

VIEWPOINT • OPEN ACCESS

HarvestStat: a global effort towards open and standardized sub-national agricultural data

To cite this article: Kyle Frankel Davis *et al* 2025 *Environ. Res. Lett.* **20** 052001

View the [article online](#) for updates and enhancements.

You may also like

- [Household, community, sub-national and country-level predictors of primary cooking fuel switching in nine countries from the PURE study](#)
Matthew Shupler, Perry Hystad, Paul Gustafson et al.
- [Spatially detailed agricultural and food trade between China and the United States](#)
Akshay Pandit, Deniz Berfin Karakoc and Megan Konar
- [Global virtual water trade and the hydrological cycle: patterns, drivers, and socio-environmental impacts](#)
Paolo D'Odorico, Joel Carr, Carole Dalin et al.



ECS The Electrochemical Society
Advancing solid state & electrochemical science & technology

247th ECS Meeting

Montréal, Canada
May 18-22, 2025
Palais des Congrès de Montréal

ECS UNITED

Unite with the ECS Community

**Register to
save \$\$
before
May 17**

ENVIRONMENTAL RESEARCH
LETTERS

VIEWPOINT

OPEN ACCESS

RECEIVED

3 April 2025

ACCEPTED FOR PUBLICATION

10 April 2025

PUBLISHED

23 April 2025

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

HarvestStat: a global effort towards open and standardized
sub-national agricultural data

Kyle Frankel Davis^{1,2,14,*} , Weston Anderson^{3,14} , Steffen Ehrmann^{4,5,14} , Rafaela Flach^{6,14} ,
Carsten Meyer^{4,5,14} , Jonathan Proctor^{7,14} , Deepak Ray^{8,14} , Liangzhi You^{9,14} , Michael Foley^{10,14} ,
Hervé Kerdiles^{11,14} , Andrew Hultgren^{12,14} , Peter Huybers^{10,14} , Endalkachew Kebede^{1,14} ,
Michele Meroni^{11,14} , Caro Park^{10,14} and Donghoon Lee^{13,14}

¹ Department of Geography and Spatial Sciences, University of Delaware, Newark, DE, United States of America

² Department of Plant and Soil Sciences, University of Delaware, Newark, DE, United States of America

³ Department of Geographical Sciences, University of Maryland, College Park, MD, United States of America

⁴ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

⁵ Institute of Biology, Leipzig University, Leipzig, Germany

⁶ Stockholm Environment Institute, Stockholm, Sweden

⁷ Faculty of Land and Food Systems, The University of British Columbia, Vancouver, Canada

⁸ Institute on the Environment, University of Minnesota Twin Cities, Minneapolis, MN, United States of America

⁹ International Food Policy Research Institute, Washington, DC, United States of America

¹⁰ Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, United States of America

¹¹ Joint Research Centre, European Commission, Ispra, Italy

¹² Department of Agricultural and Consumer Economics, University of Illinois Urbana-Champaign, Champaign, IL, United States of America

¹³ Department of Civil Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

¹⁴ All authors contributed equally.

* Author to whom any correspondence should be addressed.

E-mail: kfdavis@udel.edu

Keywords: HarvestStat, subnational, agricultural statistics, FAIR guidelines, open data

