

# SEPIA: A Scalable System for Integrated Sample Metadata Management

## Sample Essentials, Persistent Identifiers & Attributes

Mojeeb Rahman Sedeqi<sup>1,2</sup> // Rolf Krah<sup>1</sup> // Katherine Rial<sup>1,3</sup> // Heike Görzig<sup>1,2</sup>

1 Helmholtz-Zentrum Berlin für Materialien und Energie  
2 Helmholtz Metadata Collaboration  
3 ROCK-IT  
Contact: [hmc-matter@helmholtz-berlin.de](mailto:hmc-matter@helmholtz-berlin.de)  
[data-management@helmholtz-berlin.de](mailto:data-management@helmholtz-berlin.de)

### SEPIA – Sample Metadata Platform



**Global Identifiers** – Worldwide, institution-independent sample PIDs via IGSN and DataCite.



**Complete Sample Records** – Flexible metadata for provenance, keywords, contributors, and timelines.



**Beamline & Lab Integration** – Supports seamless metadata entry and retrieval across workflows.



**FAIR & Interoperable** – Enhances reproducibility and collaboration through FAIR-aligned metadata.

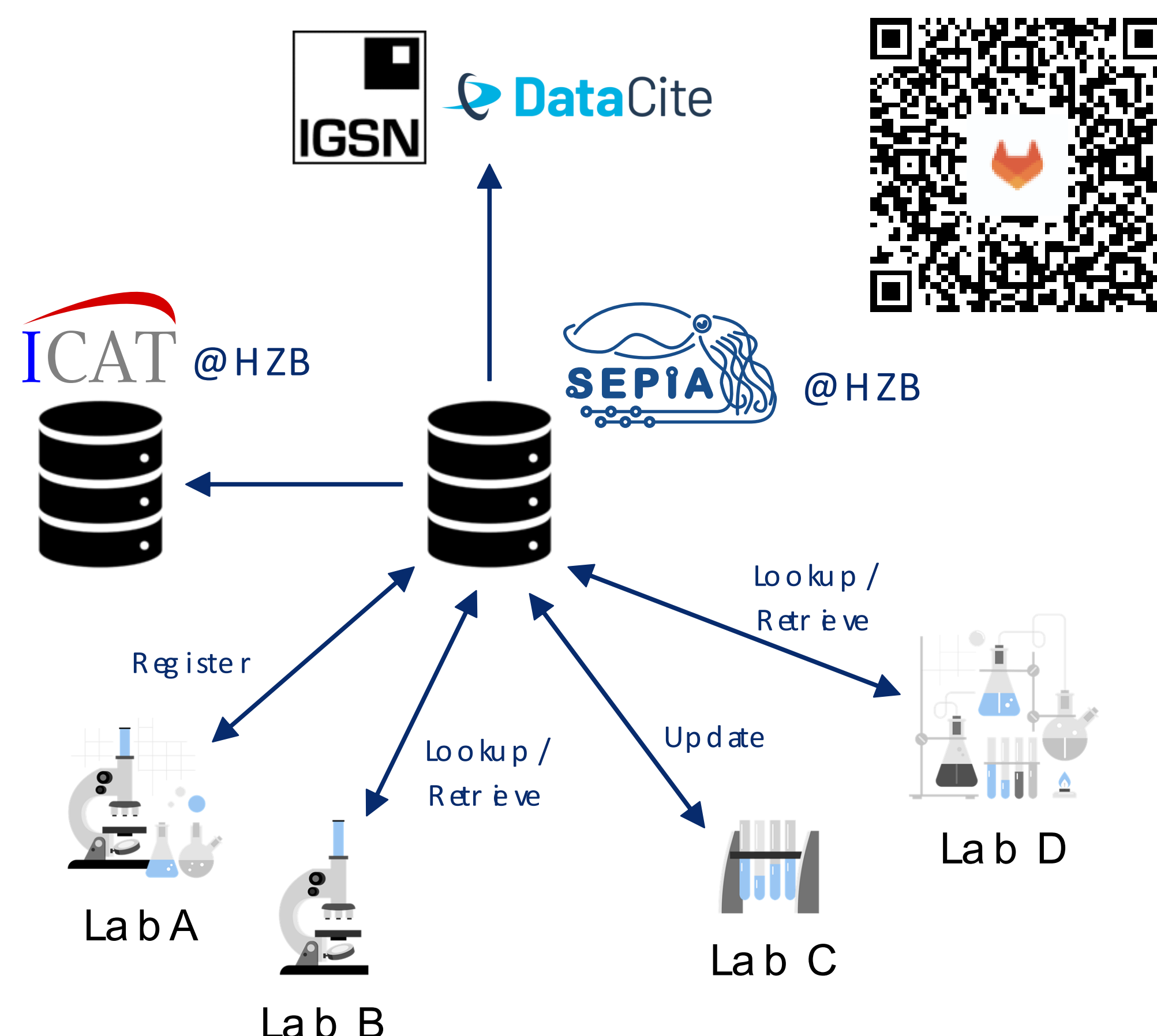
## Why SEPIA?

