# Wolfram: The AI Revolution and Implications to Project Catalyst

Milestone 2: Content Collection and Curation

Github Link: https://github.com/WolframBlockchainLabs/Wolfram-Catalyst-Navigator-Resources

## Introduction

### Background: Researching LLM with Guided Participation

Our research for this milestone centers around a community-driven question; “what resources are needed to assist a Catalyst community member with the ability to enable an AI chatbot workflow?” How can Wolfram Blockchain Labs (WBL) provide as much and as close to a step-by-step guidance through the Catalyst governance process to understand how LLMs work? Various additional questions include; what data is useful, where can I find it, how much will it cost to implement, what structure should the data take, etc. Critically, this is a reminder that the process WBL is taking is **not** the creation of an AI chatbot or AI Assistant for Catalyst. That would be an enormous undertaking at this point in time. Rather, this project is to conduct significant research and exploration to see if the useful creation of an assistant is possible—by investigating novel approaches, proofs of concept, and use cases—and if so, how to keep the operational costs low enough to provide a significant return on investment for the Catalyst community to endeavor forward. Because this project is fully open source, we believe leading this research effort would establish a strong precedent for community actors to get up to speed and build from there. The objective of this research is twofold: 1) to explore the feasibility of providing users with support throughout the stages of proposal submission, evaluation, voting and management, and 2) to understand the preferred implementation approach according to the community's needs, ethos and perspectives. This includes content curation, resource gathering, and workflow processes documented. This particular report includes that information. Lastly, and again, it's important to note that this proposal does not encompass the actual implementation of the LLM assistant, but a guide and workflow for it.

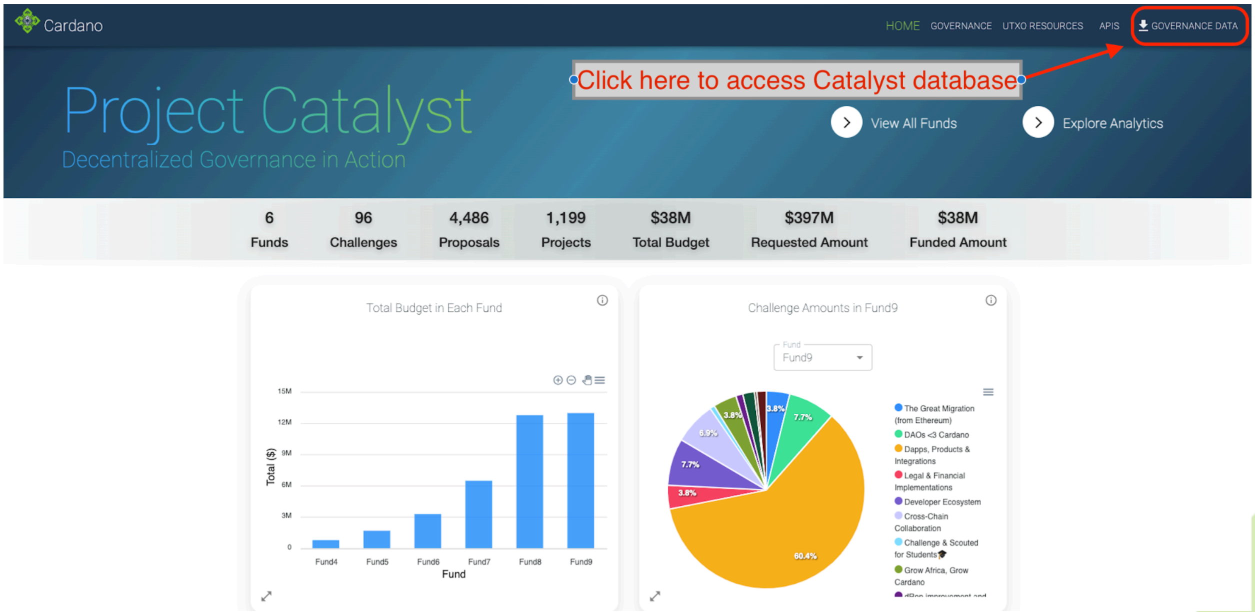
### Milestone 2: Content Collection & Curation

There have now been 11 Project Catalyst funds, and WBL has curated collections of past proposals in as much detail as possible. We have collaborated with other teams in the space, including LidoNation and MinSwap, in order to create a more robust dataset of the full catalyst proposals and accompanying data. However, for this specific use case, we attempted to curate all historical proposals and build them into a start to finish workflow. That workflow is included here, but there are still gaps in the full curation of content (as there may always be gaps). Because this is an evolving ecosystem, at each point in time there will be new rules and new additions to the Catalyst Governance process which change or supersede previous iterations, templates, standards, and datasets. So far, there has not been a definitive standard to the Catalyst proposal process, which makes combining and merging previous datasets into one coherent whole very difficult. This is an ongoing process, and it forms the basis for a more formal request from the Catalyst team to receive full, clean and structured proposal data from the Catalyst team.

