Semantic Data Management with LinkAhead

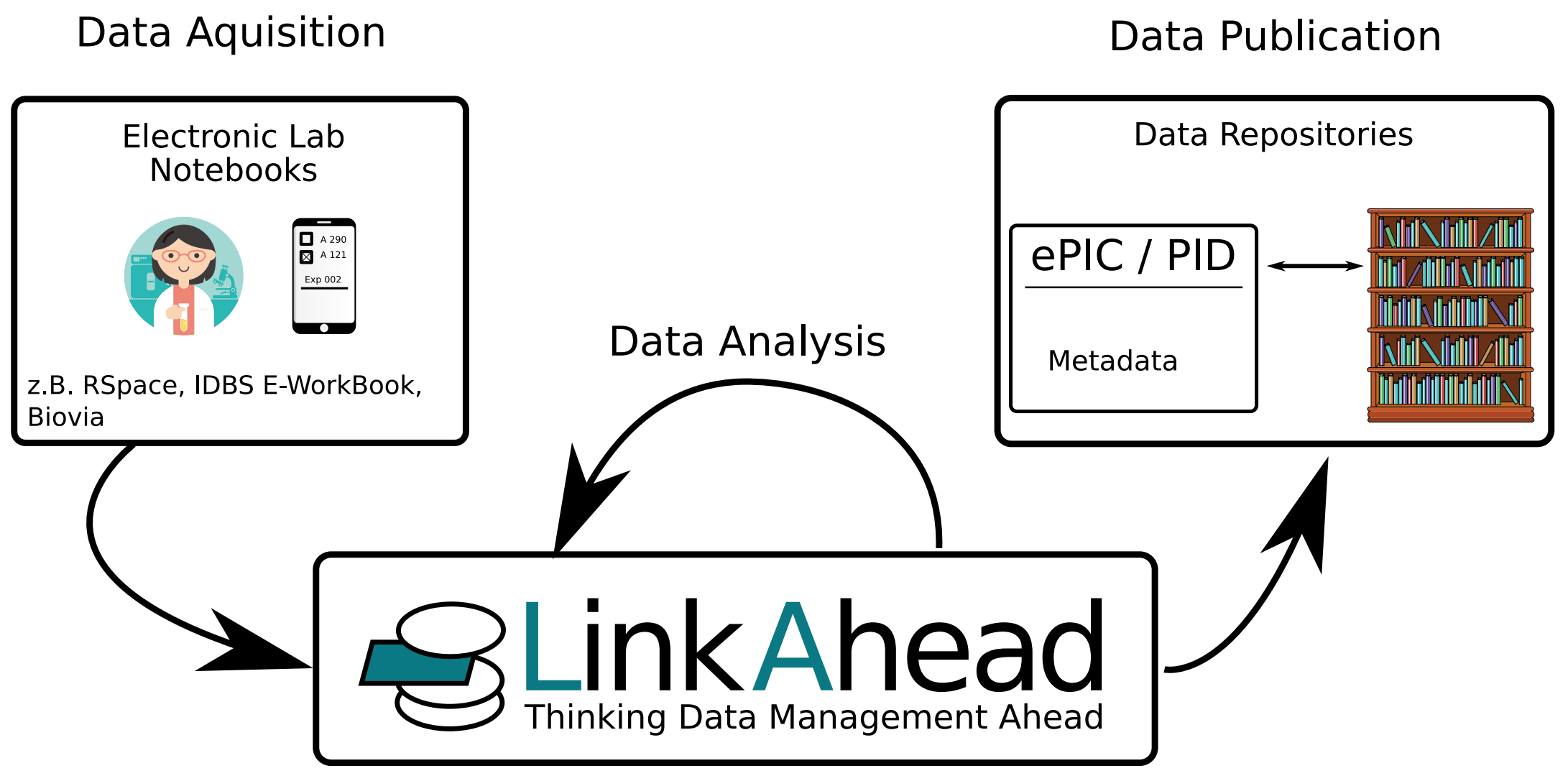
H. tom Wörden[1], A. Schlemmer[1, 2, 3], T. Fitschen[1], D. Hornung[1]

