

## KM3NeT INFRADEV – H2020 – 739560

### **Manuals for all major components of the open data access system**

**KM3NeT-INFRADEV GA DELIVERABLE: D4.9**

Document identifier:	<b>KM3NeT-INFRADEV-WP4-D4.9_v0.3</b>
Date:	<b>04/12/2020</b>
Work package:	<b>WP4</b>
Lead partner:	<b>FAU</b>
Document status:	<b>For endorsement by IB</b>
Dissemination level:	<b>Public</b>
Document link:	

#### Abstract

The KM3NeT Research Infrastructure will, over a period of at least a decade, produce a large amount of unique scientific data that are to be made available to the scientific communities concerned and to the broader general public. This requires the set-up of tools, procedures, documentation and rules to provide this service. In this report, short descriptions of, and the manuals for, all major components of the open data access system of KM3NeT are collected.

## I. Copyright notice

Copyright © KM3NeT Collaboration

## II. Delivery slip

	Names	Partner/WP	Date
Author(s)	T. Gal, S.R. Gozzini, K. Graf, S. Hallmann, J. Hofestädt, S. Raab, J. Schnabel Z. Aly, F. Huang R. G. Ruiz D. Stavropoulos	FAU  CNRS FOM NCSR-D	04/12/2020
Approved by			

## III. Document log

Issue	Date	Comment	Author/Partner
0.1	14/11/2020	First draft	T. Gal, S.R. Gozzini, K. Graf, S. Hallmann, J. Hofestädt, S. Raab, J. Schnabel / FAU Z. Aly, F. Huang / CNRS R. G. Ruiz / FOM D. Stavropoulos / NCSR-D
0.2	23/11/2020	Comments included by K. Graf, U. Katz	K. Graf, U. Katz / FAU
0.3	04/12/2020	Internal review, comments included by C. James	C. James / Curtin Univ.

## IV. Application area

This document is a deliverable for the grant agreement of the project, applicable to all members of the KM3NeT-INFRADEV project, beneficiaries and third parties, as well as its collaborating projects.



Author(s) Z. Aly, T. Gal, S.R. Gozzini, K. Graf, S. Hallmann, J. Hofestädt, F. Huang, S. Raab, R. G. Ruiz, J. Schnabel, D. Stavropoulos  
document KM3NeT-INFRADEV-WP4-D4.9  
Version 0.3 Release date:

KM3NeT 2.0 – 739560  
WP 4  
Public



## V. Terminology

<b>ADQL</b>	Astronomical Data Query Language
<b>ANTARES</b>	Astronomy with a Neutrino Telescope and Abyss environmental RESearch project
<b>API</b>	Application Programming Interface
<b>ARCA</b>	Astroparticle Research with Cosmics in the Abyss
<b>DaCHS</b>	GAVO Data Centre Helper Suite
<b>ESFRI</b>	European Strategy Forum on Research Infrastructures
<b>ODC</b>	Open Data Centre
<b>ORCA</b>	Oscillation Research with Cosmics in the Abyss
<b>OSP</b>	open science Portal
<b>REST-API</b>	Representational State Transfer – Application Programming Interface
<b>TAP</b>	Table Access Protocol
<b>VO</b>	Virtual Observatory

## VI. List of figures

Figure 1: Overview for the Open Science System.....13

## VII. List of tables

None

## VIII. Project summary

KM3NeT is a large research infrastructure that will consist of a network of deep-sea neutrino telescopes in the Mediterranean Sea with user ports for Earth and Sea sciences. Following the appearance of KM3NeT 2.0 on the ESFRI road map 2016 and in line with the recommendations of the Assessment Expert Group in 2013, the KM3NeT-INFRADEV project addresses the Coordination and Support Actions (CSA) to prepare a legal entity and appropriate services for KM3NeT, thereby providing a sustainable solution for the operation of the research infrastructure during ten (or more) years. The KM3NeT-INFRADEV is funded by the European Commission's Horizon 2020 framework and its objectives comprise, among others, the preparation of Open Data Access (work package 4).