Additionally, we must gather as much instructional content related to participating in different roles within Catalyst, like proposer, voter, reviewer, and management. It is likely that we will need to develop a significant amount of new material on how to engage in Catalyst, as a DRep or as an AI agent, for which we've hosted a meeting with several Catalyst leaders in the space (video here). Initial assessments reveal that we can't assume instructional data in training sets of LLMs from commercial entities has resulted in validated or trustworthy data regarding Project Catalyst. Therefore, our workflow uses LLM functionalities built into the Wolfram Language in the loop of exploring SQL databases to ultimately return verified, factual information based on the data that has been collected within Project Catalyst, historically speaking. Ultimately, curating data/content and developing a workflow with context and understanding of the various aspects of the Catalyst landscape is pivotal in LLM-based assistance research. Please see the following instructions for more information on the available workflow.

# Public link to data collection

1. Navigate: https://www.wolframblockchainlabs.com/dashboard/
2. Click here in the top right corner:

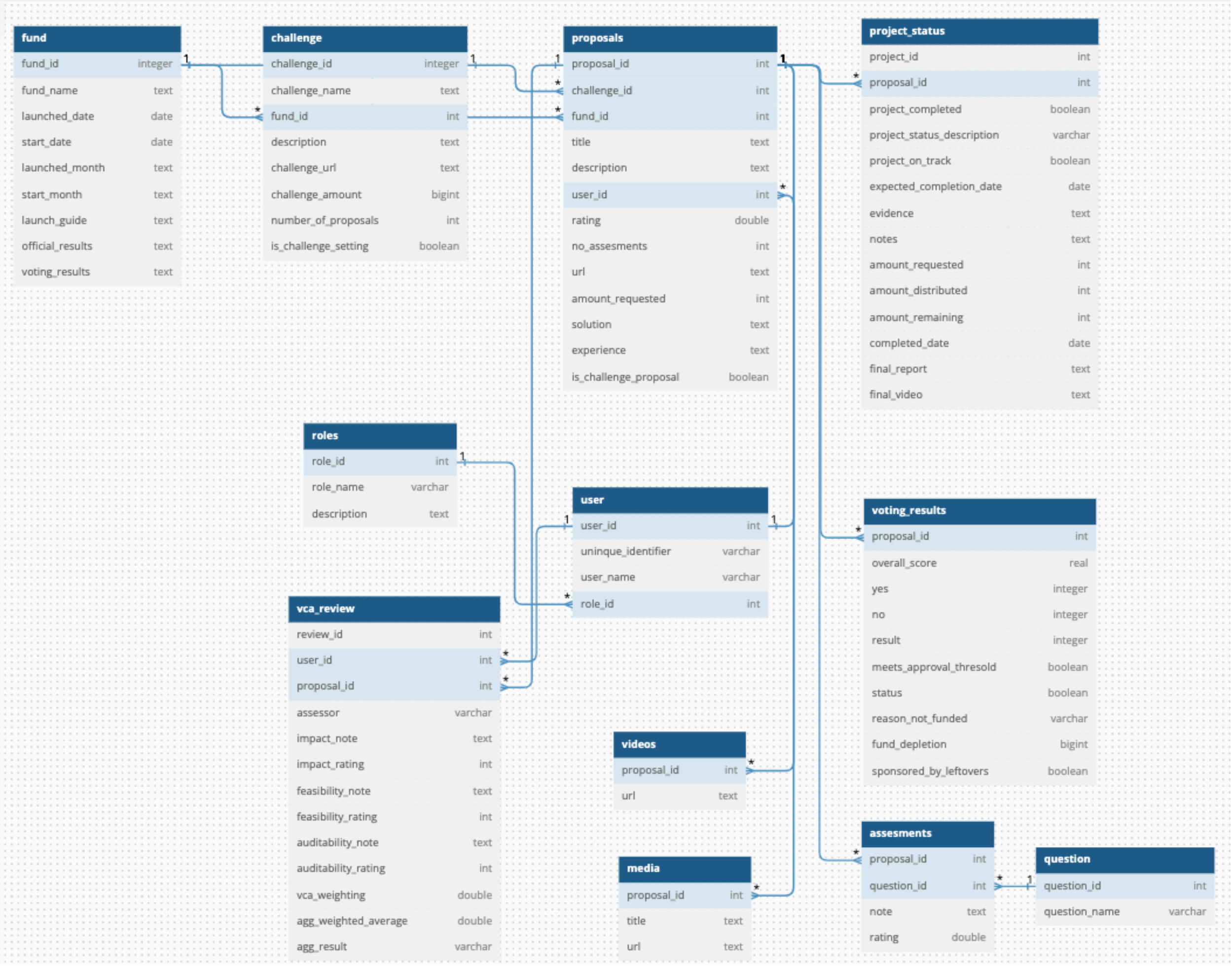

Cardano
HOME
GOVERNANCE
UTXO RESOURCES
APIS
GOVERNANCE DATA
Click here to access Catalyst database
Project Catalyst
Decentralized Governance in Action
View All Funds
Explore Analytics
Total Budget in Each Fund
15M
12M
9M
6M
3M
0
Total ($)
Fund4
Fund5
Fund6
Fund7
Fund8
Fund9
Fund
Challenge Amounts in Fund9
Fund
Fund9
3.8%
7.7%
60.4%
3.8%
7.7%
6.9%
3.8%
The Great Migration (from Ethereum)
DAOs <3 Cardano
Dapps, Products & Integrations
Legal & Financial Implementations
Developer Ecosystem
Cross-Chain Collaboration
Challenge & Scouted for Students🎓
Grow Africa, Grow Cardano

| 6 | 96 | 4,486 | 1,199 | $38M | $397M | $38M |
|---|---|---|---|---|---|---|
| Funds | Challenges | Proposals | Projects | Total Budget | Requested Amount | Funded Amount |

3. Download and unzip file to access the “Cadano SQL Dump” data

# The Database

1. Topologically, here is the database diagram: https://dbdiagram.io/d/Cardano-Governance-Database-655ddcf53be14957878114c7
2. Screenshot of database diagram:



| fund | |
|---|---|
| fund_id | integer |
| fund_name | text |
| launched_date | date |
| start_date | date |
| launched_month | text |
| start_month | text |