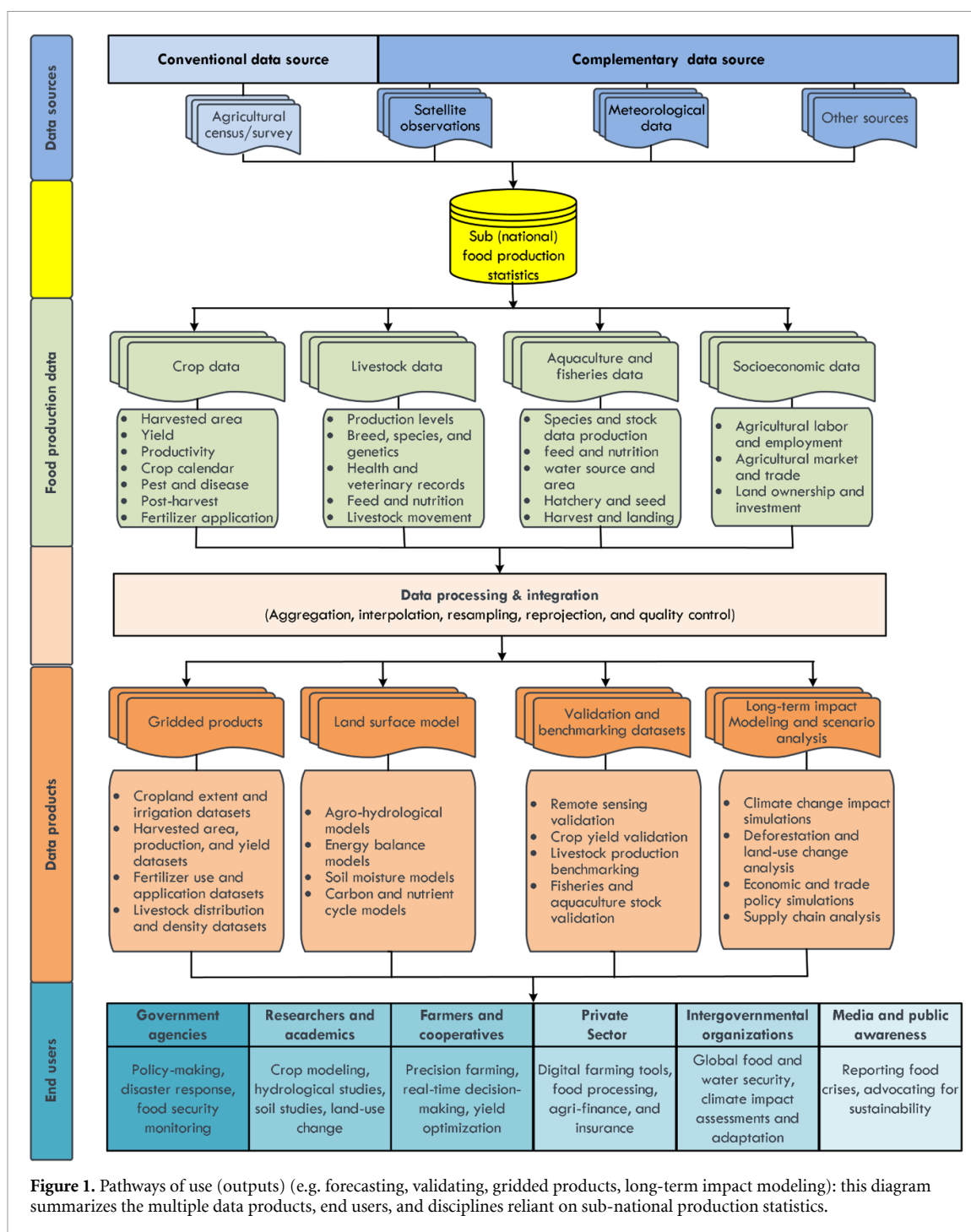
Abstract

Agricultural production statistics underpin diverse research efforts and development activities. Yet despite their critical importance, efforts to collate, update, and harmonize detailed sub-national agricultural production statistics are frequently redundant and incomplete due to the substantial time, effort, and resources required. The persisting lack of coordination and standards in the food systems data community wastes valuable resources and hinders advances in action-oriented food systems knowledge. Here we introduce the HarvestStat sub-national data consortium as an open-source, collaborative, and transparent model to overcome these challenges. HarvestStat is collaboratively producing publicly available databases and datasets for the food systems community and the broader environmental and sustainability sciences by moving beyond closed and disjointed data-gathering efforts. We are guided by core principles of complete data openness—prioritizing high standards of quality assurance; active inclusion—emphasizing involvement from local experts; and collaboration—fostering engagement across communities of data producers and users. We extend an open global call to action, inviting organizations and individuals to engage in advancing this critical agenda.

1. The central importance of sub-national
agricultural statistics

Sub-national agricultural production statistics underpin many analyses, assessments, and data products related to food security, land use, climate, and natural resource management (figure 1). These statistics and derived outputs are critical for

informing development agendas, food security and sustainability policy, humanitarian aid response, and public and private investments. The comprehensiveness, level of detail, and accuracy of agricultural statistics thus directly determine information blindspots—often in regions where food security and agricultural development needs are most pressing. In turn, interventions relying on outdated, coarse, or



inaccurate estimates may, at best, result in ineffective policy and action and, at worst, lead to adverse outcomes. While an array of sub-national agricultural data is made available by the countries that produce them—with varying formats and ease of access, a confluence of obstacles to data quality and openness has meant uneven data gathering, collation, and access across countries and organizations, thereby severely hindering the availability of harmonized, analysis-ready sub-national production statistics.

2. Challenges and the path forward

Agricultural statistics from each country present unique challenges regarding data quality, openness, and comprehensiveness. These information gaps are partly due to the substantial costs for a country to gather and produce agricultural census and survey statistics, with widely varying levels of accuracy, reliability, and willingness to share data [1]. These challenges are further exacerbated by legal restrictions on data openness, which can substantially constrain the

Table 1. Existing datasets providing multi-country sub-national agricultural production statistics.

