	Sign off on final DMSP-specified artifacts package
Clinician Co-Investigator/ Study Coordinator	Offer mappable data capture options; input & feedback	Co-curate the CRFs/data capture tools *for ease of use*	SOPs for versioning CRF changes as we revise protocol; monitor use	Iterate with team to ensure SOPs followed & data queries resolved	Document & sign off on all remaining unresolved queries	Iterate with team on each data variable's full provenance	Iterate with team on data variables DMSP metadata	Sign off on final DMSP-specified artifacts, any non-standard data's integrity
Data Mgmt/ Informatics	Compile data capture options	Co-curate the CRFs/data capture tools *for ease of data checks/queries*	SOPs for versioning EDC changes as we revise protocol; monitor use	Implement all pre-specified & novel data checks, query team to resolve	Document all remaining unresolved queries, prep & conduct lock	Iterate with team on each data variable's full provenance & descriptions	Iterate with team on data variables DMSP metadata, code to curate all	Sign off on final DMSP-specified artifacts package, DM & curation code
Data Analysts/ Biostats	Quantify Questions, & coordinate on study power	Co-curate the CRFs/data capture tools to match analysis	SOPs for versioning SAP, safety reports, & (re-)design	Conduct monitoring analyses & spec any new checks	Assess impact of unresolved queries, input on data lock	Iterate with team on full provenance of data findings	Iterate with team on all public analytic metadata, code	Sign off on final DMSP-spec'd package, ingest & analysis code

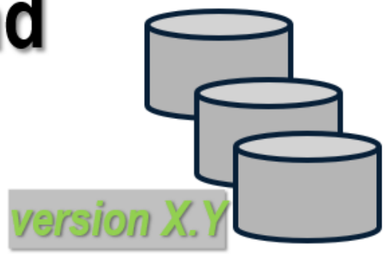
Data Management for Reproducibility: document+EDC revisions *play off one another*

Put yourself in others' shoes: *wouldn't you want your work to be readily reproduced* AND *then provide a starting point for others to* cite *it?*

To do so:* MUST specify *analysis plans in terms of* unambiguous *quantities

SO, each* REPRODUCIBLE *'research product' MUST have its own:

- ***'Statistical Analysis Plan'* *(SAP; tacit: SDM ↔ SAP/DMP)* and**
- ***corresponding 'data lock': versioned 'copy' of data that's***
  - ***non-editable***
  - ***analysis-ready***



National Institute of Diabetes and Digestive and Kidney Diseases

Data Management for Reproducibility: document+EDC revisions *play off one another*

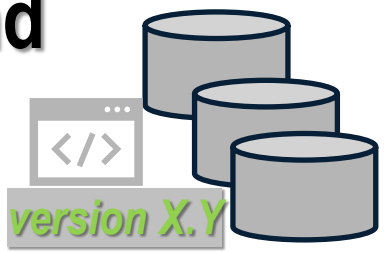
Put yourself in others' shoes: *wouldn't you want your work to be readily reproduced* AND *then provide a starting point for others to cite it?*

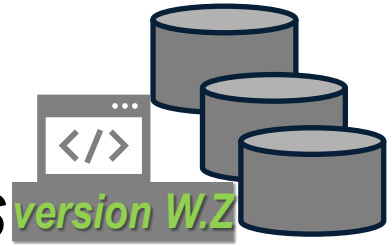
To do so:* MUST specify *analysis plans in terms of* unambiguous *quantities

SO, each* REPRODUCIBLE *'research product' MUST have its own:

- ***'Statistical Analysis Plan'* (SAP; tacit: SDM ↔ SAP/DMP) and**
- ***corresponding 'data lock': versioned 'copy' of data that's***
  - ***non-editable***
  - ***analysis-ready***

** PER EARLIER WEBINAR: analysts expect to use DE-IDENTIFIED datasets*





And thankfully a* planned, milestone-driven *timeframe includes

***SDM/SAP/DMP +'data lock'* adheres to NIH policy (& priorities) by design**

From ***Secondary*** **or** ***Retrospective*** Study Design to Collecting, Managing, Analyzing Data: **Reproducibility**

National Institute of Diabetes and Digestive and Kidney Diseases

BEST PRACTICE: all team members *work in parallel* (& *trace* how data are used) pre- & post-*curation*

Responsible Team Member	Target date 6 or more weeks out	Target date 5 or more weeks out	Target date 4 or more weeks out	Target date 3 or more weeks out	Target date 2 or more weeks out	Target date 1 or more weeks out	*Target date 1 or more weeks past*	*Target date 2 or more weeks past*
Investigator	Share pivotal product's 'concept' target date & data use: SDM	Get team input into variables' use to address aims & outline DM plan for all	Sign initial SAP. affirm all data variable's use to address aims; draft DM plan	Revise/sign SAP & affirm all data summaries' formats by aim, draft narratives	-Circulate draft for final reviews -Draft changes to NEXT pivotal SDM, SAP, DMP	FINALIZE entire pivotal product: triage edits by all co-authors/ contributors	*Iterate with team on each data variable's use in light of pivotal findings*	*Finalize team iteration on data elements' use & curation; approve DM&S*
Clinician Co-Investigator/ Study Coordinator	Provide input & feedback on feasibility re: participant interactions	Input into each data variables' *feasible* use to address aims, checks of data	Address each data variables' *queried issues* for data checks, version input	Select parts of draft of pivotal report's data summaries / narratives	Provide feedback on circulated draft & input NEXT pivotal docs	Support team's pivotal product: triage edits by all co-authors/ contributors	*Iterate on schedule for regular pre-auditing of data & SOP revisions*	*Finalize team iteration on data elements' use & curation; input on DMSP*
Data Mgmt/ Informatics	Provide input & feedback on feasibility re: data input & curation	Input into each data variables' *feasible* use via checks of data *curates/queries*	Track each data variables' *queried issues* for curated data /EDC versioning	Internal reports of all queried data issues & how resolved; versioned read-only data 'lock'	Provide feedback on circulated draft & input to NEXT SDM, SAP, DMP	Implement any data base refactoring for ease of re-use by DMS code	*Iterate on schedule for regular pre-auditing of data & data check + SOP revisions*	*Finalize team iteration on data elements' use & curation; signoff on DM items of DMSP*
Data Analysts/ Biostats	Provide input & feedback re: primary aims'	Input on checks variables' use in *analysis via SAP*	SOPs for version control of SAP, reports design;	FINAL SAP for initial version of pivotal product;	Provide feedback on circulated draft	Implement any reformatting for target venue;	*Iterate on plans for any new data/safety*	*Finalize team iteration per above on*

EXAMPLE: documenting all *as you organize* (& *mapping out* data integration)

Source Data Map *key to provenance** for each *given product* of a study

Pre-Collection/-Curation
Role of Clinical Fellow
Or other Investigator Initiates a new work product… *what principles guide how to document rigor?*
REPRODUCIBILITY

SOURCE DATA MAP: SDM

FULLY INTEGRATED
Clinical Research Database

NIH Clinical Center
America's Research Hospital

*** PROVENANCE: an audit-ready traceability of findings, ensures REPRODUCIBILITY***

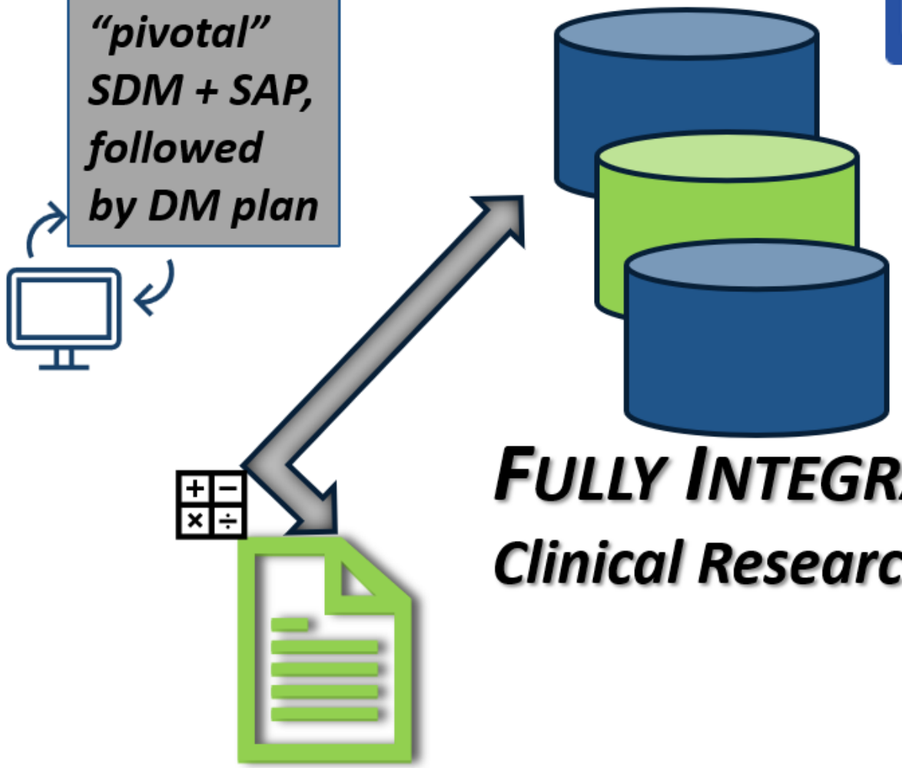
EXAMPLE: documenting all *as you organize* (& *mapping out* data integration)

Source Data Map *key to provenance** for each *given product* of a study

Pre-Collection/-Curation
Role of Clinical Fellow
Or other Investigator Initiates a new work product… *what principles guide how to document rigor?*
REPRODUCIBILITY

"pivotal" SDM + SAP, followed by DM plan

SOURCE DATA MAP: SDM



FULLY INTEGRATED
Clinical Research Database

NIH Clinical Center
America's Research Hospital

** **PROVENANCE**: an audit-ready traceability of findings, ensures **REPRODUCIBILITY***

"pivotal" study data product:

