Catalyst Cooperative 2024 Google Season of Docs Case Study

Organization or Project Name:
Public Utility Data Liberation Project (PUDL)

Season of Docs link: https://catalyst.coop/work-with-us/pudl-season-of-docs-project-proposal/

Organization Description:
The PUDL Project cleans and connects data about the US energy sector originally published by government agencies. The output is a machine-readable, analysis-ready database that researchers can dive into and begin using without engaging in the toilsome process of data preparation. This enables researchers, journalists, and non-profits to make informed assumptions about the power grid and level the playing field between utilities and intervenors in the regulatory process.

Authors:
Austen Sharpe

Summary

# Tech Writers	TW Project Hours	Budget	% Project Completed
1	160	$8,500	30%

What problem were you trying to solve? And how did you try to solve it?
Our goal was to make PUDL documentation cater to different user profiles and minimize duplicate information across documentation sources including our Read The Docs page, the website, and GitHub README.

What were the outcomes of your project?
Ultimately, the updates to the documentation were minimal and consisted mostly of spot fixes to things like links, and data source descriptions. There was no major restructuring of the documentation pages.

What went well? And what would you do differently?
What went well:
- Our technical writer was very communicative, checking in frequently about their work and feedback needs.
What would you do differently:
- Hire a technical writer with more domain knowledge about the energy sector. A large portion of our documentation describes the nature and origin of the data which is full of

energy jargon. Having someone who understood this better could have helped us refine how we convey this information to users.
- Communicate the importance of high-level documentation reform rather than just small fixes. Make sure the technical writer is capable of achieving the level of context and understanding required to make your desired changes in the time allotted for the project.
- Budget more time to work and review material from the technical writer.

What advice would you give to other projects trying to solve a similar problem with documentation?
- **Budget more time to engage with the technical writer:** One of our biggest hurdles was the time it took to review and provide feedback on content from the technical writer. We needed a much larger internal budget for working with them in order to provide the necessary feedback.

Project Description

Project Proposal

Our proposal consisted of the following core components:
- Audit existing documentation and walk through the existing process for answering three of our most common user questions: how to download a table with operational and financial characteristics about plant generators, the best method to access the latest release of the entire PUDL database, and how to set up the development environment and create a pull request to contribute to PUDL.
- Make a plan for what information should be conveyed in each location (PUDL GitHub repository, PUDL ReadtheDocs, Catalyst Cooperative website), as well as identify current duplication of information.
- Work with the Catalyst team to reorganize documentation according to the plan from the audit and eliminate duplicative information, pointing to other information sources where necessary (e.g., add link to the PUDL ReadtheDocs on the Catalyst Cooperative website as appropriate).
- Create a landing page identifying the ten most useful tables in the PUDL database, write descriptions that point new users to these tables based on the information they're looking for (e.g., for information on utility finances, head to this page).
- Work with Catalyst to develop guidelines for where to place new documentation to keep documentation organized and clear moving forward.

Link to proposal: https://catalyst.coop/work-with-us/pudl-season-of-docs-project-proposal/

Proposal Creation Process

Approximately three members of our team came up with the project proposal based on feedback from users and our own experience navigating the documentation. The whole team (9 individuals) was given the opportunity to review and provide comments on the intended scope of work.

Budget

How much money did you ask for?	$8,500

How did you come up with this estimate?	We asked for the maximum amount because we wanted our technical writer to be able to spend as much time as they needed on our lengthy and jargon-rich documentation.
How many hours of work did you budget for the project?	We did not budget a specific number of hours for the project.
How many hours of work were actually needed for the project?	The hours required for the task depends on the skill level and understanding of the person completing them.
What other expenses did you include in your budget?	NA
Did you run into any budget surprises during the project (e.g. misestimates)? If so, please explain.	None.

Tech Writer Recruitment

We put out an application for a technical writer and received submissions from people all over the world. We asked for writing samples and interviewed the candidates to see if they were a good fit. Overall, the application process went well. The writer we chose had interacted with our codebase in a meaningful way that seemed to demonstrate their ability to jump into messy code projects and make sense of them.

They stayed for the duration of the project, however, we decided to cut the project short (i.e., intentionally not finish all of our scoped line items) due to the amount of time it was taking to manage the technical writer and internal priority shifts with regard to certain documentation platforms.

Other Participants

N/A

Timeline

The project started at the beginning of the summer. Our first deadlines for reviewing the documentation were in mid June followed by a full documentation update proposal due at the end of July. Our goal was to spend August through November working on the elements of the proposal as well as the desired documentation changes stated explicitly in the scope of work.

Deliverables

Planned Deliverables

Deliverable	% Complete	Notes
Make a plan for what information should be conveyed in each location.	100	Our technical writer successfully made a plan for updating

		information in each location.
Reorganize documentation according to the plan from Phase 1	70	Some of the things we wanted to change (like Datasette) ended up being deprioritized by Catalyst due to a desire to use other tools. We also didn't have time internally to review all the pull requests for small changes to the documentation.
Reorganize documentation to have better flow through the docs for the three user cases	0	This is a difficult task to begin with because it requires a deep knowledge of the codebase and use cases that is hard to acquire in a few months of acclimation. Our technical writer was more focused on small, incremental fixes rather than big picture restructuring.
Make the data dictionary page, example notebooks, and Datasette page more clear.	0	Did not get to this for reasons similar to those stated above.
Rework table/column documentation.	0	Did not get to this for reasons similar to those stated above.
Rework data source documentation.	25	Make incremental progress on this for reasons similar to those stated above.
Create a landing page for the ten most useful tables in the PUDL database.	0	This was deprioritized by Catalyst due to our desire to have more user information before doing this.
Improve the docstrings for these tables.	0	See above.
Write a short narrative of the work done throughout the project.	100	

Unplanned Deliverables

N/A

Metrics

Intended metrics:
- Catalyst has a document with clear guidelines for what information should live in the PUDL ReadTheDocs, Github, and website.

