**Open DS4All Report**

**Open@RIT**

Sebastian Moreano Mesa, Michael Nolan, Rachel Shek

**OpenDS4All**

Xumin Liu, Andre de Waal

## Contents

[**Contents**](#_2hqe4zz6cmza) **1**

[**Executive Summary**](#_val29x5kyllx) **3**

[**Key Findings**](#_px9tgvtuuk8a) **3**

[**Process**](#_97aelewfcy14) **4**

[Research](#_xfsktpkmu9e0) 4

[About OpenDS4All](#_hq9ti72u4w7o) 4

[Reviewing the OpenDS4All Repository](#_ut8yrctcxzz1) 4

[Main Repository README.md](#_9zpnko18498j) 4

[New Module Staging Area README.md](#_3nzwtv45qdyw) 7

[Naming Conventions](#_lf4yhp3919jn) 9

[Minor Typos](#_irg2dm3uo1ry) 9

[Overview Converting Slides Into the OpenDS4All Format](#_12n70qnu9uow) 11

[Pros and Cons of using a template](#_h6avz2qlkfif) 12

[**Our Suggestions**](#_sa5abi2fn780) **13**

[Roadmap](#_32h0e7n9hgyw) 13

[Maintenance Plan](#_hkdu6aiocgkv) 14

[Other Suggestions](#_qimog8sggg4y) 15

[**Sections to consider**](#_41ze91rjgjfh) **15**

[**Resources**](#_90rzh0amswr8) **16**

[Our Links](#_3p186v4xxarm) 16

[Links that aid in contributing that should be grouped:](#_uykvuaush0lv) 17

[Helpful Links](#_4rm2b8pboepj) 17

## Executive Summary

As OpenDS4All fills the niche for offering students a combination of lectures, recitation or flipped classroom activities, and hands-on assignments (the best way for many students to learn), creating a strong base for the open source repository will allow the project to grow without needing an overhaul in the future.

## Key Findings

|  |  |
| --- | --- |
| **Finding 1** | Users, especially newer ones considering using the curriculum, don’t know how to navigate the repository. Currently, the content requires time to look through and is not easy to skim. |
| **Finding 2** | As an open source project that will grow from the community, there’s currently no structure and formatting that will be consistent and create a visual connection to OpenDS4All. |
| **Finding 3** | Users may be wary of contributing because they’re not aware of all the requirements/standards to contribute curriculum materials to the project, lack a community to connect to and want to help, etc. |
| **Finding 4** | An open source project longevity is in part reliant on the ability for that community to connect and communicate with each other. That being said, channels of communications such as Slack are needed so members of the community can communicate together outside of having direct contact with maintainers. |

