ceweb.br soyuz

Implementation Documentation

Connected Didactic ePub

Introduction

About the project

Objective

To explore the potential of the open ePub format (format for digital publications) in educational settings.

Premises

This project will make use of the ePub format, which is the standard used within the Worldwide Web Consortium (W3C). We chose this format, not only because of the use of open standards developed by the W3C, but also because it is the standard required by the National Textbook plan (PNLD) for digital books.

This project contemplates the use of Web standards (such as HTML, CSS, and JavaScript), accessibility standards (WCAG and WAI-ARIA), data standards (DWBP and JSON), an ePub 3 standards.

Selected example book

We use the materials of An Open Book of Mathematics project

An Open Book - https://umlivroaberto.org/

Fractions: https://github.com/livro-aberto/fracoes_livro_piloto
Computing for High School: https://github.com/livro-aberto/livroabertoem

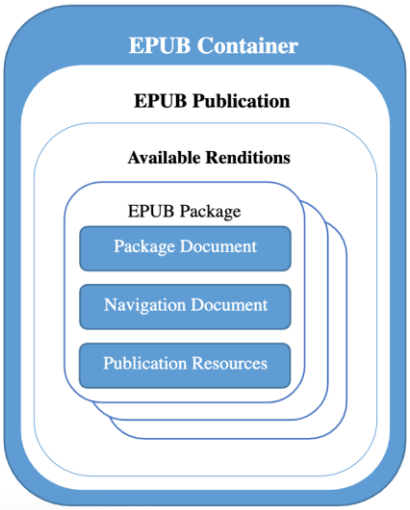
Technical development of the ePub

The ePub documents container file will follow the EPUB 3.2 format guidelines (https://www.w3.org/publishing/epub3/epub-spec.html), following the W3C specifications precisely, without modifying, enlarging, or reducing the scope of the specifications.

The specifications provide for the possibility of using scripts and multimedia elements, such as audio and video, and we can make use of these features as needed. In any case, the documents will support this type of resource/technology.

We will seek to implement the development within the proposal **EPUB for Education** of the W3C: https://w3c.github.io/publ-cg/education/epub-education.html

Figure 1: Structure of the EPUB format



Source: EPUB 3.2

xAPI and its freedoms:

1. *Freedom of statement:* The structure of “statements,” using nouns, verbs and objects, allows you to record almost any activity. Think, “I did this.”
2. *Freedom of story:* The Experience API allows LRSs to talk to each other. LRSs can share data and transcripts with each other, and their experiences can accompany you from one LRS (or organization) to another. Students can even have their own “personal data lockers” with their personal learning information inside them.
3. *Device Freedom:* Any enabled device can submit Experience API declarations (mobile phones, simulations, games, virtual reality, etc.). A constant network connection is not required - occasional connectivity is adequate.

4. *Workflow freedom:* Learning event tracking does not have to start or end in an LMS, it can start wherever the learner is and on any device they choose to use. Your content is not linked to an LMS.

Figure 2: Learning scheme with different tools

Source: xAPI

Semantics, ontologies and taxonomies

To connect ePub contents, we can use the following tools: Web semantics such as ontologies, controlled vocabularies, and taxonomies.
These tools aim to describe the content of objects contained in the ePub, allowing interoperability with external objects.

Application of metadata

Metadata will be applied in the publications to enable the connection between the works, their reuse, and the creation of new visualizations with the marked data. We are applying the metadata like the creation of a semantic library with the following possibilities:

Connected references: Metadata applied in the references of the books will allow the exposure of the data so that we can connect them into a new data source, generating examples of visualizations in Web applications.

