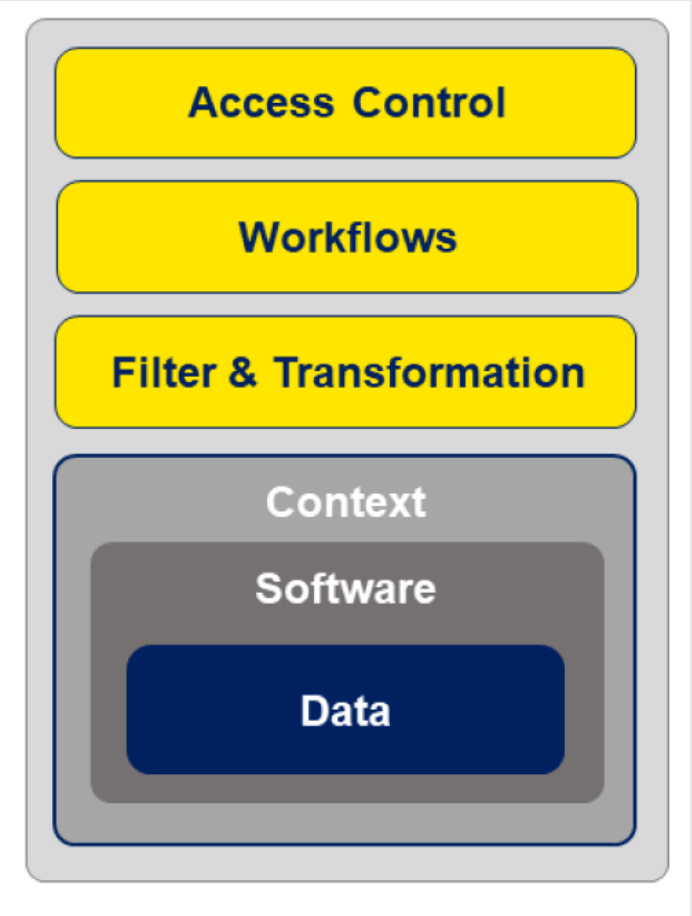
NFDIxCS: Guarantee Levels of the Research Data Management Container (RDMC)

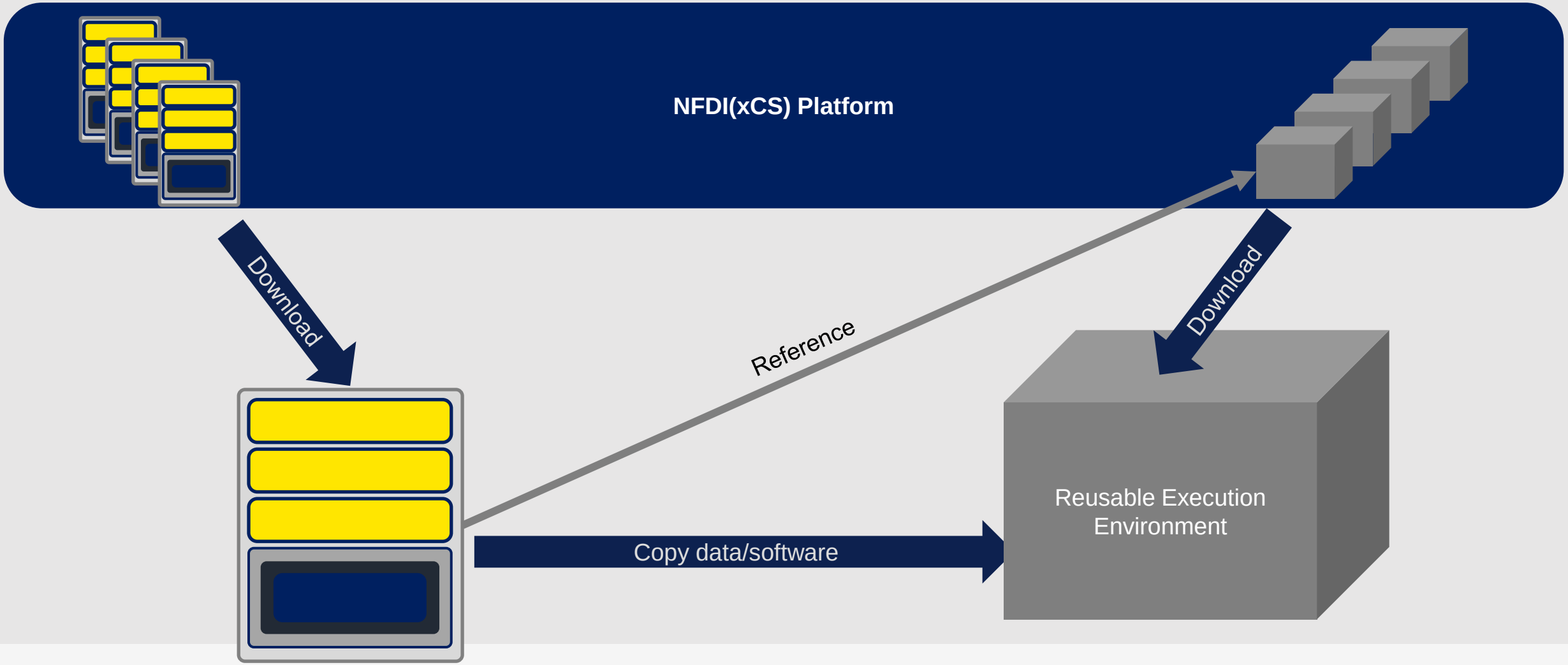
Safial Islam Ayon, Dr Firas Al Laban, Jan Bernoth, Michael Goedicke

Research Data Management Container (RDMC)

- Central Hypothesis of NFDIxCS:
  - Research Data and Research Software are not the same but it is linked together with its research context
- Solution Concept:
  - Design a **Research Data Management Containers (RDMC)** as a time capsule for research data and software
    - Digital, referenceable object
    - Describable with Metadata
    - Manage access and workflow
  - Additionally design and implement a hosting platform RDMCs