## 

## 

## Process

### Research

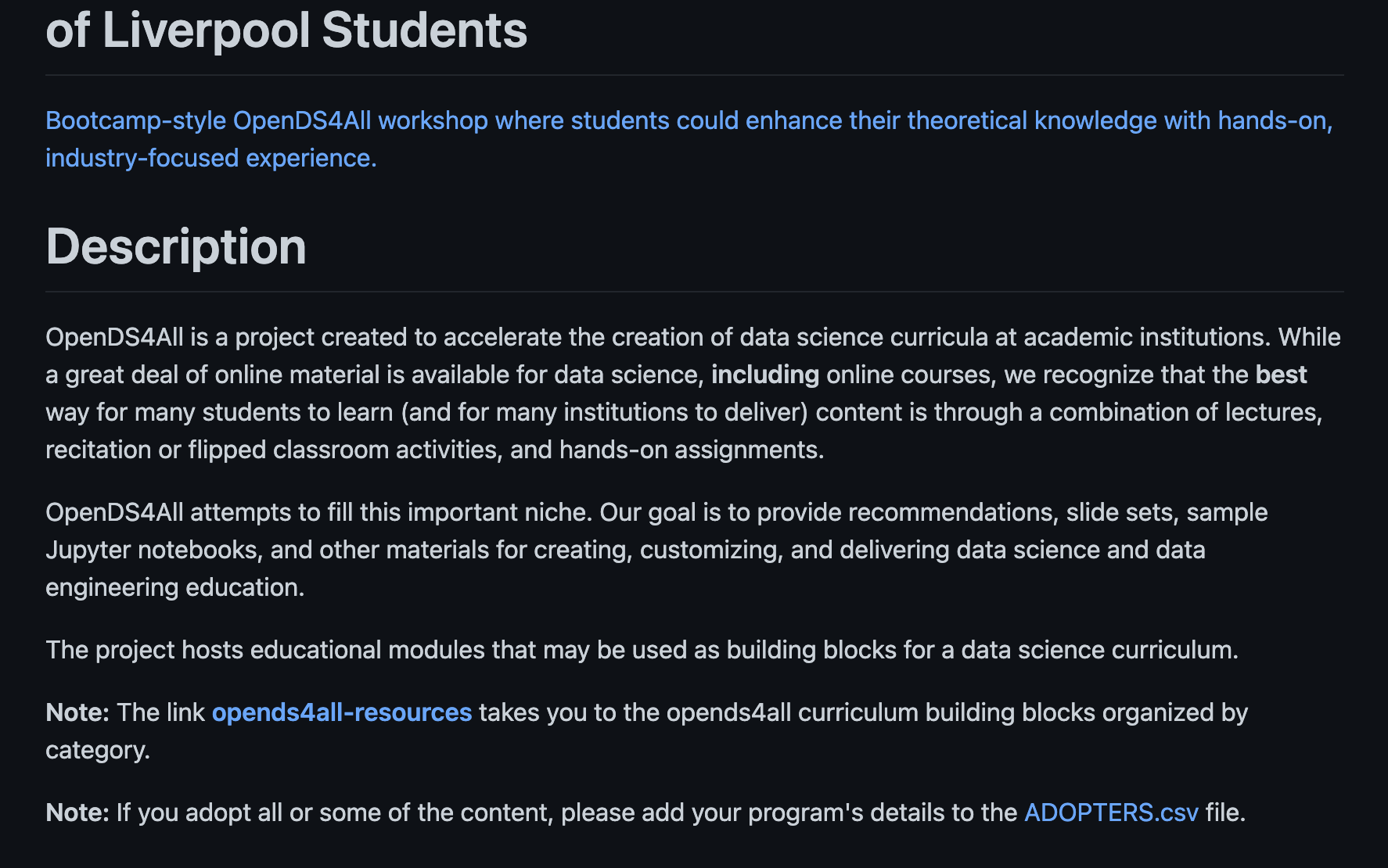
#### About OpenDS4All

OpenDS4All is an open source project aimed at providing a data science curriculum by professors for professors. In doing so, they hope to accelerate the development of data science curricula at academic institutions. The Modules within Github are what makes up the core of the curriculum. Serving as building blocks for teaching core concepts of data science, the project's goal is for schools to use these to supplement, strengthen, and start-up their data science programs.

#### Reviewing the OpenDS4All Repository

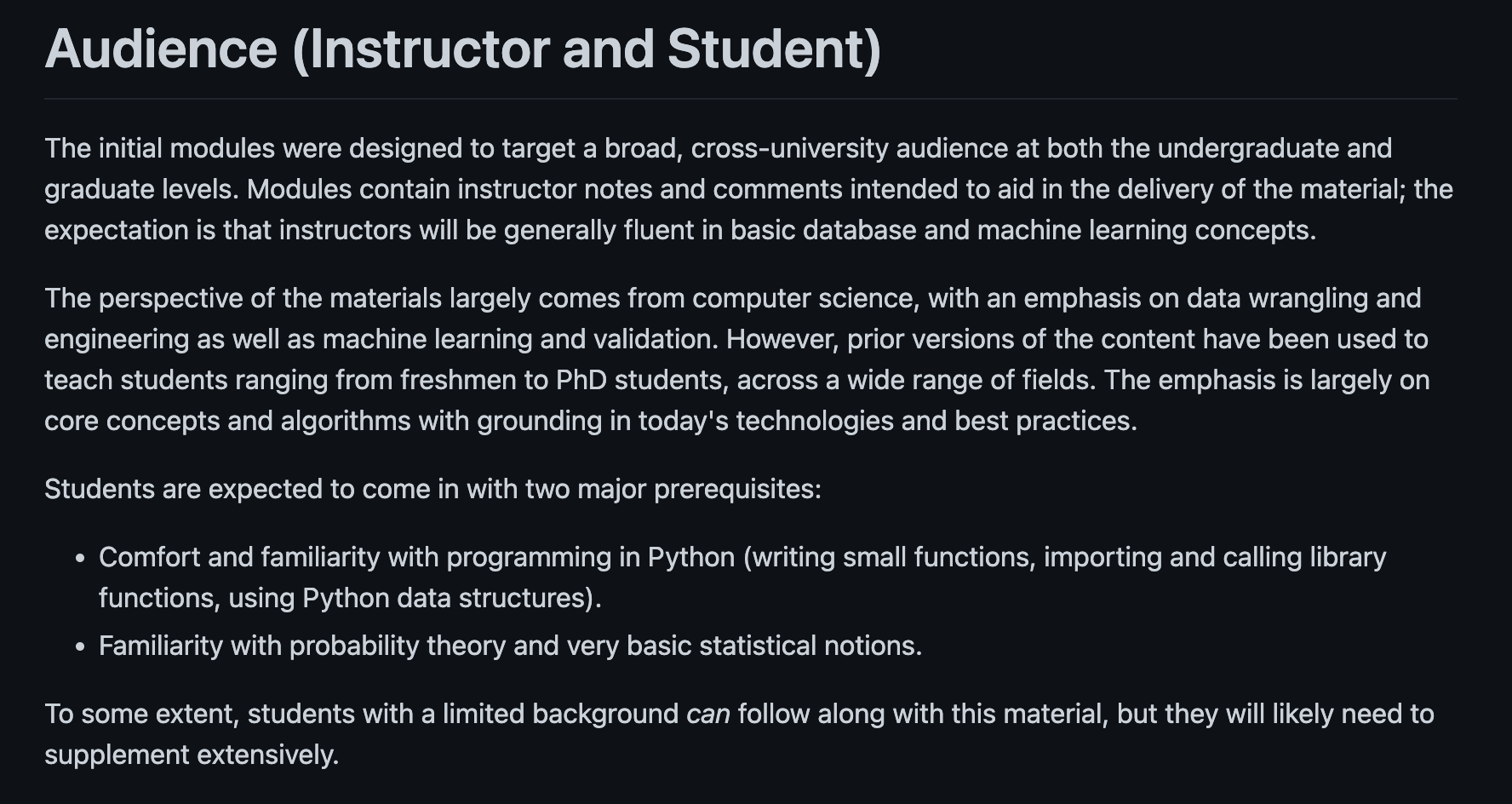
Overall the current OpenDS4All provides potential users and contributors access to a wealth of information and insights on data science, as well as providing a very useful foundational curriculum built around modules. To further improve it, we suggest doing some reorganizing/restructuring of certain areas (to make it easier to navigate and understand), adding certain information (to aid users), and taking steps towards attracting more community.