| launch_guide | text |
| official_results | text |
| voting_results | text |

| challenge | |
|---|---|
| challenge_id | integer |
| challenge_name | text |
| fund_id | int |
| description | text |
| challenge_url | text |
| challenge_amount | bigint |
| number_of_proposals | int |
| is_challenge_setting | boolean |

| proposals | |
|---|---|
| proposal_id | int |
| challenge_id | int |
| fund_id | int |
| title | text |
| description | text |
| user_id | int |
| rating | double |
| no_assesments | int |
| url | text |
| amount_requested | int |
| solution | text |
| experience | text |
| is_challenge_proposal | boolean |

| project_status | |
|---|---|
| project_id | int |
| proposal_id | int |
| project_completed | boolean |
| project_status_description | varchar |
| project_on_track | boolean |
| expected_completion_date | date |
| evidence | text |
| notes | text |
| amount_requested | int |
| amount_distributed | int |
| amount_remaining | int |
| completed_date | date |
| final_report | text |
| final_video | text |

| roles | |
|---|---|
| role_id | int |
| role_name | varchar |
| description | text |

| user | |
|---|---|
| user_id | int |
| uninque_identifier | varchar |
| user_name | varchar |
| role_id | int |

| voting_results | |
|---|---|
| proposal_id | int |
| overall_score | real |
| yes | integer |
| no | integer |
| result | integer |
| meets_approval_thresold | boolean |
| status | boolean |
| reason_not_funded | varchar |
| fund_depletion | bigint |
| sponsored_by_leftovers | boolean |

| vca_review | |
|---|---|
| review_id | int |
| user_id | int |
| proposal_id | int |
| assessor | varchar |
| impact_note | text |
| impact_rating | int |
| feasibility_note | text |
| feasibility_rating | int |
| auditability_note | text |
| auditability_rating | int |
| vca_weighting | double |
| agg_weighted_average | double |
| agg_result | varchar |

| videos | |
|---|---|
| proposal_id | int |
| url | text |

| media | |
|---|---|
| proposal_id | int |
| title | text |
| url | text |

| assesments | |
|---|---|
| proposal_id | int |
| question_id | int |
| note | text |
| rating | double |

| question | |
|---|---|
| question_id | int |
| question_name | varchar |

3. Here are the descriptions of the database diagram:

# Explanation of Database diagram

## Wolfram Governance Data - Tables and Fields

### 1. Fund

This table stores information about different funds launched by Project Catalyst.

