

Improving research data management for samples: the SEPIA Sample Database for Metadata Storage and Exchange

Mojeeb Rahman Sedeqi^{1,2,°} // Rolf Krah¹ // Katherine Rial^{1,3} // Heike Görzig^{1,2}

¹ Helmholtz-Zentrum Berlin für Materialien und Energie GmbH (HZB)

² Helmholtz Metadata Collaboration (HMC)

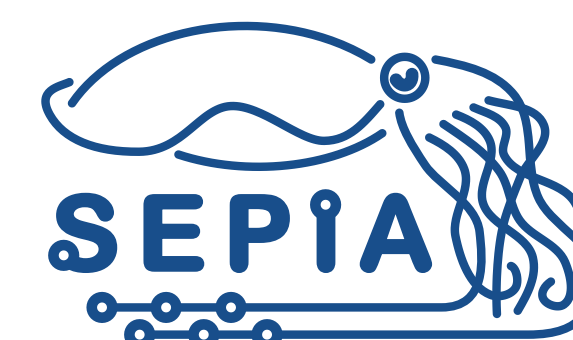
³ ROCK-IT

° Presenter: Mojeeb Rahman Sedeqi; Contact: hmc-matter@helmholtz-berlin.de

SEPIA: Enhancing Sample Metadata Management

The SEPIA (Sample Essentials, Persistent Identifiers & Attributes) system improves research data at matter facilities in Helmholtz by providing a structured approach to sample tracking and metadata management. By assigning unique identifiers, SEPIA enables researchers to track sample history, ensure data interoperability, and enhance collaborative research.

This poster presents SEPIA's architecture and functionalities, showcasing its impact on data accessibility, sharing practices, and scientific reproducibility across labs and implementation in a first use case at the Helmholtz-Zentrum Berlin (HZB).



