Title: Assessing and addressing the global state of food production data scarcity

Authors: Endalkachew Abebe Kebede[1]†, Hanan Abou Ali[1], Tyler Clavelle[2], Halley E. Froehlich[3,4], Jessica A. Gephart[5], Sarah Hartman[6], Mario Herrero[7], Hannah Kerner[8], Piyush Mehta[1], Catherine Nakalembe[9], Deepak K Ray[10], Stefan Siebert[11], Philip Thornton[12], Kyle Frankel Davis[1,13]†

Affiliations:
1 – Department of Geography and Spatial Sciences, University of Delaware, Newark, DE, USA
2 - Global Fishing Watch, Washington, D.C., USA
3 – Environmental Studies, University of California, Santa Barbara, CA, USA
4 – Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA, USA
5 – Department of Environmental Science, American University, Washington DC, USA
6 – Department of Environmental Science, Policy, and Management, University of California Berkeley, CA, USA
7 – Department of Global Development, College of Agriculture and Life Sciences, and Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, NY, USA
8 – School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA
9 – Department of Geographical Sciences, University of Maryland at College Park, College Park, MD, USA
10 – Institute on the Environment, University of Minnesota, Saint Paul, MN, USA
11– Department of Crop Sciences, University of Goettingen, Göttingen, Germany
12 – Clim-Eat, Wageningen, the Netherlands
13– Department of Plant and Soil Sciences, University of Delaware, Newark, DE, USA

†corresponding authors: endiabe@udel.edu ; kfdavis@udel.edu

Food production data scarcity is a critical and persisting information challenge preventing many countries from developing sustainable food system interventions. The lack of reliable, regularly collected, accessible, usable, and spatially disaggregated statistics limits an accurate picture of the state of terrestrial and aquatic food cultivation and harvest in many countries and has important implications for outcomes across the food system. Moreover, there is not a comprehensive understanding of where and in what ways food production information is currently lacking. Here we first conduct a comprehensive assessment of national and international food production databases to take stock of the current state of crop, livestock, aquaculture, and fisheries statistics available for each country globally. We find substantial global variation in data timeliness, granularity (both spatially and by food category), and transparency – with challenges concentrated in Central America, the Middle East, and Africa. Based on this, we then describe the technical, institutional, and political challenges that are contributing to conditions of data scarcity and discuss promising technological and policy innovations for most effectively addressing food production data scarcity. Fusing traditional and emerging data gathering techniques with coordinated governance and dedicated long-term financing will be key to overcoming current obstacles to sustained, up-to-date, and accurate food production data collection and is a foundational component of promoting and monitoring progress toward healthier and more sustainable food systems worldwide.

1. Introduction

The production of crops, livestock, and aquatic organisms is one of humanity's most important and extensive activities - providing food and fiber for the world's people and covering more than a third of the planet's land[1,2] and oceans[3]. Because of the vast reach of agriculture - and its profound role in altering Earth systems[4] and influencing human health and well-being[7], food production is a key leverage point for solving sustainability challenges across multiple dimensions and meeting UN Sustainable Development Goal (SDG) targets[5]. The sector plays a central role in determining the extent to which nations can achieve SDG2 – Zero Hunger. In addition, food production exerts important influence over numerous other SDGs through its employment of over a billion people[6] (SDG1 – No Poverty); its large diet-related global burden of disease[7] (SDG3 – Good Health and Well-being); its dominant water footprint[8] (SDG6 – Clean Water); its contribution to bioenergy[9] (SDG7 – Clean Energy); its substantial greenhouse gas emissions (GHGs)[10] (SDG13 – Climate Action); and its extensive modifications of natural systems[11] (SDGs 14+15 – Life Below Water and Life on Land). Given this multitude of linkages with multiple facets of sustainability, a detailed, accurate, and up-to-date understanding of the state of food production activities is an essential foundation for identifying where SDGs are (not) being met and for serving as a baseline upon which solutions can be built, tested, and implemented. Conversely, inaccurate food production information can, at best, lead to ineffective action and, at worst, guide actions at counter-purposes to improve a suite of sustainable development outcomes.

Underpinning much of the national and global action towards achieving these multiple SDGs is food production data (**Figure 1**), the reliability of which can vary widely across products, countries, and years. The primary sources of food production data are large-scale censuses (comprehensive data-gathering efforts meant to occur every 5 or 10 years) or surveys (more frequent and less intensive sampling), with complementary remote sensing efforts being employed in certain countries. From a perspective of equitable global development, each nation would ideally have the resources to fund, implement, and execute these comprehensive data collection efforts, to develop robust sampling strategies, to collate, standardize, and store collected data, and to make the final data available to support development, investment, and research efforts. Yet it is clear that any one (or several) of these steps can impede the ultimate provision, reporting, and publication of official food production statistics. The inaccurate, incomplete, or delayed reporting of these data can lead to a distorted understanding of food production patterns and productivity and could contribute to misinformed and poorly targeted interventions. As just one example, recent work has shown that global gridded agricultural products - which inform the efforts of a suite of global assessments and consortia (e.g., AgMIP[12], ISIMIP[13], GEOGLAM[14], CGIAR[15]) - are highly sensitive to the level of disaggregation of underlying food production statistics[16–18]. Yet, while these statistics are fundamental in shaping how humanity perceives and responds to sustainability challenges within food production systems, substantial gaps remain in the availability and accessibility of reliable, granular, and current data in many regions. This scarcity and unequal distribution of quality agriculture/food production data underscores the critical importance of identifying where and why such deficiencies exist. Quantifying and examining the root causes of insufficient agriculture/food production data is key to promoting evidence-based understanding and decision-making for sustainable food systems worldwide.

In this perspective, we take stock of the current state of global agriculture/food production data scarcity, defined here as being insufficient in terms of spatial disaggregation (detail), timeliness (recency, temporal coverage, and resolution), food item specificity, and accessibility. To evaluate this, we first quantify the current state of and recent trends in food production data availability country-by-country for the production of crops, livestock, and aquatic organisms. We then examine key technical, institutional, and policy obstacles hindering the collection and dissemination of food production data globally. We conclude by highlighting promising pathways forward for improving global food production data availability and quality. Supporting concerted and creative efforts to address these hotspots of food production data scarcity is critical for enabling holistic progress towards achieving multiple dimensions of sustainable development.

2. Current state of global food production data

Accurate food production information underpins the ability of nations to progress towards achieving the SDGs and other development goals[19]. However, the level of accuracy often varies widely between countries, food items, and years. Across crops, livestock, and aquatic production, our analysis reveals both common traits and distinct challenges in the global data landscape. Identifying these data blind spots in production statistics is an essential first step towards comprehensive and up-to-date data coverage on global food production. By evaluating the current state of data availability and deficiencies across different food sectors, we can begin targeting efforts to fill critical data gaps.

2.1. Crop production

Understanding the patterns, dynamics, and growth potential of crop production – particularly in regions of concern – is critical for assessing progress towards achieving the United Nations (UN) SDG 2 by 2030[20]. Information on the location, timing, and productivity of crop production is important for various applications, including yield forecasting, land use planning, and environmental impact assessment, among many others. Yet there remains a widely varied understanding on global patterns of crop harvest and productivity **(Figure 2).** This stems largely from reliance on census-based survey data to quantify cultivated extent and productivity. Surveys employ varied sampling methodologies and resources, constraining standardization. As surveys will likely continue serving as the primary data source on crops, targeted improvements to survey design, analysis, and data-sharing practices are essential to fill key gaps.

National government agencies (e.g., ministries of agriculture) are typically responsible for collecting, processing, and disseminating agricultural data within their countries[21], and many are working to update their methods of collecting and standardizing the agricultural statistics systems. To better standardize these efforts across countries, the Food and Agriculture Organization (FAO) of the United Nations coordinates the World Program for the Census of Agriculture (WCA), which provides agricultural census guidelines for different countries and reviews their practices[22]. FAO compiles the national agricultural census from each country and makes it publicly available through the FAOSTAT[23] database, which provides open-source agricultural data from 1961 and onwards[24]. Similarly, EUROSTAT[25], the European Union (EU) statistical organization, provides a wide range of socio-economic and environmental data for member countries of the EU through its open data portal[26]. While substantial progress has been made in gathering and sharing agricultural data through these and other national and international efforts, census methodologies and dissemination vary across different countries[24,27]. This stems from differences in resources, the importance of agriculture, data needs, and the agreements between countries, FAO, and EUROSTAT. For example, EUROSTAT can only report data for which the EU has an agreement with the member states. Thus, even if a member state collects far more detailed data, it may not enter into the database as per an agreement with the EUROSTAT agency. The FAO is similarly bound by agreements with individual countries, on whose reports they rely. Partly because of this, the categorical and spatial variation in crop area and yield in different agroecological zones is poorly captured in FAOSTAT and other international and global datasets compared to national census portals.

Crop calendars are an integral component of current and future solutions to agricultural data scarcity. Derived from censuses, models, and remote sensing applications[28], they define the dates for different stages of crop cultivation, including planting and harvest. Among other uses, crop calendars are mainly utilized when monitoring crop conditions, forecasting and estimating crop yields, and monitoring crop conditions[29]. Existing crop calendars with global coverage include those produced by Group on Earth Observations' Global Agricultural Monitoring (GEOGLAM) Crop Monitor, the United States Department of Agriculture Foreign Agricultural Service (USDA-FAS), the FAO, the European Commission Joint Research Center's Anomaly hot Spots of Agricultural Production (ASAP), Portmann et al.[18] (as part of the MIRCA2000 dataset), and Sacks et al.[30] which are typically provided at the national or subnational level (administrative levels 0 and 1). At this resolution, calendars are unable to capture regional variations at the sub-agroecological zone level that have on-the-ground impacts on cropping dates. To increase spatial detail, researchers have enhanced resolution for key crops (e.g., Becker-Reshef et al.[31]; Kotsuki and Tanaka[32]; Laborte et al.[33]). However, comprehensive higher-resolution crop calendars remain limited to major staples such as rice, soybean, and wheat. Improving the spatial granularity, crop diversity, and harvest date accuracy of published crop calendars can strengthen derived agricultural products, policies, and food aid mobilization.

To assess the current state of global agricultural statistics, we conducted a comprehensive review of each country's latest agricultural census for recency, spatial detail, and transparency (Figure 2). Through this assessment, clear regional patterns emerged, with limitations in regular censuses and data needs prevailing in Central America, the Middle East, and Africa – where nations continue to face limitations in conducting regular censuses and in meeting fundamental agricultural data needs[19,34]. For many countries in these regions, especially in Africa[35,36], our findings align with FAO assessments indicating steady declines since the 1980s in government capacity to conduct censuses and in the quality and quantity of national agricultural statistics reporting.. In countries where agricultural censuses are not available or are not carried out regularly (Figure 2a), these increasingly outdated snapshots of the magnitudes, spatial patterns, and temporal trends of crop production risk mismanaging agricultural resources and misinforming interventions in the pursuit of rural development and food security goals.