Dataset name	Agricultural products covered	Spatial coverage	Spatial granularity	Time period	Source
HarvestStat Africa	90 crop types	35 Countries	It varies by country, covering 20 countries at the Admin 1 level and 15 countries at the Admin 2 level.	1980–2022	https://github.com/HarvestStat/HarvestStat-Africa [2]
Agriwanet	14 crop types	5 Countries (Central Asia)	Country and Oblast (Admin 2)	1992–2015	https://doi.org/10.7802/2008 [3]
Regional Strategic Analysis and Knowledge Support System (ReSAKSS)	Differs among nations	26 Countries	It varies by country, Admin 1 and Admin 2 level	Varies by country; 2005–2023	https://eatlas.resakss.org/ [4]
CountrySTAT	Differs among nations ranging from 4 to more than 23	31 countries	The specific granularity of data varies by country and is available at Admin 1 and Admin 2 levels.	Varies by country; 1992–2017	www.fao.org/in-action/countrystat/national-countrystat-sites/en/ [5]
Agro-MAPS	Primary food crops: varies by country	134 countries	It varies by country: 130 countries—Admin1; 59 countries—Admin2	Varies by country and crop	https://gaez.fao.org/pages/agromaps [6]
Eurostat	79 crop classes, 21 poultry variables, 33 dairy variables, and 179 livestock variables.	European Union—27 countries	It varies by country and is disaggregated to regional levels (NUTS 1 to NUTS 3).	Varies by dataset: 1967–2024	https://ec.europa.eu/eurostat/web/main/data/database [7]
Eurocrop	A diverse range of crop types, organized using the Hierarchical Crop and Agriculture Taxonomy (HCAT)	European Union—16 countries	Georeferenced polygons at a field level, with harmonization across countries.	Varies by country; data primarily from 2018 to 2021	www.eurocrops.tum.de/index.html [8]

likelihood of organizations mandated to be hubs for food systems research and data (e.g., FAO) to share all existing information openly. As a result, substantial time, effort, and resources are required to gather, clean, and harmonize agricultural production statistics for a growing array of applications—meaning that only a handful of multi-country sub-national datasets currently exist (table 1). Yet differences in funding cycles, researcher priorities, institutional siloing, and incentive structures have led to duplications of data-gathering efforts across the food systems research community. In contrast, less privileged

researchers and teams often cannot collate or access more extensive integrated databases. Duplications in data-gathering and processing pipelines not only divert meaningful resources that could be funneled into improving accessibility but also limit the comparability of the results of modeling efforts, as the data itself may vary depending on the sourcing institution and the specific cleaning methods applied. This comparability issue is further compounded by a general lack of uncertainty and error quantification accompanying the available data, making it difficult to assess the reliability of existing agricultural statistics. This

compound challenge—a shortage of data availability, data accessibility, and data validation—results in a system where only a small subset of stakeholders can access and utilize spatiotemporally refined statistics, often with little regard to data validation. The mission of the HarvestStat consortium is to challenge this status quo by overcoming key barriers to sustained and comprehensive data openness to meaningfully address this data scarcity grand challenge of food systems.

3. Guiding principles for open production statistics

Overcoming this core challenge requires a sustained, transdisciplinary, and multi-institutional effort. The HarvestStat sub-national data consortium aims to contribute to overcoming these challenges by: (1) advocating for open subnational agricultural data and collaboration on this topic, (2) collaborating in compiling data towards a subnational crop database, and (3) creating a community for collaboration on producing quality-assured data products/datasets based on this data. The consortium is a convening space to increase collaboration and advocacy, but most importantly, it is meant to be a space for data sharing and the creation of standardized procedures. Concretely, the core guiding output of the HarvestStat consortium is a subnational dataset of agricultural production statistics produced and made available with the FAIR principles of Findability, Accessibility, Interoperability, and Reusability [9]. This dataset and the consortium at large are guided according to the following principles:

- a) Unconditional openness—in which all possible data are freely and publicly shared, with transparent documentation and sourcing.
- b) Active inclusion—in which consortium members actively seek involvement and input from local experts in all countries to ensure accurate, sensitive, and effective use of data, as well as from different research and policy communities to ensure that the datasets produced meet the needs of potential data users.
- c) Collaboration—in which consortium members prioritize providing tools, platforms, and solutions that are scalable and open to the community.

