



RESEARCH DATA **ALLIANCE**

Adoption of RDA outputs to support Reproducible Health Research

RDA Adoption Webinar Series
March 22nd, 2019

Learning objectives

1. How RDA is supporting adoption of outputs and recommendations
2. How RDA recommendations and outputs build both social and technical bridges to enable open sharing and re-use of data as exemplified by work across health research
3. How to engage with RDA to develop, adopt, or learn more about recommendations or outputs

Contributors

Anthony Juehne RDA-US, Sr. Implementation & Evaluation Analyst

Leslie McIntosh RDA-US, Executive Director

Oya Beyan Aachen University, Germany

John Graybeal Stanford University, U.S.

Mark Musen Stanford University, U.S.

Health Data IG

<https://www.rd-alliance.org/groups/health-data.html>

Reproducible Health

Data Services WG

<https://www.rd-alliance.org/groups/reproducible-health-data-services-wg>

RDA Adoption Support Task Force



Marieke Willems

Outreach & Impact
Coordinator

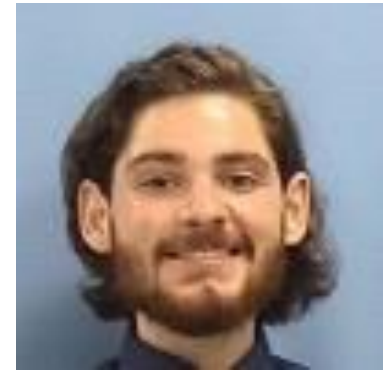
Trust-IT Services Ltd.



Daniel Bangert

Outreach & Impact
Support

Göttingen State and
University Library



Anthony Juehne

Adoption Manager

Rensselaer
Polytechnic
Institute

Agenda

- What is RDA?
- How do we support adoption and impact?
- Introduction of social and technical issues

- Health Data Interest Group & Reproducible Health Data Services WG
_____ **Question & Answer** _____
- Stanford University and the Center for Expanded Data Annotation and Retrieval (CEDAR)
- Washington University's EHR implementation of Dynamic Data Citation

- How to get involved with RDA
- How to learn more about adopting RDA outputs and recommendations

_____ **Question & Answer** _____

What is RDA?

RESEARCH DATA ALLIANCE



Vision: Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

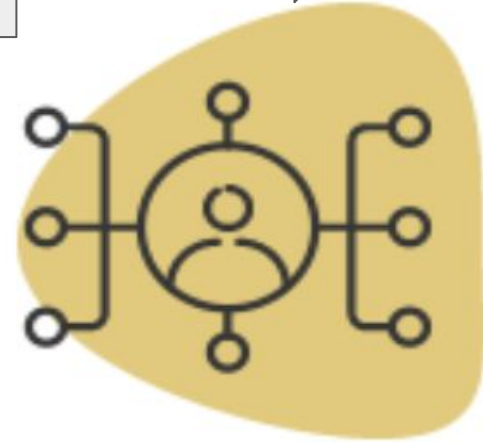
THE RDA ADOPTION STORIES

Adopters of RDA outputs share their experience and lessons learned to inspire the further uptake of RDA outputs

**FIND THE ADOPTION STORIES HERE:
[RD-ALLIANCE.ORG/RECOMMENDATIONS-
OUTPUTS/ADOPTION-STORIES](https://rd-alliance.org/recommendations-outputs/adoption-stories)**



Bridging Social and Technical



- Preservation and stewardship
- Improving usefulness of data
- Data sharing culture
- Education and training
- Data security, legal interoperability, and compliance

- Data discovery
- Provenance, reuse, and interconnection
- Issues of scale
- Governing and maintaining infrastructure
- Data access and interoperability

What are the issues surrounding data collection and data sharing for reproducible health research?

Assumptions

1. Process of sharing biomedical data not comprehensively documented
2. Effective (and ethical?) health interventions require verifiable and reproducible evidence
3. The responsibility of ensuring reproducible research involves multiple stakeholders across the research workflow

What is reproducibility?

Computational Reproducibility:

If we took your data and code/analysis scripts and re-ran it, we can reproduce the numbers/graphs in your paper

Empirical Reproducibility:

We use your exact methods and analysis, but collect new data, and we get the same results

Replicability (Results Reproducibility):

We have enough information to re-run the experiment the way it was originally conducted

Reproducible Scientific Workflows

“Reproducibility implies repetition and thus a requirement to also move back – to retrace one’s steps, question or change assumptions, and move forward again.”

Millman, K. J., & Perez, F. (2014). Developing Open-Source Scientific Practice (V. Stodden, F. Leisch, & R. D. Peng, Eds.). In Implementing Reproducible Research (CRC the R series, pp. 149-183). Boca Raton, FL: Taylor & Francis Group, LLC.

The Grand Why

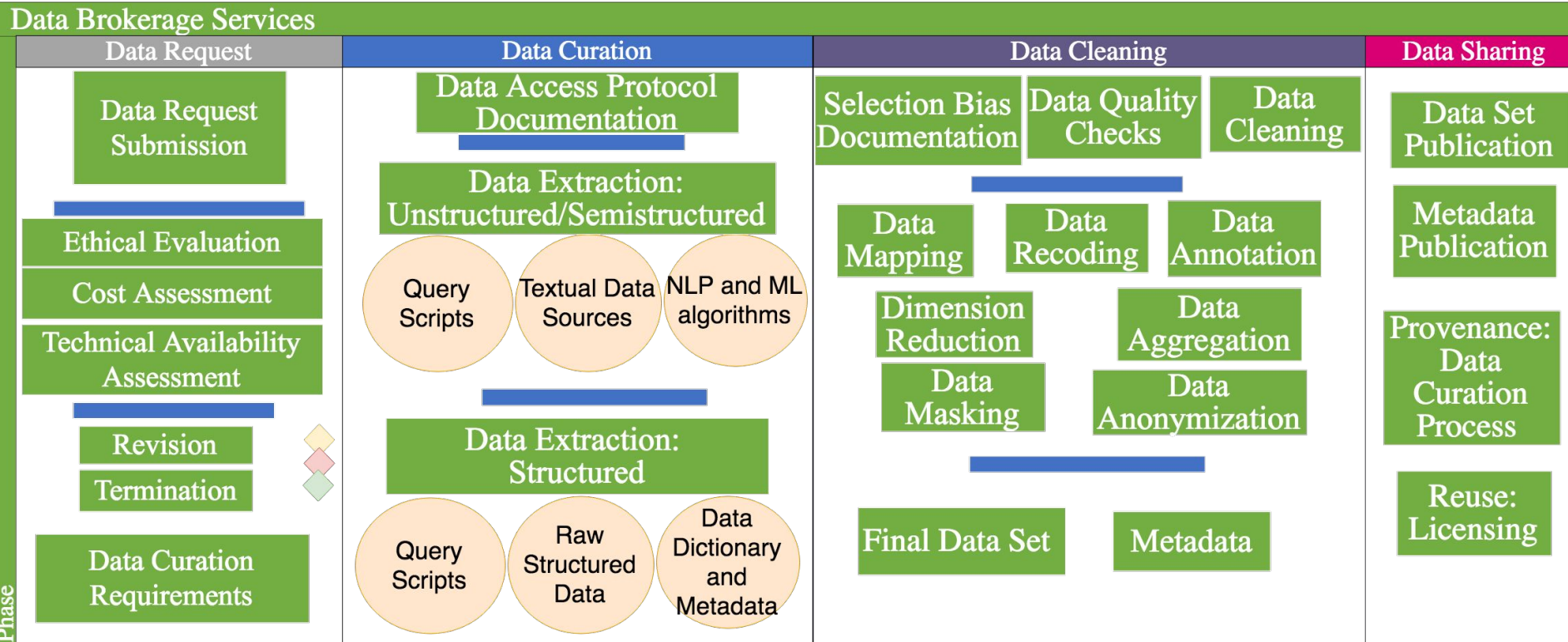
“The construction of a
scientific heritage
where anyone can validate the work of
others and build upon it.”¹

Millman, K. J., & Perez, F. (2014). Developing Open-Source Scientific Practice (V. Stodden, F. Leisch, & R. D. Peng, Eds.). In *Implementing Reproducible Research* (CRC the R series, pp. 149-183). Boca Raton, FL: Taylor & Francis Group, LLC.

The Concern

How do we ensure all stages of data collection and preparation within the scientific compendium are fully transparent and appropriately accessible to achieve reproducibility?

Reproducible Health Data Services (Social)

