Scientific Artificial Intelligence (SAI) Initiative

CRC
CENTER FOR RESEARCH COMPUTING

Human and AI Teaming for Discovery—Shared Memory

Charles F. Vardeman II and Paul Brenner

SciML@ND26

September 26, 2026

Slides: https://rb.gy/gcseh2

UNIVERSITY OF NOTRE DAME

L3D

CRC CENTER FOR RESEARCH COMPUTING

Trusted AI Knowledge (cyber)Frameworks



Our Frameworks Team

Scientific Artificial Intelligence (SAI) Initiative

Beyond Memory: A Templated Substrate for Heterogeneous Collaborative Knowledge Work with LLM Agents



P. Saboia Moreira, C. Vardeman II, and C. Sweet, “Beyond Memory: A Templated Substrate for Heterogeneous Collaborative Knowledge Work with LLM Agents,” Jul. 2026, doi: 10.5281/zenodo.21229213.

UNIVERSITY OF NOTRE DAME

($whoami)

Computational science

Molecular simulation, research computing and cyberinfrastructure
OpenMD (PhD 2009) · Notre Dame Center for Research Computing (2009-)

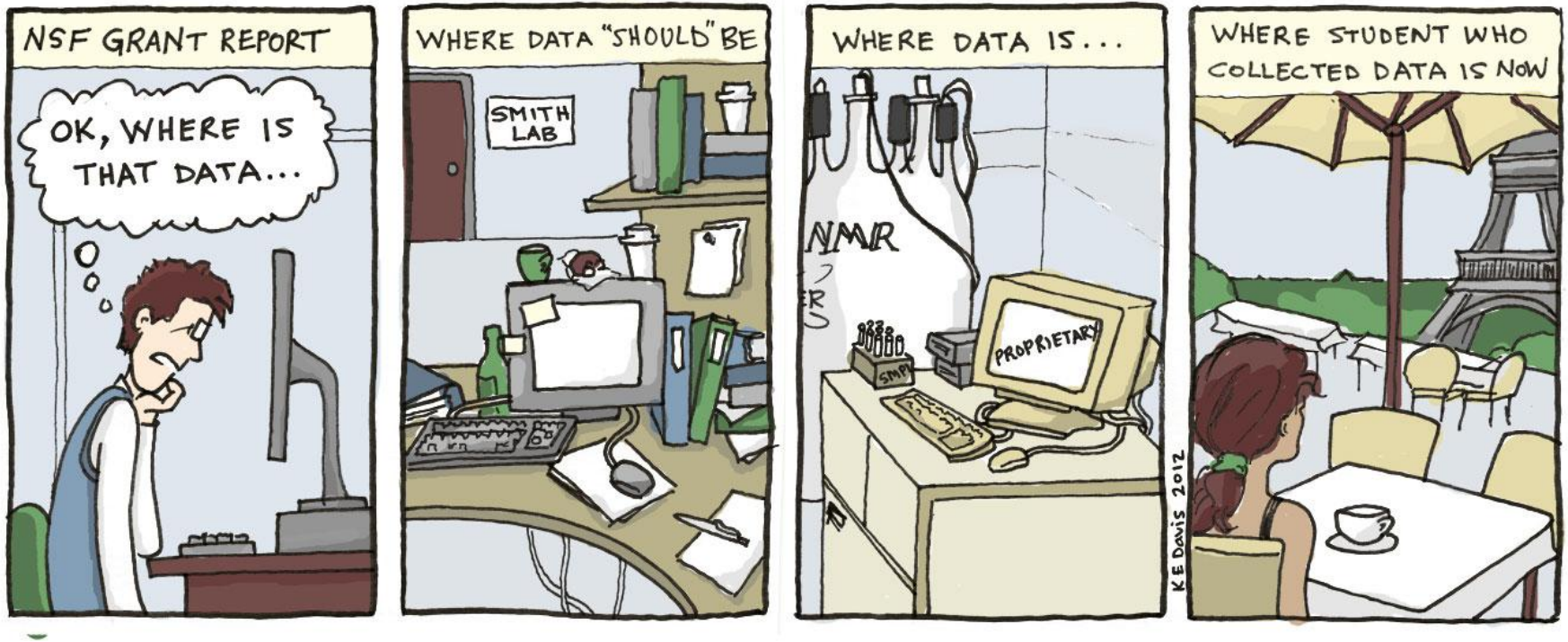
Linked and FAIR infrastructure

Preservation, domain models, and dataset discovery
DASPOS · Science on schema.org · CI-Compass

Assured agentic applications

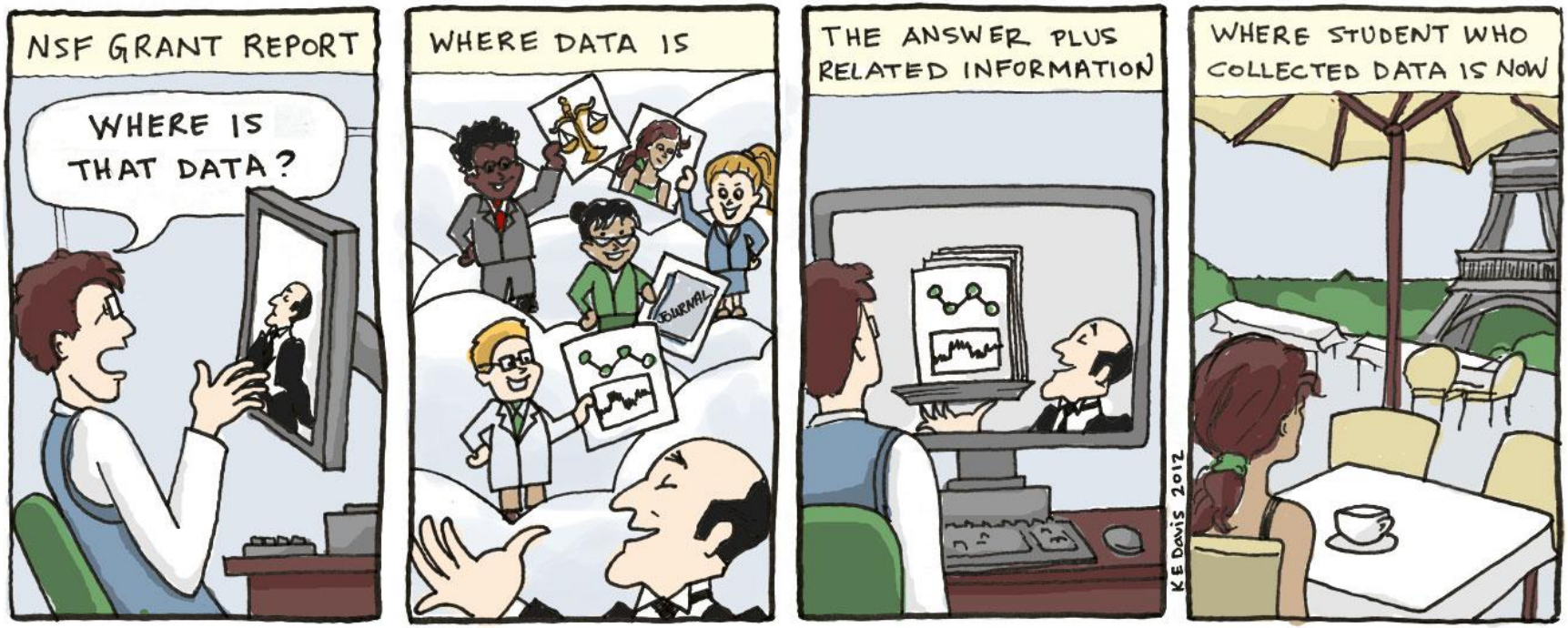
AI engineering, evaluation, and collaborative knowledge work
Trusted AI Frameworks · CRC / LA3D research context

The continuity problem, 2012



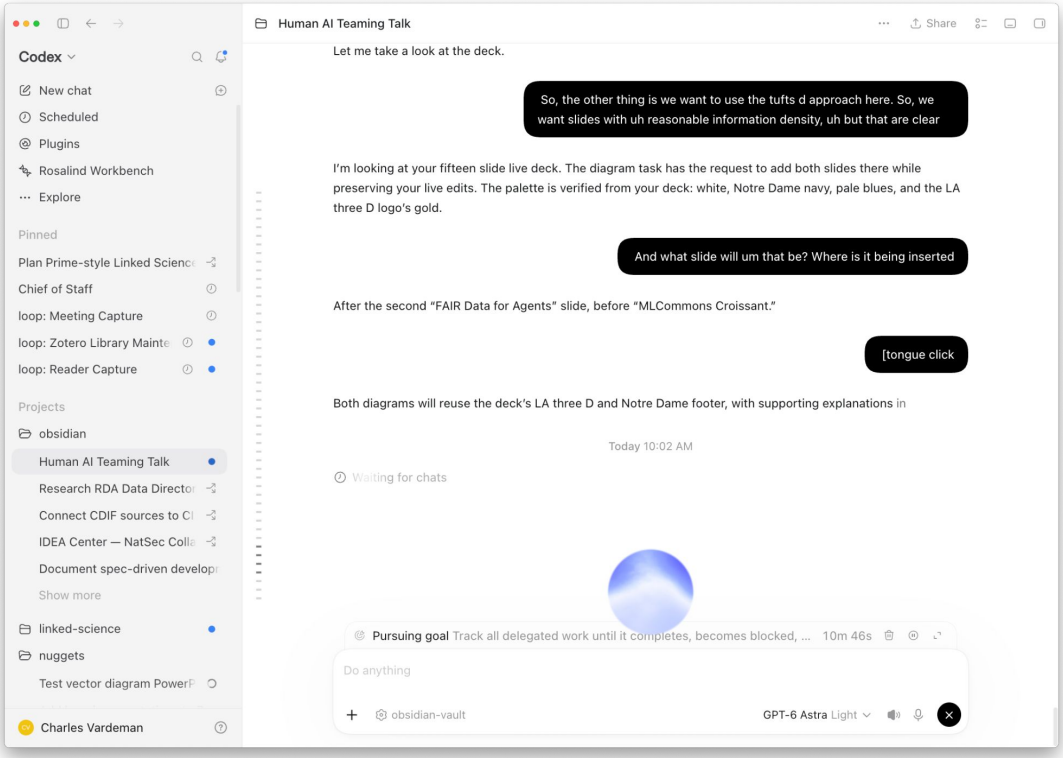
CREDIT: Kristina Davis (CRC, UND), 2012

A 2012 vision for research assistance



CREDIT: Kristina Davis (CRC, UND), 2012

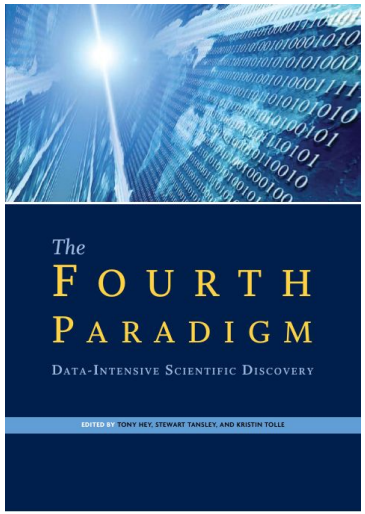
Today: collaborating with an agent through a working environment



Four paradigms of scientific discovery

Jim Gray's vision: data-intensive scientific discovery

New capabilities build on experiment, theory and simulation.



1 Empirical	**Observe / experiment** Observe and measure natural phenomena through experiments.
2 Theoretical	**Model / explain** Explain observations through models, laws and generalizations.
3 Computational	**Simulate** Simulate complex phenomena when models are difficult to solve analytically.
4 Data-intensive	**Explore data / integrate knowledge** Capture, curate and analyze data from instruments and simulations.

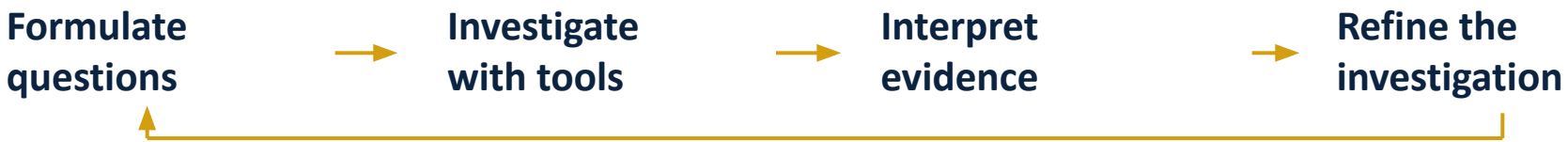
Data-intensive science unites experiment, theory and simulation through usable data.

Hey, Tansley & Tolle (eds.), The Fourth Paradigm (2009) · Gray's 2007 talk
Cover: © Microsoft · CC BY-SA 3.0 · front panel extracted

Accelerated knowledge discovery

Scientists and AI agents connect scientific capabilities across an adaptive research lifecycle.

PROPOSED SIXTH PARADIGM: accelerated knowledge discovery



Scientist: sets questions and constraints; interprets findings and reviews decisions.

