HELMHOLTZ

RESEARCH FOR
GRAND CHALLENGES

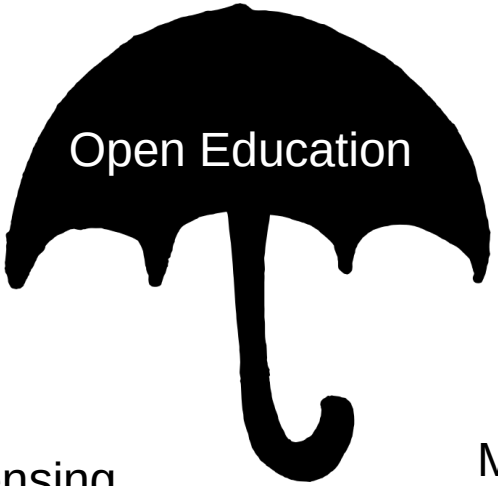
How to use GitHub for open education? A repo for ML teaching materials

Ines Schmahl

Central Library, Forschungszentrum Jülich

Open Education

Definition



Open Educational Resources (OER)

Open Practices

Open Licensing

Massive Open Online Courses (MOOCs)

Open Educational Technology

Open Educational Resources

Definition

Deutsche UNESCO-Kommission:

- All kinds of teaching materials
- Under open license
- Free access, free reuse and free redistribution

Open Educational Resources

Definition

Deutsche UNESCO-Kommission:

- All kinds of teaching materials
- Under open license
- Free access, free reuse and free redistribution

But what does open mean?

There is no clear close or open.

Open Educational Resources

Definition

Deutsche UNESCO-Kommission:

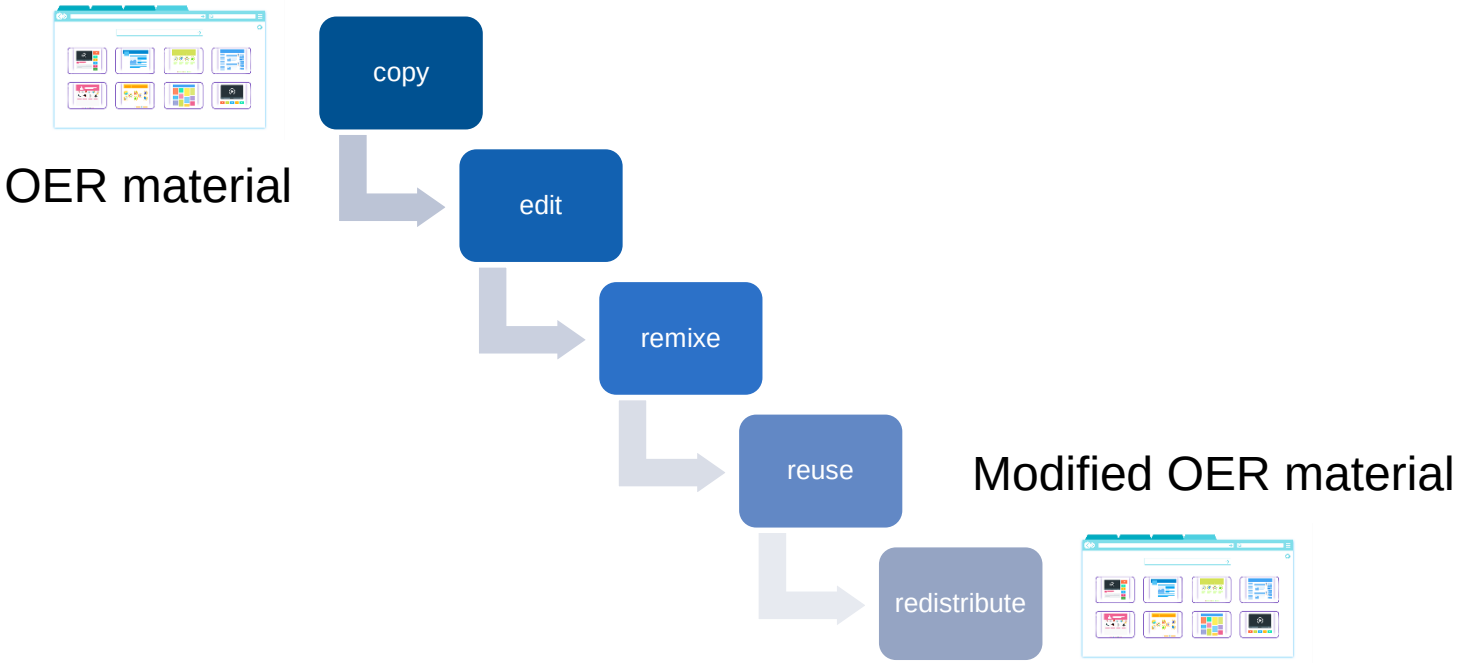
- All kinds of teaching materials
- Under open license
- Free access, free reuse and free redistribution

But what does open mean?

There is no clear close or open.

Open Educational Resources

5R Activities



Open Educational Resources

5R Activities

OER material

copy

edit

remixe

reuse

redistribute

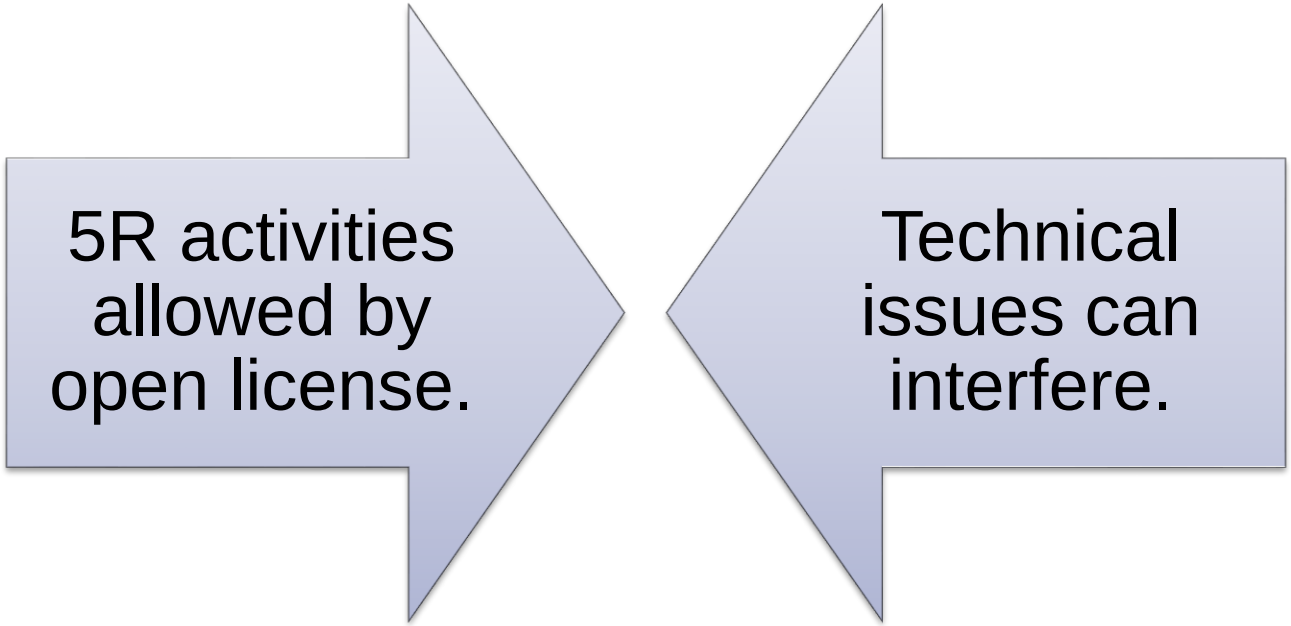
Modified OER material

Not all open licenses guarantee all those 5 rights.