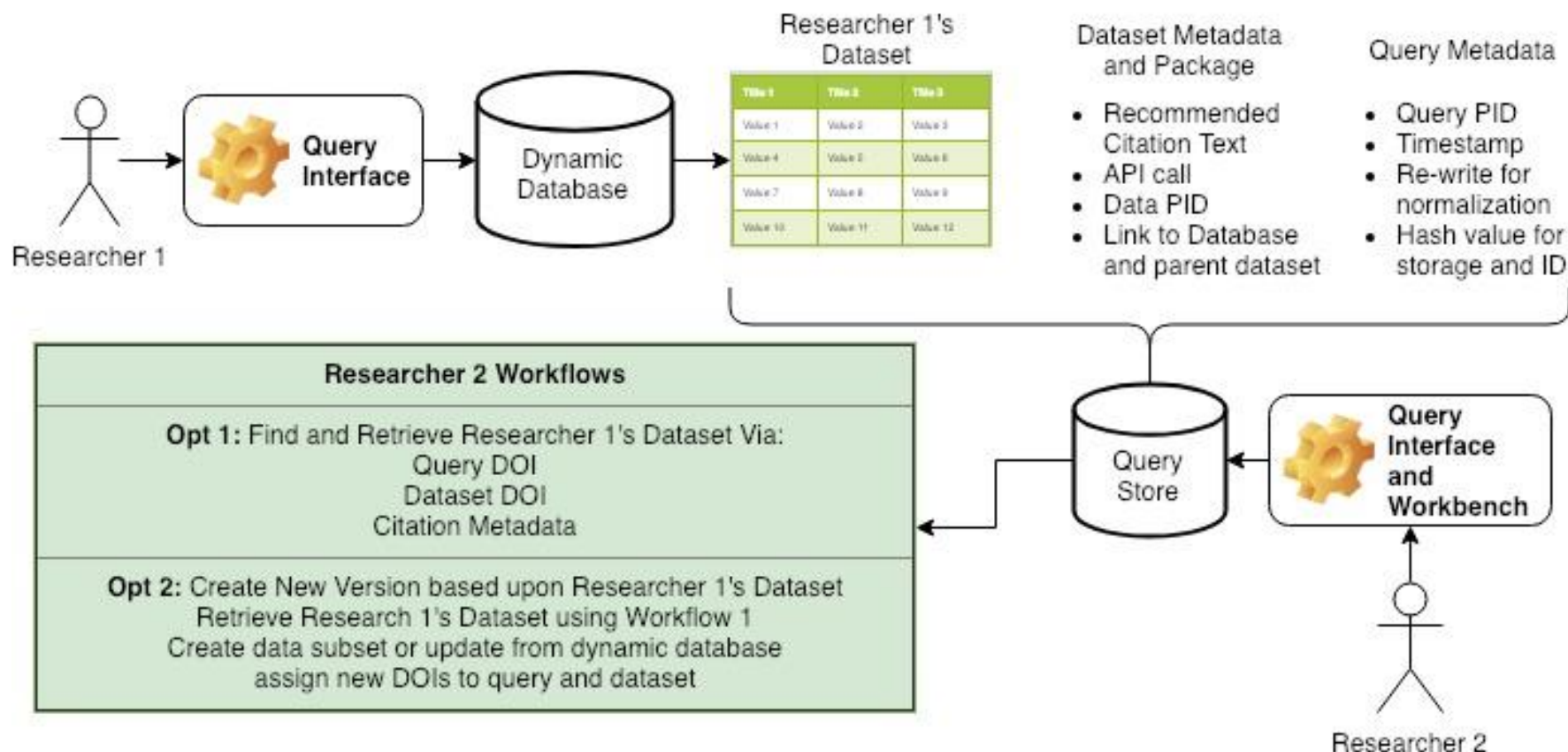


Phase

The Concern

Big Data evolves rapidly in volume, variety, and velocity. How do we ensure traceable provenance of a dataset from an evolving data source?

Scalable Dynamic Data Citation (Technical)



RDA Data Citation WG: <https://rd-alliance.org/groups/data-citation-wg.html>

RDA WG Case Statement: <https://www.rd-alliance.org/filedepot?fid=102>

Executive Summary of Output:

<https://docs.google.com/document/d/1SUr28B30Gg4yNNHfmztRawP7zLJqmx2lpEZSceeBMs/edit?usp=sharing>

Health Data IG

&

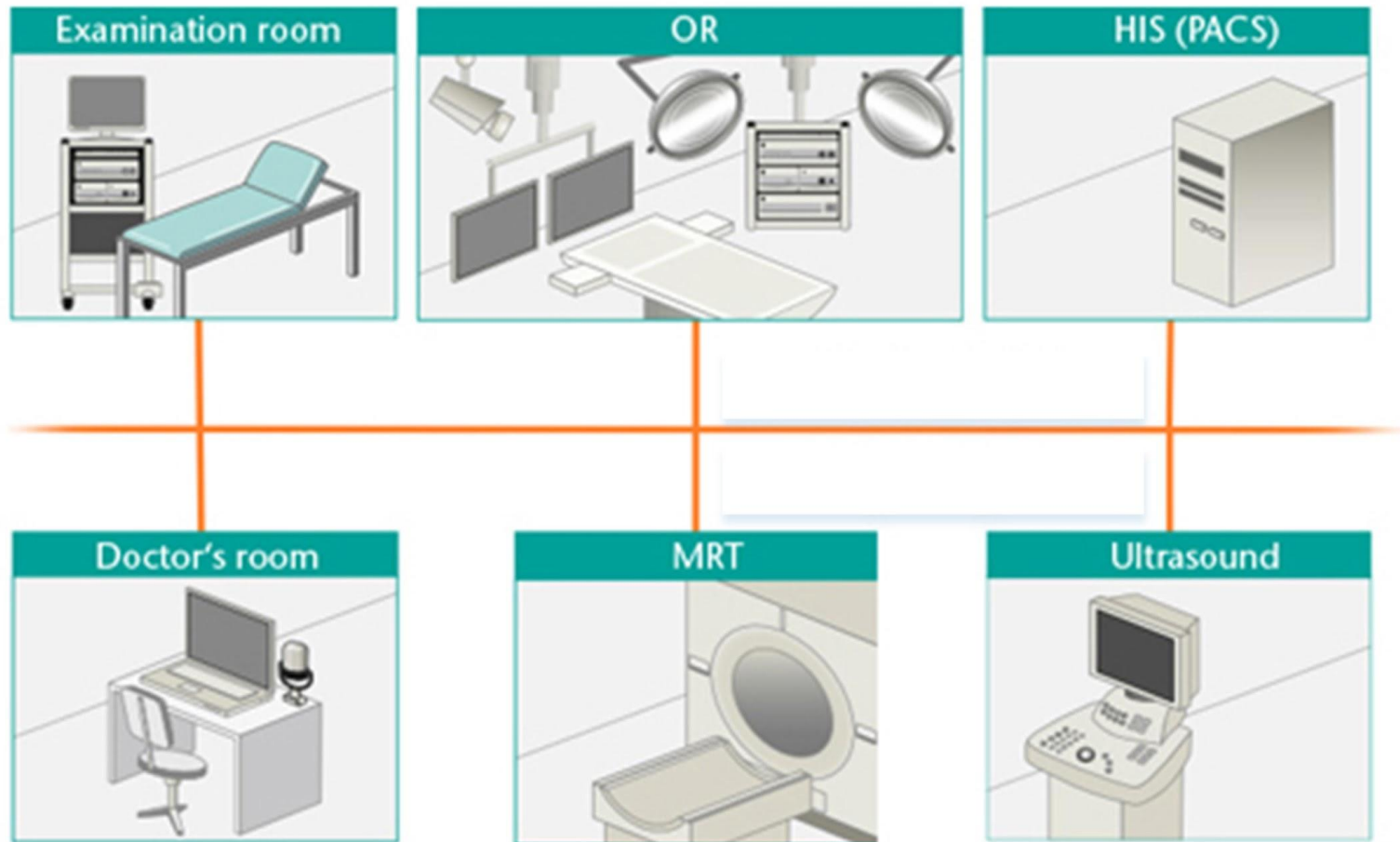
Reproducible
Health Data Services WG

Learning from bedside data..

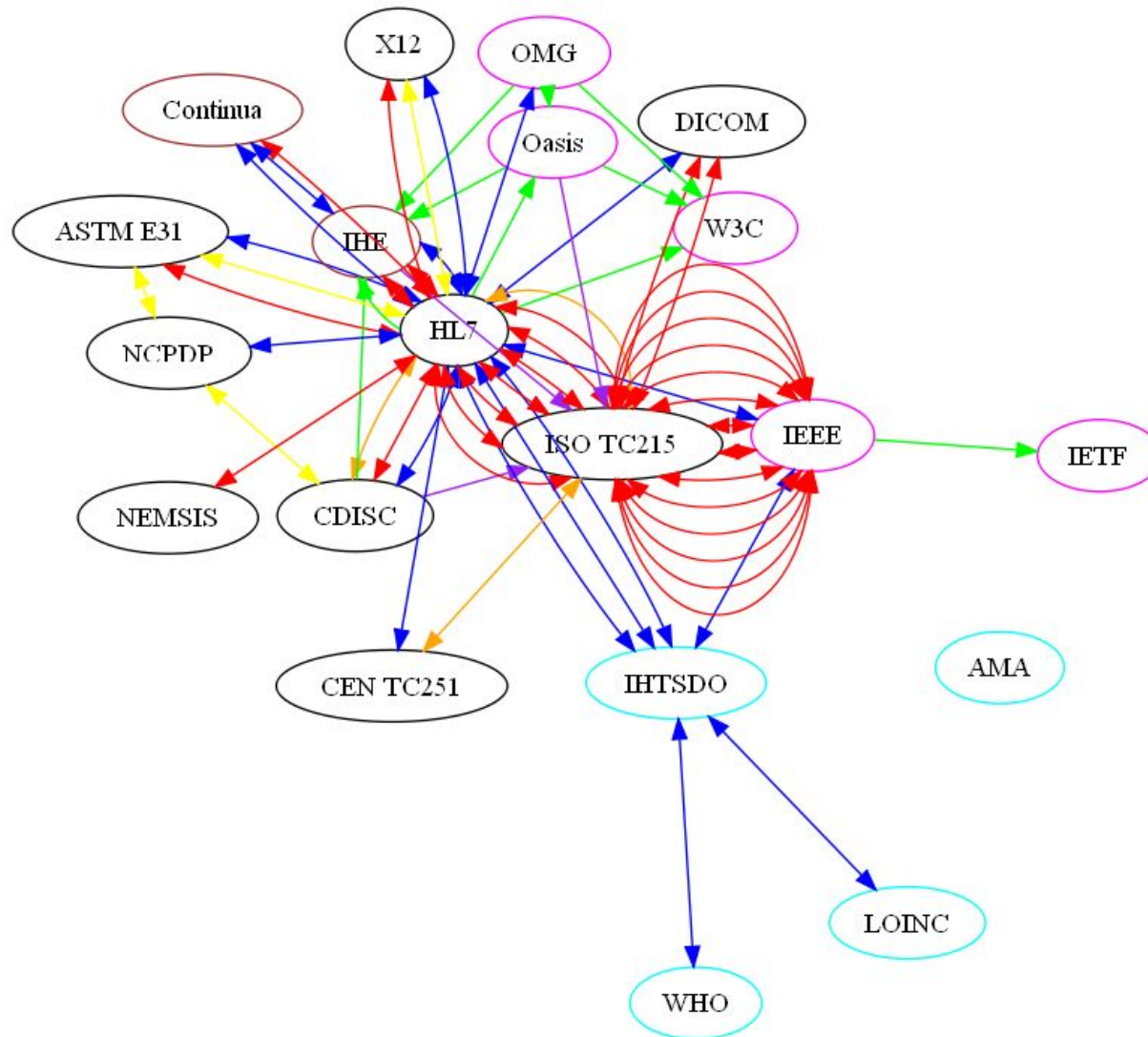
how to create FAIR research data from routine care data ?



