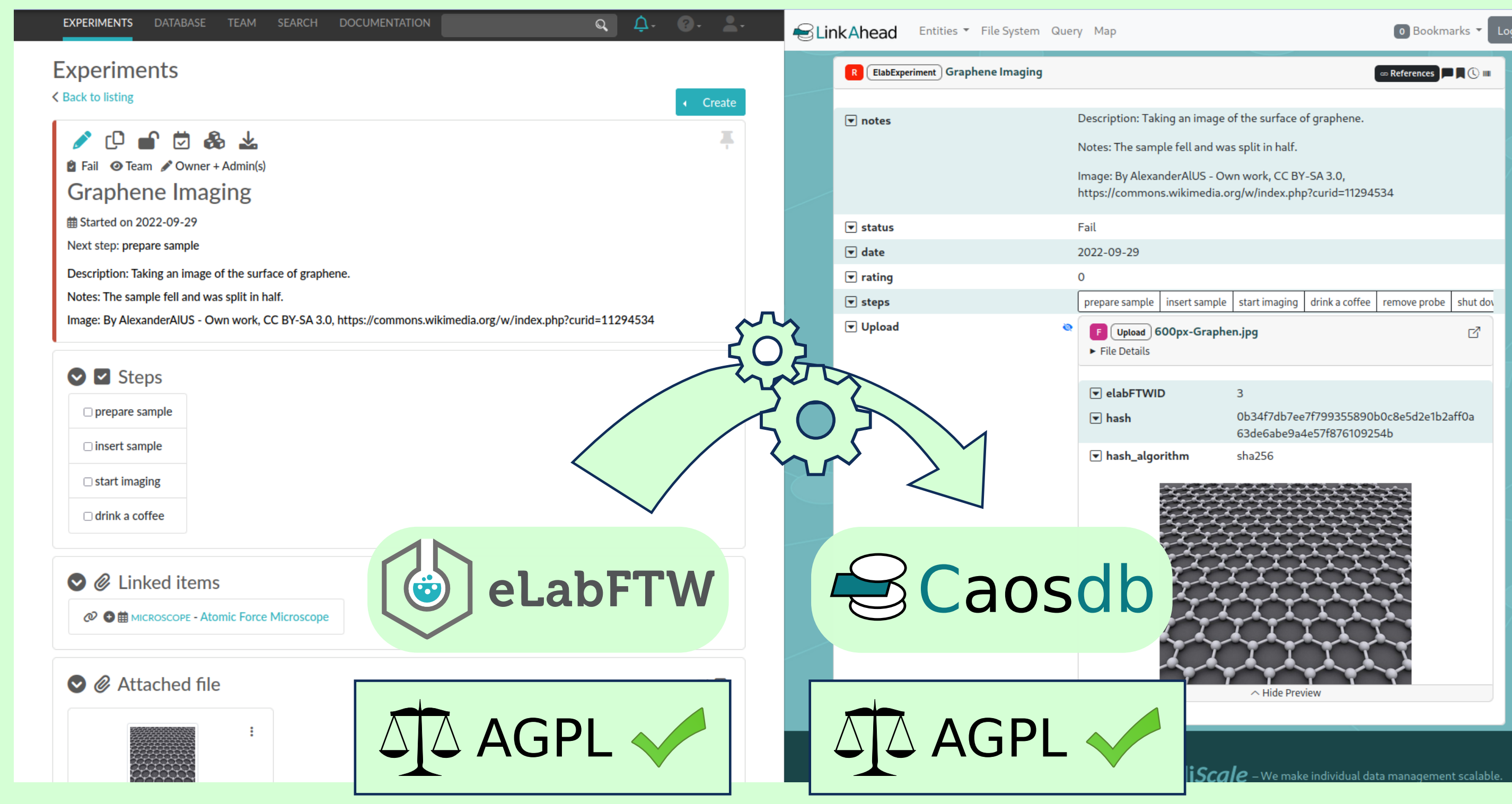


Data management challenges

FAIR data: Data should be Findable, Accessible, Interoperable, Reusable.
Data lifecycle: Data management must encompass the full data lifecycle.
Semantic data: Knowledge and data usability can be at danger if only single experts know the context in which measurements were obtained, how intermediate results should be interpreted and in general what some data "means".
Specific requirements in science: Ever-changing research questions, new devices and experimental protocols, custom-written software, good scientific practice requirements and low budgets demand solutions which allow for an agile workflow, data sharing and access control and a high degree of automation.