→ Choose license, which allows all 5 activities.

Open Educational Resources

ALMS Framework



Open Educational Resources

ALMS[1)] Framework

OER material

With which tools can I edit it? Are they open source tools?

Do I need expert skills to edit the material?

Is it technically possible to edit the material (e.g. scanned documents)?

Can I edit the material in the same format as downloaded?

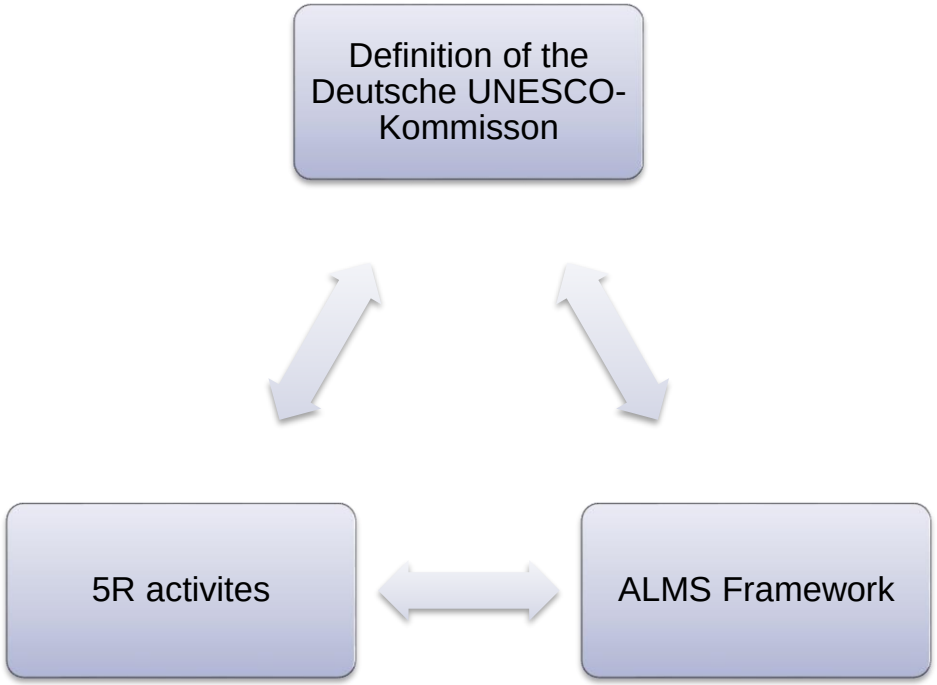
Modified OER material

[1)] *ALMS:*

- Access to Editing Tools
- Level of Expertise Required
- Meaningfully Editable
- Self-Sourced

Open Educational Resources

Overview



Open Educational Resources

Machine learning (ML) and data science (DS)

→ **Starting point:** Methods from ML and DS are used in many disciplines

Challenges

- Participants have different backgrounds
- Diverse teaching materials (slides, codes, data, exercises, images, scetches)
- Additional materials like libraries and frameworks are used

Open Educational Resources

Machine learning (ML) and data science (DS)

→ **Starting point:** Methods of ML and DS are used in many disciplines.

Challenges

- Participants have different backgrounds
- Diverse teaching materials (slides, codes, data, exercises, images, scetches)
- Additional materials like libraries and frameworks are used

Need of highly adoptable teaching materials

Open Source Educational Resources

Machine learning (ML) and data science (DS)

Present situation

- Many open courses (MOOCs)
- But no access to the sources

Developing Open Source Educational Resources for Machine Learning and Data Science

Ludwig Bothmann[1] Sven Strickroth[2] Giuseppe Casalicchio[1] David Rügamer[1] Marius Lindauer[3] Fabian Scheipl[4] Bernd Bischl[1]

Abstract

Education should not be a privilege but a common good. It should be openly accessible to everyone, with as few barriers as possible; even more so for key technologies such as Machine Learning (ML) and Data Science (DS). Open Educational Resources (OER) are a crucial factor for greater educational equity. In this paper, we describe the specific requirements for OER in ML and DS and argue that it is especially important for these fields to make source files publicly available, leading to Open Source Educational Resources (OSER). We present our view on the collaborative development of OSER, the challenges this poses, and first steps towards their solutions. We outline how OSER can be used for blended learning scenarios and share our experiences in university education. Finally, we discuss additional challenges such as credit assignment or granting certificates.

1. Introduction

Education is of paramount importance to overcome social inequalities. Globalization and broad access to the internet provide a major opportunity for allowing many more people from all over the world to access high quality educational resources. We endorse the vision of the Open Education Global initiative (OEG, n.d.a) and believe that teaching materials developed with the support of public financial resources should be openly accessible for the general public. The initiative aims at providing Open Educational Resources (OER) (OEG, n.d.b) which come with the "3R" permissions to retain, reuse, revise, remix, and redistribute the material. The UNESCO strongly promotes the concept of OER (UNESCO, n.d.), has held two world congresses on OER, in 2012 and 2017, and finally adopted a recommendation on OER (UNESCO, 2019). This recommendation is claimed to be "the only existing international standard-setting instrument on OER" (UNESCO, n.d.). A variety of Massive Open Online Courses (MOOCs) have been created in the fields of Machine Learning (ML) and Data Science (DS). These MOOCs are mostly offered by commercial platforms (e.g., Ng, 2021; Bolton et al., 2021; Malone & Thun, 2021; Google, 2021; Salmon et al., 2021; Ermenko & de Postavos, 2021) but also by university-owned platforms (e.g., MIT, 2021). Although the material itself is often freely accessible, access to the sources that are needed to reproduce, modify and reuse the material is usually not provided. Only a small fraction of courses in ML and DS actually share all their sources, such as slides sources in .ppts or LaTeX and source codes for plots, examples, and exercises. We call those "Open Source Educational Resources" (OSER) to underline this feature; positive examples include Montani (2021); Cetinkaya-Rundel (2021a,b); Vanschoren (2021).

Direct benefits of OSER from the perspective of lecturers include: 1. The material will often be of higher quality if additional experts are able to contribute and improve the material. 2. It is more efficient to develop a new course since material can be adapted and re-used from previously created courses legally. 3. Starting from an established OSER course, lecturers can focus on developing additional chapters and tailoring existing material to their audience.

We believe there will never be the one and only course on a certain topic – in fact, diversity in how topics are taught is important. Reasons include different constraints on the volume of a course defined by an institution's curriculum, different backgrounds (in the context of ML and DS courses, e.g., statistics, mathematics, computer science), different types of institutions (e.g. university vs. continuing education, undergraduate vs. (post)graduate), different substantive focus, and sometimes simply different styles. On the other hand, it seems natural that teachers for many subjects should be able to find networks of peers among which a considerable amount of content can be similar and we advo-

[1]Department of Statistics, Ludwig-Maximilians-Universität München, Germany [2]Institute of Computer Science, Ludwig-Maximilians-Universität München, Germany [3]Institute of Information Process, Leibniz University Hannover, Germany. Correspondence to: Ludwig Bothmann <ludwig.bothmann@stat.uni-muenchen.de>.

