

# Building scientific practices with LLMs in KM3NeT

## Collaboration needs

### Stale Code & Documentation

Years of collaborative work described in distributed system and poorly documented software have become hard to manage and understand.

*We need a clear information architecture, code reviews and documentation.*

### Scientific Practices

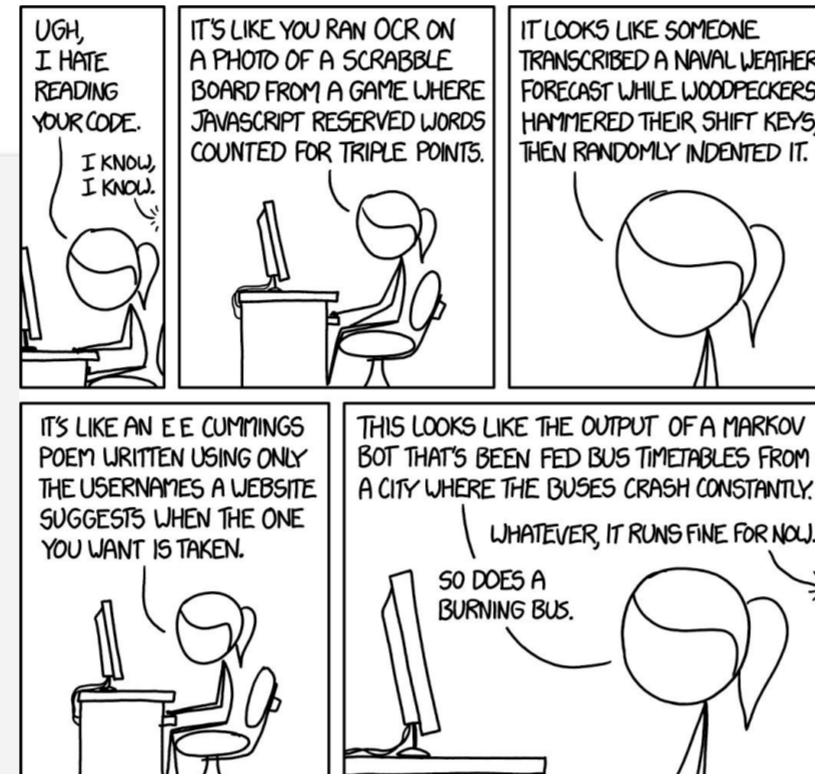
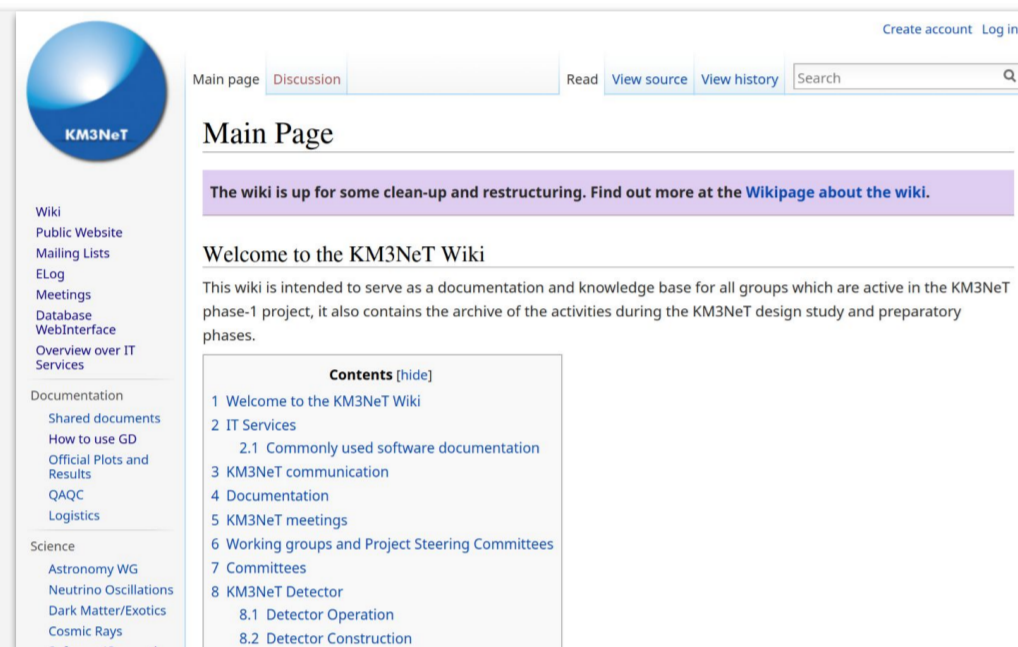
Sharing only high-level analyses outcomes as publications or conference contributions fails to incentivize the practice to invest time in reproducibility and learning and following good coding standards.