We also find that the degree of spatial disaggregation in crop statistics varies widely between countries (Figure 2b), hampering targeted action. Some countries (e.g., United States, Brazil, India, Australia) gather and provide agricultural data at fine spatial scales (i.e., county/district level) and categorical detail (i.e., distinguishing individual crops vs. aggregating in crop groups). But for many countries, publicly available agricultural data is only at coarse administrative levels (i.e., state/province or national). This coarser resolution data can fail to capture the spatial variability of crop production statistics[37], especially for countries dominated by smallholder farms[38]. Even when subnational data exists, the underlying administrative units may change over time due to renaming, splitting, merging, or aggregation. This requires reconciling spatial consistency of statistics through time, a challenge if original unit names change. Further, some developed countries exercise data privacy, restricting access to microdata critical for administrative-level estimates without agreements or payments. If administrative data risks privacy breaches, it may also be suppressed or combined across units.

Finally, we assessed each country's agricultural census transparency using the Findable, Accessible, Interoperable, and Reusable (FAIR)[39] principles (Figure 2c). A detailed description of each criterion and its corresponding score can be found in the supplementary material. Using country-specific information from the FAO's WCA portal on metadata, census reports, questionnaires, and methodological reports, we evaluated transparency based on nine defined criteria for findability, accessibility, interoperability, and reusability of each country's agricultural census (see Figure 2 caption for details on criteria). Encouragingly, most of the countries following WCA guidelines have moderate to full transparency, while a few countries (e.g., Ethiopia, Oman, Yemen, Libya, and Turkey) have overall low transparency of agricultural census reports.

Despite the existing challenges in data availability and transparency for crop production, multiple global and regional efforts have mapped spatial patterns and temporal trends of cropped areas. Perhaps most widely known are several global crop-specific harvested area and yield datasets that have been developed by combining census statistics with remote sensing data[18,40–42]. These include the most comprehensive global gridded datasets on harvested area and yields for 175 crops[41] and monthly irrigated and rainfed cropped areas (MIRCA) for 26 crop classes[18]. However, these and other global datasets are centered on the year 2000 and are becoming increasingly outdated. Yet despite the dynamic nature of crop production patterns, most recent agricultural and environmental studies still use these aging datasets due to limited or the lack of alternatives[43–45]. Updated and current, time-varying, and spatially detailed information of cropped areas and yield is urgently needed to support targeted and informed decision-making.

Some ongoing efforts – including the GAEZ[46] (Global Agroecological Zones) and SPAM[16] (Spatial Production Allocation Model) datasets – are attempting updates but face constraints from underlying statistics. Emerging remote sensing datasets also attempt to provide updated global cropland extents[47–49] at fine spatial resolutions. However, accurately estimating actual cropland area and distinguishing crop types remains challenging. Spectral and temporal similarities between cropland and grassland often cause poor cropland identification, especially in less intensified regions such as Africa[34,50]. The substantial resources required to support ground-truth data collection as well as computational needs are also considerable constraints on purely remote sensing approaches. At present, a fusion of survey-based census data, modelling, and remote sensing may offer the most promise for resolving the challenges of comprehensive global crop mapping.

2.2. Livestock production

A wide range of data on livestock populations, distributions, and production appears readily available (**Table 1)** (e.g., Livestock Data for Decisions: https://www.livestockdata.org/type/datasets). However, the livestock data landscape is far more complex than this apparent widespread availability of data products would suggest. For domesticated livestock species, country-level data on animal numbers and production levels are accessible in FAOSTAT, compiled from a wide range of sources including national censuses, surveys, and estimation procedures. FAOSTAT data offer valuable comparability across and between countries and regions, near-global coverage, and an annual time series dating back to 1961 for most variables. As such, national FAOSTAT data on livestock production have been (and will continue to be) used in innumerable analyses where data comparability, broad or global coverage, and temporal trends are deemed to be important.

While useful, FAOSTAT data has limitations. Country-level data can mask substantial subnational heterogeneity and rapid local changes between infrequent national surveys. This is increasingly the situation for rapidly expanding research and policy applications, for example, in development research, animal health, economics, and environmental adaptation, and mitigation science[51]. For more spatially explicit research, the Gridded Livestock of the World[52,53] (GLW) dataset is the global standard, mapping populations of cattle, buffaloes, horses, sheep, goats, pigs, chickens, and ducks for the years 2010 and 2015.

GLW is based on national census data downscaled to a spatial resolution of 5 arcminutes or about 10 km at the equator and allocated spatially using a set of suitability layers and other spatial predictors.[51] Beyond GLW, there are few other livestock mapping efforts, excluding highly localized studies. The analytical methods involved in the current and previous versions of GLW are described in the literature[54,55]. National livestock census data are key to efforts such as GLW, and while such data are available for many countries, their quality, resolution, and timeliness are highly variable, and considerable efforts have to be expended on collation, harmonization, and standardization before they can be used[55]. The date of census data collection is also highly variable: the census data in GLW version 4 ranges from the early 1990s to 2019, with all data at the pixel level being harmonized to the national level FAOSTAT data for the years 2010 (GLW3) and 2015 (GLW4)[53].

Other widely used global livestock datasets include the Global Livestock Production Systems (GLPS), the Global Environmental Assessment Model (GLEAM) (Opio et al. 2013; MacLeod et al. 2013; FAO 2022) and the Herrero et al. dataset on livestock biomass use, production, feed efficiencies, and greenhouse gas emissions[56,57]. Developed by Kruska et al.[58] and expanded by Robinson et al.[57], GLPS classifies livestock systems into 11 to 14 types mapped using proxies from the non-spatial livestock classification scheme of Seré and Steinfeld[59]. A limitation of GLPS is the lack of detail on mixed crop-livestock systems, partly due to inconsistencies across crop and livestock datasets. Unlike the GLPS, the farming systems mapping by Dixon et al.[60] later updated for Africa, is not derivable from spatial data. The Herrero et al.[56] dataset harmonizes livestock populations and milk and meat production data with year 2005 and 2010 FAOSTAT data and spatially downscales biomass use and GHG emissions using plausible feed rations. This dataset is a unique resource and is currently being updated to the year 2020. Beyond these datasets, few (if any) alternatives exist for comparative global or regional studies. The GLEAM model uses a similar workflow but different methods and resolution, and it does not attempt to harmonize all FAOSTAT statistics. This tool is mainly designed to assist countries in their preparation of Nationally Determined Contributions. Efforts to harmonize inputs like animal body weights and feed rations between the two models and datasets are underway.

In addition to national censuses, household survey datasets such as the Living Standards Measurement Studies (LSMS) and Rural Household Multi-Indicator Surveys[61] (RHoMIS) provide livestock information. The recent rounds of LSMS contain a comprehensive livestock module and enable a range of analyses of livestock's contribution to livelihoods[62]. RHoMIS uses a modular approach to household data collection, with modules for various agricultural activities (including livestock), and contains data for around 45,000 households across 36 countries. While useful for studies not requiring complete coverage, sampling designs can limit spatial analysis[51].

Major livestock data gaps remain, especially in lower- and middle-income countries. Ruminant diets comprise diverse feed resources - grasslands, crop residues, supplements, fodder - which have been understood through surveys, but incomplete coverage hinders many types of analyses. Gaps also exist concerning the number and distribution of different animal breeds - with GLW3 even inferring dairy and dual-purpose (milk-meat) production from other data, and there remains a persistent lack of detailed distribution datasets of even broad classes of livestock, such as cattle. Promising opportunities to fill gaps include digital data collection in real time - via mobile and social media and sensor technology - as well as the use of crowdsourcing and other participatory methods[63,64,65], in combination with remote sensing and AI tools[66].

2.3. Aquatic food production

Fish and other aquatic foods represent a highly diverse (and often understudied) food sector, comprising over 2,600 species and species groups caught and farmed in marine, brackish, and inland environments. Across this diversity, small-scale to industrial producers employ a wide range of fishing and farming methods. As a result, monitoring production across the sector requires compiling information from a wide range of actors and governmental agencies. The Coordinating Working Party (CWP) on Fishery Statistics provides standardized definitions, methods, and minimum requirements for reporting fisheries and aquaculture statistics at a global scale. Member organizations of CWP on Fisheries Statistics mainly report data through the STATLANT system of questionnaires, with data generally collected through national and regional census-based and sample-based schemes[67]. FAO reports or calculates estimates for country-level production: "When data are not reported or only partially reported, FAO implements estimates based on the best information available from alternative sources, including those from [regional fishery bodies] in the case of capture fisheries"[68]. The FAO freely provides global fishery and aquaculture data through bulk downloads and the FishStatJ computer application and summarizes this data in the biannual State of Fisheries and Aquaculture report produced by the Committee on Fisheries. Although there has been great progress in the data available on aquatic food production, there are still significant data gaps for both capture fisheries and aquaculture.

Capture fishery production data are reported for 2,647 species/species groups by marine, brackish, and inland habitat type (FAO FishstatJ). Although national statistics generally include finer resolution of where fish are caught, FAO statistics are reported as a catch within one of 19 major fishing regions. The limited geographic resolution of catch obscures the extent of distant water fishing operations – or harvest occurring outside the fishing country's waters –as fishery production is generally attributed to the fishing vessel flag state irrespective of where the fishing occurs[67]. Other efforts to spatialize catch include those which build on FAO statistics, such as Watson[69] and Zeller et al.[70], and those based on remotely detecting fishing activity, such as Kroodsma et al.[71]. where remote detection stems from thousands of fishing vessels continuously broadcasting their GPS position and identity via the Automatic Identification System (AIS) or national Vessel Monitoring Systems (VMS) on a daily basis.

Fishery catch statistics are generally reported as nominal catch, which represents the live weight equivalent of landed catch. Nominal catch aims to represent the contribution of fisheries to the economy and provision of food. It, therefore, does not include organisms caught and discarded, catch utilized prior to landing (for example, consumed by the crew or used as bait), or landings that are rejected or dumped[67]. A notable difference in the collection of catch statistics compared to other food subsectors is that catch data is a critical input for stock assessments, which inform management for many industrial fisheries and creates an incentive for collecting quality production data that is unique to fisheries. Catch data collection methods vary across industrial and small-scale fisheries, leading to differences in the comprehensiveness of catch data. Operators in industrial and semi-industrial fisheries often report collected data to a fishing authority as part of licensing and reporting requirements, which forms the basis of census-based schemes[67]. Although industrial fisheries are responsible for much of the catch, the majority of fishers are engaged in small-scale and artisanal activities, resulting in geographically dispersed catch often governed by local communities. Consequently, estimates of catch by small-scale and industrial producers are often survey-based, with uncertainty around the degree of coverage for this subsector[67]. An additional fishery catch data gap is Illegal, Unregulated and Unreported (IUU) catch, which by their nature are poorly captured by official statistics. Reconstructions of marine catch data estimate that global catch is 50% higher than reported in FAO[72], while comparisons with household surveys indicate inland catch is approximately 65% higher than reported[73]. It should be noted that FAO global values fall within the uncertainty of reconstructions[74] . FAO also engages in addressing IUU and smaller scale catch in other ways, including the Global Record of Fishing Vessels, Refrigerated Transport Vessels and Supply Vessels, development of voluntary guidelines. In addition to catch weight, catch value is often important information, but it is not currently included in international statistics[75], like it is for aquaculture.