- fund_id (integer): Unique identifier for the fund.
- fund_name (text): Name of the fund.
- launched_date (date): Date when the fund was launched.
- start_date (date): Date when the fund started accepting proposals.

- launched_month (text): Month when the fund was launched.
- start_month (text): Month when the fund started accepting proposals.
- launch_guide (text): URL for the launch guide of the fund.
- official_results (text): URL for the official results.
- voting_results (text): Stores voting results information.

### 2. Challenge

This table categorizes proposals within a fund.

- challenge_id (integer): Unique identifier for the challenge.
- challenge_name (text): Name of the challenge.
- fund_id (int): References the fund.
- description (text): Description of what proposals should address.
- challenge_url (text): Ideascale URL of the challenge.
- challenge_amount (bigint): Budget allocated for the challenge.
- number_of_proposals (int): Number of proposals submitted.
- is_challenge_setting (boolean): Indicates if this is a setting challenge.

### 3. Proposals

Contains proposals submitted for funding in each challenge.

- proposal_id (int): Unique identifier for the proposal.
- challenge_id, fund_id (int): References to challenge and fund.
- title (text): Title of the proposal.
- description (text): Detailed proposal description.
- user_id (int): References the proposer.
- rating (double): Overall proposal rating.
- no_assessments (int): Number of assessments received.
- url (text): Ideascale URL of the proposal.
- amount_requested (int): Funding amount requested.
- solution, experience (text): Provided solution and proposer's experience.
- is_challenge_proposal (boolean): Indicates if part of a challenge setting.

### 4. Voting Results

Tracks voting and status of proposals.

- proposal_id (int): References the proposal.
- overall_score (real): Overall score from reviewers.
- yes, no (integer): Upvotes and downvotes.
- result (integer): Overall votes received.
- meets_approval_threshold, status (boolean): Approval status and funding status.
- reason_not_funded (varchar): Reason if not funded.
- fund_depletion (bigint): Remaining funds after decision.
- sponsored_by_leftovers (boolean): If funded by leftover funds.

### 5. Project Status

For tracking funded projects.

- project_id (int): Unique project identifier.
- proposal_id (int): References the funded proposal.
- project_completed, project_on_track (boolean): Completion and tracking status.
- project_status_description (varchar): Status description.
- expected_completion_date, completed_date (date): Expected and actual completion dates.
- evidence, notes, final_report, final_video (text): Documentation of project progress and completion.
- amount_requested, amount_distributed, amount_remaining (int): Financial tracking.

### 6. User

Details on users involved in the governance process.

- user_id (int): Unique user identifier.
- unique_identifier (varchar): Unique identifier for user.
- user_name (varchar): Username.
- role_id (int): References the user's role.

### 7. Question

Questions for proposal assessment.

- question_id (int): Unique question identifier.
- question_name (varchar): Assessment question.

### 8. Assessments

Reviewers' assessments on proposals.

- proposal_id, question_id (int): References to proposal and question.
- note (text): Reviewer's note.
- rating (double): Given rating.

### 9. Videos

Videos related to proposals.

- proposal_id (int): References the proposal.
- url (text): Video URL.

### 10. Media

Media items related to proposals.

- proposal_id (int): References the proposal.
- title (text): Media title.
- url (text): Media URL.

### 11. Roles

Roles within the governance framework.

- role_id (int): Unique role identifier.
- role_name (varchar): Name of the role.
- description (text): Role description.

