



The Research Data Director

A Persona-Based Agentic AI Blueprint and Application for FAIR Research Data, Metadata, and Schema and Crosswalk Workflows



Raymond J. Uzwyshyn, Ph.D. · raymond@ucr.edu · rayuzwyshyn.net
Research and Technology, University of California, Riverside Libraries, USA



RDA 27th Plenary Meeting
London · 6–8 October 2026
RESEARCH DATA ALLIANCE

Six persona-based AI agents, each modeled on an established research-data professional, transform a researcher’s raw dataset into a FAIR-compliant, six-file deposit package, with human judgment and transparency governing consequential decisions.

Organizing agentic AI by professional persona brings deeper intelligence to research-data workflows, making them FAIR by design, legible by construction, and accountable at irreversible steps.

1 Introduction

Research data curation is specialized, labor-intensive work. A FAIR-compliant deposit requires:

- a schema-conformant metadata record;
- persistent identifiers for durable findability and citation;
- an auditable provenance record;
- licensing and documentation governing access, context, and reuse;
- crosswalks to adjacent standards for interdisciplinary discovery.

Few researchers have the time or specialized training to meet these requirements, and few institutions can staff every requisite role. As a result, many datasets are deposited without being findable, interoperable, or reusable, and do not meet funder, data-management, or sharing requirements.

The Research Data Director addresses this gap through a coordinated team of persona-based AI agents that produce a vetted first draft of the complete deposit documentation for human review.

Objectives

- Automate preparation of a FAIR-compliant deposit package for complex data from a single raw dataset.
- Organize agentic capacity by professional persona for legible, auditable, and accountable behavior.
- Treat schema mappings and crosswalks as primary, human-reviewed data objects.
- Preserve human governance through validation checkpoints and bounded agent authority.

