Concept and Realization of Open Educational Resources using Jupyter

Sascha Spors, Frank Schultz, Vera Erbes and Max Schröder
Institute of Communications Engineering, University of Rostock, Germany, Email: Sascha.Spors@uni-rostock.de

Introduction

Educational materials which are provided under an open license are generally referred to as Open Educational Resources (OERs). The term includes all materials aiming at the transfer of knowledge, e.g. educational books, lecture notes, reports, slides and white-papers, in any medium, e.g. on paper, video or as digital resource. OERs address the fundamental right on free information access by lowering the barrier to knowledge. In this contribution we report on the concept and realization of OERs in the field of (digital) signal processing and communication acoustics. Not only the authoring of the material is considered, but also its distribution and maintenance. We show how open and established standards, tools and workflows can be utilized in order to address all relevant aspects. First a brief look is taken at the foundation of OERs, followed by a discussion of the educational context, data provenance and existing tools. This lays the grounds for an in-depth discussion of the actual realization.

Open Educational Resources

The motivation behind OERs is to establish an alternative educational paradigm with low barriers to knowledge supported by free high-quality educational resources. Distance education [1] using electronic media is growingly pursued due to increased flexibility in different ways and socio-cultural issues in the context of further digitalization, to name two important aspects. Although OERs are by principle not limited to digital media, nowadays OER is often used as synonym for digital, at least open accessible, sometimes modifiable and redistribuatble, data that is meant for education in the context of distance education. OERs have their foundations in the early 2000s [2, Ch. 5.9], when universities started to provide video lectures and digital scripts for their courses and tutorials. This lead to the Creative Commons (CC) licensing model. On a institutional level MIT OpenCourseWare[1] can be considered as one of the early adopters offering a wide variety of courses.

The term Open Educational Resources has many working definitions. One of the first formal definitions was adopted at UNESCO's forum on the Impact of Open Courseware for Higher Education in Developing Countries 2002: *'teaching, learning and research materials in any medium, digital or otherwise, that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions. Open licensing is built within the existing framework of intellectual property rights as defined by relevant international conventions and respects the authorship of the work'* [3]. According to [4], at least five rights (known as the *5Rs*) should be granted. The right to

1. retain copies of,
2. reuse,
3. revise,
4. remix, and
5. redistribute

the contents in any sensible way. OERs may be found on any educational level, beginning on elementary till higher education, training and research materials [5]. So far, they do not seem to be widely employed in higher education. Although especially here many education materials are made available for download on the internet they are not necessarily OERs. This is often due to the fact that the licensing (model) is unclear at all or underlying materials are restrictively licensed. One reason for this might be the lacking familiarization with copyright laws and licensing models. If a proper license is given, the family of CC licenses is internationally the most widely one used for OERs.

Besides many obvious benefits of OERs, quality concerns are frequently brought up as counterargument. Traditional quality assurance systems like peer-reviews are often not in place which potentially could result in irrelevant or inaccurate contents. However, one of the most commonly used OERs – Wikipedia – shows that a reasonable quality level can be maintained in a well-designed collaborative authoring, editing and review process. We later discuss how such a process can be established for OERs using existing tools and workflows.

Educational Context and Requirements

Educational materials may be used in very diverse educational contexts giving rise to many different pedagogical concepts and realizations. In this contribution, the discussed OERs are taken from our own field of experience covering bachelor and masters courses in an electrical engineering curricula. More specifically, we deal with courses in the field of (digital) signal processing and communication acoustics. Signal processing is a sub-topic of electrical engineering focusing on the analysis, modification and synthesis of signals. Electromagnetic and electrical signals, sound or images are prominent examples for signals. The theory of signals and systems provides an abstract description of signals and their modification by systems. This allows to apply it to a wide range of engineering problems, but comes at the cost of some abstractness. In order to cope for this abstractness, illustrative examples, potentially interactive, are good practice in education. Communications acoustics is an interdisciplinary science that focuses on human communication

[1] `https://ocw.mit.edu`

by sound. It combines results from acoustics, electroacoustics, speech production, human auditory perception and signal processing. In both areas, signal processing and communication acoustics, the technical contents to be conveyed are applied mathematics (e.g. Fourier and Laplace transform), acoustics, psychophysics and numerical algorithms. Typically text, mathematical formulas, technical plots, images, animations and source code is used in the teaching materials.