##### Main Repository README.md



Expand the description to include how and why what they are doing will help create a data science curriculum.

* A section of creators and where the idea for DS4All came from would be helpful for people to gain more of a personal connection with the project and better understand the reason behind the goals of the project. This could also be a good place to highlight its recent award for AI Innovator of the Year



Defining the audience is a good starting point to show what type of relationship people in the community might have, however this can be expanded further to improve useability

* Define the roles of what Instructors and students are within the context of the curriculum
* Expectations/goals for the student/instructor.
* Instructions on how the material should and shouldn't be taught

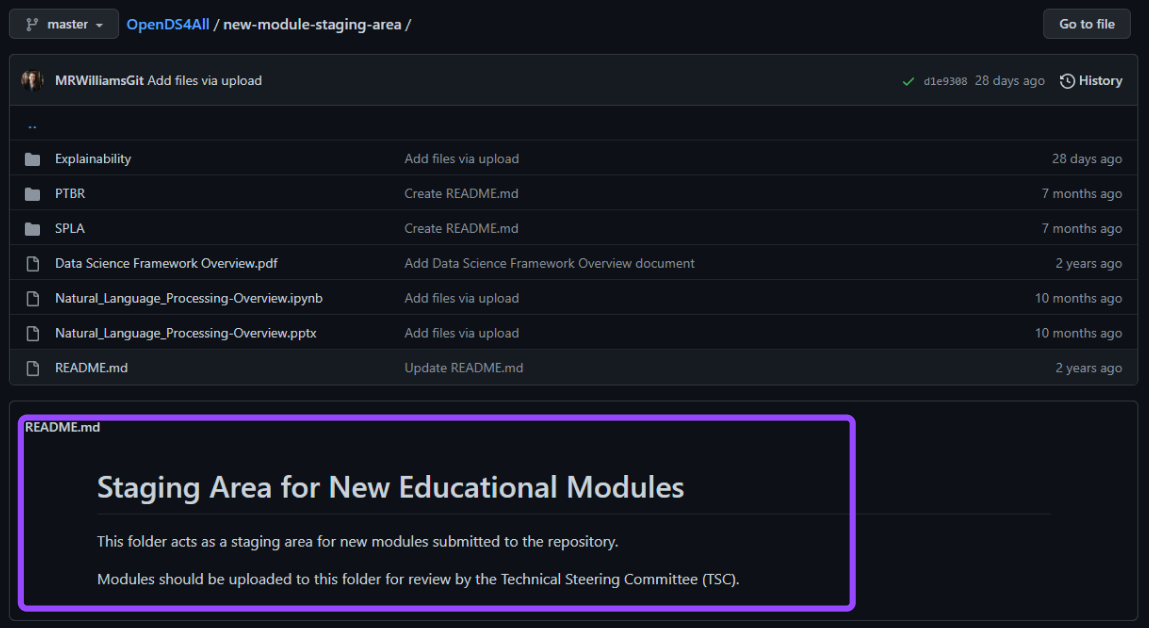


The current topology image used to show the category organization is an image that gets pushed onto the repository. With it being angled and the image shrinking the text smaller to fit all of the content, it’s hard to read and users may skim over it rather than even try.

Solutions to this may include:

* Creating a dropdown containing a list or table of all the topics (organized) – may be time consuming to do currently, but would be more manageable to do with incoming curriculums (see: [github formatting](https://docs.github.com/en/get-started/writing-on-github/working-with-advanced-formatting/organizing-information-with-tables) and [sample git](https://github.com/ras2669/test))
* Using [Mermaid](https://mermaid-js.github.io/mermaid/#/) code – would create auto-updating diagrams with code, is pretty friendly for even non coders, and [has a live editor](https://mermaid-js.github.io/mermaid-live-editor/edit#pako:eNpVkE1qw0AMha8itEohvoAXhcZOsgk0kOw8XgiPkhnC_DCWKcH23TuOCbRaSXrfewiN2AXNWOI9UTRwrZWHXF9NZZLtxVHfQlF8TkcWcMHzc4Ld5higNyFG6-8fK79bIKjG04IxiLH-Ma9S9fJ_e56gbk4UJcT2r3L9CRPsG3s2Of6_YhJn16G5UXmjoqMEFaUWt-g4ObI6nz0uBoVi2LHCMrea0kOh8nPmhqhJeK-thISlpIG3SIOEy9N373llakv5A25dzr-75Fn6) that can be worked if needed