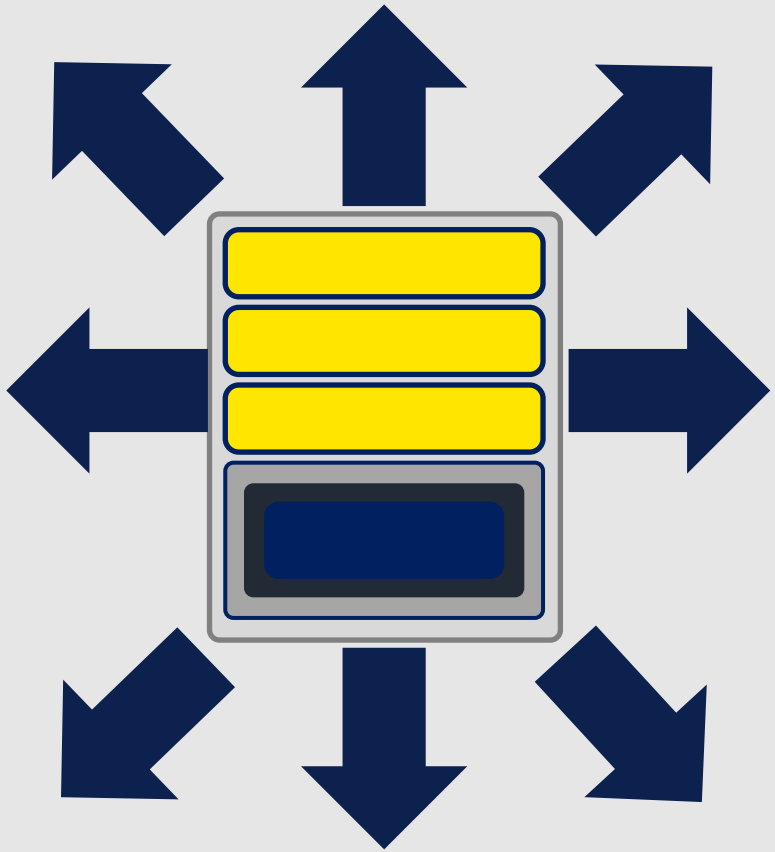


RDMC Reusable Execution Environment



Connecting Active Containers: Integration with Services and Infrastructure

- Interact with platform for e.g. artifact review processes
- Use metadata to connect to knowledge graphs
- Create an execution environment to run the software with its data
- **Integrate quality metrics by providing guarantee levels**



Guarantee in Research Software Engineering

- Research software is iterative and improves over time.
- Guarantee is important for future development.
- Verification ensures:
  - Functional correctness.
  - Alignment with system requirements.
  - Potential for reusability and reproducibility.
- Confidence Through Guarantees:
  - Data and Software behaves as expected.
  - Supports validation, feedback, and requirements.

Why Integrate Guarantees into RDMC?

- Flexibility for Diverse Needs:
  - Guarantees are shaped by chosen standards and methodologies.
  - Not all containers need long-term preservation or full standardization.
- Encourages Best Practices:
  - Guarantee responsibility lies with RDMC creators.
  - Choice of metadata, sealing, and tools impacts long-term quality.
  - Structured and well-documented artifacts enhance research integrity.
  - Hashing and sealing to maintain data quality over time.
  - Controlled access mechanisms help maintain quality and security.

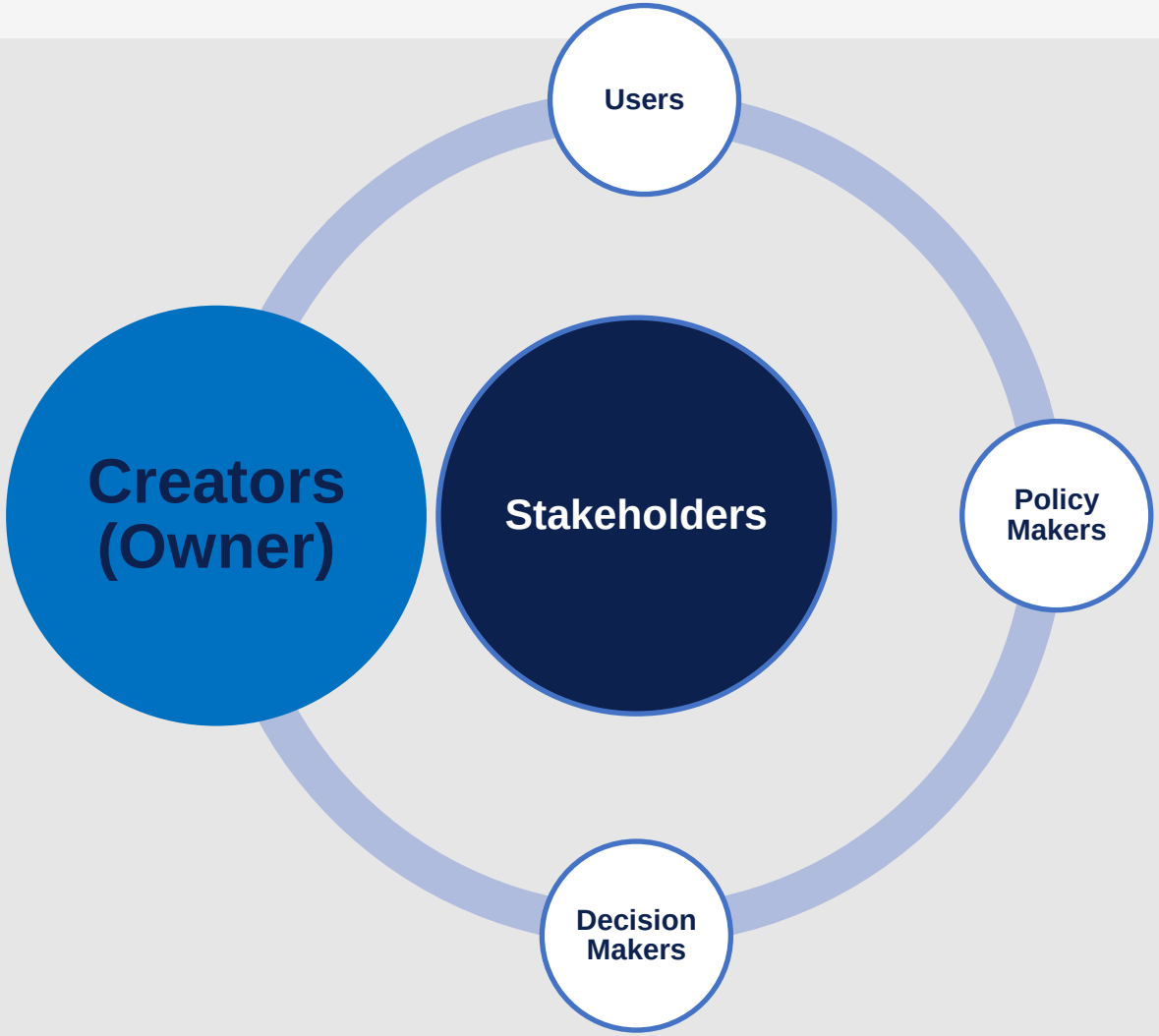
Why Integrate Guarantees into RDMC? (2)

- Informs About Missing Dimensions:
  - Highlights areas for improvement.
  - Provides opportunities to integrate best practices.
- Transparency for Container Readiness:
  - The guarantees of an RDMC depends on the methods and standards chosen by its creator.
  - RDMC ensures transparency but does not enforce specific guarantee levels.
- Supporting Decision-Making:
  - RDMC helps users make informed decisions by showing how well processes are followed.
  - Stakeholders can evaluate an RDMC’s trustworthiness based on creator-provided guarantees.