Our requirements for the realization of OERs are based on the 5Rs and our practical experience with previous sharing of educational materials with students (e.g. lecture slides, source code in MATLAB or Python). We aim at

- electronic materials featuring a seamless integration of different media,
- support for interactive behavior and media (e.g. hyperlinks, animations, ...),
- low (technical) barriers for distribution and usage,
- open, non-proprietary standards, tools and workflows, and
- open license models.

Provenance

Science relies on the transparency of its findings through documentation of processes and results, as well as validation and reproduction by third parties. The same should hold for educational materials as these form the basis of knowledge. As a matter of fact, OERs may be seen as one aspect of open science. Although non-electronic materials are explicitly included in the definition of OERs, most of the materials will likely be provided electronically, nowadays. Furthermore, educational resources for higher education (e.g. lecture notes) may contain scientific results themselves that are not documented elsewhere.

Data provenance refers to tracking the origins of data and its modifications throughout its life-cycle. It hence provides a historical record of the data. Data provenance is considered as one important prerequisite for the reproducibility of scientific results [6, 7]. Understanding OERs as an integral part of science suggests to apply the same principles to the processes of authoring and maintaining these materials. We therefore suggest the following measures

1. Revision Tracking – public tracking of revisions in order to make the authoring and maintenance process transparent.
2. Authorship Tracking – detailed tracking of authorship in order attribute new or changed material to individuals. Together with revision tracking this forms the basis for a collaborative authoring and editing process.
3. Referencing – the sources of knowledge and data have to be referenced/cited appropriately according to scientific standards. A unique identifier should be provided for the OER itself.
4. Quality Assurance Methods – include public review of materials, documentation of issues and their handling, automatic consistency checks, continuous integration techniques, documentation/freezing the computing environment.

Note, most of these measures are in place at Wikipedia.

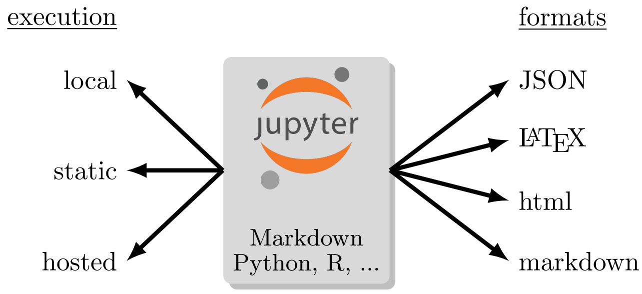

Figure 1: The Jupyter ecosystem. Logo ©2017 project Jupyter contributors

The Jupyter Ecosystem

A solution that shows many favorable properties for OERs in the technical domain is provided by *Jupyter notebooks*[2]. The term is used for both the JSON document itself and the web application for editing and executing these notebooks. The document format as well as the reference implementation of the web application are developed collaboratively by the *Project Juypter* community and published under open license models. Jupyter notebooks feature an interactive integration of text, mathematical equations, illustrations and source code together with its output. The contents are organized in *cells* within a streamlined document. All cells can be edited interactively, the code cells may be reexecuted in arbitrary order laying the grounds for exploratory computation. Documentary cells support formatting by markdown[3]. A screenshot of an exemplary notebook is given in Figure 3.

