

# Session 1: Facilitating review and data reuse across the research ecosystem

## Systematic Review Data Sharing and Standards



**Ian Saldanha**  
Johns Hopkins Bloomberg School of Public  
Health, USA

# Systematic Review Data Sharing and Standards

Ian Saldanha, MBBS, MPH, PhD

Associate Professor, Johns Hopkins Bloomberg School of Public Health

September 2, 2023

[isaldan1@jhu.edu](mailto:isaldan1@jhu.edu)

  @saldanha\_ian



# Disclosures

Director, Systematic Review Data Repository Plus (SRDR+)

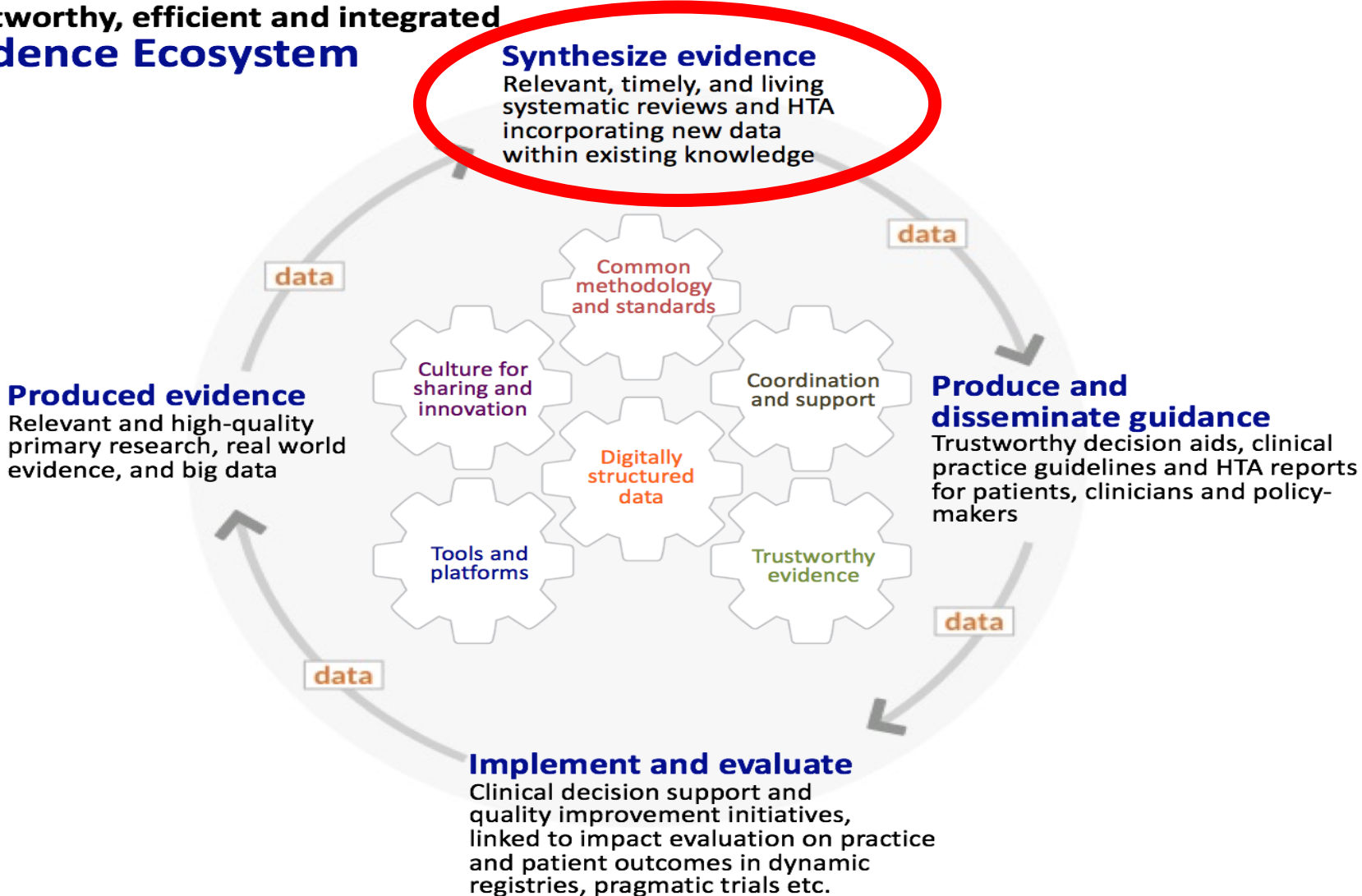


AHRQ has funded SRDR+ since platform launch (2012).