### Challenges

- **Thousands** of samples measured **every year** at **Matter facilities**.
- Metadata tracking often **limited** to safety or beamline operations.
- 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

- **Centralized, persistent, FAIR-compliant sample metadata platform.**
- Supports complete sample **tracking before, during, and after experiments**.



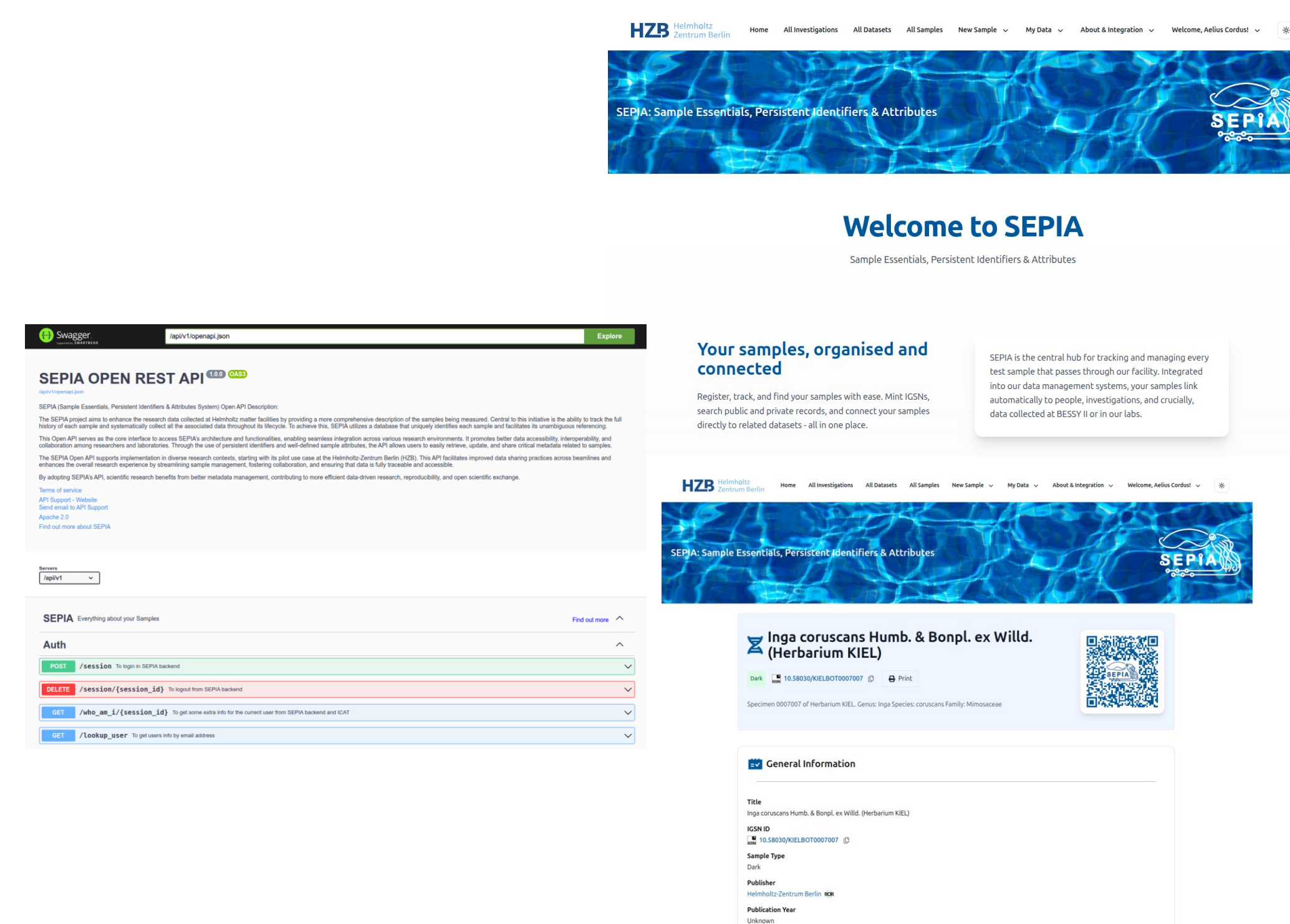
## Objectives of SEPIA

### Primary / Metadata Objectives

- **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**.