- ***Example: secondary study***

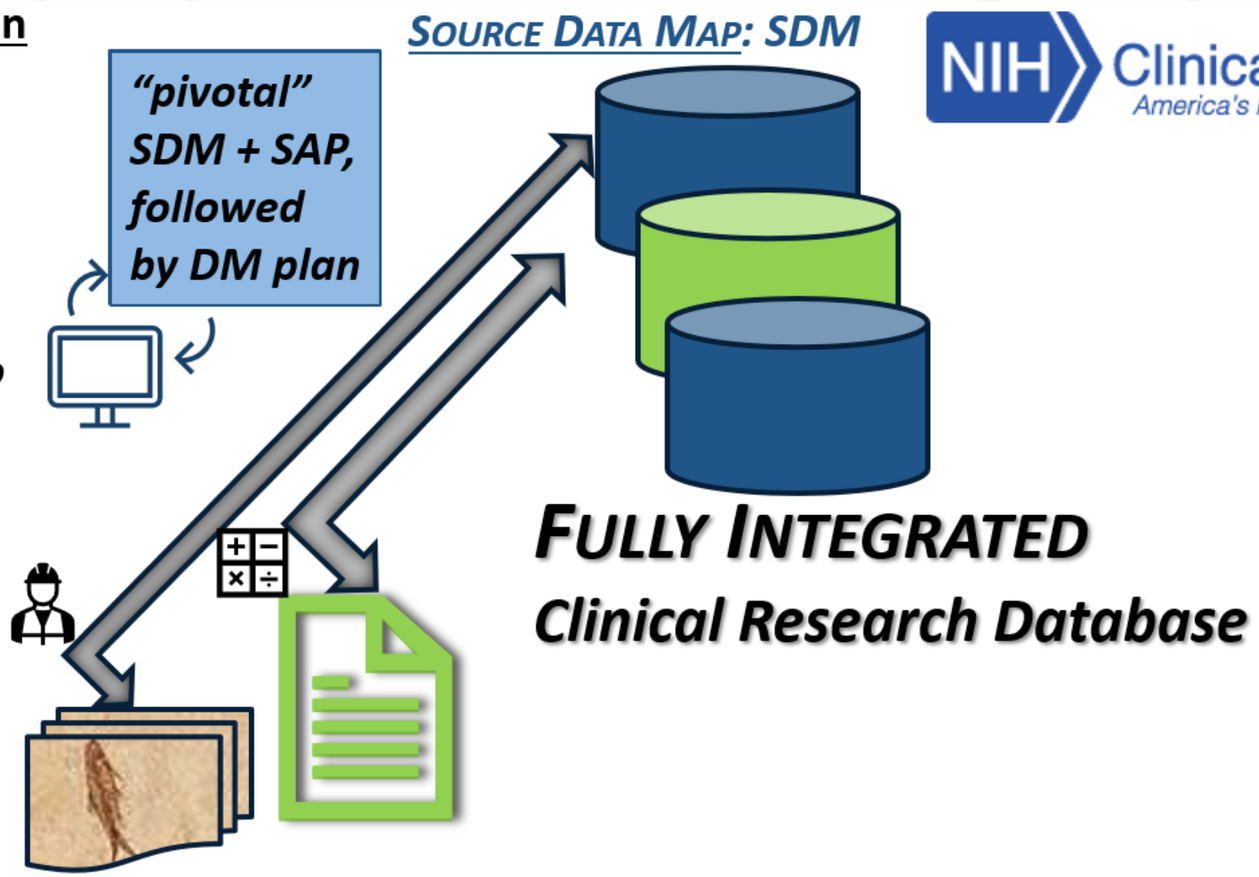
EXAMPLE: documenting all *as you organize* (& *mapping out* data integration)

Source Data Map *key to provenance** for each *given product* of a study

Pre-Collection/-Curation
Role of Clinical Fellow
Or other Investigator Initiates a new work product… *what principles guide how to document rigor?*
REPRODUCIBILITY

"pivotal" study data product:

- ***Example: secondary study***



NIH Clinical Center
America's Research Hospital

** PROVENANCE: an audit-ready traceability of findings, ensures REPRODUCIBILITY*

National Institute of Diabetes and Digestive and Kidney Diseases

From *Secondary* Study Design to Collecting, Managing, Analyzing Data: **Reproducible Doc'n**

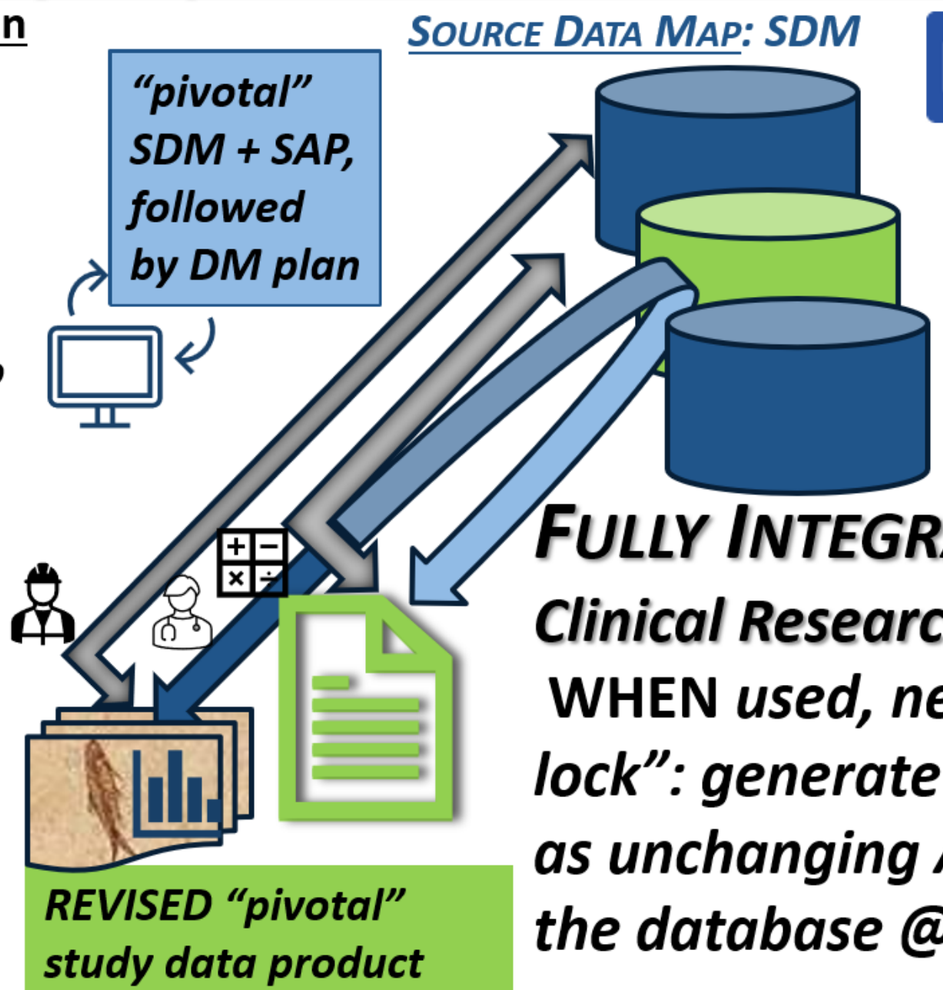
EXAMPLE: documenting all *as you organize* (& *mapping out* data integration)

Source Data Map *key to provenance** for each *given product* of a study

Pre-Collection/-Curation
Role of Clinical Fellow
Or other Investigator Initiates a new work product… *what principles guide how to document rigor?*
REPRODUCIBILITY

"pivotal" study data product:

- ***Example: secondary study***



NIH Clinical Center
America's Research Hospital

FULLY INTEGRATED
Clinical Research Database
WHEN *used, need "data lock": generate a* static *copy as unchanging* ARCHIVE *of the database @ time of use*

*** PROVENANCE: an audit-ready traceability of findings, ensures REPRODUCIBILITY***

National Institute of Diabetes and Digestive and Kidney Diseases

From *Secondary* Study Design to Collecting, Managing, Analyzing Data: **Reproducible Doc'n**

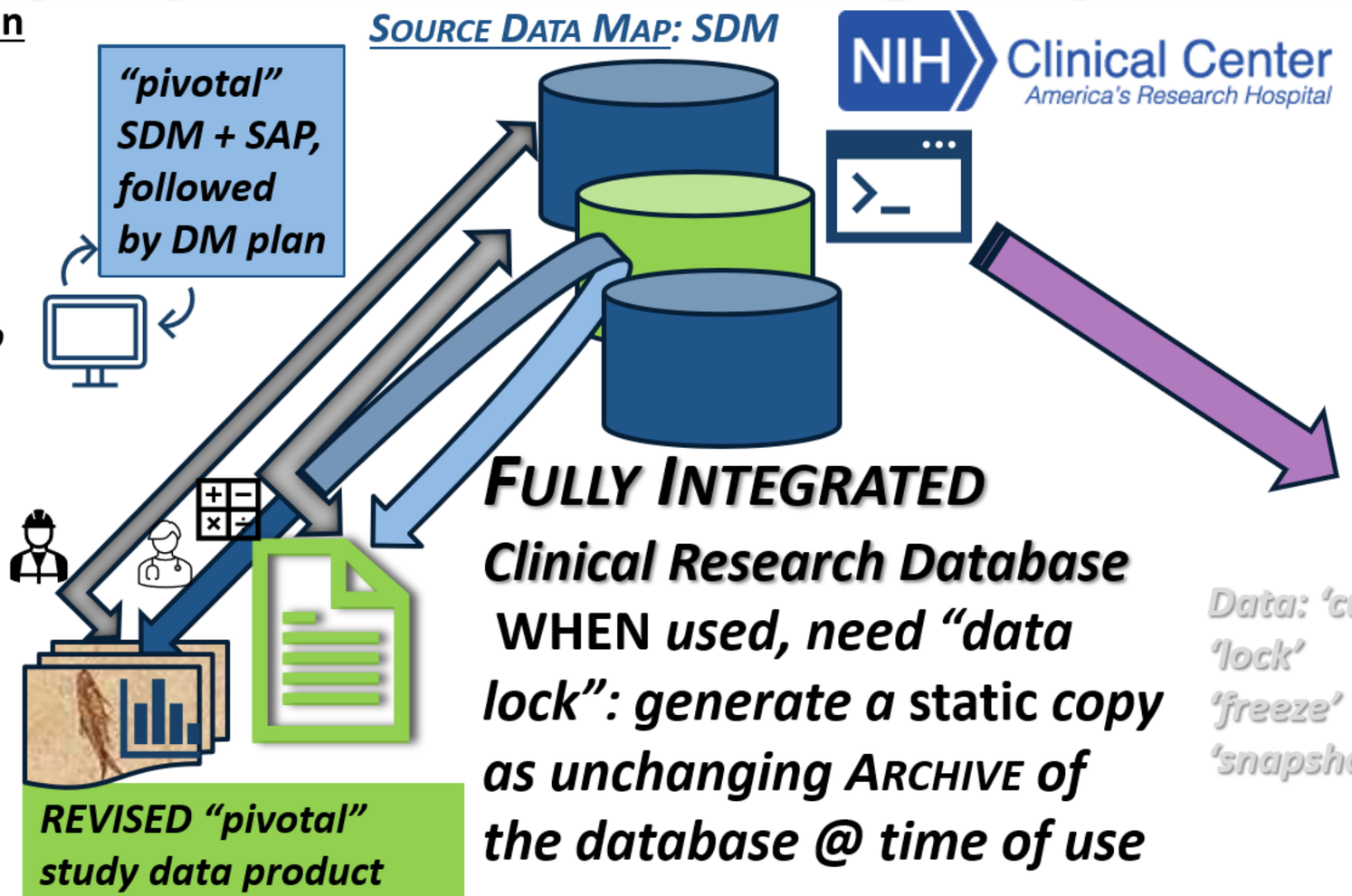
EXAMPLE: documenting all *as you organize* (& *mapping out* data integration)

Source Data Map *key to provenance** for each *given product* of a study

Pre-Collection/-Curation
Role of Clinical Fellow
Or other Investigator Initiates a new work product… *what principles guide how to document rigor?*
REPRODUCIBILITY

"pivotal" study data product:

- ***Example: secondary study***



FULLY INTEGRATED
Clinical Research Database
WHEN *used, need "data lock": generate a* static *copy as unchanging* ARCHIVE *of the database @ time of use*

Data: 'cut' 'lock' 'freeze' 'snapshot'

** PROVENANCE: an audit-ready traceability of findings, ensures* REPRODUCIBILITY

Secondary or Retrospective Study Design to Collecting, Managing, Analyzing Data: **Reproducible per *Documentation***

National Institute of Diabetes and Digestive and Kidney Diseases

BEST PRACTICE: all team members' subproducts revised & versioned *in parallel* (via *edits of* data-use documents)