https://codebase.helmholtz.cloud/hzb/research_data_management/sepia

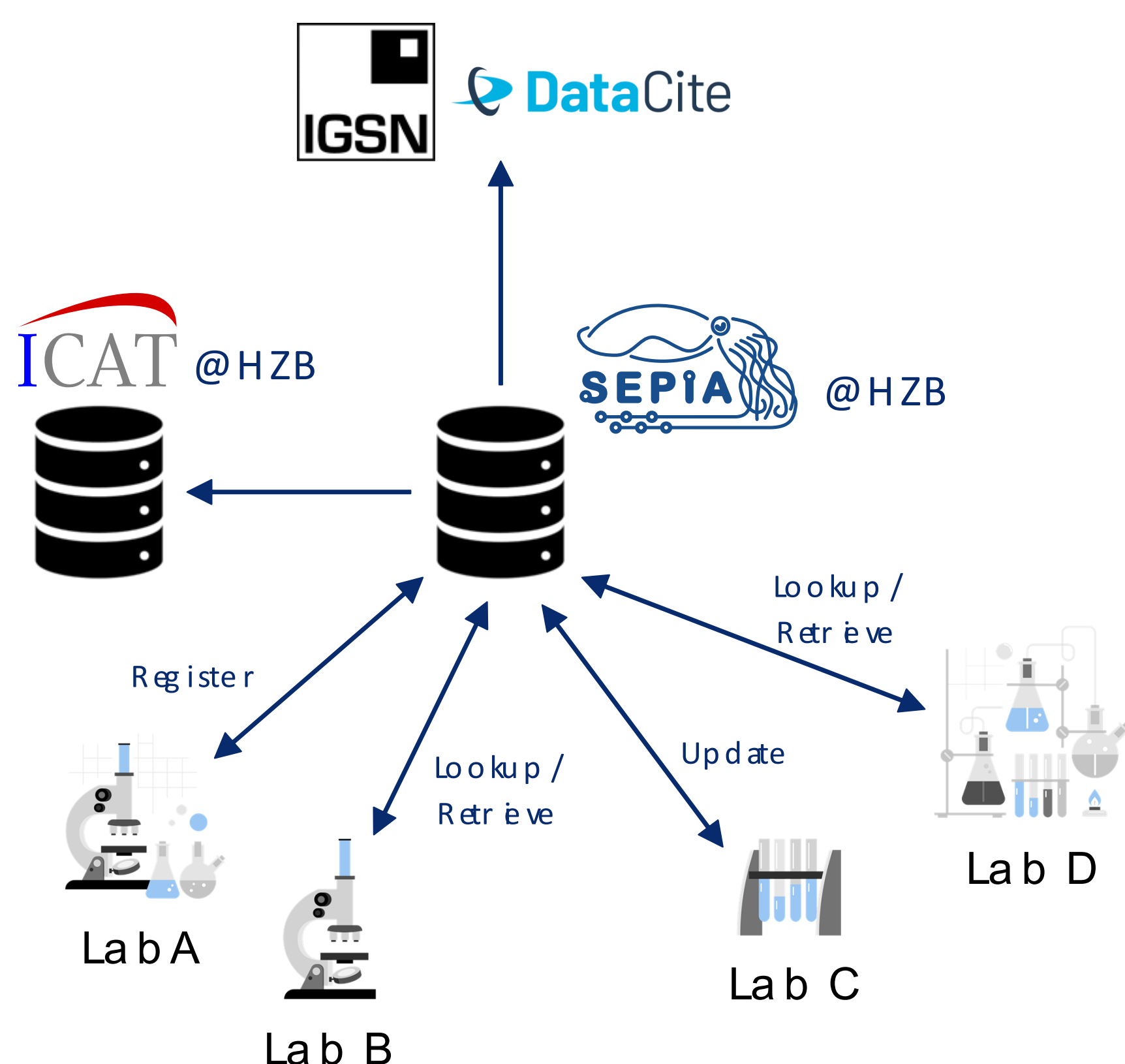
Introduction / Background

Why SEPIA?

- Thousands of samples are examined annually at **Matter facilities**, but metadata tracking is **limited** beyond safety checks.
- There is a growing need to **catalogue all chemicals, objects, and artifacts** analyzed at beamlines.
- Some samples are **also examined elsewhere or modified between measurements, making tracking crucial**.

Solution:

- SEPIA** – a database to systematically store and manage sample metadata.



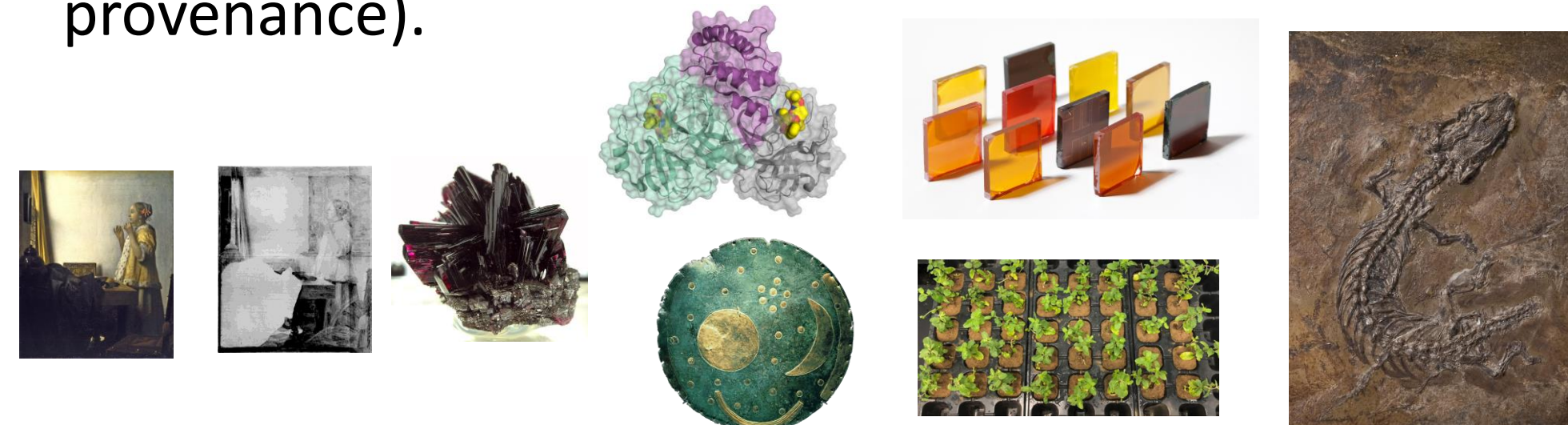
Objectives of SEPIA

- Assign Persistent Identifiers (PIDs)** to samples using **IGSN & DataCite**.
- Ensure **global sample identification**, independent of any institution.
- Accept **pre-existing PIDs** to preserve sample history.
- Track **sample modifications, location history, and experiments** over time.
- Provide **an integrated user interface** for researchers and beamline scientists

