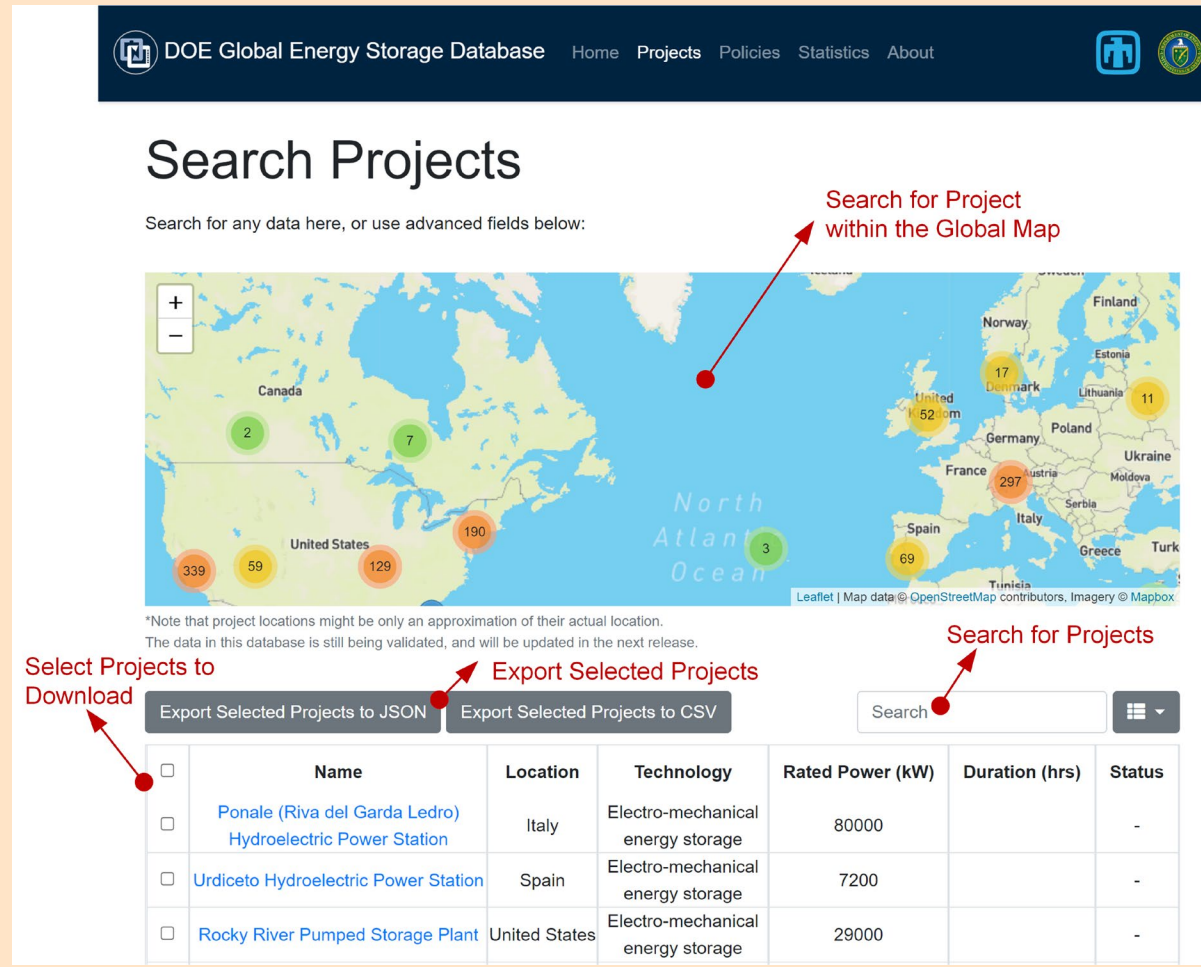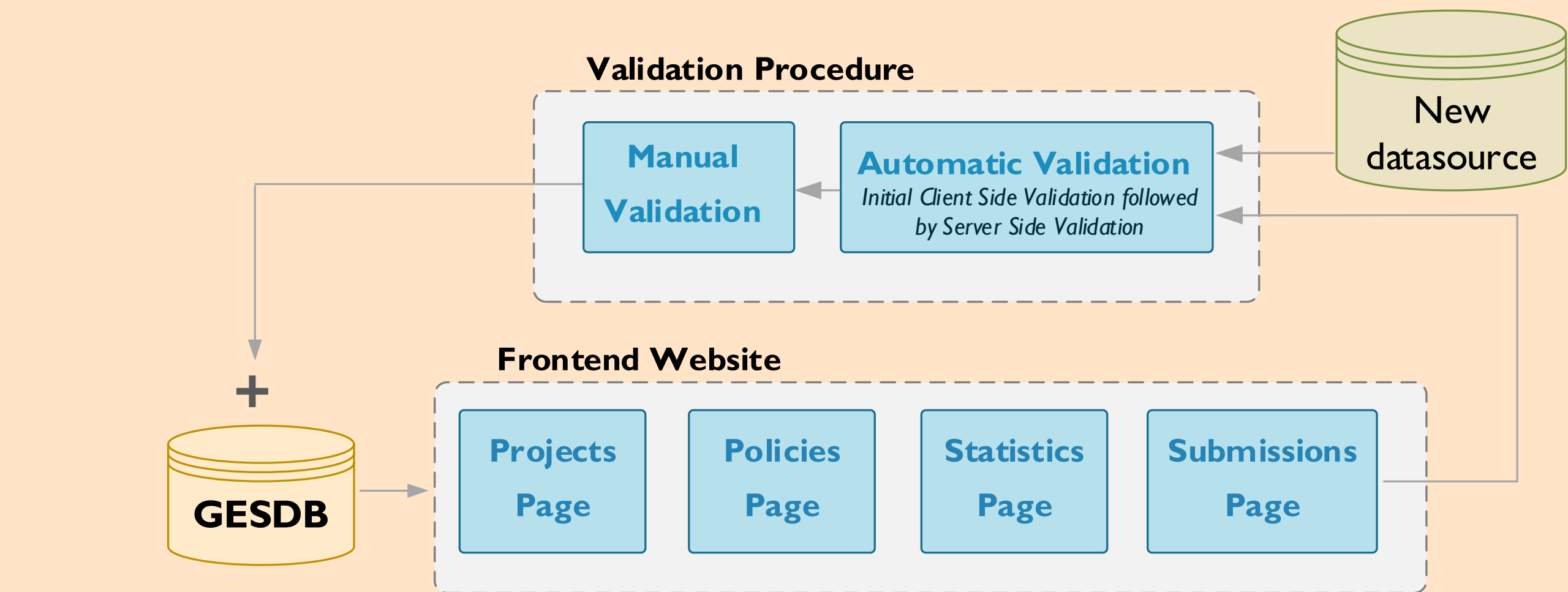*Sandia National Laboratories*