Responsible Team Member	Target date 6 or more weeks out	Target date 5 or more weeks out	Target date 4 or more weeks out	Target date 3 or more weeks out	Target date 2 or more weeks out	Target date 1 or more weeks out	*Target date 1 or more weeks past*	*Target date 2 or more weeks past*
Investigator	CONCEPT; DATA DICTIONARY => SOURCE DATA MAP (SDM); DM PLAN & SAP OUTLINES; OUTLINE SOPS	SDM REVISION PER SAP OUTLINE; DM PLAN DRAFT *outline* of EDC/ DM/DA SOPs	initial SAP, signed alongside final SDM; initial DM & DMS plans	Initial draft of pivotal report's data summaries alongside draft narratives	-Circulate draft for final reviews -Draft changes to NEXT pivotal SDM, SAP, DMS	FINALIZE entire pivotal product: triage edits by all co-authors/ contributors	*Iterate with team on each data variable's use in light of pivotal findings*	*Finalize team iteration on data elements' use & curation; approve DMS*
Clinician Co-Investigator/ Study Coordinator	DATA DICTIONARY => SOURCE DATA MAP (SDM); FEASIBILITY RPT	SDM REVISION PER SAP OUTLINE; DM PLAN DRAFT *outline* of data check/cleaning	Address each data variables' *queried issues* for data checks, version input	Select parts of draft of pivotal report's data summaries / narratives	Provide feedback on circulated draft & input NEXT pivotal docs	Support team's pivotal product: triage edits by all co-authors/ contributors	*Iterate on schedule for regular pre-auditing of data & SOP revisions*	*Finalize team iteration on data elements' use & curation; input on DMSP*
Data Mgmt/ Informatics	DATA DICTIONARY => SOURCE DATA MAP (SDM); DM PLAN OUTLINE	SDM/SAP DRAFT; DM PLAN DRAFT *outline* of data check/cleaning drafts of check-*rules/queries*	Track each data variables' *queried issues* for curated data /EDC versioning	Internal reports of all queried data issues & how resolved; versioned read-only data 'lock'	Provide feedback on circulated draft & input to NEXT SDM, SAP, DMP & DMS package	Implement any data base refactor doc'n for ease of re-use by DMS code by others	*Iterate on schedule for regular pre-auditing of data & data check + SOP revisions*	*Fully Revised Data Management & Sharing Plan (DSMP)*
Data Analysts/ Biostats	DATA DICTIONARY => SDM; SAP OUTLINE; DM PLAN OUTLINE	SDM/SAP DRAFT; DM PLAN DRAFT *code/modeling outline*	SOPs for version control of SAP, reports' code; initial version	FINAL SAP for initial version of pivotal product; draft summary	Revise formats for summaries per feedback;	Implement any reformatting for target venue;	*Iterate on plans for any new data/safety*	*Fully Revised data/safety monitoring Plan (DSMP)*

National Institute of Diabetes and Digestive and Kidney Diseases

Principles of Data Collection & Management Part 3 Topics

- Document organization and access as part of study planning: *regulatory, clinical,* and *case report forms*
- Data Management for Reproducibility
- **Data Management and Sharing Plans: "DMSPs"**
  - *Take Home Points to follow Guiding Principles*

"***Data Collection Practices*** **[have]** ***Do's and Don'ts***"
-Sai Theja, 2nd part of this webinar

"***Data Management & Sharing Practices have Plans that must be shared per NIH policy***"
-Ken Wilkins, 3rd part of this webinar

The sample Data Management and Sharing Plan below is for a proposal conducting clinical research with human participants. It is one of four examples provided by NIDDK.

NIDDK Example Data Management and Sharing Plan – Clinical Data

Element 1: Data Type:

A. Types and amount of scientific data expected to be generated in the project:
This study will collect renal dialysis data from multiple clinics. Demographic, laboratory results, clinical observations, and clinical disposition will be acquired from 250 affected participants and 250 matched healthy controls.

CORE TRUST SEAL
WORLD DATA SYSTEM

Core Trustworthy Data Repository

National Institute of Diabetes and Digestive and Kidney Diseases

From Design to Collecting/Managing/Analyzing: Data Management *and* Sharing (DMS) Plans

Hmm… heard we need a 'DMS' plan to start any research…

THERE'S A POLICY?

Yes, and it applies intramurally* AND *extramurally

many resources offered for such

SO HAVE YOUR DATA READY TO SHARE

When? @ time of dissemination… thus flesh out DMSP @ time of data specs

Ties directly to NIH policies on rigor* and *reproducibility

[EXEMPLARS FIND ELECTRONIC LAB NOTEBOOKS/MARKDOWN* DOCS HELPFUL TO DO IT]

** SEE FORTHCOMING WEBINAR "R IS FOR ALL" NEXT WEEK, 31ST OF JULY 2025: NAVIGATE TO NIDDK OR NCI'S BTEP SITES*

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to *Usable* Data: *Adhering* to NIH Data Mgmt. & Sharing Policy

Data Management & Sharing Plan (DSMP) Elements: *more answers, fewer questions*

Per NIH Policy Now in Effect, **DMS Plan Elements:**

1. **Data Type** (may evolve by data dictionary versioning)
2. **Related Tools, Software, and/or Code** (e.g., REDCap)
3. **Standards** (e.g., like Common Data Elements or CDEs)
4. **Data Preservation, Access, & Associated Timelines** *more below*
5. **Access, Distribution, or Reuse Considerations**
6. **Oversight of Data Management and Sharing**

Elements of Data Management & Sharing Plan

Per NIDDK's Public Resources, **Developing a DMS Plan follow series of stages:**

a) Evaluate study design and objectives

b) Identify data types that will be generated

c) Determine applicability of the policy to your research data

d) Consider standards and related tools appropriate for your research data

e) Select one or more repositories* by considering key facets

NIDDK-CR Resources for Research (R4R)

NIDDK Central Repository - Resources for Research (NIDDK-CR R4R) facilitates sharing of data, biospecimens, and other resources generated from studies supported by NIDDK and within NIDDK's mission by making these resources available for request to the broader scientific and research community.

* *See NIDDK-Central Repository R4R WEBINAR SERIES*

From Research Study Design to *Usable* Data: *Adhering* to NIH Data Mgmt. & Sharing Policy

NIH Policy in Action, Following DMS Plan by Example: follow by-study-type templates (as found under NIDDK's Tools & Resources)

in related news, REDCap can help by its metadata, access to CDEs

SETTING ASIDE POLICY, THINK OF *REGULATIONS*… Adherence to Human Subjects Research Protections is consequence of Good Clinical Practice (GCP)—RECALL: "DATA MANAGEMENT IS OPERATIONALIZATION OF GCP"

The sample Data Management and Sharing Plan below is for a proposal conducting clinical research with human participants. It is one of four examples provided by NIDDK.

NIDDK Example Data Management and Sharing Plan – Clinical Data

Element 1: Data Type:

A. Types and amount of scientific data expected to be generated in the project:
This study will collect renal dialysis data from multiple clinics. Demographic, laboratory results, clinical observations, and clinical disposition will be acquired from 250 affected participants and 250 matched healthy controls.

B. Scientific data that will be preserved and shared and the rationale for doing so:
Identifiable data will be de-identified prior to repository submission. Participant-level clinical data described in A will be preserved through deposition of the data in a controlled access public repository.

C. Metadata, other relevant data, and associated documentation:
The study protocol, date collection forms/case report forms, data dictionary, manual of operations, and a glossary of domain-specific terms will be submitted.

Element 2: Related Tools, Software, and/or Code:

The clinical data will be analyzed with custom R code and visualized with the ggplot2 package. R packages are all freely available via R CRAN. All code will be shared via a tagged GitHub repository and a readme.md file for the project describing the workflow, relationship between code, instructions, and parameter choices for selected tools.

Element 3: Standards:

Participant age, sex, ethnicity, height, weight, socioeconomic status, and dialysis data will be collected using the common data elements (CDEs) from the National Institutes of Health (NIH) CDE Repository.

(1) Demographics (NLM ID: Xyc4G1BHte)
(2) Standing Height (NLM ID: gaz3k9xh1da)
(3) Weight (NLM ID: ILbYoUaBc)
(4) Socioeconomic Status (NLM ID: 7kpJeKE7P)
(5) Dialysis (NLM ID: 71WP2zp2ox)

Element 4: Data Preservation, Access, and Associated Timelines:

National Institute of Diabetes and Digestive and Kidney Diseases

Principles of Data Collection & Management Part 3 Topics

- Document organization and access as part of study planning: *regulatory, clinical,* and *case report forms*
- Data Management for Reproducibility
- **Data Management and Sharing Plans**
  - ***Take Home Points to follow Guiding Principles***

"DATA MANAGEMENT IS THE OPERATIONALIZATION OF GOOD CLINICAL PRACTICE (GCP)"
- INTRODUCTION TO PRINCIPLES & PRACTICES OF CLINICAL RESEARCH (COURSE OFFERED AT NIH)

In not limited in time, can intro users to CDE API use

From Research Study Design to *Usable* Data: *Upfront* Planning per Guiding Principles

…IF you don't *FRONT-load* these decisions, issues arise too late *to address*

Get *more* return by *less* upfront effort: *avoid findings* raising more *questions*

Guiding Principles stemming from FAIR data principles*

What does *FAIR* mean, especially for research data management?

- *QUITE A BIT: Research is Reproducible, if Data are F.A.I.R.*
- *To wit: FA in place via DMS policy, IR via CDE* Permissible Values
- *Our Resources will give you each a starting point…*
- *Whole courses at NIH Library on this, AND NLM has open tutorials*
- *NIH-ecosystem-wide, ODSS has a FAIR Data & Resources Unit*

Findable
Accessible
Interoperable
Reusable

*First published in 2016

ODSS = Office of Data Science Strategy, https://datascience.nih.gov/ responsible for 2025-30 strategic plan

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to *Usable* Data: *Practical* Take Homes *if* at Clinical Center

To be glib:

wouldn't you want to avoid "GIGO: garbage in, garbage out?