Stakeholders' Interaction with RDMC Guarantees

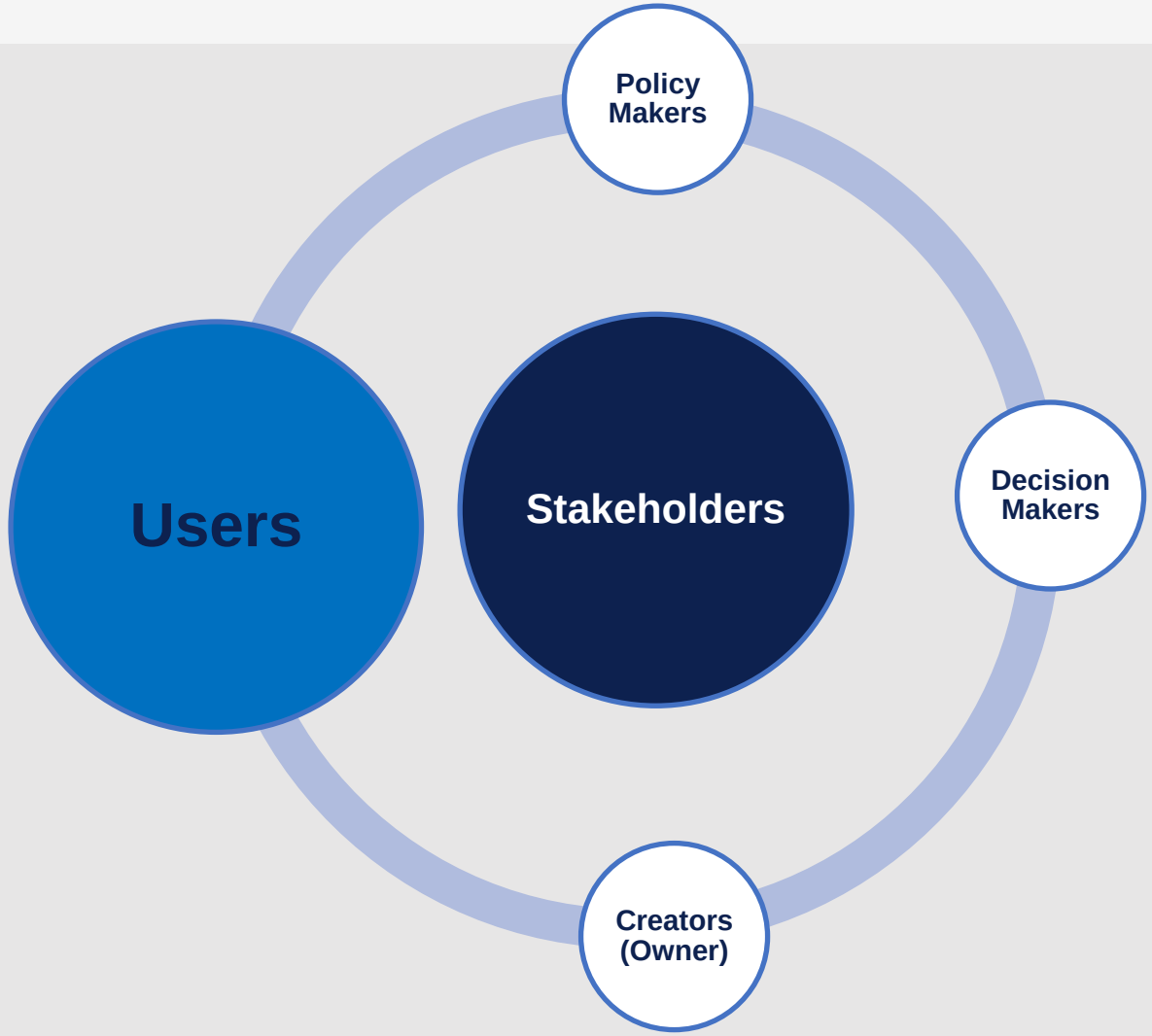
Owner's Benefits in RDMC Guarantees

- Understand how RDMCs are created and the quality of services used.
- Compare RDMCs with best practices.
- Identify gaps and improve reliability.
- Improve RDMCs by learning from guarantees.