Although Jupyter emerged from the interactive Python (*IPython*) project[4], it has broadened its focus considerably and supports a wide variety of programming languages e.g. julia, Python, R. The document and the notebook itself can be converted into various other formats. Many extensions have been developed extending the basic functionality of Jupyter, e.g. spell-checking, automatic code formatting or rating of student assignments. Jupyter is available for all major operating systems and hardware platforms. Public and commercial services allow the viewing, authoring and execution of notebooks without the need of local installation. The basic Jupyter ecosystem is illustrated in Figure 1. To date, Jupyter has become one of the most used tools for exploratory science and shows a significant growth in educational applications. This has led to an entire ecosystem supporting

[2]`https://jupyter.org`
[3]`https://en.wikipedia.org/wiki/Markdown`
[4]`https://ipython.org`

Jupyter, starting from educational materials on Jupyter itself, packaged distributions, extensions, services and scientific investigations on its usage [8], just to name a few. Due to its features, the Jupyter notebook is well suited for our technical educational materials. This is discussed in more detail in the later section on realizations.

Realization

We used the above introduced concepts as guidelines to realize OERs for the curriculum of Signal Theory and Digital Signal Processing at the Institute of Communication Engineering, Universität Rostock, Germany[5]. Table 1 provides an overview on the resources realized so far. We focus the following discussion on the OERs for the lectures on Signals and Systems and Digital Signal Processing as these are our most mature materials at the current state. The aim was to create reference OERs for both, as signal processing is an integral part of a typical electrical engineering curriculum. The materials itself, their provisioning and publication should support the ease of participation, reuse and remix. We therefore have chosen the *Creative Commons Attribution* (CC-BY) license[6] for the text and illustrations. The computational examples may be used without any restrictions and are therefore released under the *MIT license*[7]. For above given reasons we have chosen to realize all OERs as Jupyter notebooks. In order to keep the entry barriers low only the basic functionality of Juypter is used. None of the available extensions to Jupyter notebooks are used, as these may require additional installation or may not be supported by hosted services. In order to organize the contents in a modular manner, the materials are split into a series of notebooks. Notebooks belonging to specific topics are placed in respective sub-directories. As an example, the directory structure of the Signals and Systems lecture materials is shown in Figure 2. The top-level headline of each individual notebook refers to the chapter, the sub-headlines to the specific topic addressed in the notebook. This is illustrated by the screenshot of an exemplary notebook in Figure 3 located in the directory `continuous_signals/`. Although the presented contents exhibit an hierarchical structure, e.g. by building on top of each other, the contents in the individual notebooks and chapters itself are kept as independent as possible. For instance, by repeatedly introducing abbreviations in individual notebooks. The reason is that individual notebooks may serve as entry points into the OERs when addressed from the internet and re-use/remixing of chapters is envisaged directly by this structure.

The computational examples are written in Python using packages commonly included in scientific computing (e.g. numpy, matplotlib, scipy). The examples typically illustrate the theory presented before, by analytic and numerical calculus and visualization through plots. Animations and audio examples are also used when appropriate. The

[5] `https://www.int.uni-rostock.de`

[6] `https://creativecommons.org/licenses/by/4.0`

[7] `https://opensource.org/licenses/MIT`

[8] `https://en.wikipedia.org/wiki/International_Standard_Classification_of_Education`

```
.
├── continuous_signals/
├── data/
├── discrete_fourier_transform/
├── discrete_signals/
├── discrete_systems_spectral_domain/
├── discrete_systems_time_domain/
├── discrete_time_fourier_transform/
├── fourier_transform/
├── laplace_transform/
├── periodic_signals/
├── sampling/
├── systems_properties/
├── systems_spectral_domain/
├── systems_time_domain/
├── z_transform/
├── LICENSE
├── README.md
├── apt.txt
├── conf.py
├── index.ipynb
├── ipython_kernel_config.py
└── requirements.txt
```

Figure 2: Folder tree indicating the organization of the educational materials for the Signals and Systems lecture. The notebooks of a chapter are located in the respective directories, the central notebook (`index.ipynb`) with links to the chapters is located on the top level. The remaining files contain general information on the materials, the license, the documentation of software dependencies and various configurations.