Defining metadata

For interoperability to be possible, in addition to Web semantics tools, it is important that metadata standards be used to identify each type of digital object. We select the metadata based on the chosen publication, but we can only detail which metadata will be used during the development of the project, since it depends on prior the construction of the ePub document.
Below are some patterns that can be used in the project:

- EBUCore: Standard based on Dublin Core, aimed at media.
- FOAF: Standard for description of people.
- CDWA: Conceptual framework for describing and accessing information about works of art and architecture.
- VRA Core: Standard provides a categorical organization for the description of works of visual culture, as well as the images documenting them.
- Darwin Core: Metadata standard for information about the geographical occurrence of species and the existence of specimens in collections.
- EML: Standard developed for the discipline of ecology.
- IEEE LOM: Standard that specifies the syntax and semantics of Learning Object Metadata.
- CSDGM: Content standard for digital geospatial metadata maintained by the Federal Geographic Data Committee (FGDC).
- ISO 19115: “ISO 19115: 2003 Geographical Information-metadata” defines how to describe geographic information and associated services, including content, space-time purchases, data quality, access, and usage rights. It is maintained by the ISO/TC 211.
- TEI: Standard for the representation of texts in digital format, mainly in the areas of humanities, social sciences and linguistics.
- NISO MIX: “Z39.87 Data dictionary “- a technical metadata standard for digital images; (MIX) - NISO metadata for XML images that enables the management of digital image collections.
- <indecs>: Meets the need to place different creation identifiers and metadata in a framework to support intellectual property rights management.
- MARC: Standards for representing and communicating bibliographic and related information in machine-readable format.
- PBCore: Metadata and cataloging resource for public broadcasters and associated communities.

- MPEG-7: ISO/IEC standard that specifies a set of descriptors for various types of multimedia information. Developed by the Moving Picture Experts Group.
- DOI: Provides a system for the identification and management of information (content) in digital networks, providing persistence and semantic Interoperability.
- DIF: Standard for exchanging information about scientific datasets.
- RDF: A generic metadata architecture that allows you to represent information about resources on the World Wide Web (WWW or Web), such as title, author, and update date of a Web page, for example.

About ONIX

ONIX currently refers to any of the three standard XML metadata formats developed by EDItEUR for use primarily in the book trade. ONIX was originally a single standard for capturing and communicating bibliographic data related to books.

About Dublin Core

Dublin Core is a metadata schema that aims to describe digital objects such as videos, sounds, images, text, and websites on the Web. Dublin Core applications use XML and RDF.
https://www.dublincore.org/groups/education/

About Schema.org

Schema.org is a collaborative activity with a mission to "create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond."

EPub reader application

The project envisions an ePub reader application that supports the features defined in the previous items of this report. The preferential parameters for the selection of an already-existing application (which will serve as the basis) for the project are as follows:

1. Strictly follow all EPUB 3.2 specifications;
2. Support responsiveness;
3. Follow WCAG accessibility criteria.

We have selected two tool options that meet the set requirements. These are:

Lithium: EPUB Reader

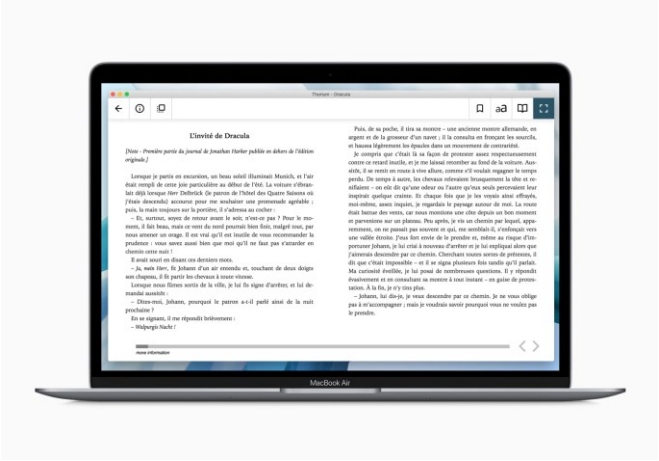
Figure 7: Lithium ePub reader demonstration

Source: Print-screen from the reader demonstration taken by the authors of this document

Description: EPub reader for Android smartphones with support for all features of this project

[Thorium Reader]

Figure 7: Lithium ePub reader demonstration



Source: https://www.edrlab.org/software/thorium-reader/

Description: EPub reader for Android smartphones with support for all features of this project

LRS application

Learning Locker is a compliant open source Learning Record Store (LRS), started in 2013 by HT2 Labs (now Learning Pool). It is a type of data repository designed to store learning activity statements generated by xAPI compliant learning activities (Tin Can).
https://github.com/LearningLocker/learninglocker

xAPI

"The experience API (xAPI) is an e-learning software specification that allows content and learning systems to communicate in a way that records and tracks all types of learning experiences.

xAPI defines web services that can be used by any plugin or APP to connect to the LMS in a standard way. The main actors in xAPI communication are:

- ***Client:** any plugin or APP that wants to use xAPI web Services*
- ***xAPI LRS (site server):** responsible for storing and serving the data issued by customers*

One good aspect about xAPI is that the amount of messages that can be shared between clients and LRS is quite limited. The main message type is called a statement, and all statements can be summarized as "an actor XX performs action YY on object ZZ."