Disclaimer: No statement in my presentation should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services (DHHS).

# Interoperability in the evidence ecosystem

Trustworthy, efficient and integrated  
**Evidence Ecosystem**





# Why share systematic review data?

- **Open science**
  - Allows data/code verification/replication
- **Good science**
  - Allows checking robustness of data
- **Efficient science**
  - Facilitates updates
  - Facilitates inclusion in overviews, guidelines

Page MJ, Nguyen P-Y, Hamilton DG, Haddaway NT, Kanukula R, Moher D, McKenzie JE. Data and code availability statements in systematic reviews of interventions were often missing or inaccurate: a content analysis. *J Clin Epidemiol* 2022. DOI: [10.1016/j.jclinepi.2022.03.003](https://doi.org/10.1016/j.jclinepi.2022.03.003)

Saldanha IJ, Smith BT, Ntzani E, Jap J, Balk EM, Lau J. The Systematic Review Data Repository (SRDR): descriptive characteristics of publicly available data and opportunities for research. *Sys Rev* 2019. DOI: [10.1186/s13643-019-1250-4](https://doi.org/10.1186/s13643-019-1250-4)

# The evidence synthesis community is not doing great in sharing our data!



Journal of Clinical Epidemiology 147 (2022) 1–10

**ORIGINAL ARTICLE**

Data and code availability statements in systematic reviews of interventions were often missing or inaccurate: a content analysis

Matthew J. Page<sup>a,\*</sup>, Phi-Yen Nguyen<sup>a</sup>, Daniel G. Hamilton<sup>b</sup>, Neal R. Haddaway<sup>c,d,e</sup>, Raju Kanukula<sup>a</sup>, David Moher<sup>f,g</sup>, Joanne E. McKenzie<sup>a</sup>

<sup>a</sup>School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia  
<sup>b</sup>School of BioSciences, University of Melbourne, Melbourne, Australia  
<sup>c</sup>Leibniz-Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, Müncheberg, Germany  
<sup>d</sup>Stockholm Environment Institute, Linnégatan 87D, Stockholm, Sweden  
<sup>e</sup>African Centre for Evidence, University of Johannesburg, Johannesburg, South Africa  
<sup>f</sup>Centre for Journalism, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada  
<sup>g</sup>School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada

Accepted 3 March 2022; Published online 10 March 2022

---

**Abstract**

**Objectives:** To estimate the frequency of data and code availability statements in a random sample of systematic reviews with meta-analysis of aggregate data, summarize the content of the statements and investigate how often data and code files were shared.

**Methods:** We searched for systematic reviews with meta-analysis of aggregate data on the effects of a health, social, behavioral, or educational intervention that were indexed in PubMed, Education Collection via ProQuest, Scopus via Elsevier, or Social Sciences Citation Index and Science Citation Index Expanded via Web of Science during a 4-week period (between November 2, and December 2, 2020). Records were randomly sorted and screened independently by two authors until our target sample of 300 systematic reviews was reached. Two authors independently recorded whether a data or code availability statement (or both) appeared in each review and coded the content of the statements using an inductive approach.

**Results:** Of the 300 included systematic reviews with meta-analysis, 86 (29%) had a data availability statement, and seven (2%) had both a data and code availability statement. In 12/93 (13%) data availability statements, authors stated that data files were available for download from the journal website or a data repository, which we verified as being true. While 39/93 (42%) authors stated data were available upon request, 37/93 (40%) implied that sharing of data files was not necessary or applicable to them, most often because “all data appear in the article” or “no datasets were generated or analyzed”.

**Discussion:** Data and code availability statements appear infrequently in systematic review manuscripts. Authors who do provide a data availability statement often incorrectly imply that data sharing is not applicable to systematic reviews. Our results suggest the need for various interventions to increase data and code sharing by systematic reviewers. © 2022 Elsevier Inc. All rights reserved.