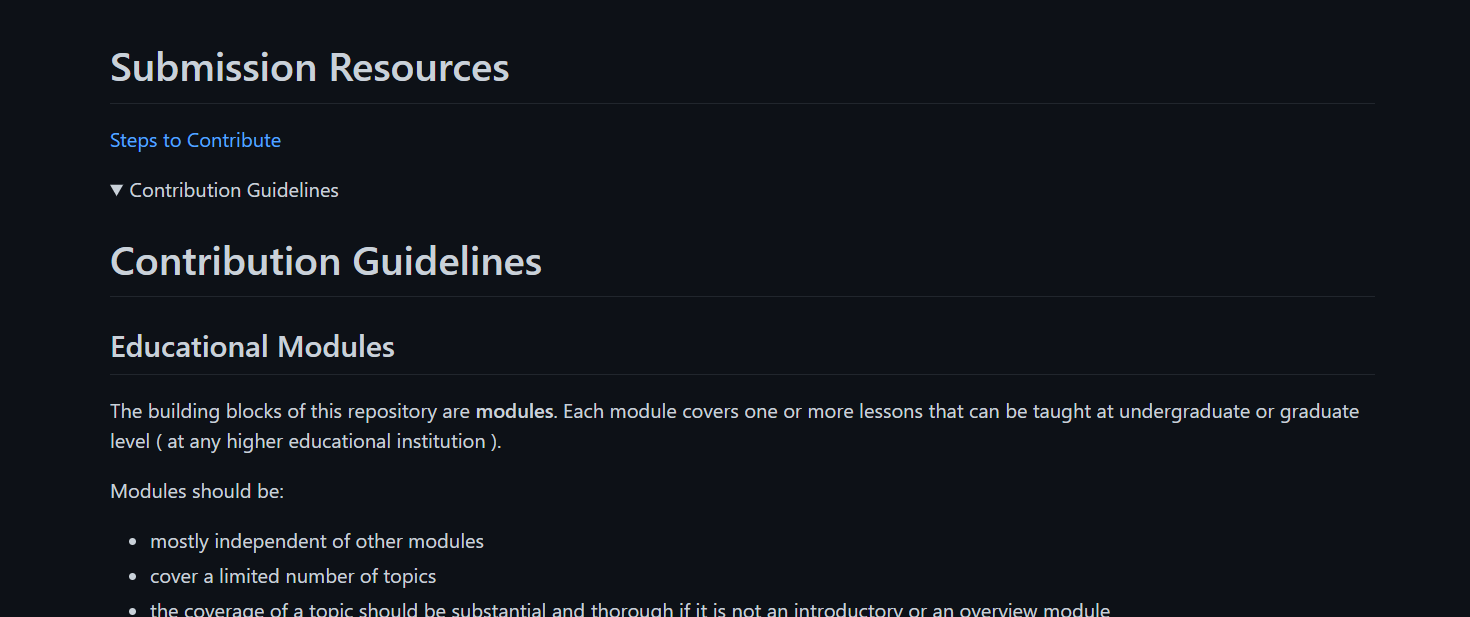
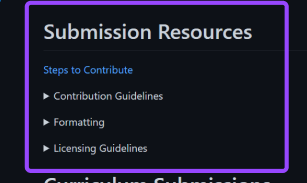
##### New Module Staging Area README.md



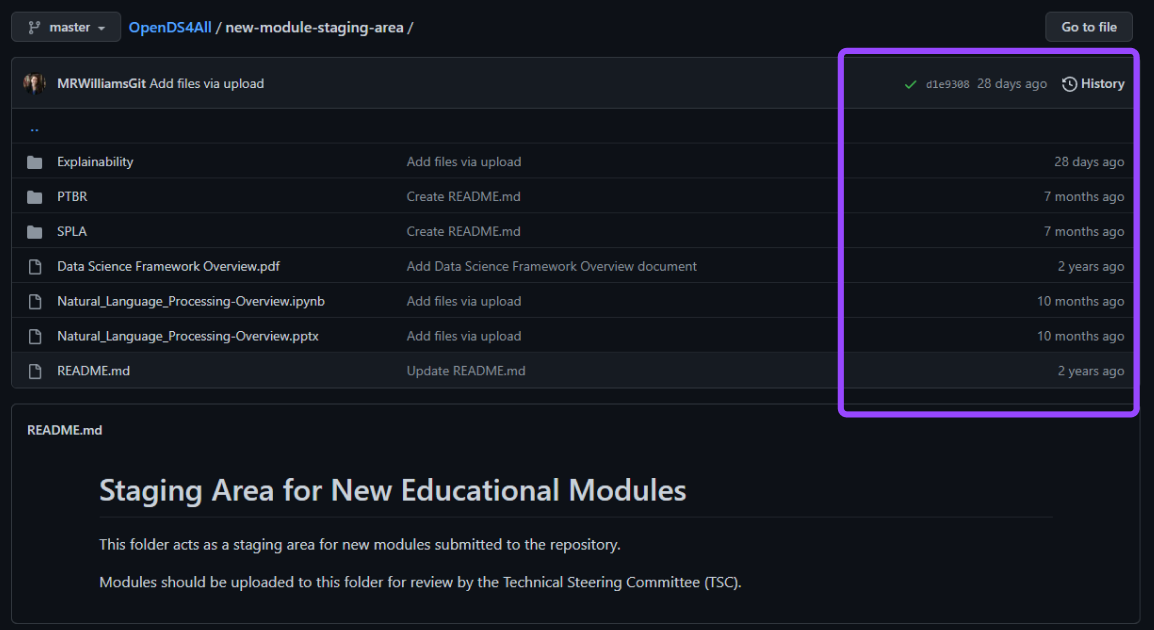
The README.md for the new module submissions is unhelpful and forces the contributor to:

* Go back to the main page to find the requirements
* Need to read through the entire directory thoroughly to find all the requirements
* Possibly submit files incorrectly and be required to come back and make changes/resubmit

Adding the requirements (see suggested links in Resources section) to this area and corresponding links will speed up the process on both ends and make it less overwhelming for contributors, and by putting them in collapsible boxes they have the ability to open and view as needed

.

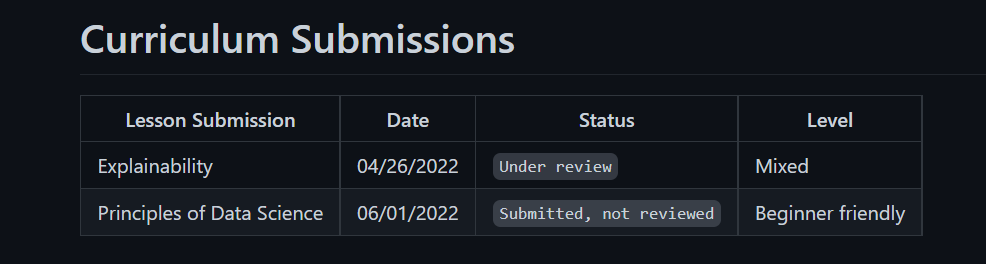
*Left, collapsed version, right, contribution guidelines tab is expanded*



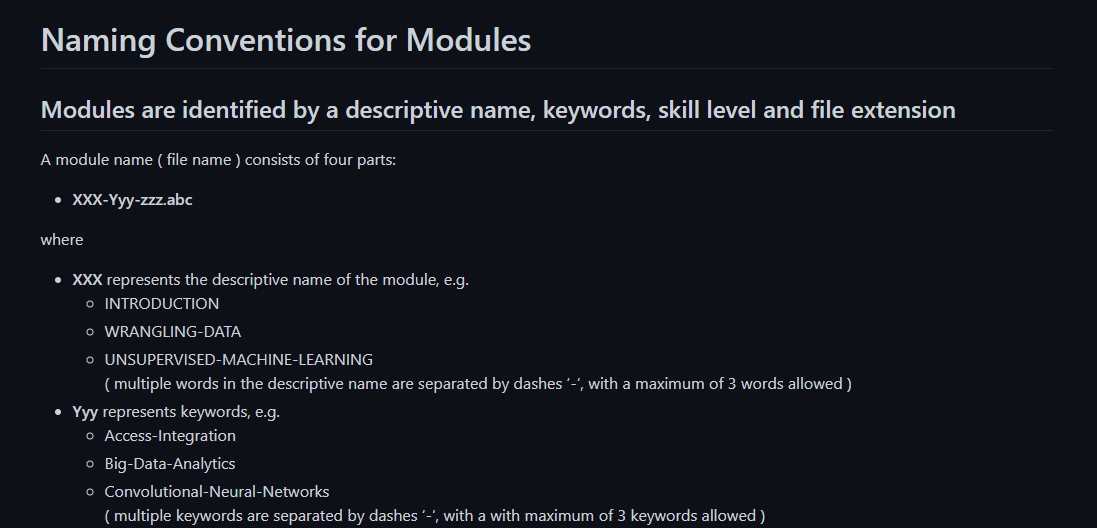
With timestamps marking each action, a contributor may be wary of taking the time and work to format their curriculum to submit because they may wonder questions like:

* When was the last time the TSC reviewed a curriculum?
* Are there people ahead of me that are not being processed or are their submissions rejected?
* If I were to submit, when would I know it’s approved?

Adding a table can be helpful to contributors and TSC to quickly be able to see what’s happening and what needs to be worked on (see example below [or on sample git](https://github.com/ras2669/test/blob/main/submission%2B.md)).