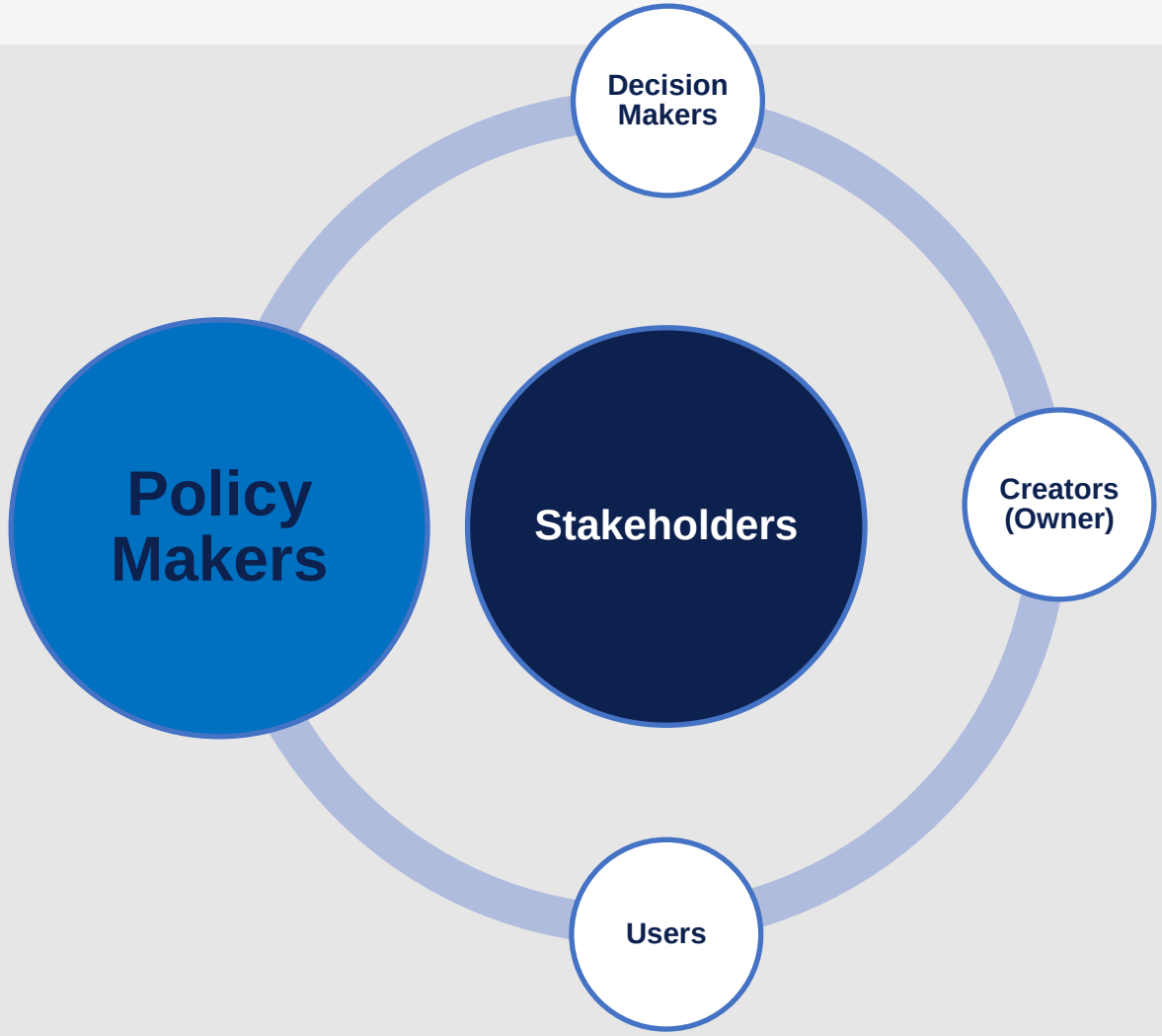
User’s Benefits in RDMC Guarantees

- Users get easy access to well-documented, structured data.
- Ensures trust about quality of artifacts.



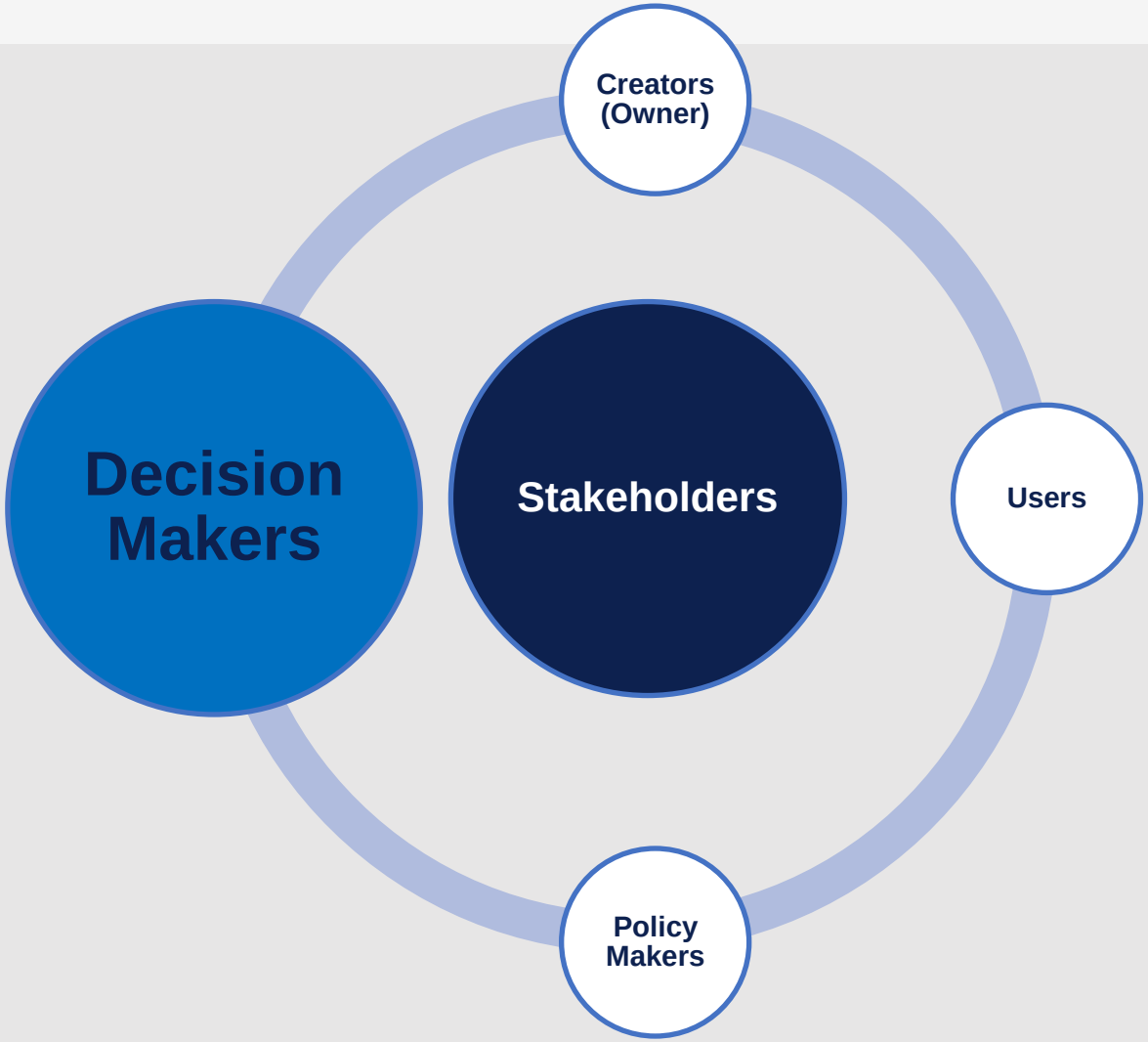
Policy Maker's Benefits in RDMC Guarantees

- Provides a provable foundation for policy development.
- Supports efficient allocation of funding.
- Aligns with FAIR principles, promoting open science.
- Ensures legal and best practice standards.