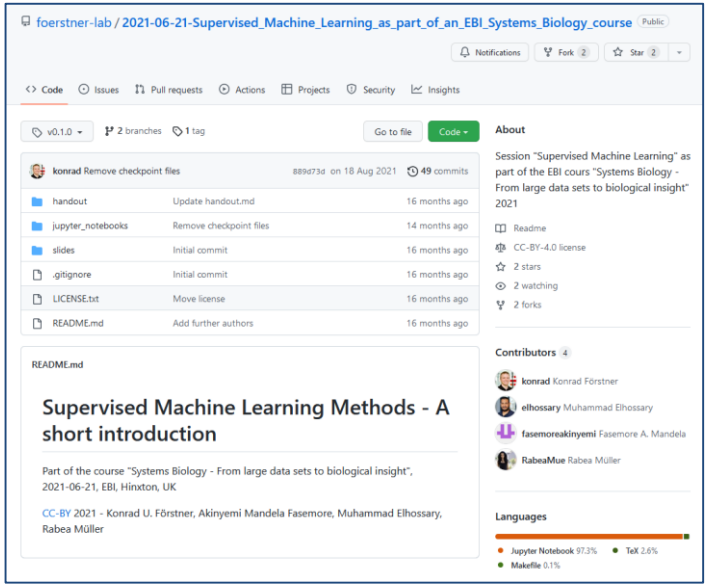
https://doi.org/10.48550/arXiv.2107.14330

From Open Eduational Resources (OER) To Open Source Educational Resources (OSER)

Session "Supervised Machine Learning Methods"

Example

On the platform GitHub



https://github.com/foerstner-lab/2021-06-21-Supervised_Machine_Learning_as_part_of_an_EBI_Systems_Biology_course/tree/v0.1.0

Teaching material	File format
Handout, README	Markdown
Code	Jupyter notebooks
Slides	PDF and .tex file
Images	PDF and .svg

Create Collection

New Approach

Increase flexibility of OSER

Instead of creating a course → Creation of a collection

Tasks

1. Create collection.
2. Create material.
3. Create GitHub repo.

Create Collection

Planning

Select content

Machine learning teaching materials

Select types of material

- Focus on images with explaining text snippets, because
  - Creation is time consuming
  - Can be easily reused across disciplines as well as languages
  - Visualisations support the understanding
- Code examples

Create Collection

Planning

Select license for materials

To fulfill the 5R activities → CC-BY 4.0 license

Note

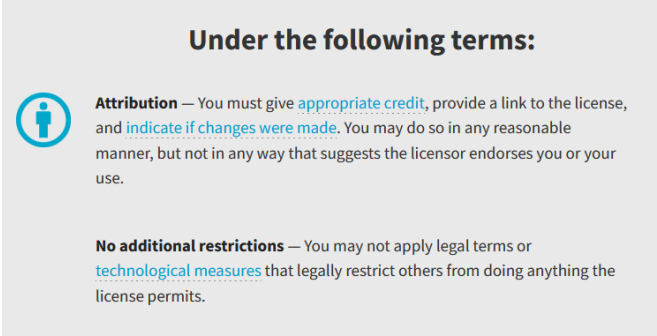
- ML needs data.
- Check the rights of use for additional materials.

Create Collection

Planning

Select metadata for materials (according to required credit)

- Title
- Creator
- Link to source
- License with link

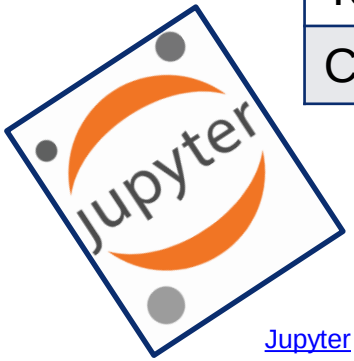


Create OER

Reusability – ALMS Framework

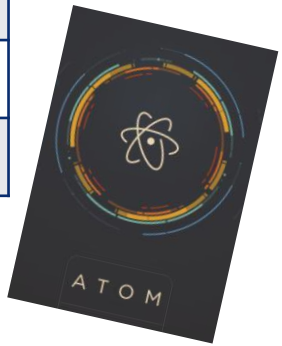
With which tools can I edit it? Are they open source tools?

Material type	Open source tools
Image	Inkscape
Text snippet	Atom (markdown editor)
Code	Jupyter Notebook



Jupyter

Inkscape



Blog (June 8, 2022): Sunsetting atom

Create OER

Reusability – ALMS Framework

Open source tools	Learning time
Inkscape	Medium
Atom	Short
Jupyter Notebooks	Short

Create OER

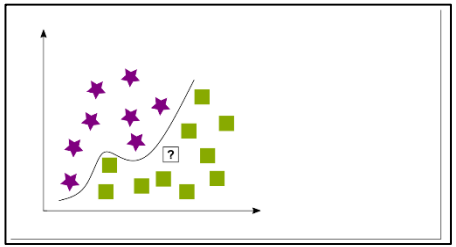
Reusability – ALMS Framework

With which tools can I edit it? Are they open source tools?

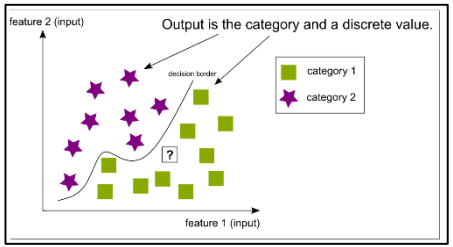
Do I need expert skills to edit the material?

Is it technically possible to edit the material (e.g. scanned documents)?

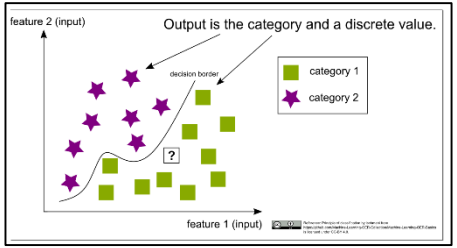
INKSCAPE
Draw Freely.



Level 1: Image



Level 2: Text



Level 3: Reference

Create OER

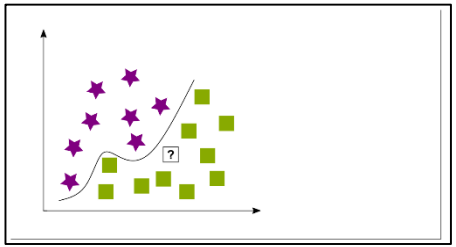
Reusability – ALMS Framework

With which tools can I edit it? Are they open source tools?

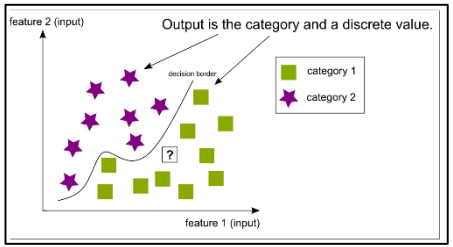
Do I need expert skills to edit the material?

Is it technically possible to edit the material (e.g. scanned documents)?