[1]Indiscale GmbH, Göttingen
[2]Max Planck Institute for Dynamics and Self-Organization, Göttingen
[3]German Center for Cardiovascular Research (DZHK), Partner Site Göttingen

h.tomwoerden@indiscale.de

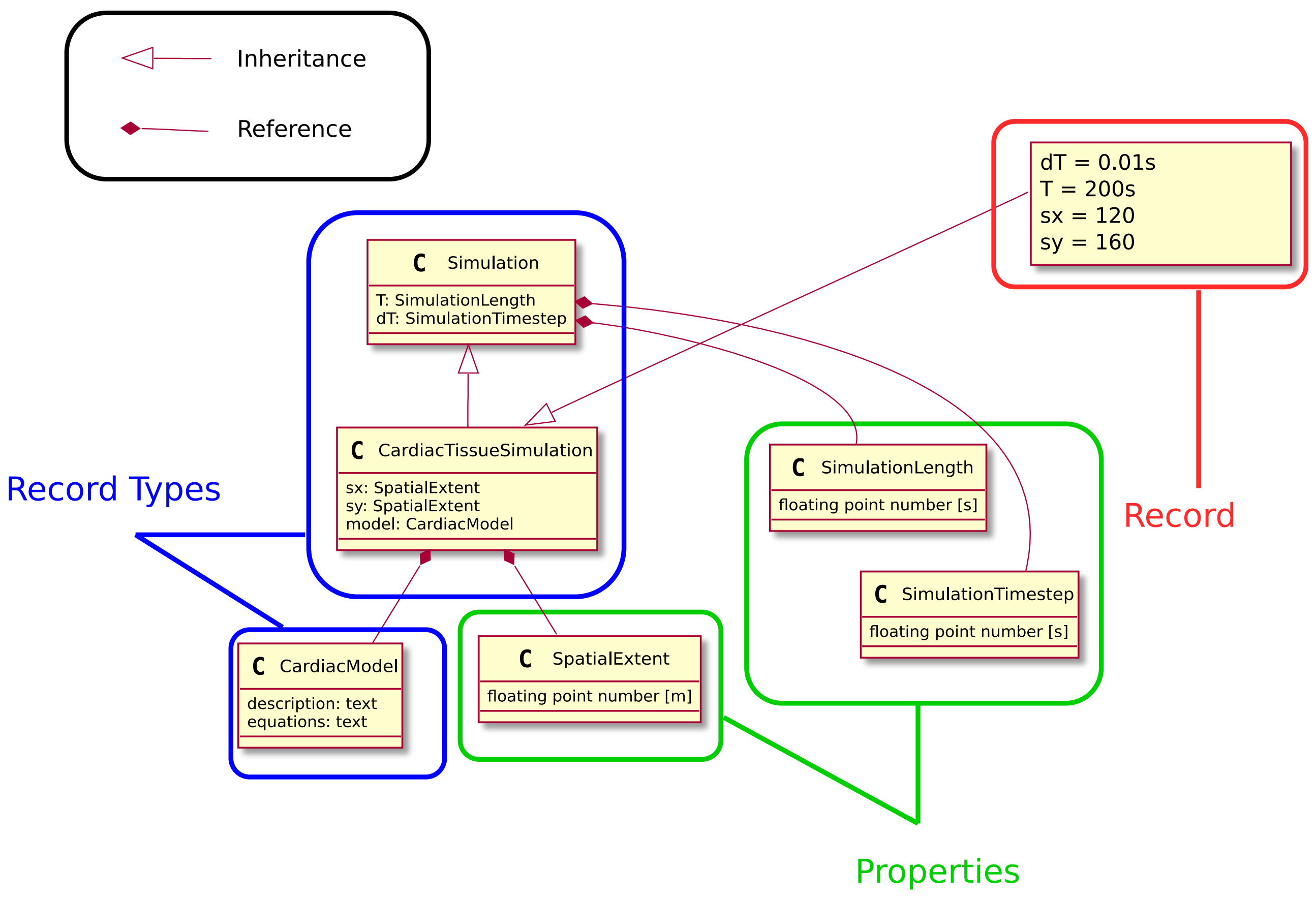
Abstract

Although scientific data has been used in digital formats for a long time, the data management is in many cases flawed and highly ineffective:
In many scientific workgroups data files are spread over many devices, hidden in impractical directory tree structures and rarely sufficiently documented or annotated with metadata. Concepts for overcoming these problems, like the FAIR data principles, receive a lot of attention, but practical solutions for data workflow management are far from commonly implemented. Here we propose a data workflow management based on LinkAhead which is able to handle big amounts of complex data. The versatile semantic data model maps various data sources and data structures, such as data from different measurement devices or computer simulation data. In particular, the software includes a powerful and intuitive query language and a system for physical units.

