

Data Foundation Systems

Mohsin Mamoon, 2018101029

Zia Bashir, 2018101006

Hasir Mushtaq, 2018102049

Publishing Datasets and Version

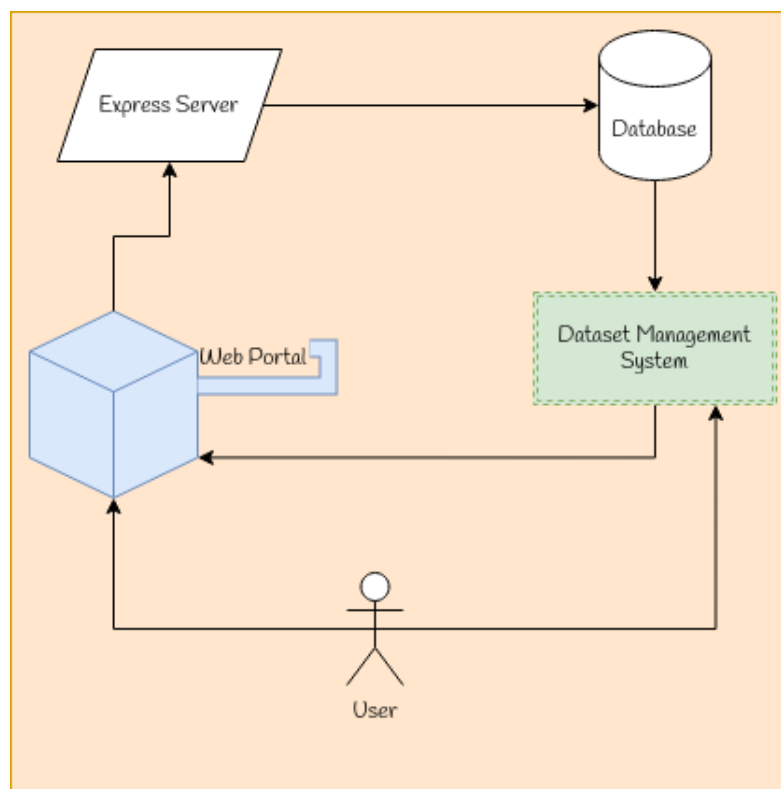
Design Document

DESIGN OVERVIEW

The project involves building a web application that will provide an interface for publishing datasets and updating their versions. Additional functionality includes the update and deletion of uploaded datasets. Further, the web application also allows consumers to raise comments or flags which the owners can review and edit as desired.

ARCHITECTURAL DESIGN

In this project, we are building a web portal that will allow dataset owners to upload their datasets onto the Dataset Management System. The portal will also provide a version control system for the datasets. In addition to this, users of the datasets will also be able to comment or raise flags for the dataset.



Subsystems:

1. Express Server: Handles the interaction between the database and the web portal
2. Web portal: Provides the ability to upload, review and manage different versions of the dataset to the users.
3. Database: Stores the datasets upload to the system
4. Dataset Management System: The dataset system that provides various functionalities regarding the datasets.

Relationships:

1. Web portal - Express Server: The express server handles the POST and GET requests from the web portal and consequently sends or receives relevant data from/to the database
2. Express Server - Database: All the interactions with the database are handled by the express server.
3. Dataset Management System - Web portal: Dataset Management System redirects to the web portal at any point where upload, editing, or commenting on any dataset is required.
4. Dataset Management System - Database: The dataset management system provides different functionalities for the datasets present in the database like dataset download, analysis, etc

High-level Approach

The user logs on to the portal and is presented with the option to either upload a new dataset or review an existing dataset. For uploading a new dataset, the user inputs the dataset metadata, which includes the dataset name, description, version, and selects a file to upload. On pressing the upload button, the dataset will be uploaded onto the system and a unique identifier will be generated for the dataset. To review a dataset, the user uses the unique identifier and version of the dataset to identify the dataset. After identification, the user can edit, delete existing versions of the dataset or upload a new version. A user can also add comments or raise flags at this screen.

Database Models

To be determined after meeting with the Data verification team and after deciding different dataset types.

Main Flows

1. Upload a new dataset that does not exist in the system:

-
- a. Owner logs on to portal
 - b. Owner selects Upload Button
 - c. Owner fills metadata of the dataset
 - d. Owner selects the dataset file
 - e. Owner presses the upload button
 2. Upload a new version of a dataset for which one or more versions are already present in the system:
 - a. Owner logs in to the portal
 - b. Owner selects Review button
 - c. Owner selects the particular dataset using the identifier.
 - d. Owner gives a new version number.
 - e. Owner presses the upload button.
 3. Replace a specific version of a dataset for which one or more versions are already present in the system:
 - a. Owner logs on to the portal
 - b. Owner selects the review button
 - c. Owner Inputs the unique identifier to retrieve details
 - d. Owner inputs the version number that needs to be updated
 - e. Owner makes any required changes to metadata
 - f. Owner clicks the upload button
 4. Delete a specific version of a dataset for which one or more versions are already present in the system:
 - a. Owner logs in to the portal
 - b. Owner selects the review button.
 - c. Owner finds the particular dataset using the identifier.
 - d. Owner deletes the specified version of the dataset by pressing the delete button.
 5. Raise flags or comments regarding the datasets:
 - a. Consumer logs on to the portal
 - b. Consumer selects the review button
 - c. Consumer inputs the unique identifier and version
 - d. Consumer adds desired comments/flags
 - e. Consumer click update button
 6. Remove flags or comments regarding the datasets that have been addressed:
 - a. Owner logs on to the portal
 - b. Owner selects the review button
 - c. Owner inputs the unique identifier and version
 - d. Owner removes desired comments/flags
 - e. Owner click update button