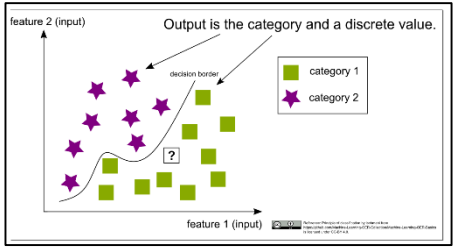
INKSCAPE
Draw Freely.



Level 1: Image



Level 2: Text

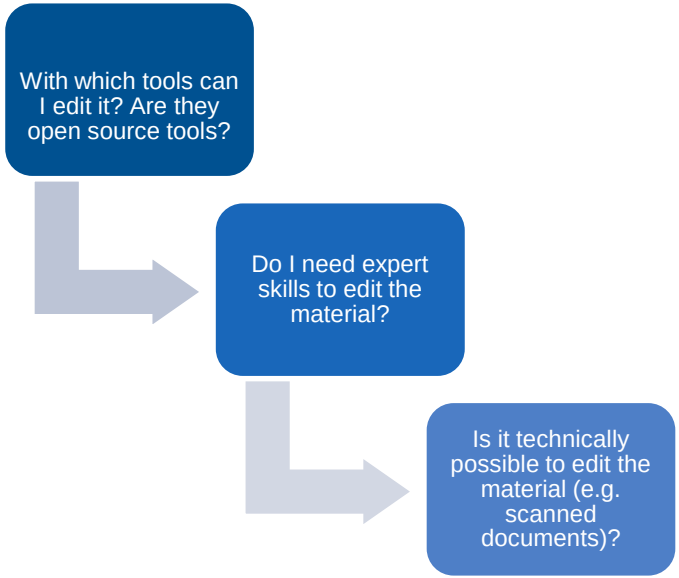


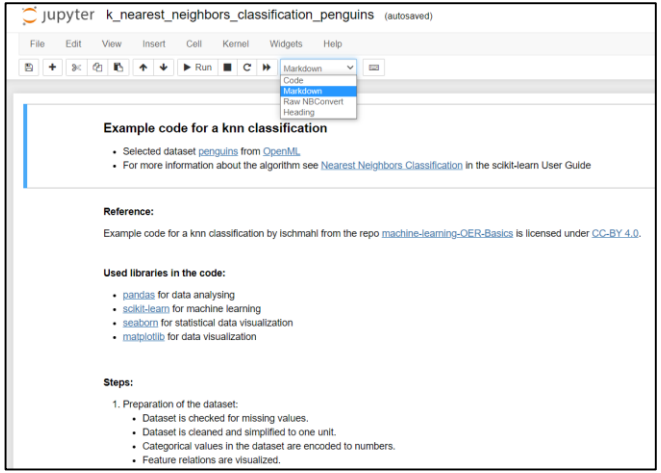
Level 3: Reference

Use features like levels in Inkscape to increase flexibility.

Create OER

Reusability – ALMS Framework

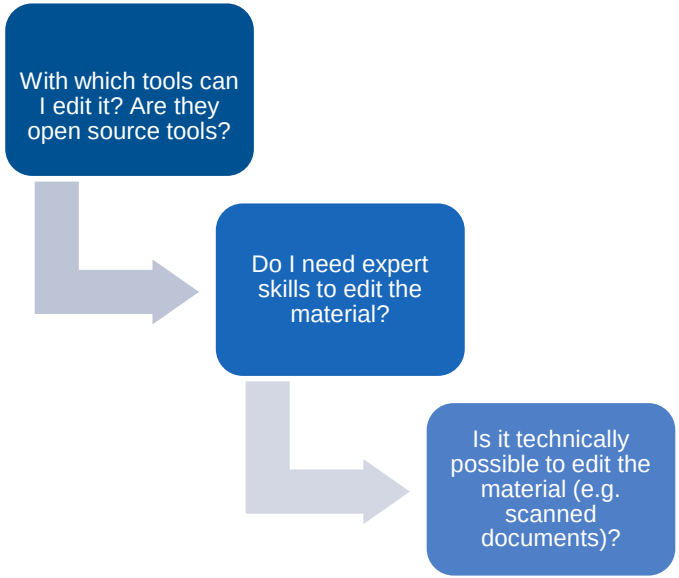




Cells in Jupyter Notebooks allows easily to add, delete or change the material.
→ Concept of modularisation

Create OER

Reusability – ALMS Framework




```
## knn algorithm

The knn algorithm predicts the category for the new dataset from the labels of the k
nearest neighbors. The value of k is set before.

<br>

>![knn algorithm](../img/principle_knn_classification.svg)

<br>

#### Code example for k=1:
    from sklearn.neighbors import KNeighborsClassifier
    knn_clf = KNeighborsClassifier(n_neighbors=1)
    knn_clf.fit(X_train, y_train)
    knn_clf.predict(X_test)
    knn_clf.score(X_test, y_test)

(Reference: Example code for a knn classification by ischmahl from the repo [machine-
learning-OER-Basics](https://github.com/Machine-Learning-OER-Collection/Machine-Learning-
OER-Basics) licensed under [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/).)

_Reference:
knn algorithm by ischmahl from the repo [machine-learning-OER-Basics](https://github.com/
Machine-Learning-OER-Collection/Machine-Learning-OER-Basics) is licensed under [CC-BY
4.0](https://creativecommons.org/licenses/by/4.0/)._
```

Easy to edit

Create OER

Reusability – ALMS Framework

With which tools can I edit it? Are they open source tools?

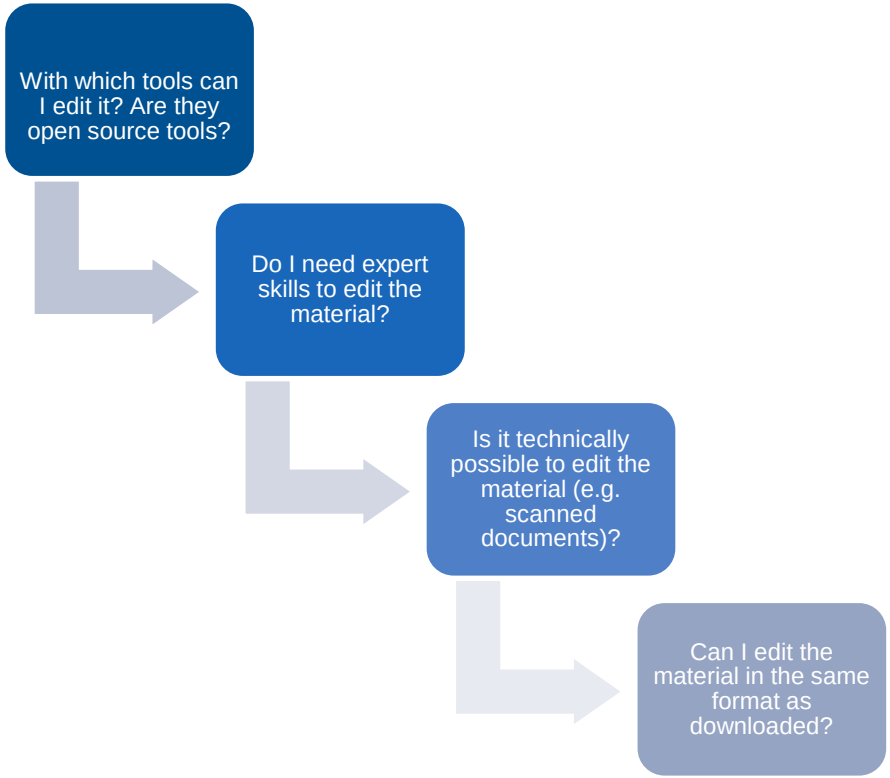
Do I need expert skills to edit the material?