##### Naming Conventions

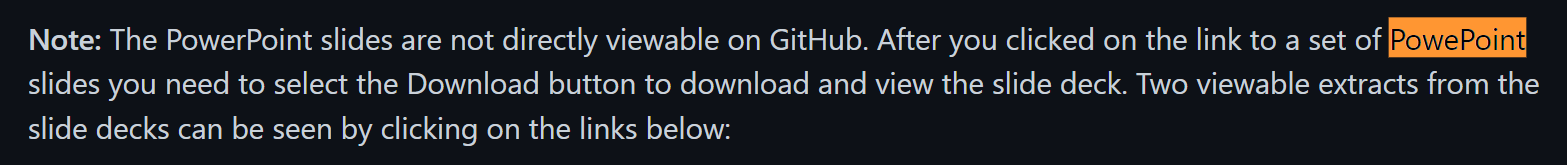
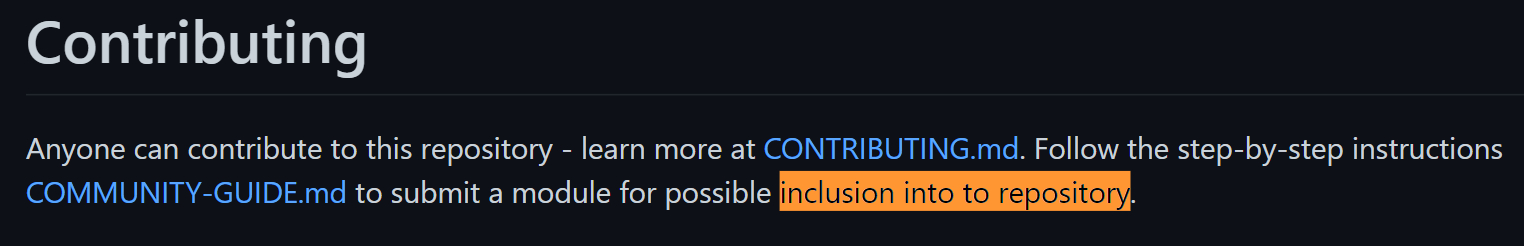


Reconsider creating a structure that is more friendly to students of different abilities. We suggest breaking subjects down further into four groups (introductory, beginner, intermediate, advanced) and creating a table or definitions for what each of these levels mean for a professor as different professors may consider their students intermediate but they may only have beginner knowledge based on your definition and categorizations of the subjects)

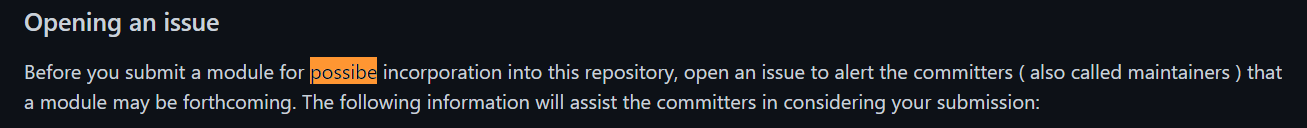
* Xumin classified her curriculum as introductory, setting a standard for future introductory content. What makes it introductory?