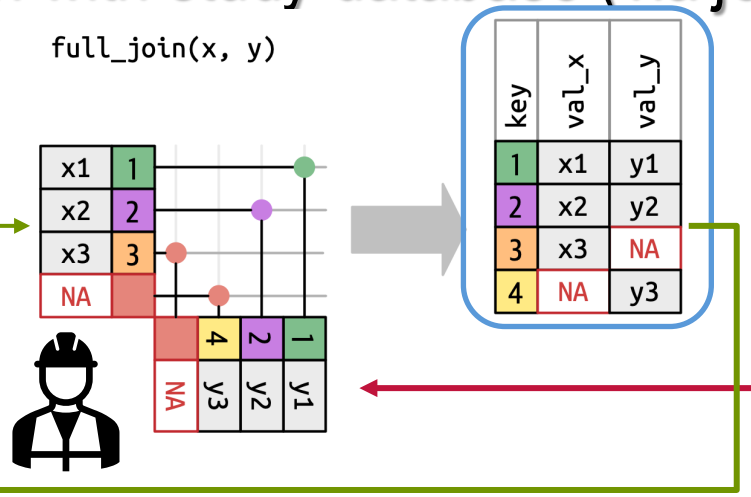
You have options if* within *the Intramural Research Program

Inherent capacity to leverage Clinical Center based infrastructure

- *Direct ingest of* entries in Clinical Research Information System (**CRIS**)
- Such 'back-end' integration with study database (via joins) allow use of *all* its varied data

REDCap
Research Electronic Data Capture

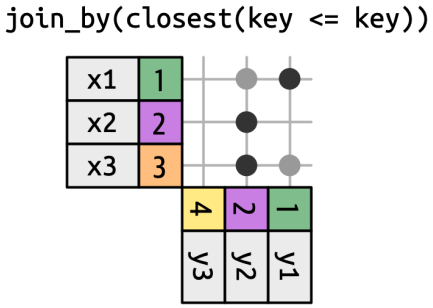
As done already within NIDDK safety reports made using tableau

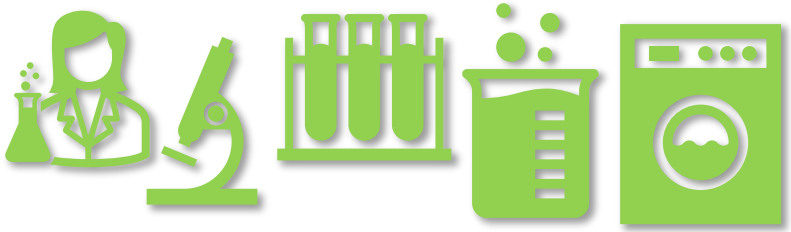


Even more options if* within *the Intramural Research Program

Inherent capacity to leverage Clinical Center based infrastructure

- *Direct ingest of* entries in Biomedical Translational Research Information System
  - Easy to search up useful data via **U**ser **F**acing **O**ntologies (**UFO**s) in its User Interface
  - Ease due to BTRIS team involving clinicians who created a Medical Entities Dictionary
- ODSS-funded upgrades in NIH CC Department of Clinical Research Informatics make *all* easier to do!

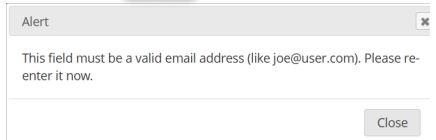




- Core facilities *bake in* own data collection/curation processes, adding to *ease* of joins!

From Research Study Design to *Usable* Data: *Practical* Take Homes on Guiding Principles

To be glib: ***wouldn't you want to avoid "GIGO: garbage in, garbage out?***



You have a few more 'glib' take home points to go by:

- ***Most salient domain expertise: specify the data variables' reasonable values***
- ***If at all possible try to 'touch' each data element's value only once: collection***
- ***If you have to 'touch' a data element > 1 time, limit to twice or thrice AND early***
- **If it ends up being used in analysis, it MUST be checked>queried>corrected *PRIOR***

- ***If you work with a data analyst to do statistics, then above take-homes will align to GCP's principles as applied to study design/analysis: Good Statistical Practice***
- ***If you work with seasoned informaticists, like Matt & Sai, you can curate FDA-ready data sets using Submission-Ready data models like CDISC's (or OMOP FOR RWD VIA FHIR)***

From Research Study Design to *Usable* Data: *OVERALL* Take Homes

To summarize: ***couldn't you now note that you've met 1-2 of below objectives?***

You could,* now *if not* after *engaging resources in NIH IRP & elsewhere (linked above)

1. **To delineate features of REDCap to support project management for research studies (e.g., how different types of studies [longitudinal vs cross-sectional etc] can be designed).**
2. **To outline steps to create detailed data collection plans which fulfill regulatory requirements.**
3. **To identify principled approaches to data collection and management.**
4. **To explain the connections between research rigor and reproducibility.**

Any remaining questions?

Questions? Comments? Other feedback or concerns?

"Data management needs to be forward looking."

[more IPPCR text's pithy take homes below]

"If data is collected in a way that it will never be examined when the original study is closed, it is realizing a fraction of its usefulness."

"Although good data management practices cannot make up for poor study design, poor data management can render a perfectly executed [study] useless."

National Institute of Diabetes and Digestive and Kidney Diseases

Other helpful resources /links within NIH SharePoint:

"During the conduct of a [study], the research should put him- or herself in the role of someone trying to discredit the study's conclusions and question every procedure that could cast doubt on the accuracy, validity, or relevance of the data collected. Every research conclusion is an argument, and the conclusions will only stand if the data stand."

"the failure of a study to produce generalizable knowledge because of bad data management carries both resource and ethical costs."

NIH NIDDK

75 YEARS 1950–2025

NIH

NIDDK

75 YEARS

1950–2025

Other helpful resources /links within NIH SharePoint:

Advancing Research & Health for All

ADDITIONAL SLIDES TO ROUND OUT PRACTICAL SIDES of *objectives*

Find MORE resources within our NIH SharePoint folder here

Webinar Objectives

1. **To delineate features of REDCap to support project management for research studies (e.g., how different types of studies [longitudinal vs cross-sectional etc] can be designed).**
2. **To outline steps to create detailed data collection plans which fulfill regulatory requirements.**
3. **To identify principled approaches to data collection and management.**
4. **To explain the connections between research rigor and reproducibility.**

Explore slides to address residual questions, YET feel free to follow up via email: WILKINSKJ@NIDDK.NIH.GOV

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to *Usable* Data: Upfront Plans per FAIR Guiding Principles

...IF you don't *FRONT-load* these decisions, issues arise too late *to address*

Get return on upfront effort: *get data that* answers your *research question*

Guiding Principles stemming from FAIR data principles, dating to 2016:

- **What does *FAIR* mean for research data management?**
  - *QUITE A BIT, yet let's give you each a starting point…*
  - *Whole courses at NIH Library on this, AND NLM has open tutorials*
  - *NIH-ecosystem-wide, ODSS has a FAIR Data & Resources Unit*

Can readily FRONT-load by RE-using existing Data Interoperability Standards:

- ***Common Data Elements (CDEs)***
- ***Common Data Models (real world data)***
- ***Data Exchange/Interchange Standards***

Findable

Accessible

Interoperable

Reusable

NIH National Library of Medicine NIH CDE Repository

NIH Endorsed CDEs Accelerate Research

OMOP pcornet The National Patient-Centered Clinical Research Network cdisc

NIH National Institutes of Health Office of Data Science Strategy

ODSS = Office of Data Science Strategy, https://datascience.nih.gov/ responsible for 2025-30 strategic plan

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to *Usable* Data: FAIR Guiding Principles, relative to AI

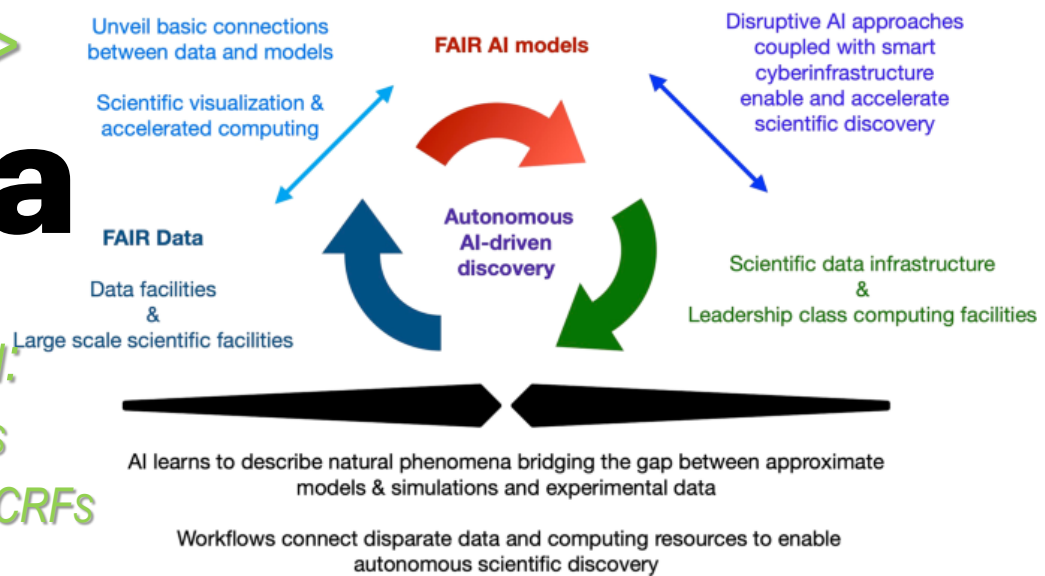
Questions to be answered: *want data to answer* your *research question?*

Guiding Principles stemming from FAIR data principles, updated to 2025

- **What does *FAIR* mean for downstream research data re-use?**
  - *Some meta-researchers say, Findable AND AI-Ready as 'FAIR'*
  - *Machine learning community proposed the Croissant Metadata standard for this*
  - *Has, in turn, led to AI models becoming FAIR=>*

scientific **data**

- *Our focus: Findable Accessible Interoperable Reusable*
- *FAIR in Regulatory Context Covered by Matt & Sai:*
  - *Clinical Data Interchange Standards Consortium (CDISC) & its*
  - *Clinical Data Acquisition Standards Harmonization (CDASH) e-CRFs*



APPENDIX on REDCap: add'l slides

REDCap
Research Electronic Data Capture

NIH NIDDK 75 YEARS 1950–2025

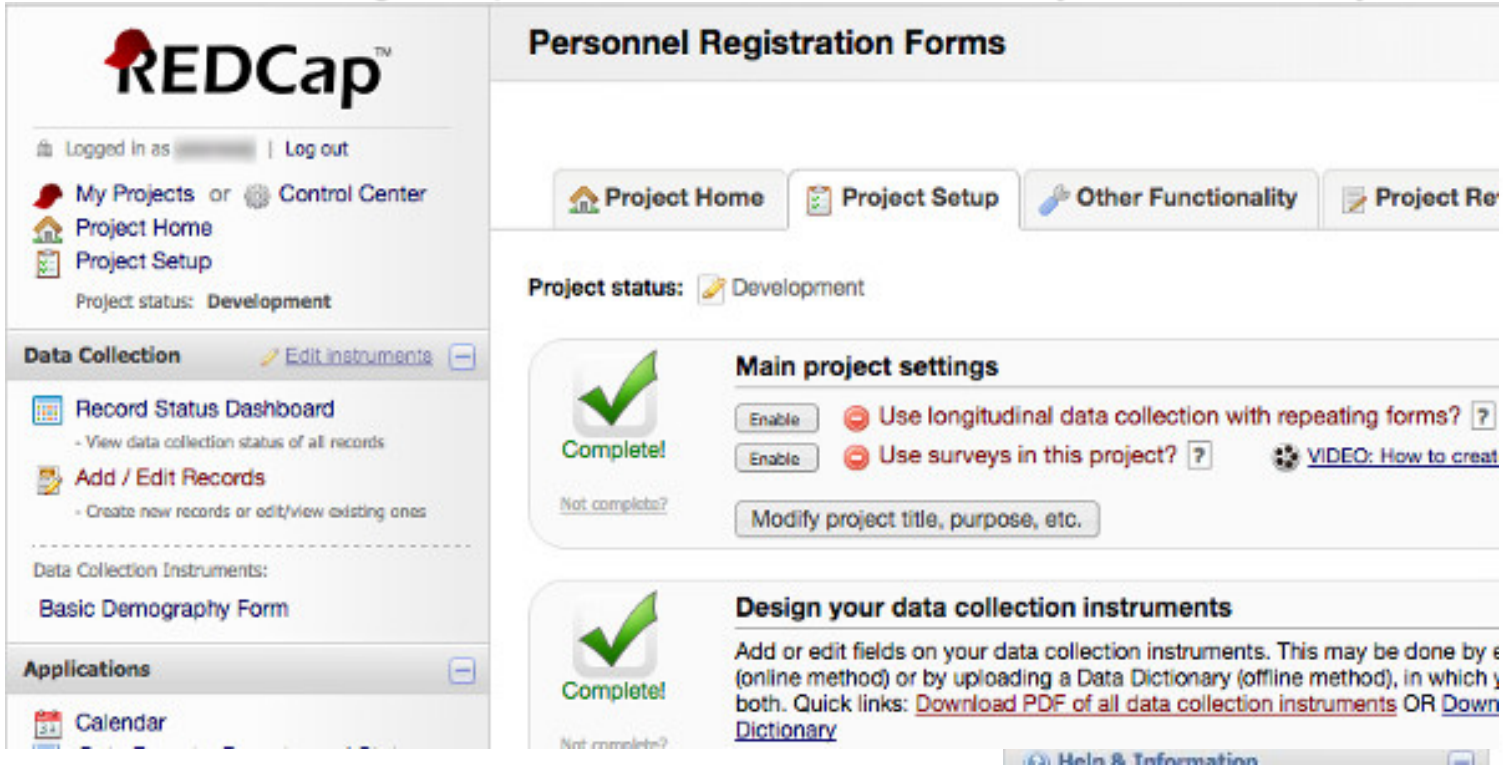
National Institute of Diabetes and Digestive and Kidney Diseases