Is it technically possible to edit the material (e.g. scanned documents)?

Easy to edit and to combine with other materials.

Create OER

Reusability – ALMS Framework



Material type	File format
Image	.svg
Text snippet	.md
Code	.ipynb

Create OER

Reusability

Link metadata and material

- Title
- Creator
- Link to source
- License

Create OER

Reusability

Link metadata and material

- Title
- Creator
- Link to source
- License

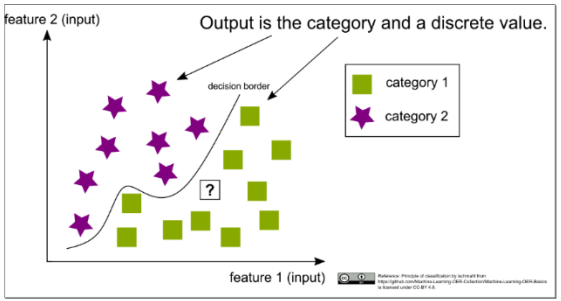
for distribution

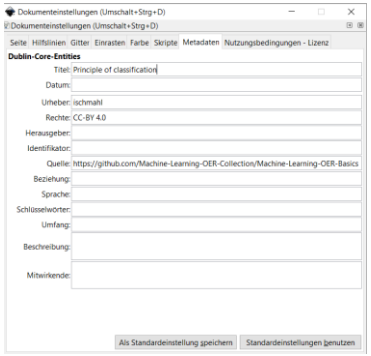
Create OER

Reusability

Link metadata and material

- Metadata included in the image
- Metadata included in the file properties

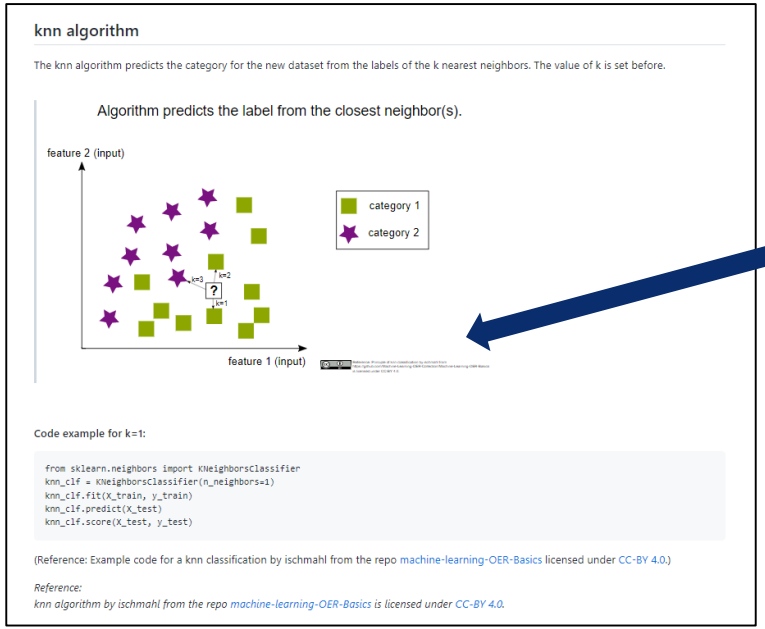


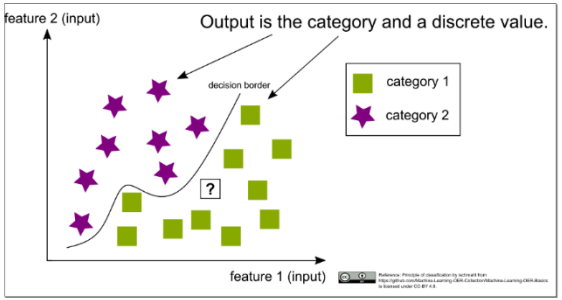


Create OER

Reusability

Mixed material

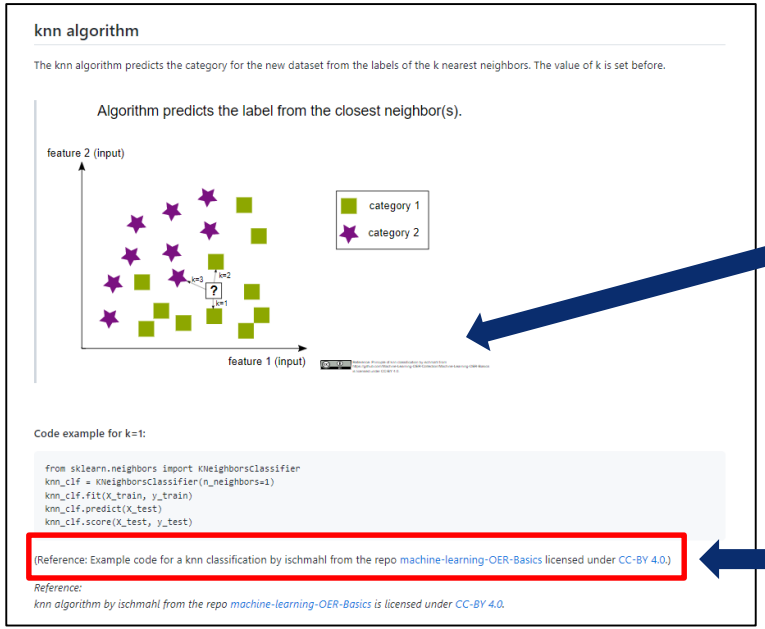


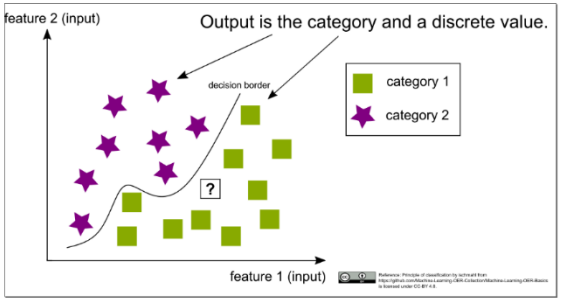


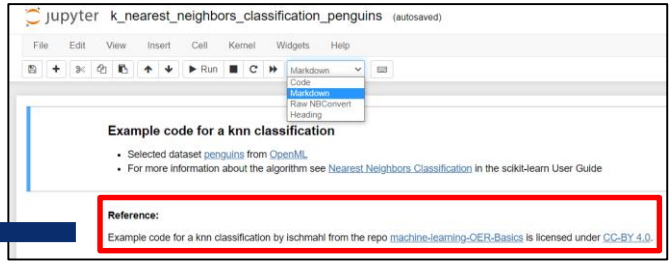
Create OER

Reusability

Mixed material: Provide references



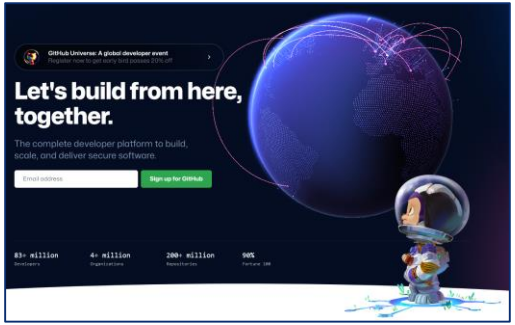




Create GitHub Repo

Key facts about GitHub

- Platform for collaborative software delevopment
- Integrates the version-control system git
- Projects are stored in repositories
- Many features for managing projects