## IX. Executive summary

The KM3NeT Research Infrastructure will, over a period of at least a decade, produce a large amount of unique scientific data that are to be made available to the scientific communities concerned and to the broader general public. This requires the set-up of tools, procedures, documentation and rules to provide this service. In this report, short descriptions of, and the manuals for, all major components of the open data access system of KM3NeT are collected. Although the manuals and documentation is directly available at all services within the open-access data system, a collection in this place will help the user with a single entry point on how to use the system. An up-to-date version of all links can be found at [openscience.km3net.de/help](https://openscience.km3net.de/help).

## X. Table of Contents

I. Copyright notice.....	2
II. Delivery slip.....	2
III. Document log.....	2
IV. Application area.....	2
V. Terminology.....	3
VI. List of figures.....	3
VII. List of tables.....	3
VIII. Project summary.....	3
IX. Executive summary.....	4
X. Table of Contents.....	4
1. Introduction.....	6
2. Manuals for users of the KM3NeT open science system.....	6
2.1. Manuals for the servers and platforms.....	6
2.1.1. Summary.....	6
2.1.2. VO server.....	7
2.1.3. KM3NeT Open Data Centre.....	7
2.2. Manuals for software.....	7
2.2.1. KM3NeT software packages.....	8
2.2.2. External software.....	8
2.3. Integration examples and workflows.....	9
2.3.1. Notebooks.....	9
2.3.2. Courses and tutorials.....	9



- 3. Manuals for KM3NeT members..... 10**
  - 3.1. How to develop software..... 10
    - 3.1.1. Python project template..... 10
  - 3.2. Standardizing the scientific workflow..... 10
    - 3.2.1. Public plot template..... 11
    - 3.2.2. Analysis template..... 11
  - 3.3. Providing data on the open platforms..... 11
    - 3.3.1. Annotating and uploading data..... 11
    - 3.3.2. Maintenance of the open science servers and platforms..... 11
- Appendix I: Architecture overview..... 12



# 1. Introduction

There are two aspects of the manuals for the KM3NeT open science system. On the one hand, manuals are needed for the external user to understand the use of the data and the interfaces, on the other hand, manuals have to be provided for KM3NeT members to actually contribute their software and data to the open science system.

However, as documentation to an evolving system - as covered in this deliverable - is valid only for a short time, this document aims at giving an overview of the different sustained sources of documentation and point to the actual currently valid versions, rather than reproducing the full content of the references.

Note that most documentation for software is built automatically from in-source documentation and human-created documentation content is primarily hosted on the KM3NeT Gitlab instance, as described in Deliverable 4.8. As the KM3NeT solution builds also on third-party software, links to the available documentation are reported here.

An up-to-date version of all links can be found at [openseience.km3net.de/help](https://openseience.km3net.de/help). For a further description of the Open Science System, see Deliverable 4.8 and the

## 2. Manuals for users of the KM3NeT open science system

For further description of the individual components, see Deliverable 4.8 or the [open science Portal \(OSP\)](#).

### 2.1. Manuals for the servers and platforms

#### 2.1.1. Summary

platform	documentation URL	content
open science Portal	<a href="https://open-data.pages.km3net.de/openseienceportal/articles/getting-started/">https://open-data.pages.km3net.de/openseienceportal/articles/getting-started/</a>	Main entry point: Quickstart to the system
Education Portal for Training Courses	<a href="https://edu.km3net.de/using-the-education-portal/">https://edu.km3net.de/using-the-education-portal/</a>	Info on access options to courses
KM3NeT Open Data Centre	<a href="https://open-data.pages.km3net.de/openserver/">https://open-data.pages.km3net.de/openserver/</a>	ODC Documentation
Virtual Observatory server	<a href="http://vo.km3net.de/static/help.shtml">http://vo.km3net.de/static/help.shtml</a>	Start page for Help
Gitlab Software Development and	<a href="https://git.km3net.de/help">https://git.km3net.de/help</a>	Built-in gitlab help system & documentation



## Integration Server

Docker  
Containerisation  
System <https://docs.docker.com/>

Getting started,  
documentation &  
Tutorials

### 2.1.2. VO server

The DaCHS server comes with auto-generated documentation to the URLs serving as end points for the services and numerous auto-generated listings of metadata, in addition to links to interface software and tutorials on their use.