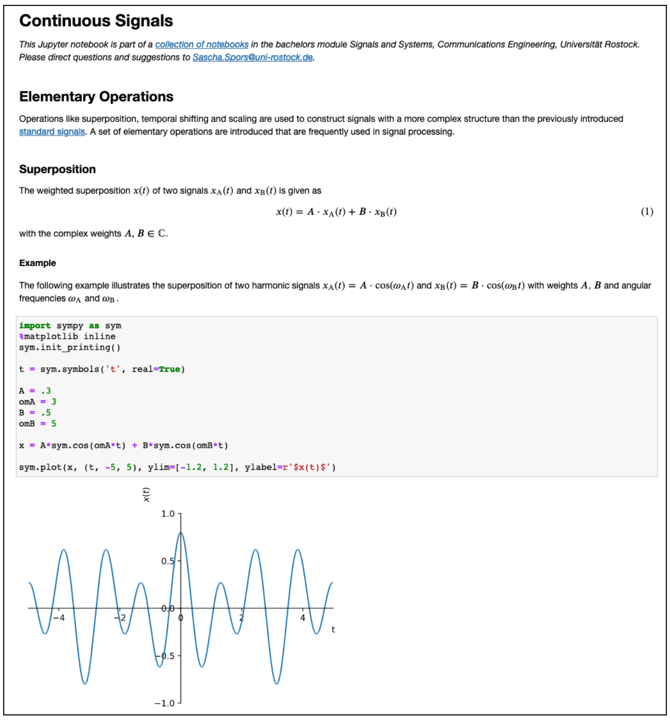

Figure 3: Exemplary Jupyter notebook illustrating elementary operations for the modification of signals. The top-level headline refers to the chapter, the sub-level headlines to the sections in this chapter. The notebook contains text including equations for the theory which is illustrated by a computational example with code generating a plot as output. The parameters in the code may be played with in order to gain understanding in their function.

module	ISCED[8]	ECTS	form	reference
Signals and Systems	6	6	lecture exercises	[9] [10]
Digital Signal Processing	7	6	lecture exercises	[11] [12]
Communication Acoustics	6	6	exercises	[13]

Table 1: Overview on realized OERs in the curriculum of Signal Theory and Digital Signal Processing, Institute of Communication Engineering, Universität Rostock, Germany.

requirements for the computational examples, in terms of Python packages, are given in the usual form of a file containing all dependencies (`requirements.txt`). This is essential in order to ensure that the OER produces the same computational results when re-evaluated. The notebooks are hosted on github.com[9] which provides a rich tool-set for provenance. Revision and author tracking is realized by the git version management system[10]. Issues and errors are tracked by issue tracking, new content is reviewed by pull/merge requests. Revisions of the entire collection of notebooks are marked as releases at reasonable intervals (e.g. semesters) and a digital object identifier (DOI) is attached by zenodo[11] for referencing. The computational examples in the notebooks are automatically executed at each revision in order to check their integrity by continuous integration. Encouraged is an interactive usage of the materials by local installation of the Jupyter web application. However, the notebooks can be visualized as a static version using nbviewer[12] and used interactively without local installation employing the binder service[13]. The collection of notebooks might also be exported as one integrated LaTeXdocument using the sphinx documentation system[14] with the nbsphinx extension[15]. Our overall experience with the Jupyter ecosystem including provenance is throughout very positive. The materials are endorsed by the students. In the course of authoring and editing the Jupyter notebooks we observed a number of shortcomings. These are related to the fact that Jupyter focuses on materials arranged in a single notebook instead of an entire collection of interlinked notebooks and the usage of markdown for document formatting. At its current state the referencing of equations across notebooks and the handling of citations is not very convenient. The JSON file format for the notebooks integrating text, code and output is not very well suited for detecting differences (aka. diffs) between different revisions of a notebook.

