

Managing Carbon Storage Data with a Living Database

Research & Innovation Center



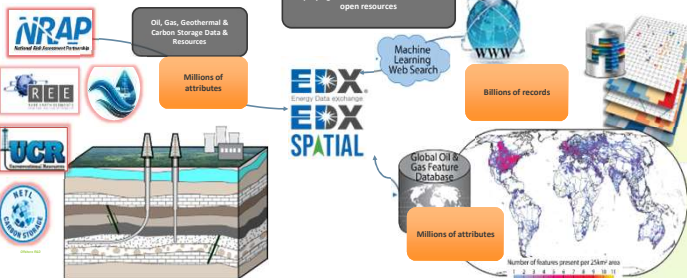
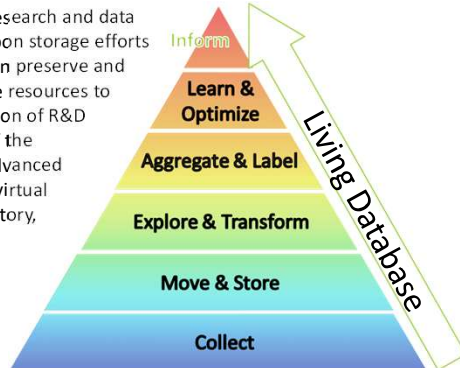
Michael Sabbatino^{1,2}; Paige Morkner^{1,2}; Abigail Choisser^{1,2}; Stephen Leveckis^{1,2}; Jennifer Bauer¹; Kelly Rose¹

¹National Energy Technology Laboratory, 1450 Queen Avenue SW, Albany, OR 97321, USA; ²NETL Support Contractor, 1450 Queen Avenue SW, Albany, OR 97321, USA

Why a LIVING Database?

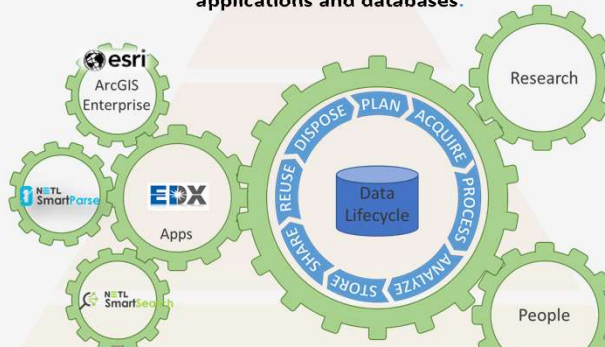
- Millions of dollars of research and data are available from carbon storage efforts
- The Living Database can preserve and efficiently access those resources to drive the next generation of R&D
- Can meet the needs of the community through advanced AI/ML tools via DOE's virtual data library and laboratory, EDX

A Living Database is KEY to future carbon storage research to integrate Machine Learning and other advanced data modeling.



The LIVING Database

The Living Database is a tool that integrates many types of data to support changing research needs through a framework of applications and databases.



The Living Database is:

- Centralized structured PostgreSQL database
- Interconnected system of applications, processes, and data stores

A Living Database can:

- Provide structured and unstructured data queries
- Leverage the capabilities of many resources and central data repositories
- Allow direct access to shared database without the need to download or extract
- Integrate spatial data with EDX Spatial using ArcEnterprise on AWS (Amazon Web Services)
- Incorporate many data sources with interactive maps and advanced ML tools
- Includes manual and automatic updates to data

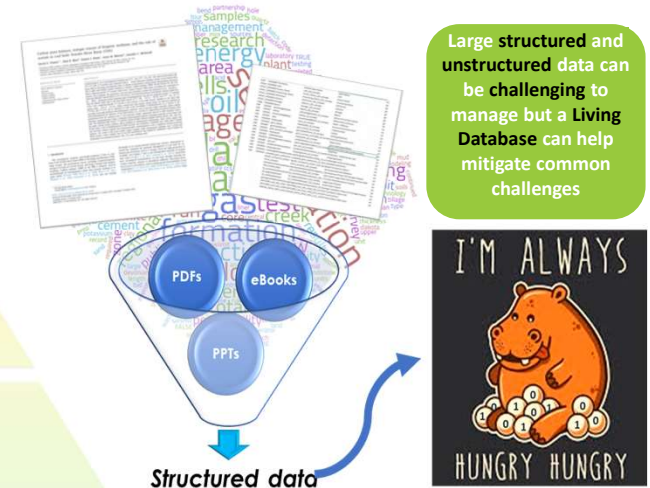
Benefits and Challenges

Challenges:

- Machine Learning Models thrive with abundance of data
- Lack of Labeled Training data for ML, AI and, NLP projects
- Living Database supports NETL research network

Benefits:

- Shift Living Database to Cloud Platform
- Expand the use of Living Database to additional projects
- More efficient Collaboration between projects



Large structured and unstructured data can be challenging to manage but a Living Database can help mitigate common challenges



Supporting Carbon Storage Data Lifecycle

Collection

- SmartSearch (artificial intelligence/machine learning (AI/ML) driven data collection)
- Expert-driven research
- EDX submissions
- Metadata development and capture

Cataloging

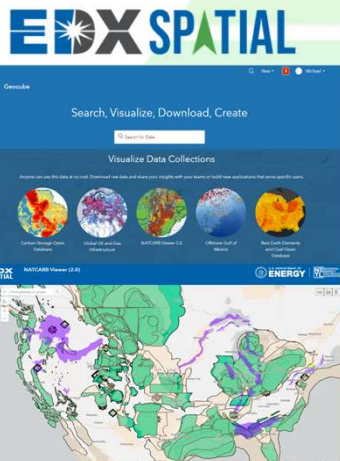
- ReadMe file development
- Natural language processing (NLP) for keywords, topic modeling, geographic association

Quality Assessment

- Data ranking
- Data assessment method scoring

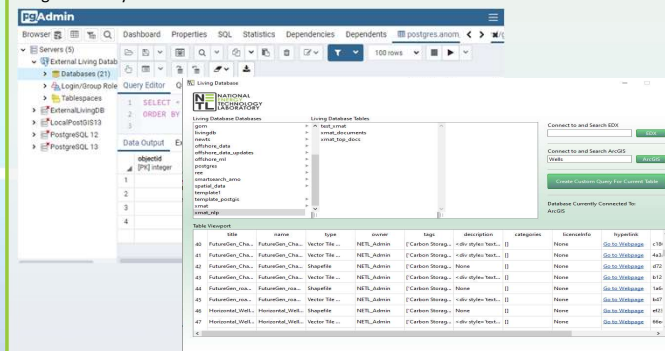
Data Organization and Publishing

- Private workspaces
- Submission packaging
- EDX Spatial



New Living Database Applications

The new Living Database application (beta version) allows users to query and download data from the on-site Living Database PostgreSQL server, EDX published data, and EDX Spatial Through ArcEnterprise. This tool expands upon the functions of data management tools and integrates many sources of data.



Front-end of the Living Database currently in beta testing.

Value Added

- Fossil Energy and Carbon Storage (CS) program investments into data curation and management has led to the development of AI/ML tools and the preservation of millions of dollars of research products which benefit ongoing and future research.
- Improved access through the integration of CS data resources on Living Database EDX into EDX Spatial, SmartSearch and SmartParse (for further searchability with spatial searches and keyword searches).
- EDX AI/ML data discovery, labeling, integration tool developments are trained to support CS and EDX4CC projects.

Disclaimer

Disclaimer: This project was funded by the United States Department of Energy, National Energy Technology Laboratory, in part, through a site support contract. Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.