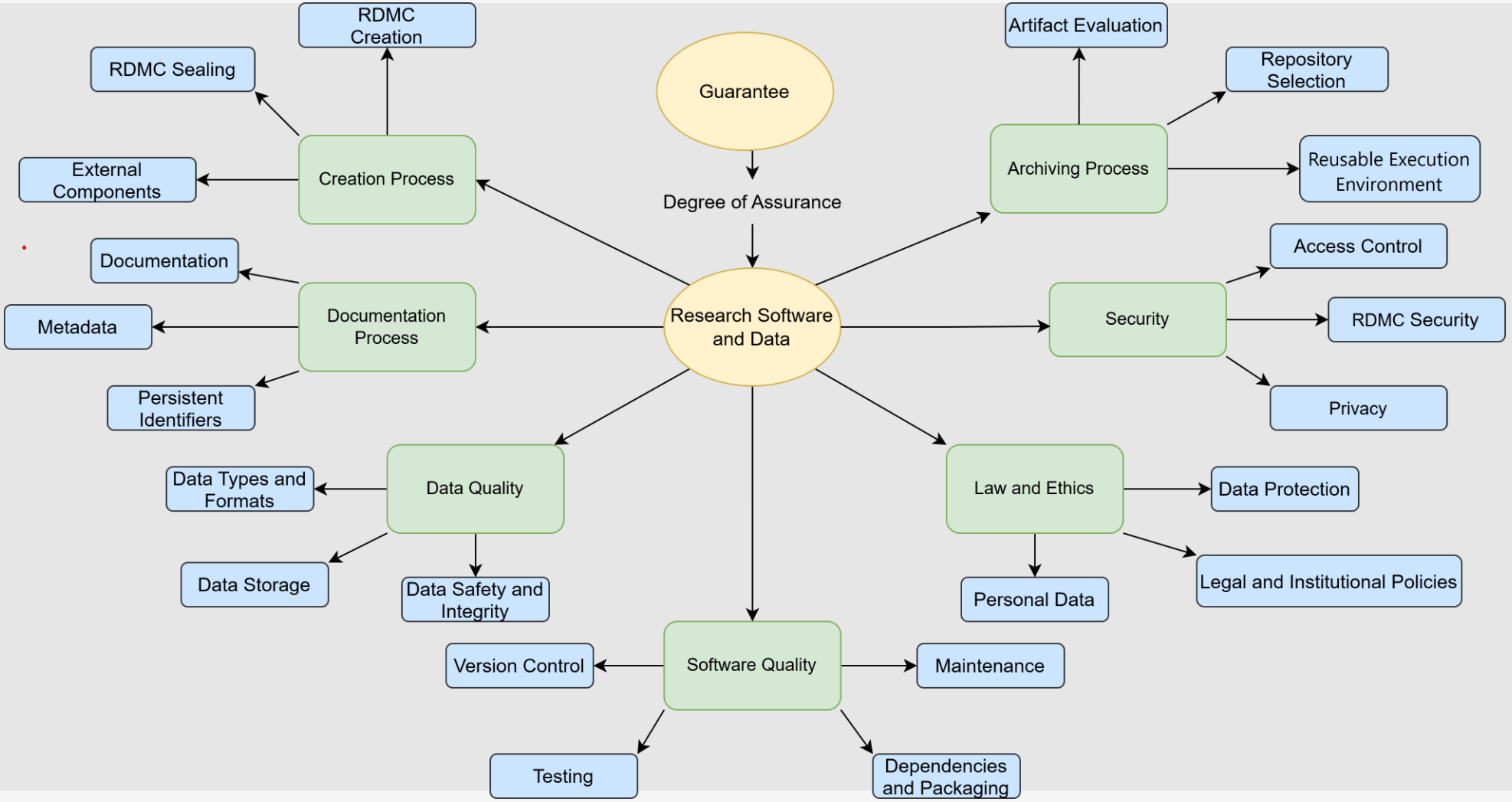


Decision Maker's Benefits in RDMC Guarantees

- Supports legal and ethical standards and FAIR data policies.
- Promotes transparency in research management.
- Supports KPI evaluation through accumulated metadata



(Possible) Guarantee Dimensions in RDMC



Guarantee Dimensions – Sealing & Metadata

Sealing

- Not Sealed
- Server Side
- Hashing (Public and Private key)

Metadata

- No Standardization
- Minimal Standardized (Dublin core, DataCite)
- Full Standardized (CodeMeta, schema.org, DCAT)
- Customize Metadata Standard

What Comes Next?

- Developing methods to define and implement guarantees
- Identifying key starting points for integration
- Developing evaluation criteria for guarantees
- Identifying key dimensions to be considered for comprehensive guarantees
- Understanding the relationship between different guarantee dimensions (e.g., how sealing guarantees affect metadata quality)

Summary

- Guarantee measures define the quality and consistency of RDMCs
- They ensure reliable, secure data management across research projects.
- Different dimensions of guarantee provide flexibility to balance effort and security.
- Guarantees make RDMCs adaptable for diverse research needs and environments.
- They play a key role in enhancing the trustworthiness and long-term usability of RDMCs.

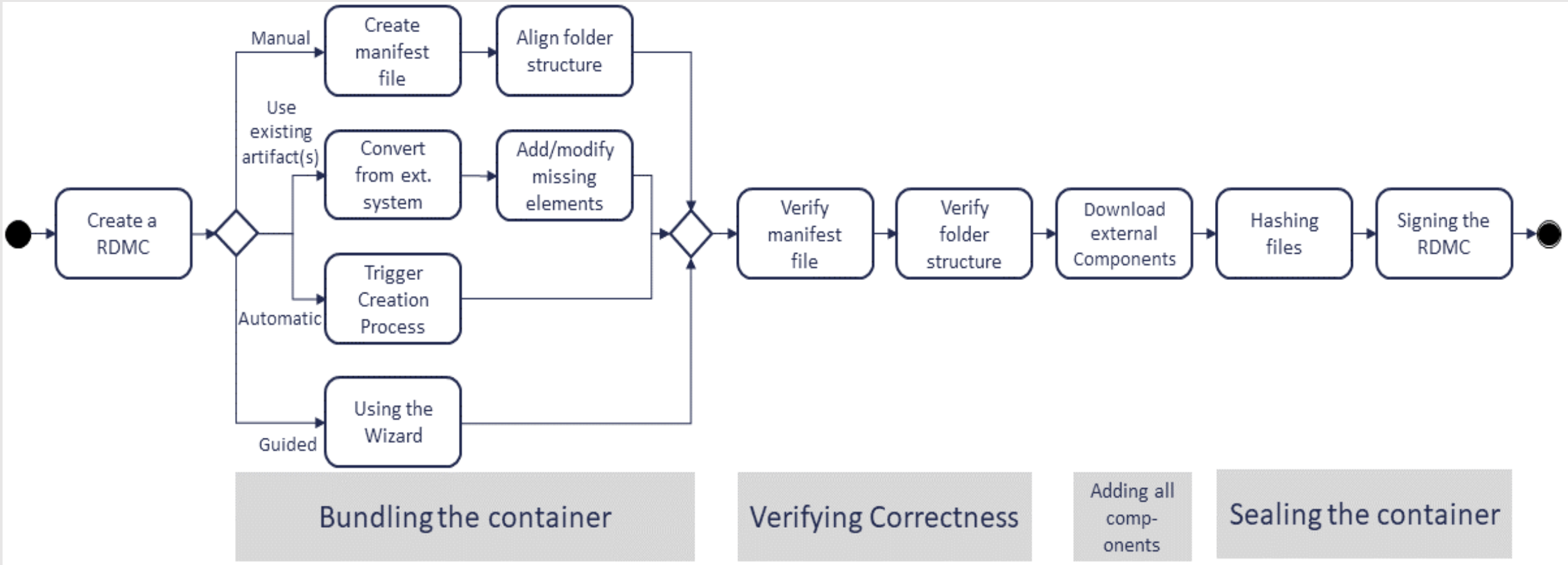
Safial Islam Ayon
Research Assistant,
NFDIxCS
University of Potsdam
ayon@uni-potsdam.de