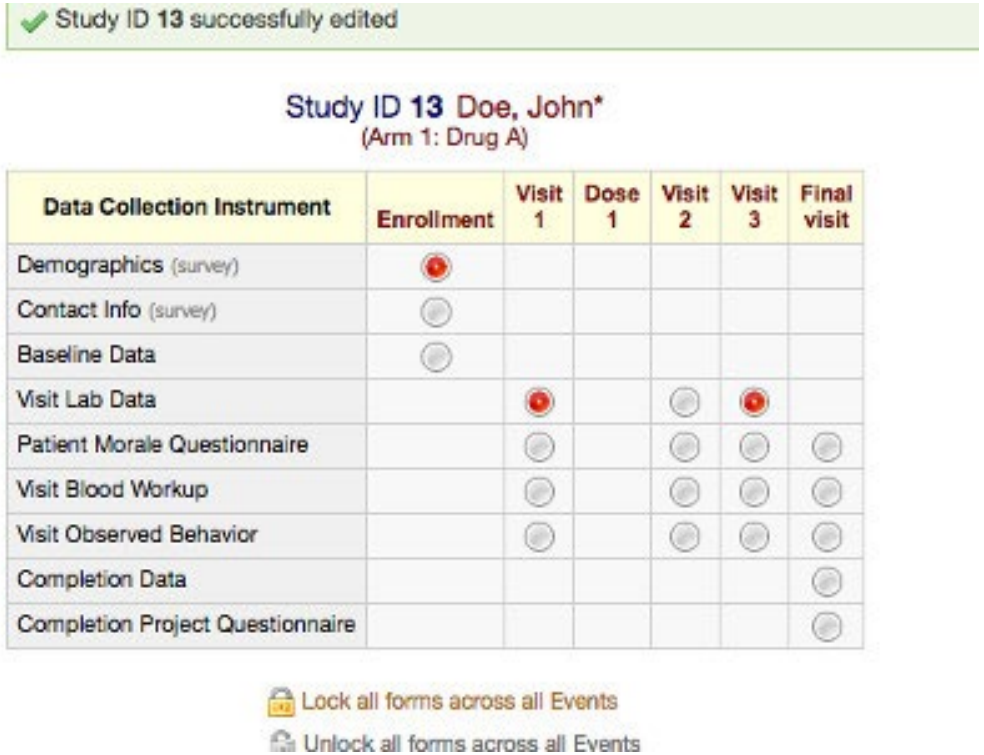
From Research Study Design to *Usable* Data: Take Home Examples on Guiding Principles

You have pre-vetted starting point* because of *Intramural Research Program REDCap

Much of our flexibility of data specifications *enabled* by how it's been configured initially

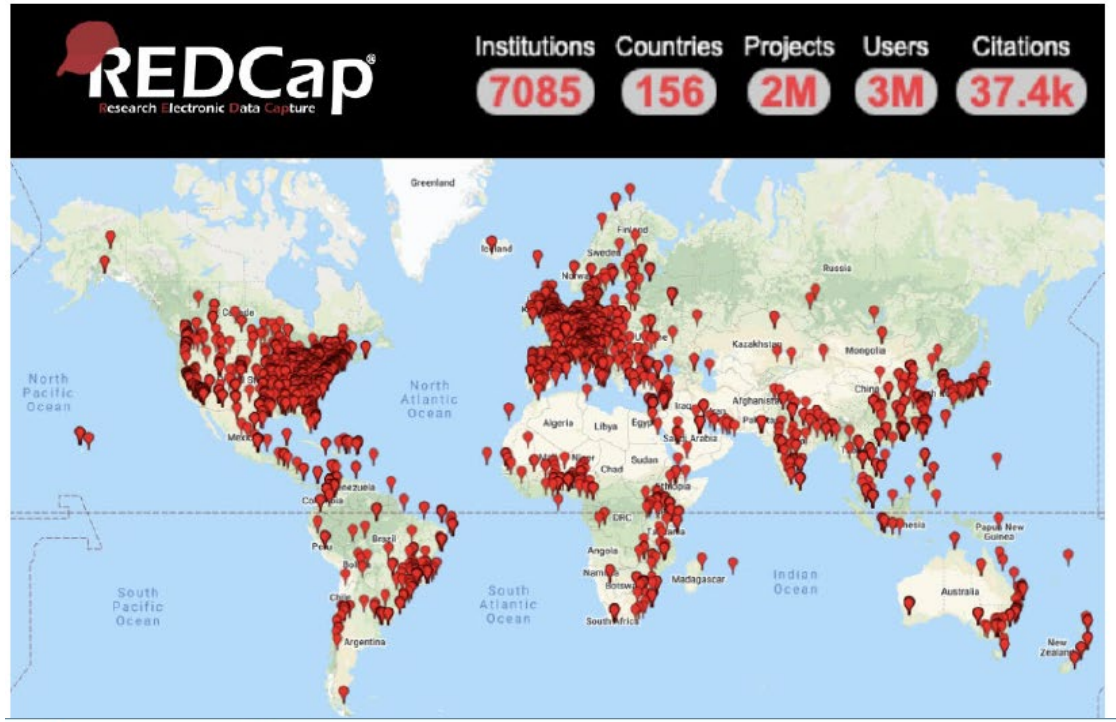
- *Some other features keep getting added centrally by a worldwide REDCap development community*
  - Can manage* operational workflows as you conduct your research, adapt CRFs for *longitudinal* reuse

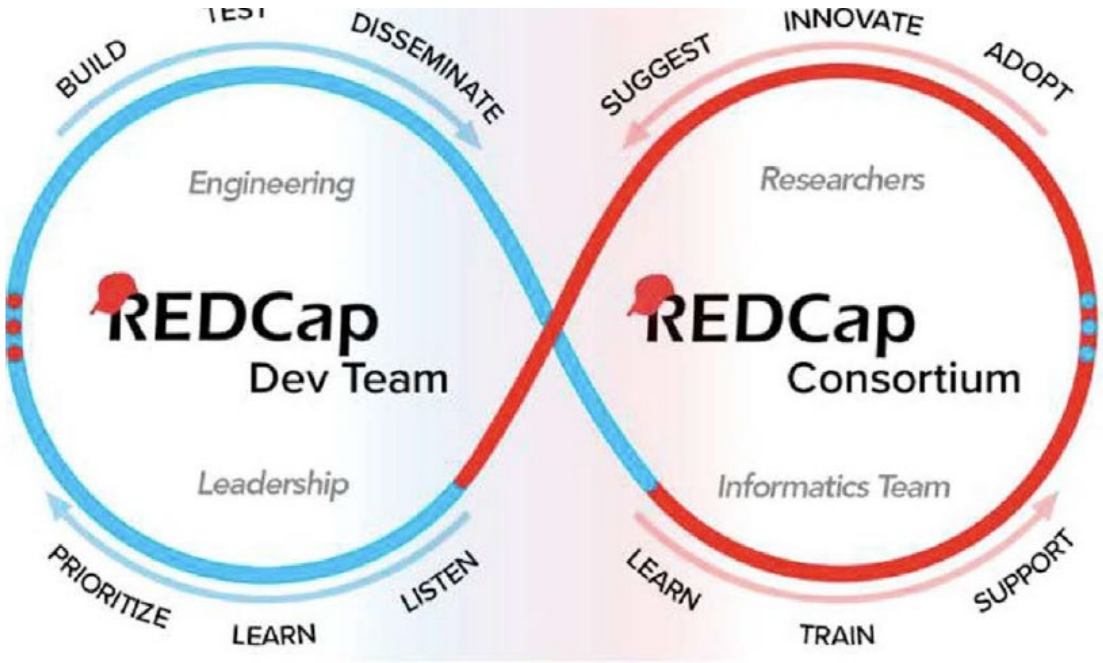




*To delineate features of REDCap to support project management for research studies/designs, as covered by Matt & Sai earlier.

You could leverage a *very* broad user community





NIH National Institutes of Health Office of Portfolio Analysis iCite

New Analysis | Global RCR Stats | How to cite | Help

Results

Citing Papers | Referenced Papers

Influence | Translation | Citations

Roll over table headers for definitions; visit the Global RCR Stats page for percentile tables

~ 1000 x average publication citation rate

Total Pubs	Pubs Per Year	Cites Per Year				Relative Citation Ratio (RCR)				Weighted RCR
		MAX	MEAN	SEM	MED	MAX	MEAN	SEM	MED	
1	1.00	1887.60	1887.60	0.00	1887.60	995.65	995.65	0.00	995.65	995.65

In the event of limited time skip to CDE API use

From REDCap founder Paul Harris presentation in March 2024

NIH NIDDK 75 YEARS 1950–2025

Leverage user community: *Regulatory* Compliance

Keyword search: cdash

Search the library

Search options:

Language: - All -

Type: show all

Minimum downloads: 0

Recent additions: show all

Shared Library
Search
Library Metrics
Consortium Activity

1 2 3 >>

Found 54 results matching your search

Didn't find what you were looking for? Suggest a validated instrument for library inclusion

Title	Downloads
CDISC|CDASHIG v2.1|Adverse Events	123
CDISC|CDASHIG v2.1|Clinical Events	48
CDISC|CDASHIG v2.1|Concomitant Medications	103
CDISC|CDASHIG v2.1|Death Details	42
CDISC|CDASHIG v2.1|Demographics	182
CDISC|CDASHIG v2.1|Disposition	56
CDISC|CDASHIG v2.1|Drug Accountability Horizontal - Dispensed Amount	25
CDISC|CDASHIG v2.1|Drug Accountability Horizontal - Returned Amount	19
CDISC|CDASHIG v2.1|Drug Accountability Vertical	25
CDISC|CDASHIG v2.1|ECG Test Results - Central Reading	26
CDISC|CDASHIG v2.1|ECG Test Results - Local Reading	31
CDISC|CDASHIG v2.1|Exposure as Collected	36
CDISC|CDASHIG v2.1|Findings About Events or Interventions	29
CDISC|CDASHIG v2.1|Healthcare Encounters	36
CDISC|CDASHIG v2.1|Inclusion/Exclusion Criteria	85
CDISC|CDASHIG v2.1|Laboratory Test Results - Central Processing	41
CDISC|CDASHIG v2.1|Laboratory Test Results - Local Processing	51
CDISC|CDASHIG v2.1|Log Form Prompt	29
CDISC|CDASHIG v2.1|Medical History	93
CDISC|CDASHIG v2.1|Microbiology Specimen Central Processing	18