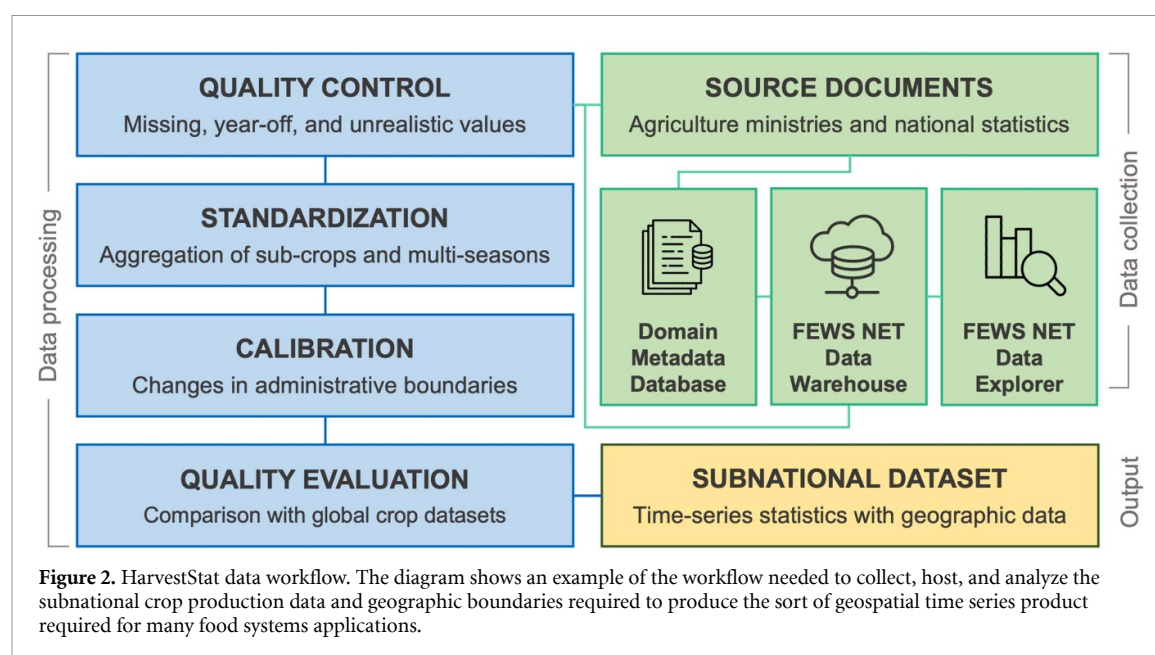
The HarvestStat group also prioritizes: standardization and validation—in which variables are reported in a consistent structure and with consistent units across diverse agricultural systems, and measures of data quality are included to guide appropriate data use; comprehensiveness—in which data from as many countries as possible is included at the

admin level or finer; and geolocation—in which global coordinates and shapefiles are linked to statistics to facilitate geographic analysis.

4. Framework and standards of HarvestStat

The HarvestStat consortium is already building an ecosystem of data and tools with FAIR-aligned principles applied along all data processing pipelines for compiling observational datasets useful for food systems researchers. For example, in Africa, where data scarcity is greatest [1], HarvestStat Africa presents a collaborative and entirely open workflow for producing subnational datasets that can be readily replicated in other regions [2]. HarvestStat Africa is a dataset built from crop production statistics collated originally in USAID's FEWS NET Data Warehouse (FDW) and mirrored to the Harvard Dataverse. The compilation and hosting of subnational statistics in a publicly-available location catalyzed the collaborative efforts of researchers and practitioners. As a sufficiently large publicly available database of subnational data, the FDW ensured that the subsequent efforts of individual research groups contributed to a product suitable for pursuing globally-relevant research questions. The data hosted in the FDW and the approach used to produce HarvestStat Africa benefits the community in three primary ways: (1) it increases the publicly available compilation of crop production statistics in a single, centralized location; (2) it provides open-source post-processing code to standardize and harmonize the available data to account for changing reporting standards and administrative boundaries; and (3) it publishes the resulting dataset in a public repository for use without restrictions (see figure 2). The effort, therefore, provides the building blocks for others to develop fit-for-purpose datasets by freely providing access to the underlying data—which was collated over many years—and the requisite code needed to process that data.

Efforts such as HarvestStat Africa, however, first require that the data and geographic boundaries be collated and stored in a centralized database. In this regard, members of the HarvestStat consortium have also produced open-source tools for community use. For instance, the census-harmonization module in LUCKINet [10] comprises a set of scripts based on the R-package arealDB [11] that harmonizes areal data, such as crop production statistics, from disparate sources. Script-based processing in the FEWS NET and LUCKINet pipelines ensures re-use by others and easy-to-implement updating when new data are added. This streamlines data collection efforts and reduces the effort required to download, standardize, harmonize, and submit data to a centralized database.