**Keywords:** Evidence synthesis; Open synthesis; Open science; Open data; Reproducibility of research; Research integrity

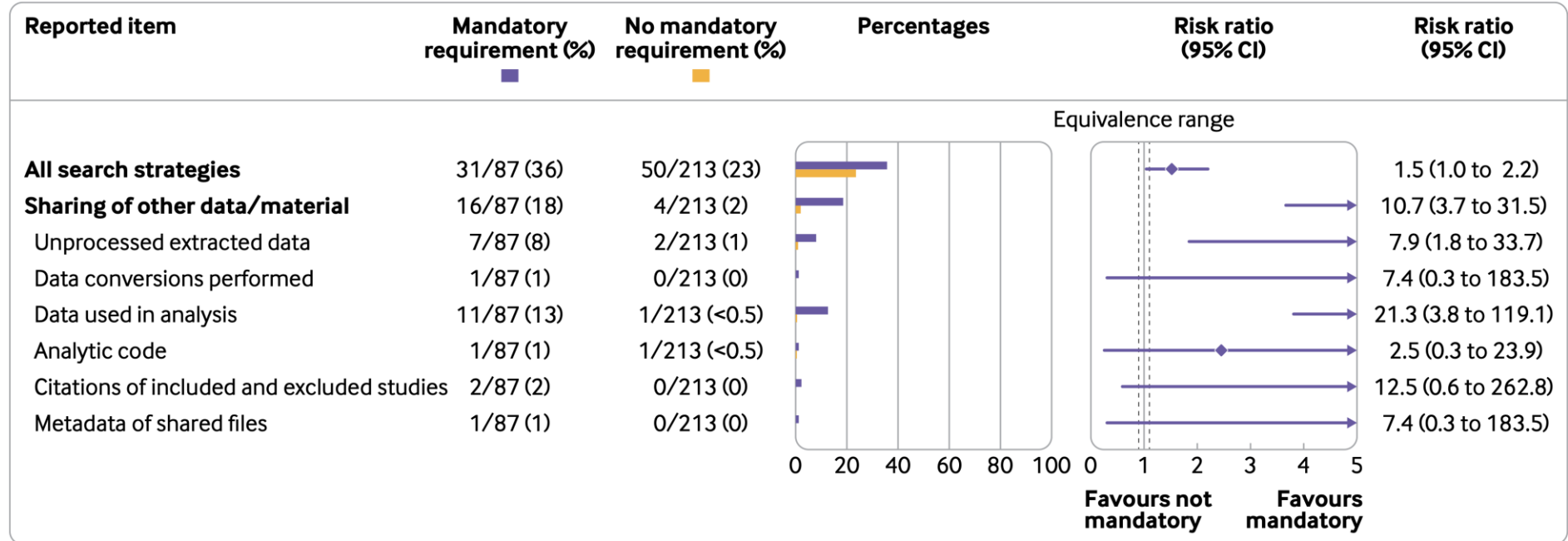
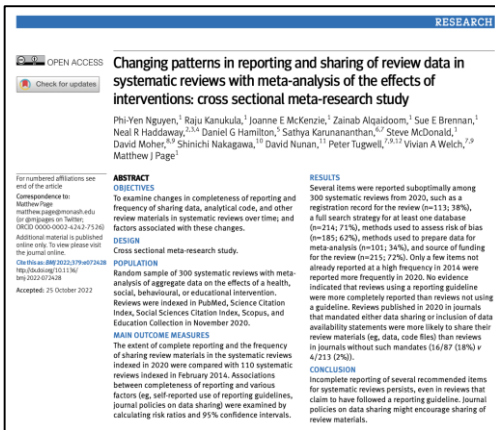
Random sample of 300 systematic reviews of interventions published in 2020

**Only 86 (29%)** had data availability statements!

**Only 12 (4%)** had data available for download (from the journal, a website, or a repository)!

Page MJ, Nguyen P-Y, Hamilton DG, Haddaway NT, Kanukula R, Moher D, McKenzie JE. Data and code availability statements in systematic reviews of interventions were often missing or inaccurate: a content analysis. *J Clin Epidemiol* 2022. DOI: [10.1016/j.jclinepi.2022.03.003](https://doi.org/10.1016/j.jclinepi.2022.03.003)

# Journals mandating data sharing or inclusion of data availability statements makes a difference!



**Fig 5 | Association between journals' data sharing requirements and reported items. Mandatory requirement=a mandatory instruction for sharing of data and materials, or in the absence of such data, a data availability statement stating why data were not shared and whether data are available on request. Equivalence range=0.9-1.1**

Nguyen P-Y, Kanukula R, McKenzie JE, et al. Changing patterns in reporting and sharing of review data in systematic reviews with meta-analysis of the effects of interventions: cross sectional meta-research study. *BMJ* 2022.