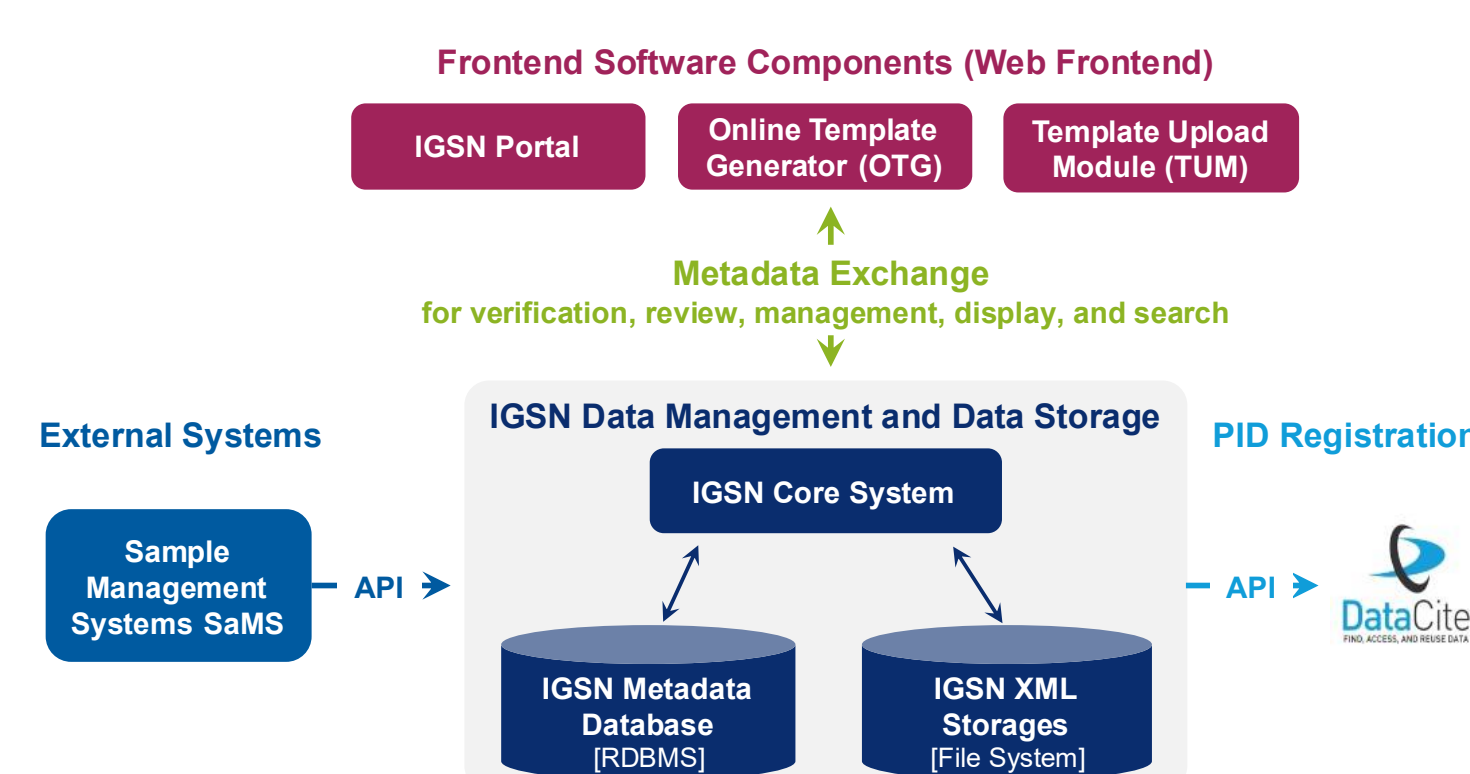
System Architecture

How SEPIA Works?

- Database for sample metadata (chemicals, objects, provenance).



- Minting of PIDs via IGSN & DataCite.