- **Record static and dynamic metadata** depending on sample type such as: **(basic info, related items, contributors & affiliations, keywords, Sample-type fields and context-dependent attributes)**

### Technical Implementation

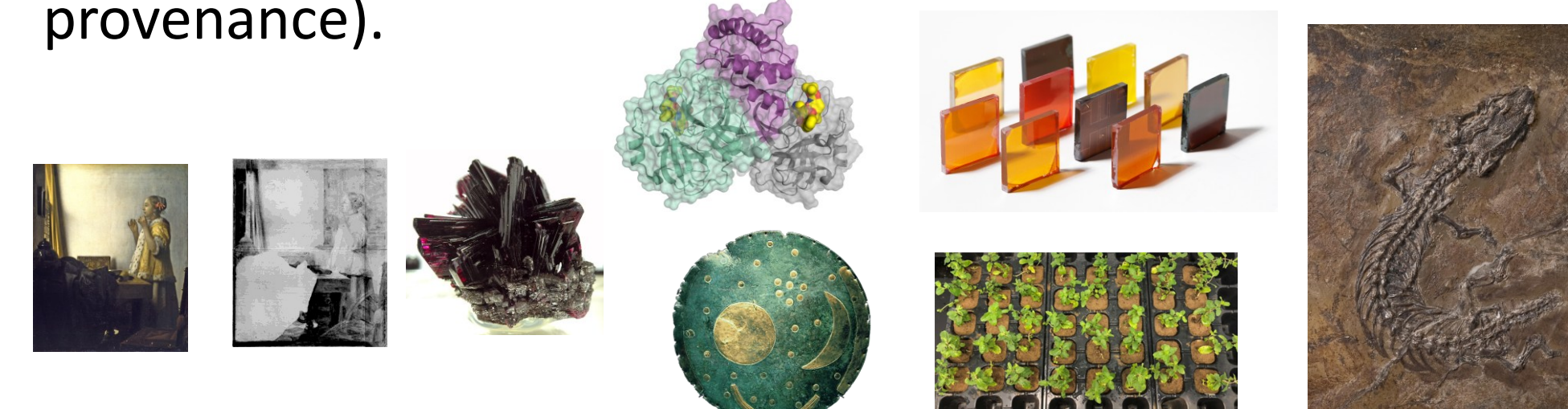
- **PostgreSQL** relational database as backend.
- **Flask REST API** following **OpenAPI** specifications.
- **Next.js, Typescript, TailwindCSS, and Shadcn UI** for the frontend.
- Tightly integrated with **ICAT**.