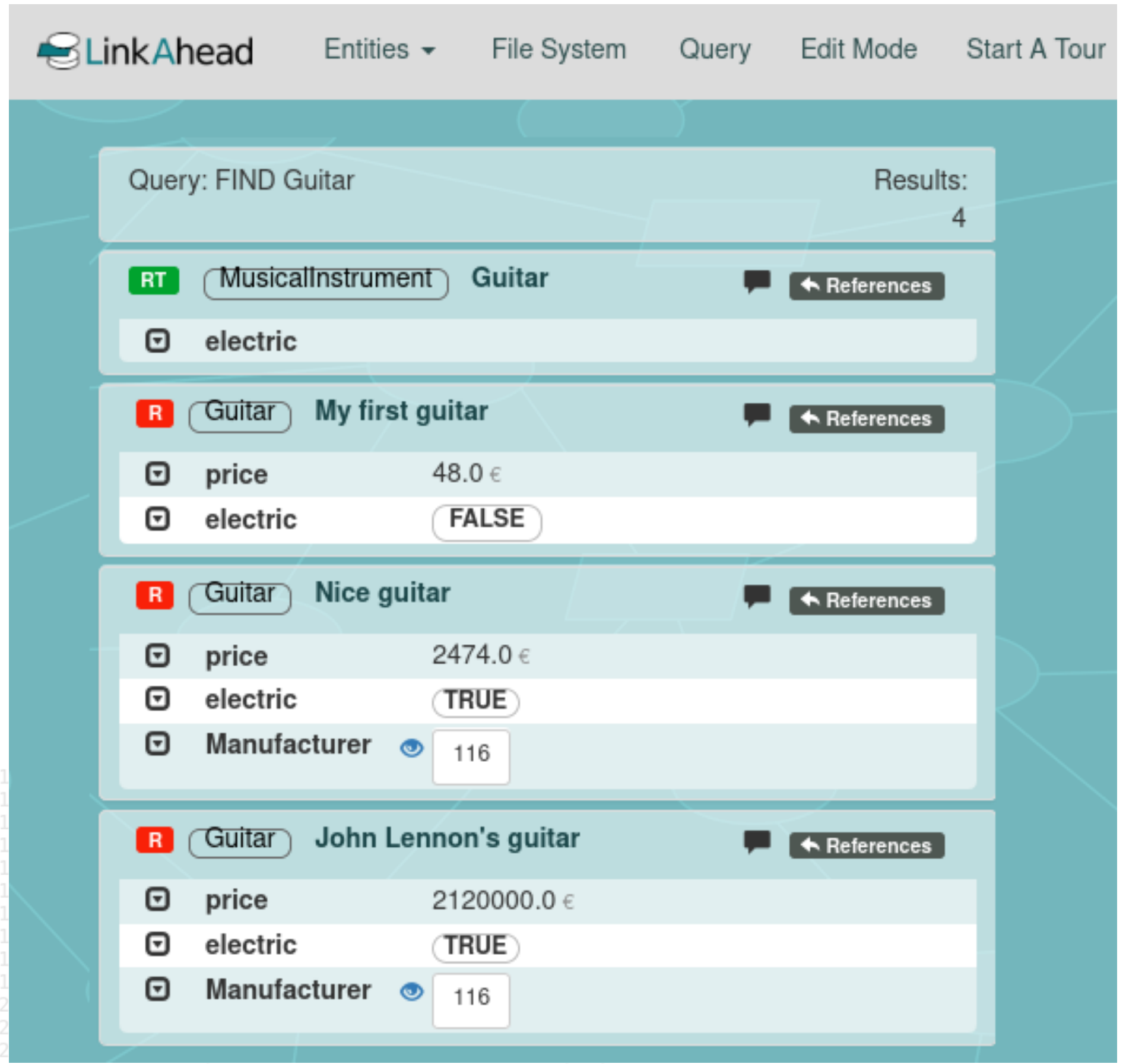


LinkAhead focuses on the daily work with data

LinkAhead does not attempt to change the way data is acquired (ELNs, Electronic Lab Notebooks), or the way data is published (Data Repositories), but concentrates on improving data management during data analysis. LinkAhead plays well