3 Results: The Architecture

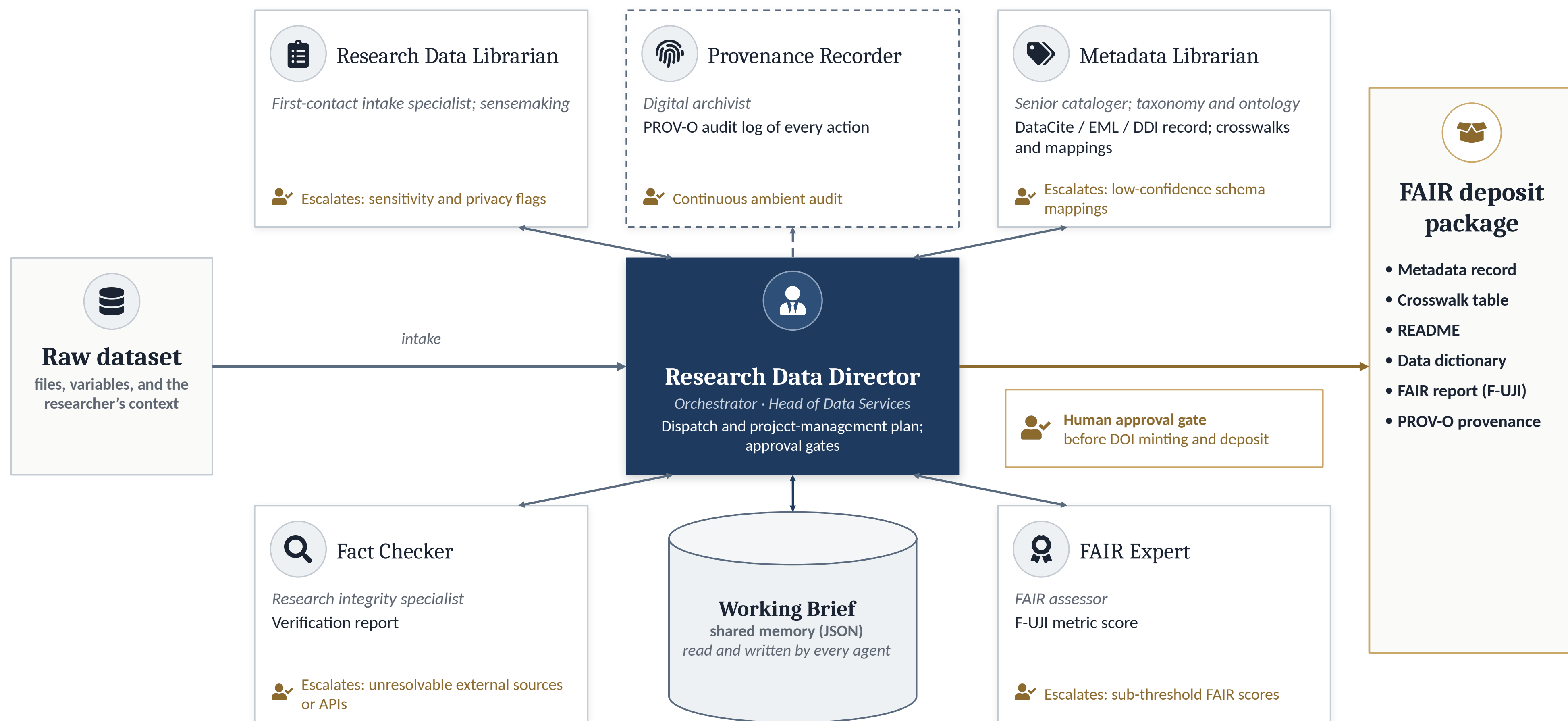


Figure 2. Orchestration, shared memory, and the six personas. The Research Data Director dispatches five specialist agents (orchestrator–workers); the Fact Checker and FAIR Expert evaluate in parallel (evaluator–critique loop); the Provenance Recorder audits continuously. All agents share a single auditable Working Brief (JSON). Each card gives the persona’s professional analogue, principal output, and escalation trigger for human review.

2 Methods

Agentic AI Persona Blueprint. This work creates a persona-based agentic AI architecture, building on the RDA/Microsoft-sponsored Data Director capabilities blueprint and recasting that foundation within a persona framework (see Acknowledgments).

Persona Specification. Each agent is modeled on a professional research-data role within a research institution or library and specified by professional analogue, scope of competence, bounded authority, principal output, escalation trigger for human review, and multi-agent pattern (orchestrator–workers dispatch, evaluator–critique loop, or continuous audit).

Iterative Application Prototyping. The blueprint pairs a conceptual architecture with a buildable implementation (agent specifications, shared memory, human checkpoints, open protocols), realized through twelve prototype iterations in July 2026: deterministic, human-centered interfaces (v1–v4); live large-language-model integration through AI APIs (v5–v9); the full persona model with a transparent deliberation layer, the “Thinking Window” (v10–v11); and experimental voice interaction (v12).

Expert review and extension. Beta prototypes and presentations were shared with the RDA Artificial Intelligence and Data Visitation (AIDV) and VANTAGE Working Groups in July 2026. Working-group feedback prompted a biomedical variant blueprint and prototype (August 2026) and a companion safety and alignment governance framework, ADAM and the AACP (September 2026).