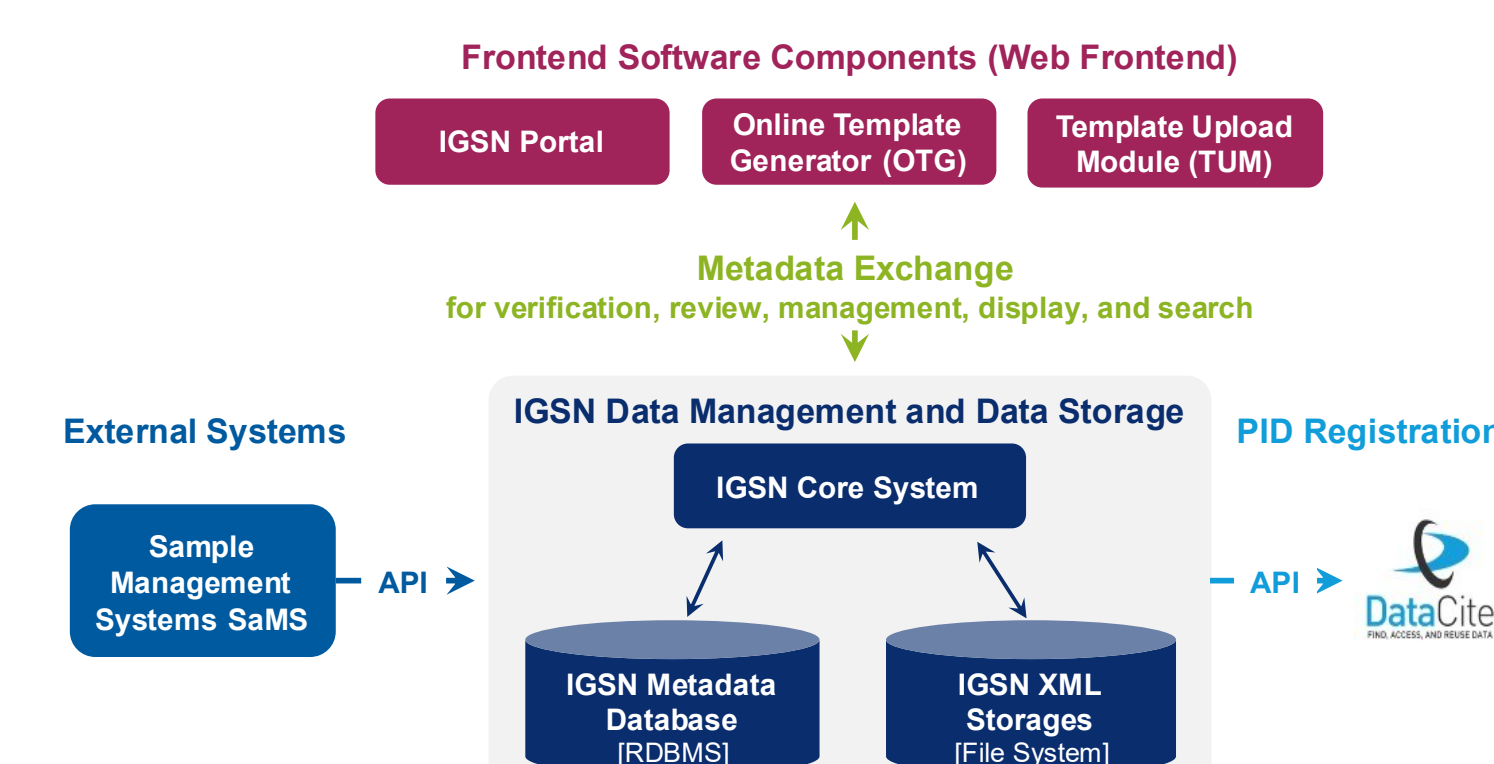
## System Architecture

### How SEPIA Works?

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



- Minting of PIDs via **IGSN & DataCite**.



- Integrable with new **HZB ID Management**.
- Tracking of sample **activities before, during, and after measurements**.

## Why SEPIA Matters

### Scientific Benefits

- Ensures **scientific reproducibility** through complete sample records.
- Provides **transparent provenance** across preparation & measurement cycles.
- Integrates with **FAIR principles**:
  - Findable (**PIDs**)
  - Accessible (**API**)
  - Interoperable (**standards**)
  - Reusable (**rich metadata**)

## Why SEPIA Matters ...

### Facility & Collaboration Benefits

- **Unifies sample tracking** across:
  - **Beamlines**
  - **Laboratories**
  - **Institutions**
- **Supports**:
  - **Cross-experiment comparison**
  - **Collaborative research workflows**

### Next Steps

- **Expand database** and workflow functionalities.
- **Enhance UI/UX** based on user feedback.
- Develop API for **automated data input/output**.

## References

- [1] Krah<sup>1</sup>, 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). <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.