The Sustainability Hub: Leveraging Cyberinfrastructure to Democratize Sustainability Data and Empower Community-Driven Research

Daniel Pittman, Alyssa Williams (Metropolitan State University of Denver); Kerstin Haring (University of Denver); Gregory Newman (Colorado State University)

Introduction

As sustainability challenges grow increasingly complex, the need for large-scale, real-time data to inform decision-making is critical. The **Sustainability Hub**, led by **Metropolitan State University of Denver**, a **Minority-Serving Institution (MSI)**, in collaboration with fellow **MSIs Colorado State University Pueblo (CSU Pueblo)** and **University of Northern Colorado (UNC)**, as well as **R1 institutions Colorado State University (CSU)** and **University of Denver (DU)**, aims to create a cloud-based platform for aggregating and democratizing sustainability data. By integrating datasets from sources such as citizen scientists, government entities, and environmental sensors, the Hub will enable communities, researchers, and policymakers to make informed, data-driven decisions.

This collaboration brings together expertise in citizen science, human-computer interaction, geospatial analysis, and community engagement to ensure broad participation and accessibility. The Hub's advanced **cyberinfrastructure** supports seamless data integration, real-time analysis, and widespread use in sustainability science. The project aligns with the goals of the **Grand Challenges for the Convergence of Computational and Citizen Science Research Workshop**, particularly in **Citizen Science Data Cyberinfrastructure, Software, Tools, and the Cloud**.

Project Overview and Alignment with Workshop Goals

The **Sustainability Hub** is being designed as a robust **cyberinfrastructure** to consolidate and democratize access to diverse sustainability datasets across Colorado. The platform aims to integrate **natural language processing (NLP)** and **large language models (LLMs)** to enhance user interaction with data, enabling automated data indexing and discovery across multiple sources. The system is intended to be deployed in a **cloud-based environment** to ensure scalability, real-time data access, and seamless automation. Below, we describe how these features align with the **Citizen Science Data Cyberinfrastructure, Software, Tools, and the Cloud** focus area of the workshop.

1. Integration of NLP and LLMs to Enhance Data Accessibility

The **Sustainability Hub** is being designed to leverage **natural language processing (NLP)** and **large language models (LLMs)** to simplify access to complex datasets. The platform will allow users to interact using **plain language queries**—for example, asking "What is the current air quality in my area?" or "How is climate change affecting water resources in Colorado?" These LLM-powered interfaces are intended to process these queries and retrieve relevant data from multiple sources. This **AI-driven interaction** is expected to lower barriers for non-expert users, enabling participation in sustainability science without needing technical expertise.

Alignment with Workshop Goals: By making data more accessible through natural language queries, the Hub aligns with the workshop's goals of developing **cyberinfrastructure tools** that enhance citizen science. The integration of NLP and LLMs is designed to enable non-experts to query, analyze, and contribute to data, expanding the reach and inclusivity of sustainability science.

2. Cloud Deployment for Scalability and Automation

The **Sustainability Hub** is being developed with a **cloud-based infrastructure** to provide the scalability needed to handle large volumes of data from diverse sources. The platform is intended to integrate real-

time environmental data from sensors, citizen-contributed data, and government datasets into a centralized cloud repository. By using **cloud environments**, the Hub is designed to scale dynamically as more data is added, ensuring it can accommodate growing data volumes and user participation across the state.

The system is also being designed to enable **automation** in data processing and indexing. As new data sources are added, the goal is for the platform to automatically index them, making them instantly searchable and accessible to users. This automation would reduce the time and effort required to maintain the platform's data repository, supporting seamless data integration from citizen science projects, researchers, and government agencies.

Alignment with Workshop Goals: The cloud deployment of the Hub aligns with the workshop's emphasis on using **cloud computing resources** to support citizen science. The automated indexing of new data from multiple sources is expected to make sustainability data widely available and easily searchable, directly addressing the workshop's goal of developing scalable and efficient tools for citizen science.

3. Democratization of Data Access Across the State

The primary goal of the **Sustainability Hub** is to **democratize access** to sustainability data. The platform is being designed to integrate diverse data sources—including real-time air and water quality monitoring systems, geographic datasets, and citizen-contributed data—into a single, cloud-based platform accessible to researchers, policymakers, and communities across Colorado. The use of NLP and LLMs will allow users to interact with the data through a **natural language interface**, making the platform accessible even to those with no technical background.

By building a **cloud-based infrastructure**, the Hub is being designed to provide **statewide access** to data, ensuring that communities in remote or underserved areas can easily analyze and contribute data related to local sustainability challenges. The platform will offer **data visualization tools** to help users explore raw data, with the future goal of adding functionality to help identify trends, patterns, and anomalies.

Alignment with Workshop Goals: The democratization of data access through a cloud-based platform, combined with the planned **natural language interface**, aligns with the workshop's objective to enhance citizen science through **cyberinfrastructure, tools, and the cloud**. The platform's intended ability to provide real-time, automated access to sustainability data will support broad participation, contributing to the workshop's goals of broadening engagement in scientific research.

Conclusion

The **Sustainability Hub** is being developed as an advanced platform that will leverage **LLMs, NLP**, and **cloud computing** to transform citizen science and enhance sustainability efforts across Colorado. The platform's design integrates AI tools, real-time data processing, and automation to ensure that sustainability data is accessible, scalable, and actionable for all stakeholders.

By aligning with the goals of the **Citizen Science Data Cyberinfrastructure, Software, Tools, and the Cloud** focus area, the **Sustainability Hub** aims to democratize data access, engage diverse communities, and support informed decision-making at both local and state levels. Participating in the **Grand Challenges for the Convergence of Computational and Citizen Science Research Workshop** will allow us to refine these approaches further and contribute valuable insights to the broader computational citizen science community.