Conclusions

We briefly discussed the professional authoring, maintenance and distribution of technical OERs for higher education. The Jupyter ecosystem together with well-established tools and workflows for data provenance turned out to be well suited for this purpose. Jupyter has a very active community supporting developers and users in many aspects, and which is constantly extending the ecosystem. For OERs its seamless integration of descriptive text, math and source code, its widespread availability, and the existing tools and services make it a favorable electronic medium for technical contents. The formatting of Jupyter notebooks bases on markdown which limits the capabilities in terms of cross-referencing and citations. However, the simplicity of Jupyter notebooks is also beneficial in terms of sustainability and ease of re-use. The vision for the presented OERs is to establish high-quality educational materials by an open and collaborative process of creation and maintenance. At the current state, all materials have been authored by a single institution. Contributions from other authors working in the field and re-usage of existing contents in curricula of third parties are intended in the future. So far, OERs are not too widespread in higher education. A reason might be insufficient rewards for educational professionals in typical performance evaluation metrics [5]. For students, the presented OERs and their distribution provide the possibility to obtain additional skills in the field of data provenance, literacy, programming, as well as good practices for documentation of scientific results.

[9] `https://github.com`
[10] `https://git-scm.com`
[11] `https://zenodo.org`
[12] `https://nbviewer.jupyter.org`
[13] `https://mybinder.org`
[14] `https://www.sphinx-doc.org`
[15] `https://github.com/spatialaudio/nbsphinx`

References

[1] U. Bernath, A. Szücs, A. Tait, and M. Vidal, Eds., *Distance and E-Learning in Transition.* John Wiley & Sons, Inc., 2009.

[2] N. Rushby and D. Surry, Eds., *The Wiley Handbook of Learning Technology.* Wiley-Blackwell, 2016.

[3] UNESCO, "Forum on the impact of open courseware for higher education in developing countries: final report," Paris, France, 1-3 July 2002, document code CI.2002/CONF.803/CLD.1.

[4] D. Wiley and J. L. Hilton III, "Defining OER-enabled pedagogy," *The International Review of Research in Open and Distributed Learning*, vol. 19, no. 4, Sep. 2018, http://www.doi.org/10.19173/irrodl.v19i4.3601.

[5] F. Miao, S. Mishra, and R. McGreal, Eds., *Open educational resources: policy, costs, transformation.* UNESCO and Commonwealth of Learning, 2016.

[6] T. Pasquier, M. Lau, A. Trisovic, E. Boose, B. Couturier, M. Crosas, A. Ellison, V. Gibson, C. Jones, and M. Seltzer, "If these data could talk," *Scientific Data*, vol. 4, 09 2017.

[7] J. M. Perkel, "A toolkit for data transparency," *Nature*, vol. 560, pp. 513–515, August 2018, http://www.doi.org/10.1038/d41586-018-05990-5.

[8] M. Schröder, F. Krüger, and S. Spors, "Reproducible research is more than publishing research artefacts: A systematic analysis of jupyter notebooks from research articles," in *German Annual Conference on Acoustics (DAGA)*, March 2019.

[9] S. Spors, "Continuous- and Discrete-Time Signals and Systems - Theory and Computational Examples," visited 7th April 2020. [Online]. Available: https://github.com/spatialaudio/signals-and-systems-lecture

[10] F. Schultz, "Continuous- and Discrete-Time Signals and Systems - A Tutorial with Computational Examples," visited 11th April 2020. [Online]. Available: https://github.com/spatialaudio/signals-and-systems-exercises

[11] S. Spors, "Digital Signal Processing - Theory and Computational Examples," visited 7th April 2020. [Online]. Available: https://github.com/spatialaudio/digital-signal-processing-lecture

[12] F. Schultz, "Digital Signal Processing - A Tutorial with Computational Examples," visited 13th April 2020. [Online]. Available: https://github.com/spatialaudio/digital-signal-processing-exercises

[13] N. Hahn and M. Geier, "Communication Acoustics - Exercises," visited 7th April 2020. [Online]. Available: https://github.com/spatialaudio/communication-acoustics-exercises