Earlier paradigms remain the foundation; AI/ML adds learned inference.

1 Empirical	2 Theoretical	3 Computational	4 Data-intensive
Observe	Explain	Simulate	Integrate data

FIFTH PARADIGM
AI/ML inference
Learned prediction

Why it matters: shorter cycles from questions to evidence and revised understanding.

What technological components make this cycle possible?

NASA vision · Ramachandran, Bugbee & Bernabe-Moreno (2026) · §§1.2–1.5

What enables accelerated knowledge discovery?

Five complementary components support the adaptive research cycle.

Generative AI / LLMs	Interpret inputs, synthesize knowledge and propose hypotheses.
Domain foundation models	Predict scientific properties and detect domain-specific patterns.
Agentic frameworks	Plan and coordinate tasks, tools and model outputs.
Explicit knowledge representation	Describe entities, relationships and scientific constraints.
Computational infrastructure	Provide the scalable compute and storage needed for execution.

NASA: an adaptive scientific inquiry cycle.
RDA: support across planning, research, sharing and impact.

Where across this lifecycle do researchers most want help?

Ramachandran et al. (2026), Table 1 · Lifecycle connection: our synthesis

Research Data Alliance (RDA)–Microsoft: priorities for research agents

Where across the research lifecycle do researchers most want help?

SURVEY RATINGS

Literature Librarian

3.35 / 4

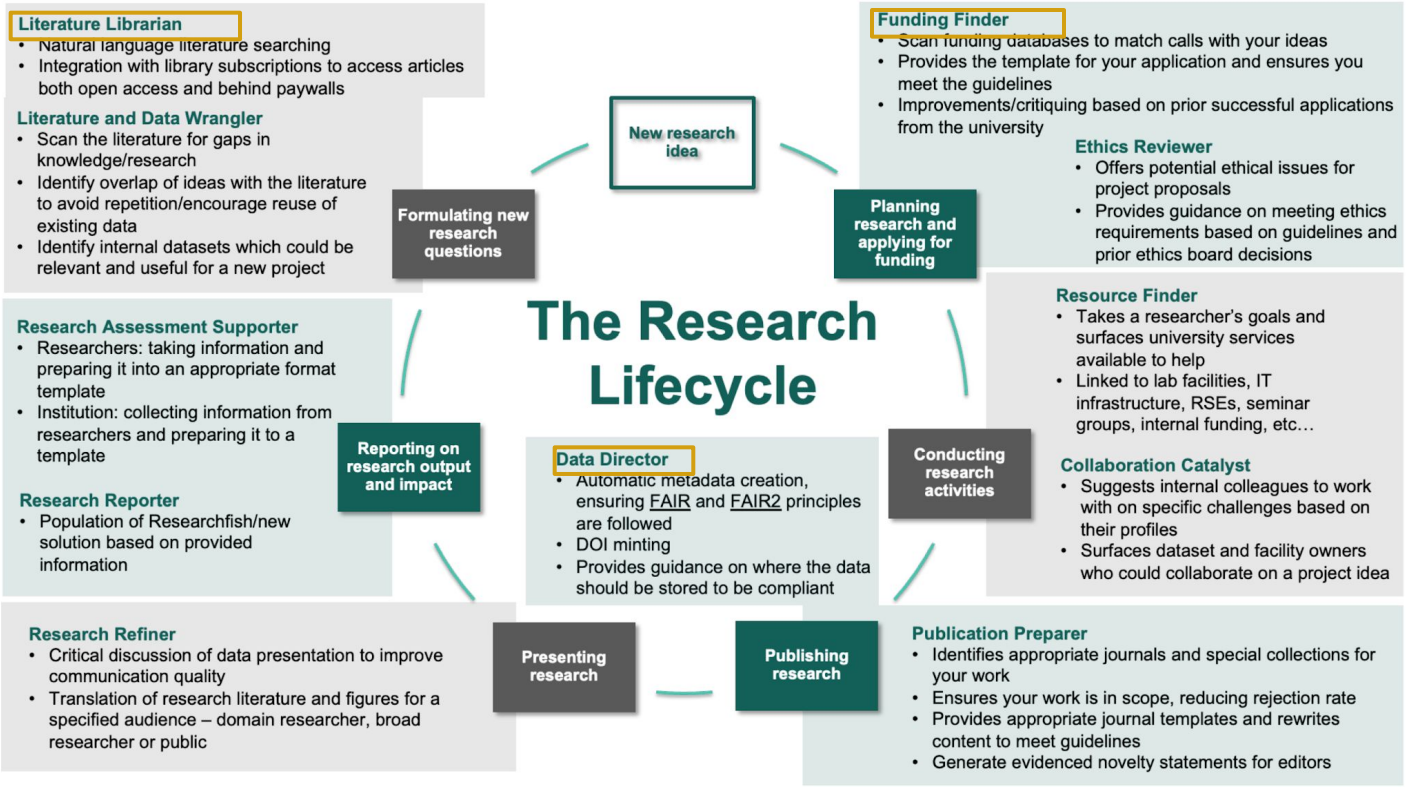
Data Director

3.24 / 4

Funding Finder

3.18 / 4

Next: the Data Director blueprint



RDA consultation, n=83; usefulness /4, unsure excluded

Clare et al. (2026) · Figure 1 + Table 1 · doi:10.15497/RDA00151

Data Director:
FAIR expertise in the workflow

Scientists supply domain meaning; the Data Director helps encode it for reuse.

RESEARCHERS PROVIDE THE MEANING

Explain what was measured and how,
including methods, context and limitations.
Review descriptions for scientific fidelity.

THE AGENT HELPS REPRESENT IT

Select standards and vocabularies;
link identifiers and prepare metadata;
check repository requirements.

Repositories, identifier services, metadata validators, workflow systems and shared vocabularies

This helps create the explicit knowledge representation from slide 10:
research objects with context, provenance and links that other agents can use.

Blueprint: capabilities + governance requirements + workflow + reference architecture

Next: a knowledge fabric connects these research objects; an agentic mesh coordinates the agents that use them.

RDA Data Director Blueprint v1.0 (2026) · specification for implementation
Wilkinson et al. (2016) · FAIR Guiding Principles

Connecting agents across distributed research infrastructure

The Data Director prepares research objects that other agents discover and use.

LABS	REPOSITORIES	COMPUTING FACILITIES
Experimental agents	**Data Director**	**Analysis / simulation agents**

AGENTIC MESH · Governed collaboration across institutions

Discovery and interoperability · accountable ownership · explicit authority · traceable actions

FAIR² · Interoperability foundation

FAIR² reuses Web standards:
shared descriptions, links and validation

→

KNOWLEDGE FABRIC

Shared semantics + metagraph:
meaning linked to distributed resources

Assemble minimum viable context for each task; governed agents use it across institutions.

Our research adaptation of Broda's architecture · FAIR² specification
RDA (2026) · Broda & Miller (2026) · Broda & Korpela (2026)

From the ecosystem to the agent

Cyberinfrastructure is the shared computing, data, software, networks and services that support research.

What expertise must an agent acquire and apply to fulfill a scientific role?

Skills encode procedures. Scientific expertise also needs connected knowledge, experience and judgment, supported by a reliable working environment.

From a shared ecosystem to an individual agent

Human goals · constraints · review

HARNESS ENVIRONMENT · permissions · execution control · observability

MODEL · reasons with selected working context

MEMORY · state across time

Working context

Plans, hypotheses, intermediate state / checkpoints

Episodic experience

Actions, decisions, failures, outcomes

Semantic knowledge

Concepts, facts and relationships linked to evidence

Personalized memory

Person / team / environment preferences, constraints

distill →

SKILLS · procedural expertise

Load task-specific instructions, examples and constraints.

PROTOCOLS · structured interaction

Discover capabilities; express actions through defined interfaces.

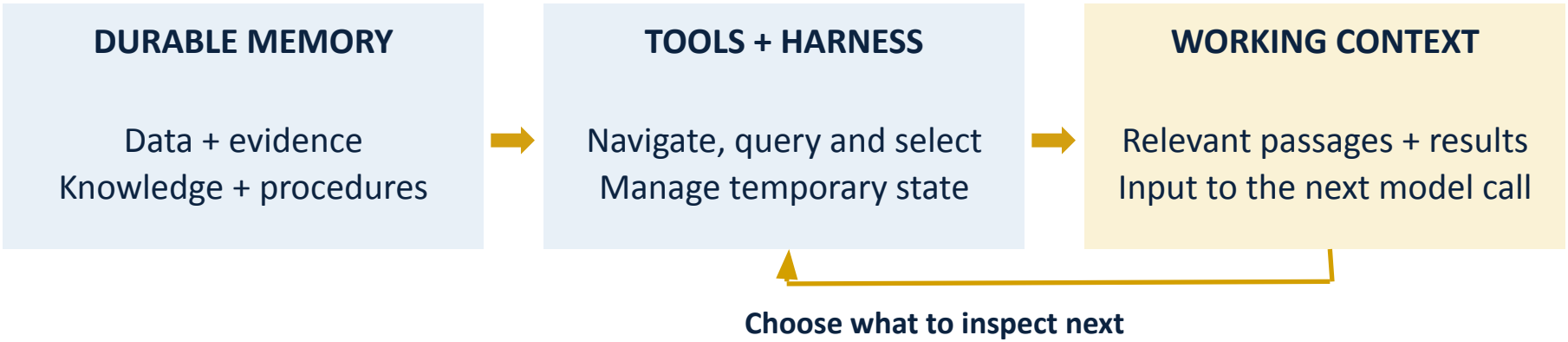
Experience can be distilled into reusable procedures.

FAIR resources describe the data; wiki memory preserves linked knowledge and experience for agents to retrieve.

Zhou et al., Externalization in LLM Agents (2026), §§3.1, 3.3, 6, 7.2 · wiki mapping: our synthesis

Agents actively manage their working context

The model generates its next tokens from the context available now.



Tool observations inform the next selection; reviewed updates persist beyond the task.

Expertise must be retrievable—and the agent must know when to bring it into context.

What must research memory preserve?

These are knowledge roles, not storage locations.

EPISODIC · what happened	Run settings and outcomes Decisions, failures, corrections
SEMANTIC · what we understand	Concepts and relationships Interpretations and open questions
PROCEDURAL · how to proceed	Methods, skills and reusable patterns Applicability, checks and escalation conditions

Link every memory to its evidence, scope and version.
The agent retrieves relevant memories into working context.

Memory roles adapted from the externalization view; research examples are our synthesis

Connecting knowledge for scientific investigation

The current scientific question guides what the agent retrieves from both.

AGENT-MAINTAINED KNOWLEDGE		SHARED SCIENTIFIC RESOURCES
Interpretations and linked explanations Procedures, lessons and reusable patterns Connected to supporting evidence		Databases, ontologies and datasets Literature and supporting evidence Identifiers, relationships and provenance

Read explanations, follow relationships, and retrieve the evidence needed next.

Expertise guides investigation; shared resources supply evidence.
Findings can update maintained knowledge.

Research-memory synthesis · builds on Verborgh, “Let’s talk about pods” (2022)

Wiki memory supports progressive discovery

An agent (claude code,codex) navigates from orientation to the detail its current task requires.

INDEX	TOPIC MAPS + CONCEPTS	INDIVIDUAL NOTES
Where should I look? Topic cues and entry points	How is this topic organized? Related ideas and note links	What detail do I need? Evidence, methods and scope

Readable prose + YAML metadata + explicit note types and typed relationships

Indexes and topic maps guide reading; defined links express support, disagreement and dependency. A defined mapping turns frontmatter into RDF for graph queries.

Keep knowledge outside the prompt;
reveal relevant detail as the investigation unfolds.

Karpathy · LLM Wiki (2026) | Conceptual hierarchy shown: our research-memory design

Keep expertise usable across tasks

People guide and review; agents maintain the knowledge and its retrieval structure.