Heterogeneous system with multiple types of data

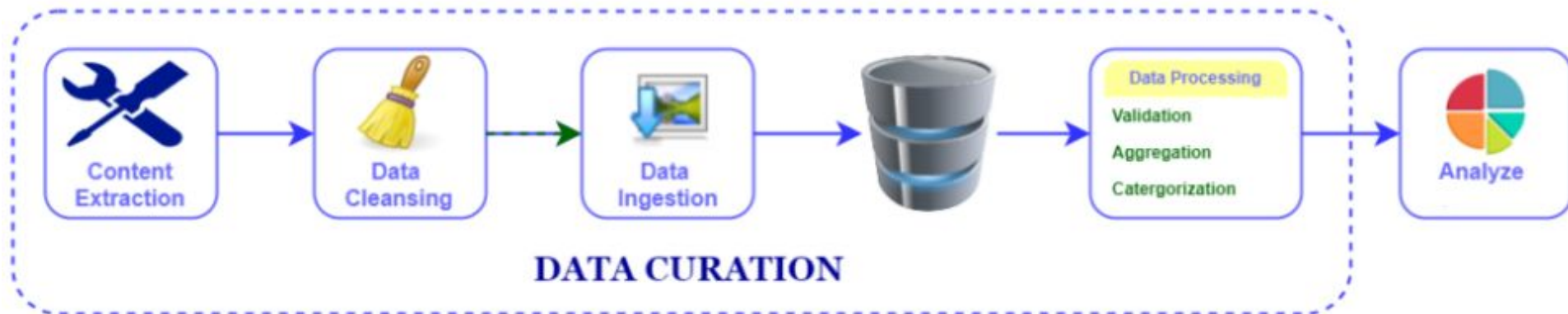


Heterogeneous standards - if any at all



Challenges of Health Data Services

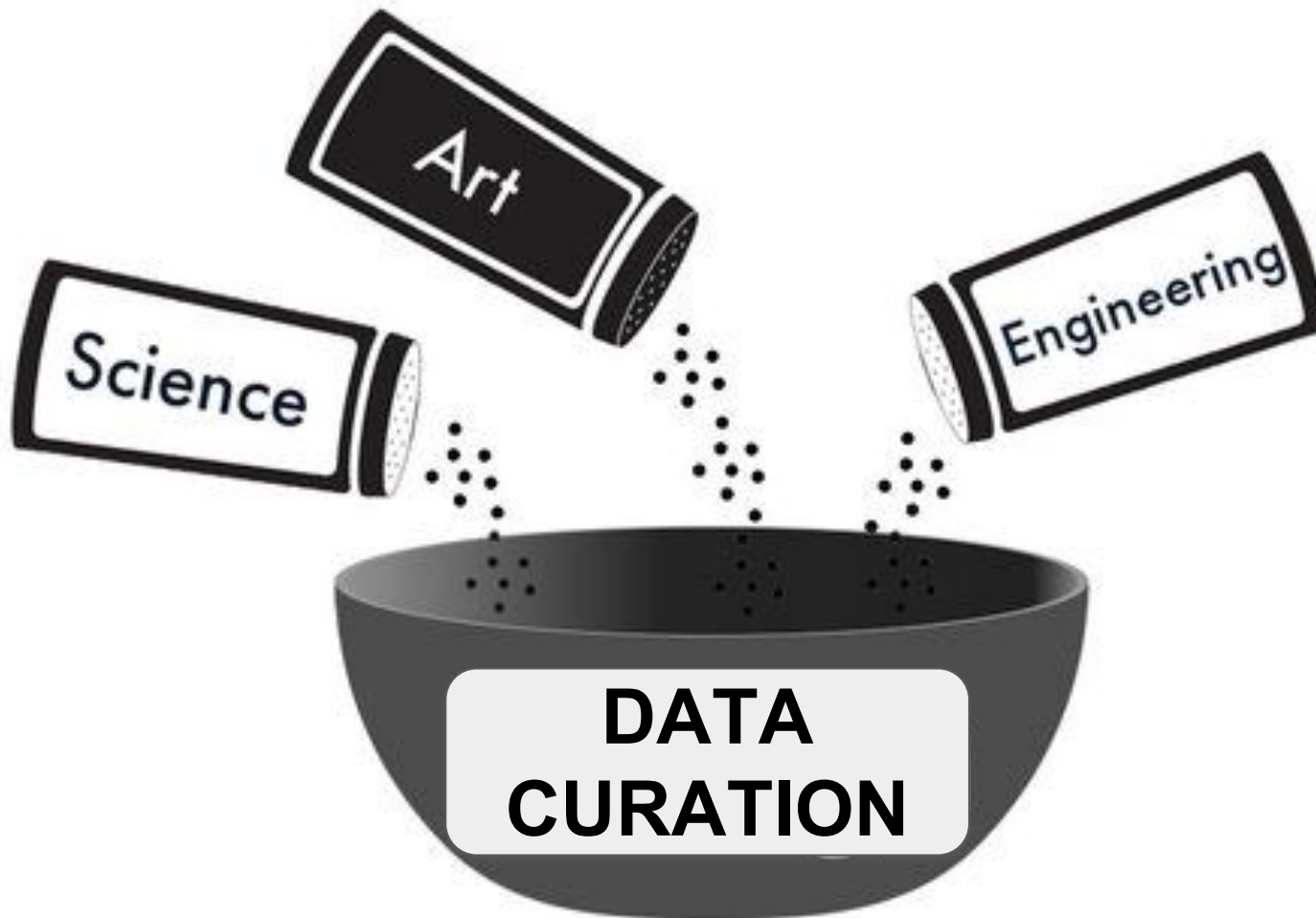
- Data curation processes
 - requires to interact with multiple systems
 - needs manual effort and crafting to map, transform, clean the data
 - typically queries and scripts are written case bases
 - ETL processes varies highly depending on the task and the curator



Challenges of Health Data Services

- Documenting metadata
 - very limited documentation if any
 - there is no recommendation / guideline on what to document
 - too time consuming
- How to capture the metadata
 - no guidance on what is useful (problem of granularity)
 - what are the available standards ?
- Sharing health data curation metadata
 - no standard way to find, access and interpret this metadata

Art of curating data



Reproducible Health Data Working Group

“to improve the reuse of health data by providing recommendations for reproducible data curation and brokerage workflow services”

FAIR + Repeatability

Fitness for Purpose

Trustability ...

Reproducible Health Data Working Group

Deliverable 1: Recommendation Statement

perform a gap analysis

- Relevance , maturity levels, ...
- Identify needs for further standards and methods
- Communicate the need with relevant groups

Reproducible Health Data Working Group

TASKS

- Identify data curation processes
 - which activities are carried out at the different phases of the data services
- Document challenges related to activities
- Explore the available standard stack
 - RDA, W3C, ISO, research communities

Reproducible Health Data Working Group

Deliverable 2:

Adoption and Training Guide

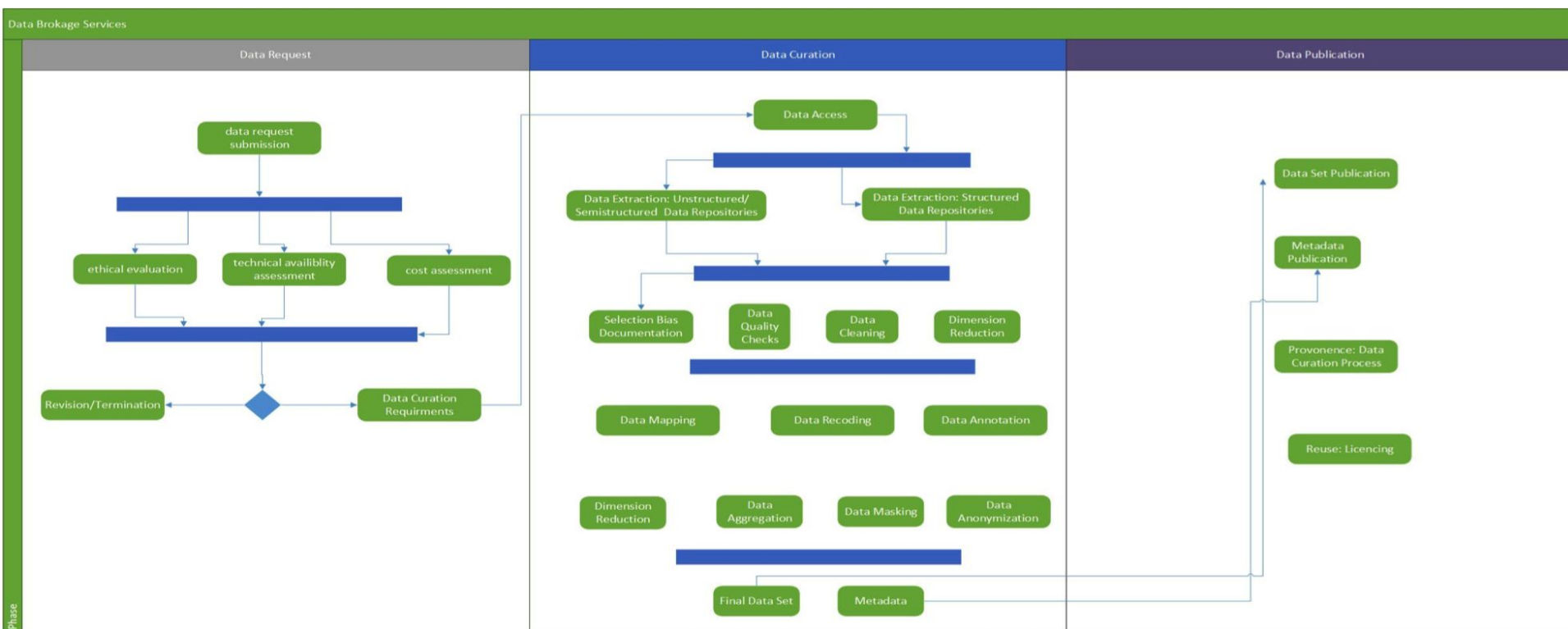
- Documentation of state-of-the-art methods and standards for clinical data curation
- Best practices for capturing and storing data curation metadata
- Identify use cases