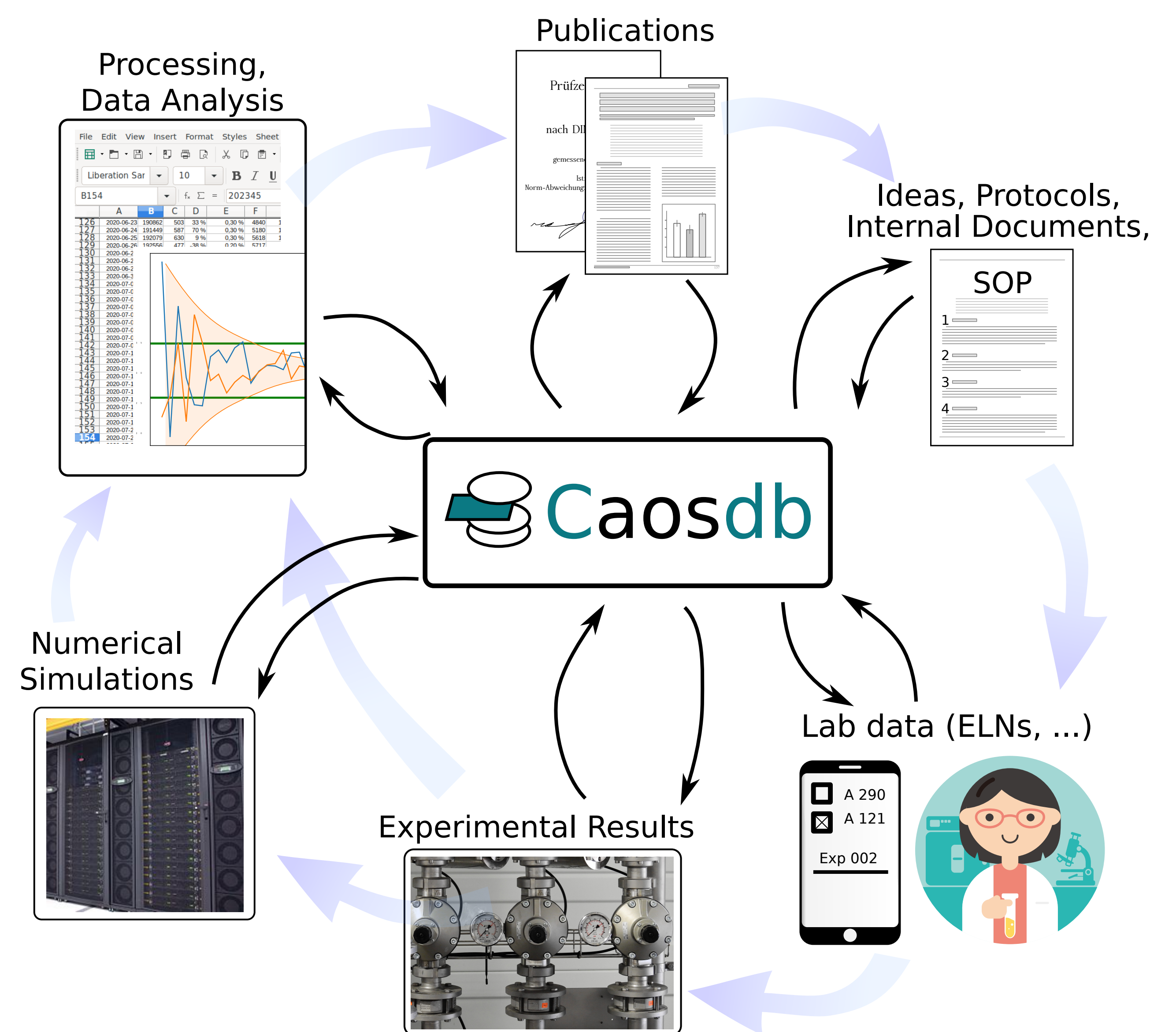
ELN integration: key to user acceptance

ELNs (Electronic lab notebooks): device and experimental settings, semi-structured **Data:** Critical for filtering and analysis
Approach: Automatic integration of ELN data into data management system