### 12. VCA Review

(Noted as unused in the database)

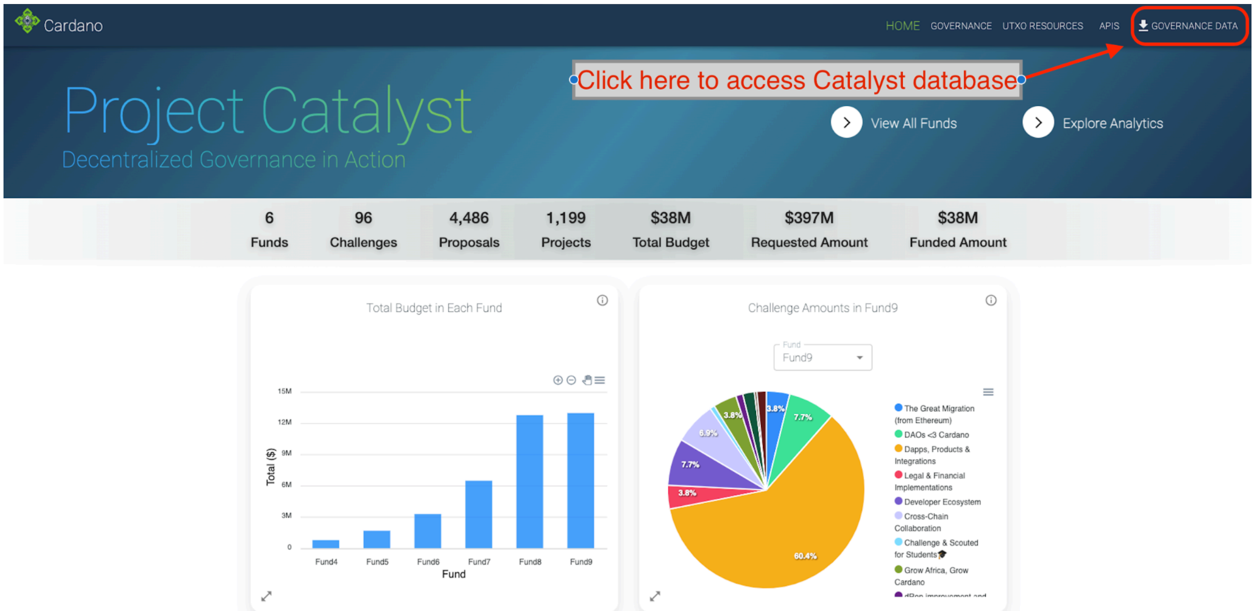
- review_id, user_id, proposal_id (int): Identifiers for the VCA review, user, and proposal.
- assessor, impact_note, feasibility_note, auditability_note (varchar, text): Reviewer details and notes.
- impact_rating, feasibility_rating, auditability_rating (int): Ratings.
- vca_weighting, agg_weighted_average (double): Weighting and averages.
- agg_result (varchar): Aggregate result.

## Table Links

Provides how tables are interconnected through various foreign keys, ensuring relational integrity across the schema. Each "Ref" line indicates a reference from a foreign key in one table to a primary key in another, establishing relationships like fund and challenge linkage, proposal to challenge and fund, user to proposal, and so forth.