FUNDED BY
DFG Deutsche Forschungsgemeinschaft German Research Foundation
under the National Research Data Infrastructure–NFDI 52/1– 501930651

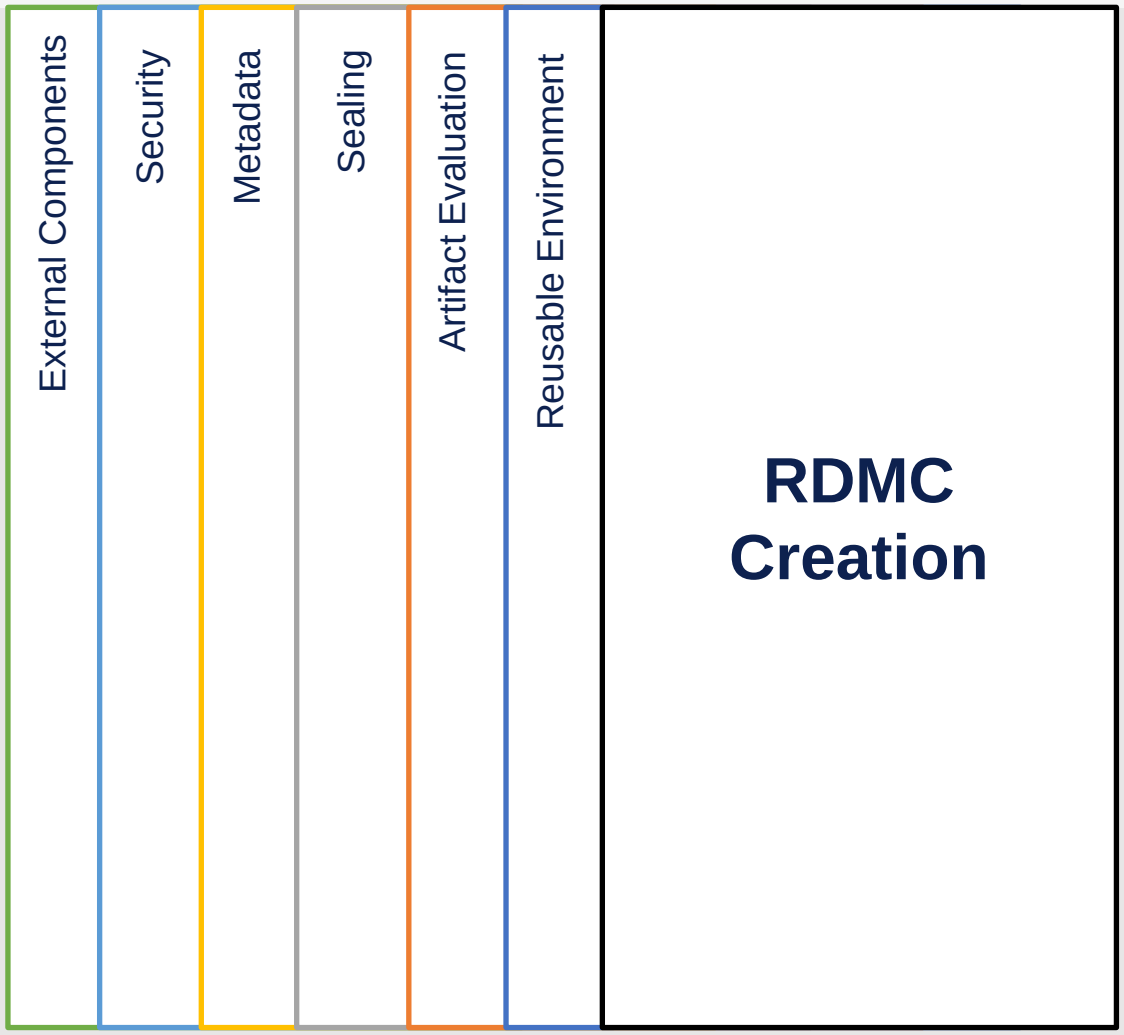
IN COOPERATION WITH
nfdi Nationale Forschungsdaten Infrastruktur