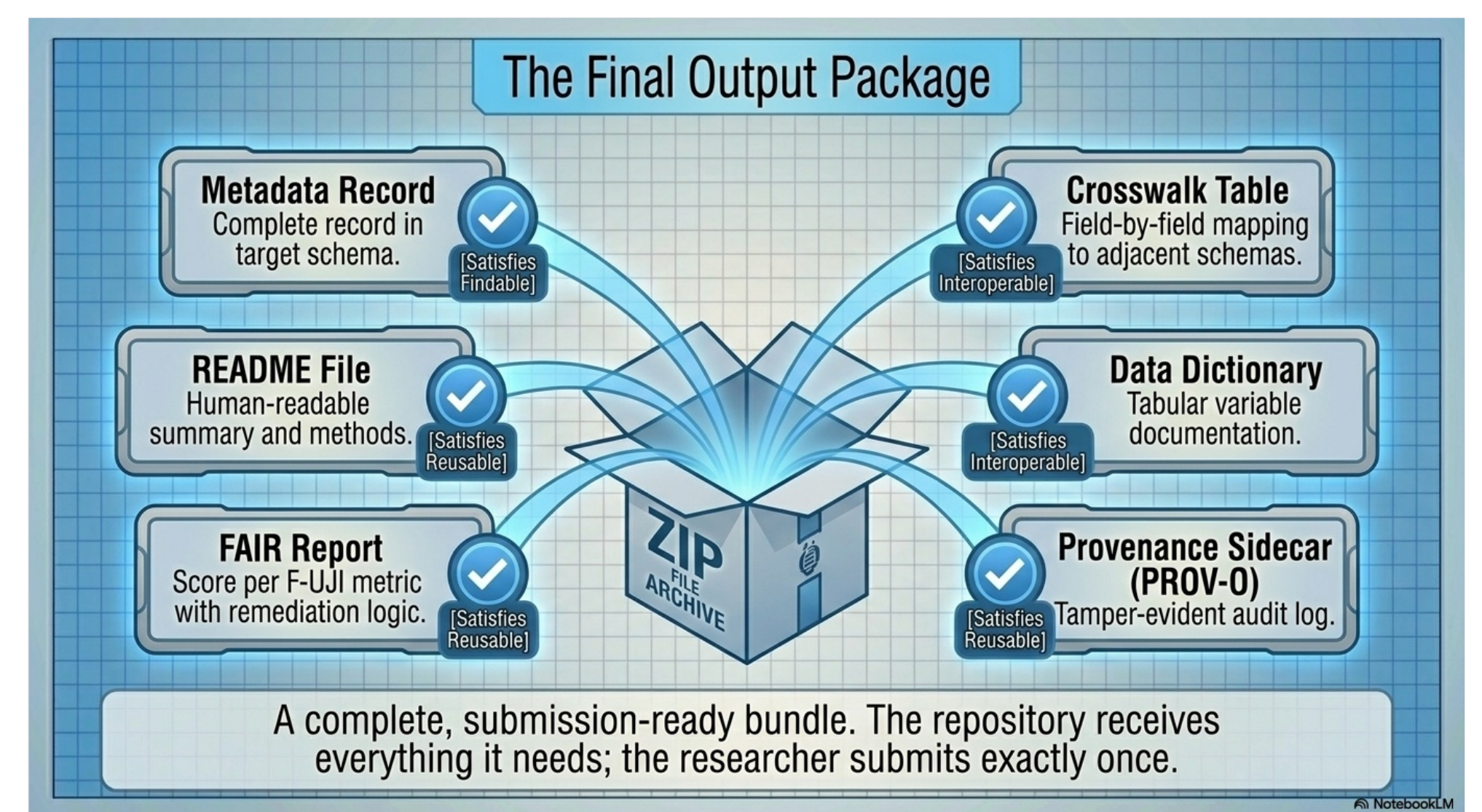


Figure 1. One input, six outputs. A single raw dataset yields six deposit-ready outputs: metadata record, crosswalk table, README, data dictionary, FAIR assessment with F-UJI score, and PROV-O provenance record.