# Barriers to data sharing

- Insufficient motivation/career incentives
- Insufficient time
- Insufficient technical expertise
- Insufficient templates to facilitate data sharing
- Concerns about data ownership
- Fear of criticism



# Recommendations from Page and colleagues

“Data sharing via supplementary files or public repositories is an effective tool to improve reproducibility of systematic reviews and should be made a standard practice.”

FigShare. The Systematic Review Data Repository (SRDR) is an example of a repository for sharing materials specific to systematic reviews.

## Box 2 Recommendations for data and code sharing for systematic reviews

### What to share:

- Data such as that referred to in [Box 1](#).
- Analytic code used to generate results.
- Metadata (such as README files describing each data file shared).

### Where to share:

- In one of the general, domain-specific, or local institutional repositories listed on <https://www.re3data.org/>. Commonly used general repositories include the Open Science Framework, Dryad, and FigShare. The Systematic Review Data Repository (SRDR) is an example of a repository for sharing

file(s) (e.g., README file).

- Assign a persistent identifier (e.g., DOI) and license outlining the terms of use (e.g., CC BY) to each file shared.

# Goals and considerations of a systematic review

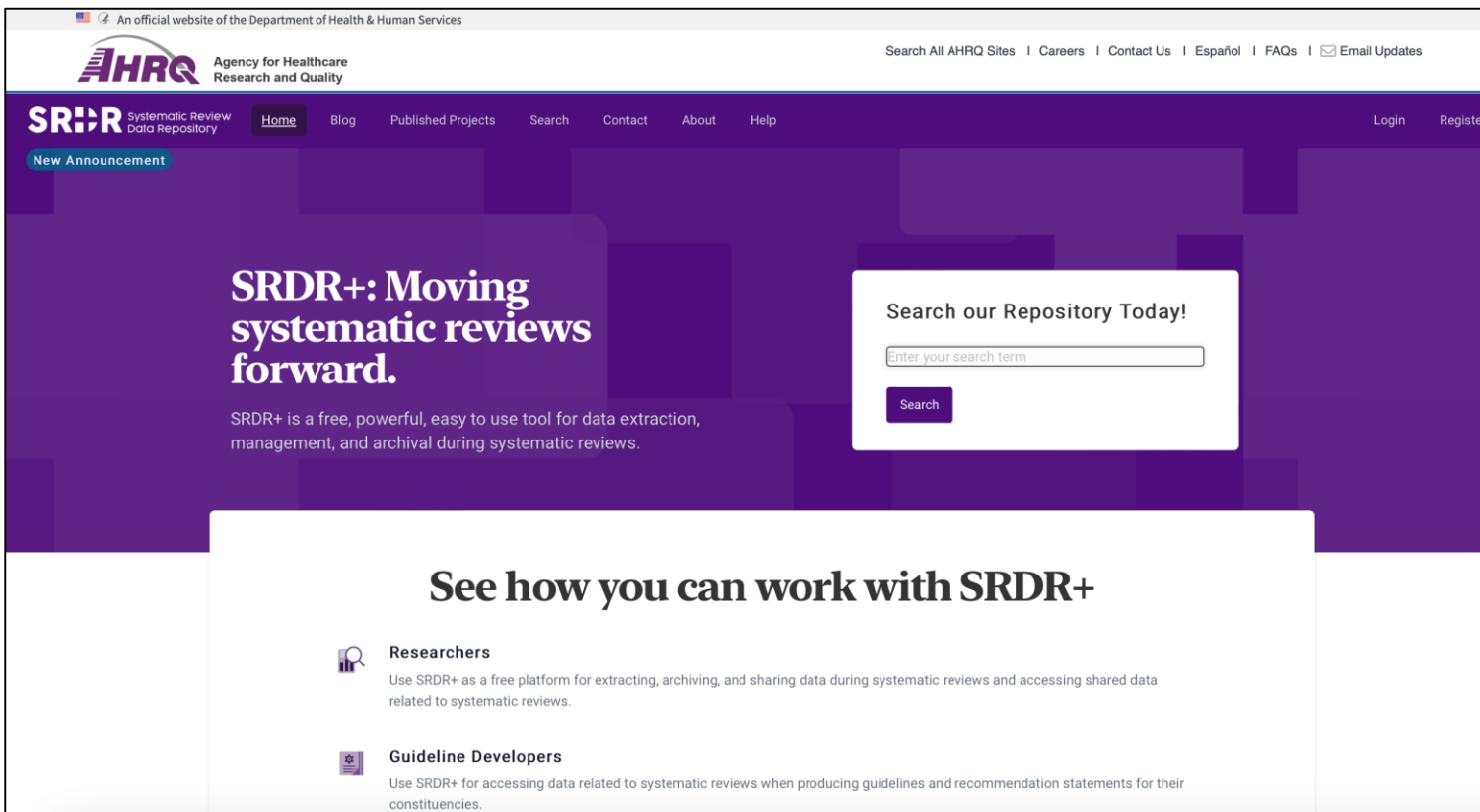
- Systematic reviewers need to conduct reviews with the goal in mind
  - Clinical/public health practice guidelines
  - Policy decisions
  - Academic interests
  - Business interests
  - Numerous other goals
- Considerations
  - Efficiency of the systematic review process
  - Use of the data by guideline developers (and others in the evidence ecosystem, e.g., clinical decision support tool developers)