GLOBAL ENERGY STORAGE DATABASE (GESDB): DATABASE UPDATES AND NEW FEATURES

Project Team: Ujjwol Tamrakar, Andres S. Lopez Ramirez, Will McNamara, Tu A. Nguyen
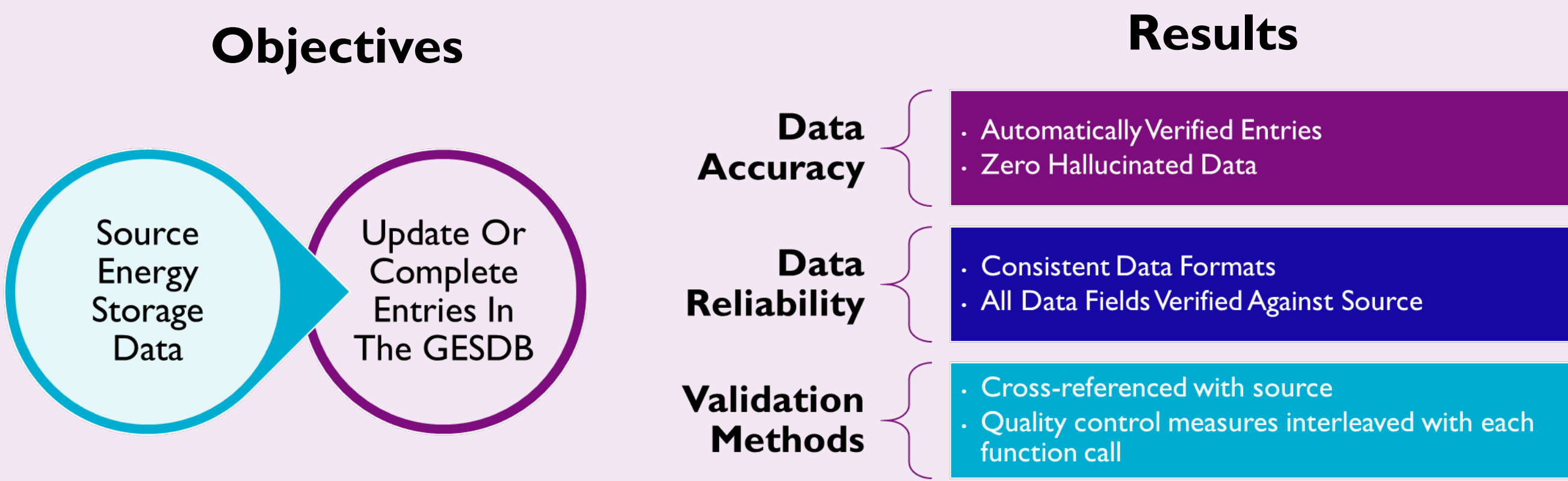Contact: utamrak@sandia.gov, tunguy@sandia.gov