Topic	documentation URL	content
TAP service	<a href="http://vo.km3net.de/__system__/tap/run/info">http://vo.km3net.de/__system__/tap/run/info</a>	How to use TAP service
ADQL	<a href="http://vo.km3net.de/__system__/adql/query/info">http://vo.km3net.de/__system__/adql/query/info</a>	How to use ADQL query
Data set descriptions	<a href="http://vo.km3net.de/__system__/dc_tables/show/tableinfo/ant20_01.main">http://vo.km3net.de/__system__/dc_tables/show/tableinfo/ant20_01.main</a>	Description of table content for ANTARES data set

### 2.1.3. KM3NeT Open Data Centre

For the KM3NeT Open Data Centre, all documentation (apart from the software framework) has to be documented by KM3NeT. This documentation must include an overview of the functions of the system, contextualization of the data offered, which is achieved via a link to the open science Portal, and an explanation of the interface options to the data. These interfaces are either provided via the REST-API, for which a description of the available endpoints is provided, or a link to the openkm3 Python package and its documentation, as it is dedicated to providing this interface in a Python-based context.

Topic	documentation URL	content
Help & Howto	<a href="http://opendata.km3net.de/help">http://opendata.km3net.de/help</a>	Landing page for server description links
API documentation	<a href="https://open-data.pages.km3net.de/openserver/rest-api.html">https://open-data.pages.km3net.de/openserver/rest-api.html</a>	Overview of API endpoints and request options
Resource view	<a href="http://opendata.km3net.de/data/&lt;type&gt;/&lt;id&gt;/view">http://opendata.km3net.de/data/&lt;type&gt;/&lt;id&gt;/view</a> , e.g. <a href="http://opendata.km3net.de/data/collections/ana20_01/view">http://opendata.km3net.de/data/collections/ana20_01/view</a>	Info on resource, stream etc, including URL to additional references

## 2.2. Manuals for software

Software documentation follows the publication guidelines for software as outlined in Deliverable 4.8. There, core documentation is required from auto-generated API documentation tools and an installation guide. Beyond this, the following additional documentation is provided by the listed packages, which are included in this list as they have



Author(s) Z. Aly, T. Gal, S.R. Gozzini, K. Graf, S. Hallmann, J. Hofestädt, F. Huang, S. Raab, R. G. Ruiz, J. Schnabel, D. Stavropoulos  
document KM3NeT-INFRADEV-WP4-D4.9  
Version 0.3 Release date: KM3NeT 2.0 – 739560  
WP 4  
Public



been used in the open data example analysis. Additional software is available from the KM3NeT Gitlab instance with automatically built project documentation at the according Gitlab pages.

### 2.2.1. KM3NeT software packages

Software	documentation URL	description	content
openkm3	<a href="https://open-data.pages.km3net.de/openkm3">https://open-data.pages.km3net.de/openkm3</a>	interface package to the ODC	user guide
km3astro	<a href="https://km3py.pages.km3net.de/km3astro/">https://km3py.pages.km3net.de/km3astro/</a>	KM3NeT specific tools for coordinate transformation	example gallery
km3pipe	<a href="https://km3py.pages.km3net.de/km3pipe/">https://km3py.pages.km3net.de/km3pipe/</a>	data processing framework	full documentation
kmeta	<a href="https://open-data.pages.km3net.de/kmeta/">https://open-data.pages.km3net.de/kmeta/</a>	metadata management	user guide

### 2.2.2. External software

#### *Using VO clients*

Software	documentation URL	description	content
Aladin	<a href="https://aladin.u-strasbg.fr/java/nph-aladin.pl?frame=downloading">https://aladin.u-strasbg.fr/java/nph-aladin.pl?frame=downloading</a>	Desktop and full sky catalogue client	Download, guidelines
ADQL	<a href="http://docs.g-vo.org/adql/html/">http://docs.g-vo.org/adql/html/</a>	Query language for VO catalogues	Introduction, API documentation
TOPCAT	<a href="http://www.star.bris.ac.uk/~mbt/topcat/">http://www.star.bris.ac.uk/~mbt/topcat/</a>	Tool for operations on catalogues and tables	Introduction & download