Web Interface



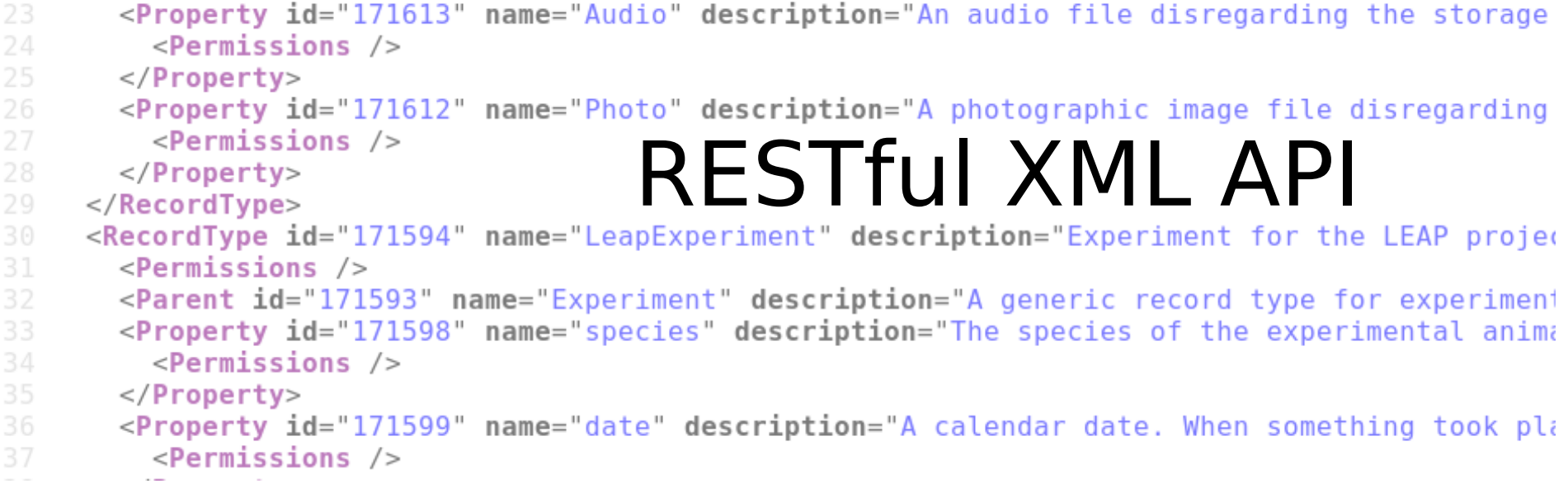
Query: FIND Guitar — Results: 4

RT MusicalInstrument Guitar
- electric

R Guitar My first guitar
- price 48.0 €
- electric FALSE

R Guitar Nice guitar
- price 2474.0 €
- electric TRUE
- Manufacturer 116

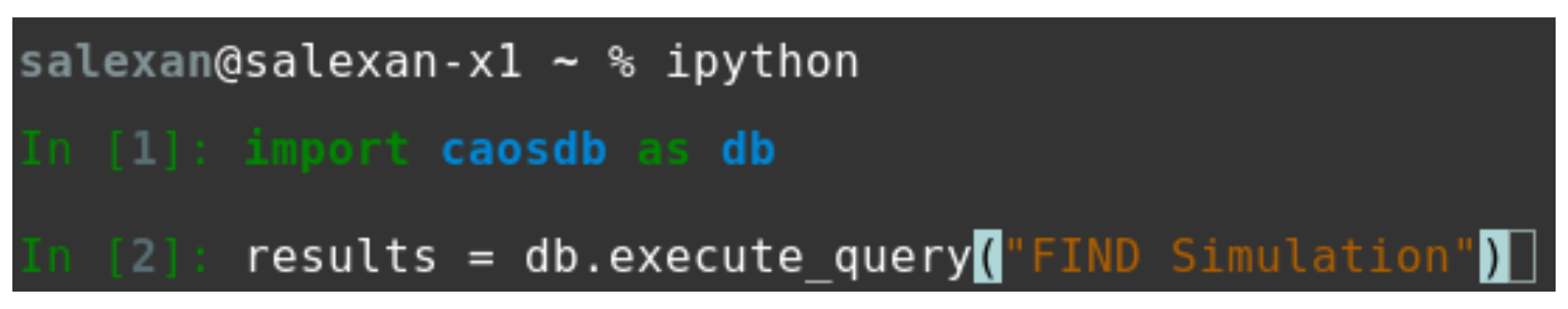
R Guitar John Lennon's guitar
- price 2120000.0 €
- electric TRUE
- Manufacturer 116



RESTful XML API

```
    <Property id="171613" name="Audio" description="An audio file disregarding the storage
      <Permissions />
    </Property>
    <Property id="171612" name="Photo" description="A photographic image file disregarding
      <Permissions />
    </Property>
  </RecordType>
  <RecordType id="171594" name="LeapExperiment" description="Experiment for the LEAP projec
    <Permissions />
    <Parent id="171593" name="Experiment" description="A generic record type for experiment
    <Property id="171598" name="species" description="The species of the experimental anima
      <Permissions />
    </Property>
    <Property id="171599" name="date" description="A calendar date. When something took pla
      <Permissions />
```

