

# Data First: The Case for FAIR-by-Design (Meta)data Standardization in New Approach Methodologies

Position Paper · Damien Huzard, PhD · Neuronautix · 11 May 2026

## Abstract

New Approach Methodologies (NAMs) are producing regulatory-grade evidence, including a human Liver-Chip study reporting 87% sensitivity and 100% specificity for drug-induced liver injury in a blinded benchmark panel [1]. Yet much NAM evidence remains trapped in incompatible formats, weakly structured metadata, and local schemas that block reuse, aggregation, and AI-assisted interpretation. This paper advances three claims. First, FAIR-by-design standardization is no longer a high-friction burden: with current tooling, it is a configuration decision with compounding returns. Second, core infrastructure already exists in practice, especially the New Approach Methodology Ontology and Schema (NAMO), a LinkML-based semantic model designed to describe NAM systems and validation context [2]. Third, the regulatory window is open now: FDA's 2026 draft NAM guidance centers context of use, technical characterization, and fit-for-purpose evidence [3], all of which are metadata-intensive requirements. We conclude with a role-specific call to action for scientists, CROs, sponsors, and regulators.

*Keywords: new approach methodologies; FAIR data; data standardization; NAMO; regulatory toxicology; ontology*

## 1. Introduction: The Pivot Moment

NAM performance is no longer hypothetical. Evidence quality, at least in selected domains, now reaches the threshold for serious regulatory consideration [1]. But evidence value is conditional on data legibility: can datasets be found, compared, pooled, and reviewed without bespoke reconstruction? In most organizations, the answer is still no. Results are captured in local spreadsheets, free-text reports, and ad hoc naming conventions that cannot be queried at scale. The data asset is created and discarded in the same operational motion.

Three trajectories have converged: NAM scientific maturity, rapid gains in AI methods that depend on structured data, and sharper regulatory focus on evidence context and traceability [3]. What has not converged is routine metadata infrastructure in laboratories and sponsor workflows. The key question is no longer whether NAM platforms can generate useful evidence. The key question is whether that evidence can travel.

## **2. The Data Crisis in NAMs**

The dominant friction in NAM programs is often not assay execution but downstream data usability. Reuse fails when endpoint definitions, experimental context, and processing provenance are missing or implicit. Interoperability fails when two laboratories encode nominally identical endpoints under incompatible variable names, units, or ontological assumptions. Historical leverage fails when legacy records remain unstructured and difficult to map into common schemas.

These are not separate failures. They are one failure expressed in three operational domains: no shared data model, no enforceable metadata contract, and no validation gate at ingestion. FAIR principles remain a useful lens because they operationalize these failures as concrete design requirements [4].

## **3. The Regulatory Imperative: Data Standard Before Artifact**

Regulators increasingly evaluate not only assay novelty but evidence package quality. FDA's 2026 draft NAM guidance organizes validation around context of use, biological relevance, technical characterization, and fit-for-purpose conclusions [3]. Each element depends on machine-readable metadata and explicit provenance, not on narrative summaries alone.

At the same time, nonclinical submission standards such as SEND are mandatory in scope but incomplete in coverage for many NAM-specific endpoints [5]. This creates a practical asymmetry: organizations are expected to submit structured evidence while many NAM outputs originate from pipelines that were never designed for structured interchange. In this setting, late conversion is expensive and brittle. FAIR-by-design collection is strategically superior to post-hoc repair.

## **4. NAMO and FAIR-by-Design: The Solution Already Exists**

FAIRification can be retrofitted or designed in. The former is often necessary for legacy corpora, but the latter is cheaper, more reliable, and easier to sustain [6][7]. NAMO provides a practical technical anchor for design-time standardization in NAM domains [2]. It is implemented with LinkML and aligned with established ontologies, enabling export pathways that serve validation, APIs, and semantic interoperability.