#### *Using the VO Python interface*

Software	documentation URL	description	content
astropy	<a href="https://docs.astropy.org/en/stable/">https://docs.astropy.org/en/stable/</a>	Community Python library for astronomy	Full documentation
pyvo	<a href="https://pyvo.readthedocs.io/en/latest/">https://pyvo.readthedocs.io/en/latest/</a>	Python library for VO, astropy affiliated	Installation, user guide





## 2.3. Integration examples and workflows

### 2.3.1. Notebooks

Workflows are made accessible as [Jupyter notebooks](#), which serve to show the integration between the data and software components. All example notebooks can be downloaded from the open science Portal as a package. The notebooks contain in-file documentation as one of their core features is the combination of code blocks and comment blocks which allow rich annotation of the executed code. Also, they demonstrate the interaction between the various platforms and the application of the software for a real-life example. All notebooks can be found at [openscience.km3net.de/notebooks](https://openscience.km3net.de/notebooks). The package includes the following notebooks, exemplifying the various applications and use of data and software components:

Name	File name	description	components
1 Week of ORCA-4	A01_recorded_rate	data taking stability, local coordinates	KM3NeT demo dataset (ODC), openkm3
Quality parameters and event selection	A02_quality_parameters_event_selection	parameter displays, selecting events of interest	KM3NeT demo dataset (ODC), openkm3
Coordinate transformations	A03_events_on_the_sky	sky coordinates of the events	KM3NeT demo dataset (ODC), openkm3, km3astro
Gravitational wave follow-up	A04_gravitational_wave_followup	search for coincidences between GW alert and ORCA neutrinos	KM3NeT demo dataset (ODC), openkm3, km3astro, <a href="#">ligo-gracedb</a>
ANTARES point source search	ANTARES_Point Source	Simple point source search from ANTARES 2007-2017 data	ANTARES 2007-17 dataset (VO), pyvo, sensitivity and background expectation (ODC), openkm3, <a href="#">gammapy</a>

### 2.3.2. Courses and tutorials

Courses at the Education portal introduce the user to complete workflows in step-by-step guides. The following courses have at this point been provided in the portal:

- Analysis and visualization with Aladin: <https://edu.km3net.de/lesson/aladin/>
- Analysis and visualization with TOPCAT: <https://edu.km3net.de/lesson/topcat-2/>
- A simple analysis in Python: <https://edu.km3net.de/lesson/a-simple-analysis-in-Python/>
- Use case 1 - Correlate neutrino data with Gamma Ray Bursts: <https://edu.km3net.de/lesson/correlating-with/>
- Use case 2 - Correlate neutrino data from different experiments: <https://edu.km3net.de/lesson/correlate-neutrino-data-from-different-experiments/>

KM3NeT members are encouraged to use the open science materials to promote the KM3NeT work with young researchers and showcase collaborative work with the open



science tools. To this end, the materials and concept for the online course described in Deliverable 4.10 are made available and offered for reuse.

The webinar was organized using Indico, a web-based management system for meetings. An [event](#) was created on Indico with open access, that means no credentials were required in order to participate. For the webinar, a video meeting room was used and linked to the description of the event page. With these materials provided at the Indico event, KM3NeT members are encouraged to hold similar seminars at their own institutes.

Link to the indico event: <https://indico.cern.ch/event/959379/>

## 3. Manuals for KM3NeT members

Manuals for KM3NeT members concern mostly two parts: on the one hand how to develop software and standardize the own workflow to meet open science criteria early on, on the other hand how to provide data on the open platforms and maintain those servers. In addition to that, KM3NeT members are encouraged to use the open courses and webinar materials to promote the use of KM3NeT open data.