CAPTURE	CONNECT	REVIEW + VALIDATE	APPLY + REVISE
Record experience Preserve sources	Link rationale State applicability	Check evidence Validate types + links	Evaluate outcomes Revise notes + indexes

PROVENANCE AND CHANGE HISTORY SUPPORT REUSE

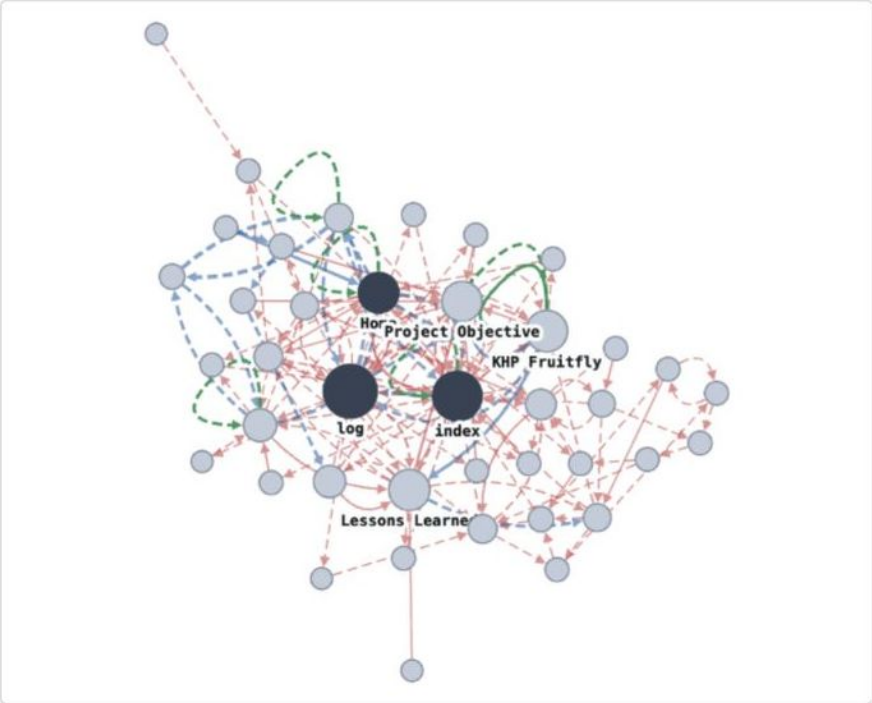
Preserve sources and applicability; inspect structural checks and review changes.
Use experience from the next investigation to improve notes, links and navigation.

Develop with the literature wiki; apply and refine through YARP.

Karpathy · LLM Wiki | Related: Beyond Memory, TrajWiki, WikiSkill

YARP: retrieving expertise behind a workflow

YARP explores reaction pathways and estimates barriers; its wiki connects methods, code and experience.



Why this software version?

Sampling paper → implementation → saved model → environment requirement

The conformer selector was saved with scikit-learn 1.3.0, explaining the cluster's matching software version.

Retrieve the procedure and its rationale

Index and topic pages guide retrieval. Linked notes connect scientific methods, model artifacts and operational lessons.

Building and consulting the YARP wiki:
red = writes; blue = reads; green = search

YARP: an investigation adds to shared memory

Ericka's question: software errors, or unintended reaction pathways?

COMPARE RUNS	RETRIEVE LESSONS	INTERPRET	UPDATE THE WIKI
Test current and stable versions	Read prior notes on outcome classification	Separate classified outcomes from errors	Link diagnosis, evidence and limits

15 of 22 flagged outcomes were classified unintended pathways; 7 were errors.

New observations joined prior expertise to explain the results.
The diagnosis became linked notes and lessons for the next investigation.

Sixth paradigm: guided investigation, with expertise that accumulates.

YARP collaboration: Chris Sweet and Ericka Miller (Brett Savoie's group)
KHP case · one run per version; xTB only; no memory/no-memory control

Where can this take us?

A Data Director connects research outputs to the context needed for reuse.

RESEARCH OBJECT	RO-Crate: research context and provenance Croissant: ML datasets, structure and semantics
FINDING	Nanopublication: a specific claim linked to evidence and attribution
WIKI CONTEXT	Interpretation, procedures and applicability linked to the research objects
SCIENTIST	Review scientific meaning and authorize publication

SHARED REPRESENTATIONS CONNECT DISTRIBUTED RESOURCES

Choose representations for the task;
assemble relevant context through their links.

Agents connect scientific capabilities

Scientists set goals, assess evidence and guide the investigation.

1 · EMPIRICAL	2 · THEORETICAL	3 · COMPUTATIONAL	4 · DATA-INTENSIVE	5 · AI/ML
Collect observations from instruments and experiments	Develop and test models that explain observations	Apply models with simulation and numerical tools	Curate, integrate, query and analyze accumulated data	Apply learned models for patterns and predictions

AGENTIC HARNESSES · tools, execution, context and oversight

Agents retrieve relevant evidence, methods and wiki experience for each task.
Scientists review traceable results; lessons inform the next investigation.

Proposed sixth paradigm:
coordinate these capabilities in an ongoing investigation.

Applied research roadmap · shared specifications, implementation and validation

Build a decentralized research ecosystem

LA3D CRC CENTER FOR RESEARCH COMPUTING

People guide agents across independently managed research resources.

LAB HARNESS	ANALYSIS HARNESS	DATA HARNESS
Experiments + instruments Tools + working context	Models + computation Tools + research memory	Discovery + stewardship Tools + shared resources

AGENTIC MESH · coordinates agents' use of resources under shared governance

KNOWLEDGE FABRIC · shared semantics + resource links ⟶ task-specific context

Metagraph: links across sites; resources remain with their providers.

Next: FAIR exchange, agent retrieval tools and SAI fellows' wiki-memory training.

Reusable expertise, traceable research,
and governed collaboration—for people and agents.

UNIVERSITY OF NOTRE DAME

Scientific Artificial Intelligence (SAI) Initiative



Thank you!

https://rb.gy/gcseh2

UNIVERSITY OF NOTRE DAME



Into the Weeds

Harnesses, protocols, and shared memory

YARP: evaluating the infrastructure in use

Applied research: assess the working system and prioritize improvements.

RESEARCH TASK	EVIDENCE	OPERATIONS	IMPROVEMENT
Complete workflows Recover prior work	Trace claims Inspect sources	Retrieve context Maintain memory	Identify failures Measure cost

NEXT: CHRIS'S ADDITIONAL YARP EVALUATION

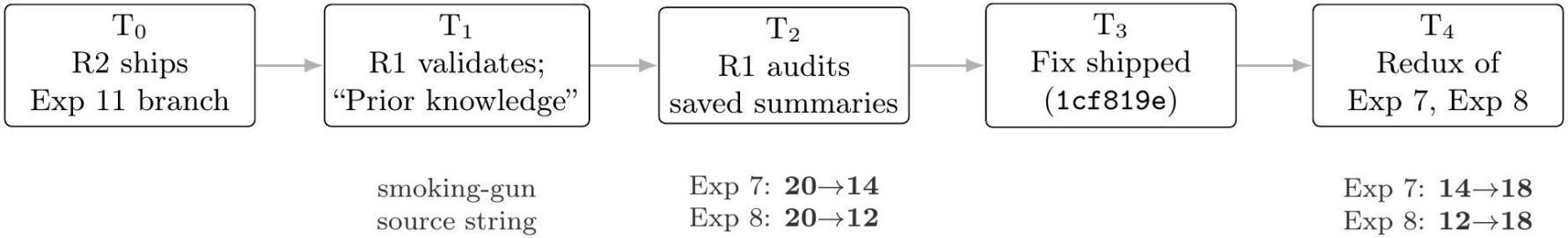
Results forthcoming: report actual tasks, configurations and measured outcomes.
Use findings to refine retrieval, maintenance and the research workflow.

Evaluation informs what we build and improve next.

Project F: audit and correction in adaptive extraction

Separate July case: saved records exposed hedges counted as successful answers

Original Figure 3 · saved records → audit → process fix → rerun



Original Table 1 · claims, audited records, and new runs with the fix

Experiment	Original claim	Audited (evidence-only)	Re-run with fix
Exp 7 (first manufacturer, deepen)	20/20	**14/20**	**18/20**
Exp 8 (second manufacturer, cross-entity)	20/20	**12/20**	**18/20**

Next: a separate research-project retrieval evaluation.

Beyond Memory v2 · §5.2, p.11 · Two-researcher case study

Graph-retrieval project: evaluate the configured system

Separate internal research project · 40 questions · one run · August presentation

Original August table · correct = rubric score 2 or 3

APPROACH	CORRECT	OVERALL (0-3)	FABRICATED
Model alone, no context	2%	0.10	2
Plain RAG over the sources	42%	1.32	10
llm-wiki read path	**90%**	**2.65**	**0**

Tasks: facts, methods, judgments, cross-source synthesis

Model alone — No project context

Plain RAG — ~200 artifacts: code, logs, results
Single-pass retrieval; fixed context budget

Wiki reader — 48 curated pages; multi-step navigation with synthesized project reasoning

Read the result with its limits

Content and navigation both differ. The raw-file navigating control was not scored.

Same-family judge; API interruption changed wiki scoring to in-session. Four factoid questions excluded.

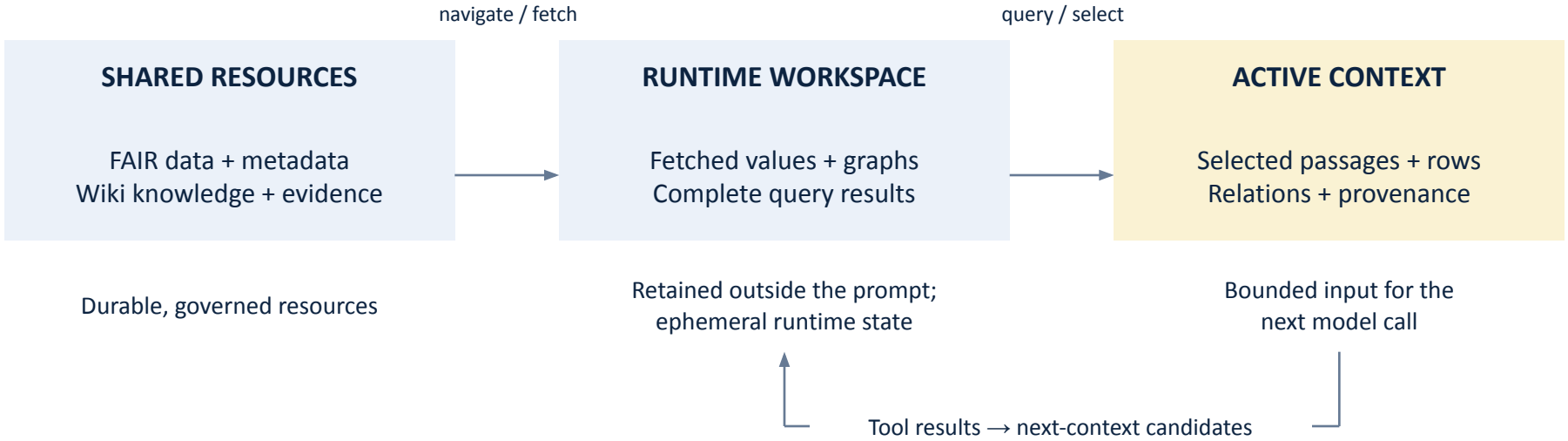
No repeated-run uncertainty. Exact model, prompts and retrieval settings are not publicly specified. Zero fabrications observed in 40.

Project-research tasks; content and navigation both differ across conditions.

Original August table · protocol and limitations: Chris Sweet, July 21, 2026

From shared data and memory to working context

Durable shared knowledge and temporary working context serve different roles.



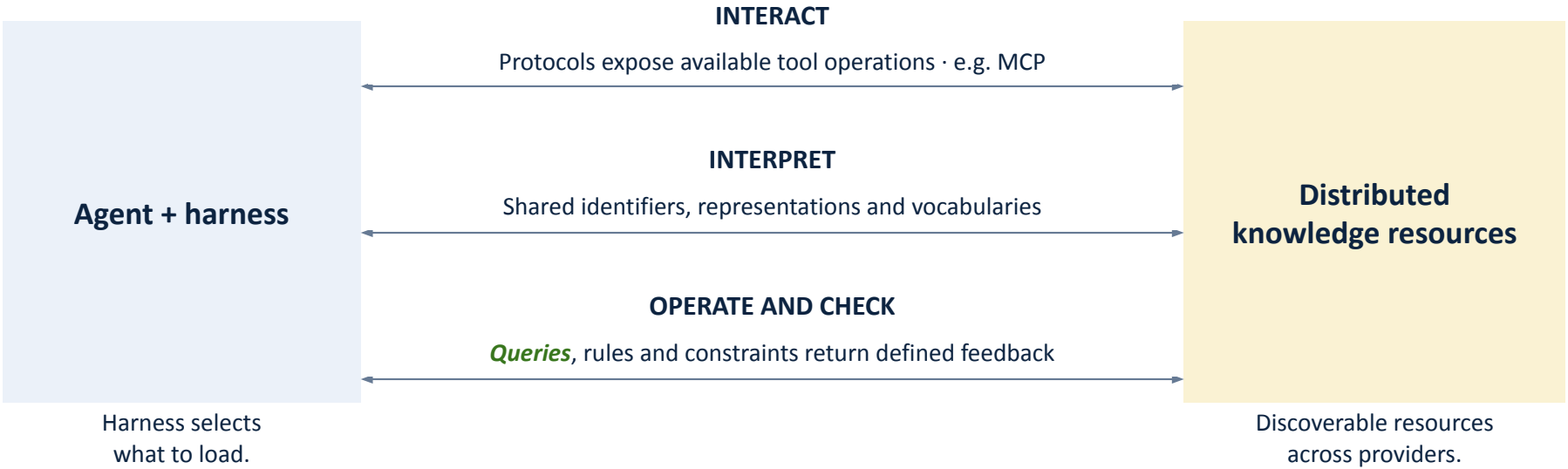
Progressive disclosure: index → relevant pages → supporting evidence.
Skills, rules and predictable page structure guide navigation.

