

The Paradox of persistence: Strategies for metadata versioning and FAIR publication in NAPMIX

What is NAPMIX?

Metadata infrastructure for NAP Physics

- NAPMIX is a metadata service architecture built around a **project-hub structure**.
- It connects project-level context with datasets, reactions, projectiles, targets, infrastructures, researchers, and related metadata.
- **Goal:** Capture richer experimental context and make it FAIR, reusable, and exportable.
- **Implementation:** Django + React architecture with a normalized relational schema.

Project is the **umbrella entity**. This is why our strategy is project-rooted rather than dataset-rooted.



Nuclear, Astro, and Particle Metadata Integration for eXperiments

Target Details

Target location: Target density: Density unit:

Sample phase: Thickness value: Thickness unit:

Rear backing: Front backing:

Target Isotopes

Isotope	Percentage	Actions
Tantalum (Ta-181)	100	Details

Target Materials

Material	Percentage	Actions
No target materials added yet.		

Dataset Details

General Researchers Formats Simulations

Name	PID	Start	End
The 48Ca+181Ta reaction: Cross section studies and investigation of neutron-deficient 96Zr93 isotopes (Dataset)	10.5281/zenodo.10959030	15/10/2015	19/09/2025

Description

Example for RDM developments:

Small result datasets given. Under development metadata schema provided

This example dataset describes alpha-decaying chains from an experiment performed at the SHE Physics group at GSI Helmholtzzentrum für Schwerionenforschung GmbH. The reaction $48\text{Ca}+181\text{Ta}$ was used to produce neutron deficient isotopes of Np, Pa and U. The 48Ca beam was delivered by the UNILAC at a variety of selected energies with 5Hz repetition rate and 5ms pulse width. The evaporation residues were implanted into the focal plane detection system, COMPASS, where their alpha-decay signatures were measured. Further details are given in the linked publication.

The data format is in comma-separated values format. Columns are as Energy(keV), Counts per energy bin, decay time (seconds) and counts per decay time bin. The decaying nucleus is given for each chain.

