## MUSICS



### Machine Learning based User friendly System Interface Creating Services for Open Science

Kilian Schwarz, Sven Karstensen, Marcel Stanitzki, Bernhard Humm DESY, FH SciComp Workshop, July 2 2024





## **Table of contents**

- 01 PIs and institutions
- 02 project idea

- 05 LLM based UI
- 06 Scientific impact

03 on site experiments at DESY

**08** Summary and Outview

04 project description

## MUSICS Pls and institutions

## **PIs and institutions**

### DESY

- Dr. Kilian Schwarz, IT/Scientific Computing
- Sven Karstensen, FTX/AST
- Dr. Marcel Stanitzki, FTX/TBT

### An application to OSCARS 1<sup>st</sup> Open Call

- Submitted March 15, 2024
- Notification of outcome: July, 2024
- Potential project run time: 2 years
- Potential funding: up to 250 k€

### **Hochschule Darmstadt**

• Prof. Dr. Bernhard Humm

# MUSICS project idea



#### **Project idea**

- Develop a user friendly service for publication and retrieval of Open Data
- Target: small experiments and test beam community at DESY
- Service shall be operable via a large language model based user interface
  - Users will be able to "talk" to the service requesting the required actions
- MUSICS combines two currently hot topics in science:
  - Open data & FAIR data management
  - Machine learning and LLMs



https://www.istockphoto.com/

## MUSICS on site experiments at DESY



### **On site experiments**

- Initially the open data services will be developed for ALPS II and the DESY II Test Beam Facility
- Next user communities:
  - MADMAX, BabylAXO, LUXE
  - Generalisation to the scientific community at large



## Schematic view of ALPS IIc experiment at DESY



## DESY. Kilian Schwarz | MUSICS | FH SciComp Workshop | July 2 2024

## On site experiments

- Any Light Particle Search II
- Light-shining-through-a-wall experiment
- Searching for axions and axion like particles (light particles as dark matter candidates)
- Photons traverse strong magnetic field, potentially undergoing conversion to axion, can thus travel through wall, inverse process again in strong magnetic field

 $\gamma$   $g_{a\gamma\gamma}$   $g_{a\gamma\gamma}$   $g_{a\gamma\gamma}$   $g_{a\gamma\gamma}$   $g_{a\gamma\gamma}$   $g_{a\gamma\gamma}$   $g_{a\gamma\gamma}$   $g_{a\gamma\gamma}$   $g_{a\gamma\gamma}$  $g_{a\gamma\gamma}$ 

Feynman diagram of the photon-axion / axion-photon coupling in ALPS II

## **On site experiments**

**DESY II Test Beam Facility** 

- DESY II Test Beam Facility offers three beam lines with electrons ranging between 1 and 6 GeV
- Each beam line can be operated independently
- 400-700 users from all over the world
- Data per test beam period from few gigabytes to ten terabytes



Test beam usage divided in several categories

## MUSICS project description

## **MUSICS**

## **Project description**

#### **User Requirements**

- Streamlined user interface
- Structured storage space
- Standardised metadata
- Metadata catalogues
- Simple data access



https://www.freepik.com

## **Supported Workflows**

- Data publication
  - Metadata creation via LLM based UI
  - Experiment specific input masks
  - Consistency checks
  - Data transfer behind the scenes
  - Assigning data sets with DOIs
  - Linking of data and publications
- Data access and processing
  - search patterns via plain text
  - Internal translation into metadata catalogue requests, URLs, DOIs, access rights
  - Data transfer behind the scenes

## **MUSICS**

### **Project description**

#### Implementation ideas

- Entry point to MUSICS: central web page with interface to LLM
- This will be combined with already existing components at DESY
  - Central storage space based on dCache
    - Clear directory structure for open data
    - Dedicated doors for internal and external data access
  - SciCat based metadata catalogue
  - Helmholtz AAI

### Implementation ideas

- Main development work: LLM
  - User can enter requirements in plain text and will receive a description how to access the wanted data sets together with links to corresponding compute and storage resources
  - Metadata generation should be automatic via Al support
  - All existing information from hardware and software supplemented by manual inputs should be used
  - Will start with a wide set of LLM models (e.g. GPT-4, BERT, NLP, ...)
  - Model selection will be an important step in the project
  - Data training initially on Maxwell Cluster

## MUSICS Scientific impact



#### **Scientific impact**

- All initial user communities are small to medium size collaborations which could profit significantly from MUSICS.
- By getting the user communities involved already in design stage experiment expertise can be directly incorporated
- Start communities are ALPS II and DESY test beam facility, then MADMAX, BabyIAXO, LUXE
- After that generalisation to scientific community at large
- Project will follow principles of open science and FAIR principles so that MUSICS can be easily adopted by other sites
- Easy access and searchability of data-sets are key ingredients for forming a successful community
- MUSICS will lower the threshold for open data usage and form a foundation for dedicated projects



shutterstock.com · 2303513463

DESY. Kilian Schwarz | MUSICS | FH SciComp Workshop | July 2 2024

## MUSICS Summary and Outview



#### **Summary and Outview**

- The MUSICS consortium applied for funding at the first OSCARS Open Call
- MUSICS plans to provide a user friendly Open Data service with LLM supported UI
- MUSICS has the potential to lower the threshold for Open Data usage especially for small and medium sized experiments



https://www.istockphoto.com/

## Thank you

#### Contact

Deutsches Elektronen-Synchrotron DESY

www.desy.de

Kilian Schwarz IT/Scientific Computing Kilian.schwarz@desy.de Phone 040 8998 2596