1 2 3 >>

Converts CDISC CDASH eCRF instrument metadata into REDCap data dictionaries for the purpose of simplifying adoption and use of CDASH instruments by research teams across the REDCap Consortium

Journal of the Society for Clinical Data Management

Cheng AC, et al. Creating and Disseminating CDASH Harmonization Electronic Case Report Forms on the REDCap Shared Data Instrument Library. *Journal of the Society for Clinical Data Management*. 2022; 2(1): 7, pp.1–5. DOI: https://doi.org/10.47912/jscdm.172

ORIGINAL RESEARCH

Creating and Disseminating CDASH Harmonization Electronic Case Report Forms on the REDCap Shared Data Instrument Library

Alex C. Cheng*, Rhonda Facile†, John Owen†, Richard Marshall†, Kathleen Mellars†, Nan Kennedy*, Brenda L. Minor*, Kyle Mcguffin* and Paul Harris*

Introduction: A guiding principle behind the development and deployment of the REDCap data management platform has always included attention to workflow design that allows easy implementation of best practices for clinical and translational researchers. CDISC standards such as CDASH have helped the clinical research community improve the efficiency, actionability, and quality of their clinical trials data, but have had limited uptake among the academic institutions.
Objective: To create a scalable methodology to convert CDISC CDASH electronic case report forms (eCRFs) instrument metadata into REDCap data dictionaries for the purpose of simplifying adoption and use of CDASH instruments by research teams across the REDCap Consortium.
Implementation: We have used our replicable methods to translate metadata from 34 CDASH Foundational eCRFs and 20 CDASH Crohn's Disease eCRFs into REDCap eCRF metadata and have made these instruments available in the REDCap Shared Data Instrument Library for widespread sharing and uptake across the REDCap Consortium. Users can import the standardized eCRFs directly into their REDCap projects for immediate use in clinical trial data collection.
Conclusion: Disseminating CDISC standards through the REDCap community will increase the accessibility of these standards for academic medical centers. Having academic clinical researchers using CDISC standards may lead to more research datasets that interoperate with pharmaceutical sponsored trials, and more discoveries from secondary use of clinical research data.

From REDCap founder Paul Harris presentation in March 2024

NIH NIDDK 75 YEARS 1950–2025

Leverage user community: *Ease of* CRF development (1)

EVEN *IF* You don't need FDA-ready CDASH e-CRFs via REDCap, re-*use* prior instruments!

Return to REDCap

Logged in as Jihad Obeid (Medical University of South Carolina)

Keyword search: depression

Search the library

Search options:

Minimum downloads: 0

Recent additions: show all

Curated by REDLOC? show all

Shared Library
Search
Library Metrics
My Activity
Institution Activity
Consortium Activity
REDLOC
Approval List
Approval History
Change Password
Logout

Found 10 results matching your search

Title	Downloads
Center for Epidemiologic Studies Depression Scale (CES-D), NIMH ★	90
Cognitive and Affective Mindfulness Scale-Revised (CAMS-R) ★	17
Geriatric Depression Scale GDS Long Form ★	15
Geriatric Depression Scale GDS Short Form ★	27
Hamilton Depression Rating Scale (HAM-D) ★	44
ICD10 MDI Major Depression Inventory ★	20
Patient Health Questionnaire 9 ★	62

Details:

View as web page
View as PDF
Import into my REDCap project

Viewing instrument as web page

Patient Health Questionnaire 9

Project Setup | Online Designer | Upload Data Dictionary

Data Collection Instruments

Whiteley 7 Scale | 8
Patient Health Questionnaire 9 | 14

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Procurement of shared data instruments for Research Electronic Data Capture (REDCap)

Jihad S. Obeid [a,*], Catherine A. McGraw [b], Brenda L. Minor [c], José G. Conde [d], Robert Pawluk [e], Michael Lin [e], Janey Wang [c], Sean R. Banks [f], Sheree A. Hemphill [g], Rob Taylor [c], Paul A. Harris [c]

[a] South Carolina Translational Research Institute, Biomedical Informatics Program, Medical University of South Carolina, 55 Bee St., MSC 200, Charleston, SC 29425, United States
[b] Center for Clinical & Translational Science and Training, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States
[c] Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University, Nashville, TN, United States
[d] School of Medicine, University of Puerto Rico Medical Sciences Campus, San Juan, PR, United States
[e] Center for Translational Science Activities, Mayo Clinic, Rochester, MN, United States
[f] Population Research Center, University of Texas at Austin, Austin, TX, United States
[g] Clinical & Translational Science Collaborative, Case Western Reserve University, Cleveland, OH, United States

ARTICLE INFO

Article history:
Received 10 August 2012
Accepted 20 October 2012
Available online 10 November 2012

Keywords:
Validated instruments
Data collection
Data sharing
Informatics
Translational research
REDCap

ABSTRACT

REDCap (Research Electronic Data Capture) is a web-based software solution and tool set that allows biomedical researchers to create secure online forms for data capture, management and analysis with minimal effort and training. The Shared Data Instrument Library (SDIL) is a relatively new component of REDCap that allows sharing of commonly used data collection instruments for immediate study use by research teams. Objectives of the SDIL project include: (1) facilitating reuse of data dictionaries and reducing duplication of effort; (2) promoting the use of validated data collection instruments, data standards and best practices; and (3) promoting research collaboration and data sharing. Instruments submitted to the library are reviewed by a library oversight committee, with rotating membership from multiple institutions, which ensures quality, relevance and legality of shared instruments. The design allows researchers to download the instruments in a consumable electronic format in the REDCap environment. At the time of this writing, the SDIL contains over 128 data collection instruments. Over 2500 instances of instruments have been downloaded by researchers at multiple institutions. In this paper we describe the library platform, provide detail about experience gained during the first 25 months of sharing public domain instruments and provide evidence of impact for the SDIL across the REDCap consortium research community. We postulate that the shared library of instruments reduces the burden of adhering to sound data collection principles while promoting best practices.

From REDCap founder Paul Harris presentation in March 2024

NIDDK

75 YEARS 1950–2025

PROMIS® (Patient-Reported Outcomes Measurement Information System), part of HealthMeasures
TRANSFORMING HOW HEALTH IS MEASURED

Leverage user community: *Ease of* CRF development (2)

Keyword search: promis

Search the library

Search options:

Language: - All -
Type: show all
Minimum downloads: 0
Recent additions: show all

Shared Library
Search
Library Metrics
Consortium Activity

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 >>

Found 1869 results matching your search — Didn't find what you were looking for? Suggest a validated instrument for library inclusion

Title	Downloads
PROMIS-29+2 Profile/Battery v2.1 (PROPr) [Battery]	204
PROMIS - 29 Profile/Battery v2.0 [Battery]	499
PROMIS - 43 Profile v2.0 (Spanish)	15
PROMIS - 43 Profile/Battery v2.0 [Battery]	36
PROMIS - 57 Profile v2.0 (Spanish)	19
PROMIS - 57 Profile/Battery v2.0 [Battery]	62
PROMIS Item Bank v1.0 - Trastornos del Sueño - Cuestionario Abreviado 6a	6
PROMIS Item Bank v1.0 - Trastornos del Sueño - Cuestionario Abreviado 8a	4
PROMIS Item Bank v1.1 - Trastorno Emocional - Enojo	6
PROMIS Item Bank v1.1- Efectos del Dolor	16
PROMIS Item Bank v1.2 - Capacidad de Funcionamiento Físico	15
PROMIS Item Bank v2.0 - Aislamiento social - Cuestionario abreviado 4a	18

PROMIS, Patient-Reported Outcomes Measurement Information System, & the PROMIS logo are marks owned by the U. S. Department of Health & Human Services.

Differences between PROMIS Measures
Computer Adaptive Tests (CATs) versus Short Forms
•Many domains offer a **computer adaptive test (CAT)** and one or more short forms. Select that type of measure that fits your needs and resources. **Learn more>>**

- CATs
  - Tailored selection of items for each respondent
  - Requires administration technology
  - High measurement precision across a wide range of symptom/function severity
- Short forms
  - All respondents answer all questions
  - No special administration technology needed
  - Degree of measurement precision varies

Example *Reproducible-by-design* CRF development via REDCap, re-*use* participant-reported outcomes instruments via PROMIS *OR* others (NIH Phenotype eXplorer Toolkit, if not NIH Toolbox®)

PhenX Toolkit

Keyword search: phenx

Search the library

Search options:

Language: - All -
Type: show all
Minimum downloads: 0
Recent additions: show all

Shared Library
Search
Library Metrics
Consortium Activity

Found 1 results matching your search — Didn't find what you were looking for? Suggest a validated instrument for library inc

Title — Downl

PhenX Toolkit [External Instrument Library]

Details:

Institution: REDLOC

Description: The PhenX Toolkit provides standard measures related to complex diseases, phenotypic traits and environmental exposures. Use of PhenX measures facilitates combining data from a variety of studies, and makes it easy for investigators to expand a study design beyond the primary research focus. REDCap Instrument Zip files are provided by the RTI PhenX team to enable use of PhenX measures in REDCap. Use of the PhenX Toolkit with or without other resources is solely the responsibility of the User/Investigator. Please review PhenX Toolkit guidance for additional information.
https://www.phenxtoolkit.org

NOTE: This is not an instrument but a third-party website that is *NOT* affiliated with REDCap in any way. The organization behind the website has made assurances that the instruments distributed on their website do not violate any copyright laws. However, neither REDCap nor Vanderbilt University nor the institution hosting your REDCap project shall be held responsible for the veracity of the foregoing statements, the conduct of PhenX Toolkit or for the content on their website (https://www.phenxtoolkit.org) in any manner whatsoever.

Go to this library

PhenX Toolkit

REDCap Instruments ZIP Files

PhenX is collaborating with REDCap to make PhenX protocols available as REDCap instruments zip files that can be uploaded directly to REDCap. More, coming soon. Click on a protocol name below to download the REDCap Zip File. Click here for REDCap "Instrument ZIP" feature.