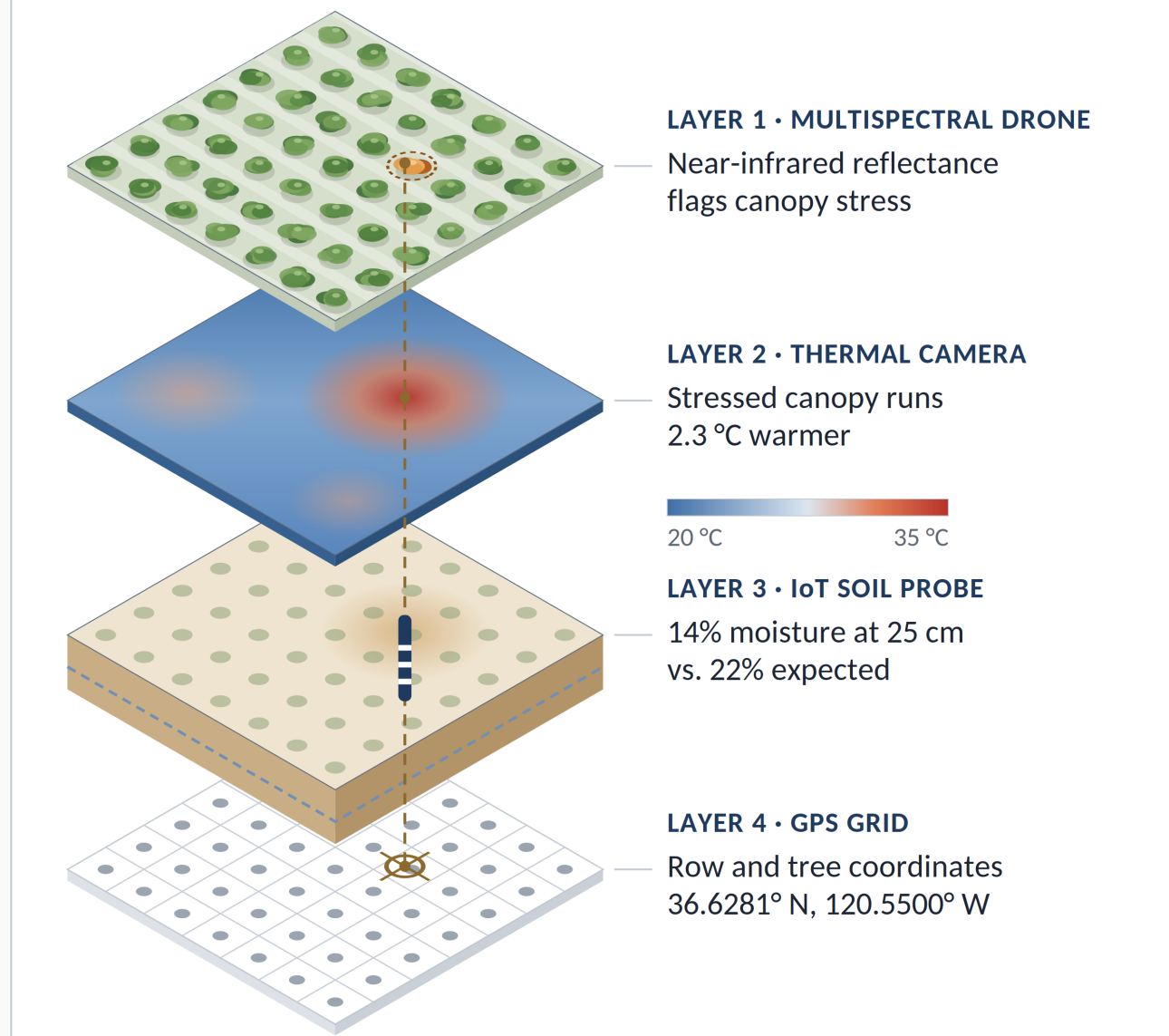


Blueprint documents. Left to right: the RDA/Microsoft Data Director Agentic AI Tool Blueprint (v1.0); the Research Data Director presentation; and the Research Data Director persona blueprint (v0.8.2).

★ Metadata Crosswalks, Mappings, Schemas and Affordances

A The Power of Synthesis

Four instruments, one answer.
A falling irrigation emitter cannot be diagnosed by eye. The diagnosis emerges only when four independent measurements are aligned in space and time.



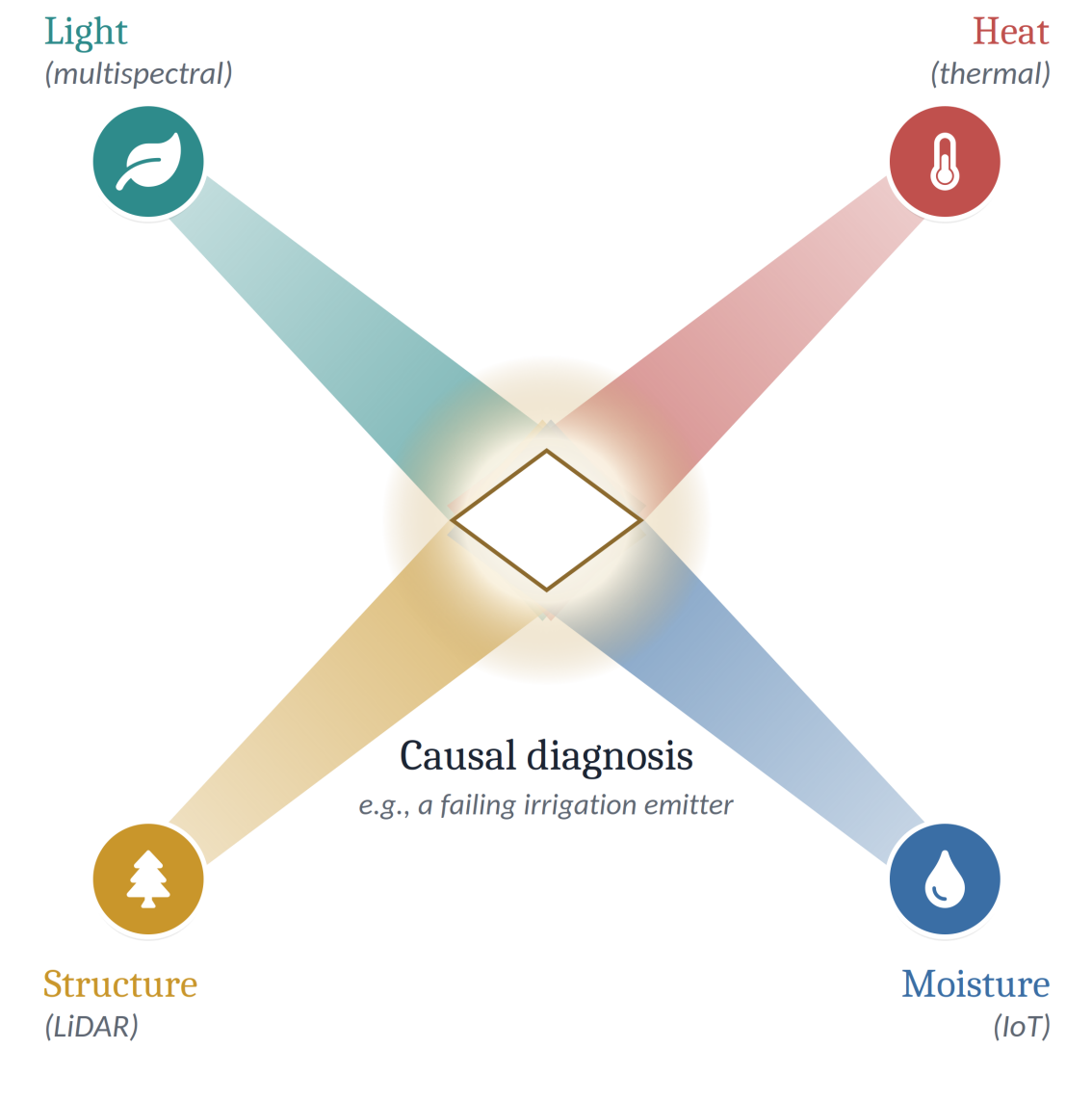
B The Power of Metadata Crosswalks: An Ecosystem of Instruments