Appendix

RDMC Creation Process

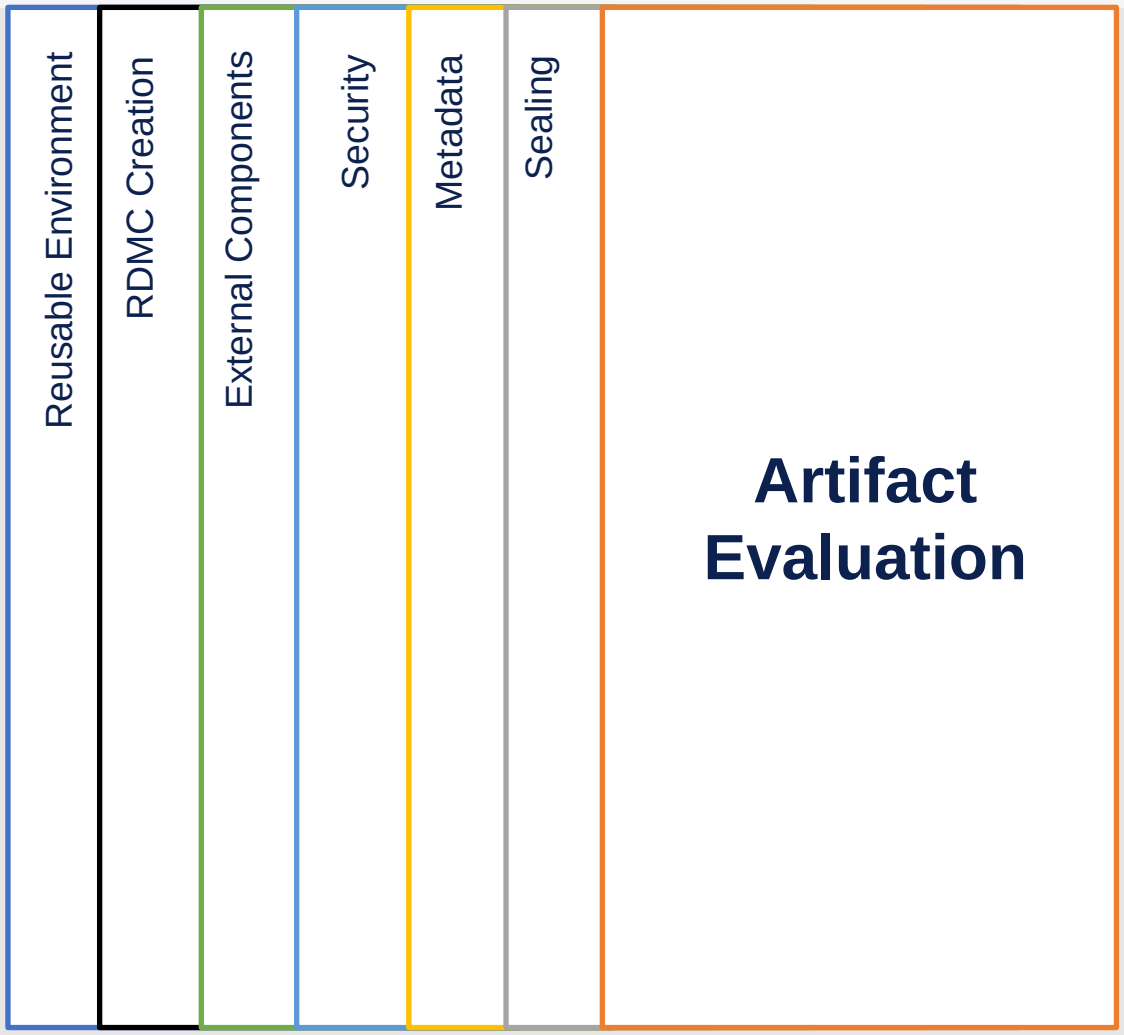


Guarantee Dimensions – RDMC Creation



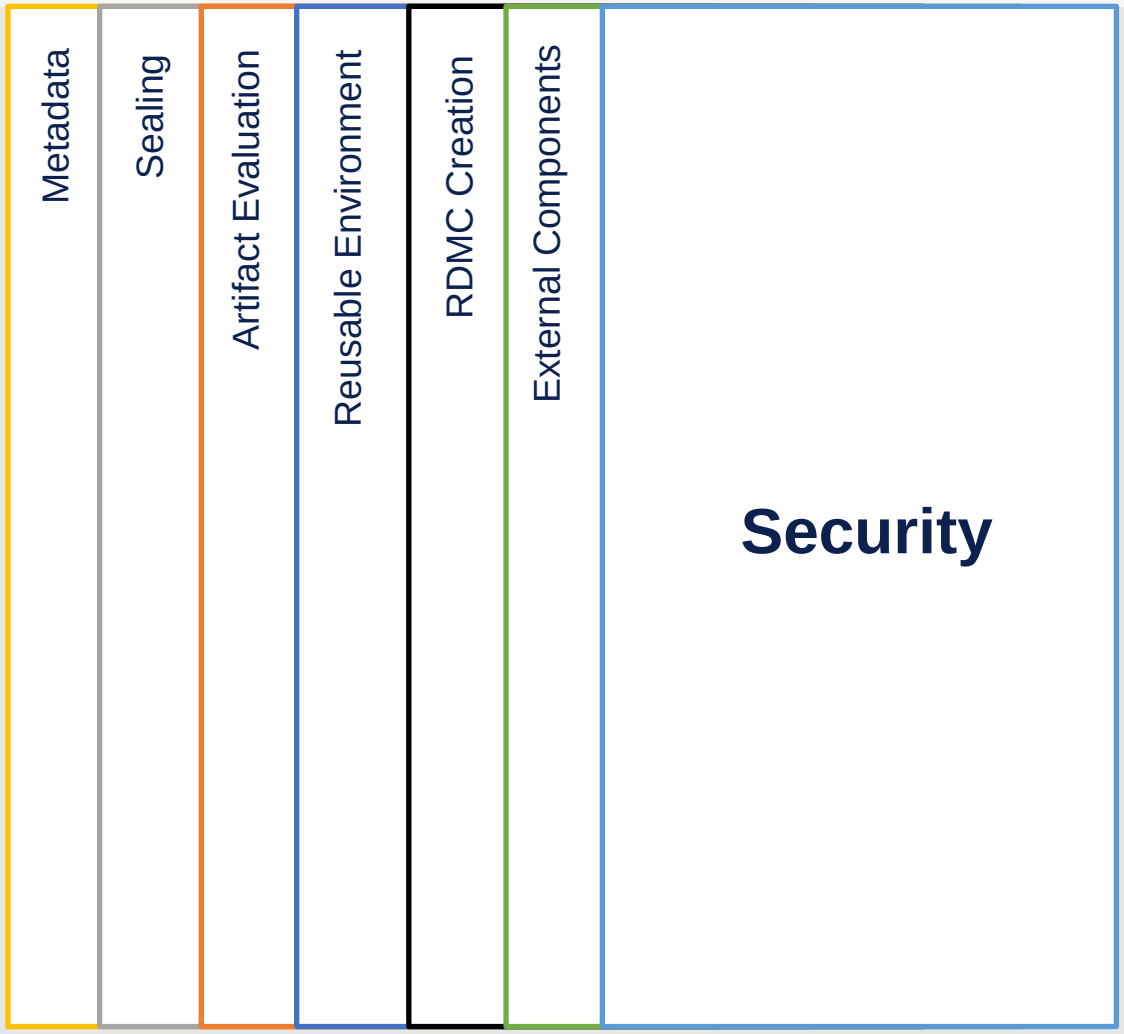
- Manual Creation
- Use Existing Artifacts
- Automatic Process
- Guided Wizard