Showing 1 to 50 of 861 protocols

Abdominal Aortic Aneurysm
Access to Health Services
Access to Health Technology
Access to Lethal Means
Activities of Daily Living (ADLs)
Acute Subjective Response to Substances - Current - General
Acute Subjective Response to Substances - Current - Specific - Alcohol
Acute Subjective Response to Substances - Current - Specific - Drugs
Acute Subjective Response to Substances - Current - Specific - Tobacco
Acute Subjective Responses to Substances - Retrospective - Alcohol
Acute Subjective Responses to Substances - Retrospective - Tobacco
Addiction Severity Index
Adequacy of Prenatal Care
Adherence to Medication Regimens

From REDCap founder Paul Harris presentation in March 2024

NIH NIDDK

75 YEARS 1950–2025

Leverage user community: *Ease of* CRF development (3)

NATIONAL CANCER INSTITUTE
Center for Biomedical Informatics & Information Technology

NIH NATIONAL CANCER INSTITUTE

Manage | View | Collaborate | Data Interchange | Personalize

>>Favorites>>NCI Standard CRFs

isplay Values | View | Delivery Options | Add to Guest User Cart

mplate Rows 1..130 of 130 1 Request

iew	1 Form Name	Public ID	Version	Owned By
	Adverse Event/Serious Adverse Event CTCAE v3 NCI Standard Template	2748814	3.00	NCI Standards
	Adverse Event/Serious Adverse Event CTCAE v4 NCI Standard Template	3265657	3.00	NCI Standards
	Adverse Event/Serious Adverse Event CTCAE v4.0 CDISC Aligned NCI Standard Template	6988567	1.00	NCI Standards
	Adverse Event/Serious Adverse Event CTCAE v5.0 CDISC Aligned NCI Standard Template	6984728	1.00	NCI Standards
	Adverse Event/Serious Adverse Event CTCAE v5.0 NCI Standard Template	6081561	1.00	NCI Standards
	Brief Pain Inventory-BPI CDISC Aligned NCI Standard Template	7093187	2.00	NCI Standards
	Concomitant Medication CDISC Aligned NCI Standard Template	6984232	1.00	NCI Standards
	Concomitant Medication NCI Standard Template	2867231	2.00	NCI Standards
	Consent CDISC Aligned NCI Standard Template	6936999	1.00	NCI Standards

Extended examples of *Reproducible-by-design* CRF development via REDCap, re-*use of longstanding Common Data Elements*, at least for *NCI's Adverse Events data elements*

[for more on NCI CDEs, see its (NIH) Enterprise Vocabulary Services or EVS, esp. its Cancer Data Standards Registry or caDSR sites]

cadsr-dev.cancer.gov/onedata/Home.jsp

NATIONAL CANCER INSTITUTE Center for Biomedical Informatics & Information Technology

Denise Warzel logged in to NCI caDSR Dev (Project: CODEMAP) (Work Area)

Define | Manage | View | Collaborate | Administer | Data Interchange | Personalize | Favorites

From REDCap founder Paul Harris presentation in March 2024

NIH NIDDK

75 YEARS 1950–2025

From Research Study Design to *Usable* Data: Take Home on "Using REDCap for FAIR data"

BONUS pre-vetted starting points* within *Intramural Research Program REDCap

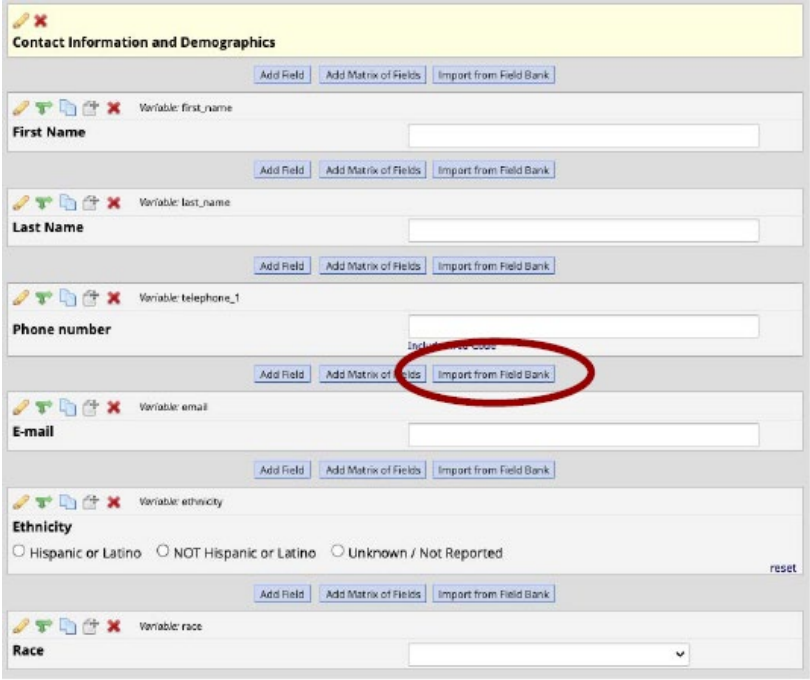
Inherent capacity to semantically search up COMMON DATA ELEMENTS *within* REDCap

- *Some that—since 2020 NIH-wide criteria were set—are designated as NIH-ENDORSED!*

[*if you don't' use REDCap*]

[*try CDEMapper toolkit*]

- Example of demographics:
  - NAME
  - TEL#
  - EMAIL
  - ETHNICITY*
  - RACE*



National Institutes of Health
Endorsed CDEs Accelerate Research

REDCap
Research Electronic Data Capture

*In the case of 'Demographics' be aware of Federal Government changes in capturing Race/Ethnicity as specified in Statistical Policy Directive #15's revision: https://spd15revision.gov/

In the event of NOT limited time, leading here, skip back to take-homes

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to *Usable* Data: Take Home on "Using REDCap for FAIR data"

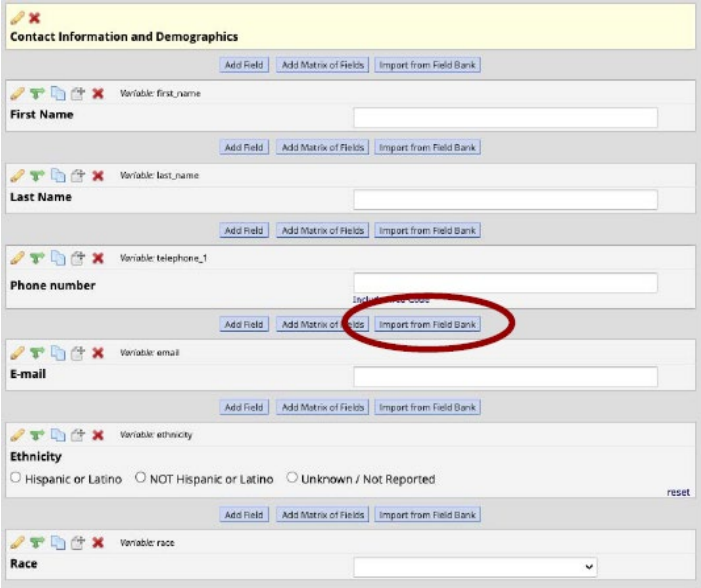
You have pre-vetted starting point because of *Intramural Research Program REDCap*

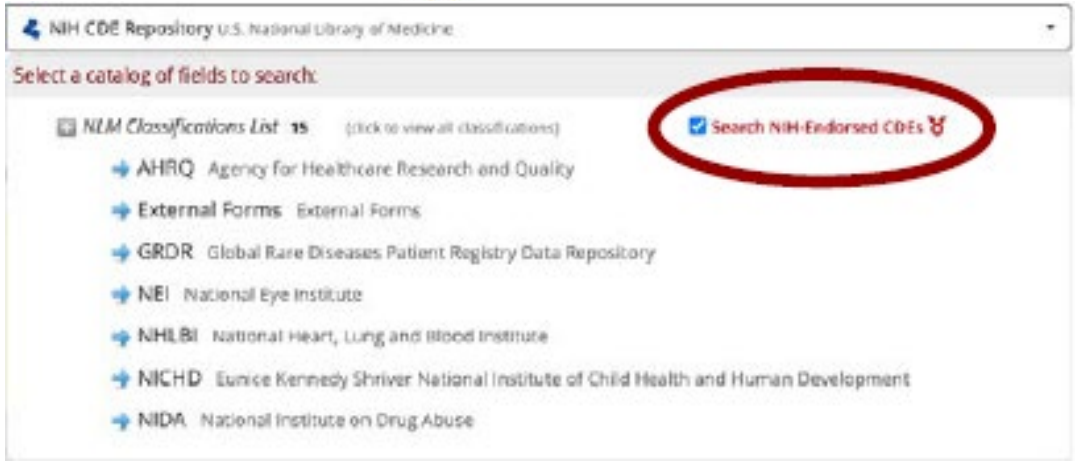
Inherent capacity to semantically search up COMMON DATA ELEMENTS *within* REDCap

- *Some that—since 2020 NIH-wide criteria were set—are NIH-ENDORSED!*

National Institutes of Health
Endorsed CDEs Accelerate Research

- Example of demographics:





REDCap
Research Electronic Data Capture

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to *Usable* Data: Take Home on "Using REDCap for FAIR data"

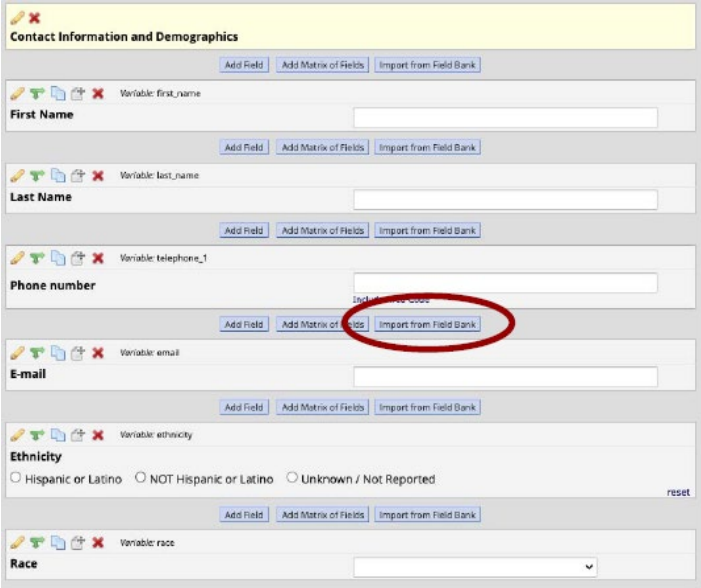
You have pre-vetted starting point because of *Intramural Research Program REDCap*

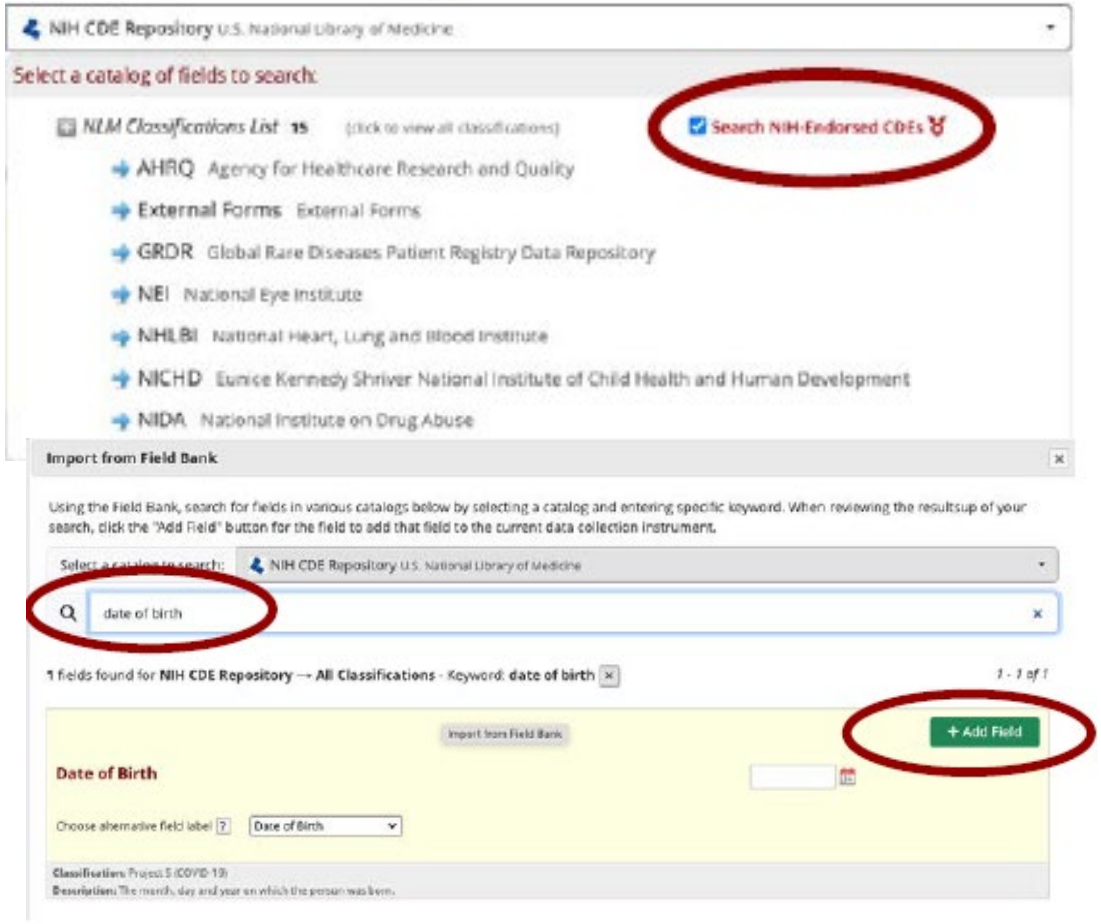
Inherent capacity to semantically search up COMMON DATA ELEMENTS *within* REDCap