The typical object in statements is a JSON element containing:

- ***Actor:** the person or group that does something*
- ***Verb:** the action performed by the actor*
- ***Object:** the object executed by the verb. There is more than one type of object that can be defined here, but for now, you can think of it as a specific "quiz attempt," for example.*
- ***Other fields:** for now, the rest of the xAPI fields have a basic data validation, so each plugin is responsible for implementing its own checks if necessary."*

https://docs.moodle.org/dev/Experience_API_(xAPI)

Security

The developed solution must comply with the Brazilian General Data Protection Law (LGPD).

Furthermore, our concern with the security aspect of the artifacts to be developed is based on three points:

SOP (Same-Origin Policy)

Same-origin policy is an important concept in the Web application security model. This policy allows scripts to run on pages that originate from the same site - a combination of schema, hostname, and port number - to access each other's DOM without specific restrictions but prevents DOM access on different sites. This policy also applies to XMLHttpRequests unless the server provides an Access-Control-Allow-Origin (CORS) header.

https://en.wikipedia.org/wiki/Same-origin_policy

Comentado [LG1]: Link substituído pela versão em inglês.

CORS (Cross-Origin Resource Sharing)

Cross-origin resource sharing (CORS) is a browser mechanism that allows controlled access to resources located outside a given domain. It extends and adds flexibility to the same-origin policy (SOP). However, it also offers potential for cross-domain-based attacks if a website's or webapp's CORS policy is poorly configured and implemented. CORS is not a protection against cross-origin attacks such as cross-site request forgery (CSRF).

https://portswigger.net/web-security/cors

Use of External Resources

The utilization of external resources can pose a potential security concern when considering the issue of malicious files that can be executed within the resources through this technology.

Technical standards used

- ePub 3.2 - https://www.w3.org/publishing/epub3/epub-spec.html
- ePub for education - https://w3c.github.io/publ-cg/education/epub-education.html
- Indexed Database API 3.0 https://www.w3.org/TR/IndexedDB/
- Web Storage API - https://html.spec.whatwg.org/multipage/webstorage.html
- HTML5 - https://html.spec.whatwg.org/multipage/
- Web Publication - https://www.w3.org/TR/wpub/
- Progressive Web App - https://developer.mozilla.org/en-US/docs/Web/Progressive_web_apps
- WCAG 2.1 - https://www.w3.org/TR/WCAG21/
- DPUB ARIA - https://www.w3.org/TR/dpub-aria-1.0/
- EPUB Type to ARIA Role Authoring Guide - https://idpf.github.io/epub-guides/epub-aria-authoring/
- Moodle - https://docs.moodle.org/dev/Experience_API_(xAPI)
- LTI - https://www.imsglobal.org/activity/learning-tools-interoperability
- Data on the Web best practices - https://www.w3.org/Translations/DWBP-pt-BR/

Controlled Vocabularies

Education

- https://www.uky.edu/~rsand1/china2018/texts/Bloom%20et%20al%20-Taxonomy%20of%20Educational%20Objectives.pdf
- http://inep.gov.br/thesaurus-brasileiro-da-educacao
- https://eur-lex.europa.eu/browse/eurovoc.html?locale=es
- https://vocabularyserver.com/tee/es/
- https://eric.ed.gov/?ti=all
- http://vocabularios.educacion.gov.ar/admin/escuelasdefrontera
- http://vocabularios.educacion.gov.ar/
- http://vocabularios.educacion.gov.ar/admin/tesauro
- http://vocabularios.educacion.gov.ar/admin/iscedp
- https://vocabularyserver.com/menvocab/index.php
- https://vocabularyserver.com/tee/en/
- https://vocabularyserver.com/tes/index.php
- http://vocabularios.educacion.gov.ar/admin/iscedf
- https://vocabularyserver.com/eurydice/en/