Similar to wild capture fisheries, data on aquaculture production are largely reported in live or wet weight, with 652 reported species units in FAO statistics[75]. As of 2020, aquaculture accounted for over half of all aquatic food production (seaweeds included), but the data from 70% of aquaculture-producing countries consisted entirely of FAO estimated species production (i.e., data are not reported or only partially reported by the producing country, requiring alternative sources); this is compared to only 22% of estimated values for wild capture, suggesting a higher level of uncertainty in the aquaculture data. Further, the amount and number of organisms farmed are likely underestimated due to data limitations[75]. Notably, aquaculture in a given country usually does not have a government-issued, standalone regional entity and instead falls into agriculture and/or a fishery agency or body, which can create data gaps and errors.

Aquaculture shares attributes (and resources) of agriculture and wild capture fisheries, including where and how it is produced. Aquaculture can operate on private property (for example, freshwater) or common-use areas (for example, oceans). Although not unique to aquaculture, there are also a myriad of ways to grow different organisms that vary between countries, farms, and species, including sourcing seed from the wild (i.e., capture-based aquaculture), using different technologies (for example, recirculating systems vs open pens), and levels of intensity (for example, extensive vs intensive) (FAO Aquaculture Methods, Klinger et al.[76]). While data on production practices is a collective challenge faced by all sectors, where it occurs can introduce unique problems because data collection tends to be a function of a region's regulatory requirements of reporting. At the most basic level, differences in what is defined as 'aquaculture' (note, FAO has a standard definition, but it does not necessarily match the regional reporting body) or what units are reported (for example, pieces vs bushels) can differ from one agency to the next, introducing data issues[77] (for example, Froehlich et al.[77]). As a result, some core measures reported in other sectors are absent from global aquaculture, in particular yield.

Yield provides a unifying measure of scale and productivity over time, but there is a dearth of information for aquaculture compared to agriculture. One major reason for a lack of yield information is likely the result of little to no information concerning the spatial location and extent of existing farms. There is currently no global map of freshwater aquaculture, and a map of most marine aquaculture around the world (excluding seaweeds) was just recently published[78], but does not account for changes over time, nor is it able to discern active sites (versus pre-leased or fallowed). Combining production values with spatial estimates from reported (agencies or farmers) or observational sources (e.g., remote sensing[79]) can help fill this gap going forward, especially in areas with high densities of aquatic farms. However, other factors such as feed, feed conversion ratio, and grow-out mortality all influence yield and a broader understanding of sustainability but remain extremely heterogeneous and underreported.

FAO fishery and aquaculture production data serves as the backbone for countless peer-reviewed papers, reports, and databases. For example, FAO marine fishery data underpins catch reconstructions in the Sea Around Us database[80]. Many limitations of fisheries and aquaculture production data noted above are related to issues with national reporting that are beyond FAO control. Nevertheless, global fishery and aquaculture data provision can be improved through more detailed metadata, particularly as it relates to data provenance, transparency in assumptions and uncertainty, and documenting changes to data through release notes. Tracing data back to its origins is critical for understanding assumptions in the data collection, detecting errors in the data, and linking data with other national data sources. Thus, relaying the data provenance in the metadata is a critical first step. In any data management and modeling exercise, there are numerous sources of assumptions and uncertainty that are important for appropriately interpreting data. In the case of fisheries and aquaculture data, more detailed flags on value estimation and reporting applied live weight conversion factors would improve transparency of assumptions, while reporting measures of uncertainty is important for users to capture uncertainty within their own applications of the data. Finally, as the data is regularly improved over time, it would be valuable for legacy data and data release notes detailing the changes to be maintained in a visible location. This would facilitate reproducibility of analyses based on older versions of the data and would improve communication of the changes and improvements to the database itself.

3. Challenges of global food production data

The current state of food production data scarcity is characterized by substantial variation in quality and detail across countries, time, and food products. While there are a growing number of publicly accessible agricultural data sources – largely due to a combination of increased global participation in cross-disciplinary research and development of agricultural and information technologies, there remain manifold technical, institutional, and policy barriers to the increased collection, dissemination, and use of agricultural data globally.

3.1. Technical challenges

Food production data is prone to quality issues such as sampling, processing, and coverage errors, which can significantly undermine the credibility of census reports. Beyond simply creating these datasets, a major challenge for their downstream use is the statistical sampling of the datasets.

Crops: Many publicly accessible agricultural datasets are collected using convenience or opportunistic sampling (or in the case of farmer surveys, from whomever responds). This results in datasets that are biased, poor quality, and not representative of the data population[63]. Not enough attention has been given to addressing this issue, with more emphasis on retroactive adjustment than anticipatory choice in the early enumeration stage[63]. For map-relevant and well-sampled datasets that can be used for proper accuracy assessment and agricultural statistics (for example, production or area), sampling techniques such as simple random sampling, stratified random sampling, or systematic sampling must be employed[96,97]. In addition, datasets should be documented with as much detail as possible about the data collection procedures, choices made, expertise of annotators, any quality assessment and control (QA/QC) performed, and other important information influencing data quality and interpretation. For geospatial and remote sensing datasets in particular, large globally distributed datasets of cropland have resulted from land cover and land use mapping initiatives, which typically include cropland as a class. While the presence of cropland (broadly defined as land used for growing crops) can in most cases be determined from inspecting high-resolution remote sensing images, there is still much inconsistency around what constitutes "cropland" across datasets. For example, some datasets define cropland to include tree crops like palm or coffee, while others do not[98]. These challenges are far greater for data collection with more fine-grained categories than the presence or absence of cropland, such as crop type, cultivation practices (e.g., tillage, cover cropping, irrigation, etc.), nitrogen or other input use, livestock stocking rates or pasture management, pests or disease, and fallow status. Annotators must visit the data locations in situ during the relevant time of the growing season to "ground truth" the observed category. For some categories – for example, crop type in intercropped fields, level of tillage, or cover crop variety—more detailed annotation is needed to effectively use the data in downstream applications. Collection of field-scale yield data is particularly challenging, very expensive, and error prone[63,99]. In some cases, such data are recorded in some form by farmers, farming equipment, or equipment companies, but these data are not typically available to the public or research community.

Livestock: Livestock presents a conundrum in the agricultural development space; despite livestock's multi-dimensional importance to the livelihoods of at least 1.3 billion people globally, the critical role of the sector has never been reflected in development assistance, research outputs or, indeed, the data landscape. From a data perspective, for instance, data collection approaches in livestock systems have not developed substantially over the last 40 years or so (Carletto et al.[63]), although the methods of data collection have certainly evolved (high-resolution remote sensing, drones, tablets, and smartphones). Even where methods have evolved, as noted above with respect to crops, higher-resolution remote sensing methods for helping to estimate livestock populations and gather data on many livestock-related management variables, for example, still face considerable challenges. These include the following. First, defining the nature and extent of grazing systems and the complexity of associated land cover (including pasture, browse, bare land, and all gradations in between) present considerable difficulties. The challenges of separating land use from land cover leads to a broad range of estimates of the extent of grazing systems locally and globally[100]. Second, in many lower- and middle-income countries, mixed crop-livestock systems predominate. These systems involve crops and livestock occupying the same or adjacent areas, and globally, 70% of farms are less than 1 hectare in size[101], further complicating the robust characterization of the livestock and crop components of small-field mixed systems. Third, unlike with crops, transboundary issues are particularly relevant for pastoral systems (nomadic, transhumant, agro-pastoral). Although there are challenges in assembling meaningful national data on livestock populations in many countries, databases on animal numbers, locations and movements are essential for preparing for, managing, and mitigating the risks of certain transboundary diseases that may have high potential impacts, such as foot-and-mouth disease[102]. Fourth, even where data do exist, there may be complex issues concerning legitimacy and accessibility in many situations, highlighting the socio-cultural challenges and power asymmetries that can militate against effective data sharing[103].

Fisheries and Aquaculture: Accurate and consistent fisheries-dependent data are required to sustainably manage fisheries of all scales[104]. However, the wide ranging, mobile, and relatively invisible nature of fisheries make such data collection challenging, time-intensive, and costly. Numerous stakeholders can be involved in data collection, and data is often recorded using paper-based logs and/or on-board observer programs that suffer from very low fleet coverage[105]. The huge variety of harvested marine species makes accurate species reporting a challenge. While FAO capture fishery data include 2,647 categories, many of the largest categories by volume are highly aggregated. In 2020, 10.5 million tonnes, or ~13% of global production, was categorized as "marine fishes not elsewhere included"[75]. Electronic monitoring (EM) is emerging as a promising strategy for comprehensive catch monitoring, including bycatch and discards. However, a review of over 100 EM trials and programs found challenges with data quality, storage, and transmission as well as overall system failure and prohibitively long data review times[106].

Vessel tracking systems have proven useful for detecting and characterizing industrial fishing activity[71,107]. Use of vessel tracking systems varies by region and fleet, but are generally biased toward larger vessels (>24m) from upper/middle income countries that have adopted AIS measures stricter than the Convention for the Safety of Life at Sea (SOLAS), the international regulation governing AIS, which explicitly exempts fishing vessels. Only around 2% of the world's 2.8 million fishing vessels and less than 0.4% of vessels under 12 meters broadcast their position over AIS[108]. AIS devices can also be manipulated or turned off, often without penalty, obscuring fishing activity and potential IUU catch[109]. Additionally, not all AIS messages that are broadcast are recorded due to variable terrestrial coverage and satellite reception. In areas with high vessel density – such as the South China Sea, Mediterranean Sea, and Gulf of Mexico – AIS messages interfere with one another, preventing them from being recorded by satellites. VMS systems, which generally carry strict penalties for tampering, are proprietary and data are rarely shared publicly in usable formats. Additionally, there is no standard format for VMS data, and efforts to merge data from multiple sources face challenges associated with different broadcast intervals, schemas, metadata, and units.