5. A call to collective action on agricultural production data

We see this effort as a new social model within Earth system science communities to serve as both a polestar—to support sub-national data gathering efforts, foster cooperation, and eliminate territoriality—and as a wellspring—to provide a central hub of equitable and transparent data access and to democratize all aspects of data gathering and sharing across the food systems research community. Here we advance the collaborative vision of having all existing data (data tables, metadata, and associated polygon files) on agricultural production activities in subnational administrative units (e.g. agricultural statistics from censuses/surveys or their immediate derivatives) comprehensively mobilized, adequately managed, harmonized, integrated, quality-assured, up-to-date, FAIR, and openly and freely accessible to all as a global public good. Progressing towards this vision will require considerable community engagement at all stages of the data life cycle (figure 2). The process of identifying and formatting crop and livestock production statistics for submission to a centralized database is best undertaken by those who are most familiar with the production systems and reporting structures of each country. Likewise, collating and digitizing subnational administrative boundaries is a time-consuming process requiring both technical and subject-matter expertise. Finally, developing methods to standardize and harmonize extant production statistics to produce usable time series datasets is still a nascent science, given the discontinuities and uncertainties inherent in the data. Given the central importance of these statistics and the meaningful progress to

date—as well as the substantial remaining effort—in addressing these key challenges, we openly welcome collaboration across the community and encourage you to contact the author team on pathways of engagement.

ORCID iDs

Kyle Frankel Davis <https://orcid.org/0000-0003-4504-1407>

Weston Anderson <https://orcid.org/0000-0003-3755-9943>

Steffen Ehrmann <https://orcid.org/0000-0002-2958-0796>

Rafaela Flach <https://orcid.org/0000-0002-6612-0668>

Carsten Meyer <https://orcid.org/0000-0003-3927-5856>

Jonathan Proctor <https://orcid.org/0000-0001-8053-8828>

Deepak Ray <https://orcid.org/0000-0002-2856-9608>

Liangzhi You <https://orcid.org/0000-0001-7930-8814>

Michael Foley <https://orcid.org/0000-0002-6747-2745>

Andrew Hultgren <https://orcid.org/0000-0002-6720-6252>

Peter Huybers <https://orcid.org/0000-0002-3734-8145>

Endalkachew Kebede <https://orcid.org/0000-0002-5212-7340>

Caro Park <https://orcid.org/0009-0002-9873-9200>

Donghoon Lee <https://orcid.org/0000-0001-5438-903X>

References

- [1] Kebede E *et al* 2024 Assessing and addressing the global state of food production data scarcity *Nat. Rev. Earth Environ.* **5** 295–311
- [2] Lee D *et al* 2025 HarvestStat Africa — harmonized subnational crop statistics for sub-Saharan Africa *Sci. Data* accepted (<https://doi.org/10.1038/s41597-025-05001-z>)
- [3] Djanibekov N and Petrick M 2020 Statistical data on farming structures, land use and agricultural output for provinces in five Central Asian countries 1992–2015 IAMO—Leibniz institute of agricultural development in transition economies Data File Version vol 1
- [4] Regional Strategic Analysis and Knowledge Support System (ReSAKSS) 2025 ReSAKSS country Atlas (available at: <https://eatlas.resakss.org/>)
- [5] Food and Agriculture Organization of the United Nations. CountrySTAT database 2025 CountrySTAT (available at: www.fao.org/in-action/countrystat/national-countrystat-sites/en/)
- [6] Food and Agriculture Organization of the United Nations 2025 Agro-MAPS database (available at: <https://gaez.fao.org/pages/agromaps>)
- [7] European Union. Eurostat database 2025 Eurostat (available at: <https://ec.europa.eu/eurostat/web/main/data/database>)
- [8] German Space Agency. Eurocrops database 2025 Eurocrops (available at: www.eurocrops.tum.de/index.html)
- [9] Wilkinson M D *et al* 2016 The FAIR guiding principles for scientific data management and stewardship *Sci. Data* **3** 160018
- [10] Ehrmann S 2025 Luckinet—build census database (Version 0.7.0) Computer software available at: https://github.com/luckinet/mdl_build_census_database
- [11] Ehrmann S, Seppelt R and Meyer C 2020 Harmonise and integrate heterogeneous areal data with the R package arealDB *Environ. Model. Softw.* **133** 104799