Testimony M Derek Mackenzie
mdm7@ualberta.ca

Agriculture Data Institute

Soil Health: the backbone of Agriculture

Soil health refers to the ability of soil to function as a living ecosystem capable of sustaining plants, animals, and humans. A healthy soil supports the growth of plants by providing essential nutrients, water, and oxygen. It also contains a diverse community of microorganisms, including bacteria, fungi, and other organisms, which play a vital role in decomposing organic matter, cycling nutrients, and improving soil structure. As a metaphor for soil function, soil health is easier for the general public to understand, and as such is a critical concept for sustainable agriculture and human health.

Soil health is critical to sustainable agriculture because functioning soils support global food security. Healthy soils can provide essential nutrients, water, and air to plants, which are necessary for their growth and development. Healthy soils also have a vibrant consortium of microorganisms, and it is the plant soil microorganism continuum that allow soils to function at the highest level. By maintaining soil health, farmers can reduce their environmental impact and promote sustainable agriculture practices that benefit both their crops and the wider ecosystem. Ultimately, sustainable agriculture practices that prioritize soil health can help to preserve agricultural productivity for future generations.

Machine Learning and Data Science

Machine learning is a type of artificial intelligence that involves the use of algorithms and statistical models to enable computer systems to learn and improve performance on a specific task without being explicitly programmed. In other words, machine learning algorithms are designed to recognize patterns and relationships in data and use this information to make predictions or decisions, which could be related to agricultural management.

Soil Health and Machine Learning

AI can play a significant role in integrating agricultural best management practices (BMP) and comprehensive soil assessment data by enabling the analysis and interpretation of large and complex datasets. One possible approach is to use machine learning algorithms to identify patterns and relationships between different types of data, which can help to identify how BMP affect soil health and function. Here are a few specific ways in which AI can help integrate molecular microbial data and comprehensive soil assessment data:

Identifying microbial species and functions: Machine learning algorithms can be trained to identify and classify microbial species based on molecular data, such as DNA sequences. These algorithms can also be trained to predict the functions of microbial communities based on their molecular profiles. By integrating molecular microbial data with comprehensive soil assessment data, machine learning algorithms can provide a more comprehensive understanding of the microbial communities in the soil ecosystem and their functions.

Predicting soil health indicators: AI can be used to predict soil health indicators based on molecular microbial data and comprehensive soil assessment data. For example, machine learning algorithms can be trained to predict soil organic matter content, water-holding capacity, and nutrient availability based on the microbial communities present in the soil. By predicting soil health indicators, AI can help to identify areas that require intervention to improve soil health.

Developing personalized soil management plans: By integrating molecular microbial data and comprehensive soil assessment data, AI can help to develop personalized soil management plans that are tailored to the specific needs of each soil ecosystem. Machine learning algorithms can be used to identify the optimal management practices for a given soil ecosystem based on its molecular microbial profile and other soil health indicators.

Overall, AI can help to integrate molecular microbial data and comprehensive soil assessment data by providing a more comprehensive understanding of the soil ecosystem and its functions. By leveraging machine learning algorithms, AI can help to identify patterns and relationships between different types of data, which can lead to more accurate predictions and personalized soil management plans. From these relationships we can create an artificial intelligence system like ChatGPT to help farmers with long range planning.

Proposed Institute

I propose that the institute be an endowed NGO with a corporate structure of 5 executives, 18 Ag scientists, 6 data scientists, 3 IT personnel, 3 administrative personnel, physical office space, and data science infrastructure. The executive would work to design strategic objectives with a Board of Governors made up of senior academic, government, industry, and farmer representatives. An independent institute is needed to manage data security and privacy, and to ensure that data remains is only used for its mandated purpose. Data sharing agreements need to be held by an single institute that can navigate the legal environment of data privacy and not for profit use of publically funded data. A draft budget for the Institute follows below.

Draft Budget

Capital Asset Cost

- Workstations/Cubicles: $1,000 per workstation, total for 35 workstations/cubicles $35,000
- Desk Chairs: $250 per chair, total for 35 desk chairs $8,750
- Filing Cabinets: $300 per cabinet, total for 35 filing cabinets $10,500
- Meeting Tables: $1,000 per table, total for 3 meeting tables $3,000
- Reception Desk: $3,500
- Computers and communications: $1,000 per employee, total $35,000
- Data Server: A high-end data server for 1000 TB $75,000
- Data Storage: A storage system for 1000 TB $50,000
- Network Equipment: Network equipment such as routers, switches, and firewalls $25,000
- Cooling and Power Equipment: Cooling and power equipment such as UPS (uninterruptible power supply) systems and air conditioning units $25,000

One-time capital asset cost: $270,750

Total Annual Cost

Office Space:

- 35 employees * 150 square feet per employee = 5,250 square feet
- 5,250 square feet * $30 per square foot per year = $157,500 per year

Employee Salaries:

- 30 non-executive employees * $120,000 per year per employee = $3,600,000 per year
- 5 executive employees * $200,000 per year per employee = $1,000,000 per year

Benefits and Insurance:

- 20% of salaries year = $920,000 per year

Total Annual Operating Costs: $5,677,500 per year

Endowment Funds

In order to cover this annual operating budget in perpetuity, an endowment fund would need to be established. As a rough estimate, assuming a conservative expected annual investment return of 4% and an inflation rate of 2%, a withdrawal rate of 3.5% would be sustainable for the endowment fund. Using the 3.5% withdrawal rate, the endowment fund would need to be approximately 28 times the annual operating budget to cover it.

Therefore, to cover an annual operating budget of $5.7 million, an endowment fund of approximately $159.6 million (28 x $5.7 million) would be required.

Next steps

1. Workshop this year to bring all stakeholders together to work on one project.
    a. Organized by MacKenzie
2. Proof of concept, show the benefits of data science for agriculture.
    a. Including projects already in progress, such as:
        i. DASH
        ii. National Soil Database
        iii. ACS Living Labs
        iv. SAFEHUB – Genome Canada Project
3. Look for grant to get initiative off the ground.
    a. Write the corporate structure, terms of reference, and operating structure
    b. Hire fund raising manager and begin fund raising.
    c. Begin head hunting the executive and BoG.
4. Once Endowment is in place, decide on location, start hiring personnel, and integrate work done to date.