Satellite imagery, while useful for large-scale detection of fishing vessels[110,111] and aquaculture farms[112], have several limitations as well as technical challenges specific to marine applications. Orbital mechanics and satellite reception result in variable spatial and temporal coverage, as most public earth observation satellites, including the important Landsat and Copernicus missions, have multi-day revisit frequencies and do not image the open ocean. Small-scale vessels are also less likely to be detectable in these imagery collections due to insufficient pixel resolution, and suitable high resolution imagery (<1m) from commercial providers like Maxar (https://www.maxar.com/) and Planet Labs (https://www.planet.com/) may be cost prohibitive at even moderate spatial scales. Synthetic Aperture Radar (SAR) is a proven method for detecting vessels at sea[111] and multiple forms of aquaculture[112,113]. However, the complexities of SAR images complicate the ability to discern vessel characteristics, such as gear type, and radar signals are reflected by the water's surface, limiting their utility for sub-surface aquaculture detection. Optical imagery, particularly at high resolution, offers increased potential for detecting and classifying at sea vessels and aquaculture but is limited by weather (e.g. clouds) and daylight. Yet, infrared imaging radiometer suite (VIIRS) day and night band (DNB) optical remote sensing images can be an effective source of information capturing vessel lights, especially fisheries that use lights as a harvest strategy (e.g., squid)[114,115]. Though the most effective, but computationally intensive approach is the combined use of a variety of observational sources (e.g., ref[116]).However, remote observations are not direct measures of production and cannot provide information on species composition unless paired with additional data, such as from logbooks.

A further technical challenge for fishery and aquaculture data is that there is no universally accepted distinction between small-scale and industrial fisheries. Definitions of the sectors, based on a range of characteristics, vary across countries, resulting in inconsistent inclusion of small-scale production in national reporting requirements[116,117]. Modern fisheries management, which developed largely in response to industrial fishing, has often deprioritized small-scale producers in data collection efforts and exempted them from self-reporting. Similarly, there is no uniform definition of small-scale aquaculture, as production methods and scales differ considerably across regions, and aquaculture development has often proceeded ad hoc in the absence of clearly defined property rights in the ocean. Efforts at defining small-scale aquatic production are complicated by its distributed nature, coupled with limited budgets and capacity for monitoring and reporting[118].

3.2. Institutional and policy challenges

Multiple institutional challenges obstruct comprehensive agricultural data collection and curation. A major challenge related to data collection is the lack of consistency and duplication of effort between various agencies and organizations, including government, research institutions, and INGOs. In some countries, there is also ambiguity about an institutional mandate for collecting and disseminating agricultural data[119]. Data collection efforts are often siloed within individual departments or institutions, and even within the same broader institution (for example, federal government), leading to a lack of coordination and sharing. For example, fisheries and aquaculture are often managed by different agencies or ministries and often not the ministry of agriculture. This creates potential reporting gaps and mismatches in the available information. Beyond this, institutions that are (or could potentially be) in charge of collecting data (for example, agricultural ministries) also often lack the capacity and expertise to adopt innovative methods for data collection or data provisioning, particularly in low-income countries[63]. Additionally, sharing of food production data between countries and/or organizations is often challenging, worsened by inconsistent data privacy protocols and platforms [120]. The lack of coordination among national and international organizations results in poorly harmonized agricultural data sources[19]. While some data collection efforts are integrated with the data users (for example, agricultural statistics agencies), in many cases, the data collector and the data scientist are also separated, leading to a gap in needs from both sides. There is a need for more cross-institution and cross-disciplinary collaborations throughout the agricultural data life cycle to address these gaps. Such collaborations could also increase the likelihood that a dataset is hosted and maintained over long periods of time, for example, beyond the duration of a funded project at a particular institution, and that the dataset is provided in formats usable by a wide range of users (i.e., interoperable).

Similarly, policy silos exist within and across countries and institutions, both in the public and private sectors. This makes it difficult to aggregate data and derive meaningful insights even within one country [121]. Policy silos also impact which data are prioritized for collection (and which are overlooked)[122] and contribute to issues related to data privacy and sharing. Different institutions have varying policies around data privacy and sharing, leading to inconsistencies in how data are collected, stored, and shared. The lack of coordination and inconsistencies in data collection, management, and analysis lead to a fragmented data landscape. For instance, some institutions may collect farmers' personal information without their consent, while others may not collect this information at all. In addition, geo-referencing data is crucial for satellite data analysis, but a lot of critical data is collected without geo-referencing, limiting their utility[122]. Moreover, some institutions do not share data with other stakeholders due to concerns about data privacy and security, while others may be more open to data sharing. This further exacerbates the problem of fragmented data, making it difficult to derive comprehensive insights into agricultural production and food security[123,124].

Perhaps the most critical obstacle hindering systematic agricultural data collection by government organizations – particularly in Sub-Saharan Africa – is the lack of steady and sufficient public funding. Without dedicated long-term financing to develop centralized data infrastructure and standards, data collection efforts remain siloed, sporadic, and disjointed across various agencies. This severely impedes the assembly of high-quality, consistent datasets needed to inform policies and interventions. While there is a growing recognition of the importance of agricultural data, many policies are not implemented due to significant funding gaps. This often hinders the implementation of mandates on data collection, management, and sharing particularly in developing countries. It also limits the adoption of new technologies, such as remote sensing and Artificial Intelligence (AI) /Machine Learning (ML), which augment traditional monitoring and assessments. Overcoming funding inconsistencies is the foremost challenge that must be addressed to improve inter-institutional coordination, reduce duplicated efforts, build capacity, and establish sustainable data systems. Concerted efforts are needed to secure enduring financial support that enables a more integrated, effective approach to agricultural data gathering across Sub-Saharan Africa. No other intervention would be as broadly impactful in overcoming current data deficiencies as establishing the consistent financing required for strengthened systems and collaboration across governmental bodies.

Agricultural research and development organizations accumulate vast amounts of data annually from numerous farmer surveys and field trials. However, despite the substantial efforts and costs involved in collecting these data and the value of these data for research re-use, minimal infrastructure currently exists to systematically document, share, and curate this information. For example, only a small fraction of the data gathered are readily FAIR (findable, accessible, interoperable, and reusable) by the wider research community. While some of the challenges are institutional, such as limited open access and interoperability[124], targeted improvements in data governance could help overcome these hurdles. Specifically, coordinated development and application of data documentation, dissemination, and formatting standards, alongside raising awareness of FAIR principles, could greatly enhance preservation and utilization of survey data assets. Initiatives are needed across local, national, and international levels to implement improved governance that simultaneously establishes community data sharing norms and builds capacity for proper data curation. By tackling obstacles in unison through governance that advances standards and education, the vast potential of accumulated survey data to inform agricultural research could be more fully realized.

Private sector investments can lead to improved productivity and efficiency, including investments in data collection that benefit both the public and private sector[125]. However, large volumes of data and insights increasing in the hands of the private sector - with no clear policies and regulations around who can collect what data and when - can disadvantage farmers in addition to exposing them to exploitation[126] (often smallholders and rural communities). Moreover, while privately collected and held data might have valuable insights and information, companies rarely share data publicly. This creates further challenges in ensuring access to relevant and accurate data, limiting the development of comprehensive policies and strategies. Further, contextually irrelevant policies that do not consider the characteristics of the target population often lead to ineffectiveness, unintended consequences, lack of buy-in, waste of resources and inequity[127] .

For livestock, aquaculture, and fisheries, substantial differences in country-specific regulations are ill-equipped to address transboundary issues, such as diseases and climate change, which can significantly affect international trade access and domestic food security[102]. Successful monitoring and control of transboundary diseases is greatly dependent on governmental cooperation, something that can be made considerably more challenging with incompatible or inconsistent surveillance systems and regulatory frameworks in the countries likely to be affected[102]. Furthermore, as Rotz et al.[128] point out, just about anything "digital" comes with critical political economy considerations – in particular, data ownership and control, as well as data security. National agricultural data governance frameworks do not always reflect farmers' concerns and expectations[129]. While aquaculture challenges largely mirror those faced by livestock with respect to transboundary and diseases-related issues, they do perhaps have a higher concern of a risk of escapees and their potential impact on local ecosystems, including wild capture fisheries (e.g., Jensen et al.[130]). Meanwhile, fisheries face additional challenges given that stocks can span multiple jurisdictions, often requiring regional management approaches. However, as many fish stocks shift due to climate change, new entities can gain access to fisheries, requiring renegotiation or creating potential conflict[131].

4. Socio-Technical Levers for Data Abundance

A variety of technological and policy innovations are potentially available or in development that offer new opportunities for overcoming many of the challenges causing food production data scarcity to persist (Figure 3). Technological innovations in digital monitoring can complement more conventional methods of data gathering, management, and sharing. For example, d'Andrimont et al.[132] demonstrated an innovative data gathering approach that used street-level images and machine learning to detect crop phenology status over large areas. However, the development and deployment of these new approaches requires an enabling policy environment with appropriate incentives, regulations, and benefits sharing[133]. As such, bundles of socio-technical innovations - that combine new technologies with coordinated policy - will be necessary for coordinating stakeholder priorities, investments, and multi-scalar governance in transforming the agricultural datascape to the benefit of all[134].

4.1. Technological opportunities

Advanced technologies have great potential to assist data collection, but use of this technology still depends on access to other resources (e.g., electricity) and governance. Such technologies can increase the scale of agricultural datasets (such as, geographic and temporal coverage, number of data points, etc.) in a number of ways. For instance, rapid acquisition and automated analysis of street-level images can be used to efficiently collect samples over a large area at low cost[132,135,136]. Commercial off-the-shelf drones or micro air vehicles are also being used to efficiently collect observations of agricultural areas at low cost[137], though the use of drones in some regions can be complicated by policy or local community restrictions or regulations. Citizen science and crowdsourcing initiatives are making use of online annotation and smartphone technologies to collect large-scale agricultural datasets globally[98,138]. However, special attention must be invested in such projects to review novice annotations and ensure high quality[98]. The growth of agricultural technology (agtech) companies also presents opportunities for partnerships that leverage data collected by companies for commercial purposes to be used for scientific research[139,140]. These innovative strategies can provide more comprehensive and diverse datasets than data collection efforts that rely on traditional data collection mechanisms such as in situ or farmer surveys. These technologies can also provide more accurate and precise measurements using high quality (yet low cost) sensors such as in situ sensors or smartphone GPS, compared to qualitative surveys or farmer recall.

Satellite remote sensing data can also be used to fill data gaps in the status and monitoring of agricultural variables; however, its utility relies heavily on high quality ground-truth data for model calibration and validation. High quality in situ datasets detailing the time and location that a particular agricultural category or quantity was observed can be paired with myriad Earth observation datasets that provide environmental and biological covariates for downstream analyses. ML technologies can be used with satellite remote sensing data to build correlative models that predict agricultural characteristics like crop type, yield, or cultivation practices from Earth observations (such as reflectance in certain wavelengths, precipitation, and temperature)[141–143]. Yet, the increasingly resolved information that these technological advances promise also poses substantial challenges in terms of data privacy and ethics. In this regard, emerging technologies such as blockchain and federated learning have the potential to address privacy concerns related to agricultural datasets (for example, anonymization) while allowing such datasets to be made available for public research. Blockchain technology can be used to create a secure and transparent system for sharing data while maintaining data privacy through anonymization and ensuring only authorized parties have access to data[144,145]. Federated learning allows for machine learning models to be trained on distributed datasets without the need for data to be centralized in a single location. This allows individual farmers or institutions to keep their data local and have control over its access and use[146]. Federated learning can also help to mitigate concerns related to data bias by ensuring that models are trained on a diverse range of datasets. Additional techniques for the anonymization of geo-referenced data (such as anonymized spatial coordinates and sample displacement) are being increasingly promoted[147].