# A data sharing platform specific to systematic reviews

Free platform with two main purposes:

1. Data management (screening, extraction)
2. Data archiving, sharing, and re-use

} **A community resource**



*User accounts = **11,006***

*Data shared publicly by systematic review authors:*

*Projects = **246***

*Studies = **21,174***

(As of July 31, 2023)

<https://srdplus.ahrq.gov>

# **FYI – Long oral presentation on SRDR+**

## **Presentation Title:**

The improved Systematic Review Data Repository Plus (SRDR+): A free, “FHIR-ed up” tool for screening, data extraction, and data sharing

**Date:** Tuesday September 5

**Time:** 11:25 am to 11:45 am

**Session Title:** Evidence synthesis innovations and technology

**Room:** Churchill

# Typical traditional data export from SRDR+

AutoSave OFF | Postpartum CoC\_June 26

Home Insert Draw Page Layout Formulas Data Review **View** Acrobat Tell me

Default | Ruler | Formula Bar | Zoom 100% | Gridlines | Headings | New Window | Arrange All | Unfreeze Panes | Freeze Top Row | Freeze First Column | Split | Hide | Unhide | Switch Windows | View Macro

AJ82 | fx

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
	Outcome	Outcome Description	Population Type	Population	Digest	Timepoint Unit	Arm Name	Arm Description	Total (N)	Events	Percentage	Arm Name	Arm Description	Total (N)	Events	Percentage	Arm Name	Arm Description	Total (N)	Events	Percentage	Arm Name	Arm Description	Total (N)	Events	Percentage	
63	al: Breastfeeding initiatic	Any breastfeeding	Categorical	All Participants	e564e74	4 mo	Video analysis		30	16	40	Peer counseling		23	12	52	Video breastfeeding		27	13	48	Standard of care		13	0	0	
64	al: Breastfeeding initiatic	Any breastfeeding	Categorical	All Participants	e564e74	2 mo	Video analysis		30	21	70	Peer counseling		4	3	75	Video breastfeeding		28	21	75	Standard of care		13	3	23	
65	al: Breastfeeding initiatic	Any breastfeeding	Categorical	All Participants	e564e74	1.4 wk	Video analysis		35	25	80	Peer counseling		32	23	72	Video breastfeeding		33	22	67	Standard of care		15	8	53	
66	al: Breastfeeding initiatic	Breastfeeding	Categorical	All Participants	65c33c0	PP	Home visit		72	61	88.4	Home visit		65	54	88.5	No home visit		100	65	76.5						
67	al: Contraceptive initiatic	Continuous	Categorical	All Participants	c93893f7	3 mo	Immediate		15	13	86.7	PP levonorgestrel		15	14	93.3	PP levonorgestrel		16	15	93.8						
68	al: Contraceptive initiatic	Continuous	Categorical	All Participants	c93893f7	6 mo	Immediate		15	13	86.7	PP levonorgestrel		15	14	93.3	PP levonorgestrel		16	15	93.8						
69	al: Breastfeeding initiatic	Any breastfeeding	Categorical	All Participants	8bf86eb9	6 mo	Immediate		15	12	80	PP levonorgestrel		15	12	80	PP levonorgestrel		16	10	62.5						
70	ncare Utilization: Unplanned	Urgent care	Categorical	All Participants	079b865	2 wk	Home visit		580	78	13	Pediatric		583	73	12											
71	ncare Utilization: Unplanned	Urgent care	Categorical	All Participants	079b865	1.5 mo	Home visit		580	163	28	Pediatric		583	182	31											
72	ncare Utilization: Unplanned	Hospital	Categorical	All Participants	77b7a98	2 wk	Home visit		580	4	1	Pediatric		583	3	1											
73	al: Mental health symptomatic	Depression	Categorical	All Participants	abc59eb	2 wk	Home visit		580	126	22	Pediatric		583	123	21											
74	al: Breastfeeding initiatic	Breastfeeding	Categorical	All Participants	73423f7	3 mo	Home visit		507	227	45	Pediatric		498	239	48											
75	al: Breastfeeding initiatic	Breastfeeding	Categorical	All Participants	73423f7	2 wk	Home visit		521	95	18	Pediatric		505	111	22											
76	ncare Utilization: Attended	Urgent care	Categorical	All Participants	b160f34f	2 wk	Home visit		508	64	12.6	Hospital		506	73	14.4											
77	ncare Utilization: Attended	Urgent care	Categorical	All Participants	b160f34f	1.5 mo	Home visit		508	131	25.8	Hospital		506	139	27.5											
78	ncare Utilization: Attended	Hospital	Categorical	All Participants	3e5183d	2 wk	Home visit		508	2	0.4	Hospital		506	1	0.2											
79	al: Mental health symptomatic	Significant	Categorical	All Participants	521672b	2 wk	Home visit		508	103	20.8	Hospital		506	86	17.9											
80	al: Breastfeeding initiatic	Any breastfeeding	Categorical	All Participants	50481b0	2 wk	Home visit		508	423	83.3	Hospital		506	424	83.8											
81	al: Breastfeeding initiatic	Exclusive	Categorical	All Participants	f296c9c0	3 mo	Text message		67	22	32.8	Beastfeeding		62	19	30.6											
82	al: Breastfeeding initiatic	Exclusive	Categorical	All Participants	f296c9c0	2 wk	Text message		71	36	50.7	Beastfeeding		56	25	44.6											