- https://vocabularyserver.com/scot/
- http://pergamum.inep.gov.br/pergamum/biblioteca/pesquisa_thesauro.php

Interdisciplinary or general subjects

- https://schema.org/
- http://vocabusp.sibi.usp.br/Vocab/Sibix652.dll/Assuntos
- http://vocabularies.unesco.org/browser/thesaurus/en/
- http://tesauros.mecd.es/tesauros/tesauros
- http://www.bnm.me.gov.ar/e-recursos/mercosur/ar/index.php
- https://www.vocabularyserver.com/lre/en/

Specific subjects

- **Mathematics**
  - https://vocabularyserver.com/tesamat/services.php
- **Biology**
  - http://doteine.uc3m.es/tesauros/biologia/services.php
  - https://vocabularyserver.com/cbdvoc/index.php
  - https://www.ufrgs.br/tesauros/index.php/thesa/terms/172
- **History**
  - http://vocabularios.caicyt.gov.ar/hitoscaicyt/index.php
  - http://vocabularios.caicyt.gov.ar/historiaargentina/index.php
  - https://www.vocabularyserver.com/historiamujeres/index.php
  - http://vocabularios.caicyt.gov.ar/ravignani/
  - https://vocabularyserver.com/historia_catalunya/
- **Geography**
  - https://www.vocabularyserver.com/toponimos/
  - http://vocabularios.educacion.gov.ar/admin/toponimos
- **Art**
  - https://vocabularyserver.com/espectaculo/index.php
  - http://www.mls.gov.br/download/voc/vocfot.pdf
  - http://139.162.178.118/pesquisa/pt/vocab/formulario.html
- **Health**
  - https://decs.bvsalud.org/E/homepagee.htm

- **Literature**
  - http://www.ufrgs.br/thesinfantojuv/
  - https://livroaberto.ibict.br/handle/1/1010
  - https://tesaurosjuventude.mdh.gov.br/vocab/?tema=292
  - https://vocabularyserver.com/motif/es/index.php
  - https://vocabularyserver.com/traces/ca/index.php
  - https://vocabularyserver.com/motif/en/index.php
  - https://vocabularyserver.com/atu/en/index.php
  - http://vocabularios.educacion.gov.ar/admin/culturainfantil

How to produce a Connected Didactic ePub

In this topic, we will explain the production process of a connected didactic ePub from the EPUB file result of our research.

What do we need?

Because it is completely based on open standards, the production of the ePub can be carried out with several tools. Here are some that we suggest:

Sigil

Sigil is a free, open source, cross-platform eBook publisher that uses Qt (and QtWebEngine). It is designed to edit books in ePub format (ePub 2 and ePub 3).
https://github.com/Sigil-Ebook/Sigil

Visual Studio Code

Visual Studio Code combines the simplicity of a code editor with what developers need for their core edit-build-debug cycle. It provides comprehensive code editing, navigation, and comprehension support along with lightweight debugging, a rich extensibility model, and lightweight integration with existing tools.
https://github.com/microsoft/vscode

Pandoc

If you need to convert files from one markup format to another, pandoc is your Swiss Army knife.
https://pandoc.org/

Ace by DAISY

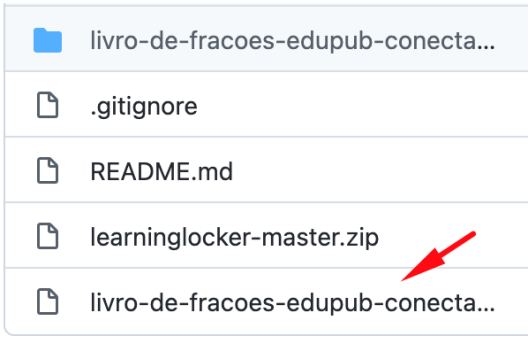
Ace by DAISY is a free, open source tool designed to check the accessibility of ePub files at any point in a publishing workflow. It was developed to assist in assessing compliance with the ePub accessibility specifications.
https://daisy.org/activities/software/ace/