### 3.1. How to develop software

The KM3NeT Collaboration has defined guidelines and recommendations for software development which help to maintain a consistent project structure and development process.

Guidelines build on the requirements for the software publication procedure and link to external best practice guides.

The full software guidelines are available in the Open Science Portal: <http://openscience.km3net.de/articles/members/#how-to-develop-software..>

#### 3.1.1. Python project template

The most commonly used programming language for open science in KM3NeT is Python. A Python project template has been defined by KM3NeT. It is publicly available under <https://git.km3net.de/templates/python-project> and is specifically designed to fit the KM3NeT GitLab continuous integration environment. It includes a skeleton Python project which will be populated with the meta information obtained during the template creation process (project name, description, authors, Git project URL, etc.) and relevant features created in the build process.

### 3.2. Standardizing the scientific workflow

The KM3NeT Collaboration has defined guidelines and recommendations for physics analysis and public plots requirements. Templates are used to reinforce KM3NeT guidelines for physics analysis and public plots production, to ensure analysis reproducibility, consistency in analysis/plots archiving and documentation.



### 3.2.1. Public plot template

The public plot template is a template to automatically create and populate a KM3NeT public plot template with all the meta information needed to archive the plot on KM3NeT gitlab with the corresponding documentation and analysis repository. In other words, the public plot template is a ready-to-use template utilizing continuous integration to create KM3NeT public plot(s).

The full description is available in the open science Portal: <http://openscience.km3net.de/articles/members/#public-plot-template>

### 3.2.2. Analysis template

The Analysis template serves to automatically create and populate a KM3NeT analysis with all the meta information needed to understand, reproduce and archive an analysis on the KM3NeT gitlab server.

Again, the full description and “How To” is available in the open science Portal: <http://openscience.km3net.de/articles/members/#analysis-template>

## 3.3. Providing data on the open platforms

The final data processing step prior to upload to the open platforms generally consists of data annotation with the corresponding metadata depending on the type of open resource. For metadata formatting and management, the kmeta Python package has been developed. The metadata information together with the according data object is then uploaded to the KM3NeT Open Data Centre. All links are internal to the KM3NeT collaboration.

### 3.3.1. Annotating and uploading data

For the first step, example workflows as Jupyter notebooks and documentation are provided in the kmeta package documentation: <https://open-data.pages.km3net.de/kmeta/>

For the second step, scripts to digest the metadata description files and manage the upload are provided (internally) in the KM3NeT Open Data Centre git project with the according README description:

<https://git.km3net.de/open-data/openserver/-/tree/master/django/maintenance>

### 3.3.2. Maintenance of the open science servers and platforms

For all components of the KM3NeT open science system, administrator-side documentation is generally provided as README to the internal gitlab projects with several supporting code examples in the according project directories. Here, specifications for the current implementation of the service are documented, and links to the full external documentation of the employed software is provided.

Service	Software	Internal documentation	External manuals
Education portal	Wordpress, LifterLMS	<a href="https://git.km3net.de/open-data/eduportal">https://git.km3net.de/open-data/eduportal</a>	<a href="https://codex.wordpress.org/">https://codex.wordpress.org/</a> <a href="https://lifterlms.com/docs/">https://lifterlms.com/docs/</a>



Open Data Centre	Django	<a href="https://git.km3net.de/open-data/openserver/">https://git.km3net.de/open-data/openserver/</a>	<a href="https://docs.djangoproject.com/en/3.1/">https://docs.djangoproject.com/en/3.1/</a>
VO Server	DaCHS	<a href="https://git.km3net.de/open-data/voserver">https://git.km3net.de/open-data/voserver</a>	<a href="https://dachs-doc.readthedocs.io/">https://dachs-doc.readthedocs.io/</a>
open science Portal	Gitlab Pages, Hugo	<a href="https://git.km3net.de/open-data/openscienceportal">https://git.km3net.de/open-data/openscienceportal</a>	Gitlab pages: <a href="https://git.km3net.de/help/user/project/pages/index.md">https://git.km3net.de/help/user/project/pages/index.md</a> Hugo: <a href="https://gohugo.io/documentation/">https://gohugo.io/documentation/</a>