Wiki memory connects interpretation to evidence so the next task can continue the work.

Anthropic, Context engineering · Zhou et al. · Linked Science REPL architecture (2026)

Shared standards make symbolic tools usable across systems.

Interact with resources, interpret their meaning, and check defined operations.



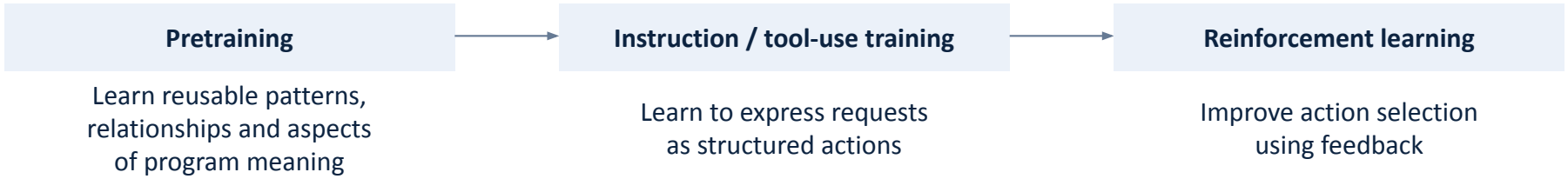
Semantic Web: web standards for linked, machine-actionable knowledge.
Examples: HTTP access · RDF / JSON-LD representations · shared vocabularies.

Learned familiarity helps the model get started; published definitions and tool feedback ground its actions.

Modern architectural extension · Ranaldi et al. (2025), arXiv:2505.15501 · MCP · W3C

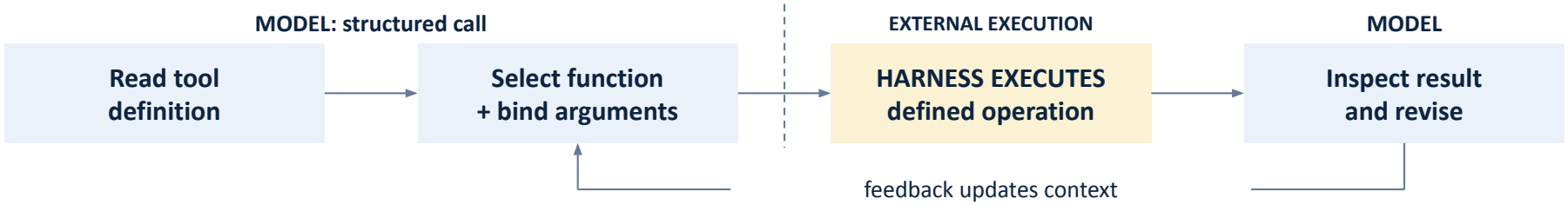
From predicting tokens to using symbolic tools

HOW CAPABILITY DEVELOPS



Simplified training pathway, not a universal recipe; RL trains the model's policy in a tool environment.

HOW THE AGENT ACTS



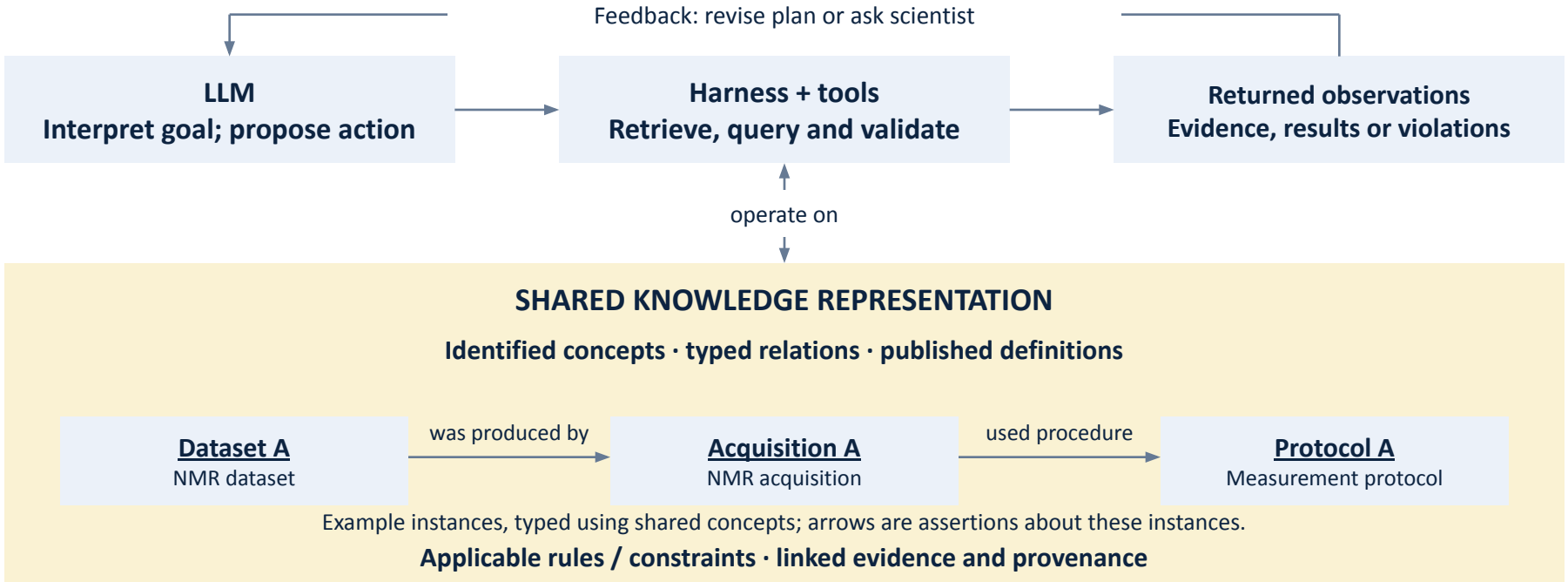
Shared identifiers, vocabularies and constraints connect actions to scientific resources and requirement checks.

Training helps the model choose and express actions; the harness connects those expressions to operations and feedback.

ULMFiT · program-meaning / binding studies · Toolformer · ToolRL · synthesis; sources in notes

A neurosymbolic agent prepares an NMR dataset

Task: prepare metadata for a selected repository; scientist retains judgment.



Shared meaning, explicit rules, and linked evidence make actions grounded, inspectable, checkable, and reproducible.

Illustrative workflow · Ranaldi et al. (2025) · RDA Data Director Blueprint v1.0 (2026)

Agents construct and maintain memory

Public plugin + attached skills + templates + typed links + correction

Sources	wiki-source	Integrate	Verify	Correct
Papers, notes and records	Ingest sources; integrate claims	Typed YAML + body links, index / log	Evidence, scope, YAML and links	Review findings; curate and recheck

Correct failures and re-run the verification gate

CONCEPT PAGE · semantic knowledge

```
---
type: concept
up: "[[Memory]]"
---
# Selective retrieval
Overview first; relevant detail on demand.
See also: [Evidence](Evidence-A)
```

wiki-init attaches the repo's GitHub wiki.
Enabled SessionStart: guidance, index, log.
Skills / rules / templates guide checked writes.

wiki-experiment + wiki-lint

Experiments preserve episodic experience.
wiki-lint checks links, claims and types;
review findings, repair and recheck.

Skills, templates and checks maintain structure for progressive disclosure.

The harness makes memory views actionable

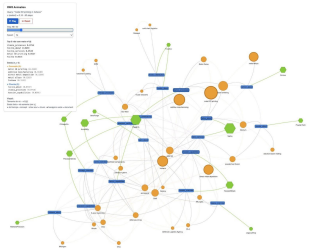
Question: Which method produced this dataset, and what evidence supports it?

EXPLANATORY SEQUENCE · execution depends on the tools exposed by the harness

1 SEARCH	2 QUERY	3 FOLLOW	4 READ	5 SELECT
Find entry points	Select relations	Expand / filter a task subgraph	Inspect supporting pages and sources	Build working context

Existing: wiki skills / file reads + separate graph scripts / SPARQL query library.

WGRA RETRIEVAL RESEARCH · graph demonstration and dense-retrieval results



WGRA · Markov demo

MEASURED · 21,100-CHUNK CORPUS · MUSIQUE TEST

Rank 1

The direct-match gold chunk. Stays at rank 1 at any corpus size.

~1,800

The 4-hop bridge chunk. Median rank of 21,100.

22%

all-gold@30 · vs 88% on a small distractor-free set.

MuSiQue · dense baseline

Structure creates navigational possibilities; skills and tools construct working context.

Chris Sweet et al. · WGRA (2026)

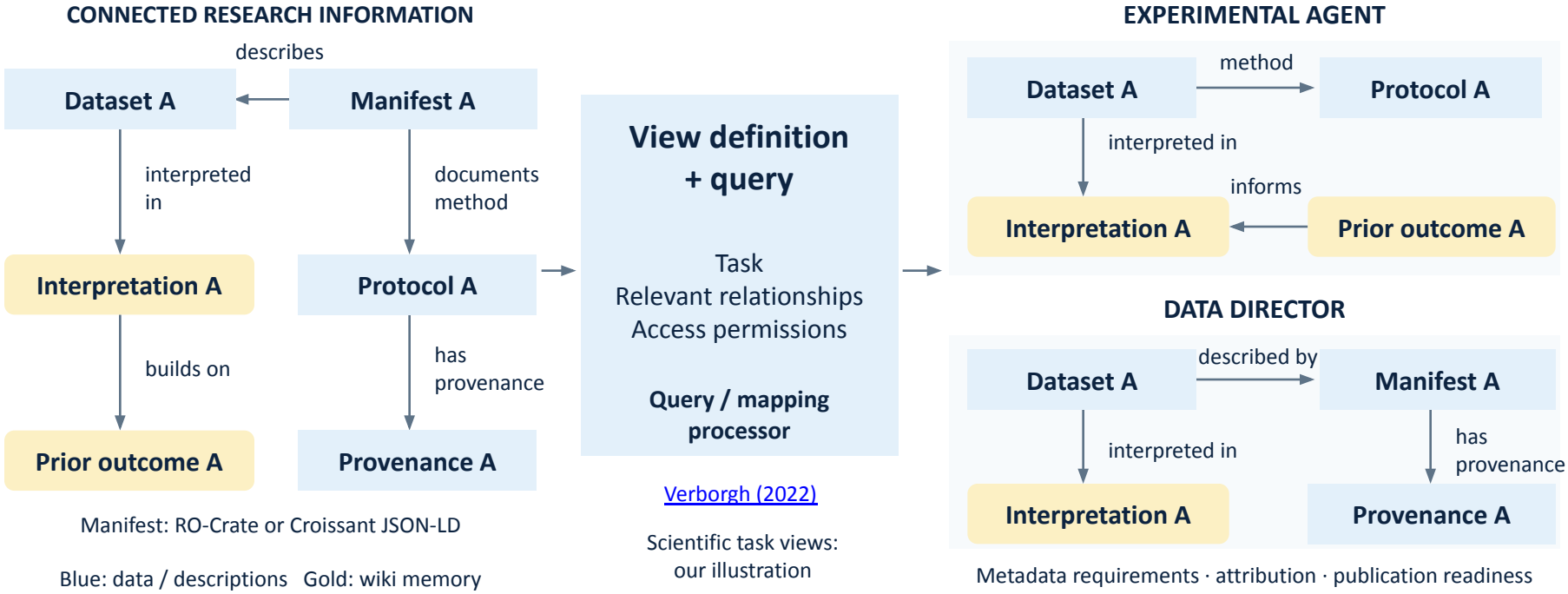
Separate methodological scopes

Random walk + restart:
concept / chunk / document demo.
Dense retrieval: 21,100 chunks;
22% all-gold@30.
No full-scale graph-walk win established by these figures.

Saboia Moreira et al., Beyond Memory (2026)

UNIVERSITY OF NOTRE DAME

Shared data and memory, different views for different tasks.



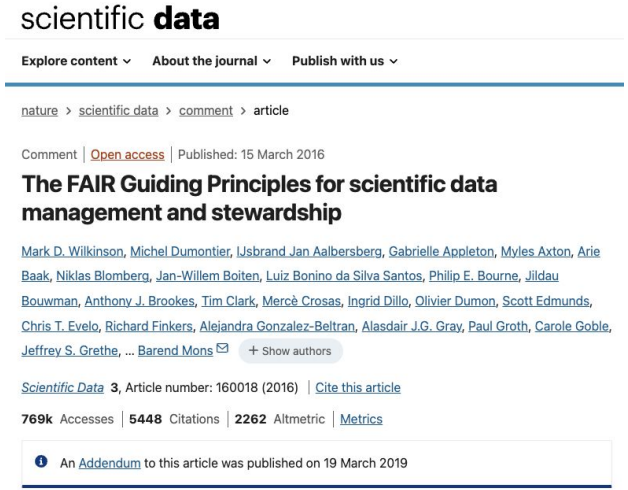
Ontologies define meaning; queries and mappings construct the view; the harness makes it available to the agent.

