

Canadian Digital Research Infrastructure Needs Assessment: An Agri-food Research Perspective from the University of Guelph

December 14, 2020

Prepared by: Emily Duncan^{1*}, Rozita Dara², Jairo Melo³, Carly Huitema⁴, & Beverley Hale⁵

*corresponding author (edunca01@uoguelph.ca)

1 Department of Geography, Environment and Geomatics University of Guelph

2 School of Computer Science, University of Guelph

3 Senior Manager of Data Strategy, Food from Thought, University of Guelph

4 Research Data Strategy, Food from Thought, University of Guelph

5 Associate Vice-President Research (Agri-food Partnership), University of Guelph

Executive summary

Digital data is essential to providing researchers the information they need to address the myriad of challenges facing society. Researchers at the University of Guelph aim to use digital agricultural data to find ways to sustainably and safely feed the world's growing population. In this paper, our intent is to provide input on the needs of our research institution related to creating a digital research infrastructure (DRI) ecosystem that is supported by tools and technologies to conduct cutting-edge research in agri-food systems and to share our vision about the future of agriculture research data management. We begin by outlining the context of agri-food research at the University of Guelph and current issues. Then we present our vision for the future state of DRI in Canada and some potential strategies to move towards a shared vision. We emphasize the need for data governance strategies to be prioritized.

University of Guelph: A Global Leader in Agri-Food Research

The University of Guelph is known as a world-class research institution for sustainable agri-food production. As “Canada’s Food University”, we are committed to improving life through innovative research. The University of Guelph is also privileged to have a unique partnership with the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) known as the Ontario Agri-food Innovation Alliance. Through this partnership, the University of Guelph manages a provincial network of fifteen research centres that promote agri-food discoveries, validate laboratory findings, stimulate future research, and provide valuable information for Ontario’s agri-food sector. This collaboration works to advance research and innovation that contributes to the success of the province’s agri-food sector and promotes rural economic development. Our unique alliance with OMAFRA positions the University of Guelph to be able to provide crucial research to ensure that Ontario solutions for healthy, safe, and sustainable food have global impact. Additionally, one of our key research programs funded by the Canadian First Research Excellence Fund is Food from Thought: Agricultural Systems for a Healthy Planet. The goal of this program is to increase the sustainability and productivity of global food production by leveraging the power of leading-edge data science, agri-food research, and biodiversity science.

One example of our state-of-the-art research facilities is the Ontario Dairy Research Centre in Elora¹. We have created a portal for daily management data produced by the facility, which enables broad access to those management data, and is soon to be replicated for our beef research facility. Data from

¹ <https://www.uoguelph.ca/alliance/research-facilities/elora-dairy>

various sensors in the dairy facility are collected and transferred securely into the servers that are managed by the Computing and Communications Services at the U of Guelph. An on-farm server gathers and validates research facility's data before sending them to the central data repository on campus. On-campus servers host relational databases and a website that enable user friendly querying and provide access to raw farm data. Data access is controlled and managed by an administrator, supporting flexible and robust data acquisition, transfer, and storage from different sensors and devices. Our long-term needs are to create an Agriculture 5.0 Observatory Decision Support System that will link to the dairy and beef portals, as well as to other sources of data generated by researchers and stakeholders, which might be of use in creating 'best net solutions'. Such data include those from the soil lysimeters at our soil research station and from the long-term agronomy trials.

In recent years, the agri-food sector has been recognized to be a key driver of Canadian economic growth. For instance, in February 2017 the Federal Government's Advisory Council on Economic Growth, led by Dominic Barton, identified the agri-food sector to be a strategic area for economic growth². This policy direction has been echoed by several subsequent reports including the Report of Canada's Economic Strategy Tables: Agri-food³, Growing Canada's value-added food sector by the Standing Senate Committee on Agriculture and Forestry⁴, and the B.C. Food Security Task Force⁵. For Canadians to benefit from this economic growth in the agri-food, the sector will need accessible digital infrastructure and increased innovations in automation and digitalization. In turn, for research institutions like the University of Guelph to support the growth of the agri-food sector, it is crucial to have a strong digital research infrastructure ecosystem.

Current Status

In outlining the issues facing in agriculture and food, we discuss the main DRI resources that are currently in use, the DRI resources that are inaccessible, and some of the key challenges with using available DRI resources.

Through Ontario Agri-food Innovation Alliance, researchers can access DRI tools at the aforementioned fifteen research centres, such as sensors, wireless networks, and in-house servers. These research stations also provide access to software for data visualization or analysis, often directly from the technology provider. The Alliance is providing some support for building a data platform for these facilities. The main DRI resources that are currently in use on an individual faculty level at the University of Guelph include advanced research computing from both Compute Canada and some in-house infrastructure. University infrastructure is managed through our Computing and Communications Services (CCS) department and is available to all researchers. Additional digital infrastructure is supported through individual grants through Tri-Council funding. Other research software for data processing is available, such as programs like Python or R.