*We need easy guidelines and tools to produce and preserve scientific work.*

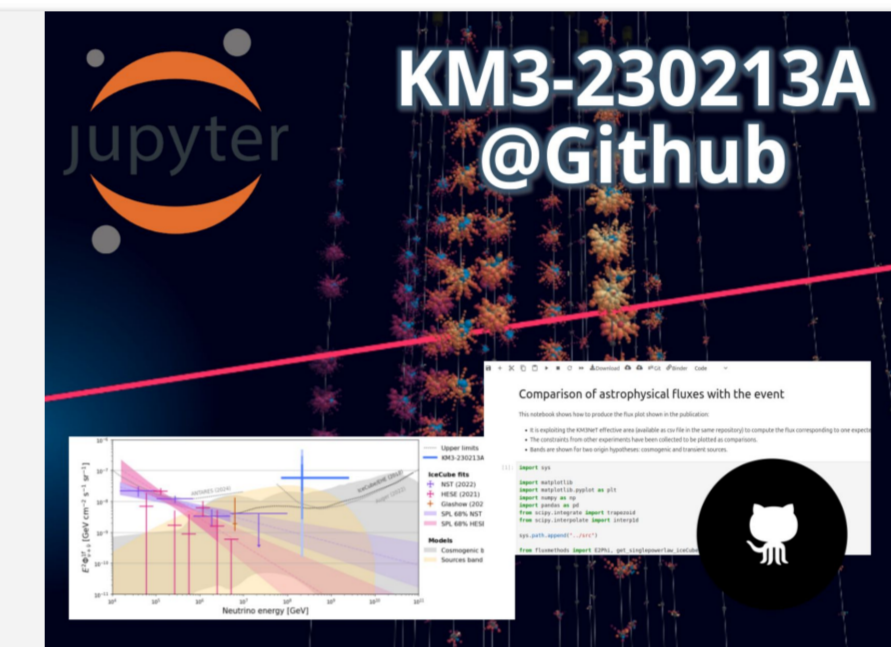
### FAIRification

Public workflows and data are issued on a analysis-specific basis which allows reproduction of some results but does not allow for generalization.

*We need ways to produce multi-purpose tools for open science and easy interfaces to our scientific results.*



[xkcd.com/1695](https://xkcd.com/1695)



## Toolbox

### Information Access

The **infobasis** python packages offers a toolbox to query ICT services of the collaboration, bulk download documents and websites from the services and store them in searchable sqlite databases.

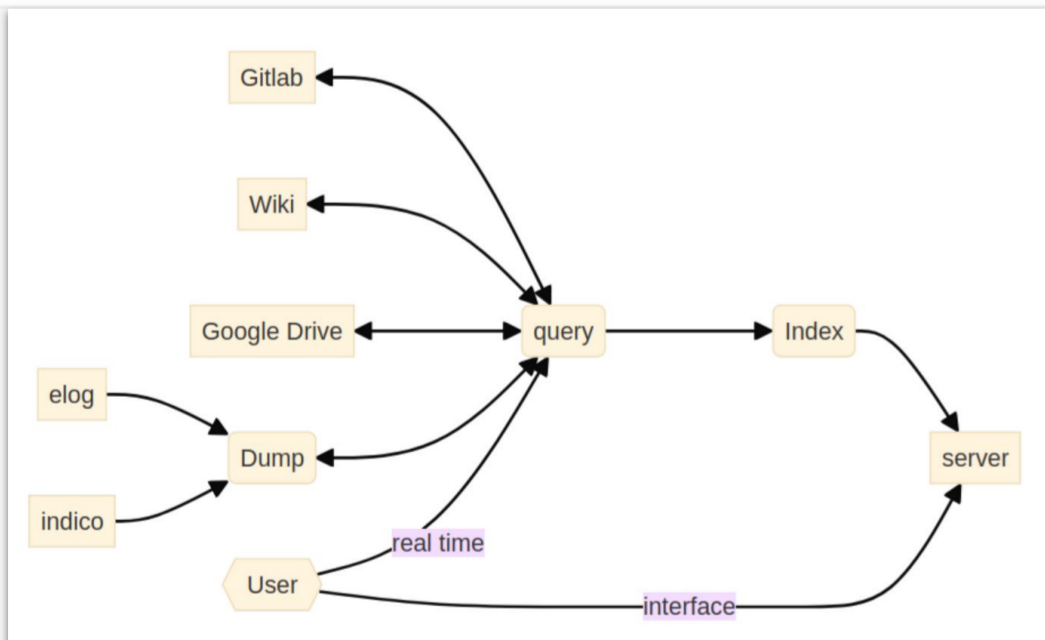
Plugins for new servers with dedicated access functions as well as custom database tables can be created easily.