Data lifecycle

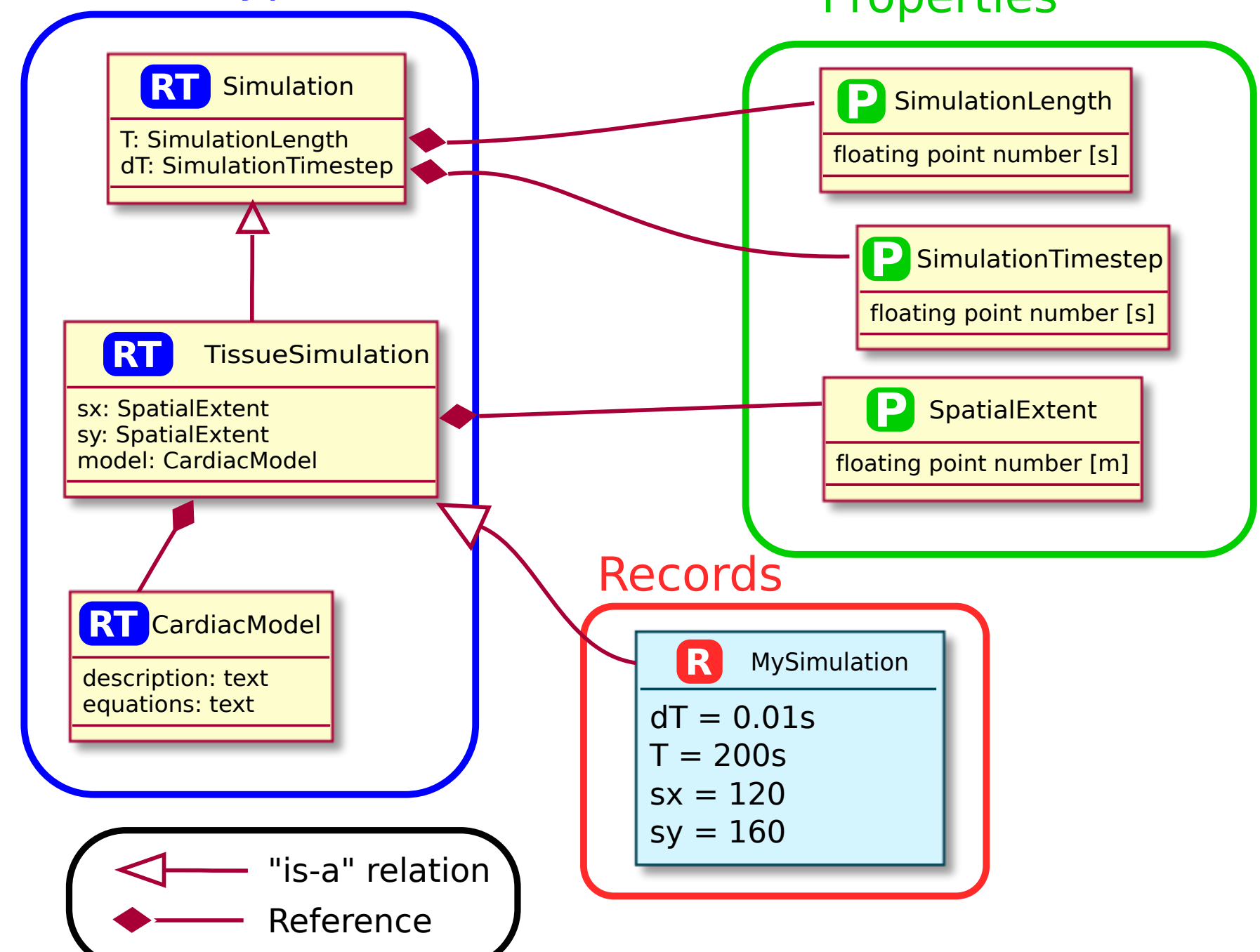
Valuable data = Interoperable, reusable data.
 CaosDB **integrates** data from all lifecycle stages and connects the data sets via **semantic links**.