URL

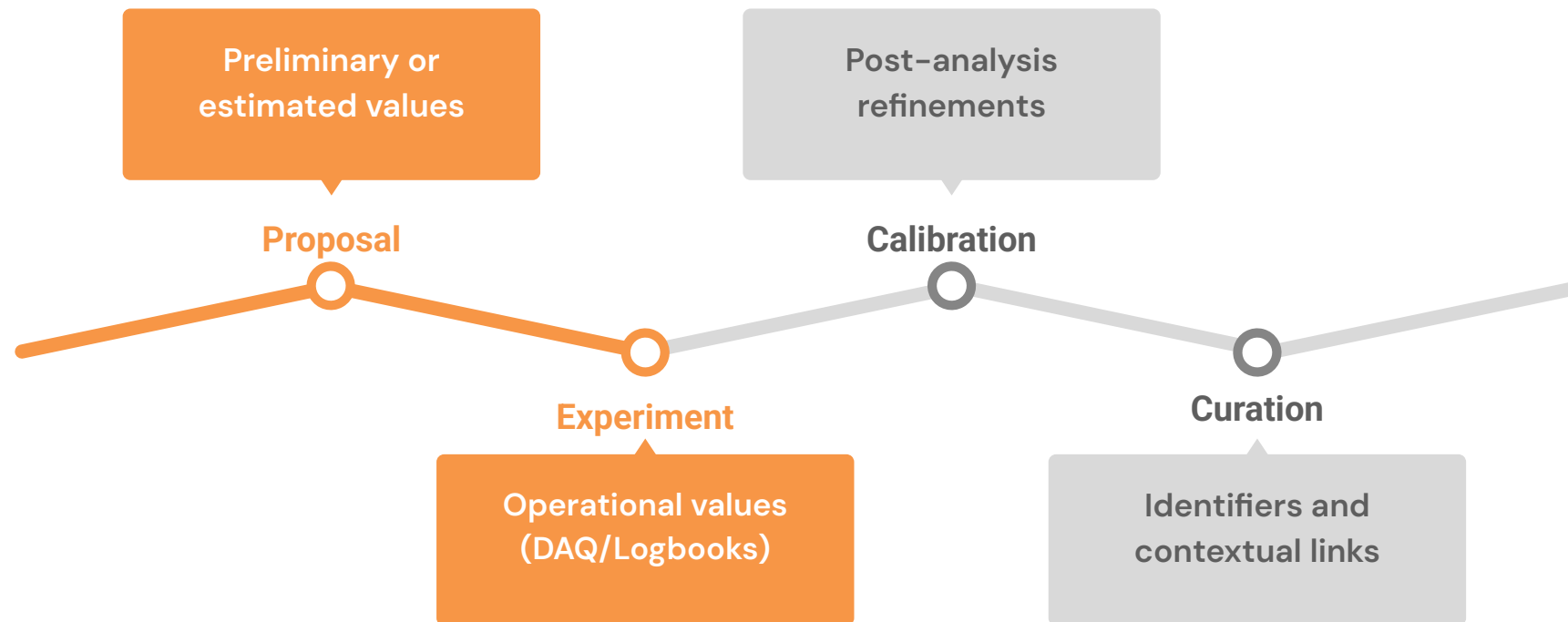
<https://zenodo.org/uploads/7270440>

Size	Version	Embargo
10kB	1	No

Data publisher	License	Category	Data collection
GSI	CC 4.0	Experiment	--

Why metadata evolves in NAP physics?

Metadata is not born final. It emerges in stages:



Result: Values such as beam energy, target composition, or experiment context may legitimately change after initial entry.

Implication: Scientific usefulness requires metadata to remain editable and improvable over time.

Scientific Reality

Metadata must evolve to stay accurate, complete, and scientifically useful.

Publication Reality

DOIs require stable, citable, reproducible records that should not drift over time.

If metadata never changes, it becomes **scientifically outdated**.

If DOI-linked metadata changes freely, **citation stability is weakened**.

Problem statement

How do we preserve a stable citation record without freezing scientific progress?

What we avoid

- Disruptive redesign of the current schema
- Indiscriminate versioning of every model/field
- Modeling multiple competing assertions in Phase 1