Reproducible Workflows for Health Data Services

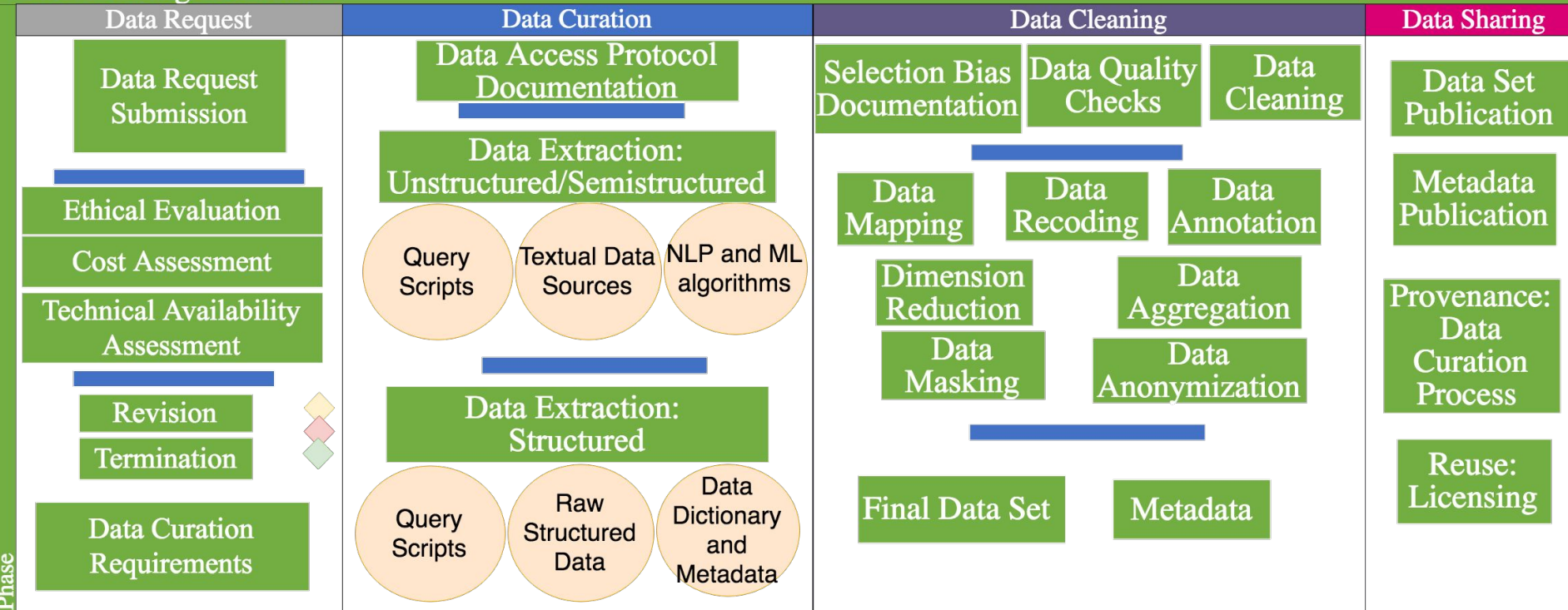
Data Request

Data Curation

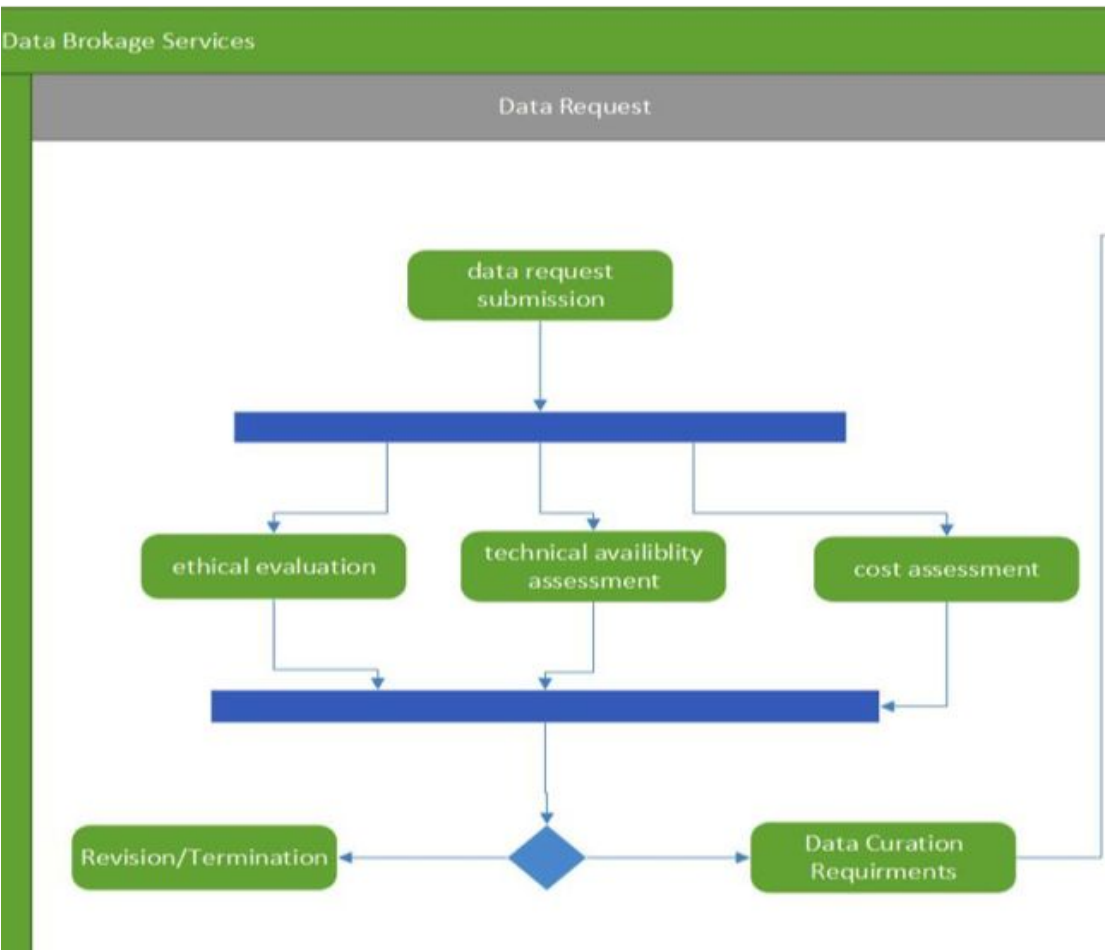
Data Sharing



Data Brokerage Services



Data Request Elements



Ethical Assessment

- IRB protocol approval
- Consent form
- Approval date
- Reviewing institution

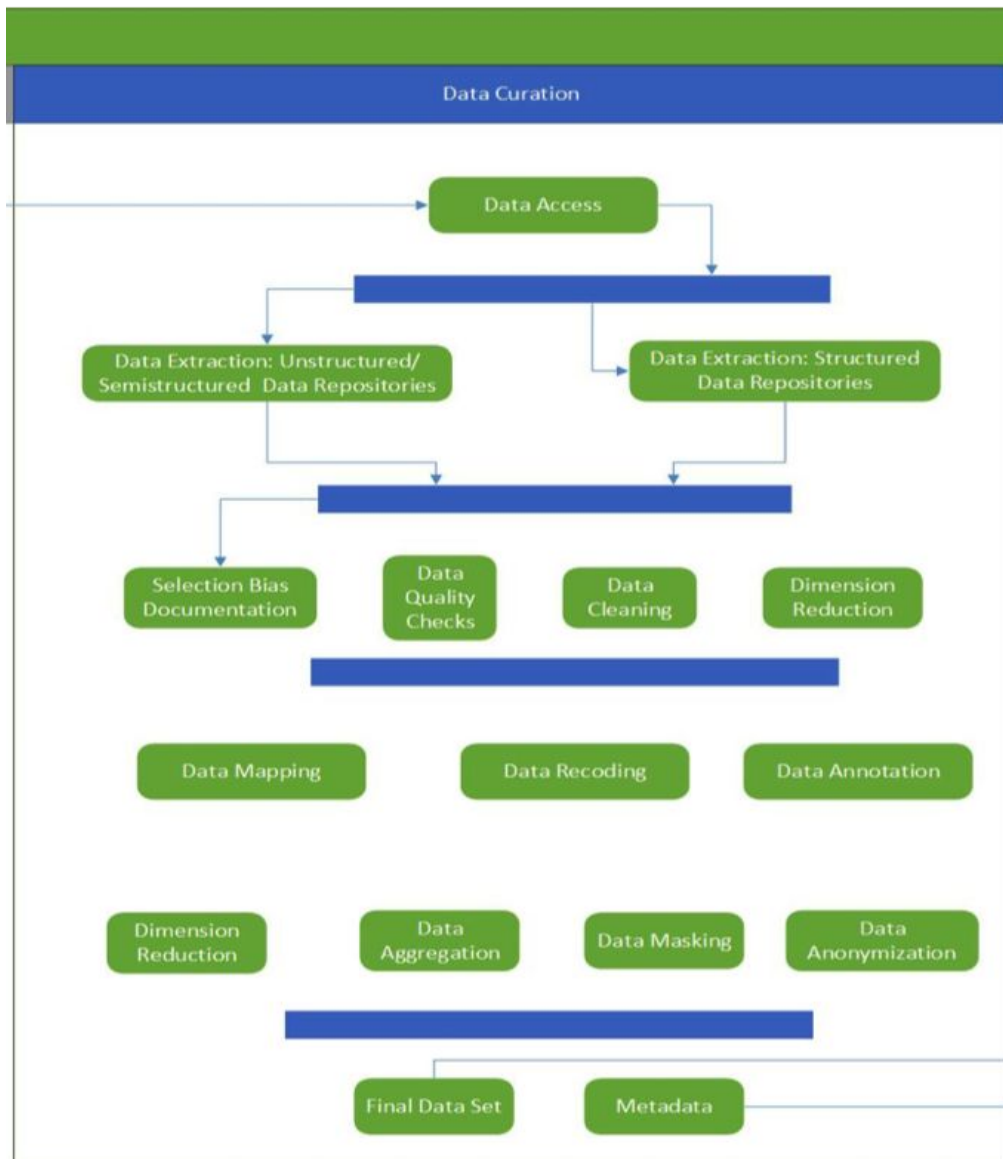
Cost Assessment

- Grant information
- Center expected hours of work

Feasibility Assessment

- Is the requested data available?
- Does it satisfy the clinical question?

Data Curation Processes



What is the source of data?

- DOI and Data citation?
- Multi-Site/Multi-registry Collection

Are query scripts FAIR?

- Interoperable
- Commented

How are participant inclusion and exclusion criteria operationalized?