## Catalyst Proposals Data Table

1. Navigate: https://www.wolframblockchainlabs.com/dashboard/
2. Click here in the top right corner:



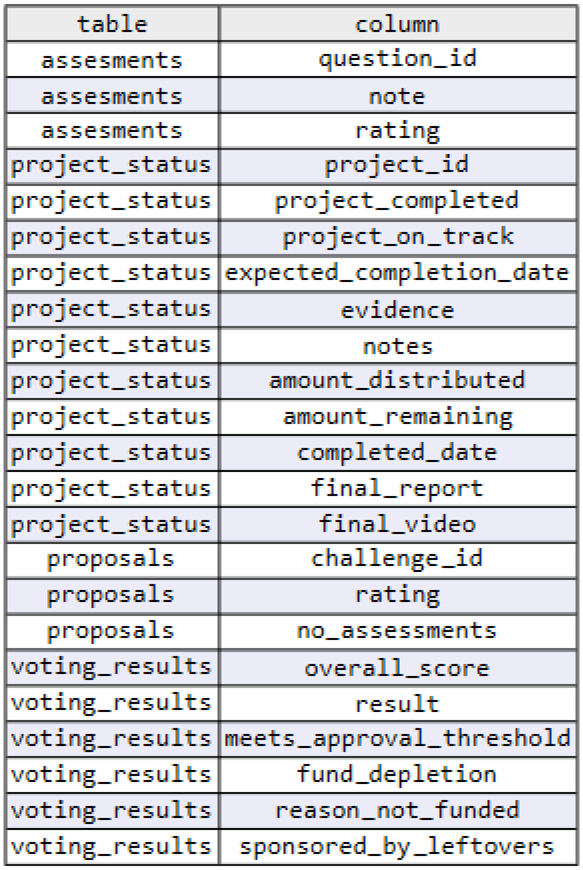
Click here to access Catalyst database

Cardano
HOME
GOVERNANCE
UTXO RESOURCES
APIS
GOVERNANCE DATA
Project Catalyst
Decentralized Governance in Action
View All Funds
Explore Analytics
Total Budget in Each Fund
15M
12M
9M
6M
3M
0
Total ($)
Fund4
Fund5
Fund6
Fund7
Fund8
Fund9
Fund
Challenge Amounts in Fund9
Fund
Fund9
3.8%
7.7%
3.8%
6.9%
7.7%
3.8%
60.4%
The Great Migration (from Ethereum)
DAOs <3 Cardano
Dapps, Products & Integrations
Legal & Financial Implementations
Developer Ecosystem
Cross-Chain Collaboration
Challenge & Scouted for Students
Grow Africa, Grow Cardano

| Funds | Challenges | Proposals | Projects | Total Budget | Requested Amount | Funded Amount |
|---|---|---|---|---|---|---|
| 6 | 96 | 4,486 | 1,199 | $38M | $397M | $38M |

3. Download and unzip file named “proposals_metadata_merged_with_full_proposals” to access the SQL data

This table contains details about each proposal submitted for funding. Importantly, this dataset was provided by LidoNation. However, assuming during the scraping phase of the gathering full proposal content from Ideascale, there were errors, bugs and limitations in the scraping mechanisms. This led to a significant number of gaps that WBL was unable to merge properly. This is an ongoing “final mile” aspect of a cleaning and structuring a secondary database of full Catalyst proposals. Some of the identified gaps between the Wolfram Governance Database and this one are in the image on the following page:



| table | column |
|---|---|
| assesments | question_id |
| assesments | note |
| assesments | rating |
| project_status | project_id |
| project_status | project_completed |
| project_status | project_on_track |
| project_status | expected_completion_date |
| project_status | evidence |
| project_status | notes |
| project_status | amount_distributed |
| project_status | amount_remaining |
| project_status | completed_date |
| project_status | final_report |
| project_status | final_video |
| proposals | challenge_id |
| proposals | rating |
| proposals | no_assessments |
| voting_results | overall_score |
| voting_results | result |
| voting_results | meets_approval_threshold |
| voting_results | fund_depletion |
| voting_results | reason_not_funded |
| voting_results | sponsored_by_leftovers |

Below are labels of the data fields and types in the Catalyst Proposal Database:

- proposal_id (int): Unique identifier for the proposal. Maps to "Id".
- user_id (int): References the user who submitted the proposal. Maps to "UserId".
- fund_id (int): References the fund to which the proposal is submitted. Maps to "FundId".
- title (text): Title of the proposal. Maps to "Title".
- url (text): Ideascale URL of the proposal. Maps to "IdeascaleLink".
- amount_requested (int): Funding amount requested. Maps to "AmountRequested".
- description (text): Includes combined information from "Problem" and "Solution".
- experience (text): Proposer's relevant experience. Maps to "Experience".
- content (text): Detailed content of the proposal. Maps to "Content".
- project_status (text): The current status of the project. Maps to "ProjectStatus".
- funding_status (text): Indicates whether the funding was approved or not. Maps to "FundingStatus".

## Voting Results Table

Tracks voting and the status of proposals.

- proposal_id (int): References the proposal. Maps directly from the Proposals table.
- yes_votes (integer): Upvotes the proposal received. Maps to "YesVotesCount".

- no_votes (integer): Downvotes the proposal received. Maps to "NoVotesCount".
- status (boolean): Derived from "FundingStatus"; could be true for funded/approved, false otherwise.

## User Table

Details on users involved in the proposals.

- user_id (int): Unique user identifier. Maps to "UserId".
- (additional fields would be needed to fully describe the user, which are not present in the CSV data)

The provided data mainly focuses on the proposals, their funding, and their voting outcomes. This restructured schema offers an approach to organizing the available information into a fuller proposal specific database, focusing on the content of proposals which still have significant leftover technical debt from the scraping process. This informs us that we should request cleaned and structured data directly from the Project Catalyst team.