https://github.com/

Create GitHub Repo

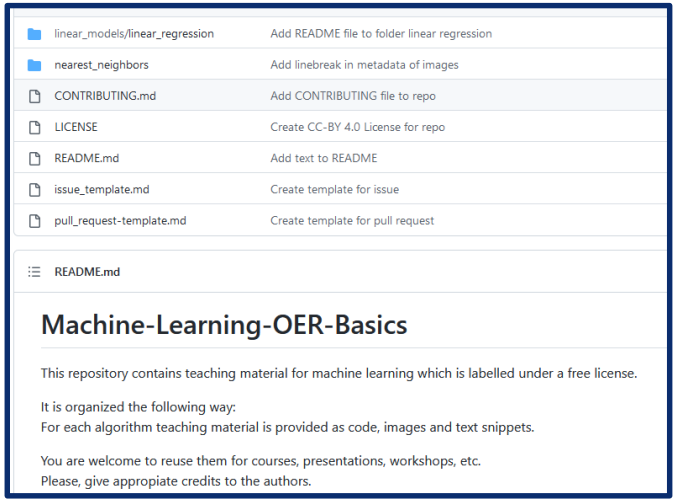
OSER Project

→ Features in GitHub can be transferred to an OSER project

For information

Write a README file including

- Short description
- How is the repo structured?
- What content is provided?
- License
- Links to further sources (optional)

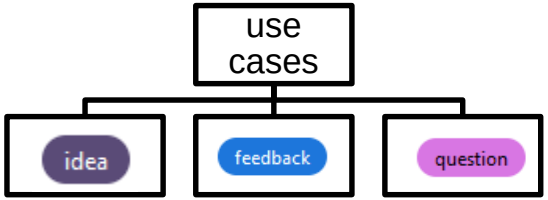


Create GitHub Repo

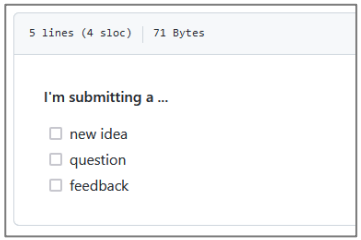
OSER Project

For communication

- Define communication use cases for the project



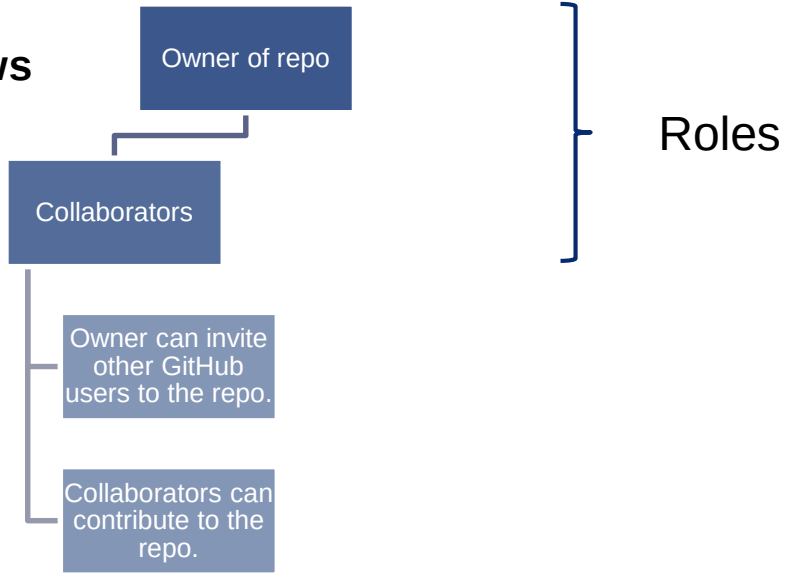
- Use features like creating an issue template and labels for a better usability



Create GitHub Repo

OSER Project

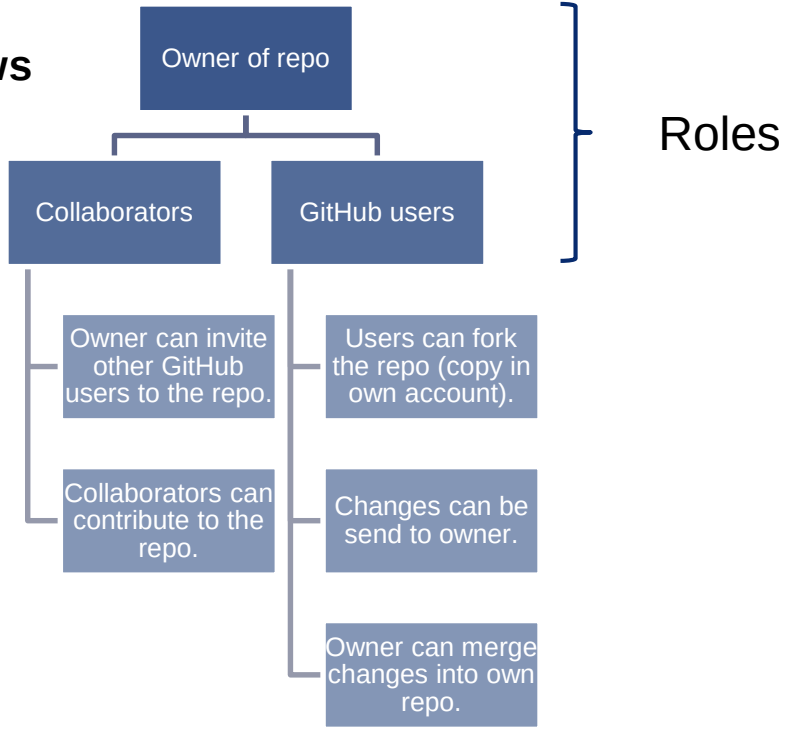
GitHub collaboration workflows



Create GitHub Repo

OSER Project

GitHub collaboration workflows



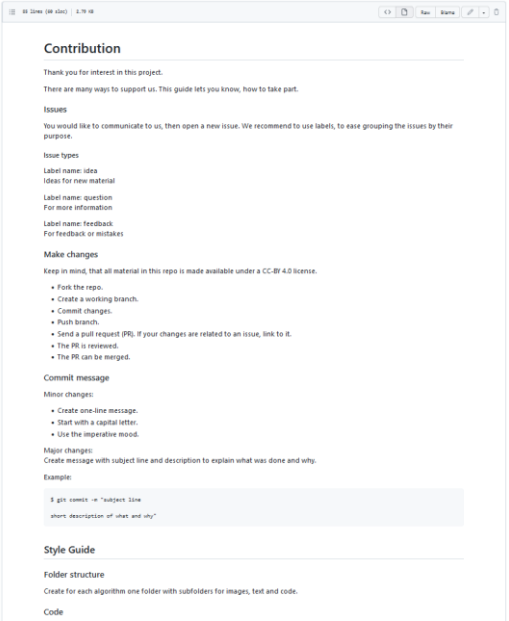
Create GitHub Repo

OSER Project

For collaboration

Write CONTRIBUTING file including

- communication workflows
- collaboration workflows
- documentation guide
- style guide for materials



Contribution

Thank you for interest in this project.