Implemented: wiki → wiki-graph → JSON-LD / Turtle → materialization + SHACL

llm-wiki-colab wiki-graph

Why FAIR enables scientific agents

FAIR explicitly includes machines as users of research objects.



scientific **data**

Explore content ▾ About the journal ▾ Publish with us ▾

nature › scientific data › comment › article

Comment | Open access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, ... Barend Mons + Show authors

Scientific Data 3, Article number: 160018 (2016) | Cite this article

769k Accesses | 5448 Citations | 2262 Altmetric | Metrics

An Addendum to this article was published on 19 March 2019

Human + machine stakeholders
Wilkinson et al. (2016)

Descriptions + persistent IDs	→	Discover
Standard access mechanisms	→	Retrieve
Vocabularies + qualified links	→	Interpret / connect
Provenance + reuse context	→	Evaluate fitness

RDA blueprint: select repositories against requirements; use domain vocabularies; validate metadata with existing tools and researcher review.

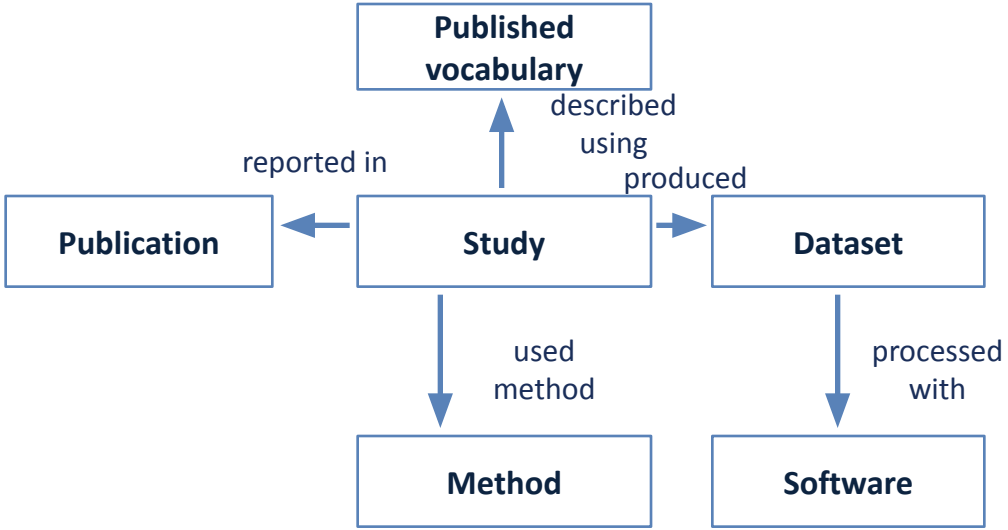
Supports reusable, interoperable agency; does not guarantee scientific correctness.

Wilkinson et al. (2016) · FAIR Guiding Principles

RDA Blueprint v1.0 (2026), R1–R4

Linked Science → FAIR
Interoperable research objects

Shared representations, FAIR vocabularies and qualified references



Illustrative research graph · our illustration, not an original ISA figure

I1 · SHARED REPRESENTATION

Formal, accessible and broadly applicable knowledge representation.

I2 · FAIR VOCABULARIES

Use vocabularies that themselves follow FAIR principles.

I3 · QUALIFIED REFERENCES

Qualify references to other data and metadata with explicit relations.

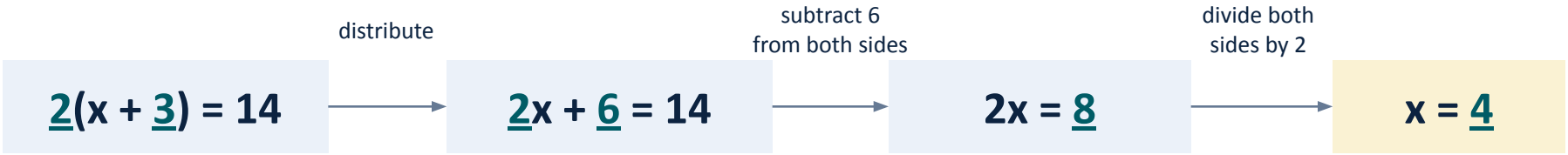
ISA evidence: structured metadata extended into RDF and linked-data publication.
RDF / linked data are implementation examples, not FAIR mandates.

These representations connect scientific knowledge.
How does an agent reason with them?

Wilkinson et al. (2016) · RDA Blueprint v1.0 (2026)

Why does an LLM need a symbolic view?

Consider why we put mathematics on paper.



Defined symbols and relationships; permitted transformations

Inspired by Peirce's account of diagrammatic reasoning

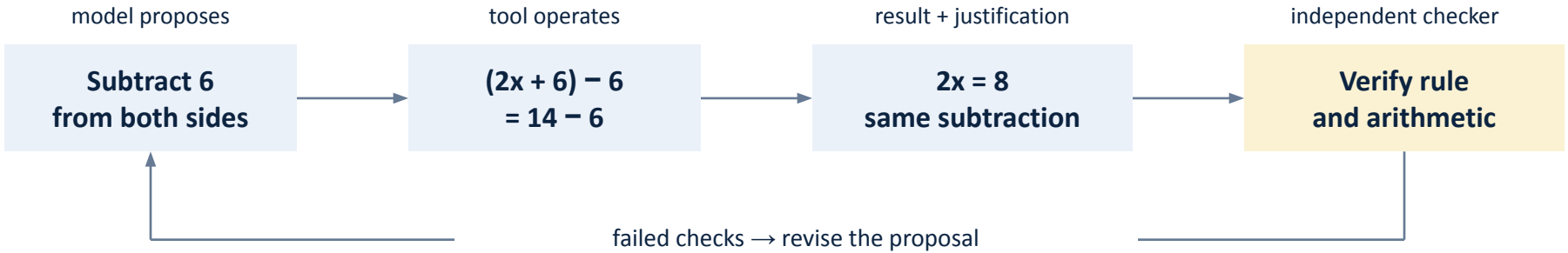
Defined concepts and relationships · Explicit rules and assumptions · Linked evidence

Underlined terms make the changes visible; named rules justify them.

Peirce, Lowell Lecture 2 (1903) · Stanford Encyclopedia of Philosophy · equation is illustrative

What makes the result trustworthy?

Each step is justified under stated assumptions.



CHECK THIS TRANSFORMATION

For real x, subtracting the same quantity preserves equality.

Check: (2x + 6) − 6 = 2x; 14 − 6 = 8

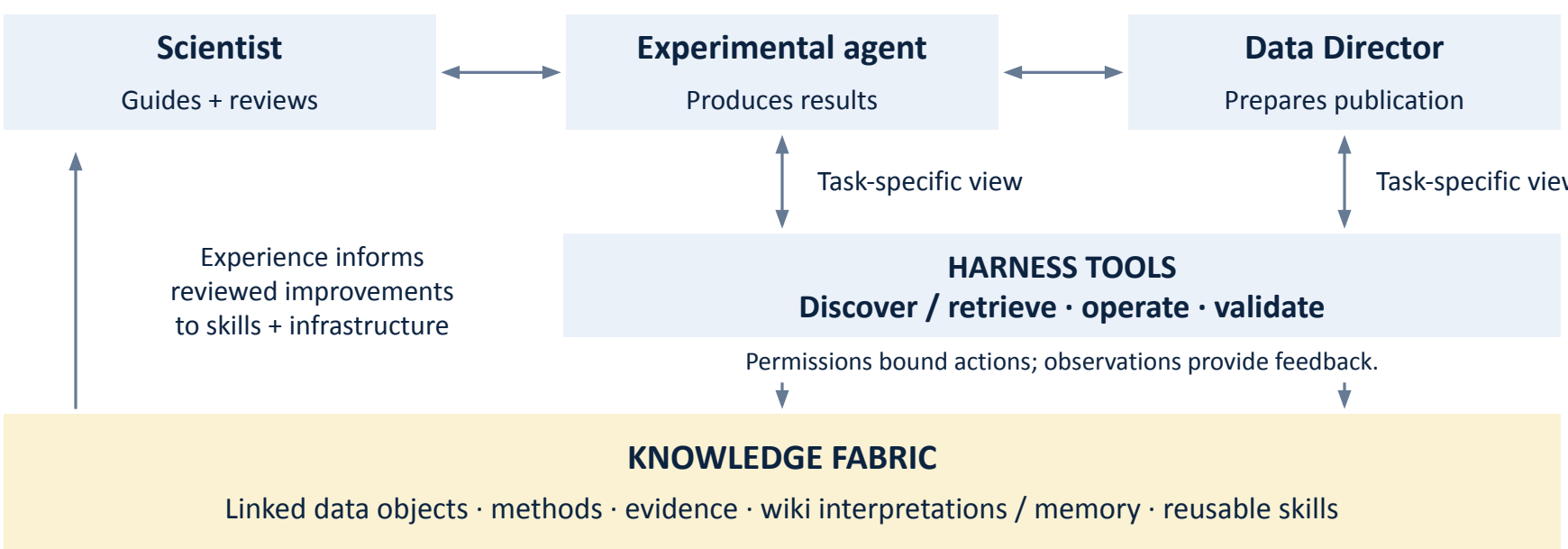
This checks the stated algebraic step; a validation report is not a mathematical proof.

Trust depends on evidence, assumptions, formalization and the checker; conformance does not establish scientific truth.

Independent checking means an explicit test—not generic model self-critique.

Shared infrastructure lets scientific work carry forward.

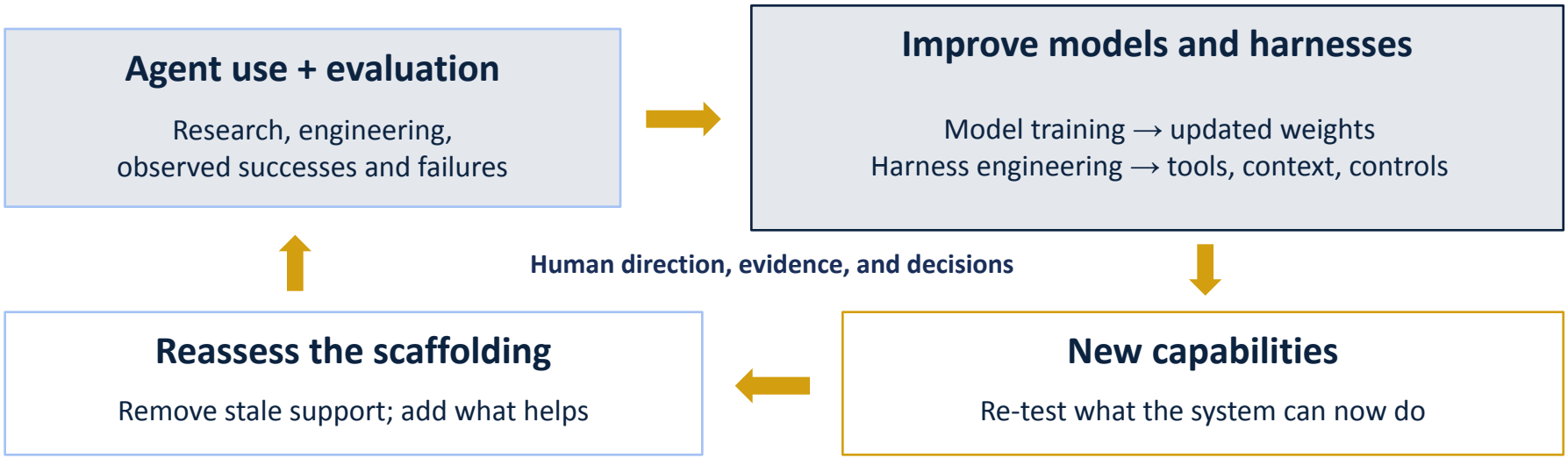
A proposed scientific integration · shared work, meaning and evidence



The mesh coordinates the work; the knowledge fabric connects its meaning and evidence—so the next participant can carry the science forward.