### Processing and Optimizing

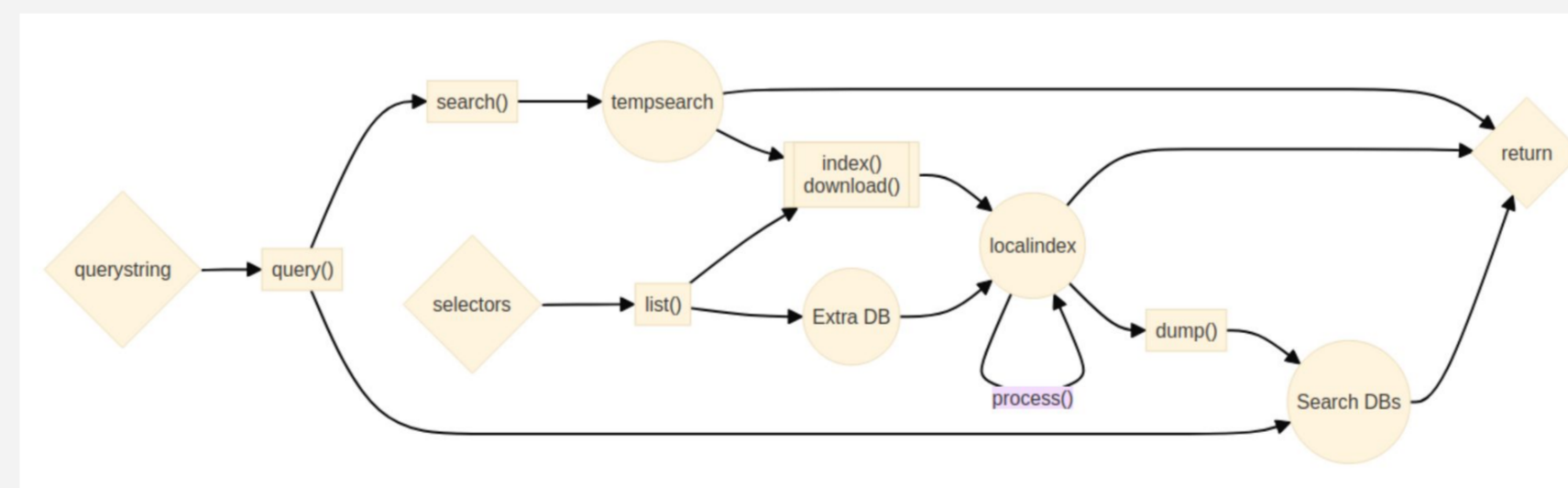
Using the infobasis package, the **LLMTuner** python package implements data retrieval pipelines for the generation of (semantic) databases for RAG-enhanced LLM use and a testing and benchmarking environment for LLM optimization.

### Interface and LLMs

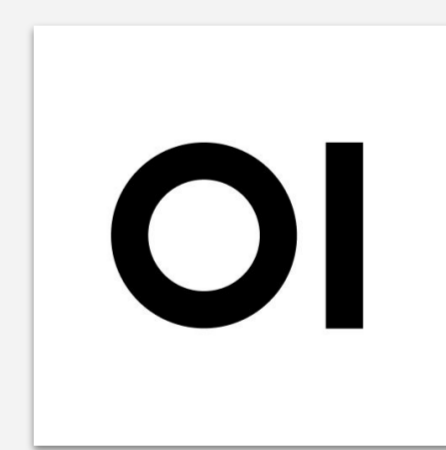
The LLMTuner interfaces are used to integrate and evaluate tools and LLMs using OpenWebUI and Ollama to provide a user-friendly environment and flexible testing of state-of-the-art models and methods.



[git.km3net.de/jschnabel/infobasis](https://git.km3net.de/jschnabel/infobasis)



[git.km3net.de/jschnabel/llmtuner](https://git.km3net.de/jschnabel/llmtuner)