CaosDB's data model

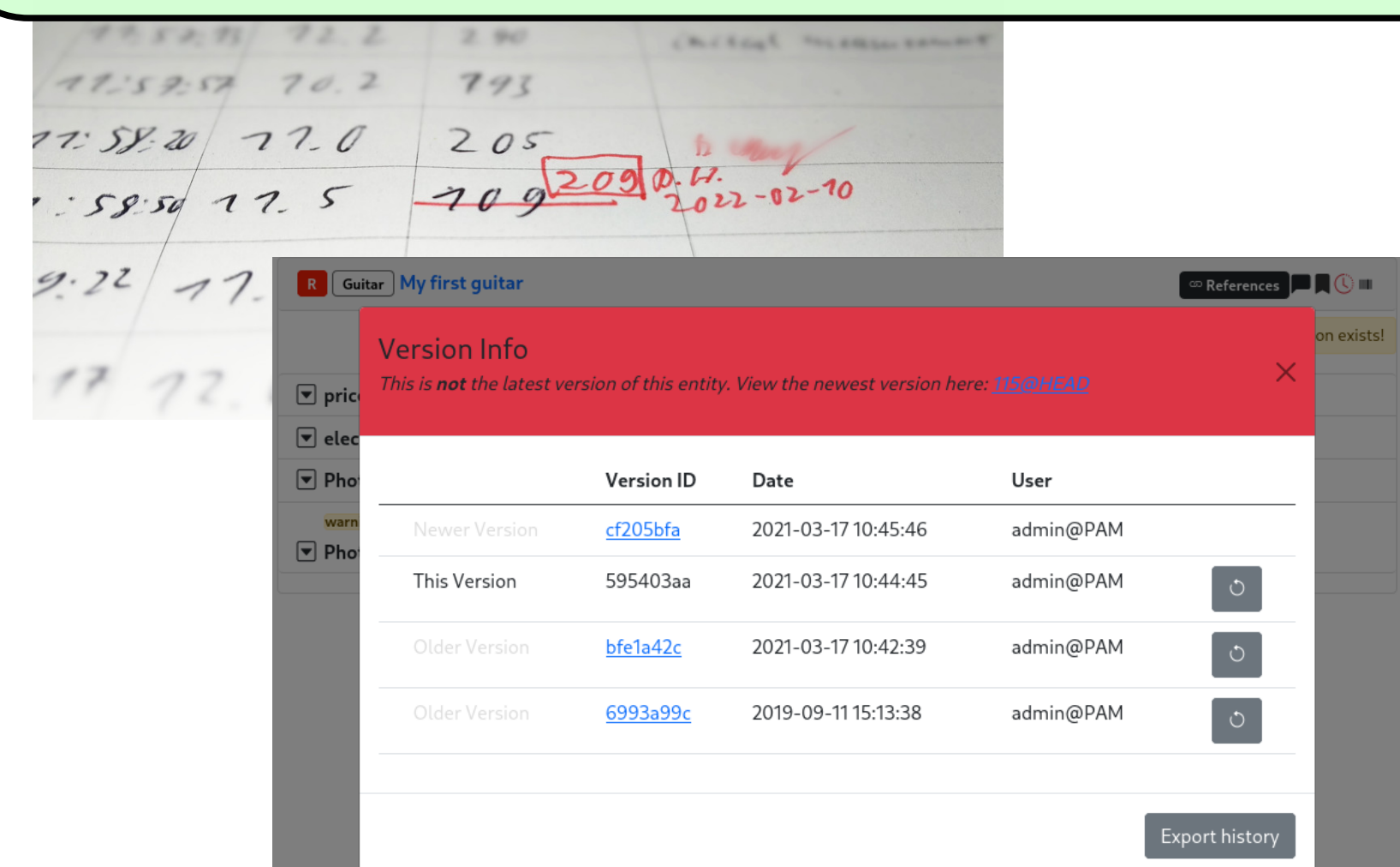
- Inheritance structure of RecordTypes.
- Records (= data) may choose to have additional Properties.
- Graded importance system for Properties.

Record Types



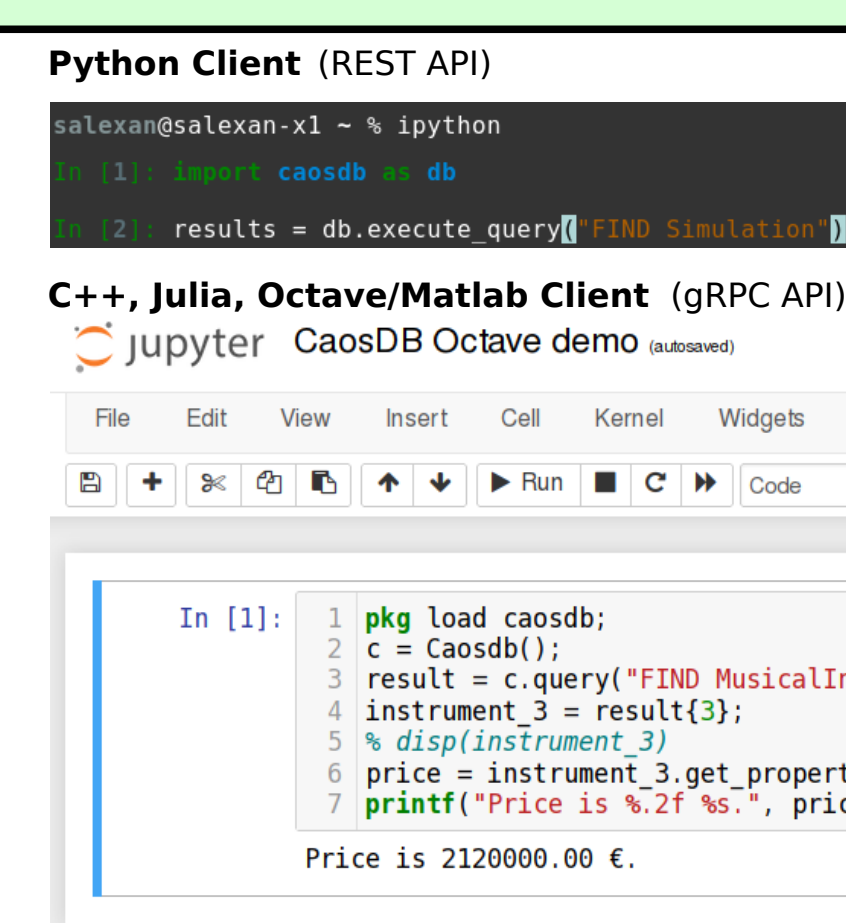
Good Scientific Practice: Retain old versions of your data