- *Some that—since 2020 NIH-wide criteria were set—are NIH-ENDORSED!*

National Institutes of Health
Endorsed CDEs Accelerate Research

- Example of demographics:





REDCap
Research Electronic Data Capture

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to *Usable* Data: Take Home on "Using REDCap for FAIR data"

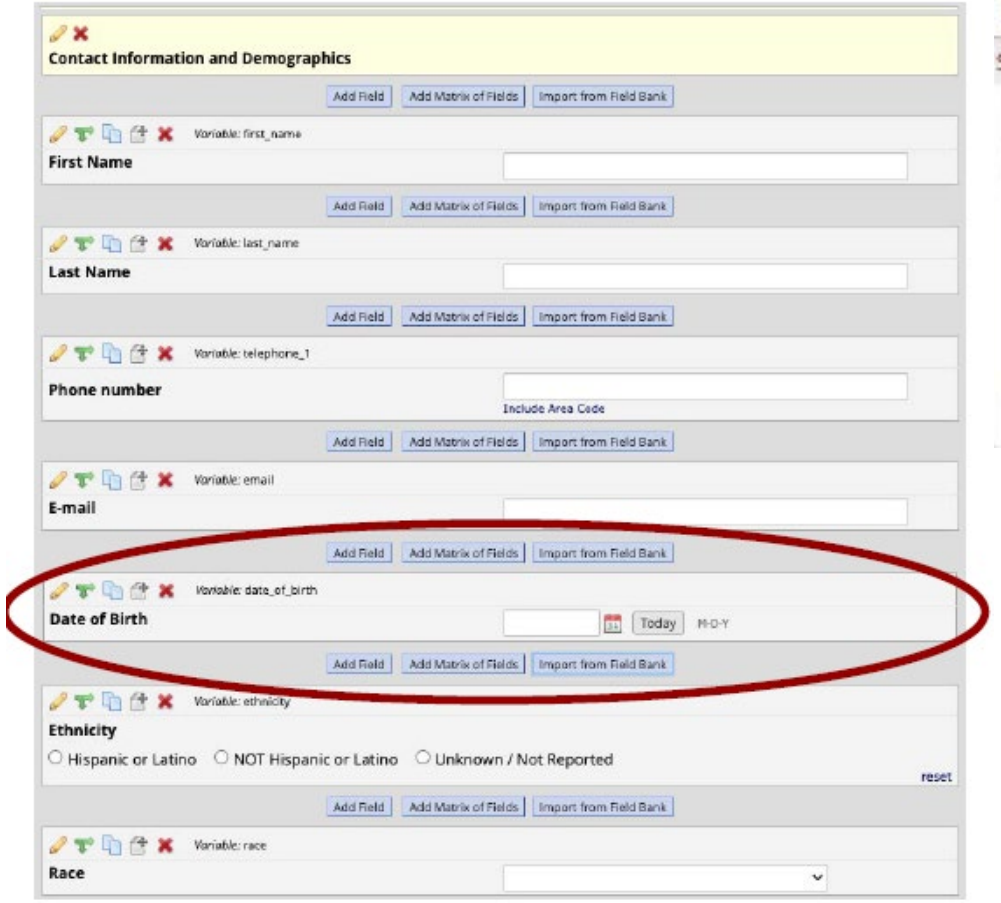
You have pre-vetted starting point because of *Intramural Research Program REDCap*

Inherent capacity to semantically search up COMMON DATA ELEMENTS *within* REDCap

- *Some that—since 2020 NIH-wide criteria were set—are NIH-ENDORSED!*

National Institutes of Health
Endorsed CDEs Accelerate Research

- Demographics:



REDCap
Research Electronic Data Capture

National Institute of Diabetes and Digestive and Kidney Diseases

From Research Study Design to *Usable* Data: Take Home on "Using REDCap for FAIR data"

You have pre-vetted starting point because of *Intramural Research Program REDCap*

Inherent capacity to semantically search up COMMON DATA ELEMENTS *within* REDCap

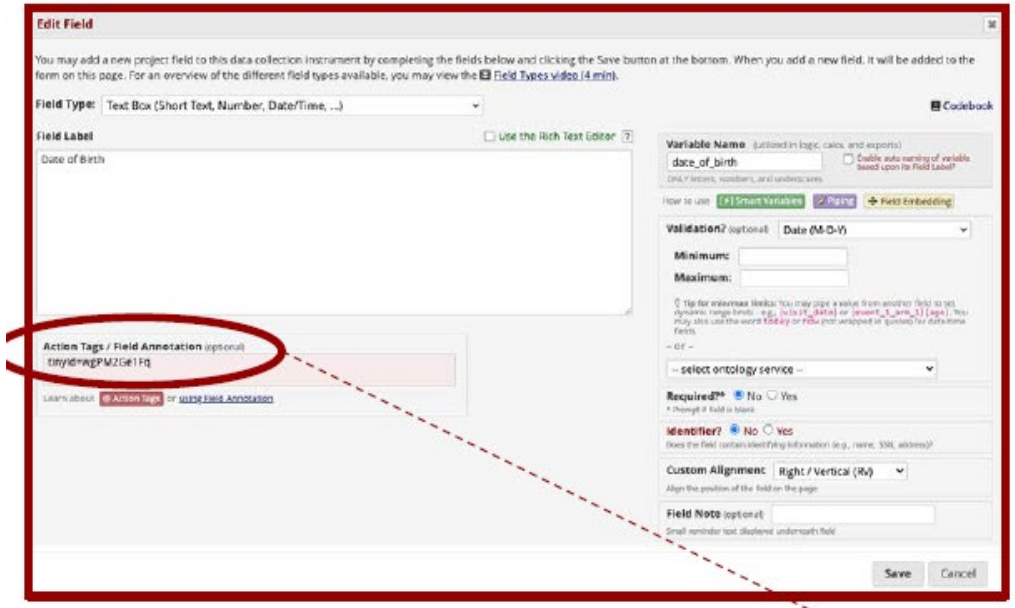
- *Some that—since 2020 NIH-wide criteria were set—are NIH-ENDORSED!*

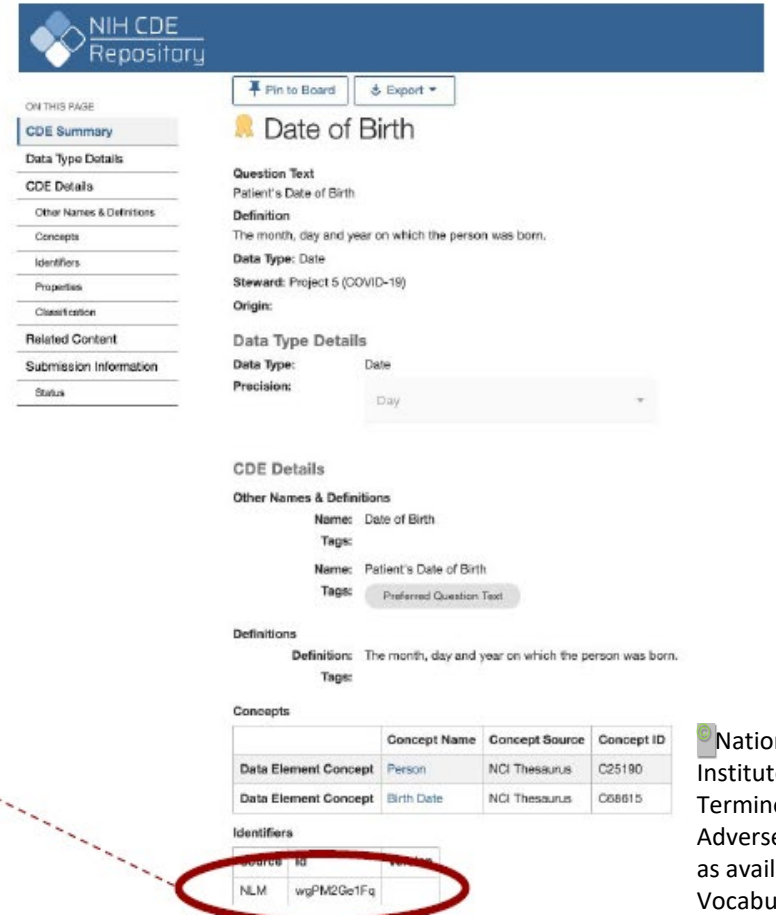
National Institutes of Health
Endorsed CDEs Accelerate Research

- Demographics:
- *Adverse Events*

CTCAE©

REDCap
Research Electronic Data Capture





© National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) as available in Enterprise Vocabulary Services (EVS)

In the event of NOT limited time, leading here, skip back to take-homes

*In the case of 'Demographics' be aware of Federal Government changes in capturing Race/Ethnicity as specified in Statistical Policy Directive #15's revision: https://spd15revision.gov/

APPENDIX FOR FURTHER LEARNING RESOURCES

Principles of good data collection IF having to resort to non-Electronic Data Capture systems like REDCap or Medidata Rave:

1. *Consistency* - processing, naming/identifiers, formats, layout and actual storage of data
2. *Proper naming* – do not use spaces (_ instead), no symbols, no capitalization, keep it less than 32 characters, shorter is better
3. *Date formatting* – Excel stores this differently in Macs vs Windows, European vs US conventions
4. *Missing info* – better to use a standard place holder for missing data, so "missing" and not "incomplete"
5. *Too much info* – No more than one thing in a cell, do not mix numeric and character
6. *Organization* – make it a "rectangle", one row or more for subjects, one column for each variable
7. *Retain metadata*–info about the data and create a "README" file
8. *No calculations* – primary/original data should contain <u>just the collected data</u>
9. *No colors or highlighting* – cannot analyze this
10. *Version control & backups* – copy your original data and write-protect it for archiving
11. *Data Validation* – build in acceptable ranges for data collection and integrate formatting requirements into data collection
12. *Use non-proprietary file formats* – your data is important and should live forever; ".csv" and "tab delimited" file formats do not require special software

To learn more about data collection, visit https://bit.ly/DataOrgSheets or use the QR code

NIH NIDDK 75 YEARS 1950–2025