Learning Locker

Learning Locker is a compliant open source Learning Record Store (LRS), started in 2013 by HT2 Labs (now Learning Pool). It is a type of data repository designed to store Learning Activity Statements generated by xAPI compliant learning activities (Tin Can).
https://github.com/LearningLocker/learninglocker

Exploring the EPUB3 format in Sigil

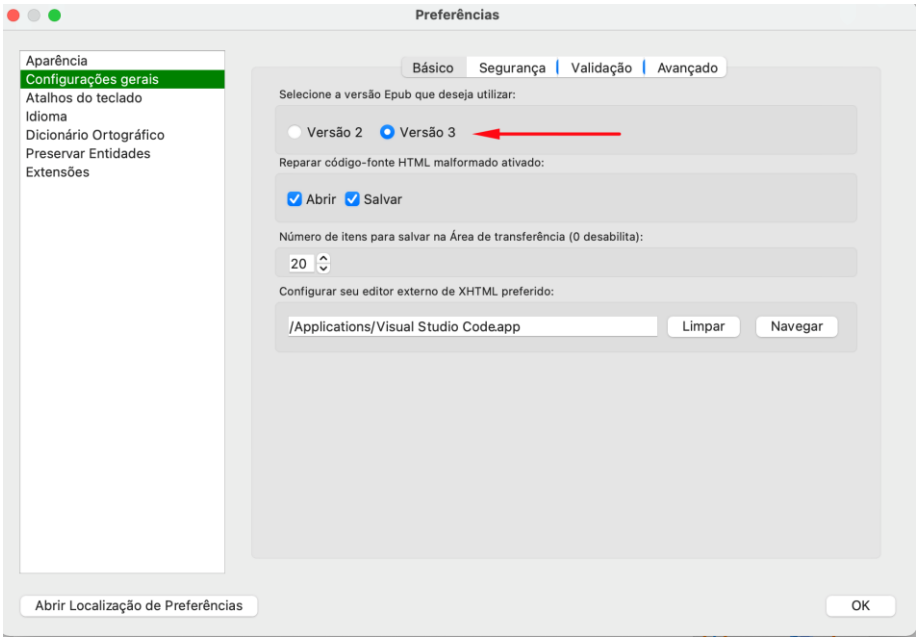
The best way to start your experiments is by looking at what we have already done. To this end, our suggestion is to download the EPUB that we created for the project in our repository: https://github.com/W3CBrasil/epub-didatico-conectado



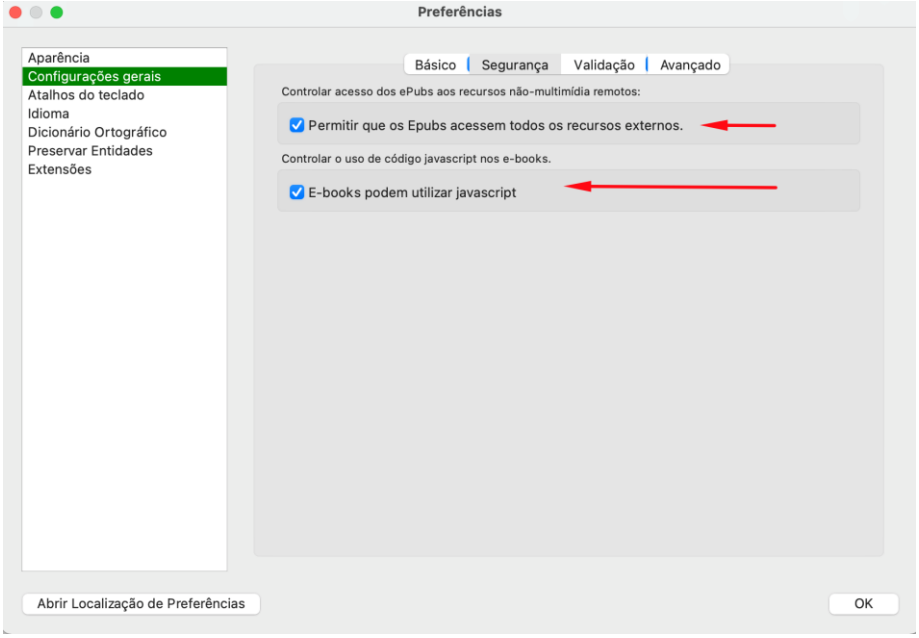
Preparing Sigil for interactive EPUB3

To allow Sigil to work correctly with the resources we need, it is important to pay attention to the following details:

In the general settings, choose the EPUB3 format as the default:



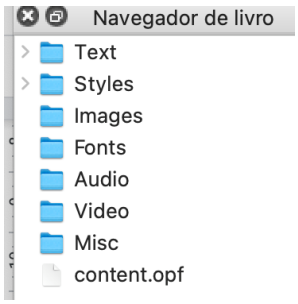
To enable interactivity, choose to run JavaScript on the EPUB:



Organization of EPUB3 in the Sigil standard

The Sigil pattern organizes files as follows:

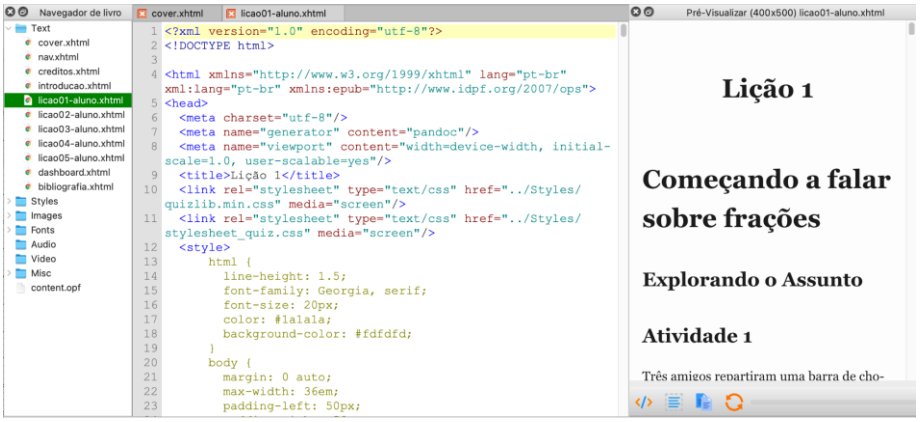
- **Text**: where the HTML files of the chapters are stored;
- **Style**: where the CSS files are stored;
- **Misc**: where JavaScript files are stored;
- **Images, Audio, Video**: where the respective media are stored.



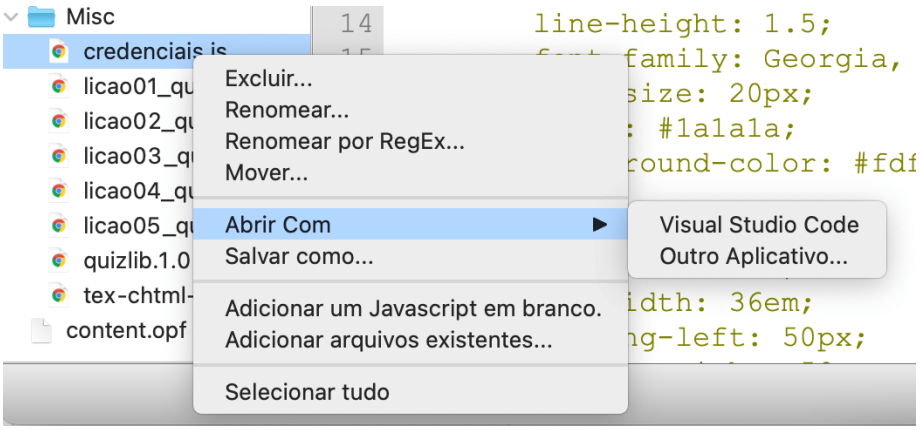
IMPORTANT: Never create subfolders in the organization. They will be ignored. Files should be on a single level. This is because EPUB is a ZIP-compressed format and unpacking multiple levels in real time is resource-intensive.

Editing the book

Sigil features three panels: navigation, editing, and preview.



Editing the files can be done in the editor itself or in an external code editor of your choice. In our case, we use Visual Studio Code as an auxiliary editor. To open a file in an external editor, right-click and indicate the editor where you want to work.



Since Sigil does not highlight JavaScript code, this is a safer way to work with this type of language.