There are many ways to support us. This guide lets you know, how to take part.

Issues

You would like to communicate to us, then open a new issue. We recommend to use labels, to ease grouping the issues by their purpose.

Issue types

Label name: idea
Ideas for new material

Label name: question
For more information

Label name: feedback
For feedback or mistakes

Make changes

Keep in mind, that all material in this repo is made available under a CC-BY 4.0 license.

- Fork the repo.
- Create a working branch.
- Commit changes.
- Push branch.
- Send a pull request (PR). If your changes are related to an issue, link to it.
- The PR is reviewed.
- The PR can be merged.

Commit message

Minor changes:

- Create one-line message.
- Start with a capital letter.
- Use the imperative mood.

Major changes:
Create message with subject line and description to explain what was done and why.

Example:

```
$ git commit -m "subject line

short description of what and why"
```

Style Guide

Folder structure

Create for each algorithm one folder with subfolders for images, text and code.

Code

Create GitHub Organization

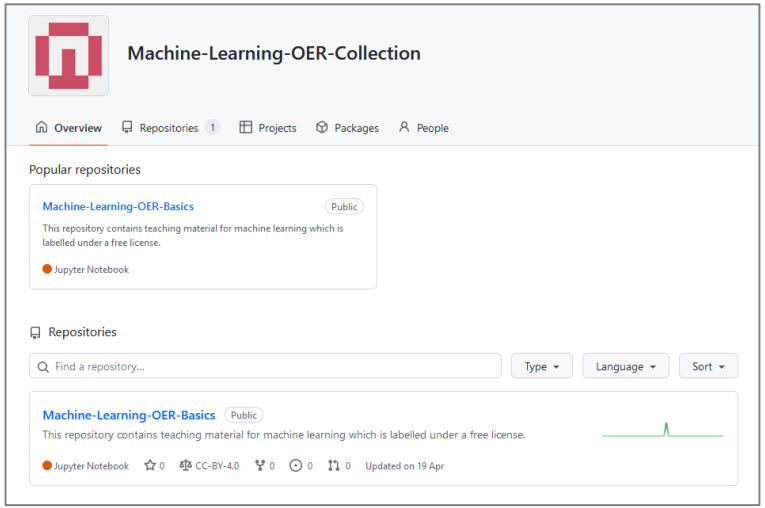
OSER Project

Enlarge project

GitHub provides a second account type:

→ Organization

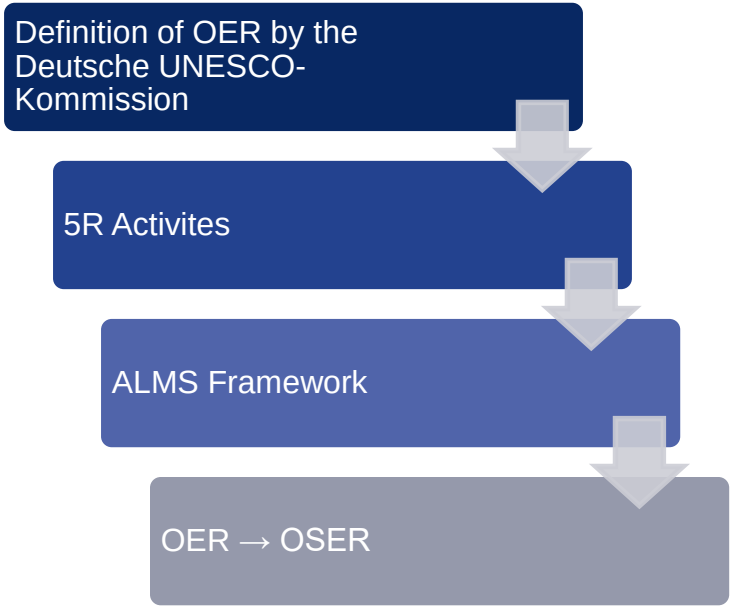
- Can manage number of repos.
- Permission management provides more roles.
- Roles can be assigned to a person or to a team.



https://github.com/Machine-Learning-OER-Collection

Lessons Learned

Summary

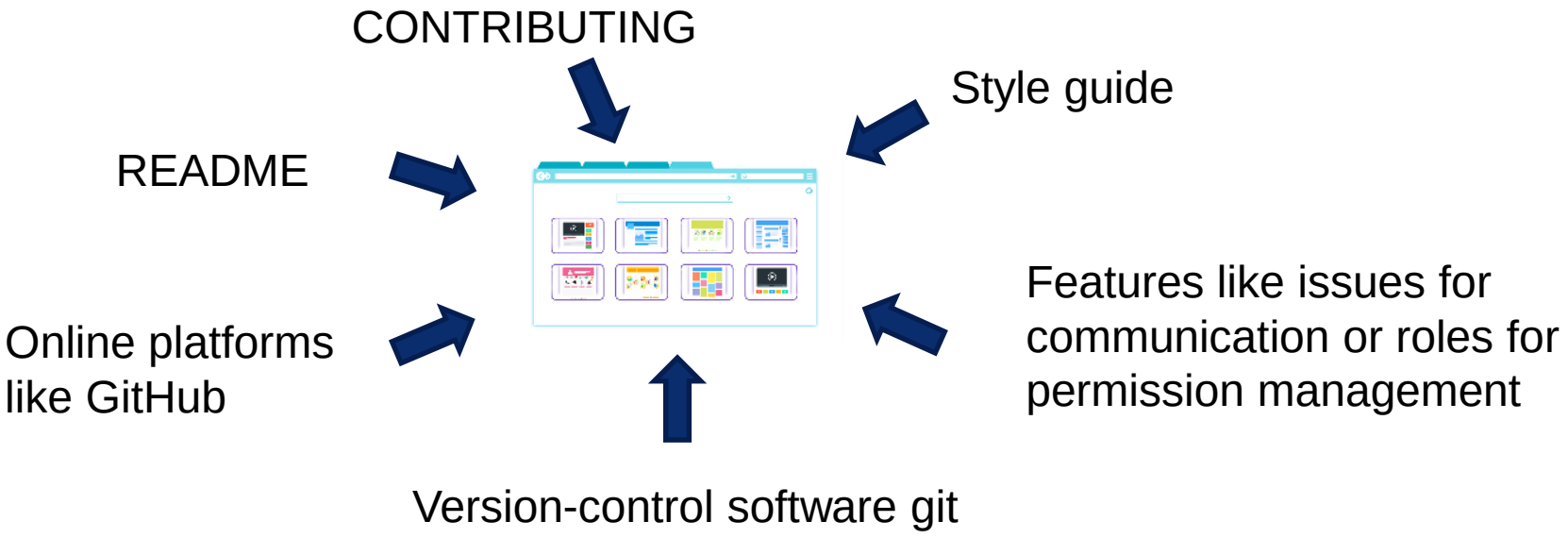


Focus on the collaborative development and improvement of open teaching materials