Abstract

The Global Energy Storage Database (GESDB) aims at providing high-quality and accurate data on energy storage projects around the globe. In this poster, we present an overview of all the features of the GESDB including recent updates to the database. The poster also highlights some backend features being developed to automate data entries into the GESDB.

Overview of GESDB
Website: https://gesdb.sandia.gov

Validation Procedure
- Manual Validation
- Automatic Validation — *Initial Client Side Validation followed by Server Side Validation*
- New datasource

Frontend Website
- Projects Page
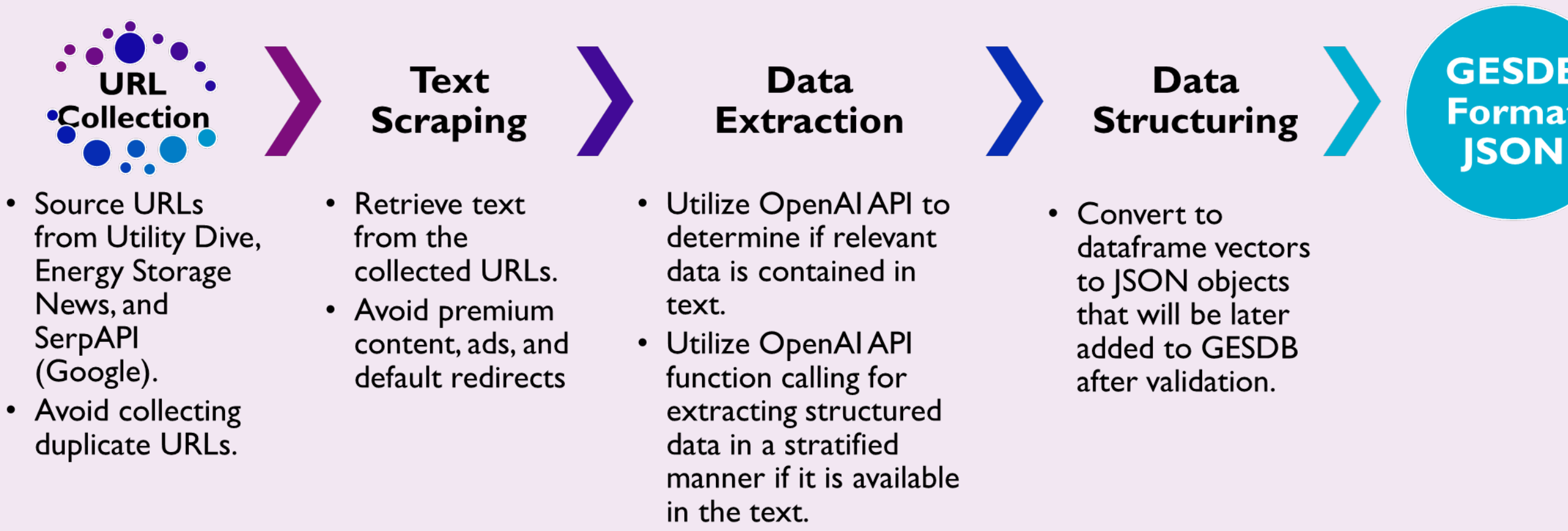- Policies Page
- Statistics Page
- Submissions Page