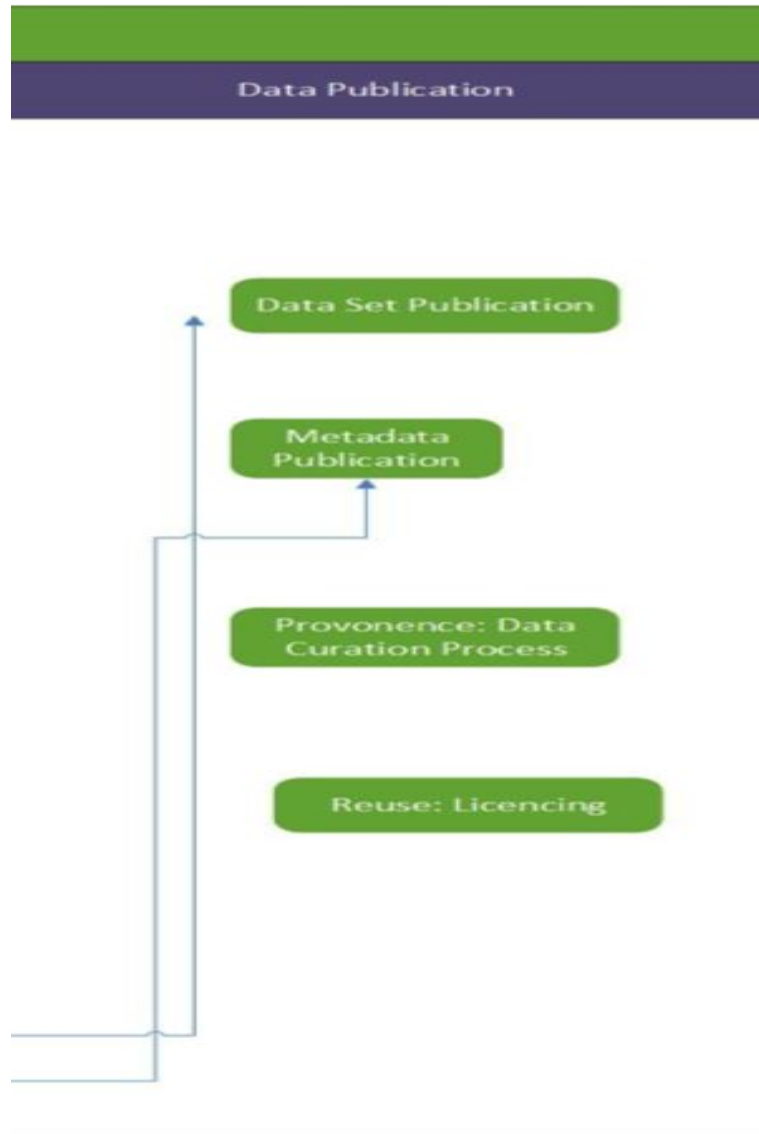
- Ontologies and standardized vocabs
- Limitations

How are data cleaned, mined, and merged?

- FAIR software code

Is there a FAIR final dataset or data dictionary?

Data Sharing and Publication



Can we apply FAIR principles to data products of the health data service workflows ?

- Rich metadata
- Persistent identifiers
- Licensing
- Common protocols to access

How to share data curation metadata together with the data?

Is it possible to have a minimum information reporting standard for data services?

How to make the data curation workflow metadata FAIR?

Draft Working Document

Current Health Data Service Center Workflow

https://docs.google.com/spreadsheets/d/1-uSocVpju4_fBcMDBgW2LxG3EpvBNRedOQuings2Cok/edit?usp=sharing

- Phases of Data Services
- Curation Activity
- Explanation
- Reproducibility Challenges
- Possible Metadata Formats
- Relevant Community Resources/Standards










Future Work

WG case statement:

<https://docs.google.com/document/d/1wrpxYnldvJHKN21J70esdFhVEabqizjaNXJePAp2SnM/edit?usp=sharing>

- Participate in the working group
- Vet the framework of elements
- Interested in adoption testing?

Stanford Adoption Case Study

	Title	Created
	Cost Assessment	1/9/18 2:15 AM
	Data Access	1/10/18 7:00 AM
	Data Extraction - Structured	1/10/18 6:13 PM
	Data Extraction_Unstructured Data	1/10/18 6:16 PM
	Ethical Assessment Review	1/9/18 1:50 AM
	Participant Exclusion Criteria	2/9/18 8:55 PM
	Participant Inclusion Criteria	1/11/18 2:32 AM
	Project Metadata	1/11/18 6:59 AM
	RDA_Health Data Service Center_Data Brokerage Workflow	1/11/18 7:21 AM

Types of Data

- **Text and numeric descriptors**
 - Names, titles, journal of publication...
- **Software files** (or associated links)
 - Query, cleaning, nlp...
- **Data files** (or associated links)
 - tabular, clinical notes, data dictionaries
- **Identifiers** (hopefully persistent)
 - DOIs, URLs, Grant IDs...
- **Clinical Ontology Variables**
 - Diagnosis, Procedure, Medication, Labs...

Bibliographic and Project Metadata



Project Metadata



Project Title



Project Principal Investigator



Institution of Project Principle Investigator

Health Data Service Center Principle Investigator or Director

Institution of Health Data Service Center

Ethical Assessment and Review



a



IRB Approving Institution



IRB Approval Date



IRB Protocol ID #



Link to IRB Protocol



Full Text of IRB Protocol

Data Access and Collection

▼ Data Access

- Database Name
- Database DOI
- Database URL
- Institution Hosting Database
- Institutional Department or Center Overseeing Database

▼ Data Extraction - Structured











- Query Script_Structured
Data_File Name
- Link to Shared Query
Script_Structured Data
- Query Script_Structured
Data_ File Format
- Query Script_Structured
Data_Software Language
- Query Script_Structured
Data_File Version #
- Query Script_Structured
Data_Author(s)
- Query_Structured
Data_Execution Date

→ Data Extraction_Unstructured Data


Learning Effective Treatment Pathways for Type-2 Diabetes from a clinical data warehouse




[Rohit Vashisht](#), PhD,¹ [Ken Jung](#), PhD,¹ and [Nigam Shah](#), MBBS, PHD¹











 rohit43 modified temp table names		Latest commit d34242a on Dec 12, 2017
 R	Update runT2DOutcomeStudy.R	4 months ago
 extras	second commit	11 months ago
 inst	modified temp table names	3 months ago
 man	updated R functions	11 months ago
 DESCRIPTION	Update DESCRIPTION	11 months ago
 DiabetesTxPath.Rproj	first commit	a year ago
 NAMESPACE	first commit	a year ago
 README.Rmd	Update README.Rmd	4 months ago
 README.html	updated R functions	11 months ago


Linking and describing cohort collection files

 **rohit43 / DiabetesTxPath**




 Watch 1  Star 0  Fork 3

 Code  Issues 0  Pull requests 0  Projects 0  Insights

Branch: master ▾ **DiabetesTxPath / R /**   

 **rohit43** Update runT2DOutcomeStudy.R Latest commit 93cd11c on Dec 2, 2017


..



 buildOutcomeCohort.R	Update buildOutcomeCohort.R	4 months ago
 createExposureCohorts.R	Update createExposureCohorts.R	4 months ago
 createNegativeControlOutcomeCohorts.R	Update createNegativeControlOutcomeCohorts.R	4 months ago

▼ Data Extraction - Structured

Query Script_Structured Data_File Name	createExposureCohorts.R
Link to Shared Query Script_Structured Data	https://github.com/rohit43/DiabetesTxPath/blob/master/R/createExposureCohorts.R
Query Script_Structured Data_File Format	.R
Query Script_Structured Data_Software Language	R, SQL, and JSON
Query Script_Structured Data_File Version #	2.0
Query Script_Structured	Rohit Vashisht, Jamie Weaver

Mapping Ontologies, Vocabularies, and Standards

Element Name	Element Description
 Participant Inclusion Criteria	Description of cohort inclusion and extraction criteria operationalized through data brok

 a 1 .. N 


Enter Field Title

ICD 9 Codes_Participant Inclusion

Enter Field Description (Help Text)

List of ICD 9 Codes Operationalizing cohort inclusion criteria

Enter Default Value


 VALUES

MULTIPLE

REQUIRED

SUGGESTIONS

HIDDEN

 INSTANCE TYPE

Name	Type	Source	Identifier	No. Values
DISEASES AND INJURIES	Branch	ICD9CM	001-999.99	-

SEARCH

→ Data Extraction_Unstructured Data

▼ Inclusion Criteria

- ICD 9 Codes_Participant Inclusion
- ICD10 Codes_Participant Inclusion
- HCFA/HCPCS Procedure Codes_Participant Inclusion
- CPT Codes_Participant Inclusion
- RXNORM Medication Codes_Participant Inclusion
- SNOMED Medication Codes_Participant Inclusion
- LOINC Laboratory Codes_Participant Inclusion

```
{"ConceptSets":[{"id":0,"name":"HbA1c_v2","expression":{"items":[{"concept":{"CONCEPT_ID":3004410,"CONCEPT_NAME":"Hemoglobin A1c (Glycated)","STANDARD_CONCEPT":"S","INVALID_REASON":"V","CONCEPT_CODE":"4548-4","DOMAIN_ID":"Measurement","VOCABULARY_ID":"LOINC","CONCEPT_CLASS_ID":"Lab Test","INVALID_REASON_CAPTION":"Valid","STANDARD_CONCEPT_CAPTION":"Standard"}}]}
```



LOINC Laboratory Codes_Participant Inclusion	
1	▼
2	▼
3	▼
4	▼
5	▼
6	▼
7	▼
8	▼
9	▼
10	▼

```
"Data Extraction_Unstructured Data": {
  "@context": {
    "Text Extraction Script_Unstructured Data_File Name": "https://schema.metadacenter.org/properties/e05ead1d-4a0b-4b94-8000-000000000000",
    "Text Extract Script_Unstructured Data_Software Language": "https://schema.metadacenter.org/properties/719ec416-4b94-8000-000000000000",
    "Query Script_Unstructured Data_File Name": "https://schema.metadacenter.org/properties/51645b80-69c3-4e51-9b94-800000000000",
    "Query Script_Unstructured Data_File Format": "https://schema.metadacenter.org/properties/59f02b9f-ea6b-4ea5-b500-000000000000",
    "Query script_Unstructured Data_Software Language": "https://schema.metadacenter.org/properties/3e9e8745-4b96-4b94-8000-000000000000",
    "Query Script_Unstructured Data_Author(s)": "https://schema.metadacenter.org/properties/ff32801e-7253-4a75-97d6-800000000000",
    "Query Script_Unstructured Data_Execution Date": "https://schema.metadacenter.org/properties/2777b740-4ef4-4fac-8000-000000000000",
    "Text Mining Script_Unstructured Data_Author(s)": "https://schema.metadacenter.org/properties/a221b43d-3832-4ed0-8000-000000000000",
    "Text Mining Script_Unstructured Data_Date Executed": "https://schema.metadacenter.org/properties/4166f856-17ba-4b94-8000-000000000000",
    "Link to Shared Query Script_Unstructured Data": "https://schema.metadacenter.org/properties/6a4d573b-a83d-4e02-8000-000000000000",
    "Link to Share Text Mining Script_Unstructured Data": "https://schema.metadacenter.org/properties/9865a4cd-b9e0-4b94-8000-000000000000",
    "Query Script_Unstructured Data_File Version": "https://schema.metadacenter.org/properties/4675ccfb-446f-4795-9000-000000000000",
    "Text Extraction Script_Unstructured Data_File Version": "https://schema.metadacenter.org/properties/26c3f13a-e05a-4b94-8000-000000000000"
  },