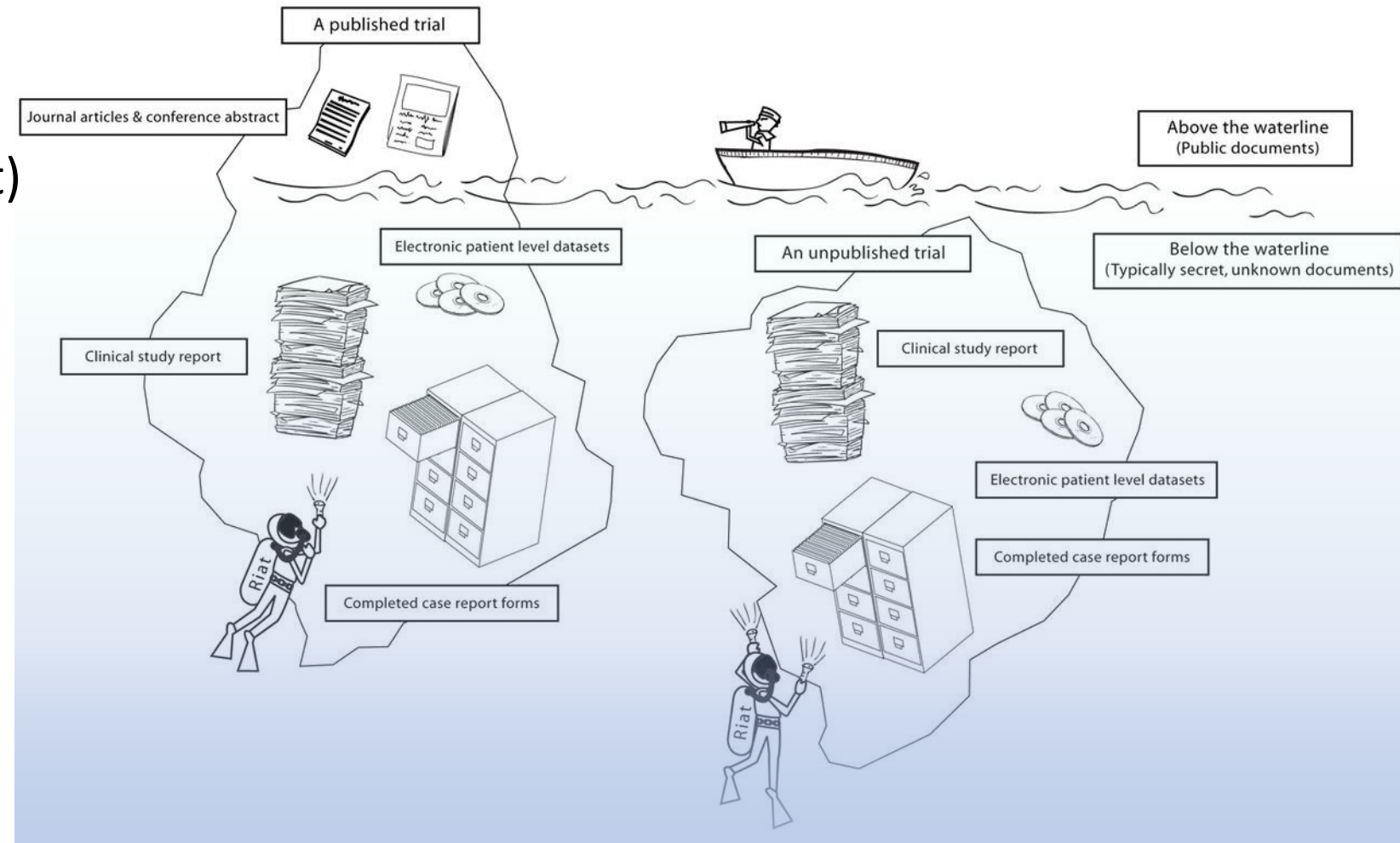
# Various sources of data for systematic reviews