Many of the technologies outlined above in relation to crop data can also be applied to livestock data, including new and improved methods of data acquisition and automated data analysis for determining livestock populations and characteristics such as species and breed. Two more livestock-specific innovations with substantial potential for providing data of great utility are animal-based methane measurement (in ruminants) and precision livestock farming. Several non-invasive methods are available to measure methane production in animals. These include microbial biomarkers in the rumen that if heritable could be used for targeting purposes, GHG emission monitoring systems using hand-held methane sensors, and ingestible methane detection capsules and other sensors that allow continuous monitoring with WiFi[148–150]. All such methods currently have some disadvantages such as cost, reliability and/or reproducibility. Wider uptake of these innovations is likely to depend to some extent on the development of appropriate validation, calibration, and standardization protocols. Precision livestock farming is a set of innovations based on the application of process engineering principles and techniques to livestock feeding to automatically monitor, model, and manage animal feeding at the individual level. The idea is not new and has been used to maximize margins for intensive livestock production for many years. However, it is developing rapidly to encompass a wide range of new monitoring and sensor technologies (the Internet of Things) and their application to the major domesticated livestock species[151–153]. For both animal-based methane measurement and precision livestock farming, the future issues around data privacy and ownership are likely to be just as challenging as for other agricultural data.

In fisheries, together with advances in big data processing and machine learning, vessel tracking data is revolutionizing the ability to provide data on fishing activity at high spatial and temporal resolution. These data now underpin numerous studies that characterize global industrial fishing activity[71,154,155], its overlap with target and non-target species[156,157], assess conservation actions[158–160], and reveal illegal fishing[111]. Similar approaches have also demonstrated successful applications to small-scale fisheries[161162]. Data from vessel tracking systems can be supplemented with vessel detections from other space-based technologies like radar, day and nighttime optical imagery[110,111,163,] and radio frequency[164] as well as with aerial surveys[165] and drones[166] to better assess fleet size and distribution and monitor IUU fishing. On board vessels, remote electronic monitoring with video cameras, instead of solely on-board observers and logbooks, can reduce cost and speed of collecting data – especially for non-multispecies fisheries – improve coverage of a fleet, and enhance compliance around fishing activities and location[106]. It may provide these benefits to small-scale fisheries as well but can be less effective filling in some of the essential data gaps around bycatch and can be affected by species type and haul size (larger catches reduce accuracy)[167]. However, there is reluctance and lack of adoption due to everything from perceived intrusion of privacy by the industry, equipment and data storage requirements, and equipment challenges in the harsh marine environment (e.g. corrosion)[163]. In aquaculture, similar hurdles exist for more “Precision” or “Smart" farming data-driven approaches, especially in poorer or more rural regions where access to electricity or the internet is not reliable[168,16]. Indeed, wealthier countries, such as the United States, likely have a greater capacity to improve their aquaculture data more quickly with the right internal coordination and support, especially across the diverse agencies tasked or interested in aquatic commodity data collection (for example, Froehlich et al. [77]). Ultimately, whether this improved data collection makes its way into FAO statistics depends on several factors, not least of them being financial support and incentives or mandates by local/regional governments to use certain technologies.

Across the food system, these and other advanced technologies hold great potential to revolutionize agricultural data creation and curation, ultimately improving food production decision-making processes and driving more sustainable and efficient food system practices.

4.2. Requisite Supporting Policy Innovations

As technology advances and new solutions emerge, policymakers must balance continuity with innovation to sustain effective agricultural data systems while exploring new approaches to address emerging challenges. A stable and effective policy framework is crucial, but policymakers must also consider new ideas and approaches that build upon existing strengths without undermining existing systems[170,171]. Thus, it is important to highlight pathways forward for improving global agricultural data availability and quality.

Collaboration and coordination among countries and institutions are essential to develop comprehensive policies and strategies for food production data collection, management, and sharing. However, the proliferation of policy frameworks and actors in the agricultural data ecosystem can lead to fragmentation and duplication of efforts, hindering progress toward common goals. Therefore, it is important to ensure policy alignment and harmonization across different levels and sectors to avoid conflicting regulations and to promote a coordinated approach[172,173]. One potential solution is to invest in greater collaboration and coordination among countries, institutions, initiatives, and research fields (agricultural economics, statistics, satellite Earth observation, and emerging AI/ML) to reduce duplication of efforts and enhance initiatives that often operate independently. Another possible solution is to establish a global platform for data sharing and coordination, enabling countries to share best practices (including standards and benchmarking) and data, which could facilitate the development of comprehensive policies and strategies for agricultural data management that consider unique contexts and characteristics of different countries and populations[174].

In addition to increasing the scope of agricultural data collection and monitoring, it is essential to have fit-for-purpose data governance systems in place. In view of the rapidly changing data landscape, the 2021 World Development Report (WDR)[175] on datasets highlights the need for a new social contract between data providers and data users of all types, founded on value (enabling the use and re-use of data for different purposes), trust (the rights and interests of all stakeholders are safeguarded), and equity (the benefits of data are shared equally). These principles of data governance at national and international levels could enforce such a social contract around data. While several elements of data governance occur primarily at the national level, resolution of some data governance challenges is possible only through international collaboration, such as combating cybercrime, reducing transaction costs by harmonizing legal and technical standards for data protection and interoperability, and (as noted above) surveilling and dealing with transboundary issues.

To ensure that the global platform for data sharing and coordination in agriculture is effective, it is important not to center goals on any single organization's self-interest but to instead develop a neutral and independent platform that is governed by a diverse set of stakeholders, including representatives from governments, private sector, academia, civil society, and farmers' organizations. This will ensure that data sharing priorities and decision-making processes are driven by the collective interest of all stakeholders, and that it remains accountable to them. In this regard, the FAO and initiatives such as GODAN and GEOGLAM could play important roles in the development and governance of a global platform for agricultural data. These organizations are well placed to help ensure that the platform is designed to promote open data sharing and collaboration for the public good, and not for the benefit of any single entity. By leveraging their networks and expertise, they are also situated to establish best practices and standards for data sharing and to facilitate the development of policies and strategies for agricultural data management that consider the unique contexts and characteristics of different countries and populations. Further, these organizations can also facilitate the responsible sharing and use of the vast amounts of data (e.g., producer surveys, field trials) which a host of national and international research institutes (e.g., CGIAR) continue to collect through hundreds of projects worldwide. These data are often collected for the purposes of a specific project but hold great (and unrealized) value for contributing to a more complete understanding of food production systems across diverse contexts. Along similar lines, an effective global platform should also reduce redundancy, build on existing infrastructure largely within FAO, increase accessibility through digital transformations such as a smartphone-based app, and explore the feasibility of user-uploaded content including crowdsourcing and citizen science.

There is widespread consensus on the need for massive expansion in what is invested in data collection. Within the context of data collection in pursuit of the SDGs[176], there is an urgent need for a comprehensive set of regulatory standards, along with the development of the necessary physical infrastructure (such as low-cost mobile broadband) and strengthened public institutions, to govern and shape digital innovations towards sustainable development of agriculture and other sectors. Indeed, recent work[177] has highlighted the necessity for rigorous monitoring of the entire food system; without it, there is little chance of being able to identify in a timely way when course corrections will be needed and holding different food system actors to account will be very difficult. As outlined above, the information available on current cropping, livestock, and aquatic systems and production is uneven in coverage and scope, and substantial investments in data collection, collation, and curation will be needed to support decision-making in the agricultural sector. To this end, national and international funding agencies are increasingly making the publication of data an integral part of the work plan and projects evaluation, which is one step in the right direction. This could be made even more useful if the major research and R4D funders helped to promote standardized data collection, documentation, and accessibility procedures, or even made them a requirement for funding.

Among the many challenges that policymakers face in confronting food production data scarcity is striking a balance between safeguarding the privacy and confidentiality of production data while also catering to the economic interests of the many stakeholders - including farmers, researchers, private and public technology, and service providers - who have an interest in accessing and using those data[129]. Of the three elements of a fit-for-purpose social contract on data[175] perhaps the most foundational is trust, something that is echoed in recent food system literature as a critical component of scalable innovation[133]. There is much that policymakers can do to help inculcate trust. Setting up wide-ranging national and international dialogues on data governance would help in understanding stakeholders' concerns and needs; though these dialogues would need to be carried out in ways that are sensitive to socio-cultural differences and the existence of power dynamics that may vary greatly in different contexts[103]. There is also the need to evaluate existing regulatory frameworks and how they can be improved, at both national and international levels[129]. There is no doubt that achieving the balancing act between ensuring the legitimacy and security of agricultural data and generating added value from their use for all stakeholders involved will be very difficult to achieve. Nevertheless, striving towards such balance is vital if the full power of data and the digitalization of our food system are to be realized for maximal public good.

5. Conclusion

Continuously developing and maintaining detailed and up-to-date food production statistics is a critical yet complex undertaking, requiring technological, institutional, and policy coordination as well as sustained efforts and dedicated resources. Yet, despite the continuing importance of the food production sector – particularly in low- and middle-income countries – and its critical role in achieving multiple sustainability targets (including those related to food security, livelihoods, and environmental protection), our assessment demonstrates that substantial data gaps persist globally. This stems jointly (and in varying degrees) from a lack of political will, insufficient capacity and funding, and ineffective or inadequate international support and prevents a comprehensive and accurate understanding of the current state of global agriculture. At best, such information deficiencies can lead to ineffective interventions and, at worst, can contribute to misinformed actions that erode the sustainability of food systems. Our assessment, therefore, sought to identify the key factors contributing to global agricultural data scarcity and the potential measures to address them.

Agricultural, livestock, and fisheries censuses will likely continue to be the primary source of information in the food production sector, and it is, therefore, essential to ensure their accuracy and reliability. Our findings reveal that restricted data accessibility and a lack of transparency in many countries are key hindrances in the global food production data landscape. Irregular intervals in census occurrence, diverse methodologies followed by each nation, and a lack of standardization make harmonizing agricultural statistics difficult, and issues of data privacy, particularly related to access to microdata, further exacerbate challenges of accessibility and interoperability. In addition to these technical challenges, diverse institutional and policy obstacles – including inadequate institutional capacity, transboundary issues, and policy silos – hinder the progress of collecting, disseminating, and utilizing food production data in a

Version of Record at: https://doi.org/10.1038/s43017-024-00516-2

consistent manner globally. A fundamental obstacle to systematic food production data collection remains the lack of dedicated and sustained long-term public funding.