```

```
"Inclusion Criteria": {
  "@context": {
    "ICD 9 Codes_Participant Inclusion": "https://schema.metadacenter.org/properties/5931e35d-4ba9-41d2-a1f4-bbda63-4b94-8000-000000000000",
    "CPT Codes_Participant Inclusion": "https://schema.metadacenter.org/properties/1b2b6f54-7f92-4efb-87b9-555640b3-4b94-8000-000000000000",
    "RXNORM Medication Codes_Participant Inclusion": "https://schema.metadacenter.org/properties/36d6470a-fe30-482b-8000-000000000000",
    "SNOMED Medication Codes_Participant Inclusion": "https://schema.metadacenter.org/properties/674e211e-a657-43d6-8000-000000000000",
    "LOINC Laboratory Codes_Participant Inclusion": "https://schema.metadacenter.org/properties/08e85ded-623c-4fec-8000-000000000000",
    "HCFA/HCPCS Procedure Codes_Participant Inclusion": "https://schema.metadacenter.org/properties/5af3c336c-5dc1-4b94-8000-000000000000",
    "ICD10 Codes_Participant Inclusion": "https://schema.metadacenter.org/properties/867ab168-46d9-48c5-8f59-941a11-4b94-8000-000000000000"
  },
  "ICD 9 Codes_Participant Inclusion": [
    {}
  ],
  "CPT Codes_Participant Inclusion": [
    {}
  ],

```


Potential Future Use Cases



Australian Government

Australian Digital Health Agency



OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



IHME Celebrating 10 years of
measuring what matters
Institute for Health Metrics and Evaluation



Treatment Pathways in Chronic Disease

Objective: The objective of this study is to characterize the prevalence of different treatment pathways for three chronic diseases: Hypertension, Type II Diabetes, and Depression. We will systematically summarize the treatment pathways observed among patients who have at least 3 years of continuous observation and persistent treatment following initiation. We will stratify the results by year to evaluate temporal trends, and will further stratify by data source to determine if treatment pathways vary by population, geography, and data capture process.

Rationale: While numerous treatment guidelines exist for chronic conditions, there is a paucity of data on the real-world treatment pathways that patients experience in practice. Understanding these pathways is essential for establishing context around questions of drug utilization, effectiveness, and adherence.

Project Leads: Patrick Ryan, Jon Duke, George Hripcsak, Martijn Schuemie, Nigam Shah

Coordinating Institution(s): Janssen R&D, Columbia University, Regenstrief Institute, Stanford University

Additional Participants:

Full Protocol: [Hypertension Treatment Pathways 12-4-2014](#)

Initial Proposal Date: 12/3/2014

Launch Date: 12/5/2014

Study Closure Date: 12/31/2014

Results Submission: [Email](#) or SFTP

Requirements

CDM: V4 or V5

Database Dialect: SQL Server, Postgres, Oracle

Software: SQL as above, R (optional)

Proposed Next Steps

- ❖ Assess impact and scalability with more use cases
- ❖ Iterative reviews with other members of the RDA to vet granularity and interoperability
- ❖ Continued collaboration with CEDAR to test and develop functionality
- ❖ Link brokerage metadata to additional scientific outputs across repositories

Washington University
Center for Biomedical Informatics
Adoption Case Study

Incorporating Data Citation in a Biomedical Repository

An Implementation Use Case

Snehil Gupta

Connie Z. Regan

Brian Romine

Daniel Vianello

Cynthia Hudson Vitale

Leslie McIntosh

March 2019



Funding Support

**MacArthur Foundation 2016 Adoption Seeds
program Foundation through a sub-contract with
Research Data Alliance**

Washington University Institute of Clinical and
Translational Sciences

NIH CTSA Grant Number UL1TR000448 and
UL1TR000448-09S1

Siteman Cancer Center at Washington University
NIH/NCI Grant P30 CA091842-14

WashU CBMI Research Reproducibility

Resources Repository

https://github.com/CBMIWU/Research_Reproducibility

Slides

<http://bit.ly/2nxjNK8>

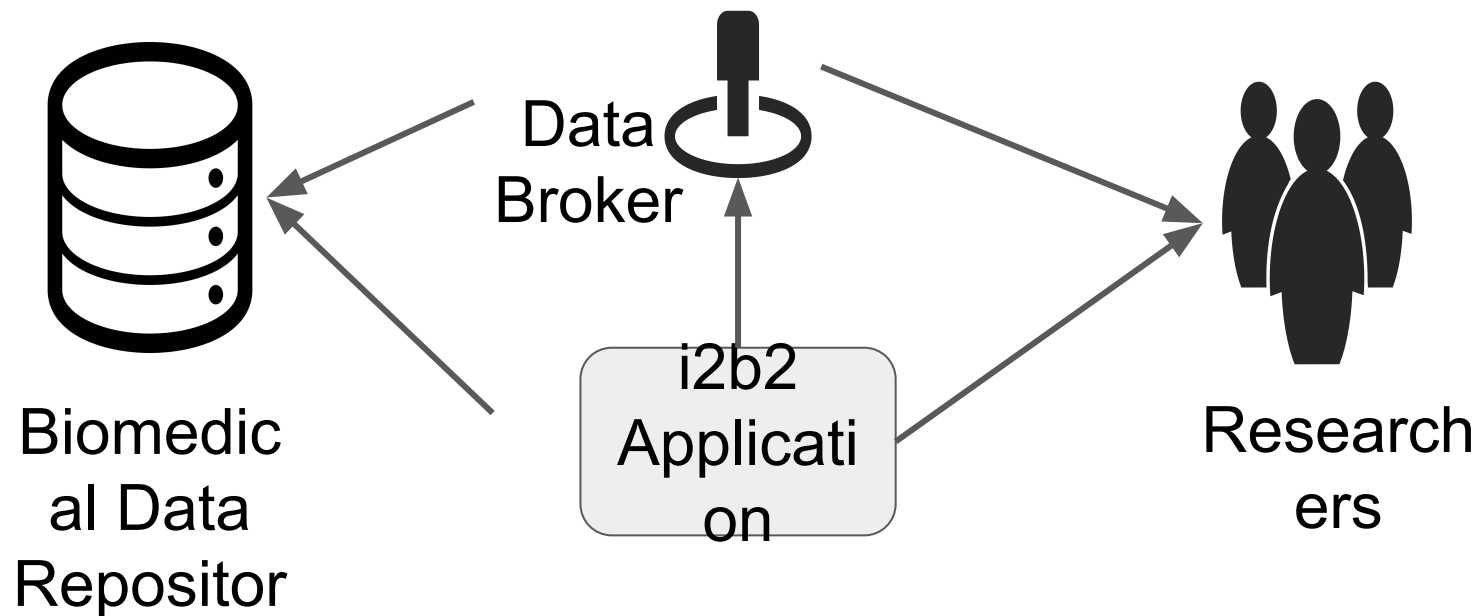
Bibliography

https://www.zotero.org/groups/biomedical_informatics_resrepo


Background

BDaaS

Biomedical Data as a Service



Move some of the responsibility of
reproducibility

Biomedical Biomedical
Researcher  Pipeline

RDA Data Citation WG Recommendations

- ▶ R1: Data Versioning
- ▶ R2: Data Timestamping
- ▶ R3, R9: Query Store
- ▶ R7: Query Timestamping
- ▶ R8: Query PID
- ▶ R10: Query Citation

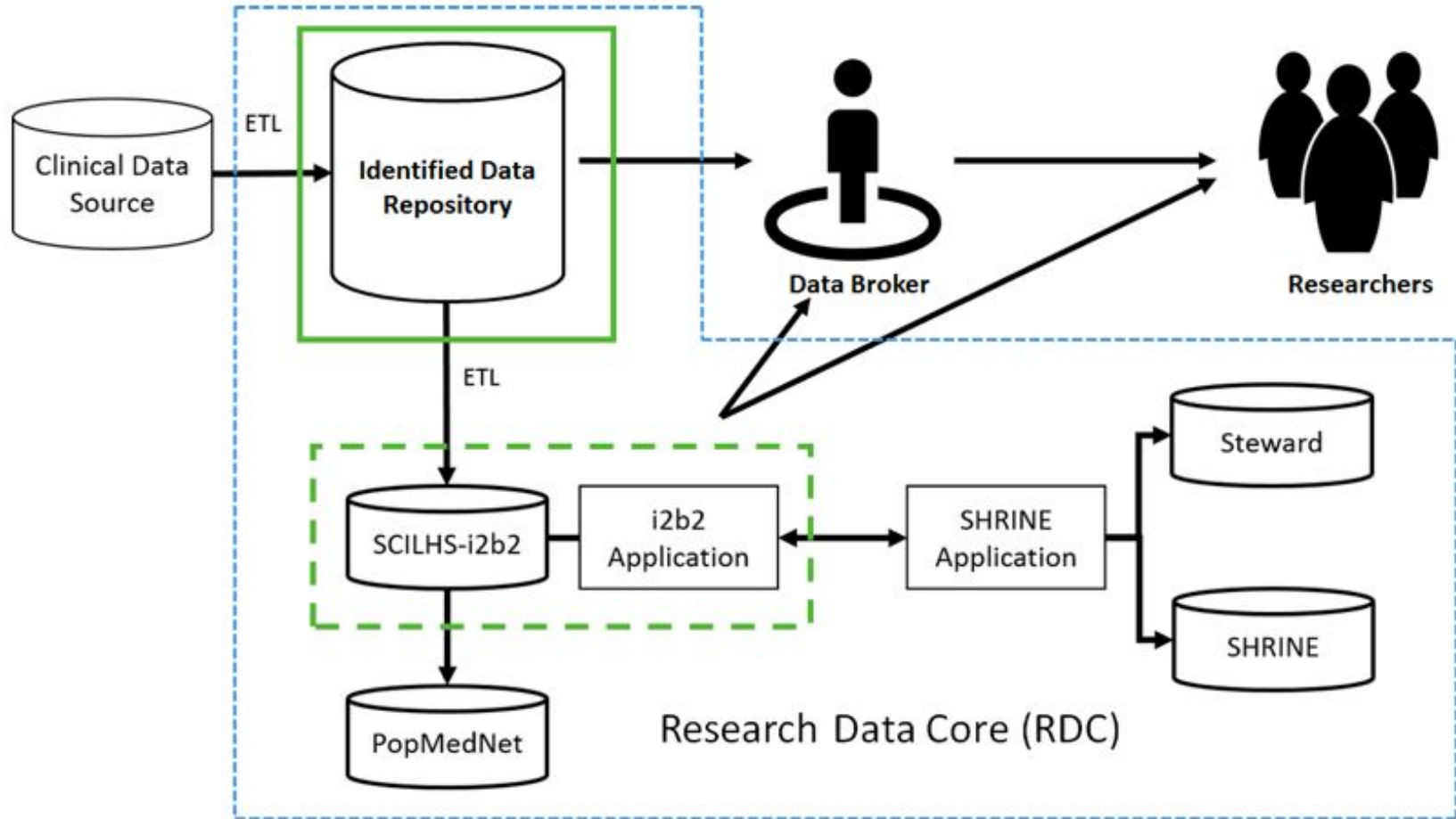
Biomedical Adoption Project Goals

- ▶ Implement RDA Data Citation WG recommendation to local Washington U i2b2
- ▶ Engage other i2b2 community adoptees
- ▶ Contribute source code back to i2b2 community