Preparing files for ePub

ePub is nothing more than a way to organize and encapsulate files in Web standard with an editorial paradigm. EPUB3 supports all the technologies that HTML5 provides. The limitations, in most cases, lie in the readers or operating systems.

Thus, each chapter of an ePub is nothing more than an HTML page. It is important, therefore, to follow the semantic organization of HTML5 so that the result is a book with the experience expected by the reader.

Several tools can convert various file formats to HTML5. Some do so faithfully, others with a less rich experience.

In the case of the sample math book we produced, the pandoc tool was used to convert the chapters of the book in LaTex format to HTML5.

```
pandoc --toc inputfile.tex -s --mathjax -o outputfile.html
```

Although it is a command line tool, its use is not complex, and the resulting files are of high quality. Obviously, as with every conversion process, an edit to clean up the code and fix inconsistencies will probably be necessary.

A document on the subject can be found here:
https://www.homepages.ucl.ac.uk/~ucahmto/elearning/latex/2019/06/10/pandoc.html

Connecting ePub as the data repository via xAPI

As previously described, xAPI is a formatted data communication API for education vocabularies. It is simple, practical, and powerful.

Before going into the details of the implementation of xAPI in ePub, let's deal with how to install an LRS server that we will use as an example for data integration.

Installing the Learning Locker

For those who have an Amazon AWS account, Learning Locker already has an image ready to install.

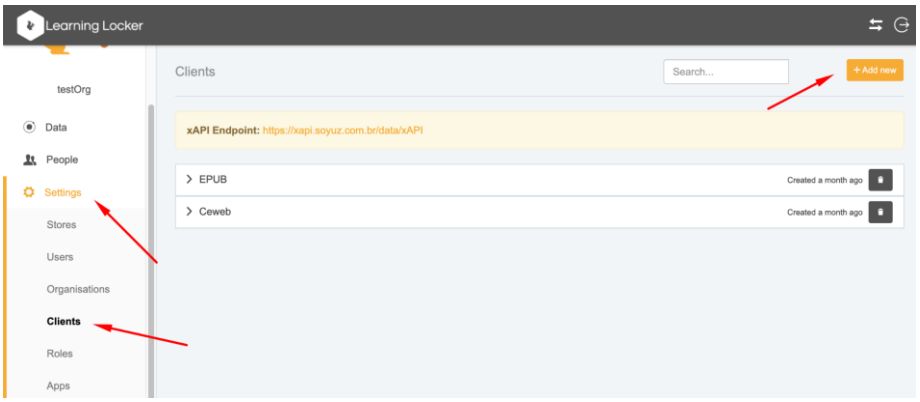
In case you prefer to install it on your own Linux server, you can do so through the following installation scripts:

wget-qO deployll.sh https://raw.githubusercontent.com/LearningLocker/deploy/master/deployll.sh & & bash deployll.sh

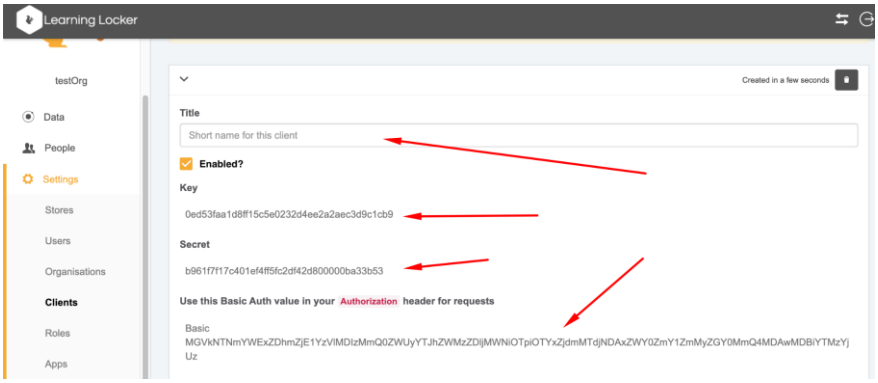
Detailed instructions on the installation process can be found in the Learning Locker document: https://docs.learninglocker.net/guides-installing/

Setting up an *endpoint xAPI* for communication with ePub

To set up communication with a customer, enter *Settings / Clients* and select *Add New*



Choose a name and save the connection credentials: **key,** often termed **login,** for the client, and **secret** is the **password**. The Basic Auth is a token with the combination of the keys. In our case, this is the information we will use.