Lessons Learned

Summary

→ **Open Education can learn from open source**



Outlook

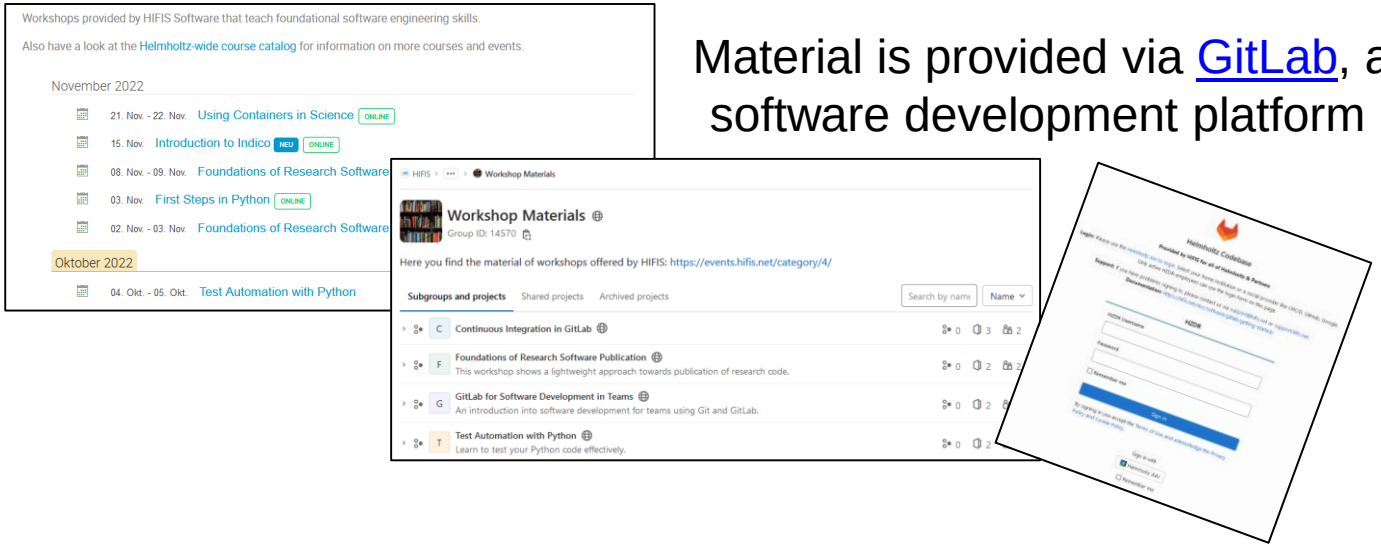
OER in HIFIS

HIFIS Workshops

Outlook

OER in HIFIS

HIFIS Workshops

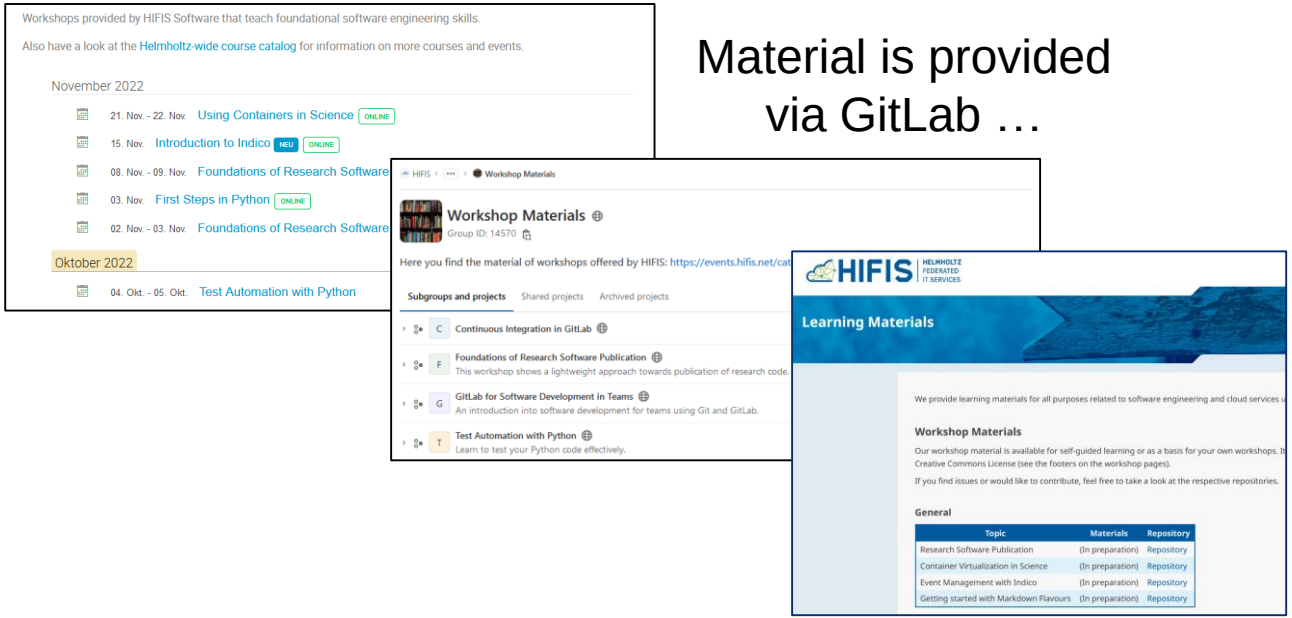


Material is provided via [GitLab](), a open source software development platform based on Git

Outlook

OER in HIFIS

HIFIS Workshops



Material is provided via GitLab …

and published on the webside Learning Materials.

Outlook

Talk: Holding Workshops is not hard - Organizing is!

Use case 1 for OER

Use case 2 for OER

Thank you!

Ines Schmahl

Central Library, Forschungszentrum Jülich

forschungsdaten@fz-juelich.de

Bibliography

Bildung | Deutsche UNESCO-Kommission. (n.d.). Retrieved 12 October 2022, from https://www.unesco.de/bildung/open-educational-resources

Bothmann, L., Strickroth, S., Casalicchio, G., Rügamer, D., Lindauer, M., Scheipl, F., & Bischl, B. (2021). Developing Open Source Educational Resources for Machine Learning and Data Science. CoRR, abs/2107.14330. https://arxiv.org/abs/2107.14330

Müller, R., Fasemore, A. M., Elhossary, M., & Förstner, K. U. (2022). A lesson for teaching fundamental Machine Learning concepts and skills to molecular biologists. In K. M. Kinnaird, P. Steinbach, & O. Guhr (Hrsg.), Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop (Bd. 170, S. 68–72). PMLR. https://proceedings.mlr.press/v170/muller22a.html

What We Do – OEGlobal. (n.d.). Retrieved 12 October 2022, from https://www.oeglobal.org/about-us/what-we-do/

Wiley, D. (n.d.). Defining the ‘Open’ in Open Content and Open Educational Resources. Defining the ‘Open’ in Open Content and Open Educational Resources. Retrieved 12 October 2022, from https://opencontent.org/definition/

Image list

200degrees. (2016). Browser internet tab. https://pixabay.com/de/vectors/browser-internet-tab-1666995/

Clker-Free-Vector-Images. (2014). Menschen Gruppe Menge. https://pixabay.com/de/vectors/menschen-gruppe-menge-linie-312122/

Mello, J. (n.d.). Open Educational Resources. https://www.unesco.de/bildung/open-educational-resources

wanglizhong. (2014). Umbrella-revolution-symbol. https://openclipart.org/detail/202690/umbrella-revolution-symbol-3