Python Client



```
salexan@salexan-x1 ~ % ipython
In [1]: import caosdb as db

In [2]: results = db.execute_query("FIND Simulation")
```

User Interfaces

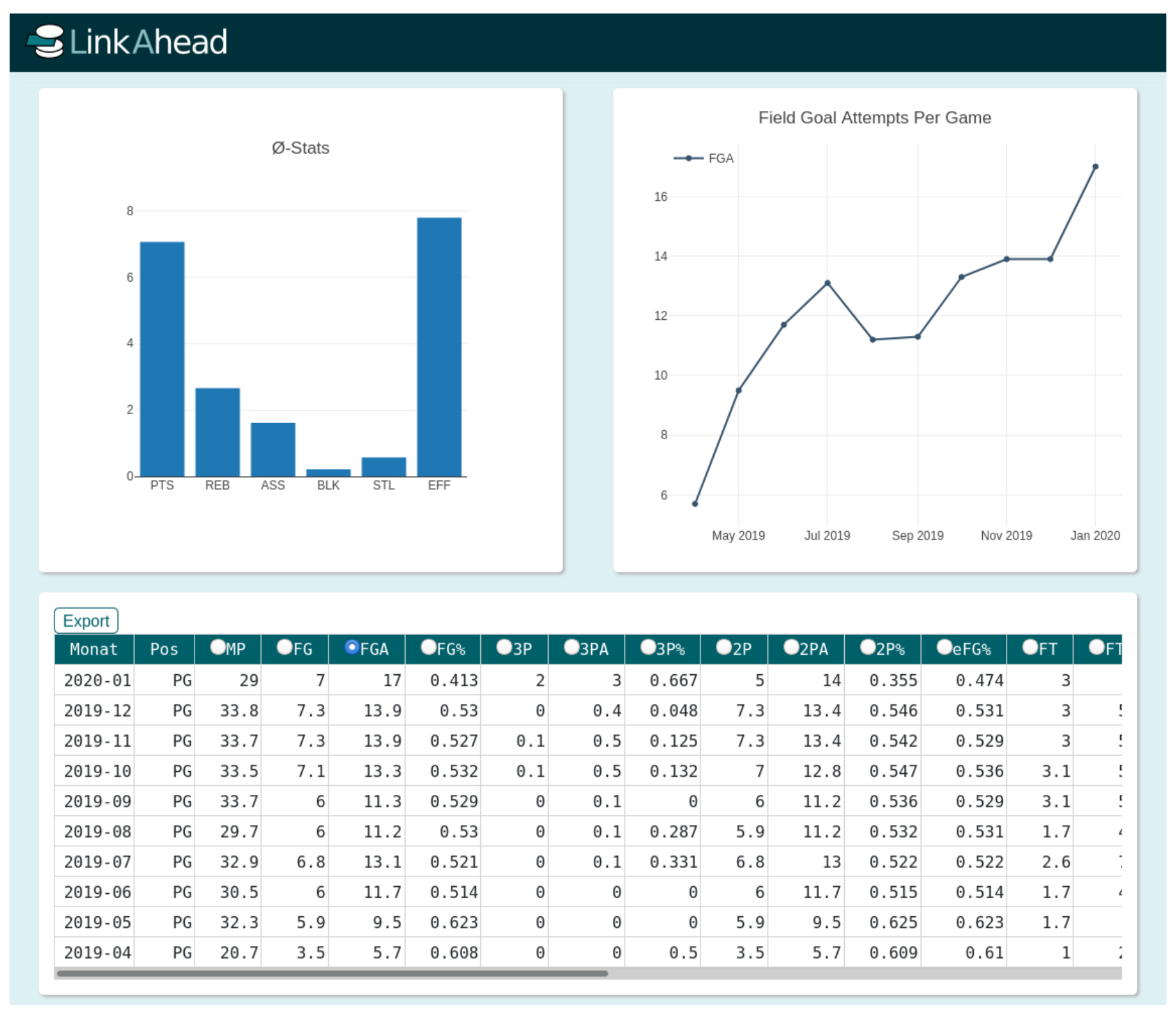
LinkAhead can be controlled using multiple user interfaces.

For high level data access, e.g. browsing, querying and viewing data, users access LinkAhead via the **WebUI** which is a web frontend to LinkAhead. It also integrates a graphical editor for the data model.

For data analysis and automation the **Python Client** is used.