DRI ecosystem components that are currently lacking include tools, apps, and devices for data generation. Agricultural data is unique in that it captures a time-sensitive snapshot of the natural world in settings that are often difficult to replicate, like long-term crop yield trials or methane emissions from dairy cattle with specific genetic traits. Therefore, it is imperative to have the necessary technologies to be able to capture and share these types of agricultural data. At a more rudimentary

² <https://www.budget.gc.ca/aceg-ccce/pdf/key-sectors-secteurs-cles-eng.pdf>

³ [https://www.ic.gc.ca/eic/site/098.nsf/vwapj/ISED_C_Agri-Food_E.pdf/\\$file/ISED_C_Agri-Food_E.pdf](https://www.ic.gc.ca/eic/site/098.nsf/vwapj/ISED_C_Agri-Food_E.pdf/$file/ISED_C_Agri-Food_E.pdf)

⁴ https://sencanada.ca/content/sen/committee/421/AGFO/reports/AGFO_SS-5_Report_FinalV3_e.pdf

⁵ <https://engage.gov.bc.ca/app/uploads/sites/121/2020/01/FSTF-Report-2020-The-Future-of-Food.pdf>

level, gaps in connectivity prohibit extensive agricultural data connection. Additionally, there is the requirement for access to secure and high-performance computing. Finally, once data are collected, there is a need for data management tools. These data management tools can aid in meta-data creation, anonymization, and standardization of data. To summarize, despite advances we have made with our data platform, important issues include: access to infrastructure such as data platforms, access to standardized data, capability to share data, availability of meta-data, and research data governance challenges such as determining intellectual property through data agreements.

There are two pressing challenges to be able to optimize the use of currently available DRI resources and to be able to fully utilize tools that may be accessible in the future: the lack of qualified personnel and the lack of data governance mechanisms. Following the flow of data through the research ecosystem illustrates these challenges. First, data is generated at research stations or labs, however there is a need for skilled people who understand how to use the existing resources and infrastructure, to integrate software tools, and to analyze data for research teams. Qualified personnel must be equipped with interdisciplinary skills to use the DRI ecosystem and understand the specific needs of the agricultural sector. Secondly, data flows to individual faculty for their research, but there is a challenge around governance mechanisms. Lack of standardization around data, and specifically meta-data, is a general limitation to research across sectors. The lack of taxonomies and ontologies for agricultural data limits integration and accessibility. Researchers might spend large amount of time and resources manipulating data into the forms needed for their research. Then researchers may want to share data, either raw or aggregated, with collaborators but the absence of processes and protocols for data sharing, anonymization, data exchange, and ownership leads to limited access.

Particularly, in agricultural studies, researchers often rely on collaborations with industry partners for access to data, but the lack of protocols around confidentiality can undermine the trust of collaborators to allow researchers to store and process their data through DRI tools. In some cases, this can lead to collaborators only allowing researchers access to aggregated data which is less preferable than access to raw data. Boundaries around data ownership are obscured when technology and infrastructure at research centres produce raw data, but after the data has been altered by researchers, it is unclear who owns it. Lack of transparency around the data life-cycle leads to the creation of silos in research, data duplicity, and inconsistency.

Future DRI State

Our vision for a cohesive Canadian DRI ecosystem that would fulfill our research needs includes strong leadership in data governance mechanisms. Standardization of DRI resources is necessary to facilitate the use of data in research. In an ideal DRI ecosystem, there would be clear communication about data availability across the research community to enable collaborations and efficient use of resources. Equitable access to integrated and reliable infrastructure for all disciplines is a key principle for a functional DRI ecosystem. The infrastructure should be the basis for enabling FAIR data practices⁶ (findable, accessible, interoperable, and reusable). The final step for designing cohesive DRI is automation – data loading should be consistent through automation, data schemas for meta-data can automatically be generated, and updates to data lifecycles should be automated through data retention policies.

⁶ Wilkinson, D. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018–160018. <https://doi.org/10.1038/sdata.2016.18>

One of the current challenges is organizational resistance due to the lack of integrated tools with governance and policies. Researchers need to be incentivized to share data; however, the current infrastructure makes this task onerous due to the lack of interoperability and potentially risky due to lack of clear ownership. Integrated DRI resources will require the development of a systemized process to request access to data through authorization requests.

How to Bridge the Gap

In order to achieve a desirable future state of DRI in Canada, there is a demonstrable need for qualified personnel dedicated to developing the protocols and processes for data storage, processing, and standards for data, meta-data, and tools. While NDRIO can provide support in the form of funding for the infrastructure required to develop a more FAIR data ecosystem, we see NDRIO fulfilling a crucial gap by providing leadership in data governance mechanisms and defining best practices. One example could be defining systematic best practices for building research facilities, which would provide guidance around how new facilities can be interoperable and what updates existing facilities might require.

Not only can NDRIO provide collaboration among research in Canada, but it can also act to facilitate collaboration between international communities. As we work to build data governance mechanisms nationally, it is key to ensure that the standards we develop are interoperable with international collaborators. We emphasize that this is highly important for agricultural datasets, as what we learn here from agricultural research has the potential to be applied elsewhere expanding research impact.