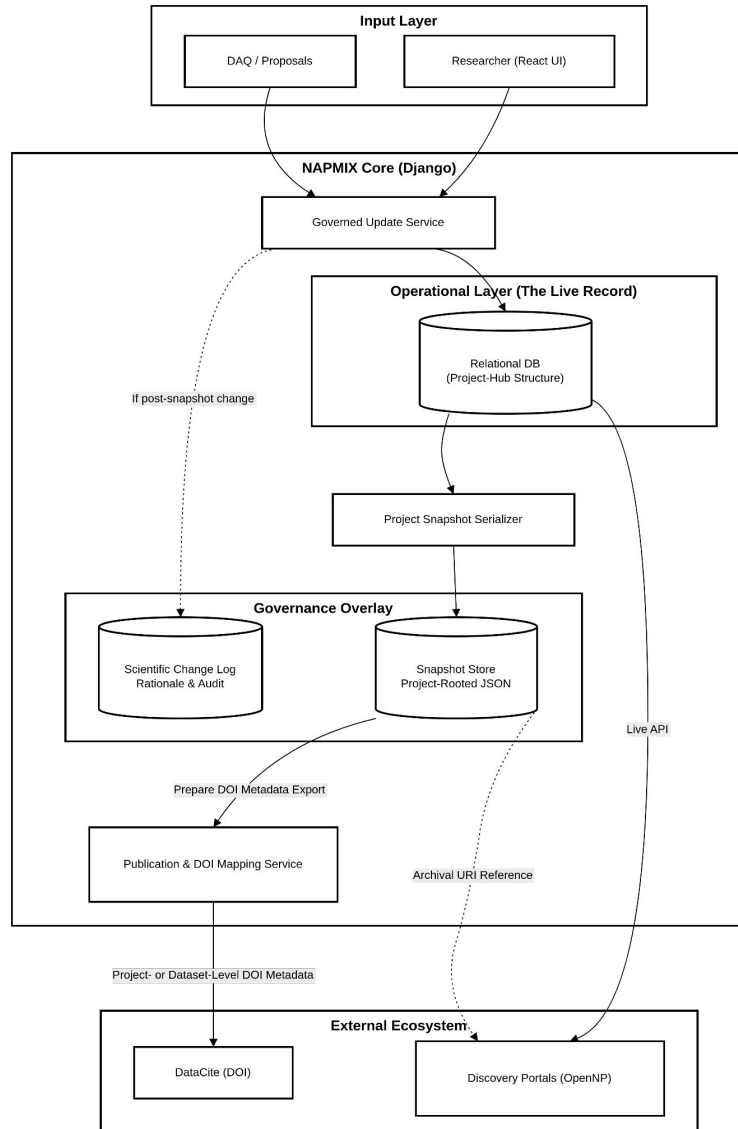
Our Choice: Overlay

- Keep the live operational model as it is
- Add a governance overlay around it
- Selectively govern high-impact physics metadata

Strategic shift

Governance instead of versioning

Governance Overlay Strategy



Functional Flow: the governed update service mediates between live metadata and frozen governance artifacts

- 01

Governed Update Service (Mediator)

The central entry point controlling the flow of information from DAQ/proposals and the React UI, ensuring data integrity before database persistence.

- 02

Layer Separation

A clear distinction between the Operational Layer (live data in a project-hub structure) and the Governance Overlay (immutable JSON snapshots and scientific change logs)

- 03

Publication & DOI Mapping (The Lens)

A dedicated service that maps frozen, project-rooted snapshots into stable DOI metadata at either the project or dataset level.

Metadata remains live for science, frozen for citation, and explainable through governed change logs.

The campaign suitcase

Snapshots are rooted in **Project**, the umbrella entity in NAPMIX.

- A snapshot freezes the project together with its governed datasets, reactions, and experimental context.
- Frozen JSON is the **authoritative archival record** for that moment.
- Publication-specific views are derived on demand.

Publication & DOI Mapping Service maps project snapshots into DataCite-ready metadata payloads