## Appendix I: Architecture overview



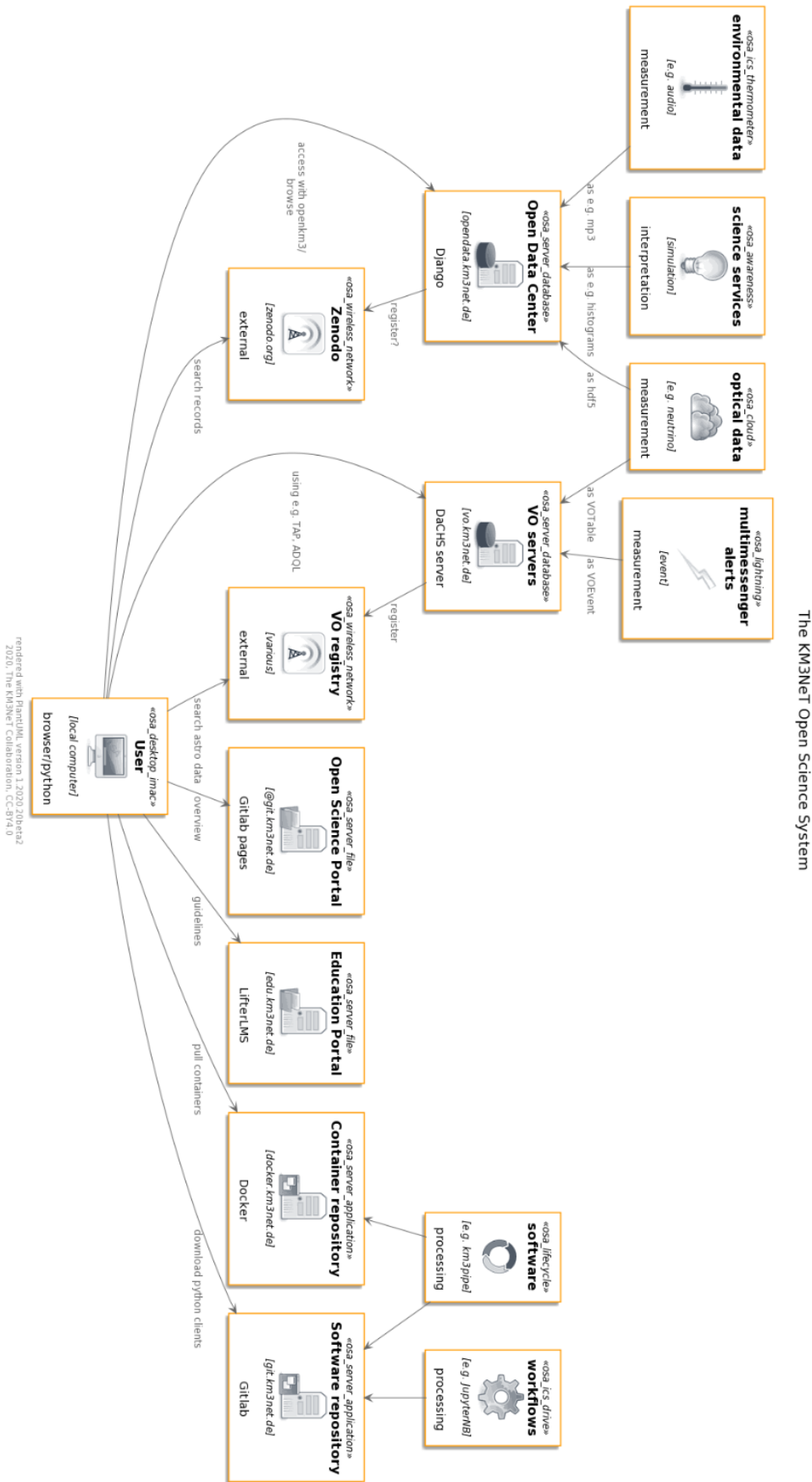


Figure 1: Overview for the Open Science System