GESDB

- **Project Page-** Users can navigate, search, and download a particular project or a set of projects
  - 2300 projects → >90 GW of energy storage projects worldwide
- **Policies Page-** Relevant federal and state policy documents
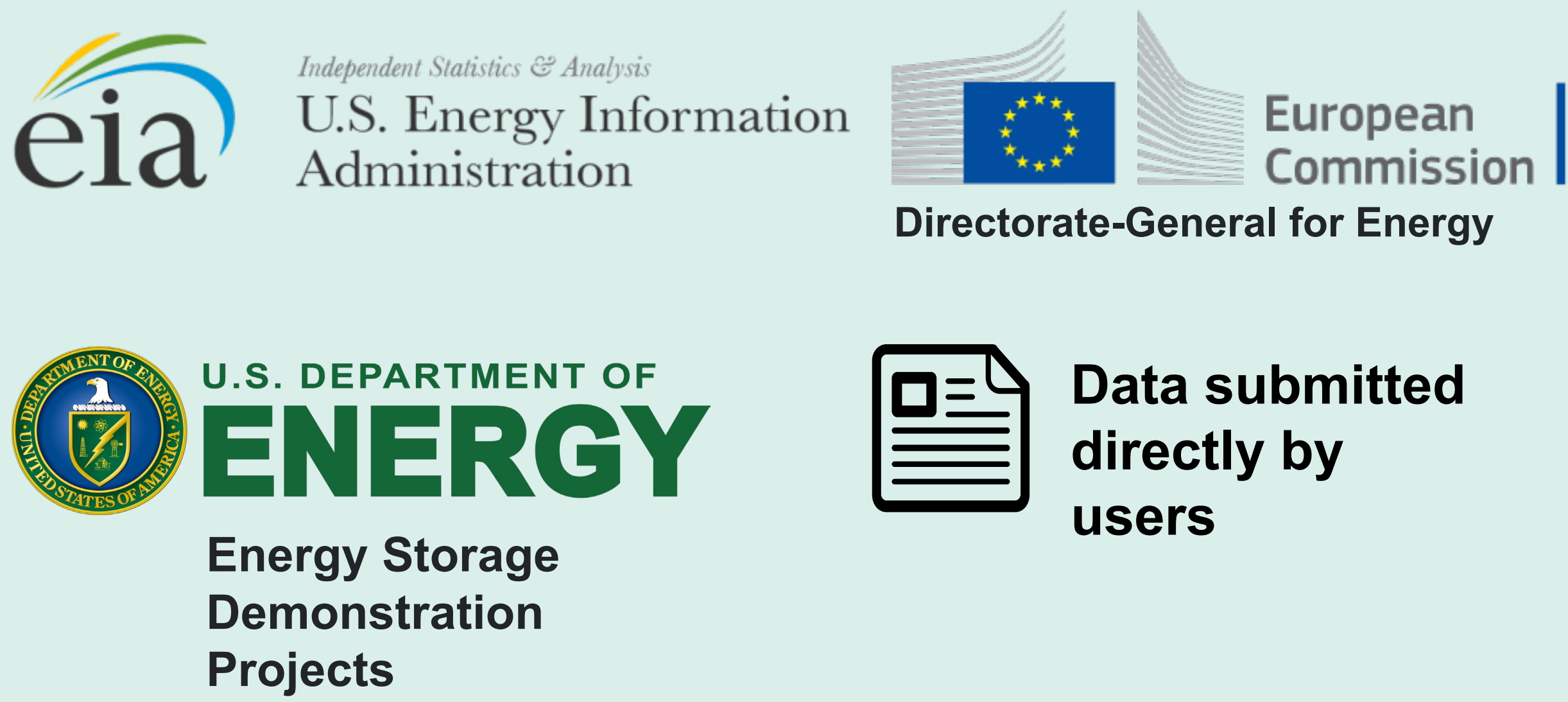- **Statistics Page-** Interactive statistics and visualizations

Web Scraping Framework

Objectives
Source Energy Storage Data → Update Or Complete Entries In The GESDB

Results
Data Accuracy	• Automatically Verified Entries • Zero Hallucinated Data
Data Reliability	• Consistent Data Formats • All Data Fields Verified Against Source
Validation Methods	• Cross-referenced with source • Quality control measures interleaved with each function call

Web Scraper Pipeline

URL Collection
- Source URLs from Utility Dive, Energy Storage News, and SerpAPI (Google).
- Avoid collecting duplicate URLs.

Text Scraping
- Retrieve text from the collected URLs.
- Avoid premium content, ads, and default redirects.

Data Extraction
- Utilize OpenAI API to determine if relevant data is contained in text.
- Utilize OpenAI API function calling for extracting structured data in a stratified manner if it is available in the text.

Data Structuring
- Convert to dataframe vectors to JSON objects that will be later added to GESDB after validation.