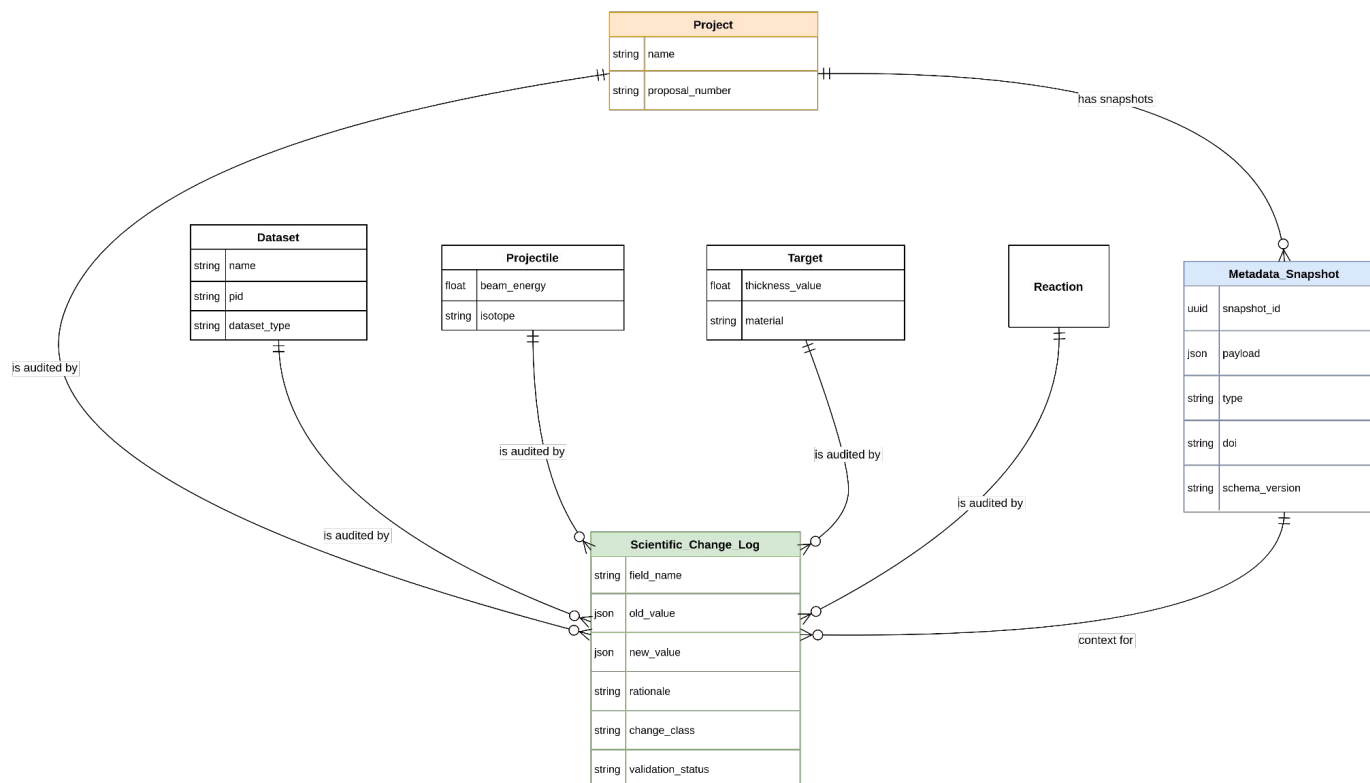
"One snapshot, many citation views"

Why Project-rooted? Keeps datasets anchored to the wider experimental campaign instead of isolation.

The DOI mapping service maps the same project snapshot into DataCite-ready metadata at either Project or Dataset level.

Project-rooted governance does not prevent dataset-level publication; it gives it stable context.

Technical implementation



The Governance Overlay

Non-Invasive Integration

Leverages Generic Relations (Generic Foreign Keys) to attach governance logic to existing schema entities without requiring a database refactor.