Schema of the software components with the frontend modules on top IGSN Portal

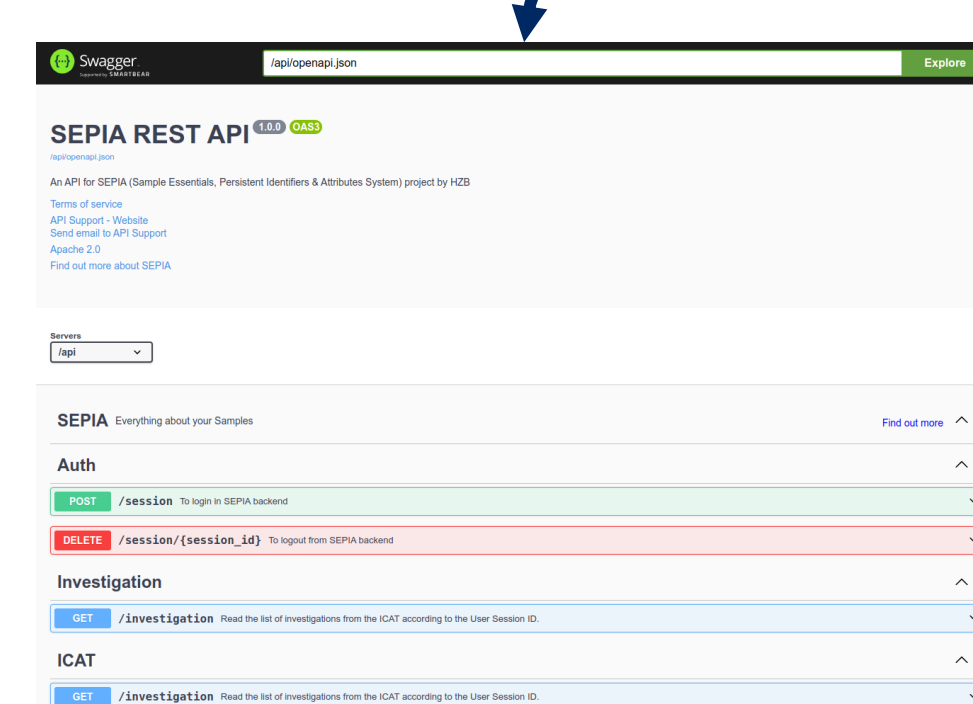
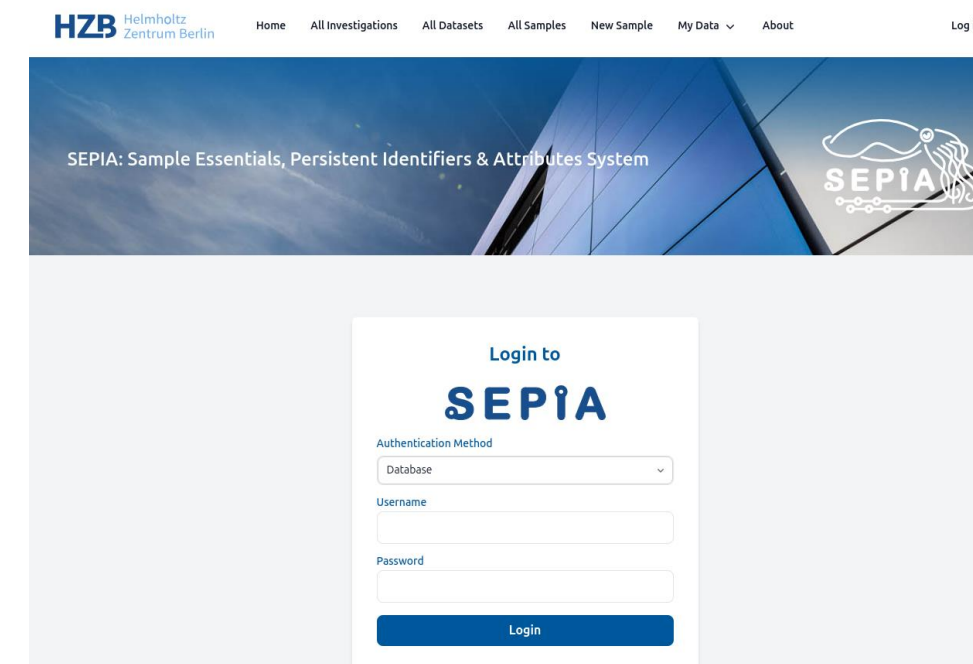
- Integration with HZB ID Management.
- Tracking of modifications before, during, and after measurements.

Implementation

- SEPIA is implemented based on a relational database (PostgreSQL).
- The backend implements a REST API using the OpenAPI specification.
- The frontend build using the NextJs, tailwindcss and Shadcn UI.

Key Functionalities

- Unique Sample IDs (PIDs)
- Metadata Storage
- Interoperability
- User Interface
- Data Sharing



Impact & Future Development

Why SEPIA Matters

- Improves scientific reproducibility through complete sample histories.
- Aligns with FAIR data principles, making research more transparent.
- Enhances collaboration between beamlines, labs, and institutions.
- Enables long-term sample tracking, supporting future studies.

Next Steps:

- Expand database functionalities.
- Improve UI for better usability.
- Develop API for automated data input/output.
- Implement feedback from researchers & labs.

References

- [1] Krah, R., Sedeqi, M. R., & Rial, K. (2024, November 5). Enhancing Research Data Annotation: The SEPIA Sample Database for Metadata Storage and Exchange. Helmholtz Metadata Collaboration Conference 2024 (HMC 2024), online. Zenodo. <https://doi.org/10.5281/zenodo.14041060>
- [2] FAIR Data Maturity Model WG. (2020). doi: [10.15497/rda00050](https://doi.org/10.15497/rda00050).
- [3] Collaboration, T. I. C. A. T. (2014). The ICAT Project. The ICAT Collaboration. <https://doi.org/10.5286/SOFTWARE/ICAT>

Acknowledgments

This work was supported by the Helmholtz Metadata Collaboration (HMC), an incubator-platform of the Helmholtz Association within the framework of the Information and Data Science strategic initiative; the NFDI consortium DAPHNE4NFDI funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460248799; and the Helmholtz Digital Research Infrastructure ROCK-IT.