- No more than 25% of our PUDL Office Hours visits and GitHub discussion questions are from new users who cannot find one of the ten most useful tables
- The number of total PUDL users increases by 20%
- The number of pull requests from community contributors increases by 10%

It's too soon to measure any of the metrics, but because the documentation didn't undergo any substantial changes there may not be any observable results. We plan to continue monitoring these metrics moving forward regardless and are investing in more infrastructure to collect user information and feedback.

Analysis

PUDL is a niche tool for a niche audience and this made it hard for a technical writer to jump in and understand what was going on.

Unfortunately the vast majority of our intended projects remain uncompleted; there are a few reasons for this. First, Catalyst had insufficient funds to oversee the technical writer and provide the level of feedback necessary to complete projects to our satisfaction. It was difficult for us to balance granular feedback needs with a desire to move the project forward at a high level. We were hoping for a holistic restructuring of the documentation and got bogged down by small edits to wording, links, and description.

Second, we should have found a technical writer with more domain knowledge of the energy sector. After conveying our desire for high-level structural changes to the documentation, it was difficult for our technical writer to make tractable suggestions due to their limited understanding of how one might interact with the data. Ultimately we didn't feel like they were prepared to take on this level of documentation change without deeper knowledge of the energy sector or the codebase.

Third, Catalyst deprioritized several projects due to internal priority changes over the course of the summer and fall. We decided to pursue a different data cataloging tool and therefore stopped work on all the Datasette updates. We also decided not to build out a dashboard for ten important tables until we had more information about which of our current tables users were actually interacting with most. We didn't have definitive decisions about our goals for these tools until about halfway through the Season of the Docs timeline which was not ideal.

Lessons Learned

What went well?

Topic	What we did	Lesson Learned *(+1 for relevance to existing takeaways:* [*https://github.com/google/season-of-docs/blob/main/resources/best-practices.md*](https://github.com/google/season-of-docs/blob/main/resources/best-practices.md)*)*
Budget	We correctly anticipated that it would take	Ensure adequate time for documentation

	the technical writer a while to familiarize themselves with the code and dedicated almost half of their time in the beginning of the project to running through the documentation under the guise of different user profiles.	exploration.
Communication	Bi-weekly check-ins, slack channel, shared google documents, GitHub pull requests.	+1
Mentorship	Walk through the codebase with the technical writer and have them get to know several team members of whom they can ask questions.	+1
Metrics	Contemplate what values we can actually measure and how they relate to our goals with the documentation.	It may take a long time to see if the changes make a difference, and at that point it could be due to unrelated changes to the codebase or documentation
Onboarding	Have them look at our good first issues, familiarize themselves with our different user profiles and attempt to navigate the data pretending to be each one.	+1
Participants	A subset of the team is involved in coaching and directing the technical writer. Intended to have beta testers for the new documentation, but it didn't change enough to warrant this.	+1
Project Deliverables	A spreadsheet containing a description of a task and the desired deadline for that task.	+1
Project Management	Two members of Catalyst were tasked with overseeing the project and checking in bi-weekly with the technical writer	+1
Recruiting & Hiring	Request writing samples and have multiple team members involved in the hiring process.	+1

What could be improved?

Topic	What we did	What we would do differently	Lesson Learned (+1 for relevance to existing takeaways: https://github.com/google/season-of-docs/blob/main/resources/best-practices.md)

Budget	Rely solely on the volunteer stipend for internal review of the documentation updates.	Budget more time to work with the technical writer.	Working with the technical writer takes more time than we thought!
Communication	Bi-weekly video calls and written updates.	Use a spreadsheet to track progress more rigorously.	There was a slight language barrier so we ultimately relied on written communication more heavily than verbal communication.
Mentorship	Same as above	We didn't have time to provide a high-level of mentorship, so we should have hired someone that was more familiar with the codebase and the energy sector.	Consider how much time your organization has to dedicate to working with the technical writer before you decide what you want them to do.
Metrics	Our metrics were based on a high-degree of change to the current documentation.	We could have developed certain metrics that were smaller or more subjective. For example, having someone already familiar with the documentation read through them and see if the changes were an improvement.	Develop metrics can apply to large-scale and small-scale changes to the documentation.
Onboarding	-	-	-
Participants	The only participants were our technical writer and internal Catalyst overseers.	It would have been nice to involve a user in the discussion of how to restructure the documentation.	+1
Project Deliverables	Define a list of clear tasks in a spreadsheet.	Write up a high-level description of our intended changes in addition to a granular depiction of tasks.	Clarify high-level deliverables as well as granular deliverables. If your project line items are too specific it might overshadow the big picture changes you are looking for and lead your technical writer down a path of small tweaks as opposed to systemic changes.
Project Management	1-2 members of Catalyst met with the technical writer bi-weekly. The project was in its own silo and not a part of any other projects Catalyst was working on at the time.	Note key uncertainties or potential priority shifts as soon as they become apparent.	Figure out how to successfully de-scope projects that are no longer relevant or within the capabilities of your technical writer's time or ability.

Recruiting & Hiring	Members of our business development team worked on hiring the technical writer.	Could have decided earlier in the process who would work with the technical writer and have them more involved in the hiring process.	Involve technical writer overseer/mentor in the hiring process.