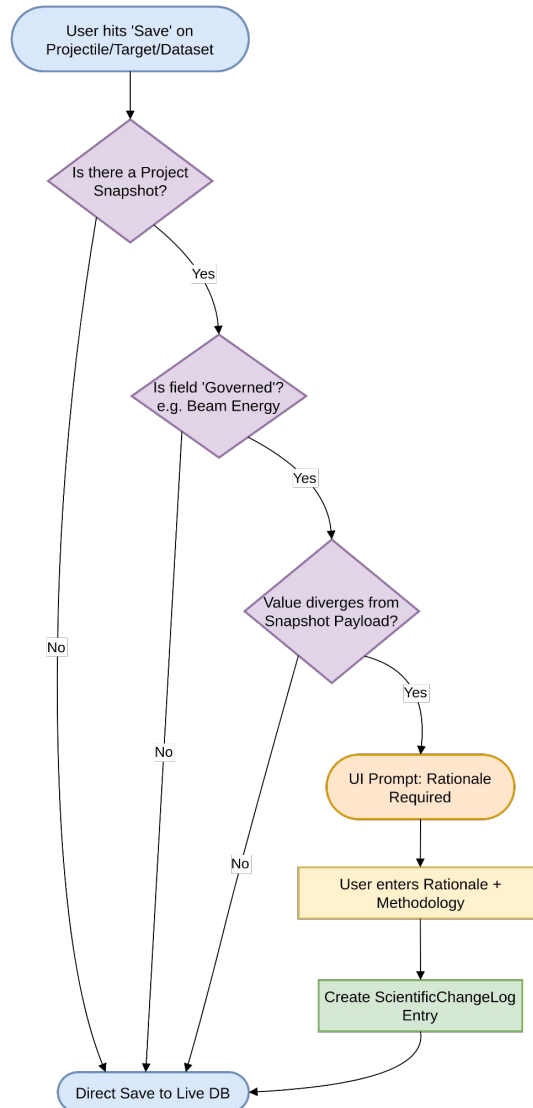
Metadata Snapshot

Directly anchored to the Project; captures the immutable, self-describing JSON payload of the entire experimental campaign.



Scientific Change Log

Acts as a central hub using Generic Foreign Keys to audit all governed entities (Dataset, Projectile, Target, Reaction) while referencing snapshots for scientific context.



The Divergence Workflow

Step 1

Context Awareness

The system checks for an existing Project Snapshot before applying any governance logic.

Step 2

Selective Governance

Only edits to high-impact Governed Fields (e.g. Beam Energy, Target Thickness) trigger the review process.

Step 3

Divergence Detection

The live update is compared against the Snapshot Payload; if the physics hasn't drifted, the save is silent.

Outcome

Transparent Audit

If values diverge, the UI prompts for Rationale and Methodology, which are then captured in the Scientific Change Log

We do not version everything

Phase 1 selectively governs high-impact physics fields and selected relations after a snapshot is frozen:

`Projectile.beam_energy`
`Projectile.percentage`
`Projectile.frequency`
`Reaction.reaction_type`
`Reaction.source_type`
`Target.target_thickness_value`
`Target.target_density`
`Target.sample_phase`
...

Why selective?

It avoids governance fatigue. Scientists are only prompted for meaningful divergence.

What remains live?

The operational record continues to evolve normally. Only governed changes after a snapshot trigger additional control.

Phase 1 philosophy

Stable citation where needed, flexible science everywhere else.

Phase 1 implementation scope



In Scope	Out of Scope / Deferred
Introduce two governance models: MetadataSnapshot and ChangeLog	Full versioning of all models and relations
Project snapshot serializer	Multiple concurrent assertions
Publication and DOI mapping service	Consensus / Dispute workflows
Divergence checks (governed fields)	Full ontology-backed registry

Implementation principle

Adoptable changes around the current system, not a ground-up redesign.



Stable Context

Project-rooted records for DOI use.



Live Science

Continued metadata improvement.

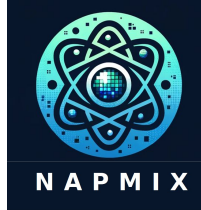


Explainability

Rationale for every critical change.

- Stable, project-rooted frozen records for DOI and collaboration use
- Dataset-level publication views that remain anchored to campaign context
- A clear explanation of why important metadata changed after a snapshot
- Better reproducibility and auditability without blocking live metadata improvement
- A path to future extensions, such as richer review logic or assertion support

NAPMIX solves the paradox by separating live metadata, frozen snapshots, and publication mappings.



Conclusion

NAPMIX is moving from a purely operational metadata platform toward **governed persistence**.

1. GOVERNED PERSISTENCE

Separates the **live scientific record** (operational truth) from the **frozen archival record** (citation truth).

2. PROJECT-ROOTED SNAPSHOTS

Preserves a **stable archival context** by capturing the metadata of the entire experimental campaign, not just isolated datasets.

3. SCIENTIFIC CHANGE LOG

Provides a transparent audit trail that **explains post-snapshot evolution** through controlled scientific rationale and methodology.

4. PUBLICATION MAPPING

Acts as the **lens** that focuses archival snapshots into stable, citable Project-level or Dataset-level DOI metadata.

Metadata remains live for science, frozen for citation, and explainable through governed change logs.

Thank you!

open-science@gsi.de

i.knezevic@gsi.de