Each instrument, like each metadata schema, records part of the object and is blind to the rest. Crosswalks align them, so that what one misses, another records.

Instrument	Measures	Reveals	Characteristic blindness
RGB camera (drone)	Reflected visible light, cm resolution	3D geometry via photogrammetry	Cannot see biochemistry, temperature, or sub-canopy structure
Multispectral (drone)	5–10 narrow bands (visible and NIR)	Plant biochemistry; NDVI	Detects stress but cannot diagnose its specific cause
Thermal (drone)	Emitted infrared radiation	Heat anomalies; transpiration	Cannot distinguish among causes of a temperature difference
LiDAR (drone)	Laser-pulse return time	3D bare-earth structure beneath the canopy	Blind to chemistry, temperature, and spectral composition
Optical / SAR (satellite)	Reflected light; microwave backscatter	Daily global coverage; mm-scale surface motion	Resolution too coarse for analysis of individual features
IoT sensors (ground)	Point data (moisture, temperature)	Continuous subsurface ground truth	Cannot extrapolate spatially without modeling

C The Triangulation Principle

When distinct instruments are aligned in a shared spatial and temporal framework, the relationships among their measurements encode information that resides in no single measurement.



Crosswalks: interoperability, FAIR, and discovery. The Metadata Librarian maps each native record, field by field, to adjacent schemas and assesses confidence for every mapping. Low-confidence mappings are escalated for human adjudication, and the crosswalk is exported as a first-class artifact for discovery. A biomedical variant extends this approach to clinical and imaging standards.

What persona-based design affords

- **Transparency.** The Thinking Window exposes agent deliberation, so outputs can be reviewed as one would review a professional colleague’s work.
- **Contestability.** Researchers may intervene in real time and audit the record.
- **Interdisciplinarity.** Personas can be customized and re-parameterized across domains, from general research data to biomedicine.
- **Discovery.** Crosswalks and mappings enable data to form larger bridges and analogues with adjacent schemas, enriching the possibilities for scientific discovery.
- **Governance and accountability.** Escalation, transparency, and privacy triggers are role-specific and recorded in a PROV-O audit trail.

Figure 3. Crosswalks and affordances. (A) four instruments aligned in space and time; (B) each instrument’s characteristic blindness; (C) triangulation yields what no single measurement holds.