Leveraging advanced technologies can offer an effective means of data collection and dissemination. Estimates derived from satellite remote sensing, in-situ sensors, and digital monitoring using AI/ML and other modeling techniques can be a promising complement for filling data gaps in censuses. Complements to censuses become especially important in times of crisis such as wars or pandemics, which can delay or halt on-the-ground data collection entirely. However, while these technologies offer a variety of potential benefits, the need for intensive computational resources as well as very real ethical and privacy concerns that these approaches can potentially present, must be critically considered to ensure their responsible use. Issues of data privacy should be addressed through clear data collection and sharing policies and the application of emerging technologies such as blockchain and federated learning. Enabling institutional arrangements and policy environments will also be essential to promote collaboration and partnership among stakeholders, to foster responsible data governance, and ensure equitable benefits sharing across all food system actors. Addressing the key factors contributing to food production data scarcity is a fundamental step towards realizing the full potential of food systems for achieving multiple sustainability goals.

References

1. Loizou, E., Karelakis, C., Galanopoulos, K. & Mattas, K. The role of agriculture as a development tool for a regional economy. *Agric Syst* **173**, 482–490 (2019).

2. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).

3. Kroodsma, D. A. *et al.* Tracking the global footprint of fisheries. *Science (1979)* **359**, 904–908 (2018).

4. Carpenter, S. R., Booth, E. G. & Kucharik, C. J. Extreme precipitation and phosphorus loads from two agricultural watersheds. *Limnol Oceanogr* **63**, 1221–1233 (2018).

5. Zhang, J., Wang, S., Zhao, W., Meadows, M. E. & Fu, B. Finding pathways to synergistic development of Sustainable Development Goals in China. *Humanit Soc Sci Commun* **9**, 21 (2022).

6. Searchinger, T. *et al. World resources report 2013-14 : interim findings To cite this version : Creating a Sustainable Food Future*. (2020).

7. Willett, W. *et al.* Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *The Lancet* **393**, 447–492 (2019).

8. Hoekstra, A. Y. & Mekonnen, M. M. The water footprint of humanity. *Proceedings of the National Academy of Sciences* **109**, 3232–3237 (2012).

9. Saleem, M. Possibility of utilizing agriculture biomass as a renewable and sustainable future energy source. *Heliyon* **8**, e08905 (2022).

10. Poore, J. & Nemecek, T. Reducing food’s environmental impacts through producers and consumers. *Science (1979)* **360**, 987–992 (2018).

11. Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A. & Hansen, M. C. Classifying drivers of global forest loss. *Science (1979)* **361**, 1108–1111 (2018).

12. Agricultural Model Intercomparison and Improvement Project (AgMIP) . AgMIP. *https://doi.org/10.15482/USDA.ADC/1212378. Accessed 2023-09-03* (2015).

Accepted Manuscript
Version of Record at: https://doi.org/10.1038/s43017-024-00516-2

13. Warszawski, L. *et al.* The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework. *Proceedings of the National Academy of Sciences* **111**, 3228–3232 (2014).

14. Becker-Reshef, I., Justice, C., Whitcraft, A. K. & Jarvis, I. Geoglam: A Geo Initiative on Global Agricultural Monitoring. in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* 8155–8157 (2018). doi:10.1109/IGARSS.2018.8517575.

15. Sellitti, S. Evaluation of CGIAR Platform for Big Data in Agriculture. in (2021).

16. Yu, Q. *et al.* A cultivated planet in 2010-Part 2: The global gridded agricultural-production maps. *Earth Syst Sci Data* **12**, 3545–3572 (2020).

17. Fischer, G. *et al. Global Agri Ecological Zones V4 Model Documentation*. (2021).

18. Portmann, F. T., Siebert, S. & Döll, P. MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochem Cycles* **24**, n/a-n/a (2010).

19. World Bank. *Global Strategy to Improve Agricultural and Rural Statistics Report of the Friends of the Chair on Agricultural Statistics. Report of the Friends of the Chair on Agricultural Statistics .United Nations Economic and Social Council Statistical Commission Forty-first session* . (2010).

20. Ray, D. K. *et al.* Crop harvests for direct food use insufficient to meet the UN's food security goal. *Nat Food* **3**, 367–374 (2022).

21. Weersink, A., Fraser, E., Pannell, D., Duncan, E. & Rotz, S. Opportunities and Challenges for Big Data in Agricultural and Environmental Analysis. *Annu Rev Resour Economics* **10**, 19–37 (2018).

22. Food and Agriculture Organization of the United Nations. *Global review of agricultural census methodologies and results (2006 – 2015).World Programme for the Census of Agriculture 201.0 FAO Statistical Development Series 18. Global review of agricultural census methodologies and results (2006 – 2015)* (2021) doi:10.4060/cb2650en.

23. Food and Agriculture Organization of the United Nations. FAOSTAT Statistical Database.[Rome]:FAO, 1997. Accessed November 2022. (2022) doi:10.1038/s41597-022-01675-x.

24. Food and Agriculture Organization of the United Nations. Conducting Agricultural Censuses and Surveys (FAO Statistical Devel- opment Series No. 6). (1996).

25. EUROSTAT. Statistical Office of the European Union. (2023). https://ec.europa.eu/eurostat (2023).

26. Lahti, L., Huovari, J., Kainu, M. & Biecek, P. *Retrieval and analysis of Eurostat open data with the eurostat package*. http://ropengov.github.io/eurostat.

27. Food and Agricultural Organization of the United Nations. *World Programme for the Census of Agriculture 2020 Volume 1 Programme, concepts and definitions FAO Statistical Devlopment Series 15*. (2017).

28. Maria, D., Michele, M. & Felix, R. Development of a national and sub-national crop calendars data set compatible with remote sensing derived land surface phenology. in (2018).

29. Fritz, S. *et al.* A comparison of global agricultural monitoring systems and current gaps. *Agric Syst* **168**, 258–272 (2019).

30. Sacks, W. J., Deryng, D., Foley, J. A. & Ramankutty, N. Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography* **19**, 607–620 (2010).

31. Becker-Reshef, I. *et al.* Crop Type Maps for Operational Global Agricultural Monitoring. *Sci Data* **10**, 172 (2023).

32. Kotsuki, S. & Tanaka, K. SACRA – a method for the estimation of global high-resolution crop calendars from a satellite-sensed NDVI. *Hydrol Earth Syst Sci* **19**, 4441–4461 (2015).

33. Laborte, A. G. *et al.* RiceAtlas, a spatial database of global rice calendars and production. *Sci Data* **4**, 170074 (2017).

34. See, L. *et al.* Improved global cropland data as an essential ingredient for food security. *Glob Food Sec* **4**, 37–45 (2015).

35. Food and Agricultural Organization of the United Nations. *Independent External Evaluation of the United Nations Food and Agricultural Organization*. (2005).

36. Food and Agriculture Organization of the United Nations. *Independent External Evaluation of FAO's Role and Work in Statistics*. (2008).

37. Iizumi, T. *et al.* Historical changes in global yields: major cereal and legume crops from 1982 to 2006. *Global Ecology and Biogeography* **23**, 346–357 (2014).

38. Gangopadhyay, P. K., Shirsath, P. B., Dadhwal, V. K. & Aggarwal, P. K. A new two-decade (2001–2019) high-resolution agricultural primary productivity dataset for India. *Sci Data* **9**, 730 (2022).

39. Wilkinson, M. D. *et al.* Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, (2016).

40. Leff, B., Ramankutty, N. & Foley, J. A. Geographic distribution of major crops across the world. *Global Biogeochem Cycles* **18**, (2004).

41. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochem Cycles* **22**, 1–19 (2008).

42. Ramankutty, N., Evan, A. T., Monfreda, C. & Foley, J. A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochem Cycles* **22**, (2008).

43. Deutsch, C. A. *et al.* Increase in crop losses to insect pests in a warming climate. *Science (1979)* **361**, 916–919 (2018).

44. Lombardozzi, D. L., Bonan, G. B., Levis, S. & Lawrence, D. M. Changes in Wood Biomass and Crop Yields in Response to Projected CO2, O3, Nitrogen Deposition, and Climate. *J Geophys Res Biogeosci* **123**, 3262–3282 (2018).

45. Matteo Rolle, Stefania Tamea & Pierluigi Claps. Improved large-scale crop water requirement estimation through new high-resolution reanalysis dataset. in *EGU General Assembly* (2020).

46. Fischer, G. , *et al. Global Agroecological Zones (GAEZ v3.0), IIASA, Laxenburg, Austria and FAO, Rome, Italy, 2012*. (2012).

47. Bartholomé, E. & Belward, A. S. GLC2000: a new approach to global land cover mapping from Earth observation data. *Int J Remote Sens* **26**, 1959–1977 (2005).

48. Potapov, P. *et al.* Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat Food* **3**, 19–28 (2022).

49. Klein Goldewijk, K., Beusen, A., van Drecht, G. & de Vos, M. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography* **20**, 73–86 (2011).

50. Kerner, H. *et al.* How accurate are existing land cover maps for agriculture in Sub-Saharan Africa? (2023).

51. Meisner, J. *et al.* A time-series approach to mapping livestock density using household survey data. *Sci Rep* **12**, 13310 (2022).

52. Gilbert, M. *et al.* Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci Data* **5**, 180227 (2018).

53. Gilbert, M. *et al.* Global cattle distribution in 2015 (5 minutes of arc). *Harvard Dataverse, V1* (2022).

54. Da Re, D. *et al.* Downscaling livestock census data using multivariate predictive models: Sensitivity to modifiable areal unit problem. *PLoS One* **15**, e0221070- (2020).

55. Nicolas, G. *et al.* Using Random Forest to Improve the Downscaling of Global Livestock Census Data. *PLoS One* **11**, e0150424- (2016).

56. Herrero, M. *et al.* Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. *Proceedings of the National Academy of Sciences* **110**, 20888–20893 (2013).

57. Robinson, T. P. *et al. Global livestock production systems*. (FAO and ILRI, 2011).

58. Kruska, R. L., Reid, R. S., Thornton, P. K., Henninger, N. & Kristjanson, P. M. Mapping livestock-oriented agricultural production systems for the developing world. *Agric Syst* **77**, 39–63 (2003).

59. Seré Rabé, C. & Steinfeld, H. World livestock production systems. *FAO Animal Production and Health Paper* (1996).

60. Dixon, J. A., Gibbon, D. P. & Gulliver, A. *Farming systems and poverty: improving farmers' livelihoods in a changing world*. (Food & Agriculture Org., 2001).

61. Hammond, J. *et al.* The Rural Household Multi-Indicator Survey (RHoMIS) for rapid characterisation of households to inform climate smart agriculture interventions: Description and applications in East Africa and Central America. *Agric Syst* **151**, 225–233 (2017).

62. Zane, G. & Pica-Ciamarra, U. The contribution of livestock to household livelihoods in Tanzania and Uganda: measuring tradable and non-tradable livestock outputs. *Trop Anim Health Prod* **53**, 304 (2021).

63. Carletto, C. Better data, higher impact: Improving agricultural data systems for societal change. *European Review of Agricultural Economics* **48**, 719–740 (2021).

64. Carletto, C., Dillon, A. & Zezza, A. Chapter 81 - Agricultural data collection to minimize measurement error and maximize coverage. in *Handbook of Agricultural Economics* (eds. Barrett, C. B. & Just, D. R.) vol. 5 4407–4480 (Elsevier, 2021).