## Public data sources

- Journal article
- Short report (e.g., letter, abstract)
- Trial registration
- Results on trial registry
- Information from regulators

## Non-public data sources

- Unpublished manuscript
- Individual participant data
- Grant proposal
- Study protocol
- Case report form
- Memos and emails



# But what data items should be shared with others in the ecosystem?

## AHRQ Evidence-based Practice Center Program-identified minimum data items for systematic review data sharing

- For all studies in AHRQ-funded systematic reviews of comparative effectiveness, comparative harms, or diagnostic accuracy
- Contractual deliverable for each systematic review contract
- Structured dataset that is separate from the systematic review report and journal articles

# Minimum items for data sharing – Level 1: Study level

## Study Citation Information (below)

- Title
- Authors
- Year of Publication (or of presentation if conference abstract only)
- Journal (or conference name if conference abstract only)
- Volume
- Issue
- Page Numbers (or abstract [number](#), if conference abstract only)
- PubMed ID (if available)
- Digital Object Identifier (DOI)
- Registration number (e.g., NCT number for ClinicalTrials.gov)

## Study Characteristics

- Study design (RCT or not RCT)
- Funder type

## Arm Details

- Name of each arm (group), e.g., intervention, diagnostic test

## Sample Characteristics

- Country
- Overall sample size at baseline
- Sample size at baseline for each arm (group)
- Age
- Sex
- Race/ethnicity

## Risk of Bias

- Risk of bias rating – for each item/domain in tool
- Risk of bias rating – overall (for main outcome of the report)

## Outcomes

- Name of each extracted outcome (*only* outcomes prioritized for Strength of Evidence tables)



# Minimum items for data sharing – Level 2: Review level

## Meta-Data

- Project Name
- Attribution (e.g., EPC name)
- Authors of report (when available)
- Methodology description (Systematic review vs. rapid review vs. technical brief vs. evidence map)
- PROSPERO Registration ID
- DOI of AHRQ report (when available)
- Funding source

## PICODS for each Key Question

- Populations
- Interventions/Exposures
- Comparators
- Outcomes
- Study Designs
- Settings

## Summary of evidence (for each Key Question)

- Strength of evidence tables (summary of prioritized outcomes, findings, and strength of evidence)

## Meta-analysis results (for each Key Question)

- Estimates from pair-wise meta-analyses (*only* outcomes prioritized for Strength of Evidence tables) (if conducted)
- Estimates from network meta-analyses (*only* outcomes prioritized for Strength of Evidence tables) (if conducted)

# Making evidence synthesis data interoperable with other platforms

For interoperability, the data must be computable/machine-readable

- For that, the data structure needs to follow a **standard**.



There are now rigorously developed standards (**FHIR**) for most information relevant to systematic reviews.