Sending ePub information to LRS

In the case of our sample book, we set up an action so that the result of the exercises would be sent to the server each time students press the correction button.

To this end, we created a JavaScript routine inside the file responsible for correcting the activities.

```
fetch(xapi.endpoint, {
  method: 'POST',
  headers: xapi.headers,
  body: JSON.stringify(data),
})
.then((response) => response.json())
//Then with the data from the response in JSON...
.then((data) => {
  console.log('Success:', data);
})
//Then with the error genereted...
.catch((error) => {
  console.error('Error:', error);
});
```

Because it is a standard REST API, the xAPI can be accessed using a standard HTTP data communication routine, already natively supported by most browsers.

Let's dissect the anatomy of this request:

```
var xapi ={
    headers: {
       "Content-Type": "application/json",
       "X-Experience-API-Version": "1.0.3",
       "Authorization": "Basic M2QzYTZiNWJkY2UyOWQ",
      },
      endpoint: "https://xapi.xyz/data/xAPI/statements"
}
```

We created a variable to make it easier to manipulate the connection data. **Headers** are transmitted with the authorization information and with the relative protocols.

Endpoint is the address where the data will be sent. It is a two-way street, but in this case, we will only send data, via a POST request.

```
fetch(xapi.endpoint, {
  method: 'POST',
  headers: xapi.headers,
  body: JSON.stringify(data),
})
```

In this first part, we inform the connection address **xapi.endpoint.** HTTP transmission method: **POST,** the **xapi.headers** headers, and finally, the data of the xAPI **statements**: **date.**

Anatomy of an xAPI Statement

```
var  data = {
  "actor": {
    "name": aluno.nome,
    "mbox": "mailto:"+aluno.email
  },
  "verb": {
    "id": "http://adlnet.gov/expapi/verbs/answered",
    "display": {"pt-BR": "respondeu" }
  },
  "object": {
    "id": "epub://livro-de-fracoes/OEBPS/licao01-aluno.xhtml#"+quiz.container.id,
    "definition": {
      "type": "http://adlnet.gov/expapi/activities/question",
      "name": { "pt-BR": licao+" - "+atividade+" - Questão "+(no+1)},
      "description": {
        "pt-BR": question.querySelector('.quizlib-question-title').innerText
      }
    }
  },
  "result": {
    "success": correct,
    "response":question.querySelector('input').value
  },
  "context": {
    "instructor": {
      "name": professor.nome,
      "mbox": "mailto:"+professor.email
  }},
  "timestamp": new Date().toISOString(),
  "authority": {
    "name": professor.nome,
    "mbox": "mailto:"+professor.email
  }
};
```

An xAPI statement is a set of data in JSON format that follows a specific vocabulary for educational data storage. It is powerful and flexible. It is composed of three basic parts:

- **Actor**: the person carrying out the action. In our case, the student identification.
- **Verb**: this is what is expected from the action, in our case, to answer a question.
- **Object**: the didactic exercise or object where the action is being performed.

The other data, such as results, context, etc., are added according to the purpose of the activity and the metrics to be stored.

You can delve even deeper into the concept by taking a look at the following link:
https://xapi.com/statements-101/

Accessibility in EPUB3

By following Web standards, EPUB3 also contemplates compliance with the Web Content Accessibility Guidelines (WCAG).

To this end, the DAISY consortium, which has been providing tools to promote accessibility for digital publications for many years, has created a tool that greatly facilitates the verification of the ePub accessibility guidelines: Ace.

With Ace, you just need to "throw" the .epub file inside the tool and it will produce a complete compliance report:

Resumo

Tipo	Crítica	Séria	Moderada	Baixa	Total
wcag2a	0	0	0	0	0
wcag2aa	0	0	0	0	0
EPUB	0	0	0	0	0
Boas Práticas	0	0	0	0	0
Outro	0	0	0	0	0
Total	0	0	0	0	0

Although extremely useful, the tool fails to evaluate all accessibility features. Some caution is needed to ensure that the publication is truly accessible:

- Images must have alternative text
- Forms must have labels related to their fields
- Check the color contrast between text and background