65. Duncan, A. J., Lukuyu, B., Mutoni, G., Lema, Z. & Fraval, S. Supporting participatory livestock feed improvement using the Feed Assessment Tool (FEAST). *Agron Sustain Dev* **43**, 34 (2023).

66. Fritz, S. *et al.* A global dataset of crowdsourced land cover and land use reference data. *Sci Data* **4**, 1–8 (2017).

67. Food and Agriculture Organization of the United Nations. *Coordinating Working Party on Fishery Statistics (CWP) Handbook.*

https://www.fao.org/cwp-on-fishery-statistics/handbook/introduction/en/ . (2020).

68. Food and Agriculture Organization of the United Nations. The State of World Fisheries and Aquaculture(SOFIA). https://www.fao.org/3/cc0461en/online/sofia/2022/world-fisheries-aquaculture.html (2022).

69. Watson, R. A. A database of global marine commercial, small-scale, illegal and unreported fisheries catch 1950–2014. *Sci Data* **4**, 1–9 (2017).

70. Zeller, D. *et al.* Still catching attention: Sea Around Us reconstructed global catch data, their spatial expression and public accessibility. *Mar Policy* **70**, 145–152 (2016).

71. Kroodsma, D. A. *et al.* Tracking the global footprint of fisheries. *Science (1979)* **359**, 904–908 (2018).

72. Pauly, D. & Zeller, D. Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nat Commun* **7**, 10244 (2016).

73. Fluet-Chouinard, E., Funge-Smith, S. & McIntyre, P. B. Global hidden harvest of freshwater fish revealed by household surveys. *Proceedings of the National Academy of Sciences* **115**, 7623–7628 (2018).

74. Ye, Y. *et al.* FAO's statistic data and sustainability of fisheries and aquaculture: Comments on Pauly and Zeller (2017). *Mar Policy* **81**, 401–405 (2017).

75. Food and Agricultural Organization of the United Nations. *Fishery and Aquaculture Statistics. Global production by production source 1950-2020 (FishStatJ). In: FAO Fisheries and Aquaculture Division [online]. Rome. Updated 2022*. (2022).

76. Klinger, D. H. *et al.* Moving beyond the fished or farmed dichotomy. *Mar Policy* **38**, 369–374 (2013).

77. Froehlich, H. E. *et al.* Piecing together the data of the US marine aquaculture puzzle. *J Environ Manage* **308**, 114623 (2022).

78. Clawson, G. *et al.* Mapping the spatial distribution of global mariculture production. *Aquaculture* **553**, 738066 (2022).

79. Ottinger, M., Bachofer, F., Huth, J. & Kuenzer, C. Mapping aquaculture ponds for the coastal zone of Asia with Sentinel-1 and Sentinel-2 time series. *Remote Sens (Basel)* **14**, 153 (2021).

80. Pauly D., Z. D. , & Palomares M.L.D. Sea Around Us Concepts, Design and Data (seaaroundus.org). *Design and Data (seaaroundus. org)* **551**, 552 (2020).

81. Cassidy, E. S., West, P. C., Gerber, J. S. & Foley, J. A. Redefining agricultural yields: from tonnes to people nourished per hectare. *Environmental Research Letters* **8**, 034015 (2013).

82. Fischer G *et al. Global agro-ecological zone V4 – Model documentation*. *Global agro-ecological zone V4 – Model documentation* https://gaez.fao.org/ (2021) doi:10.4060/cb4744en.

83. Iizumi, T. & Sakai, T. The global dataset of historical yields for major crops 1981–2016. *Sci Data* **7**, (2020).

84. Franke, J. A. *et al.* The GGCMI Phase 2 experiment: Global gridded crop model simulations under uniform changes in CO2, temperature, water, and nitrogen levels (protocol version 1.0). *Geosci Model Dev* **13**, 2315–2336 (2020).

85. Jägermeyr, J. *et al.* Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat Food* **2**, 873–885 (2021).

86. Müller, C. *et al.* The Global Gridded Crop Model Intercomparison phase 1 simulation dataset. *Sci Data* **6**, 50 (2019).

87. The Global Yield Gap and Water Productivity Atlas. (GYGA). http://www.yieldgap.org/ (2022).

88. Mueller, N. D. *et al.* Closing yield gaps through nutrient and water management. *Nature* **490**, 254–257 (2012).

89. Jackson, N. D., Konar, M., Debaere, P. & Estes, L. Probabilistic global maps of crop-specific areas from 1961 to 2014. *Environmental Research Letters* **14**, (2019).

90. Ray, D. K. *et al.* Climate change has likely already affected global food production. *PLoS One* **14**, e0217148- (2019).

91. International Food Policy Research Institute. *Global Spatially-Disaggregated Crop Production Statistics Data for 2000 Version 3.0.7*. https://doi.org/10.7910/DVN/A50I2T (2019).

92. International Food Policy Research Institute (IFPRI); International Institute for Applied Systems Analysis (IIASA). *Global Spatially-Disaggregated Crop Production Statistics Data for 2005 Version 3.2*. https://doi.org/10.7910/DVN/DHXBJX (2016).

93. International Food Policy Research Institute. *Global Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 2.0*. https://doi.org/10.7910/DVN/PRFF8V (2019).

94. West, P. C. , *et al. Leverage points for improving global food security and the environment. Science, 345(6194), 325-328.* http://science.sciencemag.org/content/345/6194/325 .abstract (2014).

95. Clawson, G. *et al.* Mapping the spatial distribution of global mariculture production. *Aquaculture* **553**, 738066 (2022).

96. Olofsson, P. *et al.* Good practices for estimating area and assessing accuracy of land change. *Remote Sens Environ* **148**, 42–57 (2014).

Accepted Manuscript
Version of Record at: https://doi.org/10.1038/s43017-024-00516-2

97. Stehman, S. V & Foody, G. M. Key issues in rigorous accuracy assessment of land cover products. *Remote Sens Environ* **231**, 111199 (2019).

98. Laso Bayas, J. C. *et al.* A global reference database of crowdsourced cropland data collected using the Geo-Wiki platform. *Sci Data* **4**, 1–10 (2017).

99. Gourlay, S., Kilic, T. & Lobell, D. B. A new spin on an old debate: Errors in farmer-reported production and their implications for inverse scale-Productivity relationship in Uganda. *J Dev Econ* **141**, 102376 (2019).

100. Phelps, L. N. & Kaplan, J. O. Land use for animal production in global change studies: Defining and characterizing a framework. *Glob Chang Biol* **23**, 4457–4471 (2017).

101. Lowder, S. K., Sánchez, M. V & Bertini, R. Which farms feed the world and has farmland become more concentrated? *World Dev* **142**, 105455 (2021).

102. van Andel, M., Tildesley, M. J. & Gates, M. C. Challenges and opportunities for using national animal datasets to support foot-and-mouth disease control. *Transbound Emerg Dis* **68**, 1800–1813 (2021).

103. Abebe, R. *et al.* Narratives and counternarratives on data sharing in Africa. in *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* 329–341 (2021).

104. Beddington, J. R., Agnew, D. J. & Clark, C. W. Current Problems in the Management of Marine Fisheries. *Science (1979)* **316**, 1713–1716 (2007).

105. Bradley, D. *et al.* Opportunities to improve fisheries management through innovative technology and advanced data systems. *Fish and Fisheries* **20**, 564–583 (2019).

106. van Helmond, A. T. M. *et al.* Electronic monitoring in fisheries: lessons from global experiences and future opportunities. *Fish and Fisheries* **21**, 162–189 (2020).

107. Seto, K. L. *et al.* Fishing through the cracks: The unregulated nature of global squid fisheries. *Sci Adv* **9**, eadd8125 (2023).

108. Taconet, M. *et al. Global atlas of AIS-based fishing activity : challenges and opportunities*. (2019).

109. Welch, H. *et al.* Hot spots of unseen fishing vessels. *Sci Adv* **8**, eabq2109 (2023).

110. Kroodsma, D. A. *et al. Title: Revealing the Global Longline Fleet with Satellite Radar*. (2022).

111. Park, J. *et al.* Illuminating dark fishing fleets in North Korea. *Sci Adv* **6**, eabb1197 (2023).

112. Ottinger, M., Clauss, K. & Kuenzer, C. Large-scale assessment of coastal aquaculture ponds with Sentinel-1 time series data. *Remote Sens (Basel)* **9**, (2017).

113. Xu, Y. *et al.* Mapping aquaculture areas with multi-source spectral and texture features: A case study in the pearl river basin (guangdong), China. *Remote Sens (Basel)* **13**, (2021).

114. Cozzolino, E. & Lasta, C. A. Use of VIIRS DNB satellite images to detect jigger ships involved in the Illex argentinus fishery. *Remote Sens Appl* **4**, 167–178 (2016).

115. Li, J. *et al.* Satellite observation of a newly developed light-fishing “hotspot” in the open South China Sea. *Remote Sens Environ* **256**, 112312 (2021).

116. FAO, D. U. & W. *Illuminating Hidden Harvests: The contributions of small-scale fisheries to sustainable development – Executive summary. Rome.*

https://www.fao.org/3/cc4576en/cc4576en.pdf (2023).

117. Halim, A. *et al. Developing a functional definition of small-scale fisheries in support of marine capture fisheries management in Indonesia Developing a functional definition of small-scale fisheries in support of marine capture 1 fisheries management in Indonesia 2*. (2018).

118. Smith, H. & Basurto, X. Defining Small-Scale Fisheries and Examining the Role of Science in Shaping Perceptions of Who and What Counts: A Systematic Review. *Front Mar Sci* **6**, (2019).

119. Carletto, C., Jolliffe, D. & Banerjee, R. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *Journal of Development Studies* **51**, 133–148 (2015).

120. Global Food for Thought. Could a Data Sharing Protocol be Agriculture's Missing Link? https://globalaffairs.org/commentary-and-analysis/blogs/could-data-sharing-protocol-be-agricultures-missing-link (2021).

121. Fisher, A. & Fukuda-Parr, S. Introduction—data, knowledge, politics and localizing the SDGs. *J Human Dev Capabil* **20**, 375–385 (2019).

122. Montenegro de Wit, M. & Canfield, M. Feeding the world, byte by byte': emergent imaginaries of data productivism. *Journal of Peasant Studies* (2023) doi:10.1080/03066150.2023.2232997.

123. Wolfert, S., Ge, L., Verdouw, C. & Bogaardt, M.-J. Big Data in Smart Farming – A review. *Agric Syst* **153**, 69–80 (2017).

124. Spanaki, K., Karafili, E. & Despoudi, S. AI applications of data sharing in agriculture 4.0: A framework for role-based data access control. *Int J Inf Manage* **59**, 102350 (2021).

125. Brinkerhoff, D. W. & Brinkerhoff, J. M. Public–private partnerships: Perspectives on purposes, publicness, and good governance. *Public administration and development* **31**, 2–14 (2011).