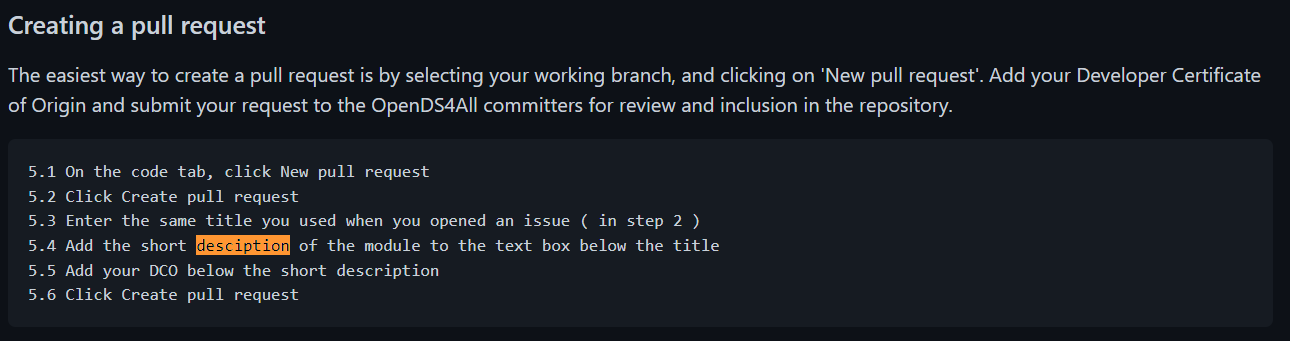
##### Minor Typos

Location: <https://github.com/odpi/OpenDS4All/blob/master/README.md>

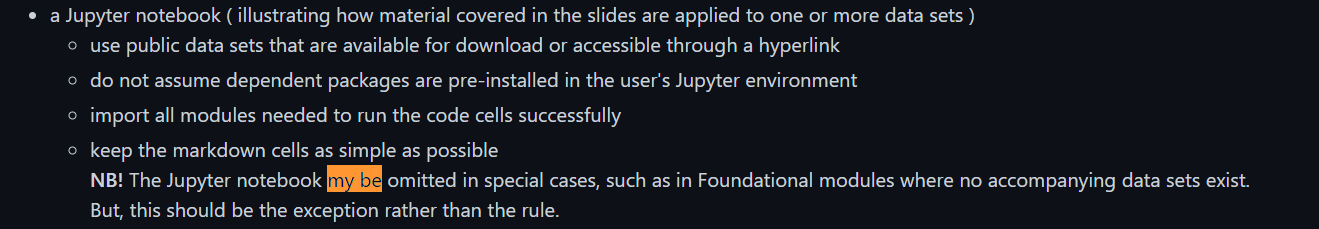


Location: <https://github.com/odpi/OpenDS4All/blob/master/COMMUNITY-GUIDE.md>

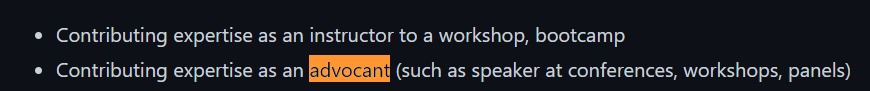




Location: <https://github.com/odpi/OpenDS4All/blob/master/CONTRIBUTING.md>



Location: <https://github.com/odpi/OpenDS4All/blob/master/CONTRIBUTORS.md>



**Can the Code of Conduct be linked directly from the main directory?**

Currently it goes from: the [main directory](https://github.com/odpi/OpenDS4All) to [what a user thinks is the code of conduct but is a file with a link to the code of conduct](https://github.com/odpi/OpenDS4All/blob/master/CODE_OF_CONDUCT.md) to [the code of conduct](https://github.com/odpi/tsc/blob/master/CODE_OF_CONDUCT.md) which adds an unnecessary step.

#### Overview Converting Slides Into the OpenDS4All Format

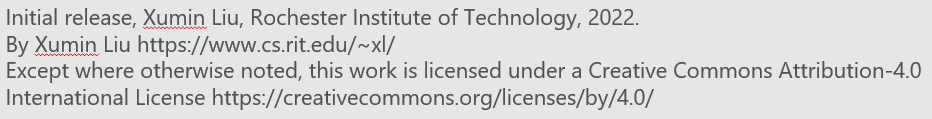
We used Professor Xumin Liu’s Data Science Curriculum to test out having professors convert their slides into DS4All templates. The conversion of a single set of slides took anywhere from 30 minutes to around one hour (or slightly longer, depending on the lesson slide length) and her curriculum was 15 powerpoints long. This doesn’t include the time Professor Liu took to add teaching notes to the slides or time taken for the final powerpoints to be reviewed by the committee.

**Things we encountered in our experience reformatting:**

* Basic formatting issues (IE the word Yes being shown as Ye (line break) s)
* Needing to add a contents slide and slides for topic changes + needing to decide what was important and what wasn’t
* Body content not fitting on the slide if using a consistent size/needing to push the title up on the slide
* If unfamiliar with Powerpoint, needed to research how to edit the master slide template in order to add credits
* Links such as to the creator or to the Creative Commons are not clickable as being part of the template
* Powerpoints distort in Google Drive due to different size slides (when downloaded and used in Powerpoint they’re fine) – this won’t apply to future Powerpoints as they will be submitted through Github, but is noted here for this situation

**Some of these issues are solvable:**

* Follow a [quick tutorial](https://support.microsoft.com/en-us/office/customize-a-slide-master-036d317b-3251-4237-8ddc-22f4668e2b56) on how to edit the master template
* Allow for content to take up more than one slide when needed
* To make sure viewers can go to links for the creator/Creative Commons, add them in the speaker's notes of the first slide (see below)



Sample copyright text in the speaker’s notes/first slide of each Powerpoint

##### Pros and Cons of using a template

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| * Creates uniformity * Easier for contributors to look at references * Streamlined and ready to use as foundation for any curriculum * With guidelines will ensures that the content is legible in the classroom | * Is an extra step in formatting requirements * Can be time consuming * Pasting content from premade slides can have distortion * Can be difficult if doesn't already follow a standard academic powerpoint structure * Requires the committee to heavily review the slides and make sure it meets requirements * Requires learning if unfamiliar with Powerpoint and master slides * To edit credits to the contributor, requires the user to edit multiple pages of the master template |

## 

## 

## Our Suggestions

#### Roadmap

We have created a Roadmap for DS4All of guidelines, suggestions and changes that should be implemented. These are a combination of long term and short implementations that in the long run will give the project longevity and prospective to grow the community. [Here is an example](https://github.com/KirstieJane/STEMMRoleModels/issues/1) of a roadmap on GitHub so that contributors can see the project’s status.

**Suggestions/Changes**

**Short Term** :

* Restructuring the README.md (see our suggestions for the [main repository](#_9zpnko18498j) and the [new staging module area](#_3nzwtv45qdyw))
* Edit [minor typos](#_irg2dm3uo1ry)
* Decision on whether or not the Powerpoint template will be a permanent use ([consider the pros and cons mentioned here](#_h6avz2qlkfif))
* Update Topology (see directory [example 1](https://github.com/ras2669/test/blob/main/topology.md) and [example 2](https://github.com/seb98mm/test/blob/main/topology.md))
* More forms of communications.
  + As it stands right now there is only one form of contact which is an email to contact the maintainers.
  + There should be some sort of group chat channel such as a slack channel, IRC or a discord. Something that allows the community to have communication between each other in a monitored space.
  + A contact form that could help disperse where inquiries go for better organization

**Long Term** :

* Create a [Maintenance Plan](#_hkdu6aiocgkv)
* Continue to improve contribution experience
  + Collect all necessary guidelines into one group and make them accessible in necessary areas
  + Decide on whether or not using OpenDS4All’s template will be a requirement
    - If yes, create a blank sample and link or create an image step by step on how to edit the template’s credits
  + Create transparency and show a path for contributor’s work (letting them know the steps that their work gets processed, IE they submit, it is reviewed, it’s either rejected, processed, or asked for edits, etc.). Let them know what will happen with their work (IE a curriculum may be split and sorted into corresponding subjects)
* Consider using releases
  + Questions to consider and document the answers to:
    - When should you do a release? (updated content wise)
    - What will need to be updated in a release?
    - Where can people see a list of changes?
    - Who will be responsible for keeping track of releases?
    - How is a release passed?