(e.g., citation information, study design, risk of bias, results)

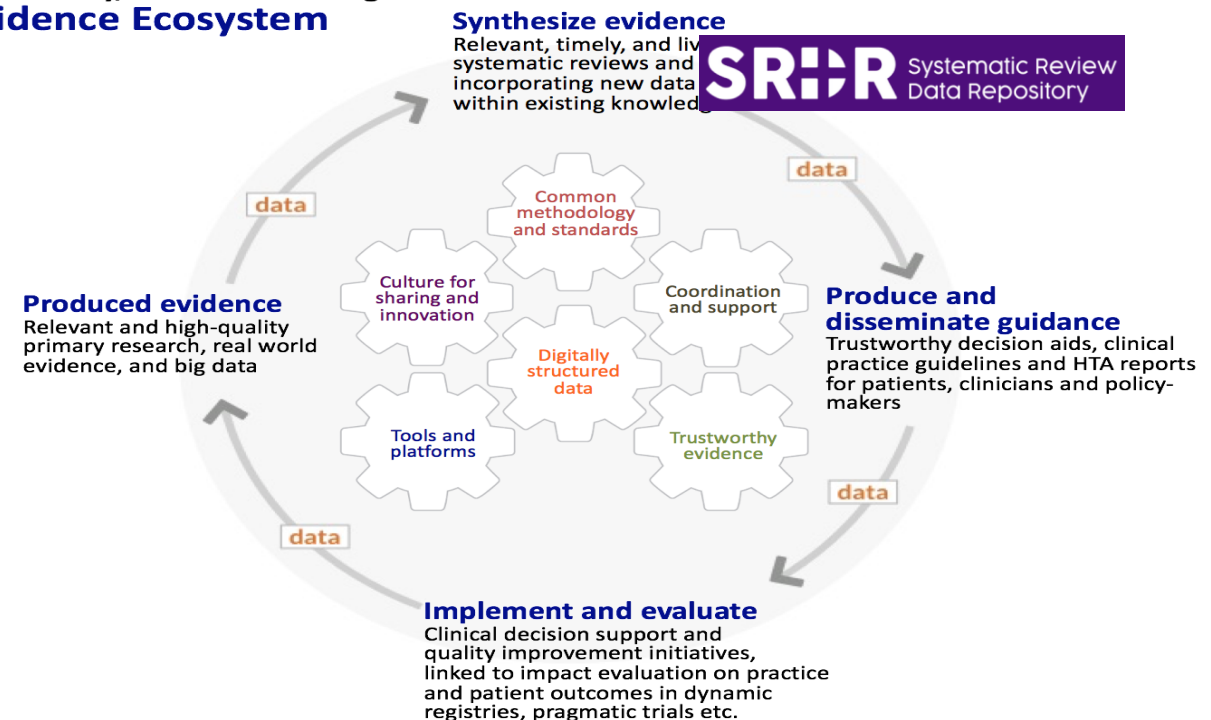
*Great opportunity for systematic review data to be interoperable among platforms!*

# So, where is SRDR+ with FHIR standards?

We are making the data in SRDR+ fully compliant with **FHIR** standards

- Data will be usable by other platforms in a machine-readable way using “API endpoints”
- Preliminary version done  
→ Being refined
- Other platforms are also working on this

Trustworthy, efficient and integrated  
**Evidence Ecosystem**



# Some broad opportunities for interoperability of structured data in the ecosystem

## 1. Exporting data among platforms (for systematic review updates)

One data extraction platform → Another data extraction platform

## 2. Exporting risk of bias ratings and meta-analysis results

Data extraction platform → MAGICapp (a guideline authoring platform)

## 3. Importing information from systematic review protocol registers

PROSPERO → Data extraction platform

## 4. Importing study information from registries (*Lene Seidler's talk will discuss this*)

ClinicalTrials.gov → Data extraction platform

# My take-home messages

1. Systematic reviews are not done in a vacuum. They are part of an evidence ecosystem and should be done with that goal in mind.
2. Prioritize bidirectional communication (with guideline developers and trial data generators).
3. The future calls for structured data outputs and data sharing, which can
  - Maximize utility of the evidence
  - Facilitate guideline development
  - Help reduce research waste
  - Contribute to open science.

# Thank you!



Ian Saldanha  
[isaldan1@jhu.edu](mailto:isaldan1@jhu.edu)



@saldanha\_ian