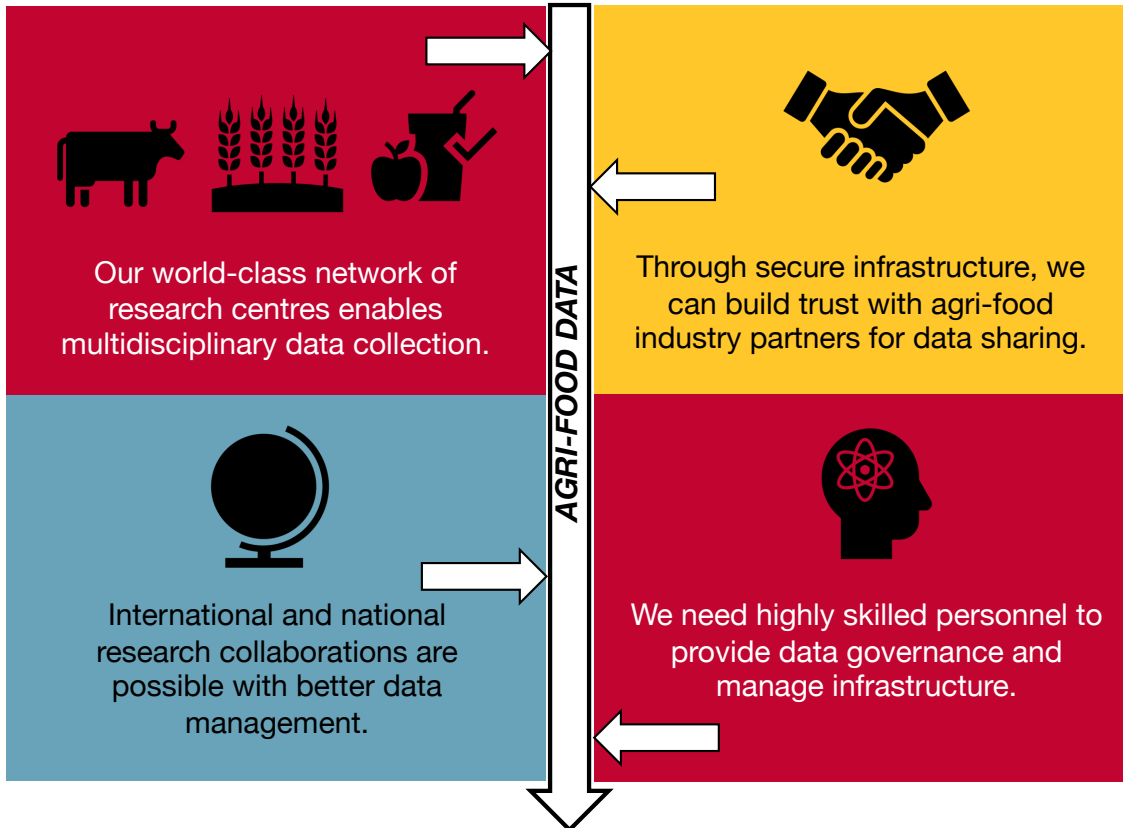
Another area where NDRIO can bridge gaps in research infrastructure, particularly for the agri-food sector, is to enable collaboration with the private sector. In enhancing data governance mechanisms, research institutions will be able to build trust regarding data storage and management with private sector partners. We foresee the possibility of creating a national consortium for agriculture data management and standardization which would create specific policies for agri-food data governance. NDRIO would be a valuable partner in establishing such a venture. Agri-food Data Canada, supported by the Food from Thought initiative at the University of Guelph, aims to build an advanced management system which will allow authorized users to access and share data in a secure environment.

Conclusion

As agri-food research becomes highly dependent on digital technologies to address important questions, such as how to achieve sustainable production, research institutions like the University of Guelph require a digital research infrastructure ecosystem that facilitates innovation through effective data governance mechanisms. Our current DRI is shaped by pioneering platforms such as Agri-food Data Canada from the Food from Thought research program, the Ontario Agri-Food Innovation Alliance, and access to resources through our Computing and Communication Services. We envision that through our leadership in agri-food research, we will develop an agri-food data platform that will bring together stakeholders and facilitate innovation in the Canadian agri-food sector. While interoperability persists as an issue for digital research across sectors, we see NDRIO addressing this challenge by developing data governance mechanisms in collaboration with input from research institutions, both nationally and internationally, and with involvement from the private sector. Efforts to implement advanced research computing, data management strategies, and new research software need to be supported by the accompaniment of highly qualified personnel with interdisciplinary skillsets to ensure that DRI resources can be utilized to their full capacity. Agricultural data and

technologies hold benefits for all Canadians through creating a more sustainable and transparent food system. NDRIO can support an improved food system by developing the tools, technologies, and data governance mechanisms that will support leading-edge research by institutions like the University of Guelph and create a more sustainable agri-food data management ecosystem.

University of Guelph: Our Vision for a Comprehensive Agri-food Data Ecosystem



Agri-Food Data Canada AT THE UNIVERSITY of GUELPH

Supported by the Ontario Agri-food Alliance, Food from Thought, and through a potential future opportunities with NDRIO, Agri-Food Data Canada will provide a data management system to fuel innovation in the Canadian agri-food sector.