The strategic advantage of schema-first design is that it converts recurring harmonization projects into routine validation. Required fields, controlled terms, units, and provenance are enforced once and reused everywhere. This pattern follows the historical arc of successful biomedical ontology adoption, where shared vocabularies turned isolated records into computable, queryable knowledge assets [8][9].

## 5. Implementation: Lower Barrier Than You Think

In operational terms, FAIR-by-design is a setup decision. Teams choose a schema, define required metadata, wire ingestion and validation, and then run normal scientific workflows with stronger downstream reuse. The scientist's daily work does not need to become ontology engineering. The infrastructure absorbs that burden.

The non-trivial step is vocabulary mapping. Local terms for model systems, endpoints, and assay conditions must be aligned to shared terms. That step is increasingly tractable with AI-assisted curation tools such as SPIRES and related ontology-grounded extraction approaches, provided human validation remains mandatory [10]. Legacy remediation is harder but still viable: partial structured recovery of historical datasets creates immediate incremental value, especially where reuse decisions can be audited.

## 6. Call to Action: Stakeholder Responsibilities

**NAM scientists:** treat data and metadata as first-class outputs. Capture context of use, assay conditions, and provenance at source.

**CROs, biotechs, and sponsors:** contract for schema-compliant deliverables and enforce validation gates before data acceptance.

**Regulators:** continue clarifying structure expectations for NAM submissions and support interoperable pathways that reduce avoidable conversion burden.

## 7. Conclusion: Compound Now or Repair Later

NAM science has crossed an important credibility threshold. AI and regulatory review systems are converging toward structured evidence expectations. Organizations that standardize now will accumulate reusable, auditable, AI-ready assets. Organizations that delay will accumulate unstructured liabilities and growing remediation debt. The direction is clear. Timing determines advantage.

## Limitations

This paper is a position argument, not a prospective implementation trial. It does not report measured adoption costs across laboratories or controlled comparisons between design-time and post-hoc FAIRification. Several references in the long bibliography still require field-level verification of author order or edition details before journal submission. Claims about broad transferability beyond the cited contexts should be interpreted with caution.

## Declarations

**Data availability:** No new primary data were generated for this position paper. Sources cited are publicly accessible.

**Ethics:** Not applicable.

**Conflict of interest:** To be finalized for submission venue requirements.

**Funding:** To be finalized.

**AI usage disclosure:** Drafting support included AI-assisted literature organization and manuscript structuring; all claims and citations require final author verification before submission.

## Selected References

- [1] Ewart et al. (2022). Performance assessment and economic analysis of a human Liver-Chip for predictive toxicology. *Communications Medicine*.  
<https://doi.org/10.1038/s43856-022-00209-1>
- [2] Monarch Initiative. NAMO: New Approach Methodology Ontology and Schema.  
<https://monarch-initiative.github.io/namo/>
- [3] U.S. FDA (2026). Draft guidance on NAM use in drug development.  
<https://www.fda.gov/news-events/press-announcements/fda-releases-draft-guidance-alternatives-animal-testing-drug-development>
- [4] Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. <https://doi.org/10.1038/sdata.2016.18>
- [5] CDISC. Standard for Exchange of Nonclinical Data (SEND).  
<https://www.cdisc.org/standards/foundational/send>
- [6] Rocca-Serra et al. (2023). FAIR in action - A flexible framework to guide FAIRification.  
<https://doi.org/10.1038/s41597-023-02167-2>
- [7] Rocca-Serra et al. (2023). The FAIR Cookbook. <https://doi.org/10.1038/s41597-023-02166-3>
- [8] Ashburner et al. (2000). Gene Ontology: Tool for the unification of biology.  
<https://doi.org/10.1038/75556>
- [9] Mungall et al. (2017). The Monarch Initiative. <https://doi.org/10.1093/nar/gkw1128>
- [10] Caufield et al. (2024). SPIRES. <https://doi.org/10.1093/bioinformatics/btae104>