All LinkAhead Clients communicate with the LinkAhead Server using the XML API.

Specialized Views



Monat	Pos	○MP	○FG	●FGA	○FG%	○3P	○3PA	○3P%	○2P	○2PA	○2P%	○eFG%	○FT
2020-01	PG	29	7	17	0.413	2	3	0.667	5	14	0.355	0.474	3
2019-12	PG	33.8	7.3	13.9	0.53	0	0.4	0.048	7.3	13.4	0.546	0.531	3
2019-11	PG	33.7	7.3	13.9	0.527	0.1	0.5	0.125	7.3	13.4	0.542	0.529	3
2019-10	PG	33.5	7.1	13.3	0.532	0.1	0.5	0.132	7	12.8	0.547	0.536	3.1
2019-09	PG	33.7	6	11.3	0.529	0	0.1	0	6	11.2	0.536	0.529	3.1
2019-08	PG	29.7	6	11.2	0.53	0	0.1	0.287	5.9	11.2	0.532	0.531	1.7
2019-07	PG	32.9	6.8	13.1	0.521	0	0.1	0.331	6.8	13	0.522	0.522	2.6
2019-06	PG	30.5	6	11.7	0.514	0	0	0	6	11.7	0.515	0.514	1.7
2019-05	PG	32.3	5.9	9.5	0.623	0	0	0	5.9	9.5	0.625	0.623	1.7
2019-04	PG	20.7	3.5	5.7	0.608	0	0	0.5	3.5	5.7	0.609	0.61	1

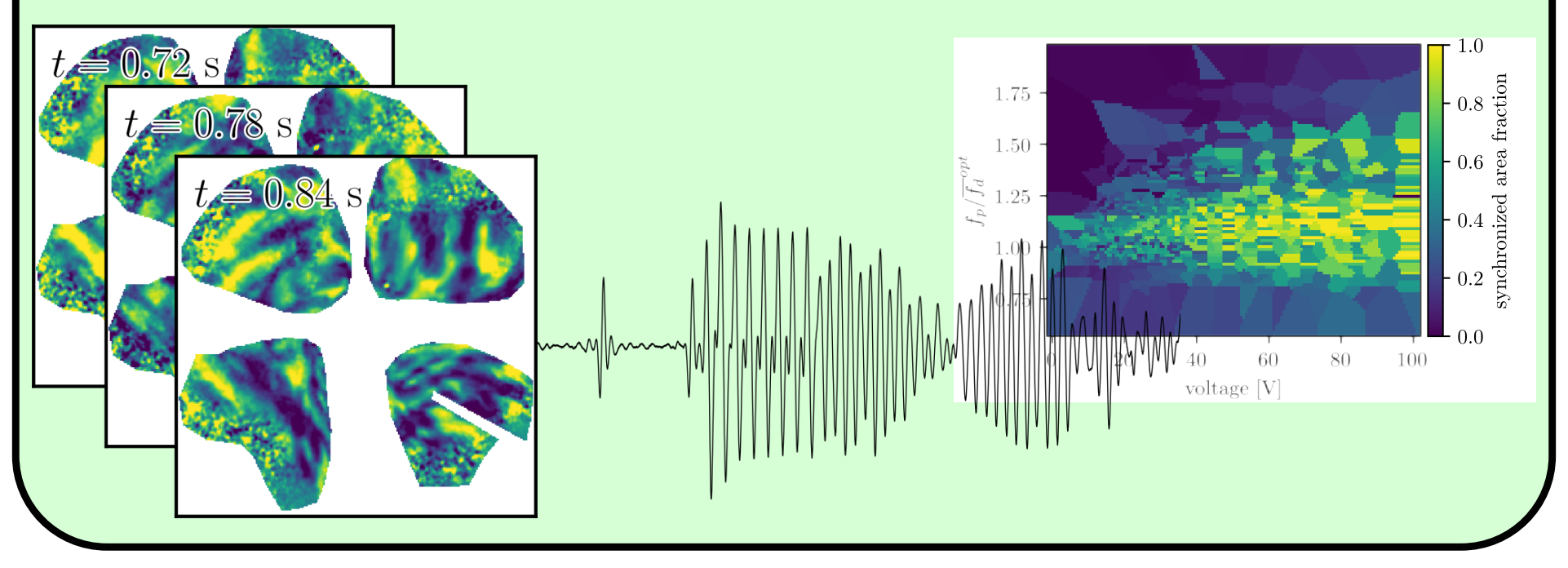
LinkAhead
Thinking Data Management Ahead

SQL Backend

LinkAhead makes use of robust technology to store its data and data model. Internally the data model and the data (except for files) are mapped into a relational database structure with an SQL backend (MariaDB).