## Application

### Code Review & Documentation

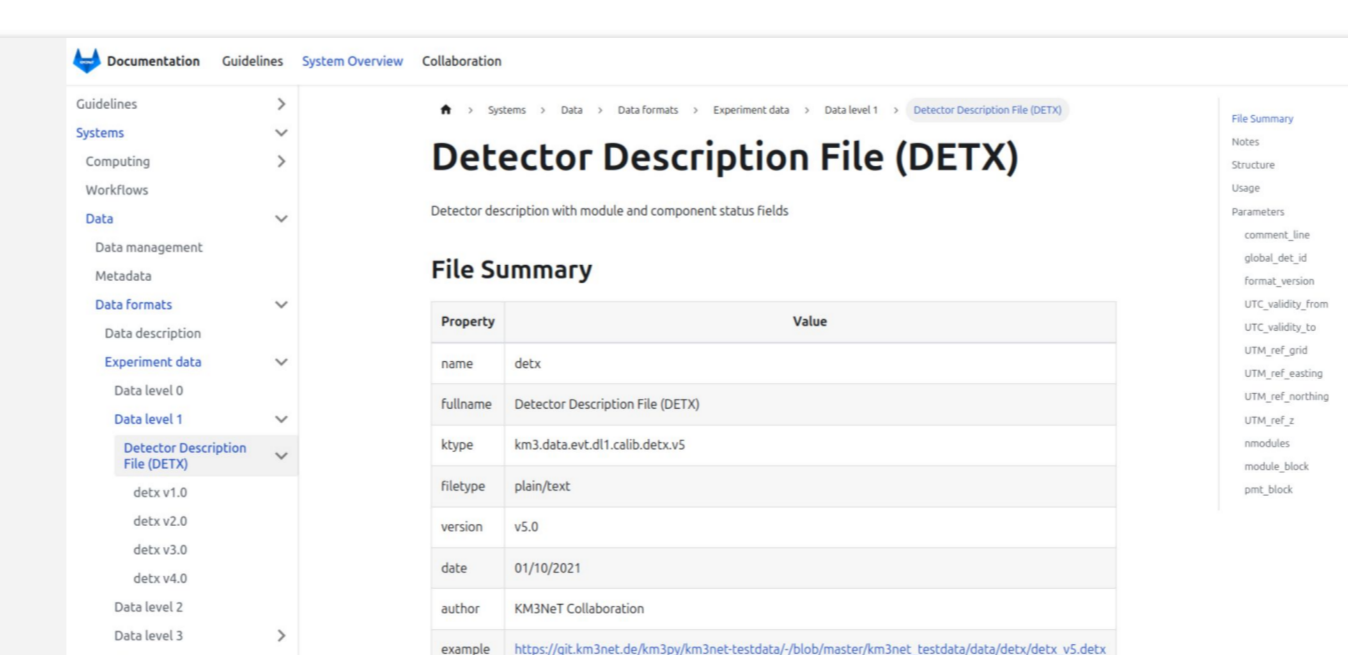
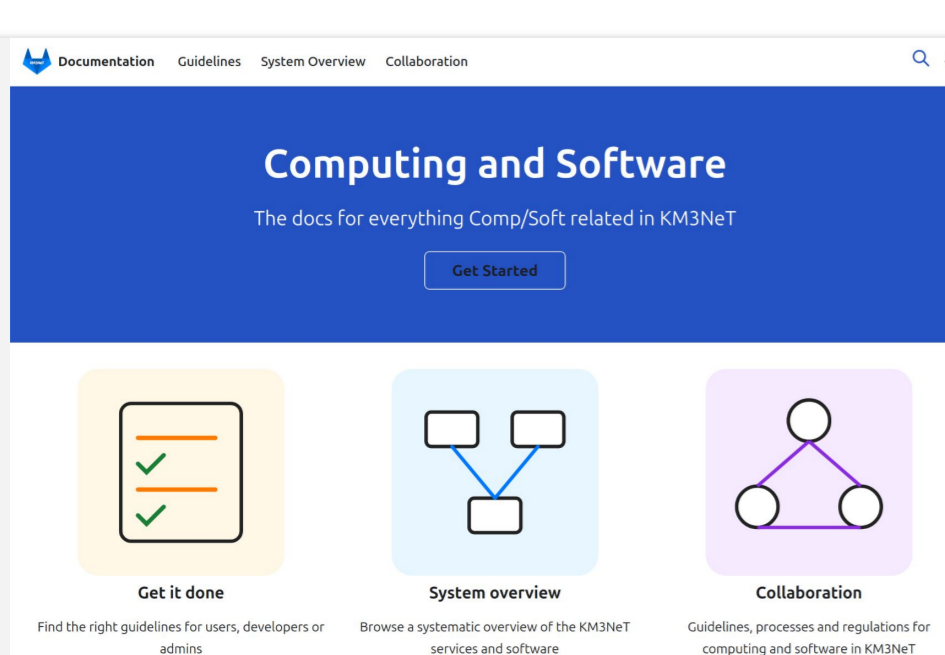
- ✔ Automated synchronization of documentation from other sources using infobasis
- ✔ LLM processing to enhance code style and documentation
- ✔ Easy re-formatting of structured information using LLMs

### Scientist & Developers Support

- ! Dedicated LLM setup for collaboration specific search on best practices and standards
- ! LLM support for code clean-up and migration projects

### Open Interfaces

- ✔ Information retrieval for scientific workflows from km3net interfaces using infobasis (*openkm3 python package*)
- ! Enhanced interfaces for integrated multi-messenger analyses using LLMs



[git.km3net.de/open-data/openkm3](https://git.km3net.de/open-data/openkm3)