Models and harnesses evolve together

Agent behavior changes as capabilities and the surrounding system improve



Coding-agent foundations increasingly support general-purpose work

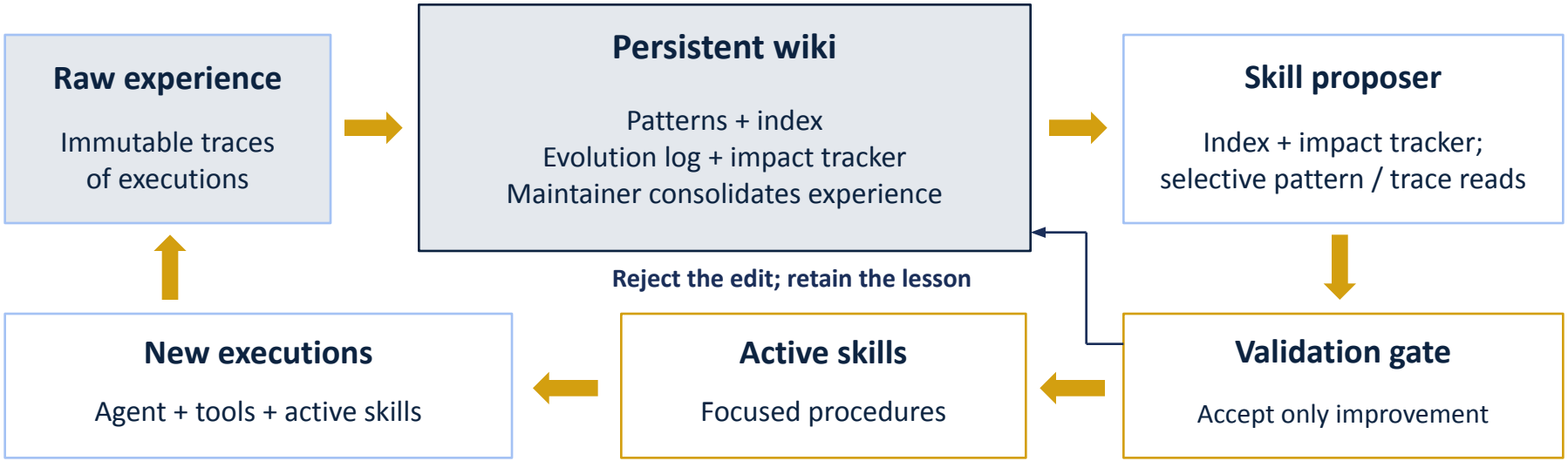
Engineering opportunity: reuse useful components in domain-specific harnesses

As general capabilities improve, scientific teams still supply goals, domain knowledge, evidence, and working history.

OpenAI: Research acceleration (Sep 6); Agents API (Sep 10); Codex for every role (Jun 2), 2026
Anthropic: Harness design for long-running application development (2026) · links in notes

From knowledge to skills

Knowledge accumulates; skills are distilled and validated.



Focused instructions; linked memory retains experience and failed attempts.

In this study: inference gets all active skills injected; it cannot read the wiki.

Record evidence, proposed changes, and validation outcomes

From experiment to publication

Data Director: a community specification for one infrastructure participant

Scientific work upstream — scientist + research harness
Evidence package → publication preparation

Confirm governance, consent, and sharing constraints before proceeding.

1 Choose repository	2 Discover rules	3 Prepare + validate	4 Approve deposit	5 Enrich + reuse
Needs + costs	Schema + terms	Metadata + files	Human decision	Links + updates

Original workflow excerpt: validate metadata before submission (R4)

Validate metadata quality and produce standardised assessments, recommendations and standards

Guidance note:
Metadata quality should be validated before the researcher submits the data, not after it has been deposited.

R4:
Validate metadata quality and produce standardised records assessed against community best practices, recommendations and standards

Repository choice comes before metadata generation.

RDA Data Director Blueprint v1.0 · §§7–8, Figure 5 · 3 Sep 2026

Expertise becomes publication support

Human oversight: consequential approvals and metadata rollback

Orchestration + tools + knowledge + context

Planning
DMP + documentation

Linking
Related outputs + PID-ready links

Metadata
Domain-aligned descriptions

Repository
Destination + requirements

Provenance
Actions + lineage

Validation
Schemas + governance checks

Operational memory
Context + checkpoints

Immutable audit evidence
Actions + outcomes

Identity + permissions + repository integration

Scientific judgment stays with the scientist, outside this publication boundary.

Simplified from RDA Figure 7 + Tables 7–8 · full source in notes

Make requirements executable

Descriptions, profiles, and validation rules make publication requirements usable by tools.

PURPOSE	LINKED STANDARDS	RDA CONNECTION
Find + select	DCAT: catalogs and data services Schema.org: Web discovery metadata	R2–R3, R6
Package + describe	RO-Crate: research package Croissant / FAIR²: dataset descriptions	R2, R4, R6
Interpret + attribute	QUDT: units CRediT / CRO: contributions	R2–R3
Trace + connect	PROV: actions and lineage Links to related research outputs	R10–R11
Discover + check rules	PROF: profiles and their resources SHACL: metadata constraints	R3–R4, R6

FAIR²: AI-reuse dataset descriptions.
CDIF: profiles for cross-domain interoperability.

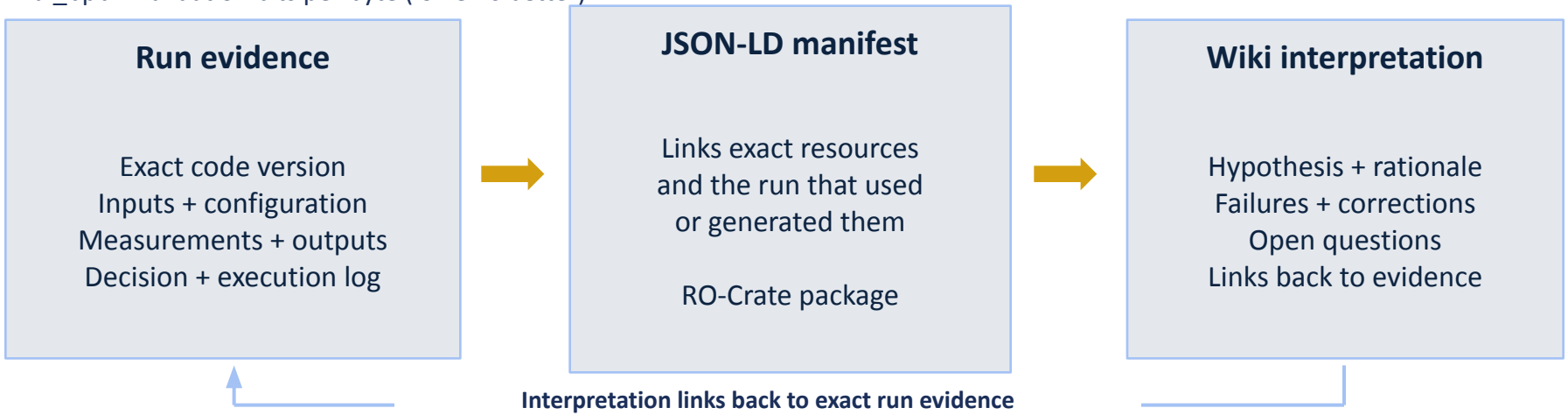
RDA R2–R4, R6, R10–R11 · FAIR², DCAT, RO-Crate, PROF, SHACL · links in notes

Preserve the experiment and its interpretation

Proposed integration: autoresearch outputs become linked research objects

Autoresearch: code change → 5-minute GPU run → evaluate val_bpb → keep / discard

val_bpb = validation bits per byte (lower is better)

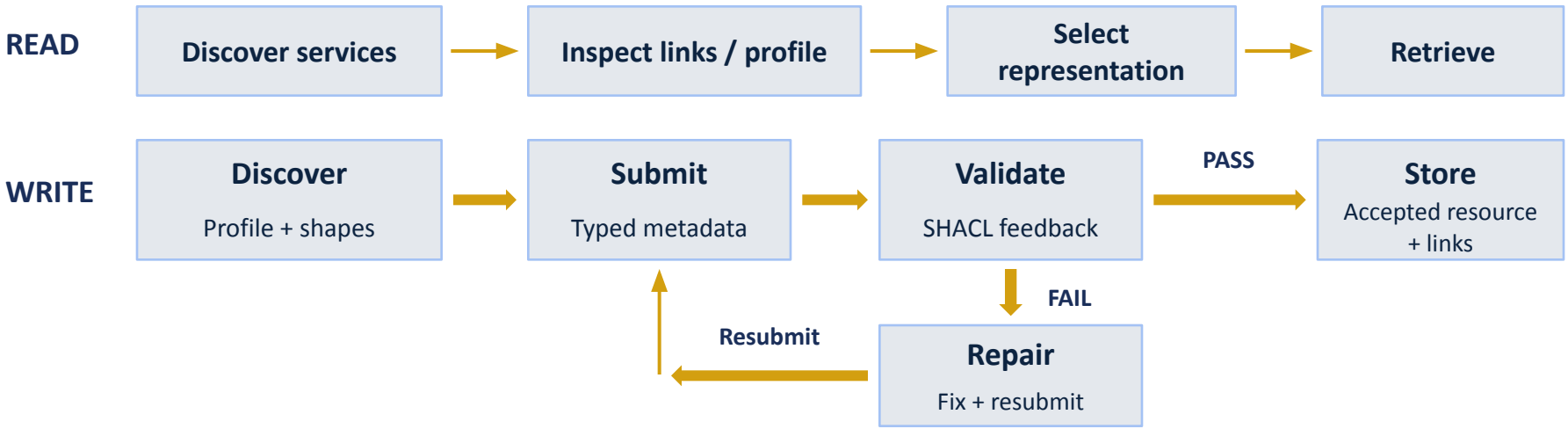


Dataset resources: Croissant / FAIR² descriptions
Different object profiles, common linked resources.
Retain evidence even when a proposed change is rejected.

Karpathy autoresearch · RO-Crate / Workflow Run RO-Crate · proposed extension

An agent can discover how to read and write shared resources

LWS prototype: discoverable representations, application profiles, and checked writes



ONE STORAGE SUBSTRATE · APPLICATION PROFILES AS DATA

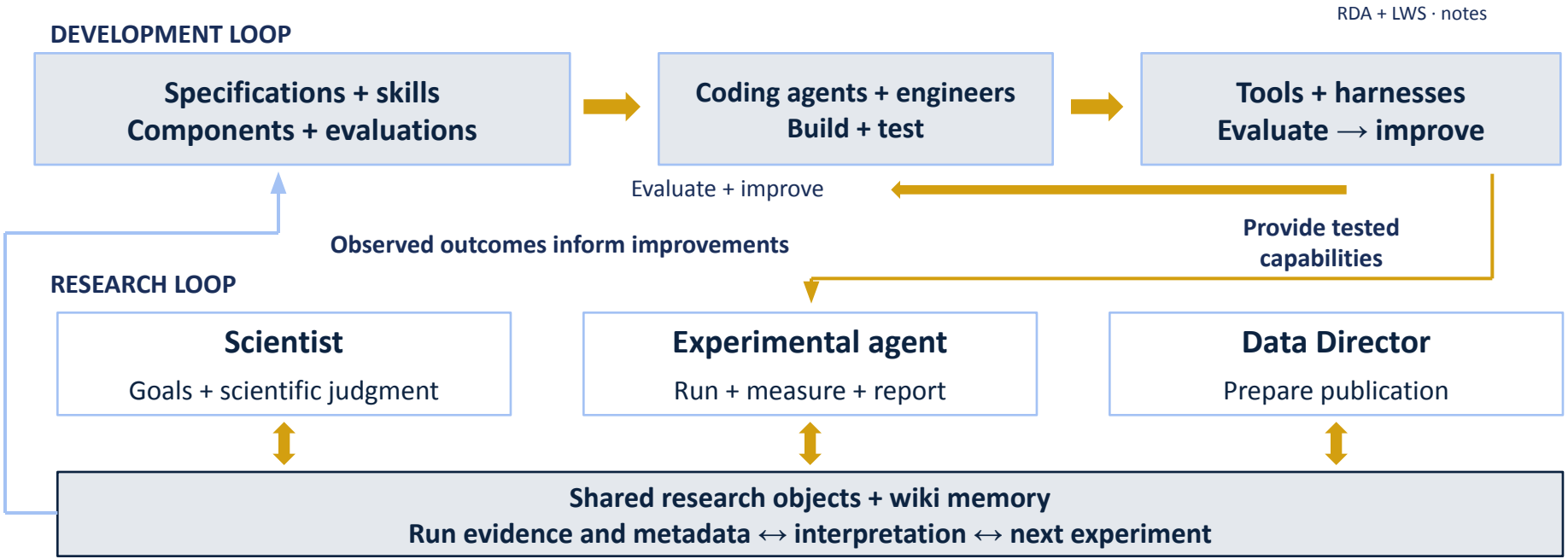
Wiki content	Typed links	Catalog resources
Markdown synthesis	Separate JSON-LD relations	DCAT descriptions

MCP tool interface + HTTP discovery / negotiation + permissions

SHACL checks graph constraints, not prose truth.

Build the environment.
Carry scientific work forward.

Proposed integration: build, use, and improve the research environment



Coding agents help build the components. Shared specifications and resources let people and agents interpret, act, and continue the work.

Can the next collaborator continue from here?