File System

LinkAhead does not store the data files itself inside a database, but **indexes existing file systems**. This has several advantages:
- **Interoperability** with other software operating on the file system
- Existing data can be directly used
- **No change in established scientific workflows required**

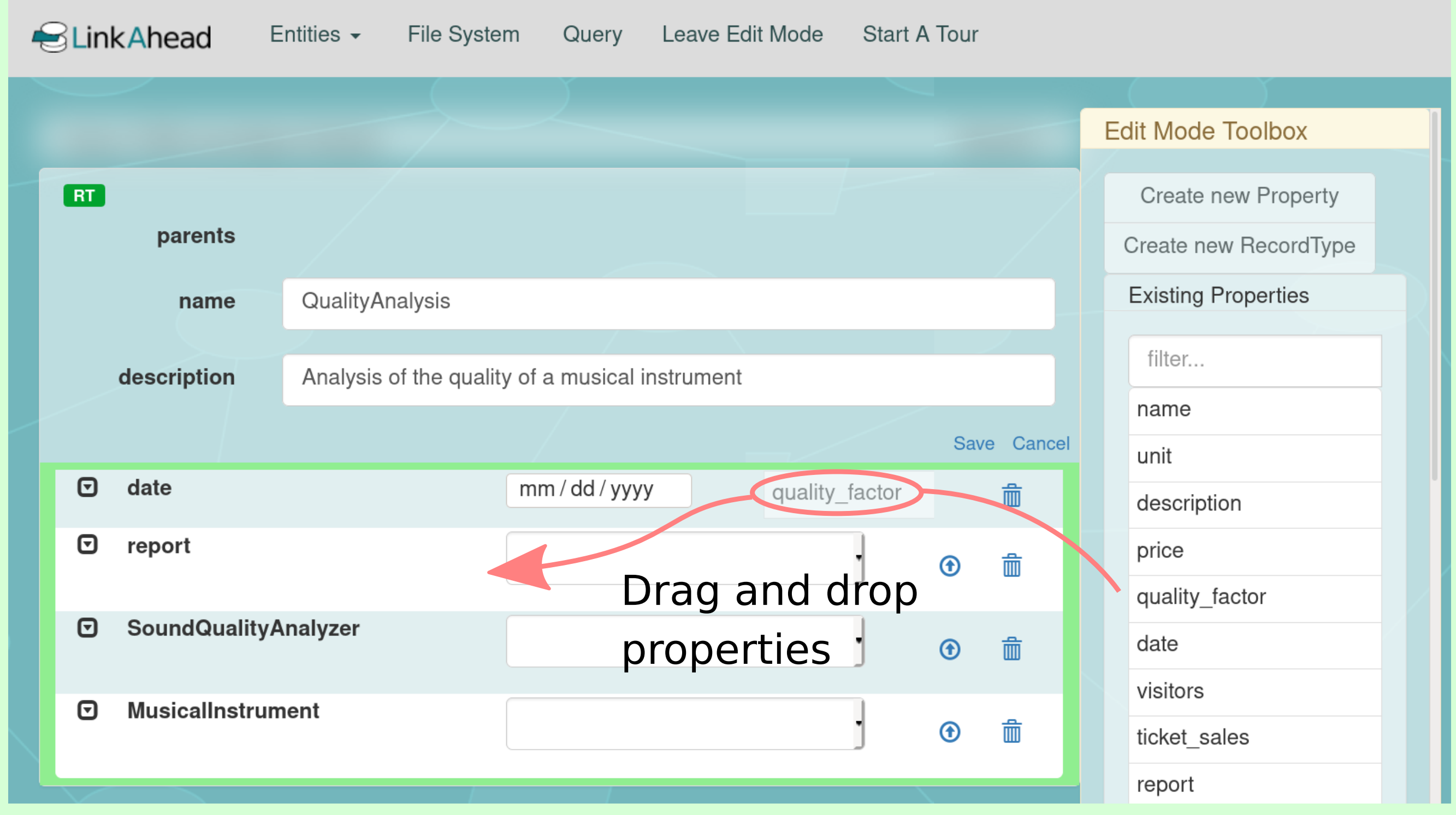


Flexible Data Model

LinkAhead implements a flexible semantic data model. Flexible means, that the model can be changed, adapted, improved at any time.The semantic model used by LinkAhead is similar to an object oriented model and based on three basic concepts:

- **Properties** describe the basic variables which are stored in the database. They have a name and a data type which can be a base type (INTEGER, TEXT, ...) or a REFERENCE to another RecordType.
- **RecordTypes** combine multiple Properties and allow for inheritance. They are similar to classes in object oriented programming.
- **Records** store the actual data by setting values required by their RecordType.