Guarantee Dimensions - Artifact Evaluation



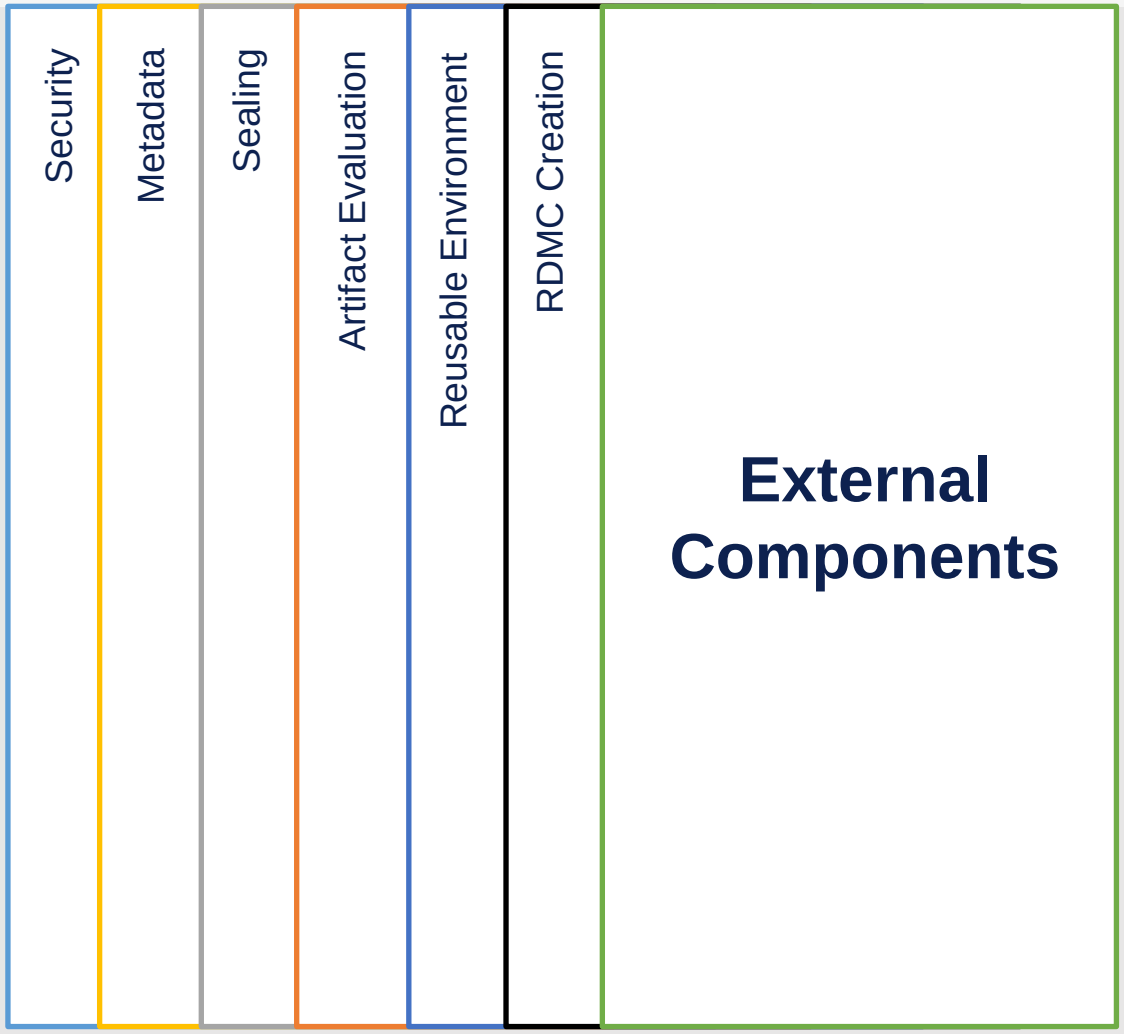
- No Badge
- Artifacts Evaluated (Functional and Reusable)
- Artifacts Available
- Results Validated (Reproduced and Replicated)

Guarantee Dimensions - Security



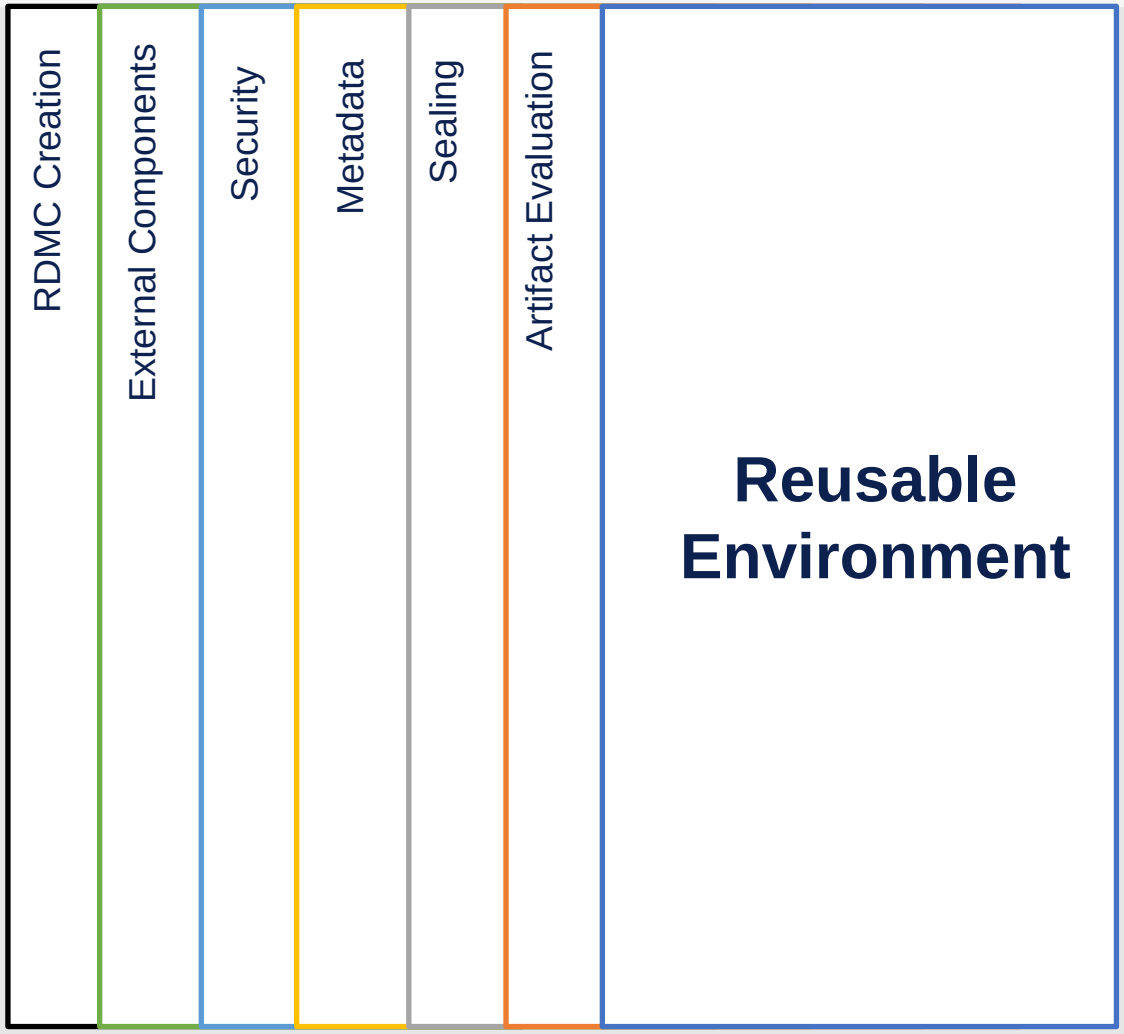
- No security, no encryption
- Minimal security, possible encryption
- Full encryption, secure access

Guarantee Dimensions - External Components



- External links only
- Some local, some external
- All resources stored locally

Guarantee Dimensions - Reusable Environment



- Not Includes
- Recording
- Docker
- Ngnix

Future dimensions that can be consider

- Cross-platform integration
- Community-based reviews & feedback
- Automated metadata extraction
- Scalability (for handling large datasets or complex workflows).
- Interdisciplinary compatibility.
- Data management plan and Software management plan.

Future Prospects of Guarantee Services in RDMC

- Adaptability:

  RDMCs remain flexible to future advancements.

- Integration with New Standards:

  Adaptation to new tools and standards.

- Continuous Improvement:

  Enable ongoing improvements in security and quality.

- Reliable Research Outcomes:

  Standardized guarantees will support consistent and verifiable results.

- Facilitated Collaboration:

  Guarantees enable researchers to confidently share and reuse data.