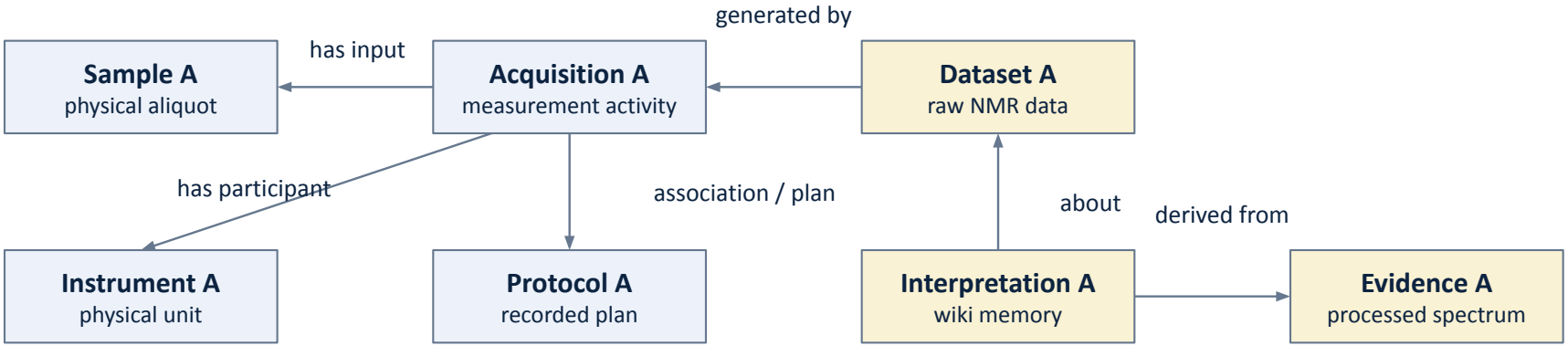
Deep in the Weeds

An illustrated technical walkthrough

Can this NMR dataset support the interpretation?

Can this NMR dataset support the proposed interpretation—and what evidence is missing?

1 / 10 · SHARED GRAPH · NMR = nuclear magnetic resonance



Readable labels map to SIO / PROV / Schema.org relations; protocol is a two-edge path.

MODEL
Interpret the question

→

TOOLS
Query · calculate · validate

→

OBSERVATIONS
Selected tool outputs

Observations are tool returns to the model; Acquisition A is the NMR measurement.

Neural choices operate over explicit resources and executable checks.

PROV-O · SIO · Schema.org Dataset · illustrative linked NMR environment

What must be known before interpreting Dataset A?

2 / 10 · Nominal proton (^{1}H) observation frequency is 80 MHz—not spectral resolution.



Bruker manufactures the model; the physical unit's use here is illustrative.

REPRESENT THE QUANTITY

QUDT numericValue: 80.0
hasUnit: unit:MegaHZ
hasQuantityKind: Frequency
Record nucleus + nominal status.

CONNECT THE RECORD

Dataset A ⟶ Acquisition A
⟶ documented Protocol A.
Pin Bioschemas discovery
requirements and executable checks.

ASSESS SUITABILITY

0.1 ppm × 80 MHz = 8 Hz
At 400 MHz: 40 Hz.
Linewidth, shimming, acquisition
and processing still matter.

qudt = http://qudt.org/schema/qudt/ unit = http://qudt.org/vocab/unit/
quantitykind = http://qudt.org/vocab/quantitykind/ · Frequency is the quantity kind.

Correct fields and units support judgment; they do not establish adequacy.

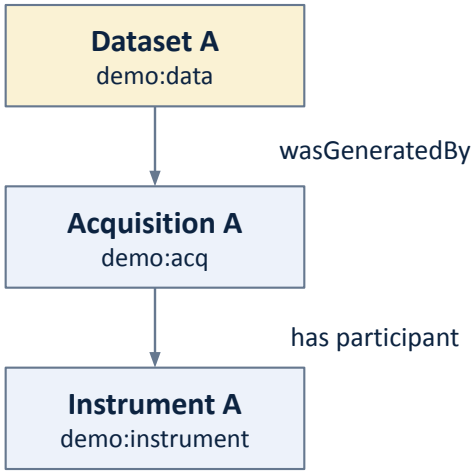
Bruker Fourier 80 Duo · QUDT · IUPAC chemical shift · Bioschemas Dataset 1.0-RELEASE

Inspect the linked record; follow selected resources

3 / 10 · demo = https://example.org/ (unpublished examples)

JSON-LD · Dataset A in the same shared graph

```
{"@context": {
 "demo": "https://example.org/",
 "schema": "https://schema.org/",
 "prov": "http://www.w3.org/ns/prov#",
 "sio": "http://semanticscience.org/resource/"},
 "@id": "demo:data", "@type": "schema:Dataset",
 "prov:wasGeneratedBy": {
  "@id": "demo:acq",
  "sio:SIO_000230": {"@id": "demo:sample"},
  "sio:SIO_000132": {"@id": "demo:instrument"},
  "prov:qualifiedAssociation":
    {"@id": "demo:association"}},
 "schema:subjectOf": {"@id": "demo:memoryIndex"}}
```



Then inspect model, manufacturer, protocol association and wiki links.

Verified ORCID identifies a person; contribution needs separate evidence.

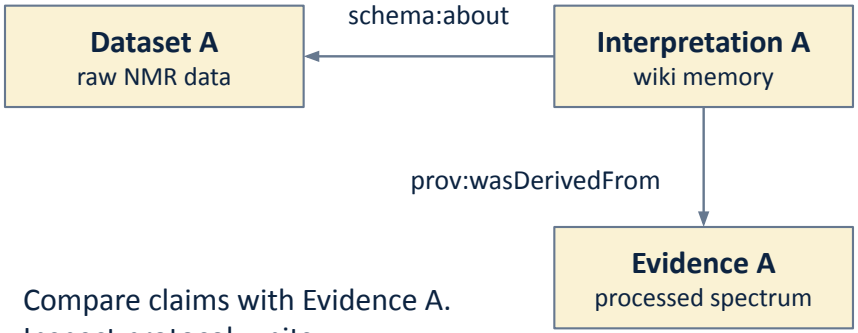
Select descriptions to fetch; persistent IDs need maintained resolution.

JSON-LD · Schema.org · PROV-O · ORCID · w3id and LWS deployment notes

Two tasks select views of the same Dataset A

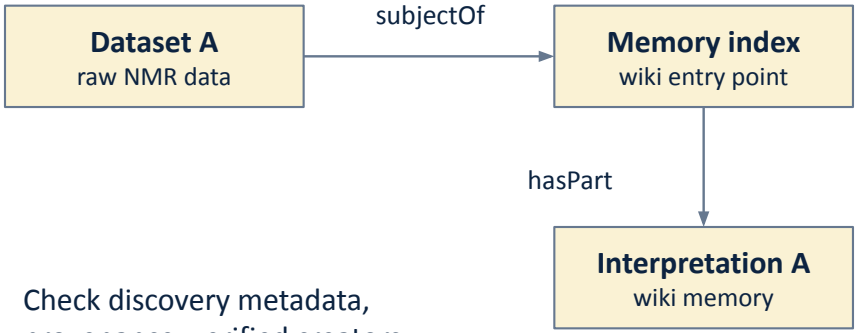
4 / 10 · Shared data + wiki memory; each task retrieves the relationships it needs.

EXPERIMENTAL INTERPRETATION



Compare claims with Evidence A.
Inspect protocol, units
and unresolved alternatives.

DATA DIRECTOR · PUBLICATION



Check discovery metadata,
provenance, verified creators,
access and license requirements.

Pin Bioschemas Dataset 1.0-RELEASE; execute selected publication checks.

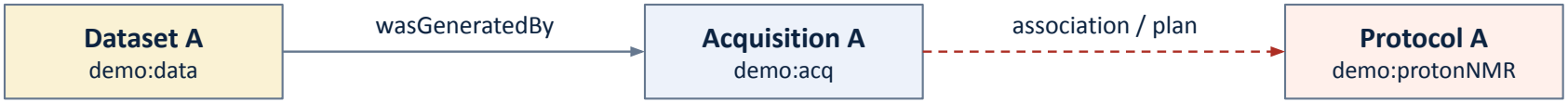
schema = https://schema.org/ prov = http://www.w3.org/ns/prov#

Both views reuse Dataset A and its interpretation; the evidence stays linked.

Bioschemas Dataset profile · Schema.org about / subjectOf / hasPart · PROV-O

A missing protocol link becomes actionable feedback

5 / 10 · Validate Dataset A against one selected requirement.



Missing association → Protocol A link

SHACL · require one documented protocol

```
demo:ProtocolShape a sh:PropertyShape ;
 sh:targetNode demo:data ;
 sh:path (prov:wasGeneratedBy
   prov:qualifiedAssociation prov:hadPlan) ;
 sh:class sio:SIO_001043 ;
 sh:minCount 1 ; sh:maxCount 1 ;
 sh:message "Find protocol; link and recheck." .
```

demo=https://example.org/ prov=http://www.w3.org/ns/prov#
sio=http://semanticscience.org/resource/ sh=http://www.w3.org/ns/shacl#

TOOL OBSERVATION · sh:Violation

Focus: Dataset A
Shape: demo:ProtocolShape
Missing: recorded protocol plan

Retrieve the documented plan.
If unavailable, ask; do not invent it.
Add the supported link; revalidate.

Passing this path check is distinct from proving that the protocol was followed.

SHACL Recommendation · sequence paths · SIO experimental protocol SIO_001043

Derive a relation; keep validation separate

6 / 10 · Derivation needs a recorded dependency—not merely a participant or input.



Processing A constructed Evidence A from Dataset A; preserve that documented provenance.

DERIVED RELATION · materialized shortcut

```
CONSTRUCT {
 ?out prov:wasDerivedFrom ?source
} WHERE {
 ?out prov:qualifiedDerivation ?d .
 ?d a prov:Derivation ;
    prov:entity ?source .
}
```

prov = http://www.w3.org/ns/prov# sh = http://www.w3.org/ns/shacl#

SEPARATE VALIDATION RESULT

Missing Protocol A ⟶ sh:Violation
Documented repair ⟶ conforms

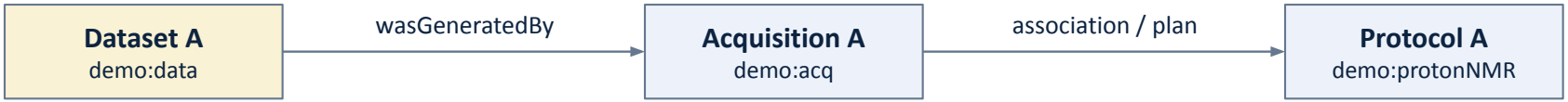
Keep asserted graph, derived triples, report, tool version and evidence. Neither output establishes truth.

One observation adds a relation; another reports an unmet requirement.

PROV-O qualified derivation · SPARQL 1.1 CONSTRUCT · SHACL

Experts and agents maintain the reusable procedure

7 / 10 · Turn the missing Protocol A finding into a reviewed check and skill.



SCIENTIST / DOMAIN EXPERT

Define suitability and evidence.
Approve the protocol requirement.
Correct frequency ≠ resolution.
Review interpretation uncertainty.

CODING AGENT + TOOLS

Reuse SIO / QUDT / PROV.
Pin Bioschemas requirements.
Build shapes, queries and tests.
Check missing / valid / wrong cases.

VERSIONED SKILL

Inspect Dataset A's task view.
Run checks; retrieve missing facts.
Record evidence and feedback.
Recheck before proposing a claim.

Question → requirement → executable test → reviewed release → run record

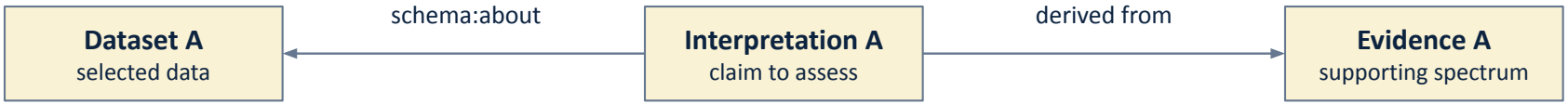
NMR motivation: a reported analysis revised its interpretation after follow-up evidence.

Keep scientific corrections in maintained resources, checks and procedures.

SHACL · Bioschemas · Anthropic NMR case · expert-reviewed construction workflow

The harness assembles evidence for the scientist

8 / 10 · Can Dataset A support Interpretation A—and what evidence is missing?



QUESTION + TASK VIEW

Scientist sets the goal.
Harness selects relationships,
wiki pages, source data
and unresolved requirements.

EXTERNAL OPERATIONS

RLM-style workspace holds data.
Code queries, filters and computes.
Tools validate the protocol path.
Focused model calls inspect text.

SELECTED OBSERVATIONS

Return evidence + missing plan.
Keep 80 MHz in its proper role.
State uncertainty and next checks.
Scientist reviews the conclusion.

External state: graph + files + wiki + tool results. Only selected returns enter model context.

Observations = tool outputs. LWS serves resources; MCP exposes operations.

This is a proposed NMR integration; access to data does not mean it was all read.

Recursive Language Models · predict-rlm · LWS · Model Context Protocol

Save the correction; evaluate the revised skill

9 / 10 · Dataset A's missing protocol and frequency correction become reusable experience.



EXTERNAL ARTIFACT REFINEMENT

WikiSkill: evaluate the candidate; retain or roll back.
GEPA / predict-rlm: optimize strategies and skills.
Use held-out tasks for protocol, units, evidence coverage and calibrated uncertainty.
Model weights can remain fixed.