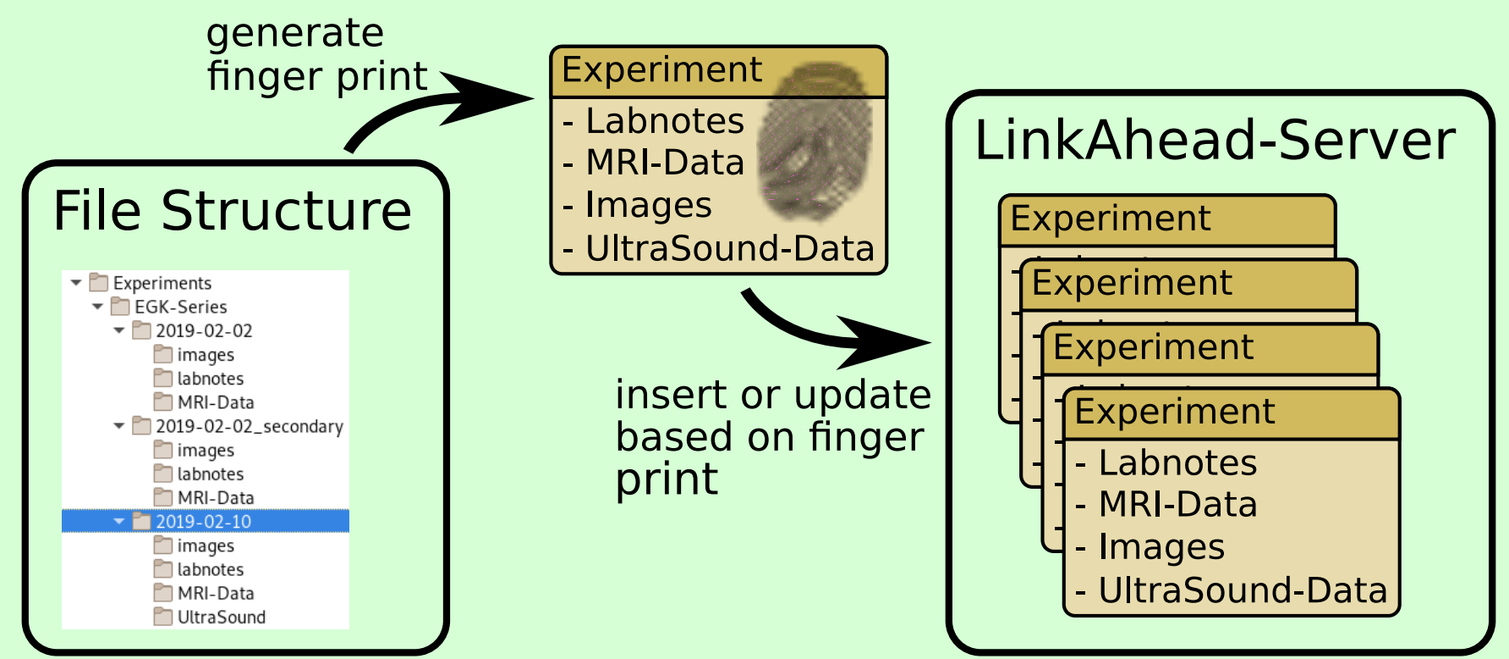
The Data Model can be changed in the Webinterface



RT
- parents
- name: QualityAnalysis
- description: Analysis of the quality of a musical instrument
- date: mm / dd / yyyy
- report
- SoundQualityAnalyzer
- MusicalInstrument

Automatic Data Insertion

A **crawler** that has to be adapted to the requirements of the use case can periodically run and automatically index files on the file system.
LinkAhead stores links to files and file hashes in order to be able to check the file system for consistency.



LinkAhead is OpenSource!

LinkAhead is the professionally maintained version of the community project CaosDB. The license is AGPLv3. The concepts were published in the journal data.

Article in data

Sourcecode on gitlab.com

Demo

IndiScale GmbH
Philipp-Reis-Straße 2a
37075 Göttingen

Geschäftsführung: Henrik tom Wörden
Sitz: Göttingen

Ust-ID: DE 325018726
Registergericht: AG Göttingen
Registernummer: HRB 205721

LinkAhead
Thinking Data Management Ahead