Approach

1. Assess the Center's data infrastructure
2. Conduct gap analysis of the infrastructure against RDA-DC recommendations
3. Define our local requirements
4. Define and evaluate our approach

Infrastructure



Gap Analysis

Table 2: Gap Analysis Summary

Database		Data Versioning (R1)	Data Timestamp (R2)	Query Store (R3/R9)	Query Timestamp (R7)	Query (R8)	PID	Citation Text (R10)
Identified Repository i2b2 (Local)	Data	Yes (default)	Yes (default)	No	No	No		No
		No	No	Yes (i2b2 default)	Yes (i2b2 default)	Yes (i2b2 default)		No
		No	No	Yes (i2b2 default)	Yes (i2b2 default)	Yes (i2b2 default)		No
		No	No	No	No	No		No

Internal Implementation Requirements

- ▶ Scalable
- ▶ Available for PostgreSQL
- ▶ Actively supported
- ▶ Easy to maintain
- ▶ Easy for data brokers to use

Implementation

R1 and R2 Implementation

1



PostgreSQL
Extension
"temporal_tables"

RDC.ta

c1	c2	c3



sys_per
iod

sys_per iod

2

trigg
ers

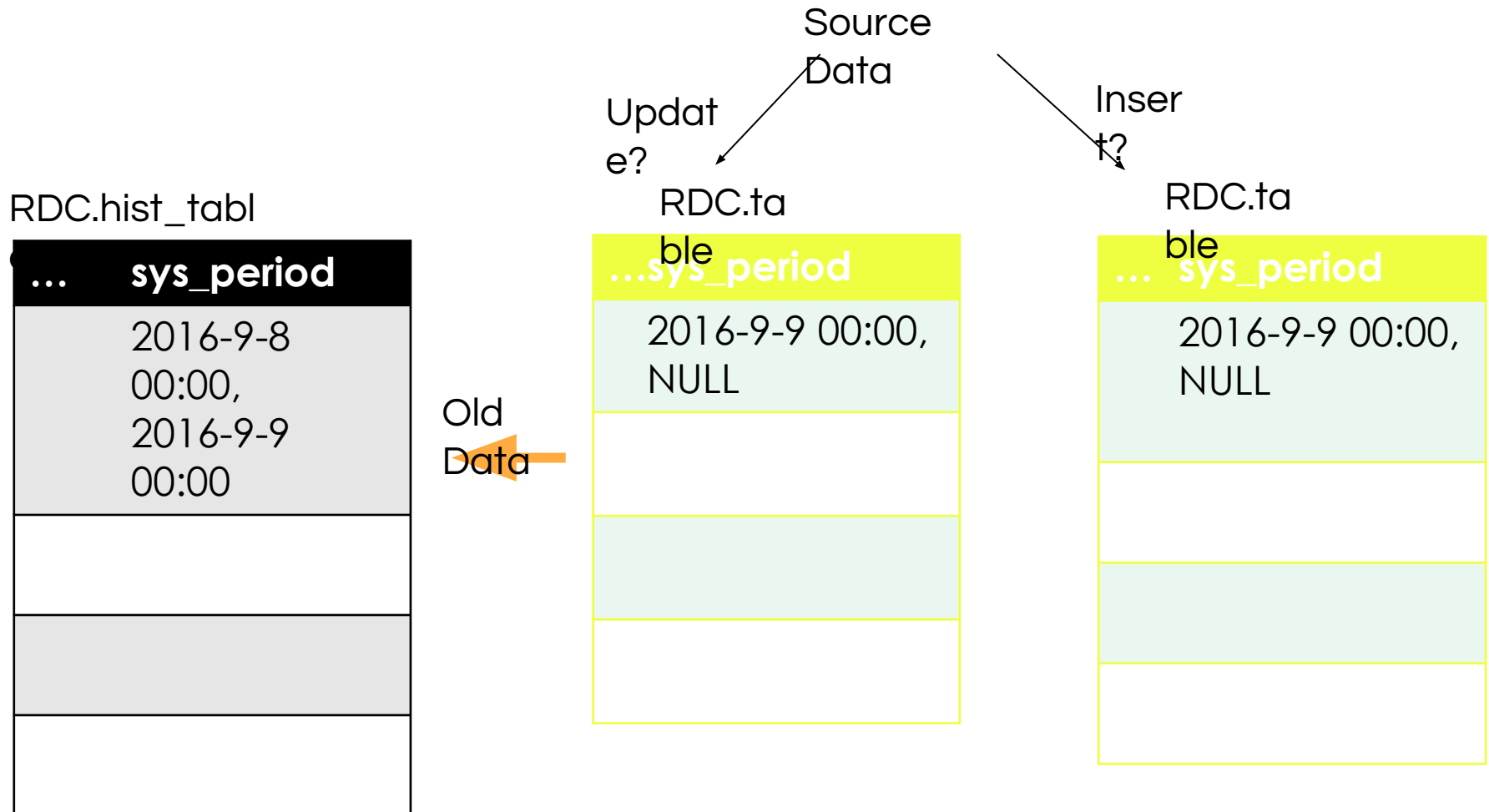
3

RDC.hist_ta

c 1	c 2	c 3	sys_per iod

*stores history
of data
changes

ETL Incrementals



R3, R7, R8, R9, and R10 Implementation

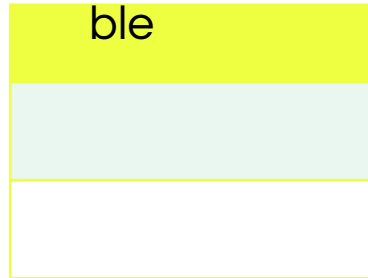
1



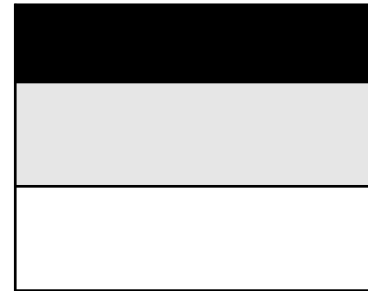
PostgreSQL
Extension
"temporal_tables"

2

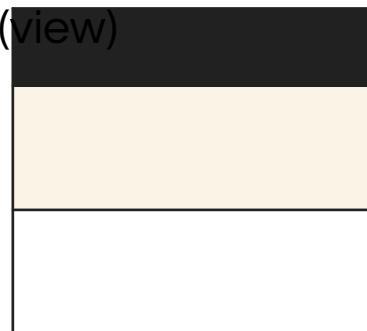
RDC.ta
ble



RDC.hist_t



RDC.table_with_history
(view)

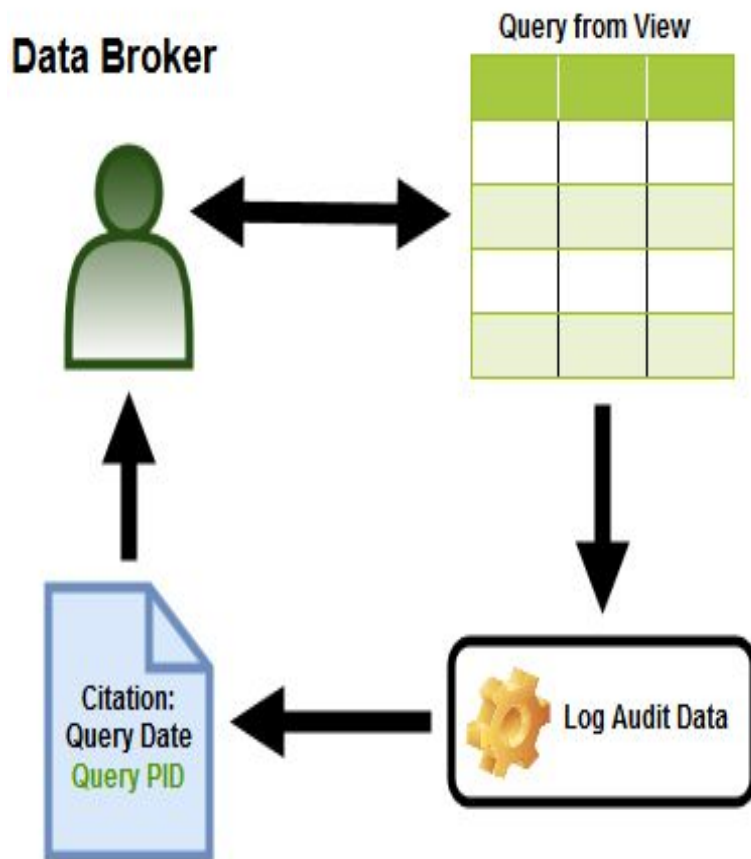


3

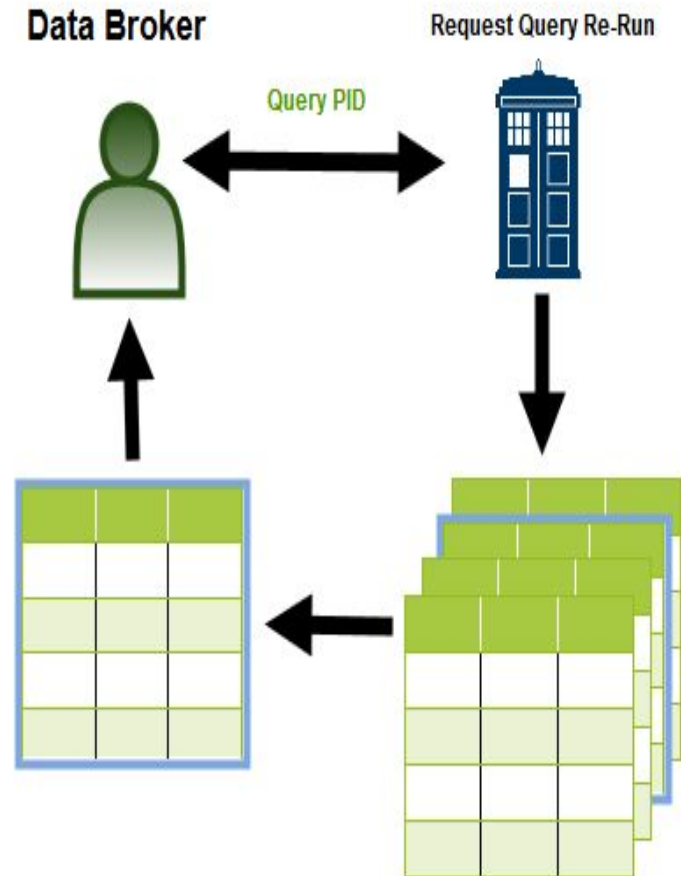
- functions
- triggers
- query audit tables