GESDB Format JSON

Current Data Sources

eia — Independent Statistics & Analysis — U.S. Energy Information Administration

European Commission — Directorate-General for Energy

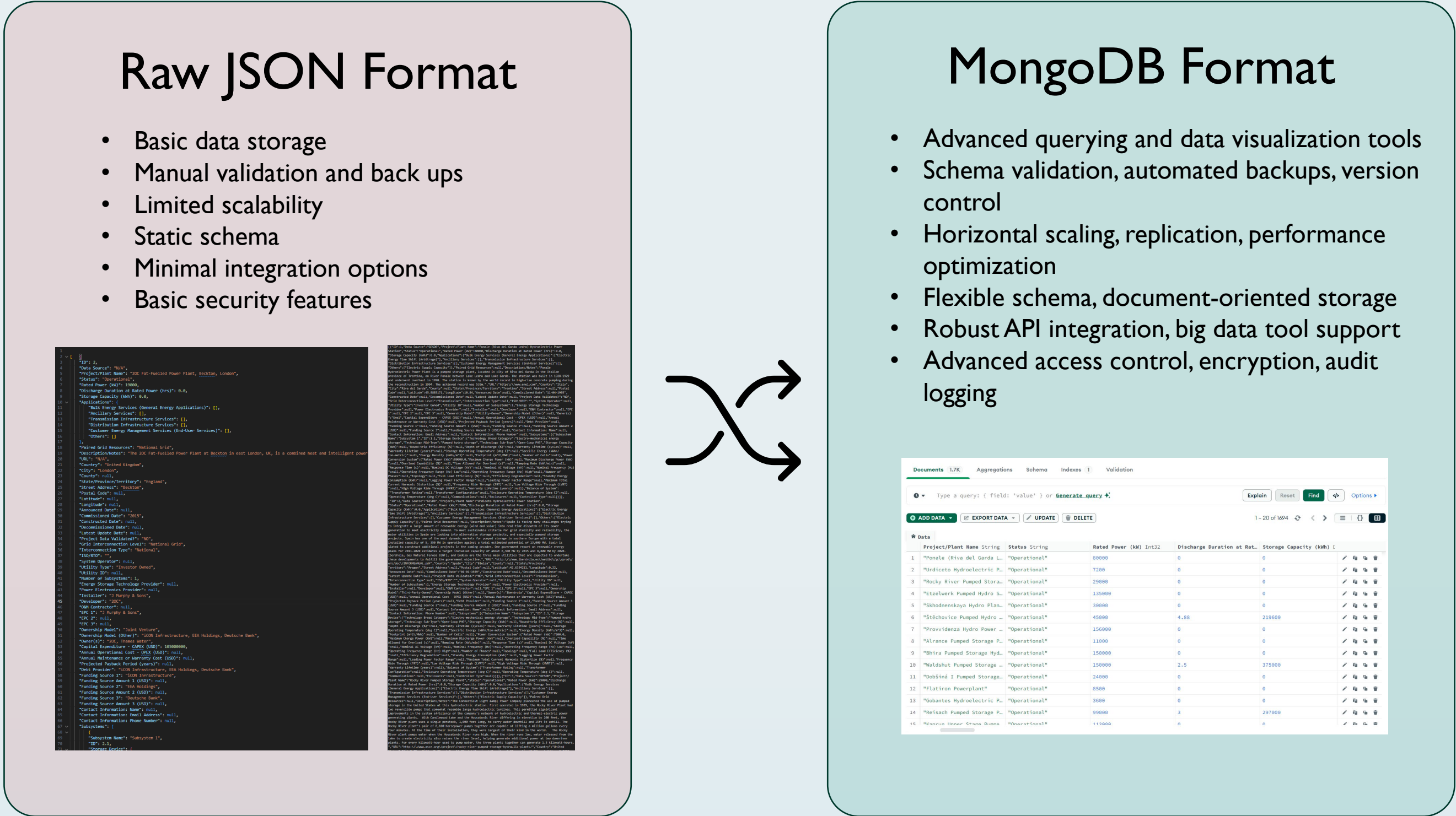**U.S. DEPARTMENT OF ENERGY** — Energy Storage Demonstration Projects

Data submitted directly by users

- Several new data sources now incorporated into GESDB
- **About 600 entries added / supplemented**
- Web-scraping tool leveraging OpenAI API in development to track new projects

Transition to Formal Database Management System

MongoDB is a source-available, cross-platform, document-oriented database program

Raw JSON Format
- Basic data storage
- Manual validation and back ups
- Limited scalability
- Static schema
- Minimal integration options
- Basic security features

MongoDB Format
- Advanced querying and data visualization tools
- Schema validation, automated backups, version control
- Horizontal scaling, replication, performance optimization
- Flexible schema, document-oriented storage
- Robust API integration, big data tool support
- Advanced access control, encryption, audit logging

Future Work
- Deploy web-scraping tool in coordination with the validation process to automate data entry procedure
- Continue to incorporate new data sources and update existing sources
- Deploy a chat-bot experience to automate data extraction and visualization

Sandia National Laboratories