- Data can be changed, but old versions are kept.
- Unique identifier for each version: reference can be either to HEAD or to specific version.



Permissions & Roles

- Keep private data protected, publish selected data.
- Fine-grained permissions system, selectively for viewing, modifying, deleting, creating, ...
- Role-based access control and ACLs for users.

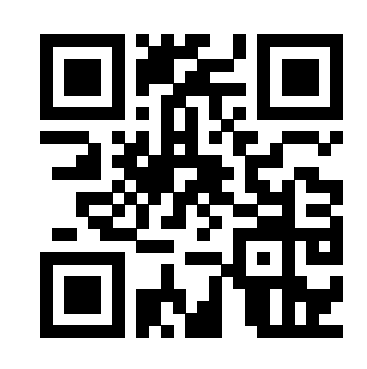


CaosDB is OpenSource!

License: AGPLv3, sources: <https://gitlab.com/caosdb>



Code review and security audits by IndiScale GmbH, which also provides CaosDB distribution "LinkAhead".

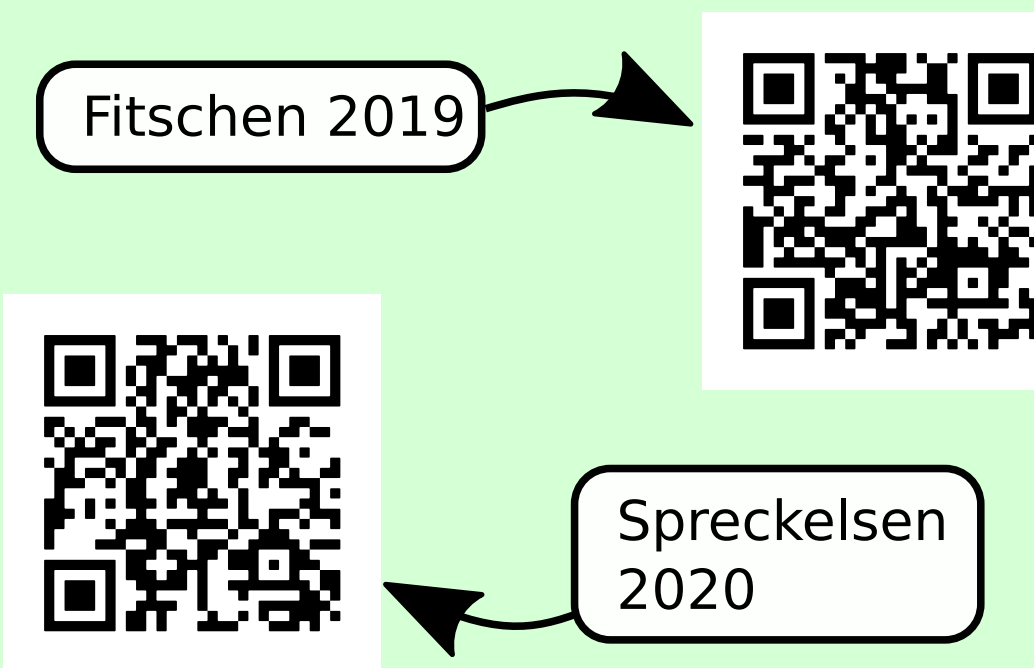


Flexible file backend

- CaosDB does not natively store file content, only **references to files**.
- **Interoperable** with other software.
- Use existing data as-is.
- Scientists can **continue to use their workflows**.
- [WIP] **Abstract object store** for repositories, S3 buckets, ...

Publications

1. CaosDB — Research Data Management ..., *Fitschen et al.*, Data, 2019
2. Guidelines for a Standardized ..., *Spreckelsen et al.*, Data, 2020



Automated data crawler

- Framework for **automated data integration**:
- Find new or changed raw data.
- Simplifies integration with **existing systems**.
- Based on Python library.