Data Reproducibility Workflow

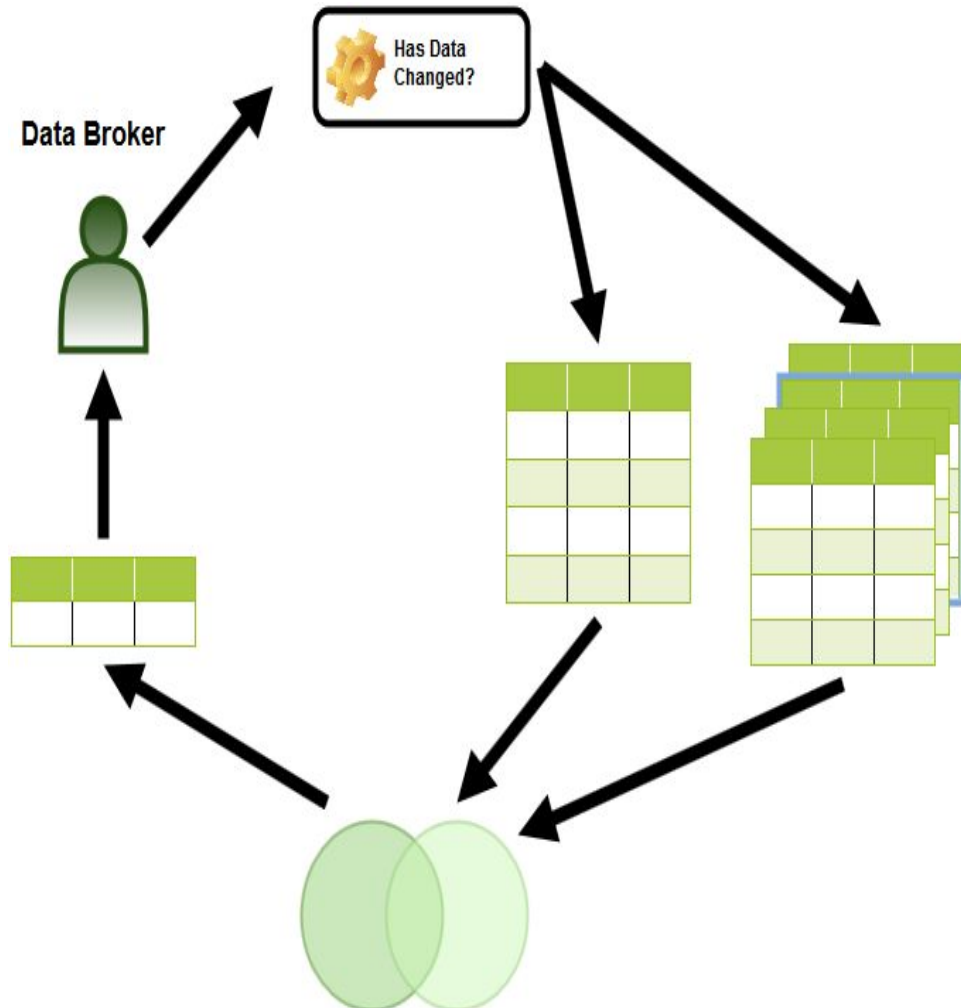
TODAY



SOME TIME LATER



Bonus Feature: Determine if Change Occurred



Implementation in Practice

Temporal tables

For RDA compliance, we've instituted Temporal tables

This means that each of these tables has a historical analogue.

The naming convention is **hist_domain**

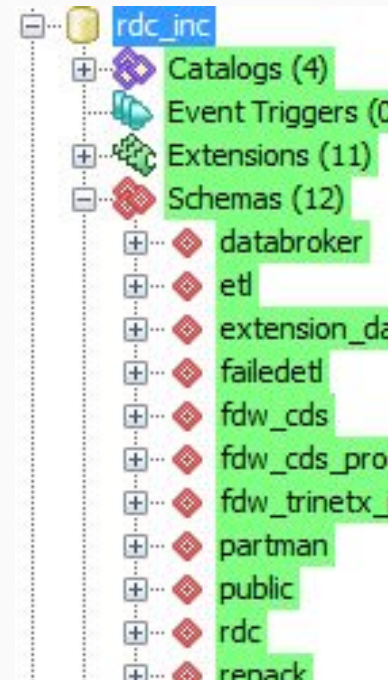
These aren't very helpful for data brokers

You'll use the view **domain_with_history**

Temporal query workflow

Currently all RDC data is in the RDC schema.

The plan is for databrokers to have access to the databroker schema that will automatically log queries run on it.



Using the databroker schema

```
1 SELECT * FROM databroker.encounter
2 LIMIT 2
3
4
```

!!!

Output pane

Data Output

Explain

Messages

History

	encounterid bigint	patientaccountnumber character varying(2000)	encounterstartdate timestamp without time zone	encounterenddate timestamp without time zone	encounterstatus character varying(50)
1	235[REDACTED]	1257[REDACTED]	200[REDACTED] 00:00	200[REDACTED] 23:59:00	Final
2	235[REDACTED]	272[REDACTED]	200[REDACTED] 00:00:00		In Progress

!!!

Using the databroker schema

The screenshot shows a SQL Editor interface with two main panes: the SQL Editor and the Output pane.

SQL Editor: The top pane contains a SQL query:

```
1 SELECT * FROM databroker.query_tracker
2 WHERE user_name = 'bromine'
3 AND tracktime LIKE '2016-12-19%'
4
```

Output pane: The bottom pane displays the results of the query in a table format. The table has the following columns:

	broker_id bigint	query_id bigint	tracktime timestamp with time zone	query text	user_name text	row_count bigint	formal_citation text
1	527	124	2016-12-19 15:52:32.066929-06	select * from databroker.encounter	bromine	1	

Using the databroker schema

previous queries

```
1 SELECT databroker.has_my_data_changed('previous', 124)
2
3 |
```

output pane

Data Output Explain Messages History

	has_my_data_changed boolean
1	f

Rerunning a historical query

We have a function `rerun_historical_query()`

For the earlier example we'll just use these commands

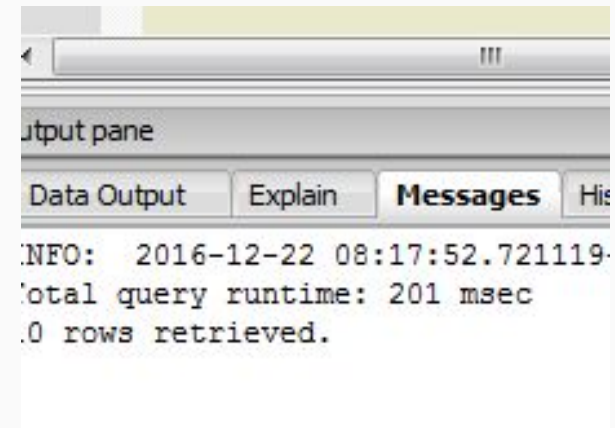
```
select  
databroker.rerun_historical_query('cursorname', 124); fetch all from cursorname;
```

These need to be run together, the first creates a cursor and the second displays it. They need to be highlighted and run at once.

Citations for queries

After running your query, just click on the messages tab in the output pane on PG admin. There is an info: note raised that will contain the information necessary to cite the dataset you've just generated.

Include this when you deliver, so the researcher can cite it. It also contains the information needed to easily re-run the query.



Future Developments

- ▶ Develop a process for sharing Query PID with researchers in an automated way
- ▶ Resolve Query PIDs to a landing page with Query metadata
- ▶ Implement research reproducibility requirements in other systems as possible

Outcomes and Support

Return on Investment (ROI) - Estimated

- ▶ 20 hours to complete 1 study
- ▶ \$150/hr (unsubsidized)
- ▶ \$3000 per study
- ▶ 115 research studies per year
- ▶ **14 replication studies**

Center for Biomedical Informatics @WUSTL

Teams for Reproducible Research
NIH-NLM Supplement

Leslie McIntosh

Cynthia

Hudson-Vitale

Anthony Juehne

Rosalia Alcoser

Xiaoyan 'Sean' Liu

Brad Evanoff

RDA Collaborators

Andreas Rauber

Stefan Pröll

Research Data
Alliance

Leslie McIntosh

Cynthia

Hudson-Vitale

Anthony Juehne

Snehil Gupta

Connie Zabarovskaya

Brian Romine

Washington University in St. Louis

WashU CBMI Research Reproducibility

Resources Repository

https://github.com/CBMIWU/Research_Reproducibility

Slides

<http://bit.ly/2nxjNK8>

Bibliography

https://www.zotero.org/groups/biomedical_informatics_resrepo

JOIN OUR COMMUNITY TODAY!



Individual Membership

<https://www.rd-alliance.org/get-involved/individual-membership.html>

RDA Recommendations and Outputs

<https://rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs>

Adoption Stories Series

<https://www.rd-alliance.org/recommendations-outputs/adoption-stories>

Webform to Submit Adoption Stories

<https://rd-alliance.org/add/adoption-stories>



TELL YOUR ADOPTION STORY

**RDA is inviting you to share your story
and inspire further adoption of RDA
outputs**

[RD-ALLIANCE.ORG/ADD/ADOPTION-STORIES](https://rd-alliance.org/add/adoption-stories)



**SHARE YOUR
ADOPTION
STORY TO
INSPIRE
FURTHER
UPTAKE OF
RDA OUTPUTS**

Image: Florian Klauer on Unsplash

SUBMIT YOUR STORY HERE:
[RD-ALLIANCE.ORG/ADD/ADOPTION-STORIES](https://rd-alliance.org/add/adoption-stories)



Questions & Discussion