##### Maintenance Plan

A maintenance is what keeps a project and allows it to stay accessible to the community. This is related to the roadmap but is not a roadmap. The roadmap being what work needs to be done, while maintenance can be viewed as a set of guidelines on how to go about doing such work as well as defining roles and processes to maintain, update and what to do when you deprecate or sunset and archive your project. Here is an example of how to set up a [Project Maintenance Plan.](https://mozilla.github.io/open-leadership-training-series/articles/open-project-maintenance/open-project-maintenance/) The guidelines below are some general thoughts to keep in mind when creating a project maintenance plan.

Guidelines

* Always provide clear access and navigation
* Provide a comfortable environment for people to communicate
* Review documentation and content for areas that might not be clear to newer members or outside users
* Keep contact information updated and set a time and day if needed to make sure it is being checked
* When reviewing submitted lessons, be specific when possible or give examples if something needs to be changed

#### Other Suggestions



Add a background to the header image for legibility in dark mode

## **Sections to consider**

**Welcome Statement**- A welcome statement helps people quickly understand the scope, goals and the community behind the project.

**Table of Content-** Is crucial to having in the very beginning to allow anyone browsing to have a quick understanding of the project as well as being able to navigate easily. Furthermore it provides at a glance how the Github repository is structured.

**Glossary** - With modules ranging from beginner to experienced, there will also be people browsing with varying skill levels. As such it would be advantageous to define important vocabulary that is used so the community doesn't get stuck or confused when something unfamiliar comes up.

**Get Involved/What We Need/Expand Contribution** - This would all really fall under Contribution, But having a quick overview of the contributing process would be good so people interested have an idea of the general expectations and path to contribution. Then they can delve deeper into different sections related to contributing and the specifics in new module submissions. This should also be linked here

**Find Out More** - DS4All is a open source project, it likely also used other open source and non open source tools/materials/references or is relevant to it. That being said a find out more is a good place to link all of this so people can have other material to look at and aid them in data science

**Road Map** - Here is where people can investigate issues that they contribute to and view milestones. We have a more detailed Road Map down below for what should be done for the project down the road.

**Thank You** - It is always nice to thank people for taking time to look through, use, contribute and join the community, these actions are what keeps open source projects running.

## Resources

#### Our Links

**Our FigJam notes file:** <https://www.figma.com/file/iXdp2pHslsc4Yl61yyybsr/Initial?node-id=0%3A1>

**Our Github Samples:**

* [Topic Folder structure](https://github.com/ras2669/test/tree/main/opends4all-resources/principles-of-data-science)
* [new-module-staging-area README.md](https://github.com/ras2669/test/blob/main/submission%2B.md)
* [Sample Mermaid diagrams to show directory structure](https://github.com/ras2669/test/blob/main/topology.md)
* [Sample Mermaid to show module structure](https://github.com/seb98mm/test/blob/main/topology.md)

**Sample Curriculum Repositories:**

* <https://github.com/microsoft/Web-Dev-For-Beginners>
* <https://github.com/HackYourFuture/curriculum>
* [STEM Role Model App](https://github.com/KirstieJane/STEMMRoleModels)

##### Links that aid in contributing that should be grouped:

* <https://github.com/odpi/OpenDS4All/blob/master/NAMING-CONVENTIONS.md>
* <https://github.com/odpi/OpenDS4All/blob/master/CONTRIBUTING.md>
* <https://github.com/odpi/OpenDS4All/blob/master/COMMUNITY-GUIDE.md>
* <https://github.com/odpi/tsc/blob/master/process/contribution_guidelines.md#developer-certificate-of-origin>

#### 

#### Helpful Links

* [About Mermaid, the open source JS tool that lets you create diagrams and visualizations using text and code](https://mermaid-js.github.io/mermaid/#/)
* [Mermaid Live Editor](https://mermaid-js.github.io/mermaid-live-editor/edit#pako:eNpVkE1qw0AMha8itEohvoAXhcZOsgk0kOw8XgiPkhnC_DCWKcH23TuOCbRaSXrfewiN2AXNWOI9UTRwrZWHXF9NZZLtxVHfQlF8TkcWcMHzc4Ld5higNyFG6-8fK79bIKjG04IxiLH-Ma9S9fJ_e56gbk4UJcT2r3L9CRPsG3s2Of6_YhJn16G5UXmjoqMEFaUWt-g4ObI6nz0uBoVi2LHCMrea0kOh8nPmhqhJeK-thISlpIG3SIOEy9N373llakv5A25dzr-75Fn6)
* [Editing the Master Slide in Powerpoint](https://support.microsoft.com/en-us/office/customize-a-slide-master-036d317b-3251-4237-8ddc-22f4668e2b56)
* [Mozilla Open Leadership Series](https://mozilla.github.io/open-leadership-training-series/articles/readme/)