OPTIONAL WEIGHT TRAINING

Separate training changes model parameters.
Legal-domain RLM result: 29.9% → 63.0% mean rubric-criteria pass rate; 50 held-out synthetic data rooms; fixed sub-agents.
This is not an NMR result.

Next run: versioned skill + retained evidence + explicit checks → evaluate again

Engram combines editable notes, learned memory and raw-document verification.

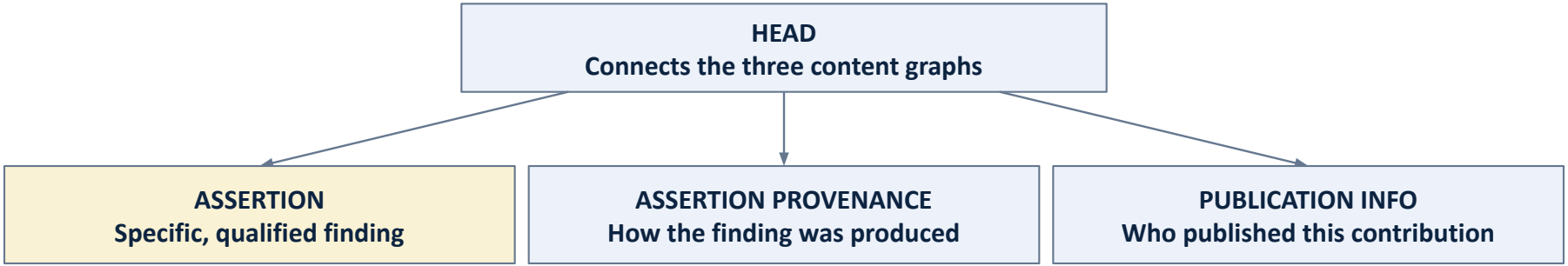
Evaluate the revised skill; retain failures, evidence and prior versions.

WikiSkill · predict-rlm / GEPA · Harvey RLM / Tenet · Engram

From working memory to citable findings

10 / 10 · Publish small, attributable results that later syntheses and papers can build on.

LWS resources + analysis + wiki → selected assertion → checks + scientist authorization



Illustrative claim:
"Interpretation A concerns Dataset A; adequacy remains unresolved without Protocol A."

Dataset A + version
Protocol / analysis procedure
Evidence A + wiki revision
Scope and uncertainty

Verified creator identity
Timestamp + license
Human ORCID is distinct from software identity

FAIR: identity / discovery · access · vocabularies / units · provenance / license / scope

Trusty URI verifies content—not truth or access; w3id provides maintained redirects.

Field study: 15 submissions, 18 authors; journal integration demonstrated.

Nanopub Guidelines · Groth et al. (2010) · Bucur et al. (2023) field study

Knowledge extraction for what?

Graphs and ontologies support views built for a purpose.

PURPOSE / QUESTION shapes extraction, representation and view selection.

Literature + data	→	Extract + curate	→	Shared typed graph
Documents have boundaries.		Agents + ontology/schema Checks + human review		Connected claims + entities Source links + provenance

Ontologies name shared entities and relations; graphs connect evidence across documents.

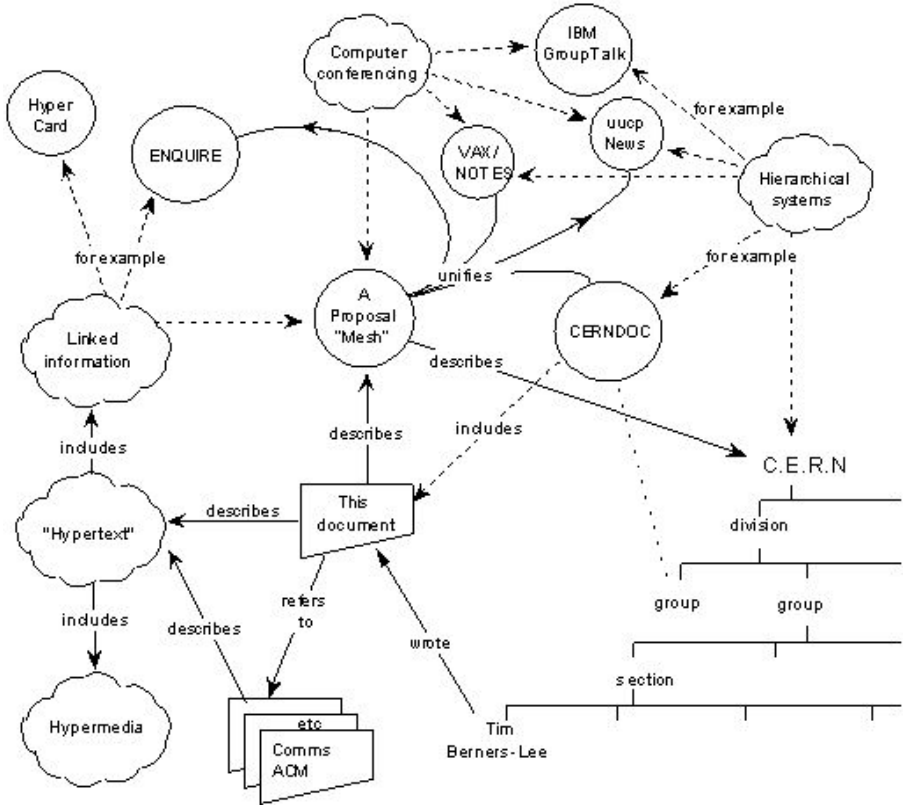
Select overlapping subgraphs for each question—reuse the same identifiers and evidence.

DISEASE MECHANISMS	PROTEIN KNOWLEDGE	MODEL-TRAINING DATA
DisMech · implemented	**UniProt · curation precedent**	**Proposed application**
Which processes connect disease and phenotype? What evidence?	Which function, organism and supporting evidence?	Which examples, features and labels are eligible? How are splits defined?

New consumers reuse shared knowledge without inheriting one application's document organization.

DisMech · UniProtKB · Verborgh (2022) · architectural synthesis; examples are distinct systems

The Web began as a continuity problem



Original Web proposal at CERN
March 1989

People rotated through.
Projects evolved.
Recorded information became hard to find.

Link people, projects, documents and systems as the organization changes.

Tim Berners-Lee, Information Management: A Proposal (1989)

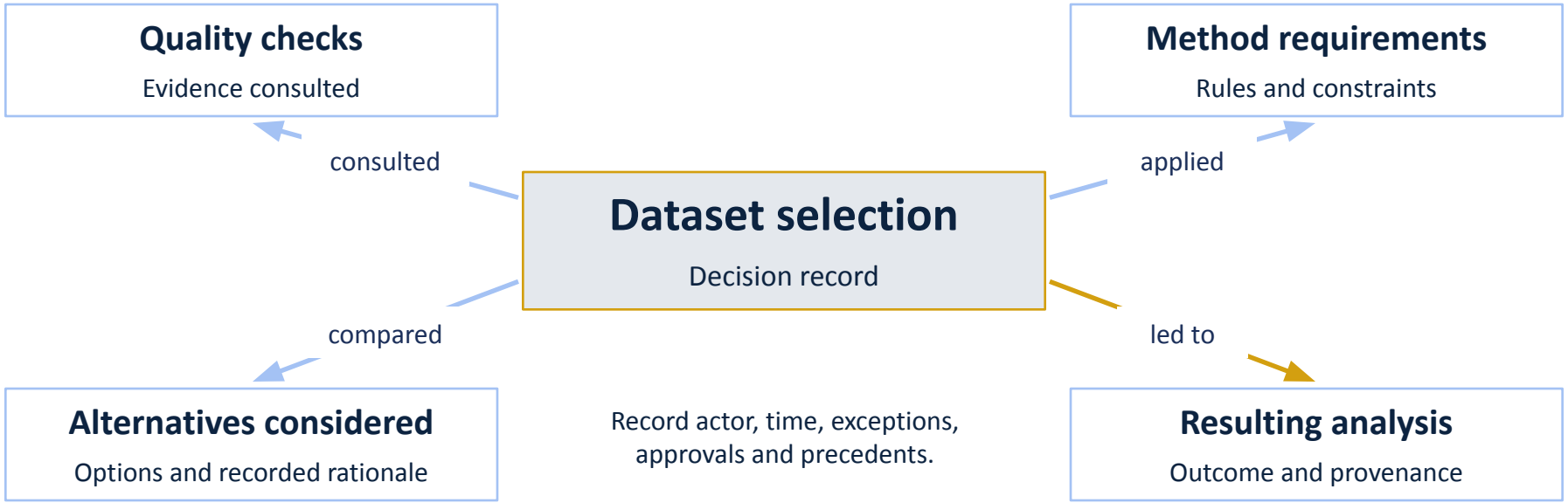


"The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users."

Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284 (5): 34–43. https://www.jstor.org/stable/26059207. PDF: https://www.lassila.org/publications/2001/SciAm.pdf

Context graphs

From working context to recorded decision context



Harnesses use context during execution; linked records retain evidence, constraints, decisions, outcomes and stated rationale.

For future work: stated rationale, not hidden model reasoning.

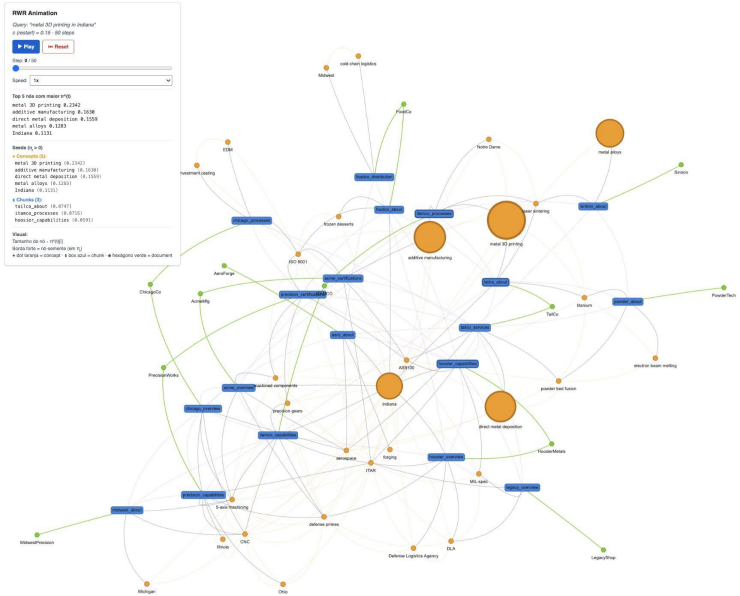
Recording is selective and must be designed.

Scientific example: our synthesis. Gupta & Garg, Foundation Capital (2025)

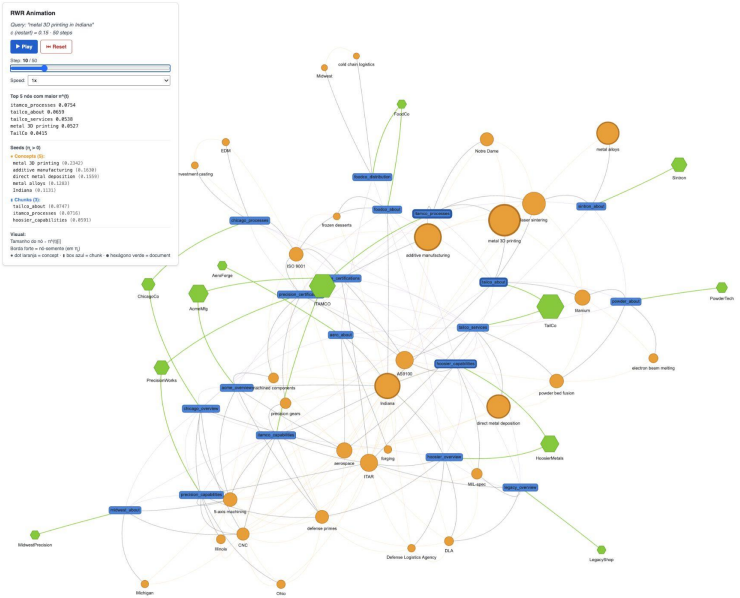
Following connections beyond the first match

Original simulation: seed weight spreads, with repeated returns to the query

Seed state · step 0



Spread + restart · step 10



Concept / chunk / document graph; demonstration, not benchmark results.

Original rendering · Chris Sweet et al., WGRA random-walk demo

Why those connections matter

Original result panel: dense retrieval finds the match but misses bridges

MEASURED · 21,100-CHUNK CORPUS · MUSIQUE TEST

Rank 1

The direct-match gold chunk. Stays at rank 1 at any corpus size.

~1,800

The 4-hop bridge chunk. Median rank of 21,100.

22%

all-gold@30 · vs 88% on a small distractor-free set.

The evidence problem

The first match is easy.
Later bridge evidence can sit far below top-K.

These results motivate structured retrieval.
A graph-walk win at full scale was not yet shown.

Source-reported MuSiQue results; no new benchmark or reanalysis.

Original result panel · Chris Sweet et al., WGRA talk (2026)