4 Conclusions

- **Mechanistic and Human Interpretability.** Mapping agentic AI onto an established professional division of labor makes multi-agent behavior intelligible to domain specialists: librarians, archivists, curators, researchers, and practitioners. Persona framing also draws on the shared human expertise encoded in model parameters and training data.
- **Interoperability.** Treating crosswalks as governed, human-reviewed outputs addresses the FAIR principle most resistant to reliable automation and widens the scope for discovery and FAIR data.
- **Feasibility.** Working prototypes demonstrate that a complete deposit package can be assembled from a single raw input, with human authority retained over verification, irreversible actions such as DOI minting, and low-confidence determinations.
- **Distinctiveness.** The Research Data Director is the first persona-based Data Director architecture realized in working prototypes. Capabilities specify what a system must do; personas specify who is accountable and who does what.

5 Future Directions (2027–2028)

Deployment contexts

- **Completion of Persona-Based Model:** backend hook-up of various APIs for validation, user testing and feedback, local/proprietary model public use testing, further versioning iteration.
- **Sensitive data (AIDV Working Group):** governed deposit within secure enclaves and processing environments, addressing the ethical, legal, and technical conditions of AI access to protected data.
- **Biomedical research (VANTAGE Working Group):** a Medical Research Data Director variant drawn from the biomedical blueprint, focused on crosswalks across multimodal clinical and imaging data.
- **General research data (RDA Data Director implementation group):** the persona-based Research Data Director, offered as a working basis for a shared RDA reference implementation, general or capabilities-based, with common crosswalks.

Governance, alignment, and transparency: ADAM and the AACP

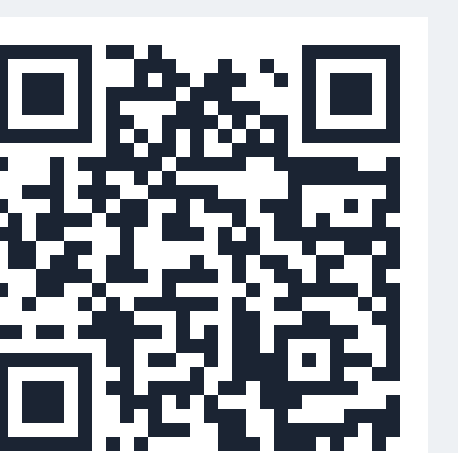
Agentic Directed AI Modeling (ADAM) and the Agentic AI Checkpoint Protocol (AACP) will form the governance layer for persona-based agents, operationalizing AI alignment, safety, research and data ethics, and transparency through tiered human checkpoints, model-and-human verification reports, and validation of DOIs, author identifiers, and provenance, so that growing agentic autonomy is matched by commensurate accountability.

Acknowledgments. With gratitude to Connie Clare (Community Coordinator, Research Data Alliance, UK); Becky Grady and Maria Praetzelis (California Digital Library, US); and AIDV and VANTAGE colleagues Patricia Buendia, Seonyoung Kim, and Francis Crawley for their support of this project.

Project origins. This work originated in, and builds on, the RDA Global Agentic AI Initiative Data Director cohort, sponsored by RDA and Microsoft. Between April and June 2026, the initiative convened some 90 participants worldwide in an eight-week program, consolidating into two regional cohorts of about 16 data librarians, research scientists, metadata specialists, and agentic AI developers, whose community brought together the domain expertise and multidisciplinary perspectives underlying research-data lifecycle workflows. The author continues to participate in a beginning RDA capabilities implementation group.

Selected references

1. Uzwyshyn, R. J. (2026). *The Research Data Director: A Persona-Based Agentic AI Architectural Blueprint for Research Data, Metadata, Schema Creation and Data Repository Publication* (v0.8.2). Research Data Alliance.
2. Uzwyshyn, R. J. (2026). *Research Data Director Agentic AI Persona Appendix: Six Personas* (v0.9). ResearchGate.
3. RDA Global Agentic AI Initiative. (2026). *The Data Director Agentic AI Tool Blueprint*. Research Data Alliance (endorsed September 2026).
4. Uzwyshyn, R. J. (2026). *The Research Data Director, Biomedical Variant: A Persona-Based Agentic AI Architectural Blueprint for Biomedical Research Data, Metadata, Crosswalks, and Governed Publication*. RDA VANTAGE Working Group.
5. Uzwyshyn, R. J. (2026). *ADAM and AI: Agentic Directed AI Modeling and the Agentic AI Checkpoint Protocol. A Framework for AI Alignment, Safety, Ethics and Transparency in Human–AI Research* (v1.0). Research Data Alliance.



Scan for blueprints, working prototypes, video overviews, and presentation materials.