126. Wiggins, S., Kirsten, J. & Llambí, L. The future of small farms. *World Dev* **38**, 1341–1348 (2010).

127. Piñeiro, V. *et al.* A scoping review on incentives for adoption of sustainable agricultural practices and their outcomes. *Nat Sustain* **3**, 809–820 (2020).

128. Rotz, S. *et al.* The politics of digital agricultural technologies: a preliminary review. *Sociol Ruralis* **59**, 203–229 (2019).

129. Jouanjean, M.-A., Casalini, F., Wiseman, L. & Gray, E. Issues around data governance in the digital transformation of agriculture: The farmers' perspective. (2020).

130. Jensen, Dempster, T., Thorstad, E. B., Uglem, I. & Fredheim, A. Escapes of fishes from Norwegian sea-cage aquaculture: Causes, consequences and prevention. *Aquaculture Environment Interactions* vol. 1 71–83 Preprint at https://doi.org/10.3354/aei00008 (2010).

131. Pinsky, M. L. *et al.* Preparing ocean governance for species on the move. *Science (1979)* **360**, 1189–1191 (2018).

132. d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L. & Van der Velde, M. Monitoring crop phenology with street-level imagery using computer vision. *Comput Electron Agric* **196**, 106866 (2022).

133. Herrero, M. *et al.* Innovation can accelerate the transition towards a sustainable food system. *Nat Food* **1**, 266–272 (2020).

134. Barrett, C. B. *et al.* Bundling innovations to transform agri-food systems. *Nat Sustain* **3**, 974–976 (2020).

135. Paliyam, M., Nakalembe, C., Liu, K., Nyiawung, R. & Kerner, H. Street2sat: A machine learning pipeline for generating ground-truth geo-referenced labeled datasets from street-level images. in *ICML 2021 Workshop on Tackling Climate Change with Machine Learning* (2021).

136. Yan, Y. & Ryu, Y. Exploring Google Street View with deep learning for crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing* **171**, 278–296 (2021).

137. van der Merwe, D., Burchfield, D. R., Witt, T. D., Price, K. P. & Sharda, A. Drones in agriculture. *Advances in agronomy* **162**, 1–30 (2020).

138. d'Andrimont, R. *et al.* Crowdsourced street-level imagery as a potential source of in-situ data for crop monitoring. *Land (Basel)* **7**, 127 (2018).

139. Kerner, H. R. *et al.* Phenological normalization can improve in-season classification of maize and soybean: A case study in the central US Corn Belt. *Science of Remote Sensing* **6**, 100059 (2022).

140. Wang, S. *et al.* Mapping crop types in southeast India with smartphone crowdsourcing and deep learning. *Remote Sens (Basel)* **12**, 2957 (2020).

141. Tseng, G., Kerner, H., Nakalembe, C. & Becker-Reshef, I. Learning to predict crop type from heterogeneous sparse labels using meta-learning. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 1111–1120 (2021).

142. Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H. & Islam, N. A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sens (Basel)* **14**, 1990 (2022).

143. Deines, J. M., Wang, S. & Lobell, D. B. Satellites reveal a small positive yield effect from conservation tillage across the US Corn Belt. *Environmental Research Letters* **14**, 124038 (2019).

144. Ferrag, M. A., Shu, L., Yang, X., Derhab, A. & Maglaras, L. Security and privacy for green IoT-based agriculture: Review, blockchain solutions, and challenges. *IEEE access* **8**, 32031–32053 (2020).

145. Rahman, M. U., Baiardi, F. & Ricci, L. Blockchain smart contract for scalable data sharing in IoT: a case study of smart agriculture. in *2020 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)* 1–7 (IEEE, 2020).

146. Durrant, A. *et al.* The role of cross-silo federated learning in facilitating data sharing in the agri-food sector. *Comput Electron Agric* **193**, 106648 (2022).

147. UNSC. *Spatial anonymization: guidance note for the Inter-Secretariat Working Group on household surveys, United Nations, New York.* https://unstats.un.org/unsd/statcom/52nd-session/documents/BG-3l-Spatial_Anonymization-E.pdf. (2021).

148. Tedeschi, L. O. *et al.* Quantification of methane emitted by ruminants: a review of methods. *J Anim Sci* **100**, skac197 (2022).

149. Ramayo-Caldas, Y. *et al.* Identification of rumen microbial biomarkers linked to methane emission in Holstein dairy cows. *Journal of Animal Breeding and Genetics* **137**, 49–59 (2020).

150. Han, C. S. *et al.* Invited review: Sensor technologies for real-time monitoring of the rumen environment. *Journal of Dairy Science* vol. 105 6379–6404 Preprint at https://doi.org/10.3168/jds.2021-20576 (2022).

151. Tullo, E., Finzi, A. & Guarino, M. Review: Environmental impact of livestock farming and Precision Livestock Farming as a mitigation strategy. *Science of The Total Environment* **650**, 2751–2760 (2019).

152. Chase, L. E. & Fortina, R. Environmental and Economic Responses to Precision Feed Management in Dairy Cattle Diets. *Agriculture (Switzerland)* vol. 13 Preprint at https://doi.org/10.3390/agriculture13051032 (2023).

153. Mackenzie, S. The Potential Contribution of Smart Animal Nutrition in Reducing the Environmental Impacts of Livestock Systems. . *In Smart Livestock Nutrition (pp. 311-336). Cham: Springer International Publishing.* (2023).

154. Sala, E. *et al.* The economics of fishing the high seas. *Sci Adv* **4**, eaat2504 (2023).

155. Seto, K. L. *et al.* Fishing through the cracks: The unregulated nature of global squid fisheries. *Sci Adv* **9**, eadd8125 (2023).

156. White, T. D. *et al.* Predicted hotspots of overlap between highly migratory fishes and industrial fishing fleets in the northeast Pacific. *Sci Adv* **5**, eaau3761 (2023).

157. Queiroz, N. *et al.* Global spatial risk assessment of sharks under the footprint of fisheries. *Nature* **572**, 461–466 (2019).

158. White, T. D. *et al.* Assessing the effectiveness of a large marine protected area for reef shark conservation. *Biol Conserv* **207**, 64–71 (2017).

159. McDermott, G. R., Meng, K. C., McDonald, G. G. & Costello, C. J. The blue paradox: Preemptive overfishing in marine reserves. *Proceedings of the National Academy of Sciences* **116**, 5319–5325 (2019).

160. Cabral, R. B. *et al.* Rapid and lasting gains from solving illegal fishing. *Nat Ecol Evol* **2**, 650–658 (2018).

161. Behivoke, F. *et al.* Estimating fishing effort in small-scale fisheries using GPS tracking data and random forests. *Ecol Indic* **123**, 107321 (2021).

162. Tilley, A., Dos Reis Lopes, J. & Wilkinson, S. P. PeskAAS: A near-real-time, open-source monitoring and analytics system for small-scale fisheries. *PLoS One* **15**, e0234760- (2020).

163. Snapir, B., Waine, T. W. & Biermann, L. Maritime vessel classification to monitor fisheries with SAR: Demonstration in the North Sea. *Remote Sens (Basel)* **11**, (2019).

164. Sarda, K. , CaJacob, D. , Orr, N. , & Zee, R. Making the invisible visible: precision RF-emitter geolocation from space by the hawkeye 360 pathfinder mission. (2018).

165. Iacarella, J. C. *et al.* Application of AIS- and flyover-based methods to monitor illegal and legal fishing in Canada's Pacific marine conservation areas. *Conserv Sci Pract* **5**, e12926 (2023).

166. Prayudi, A., Sulistijono, I. A., Risnumawan, A. & Darojah, Z. Surveillance System for Illegal Fishing Prevention on UAV Imagery Using Computer Vision. in *2020 International Electronics Symposium (IES)* 385–391 (2020). doi:10.1109/IES50839.2020.9231539.

167. Bartholomew, D. C. *et al.* Remote electronic monitoring as a potential alternative to on-board observers in small-scale fisheries. *Biol Conserv* **219**, 35–45 (2018).

168. Antonucci, F. & Costa, C. Precision aquaculture: a short review on engineering innovations. *Aquaculture International* **28**, 41–57 (2020).

169. Rastegari, H. *et al.* Internet of Things in aquaculture: A review of the challenges and potential solutions based on current and future trends. *Smart Agricultural Technology* 100187 (2023).

170. Cervantes-Godoy, D. *et al. The future of food and agriculture: trends and challenges*. *The future of food and agriculture: trends and challenges* vol. 4 (2014).

171. Turnheim, B. *et al.* Evaluating sustainability transitions pathways: Bridging analytical approaches to address governance challenges. *Global environmental change* **35**, 239–253 (2015).

172. Dawes, S. S. Stewardship and usefulness: Policy principles for information-based transparency. *Gov Inf Q* **27**, 377–383 (2010).

173. Xie, W. *et al.* Crop switching can enhance environmental sustainability and farmer incomes in China. *Nature* (2023) doi:10.1038/s41586-023-05799-x.

174. Kochupillai, M., Kahl, M., Schmitt, M., Taubenböck, H. & Zhu, X. X. Earth Observation and Artificial Intelligence: Understanding emerging ethical issues and opportunities. *IEEE Geosci Remote Sens Mag* (2022).

175. World Bank. *World development report 2021: Data for better lives*. (2021) doi:10.1596/978-1-4648-1600-0.

176. Sachs, J. D. *et al.* Six Transformations to Achieve the Sustainable Development Goals. Nature Sustainability, 2, 805-814. Preprint at (2019).

177. Fanzo, J. *et al.* Viewpoint: Rigorous monitoring is necessary to guide food system transformation in the countdown to the 2030 global goals. *Food Policy* **104**, 102163 (2021).

Data availability

The data used to support the findings of this study are freely available in the Supplementary Information.

Acknowledgments

KFD and EAK were supported by USDA Grant no. 2022-67019-37180. HEF recognizes support from the University of California, Santa Barbara. JAG was supported by NSF #2121238. SH was supported by NSF Digital Transformation of Development (#2125913) and NSF Graduate Research Fellowship Program (#1752814).

Competing interests

The authors declare no competing interests.

Author contributions

KFD led the study. EAK and KFD conceived the study and designed all analyses. EAK, SH, PM, DR, and SS drafted the crop production section. MH and PT drafted the livestock and policy innovation sections. TC, HF, and JG drafted the fisheries and aquaculture section. HK, CN, and HAA drafted the technology section. EK, HAA, SH, and KFD compiled and edited the initial manuscript. EAK collected the food production data. EAK and SH produced graphical representations of the results. All authors contributed to writing and revising the paper.

Accepted Manuscript
Version of Record at: https://doi.org/10.1038/s43017-024-00516-2

Display Items

Table 1: List of global open-access data portals and data products on food production

Product name	Type of product	Sector	Data sources/ methods	Variable(s)	Item Specificity	Timeliness	Spatial Disaggregation	Identified strengths	Identified limitations
Cassidy et al.[80]	Gridded dataset	Crop	Census and model	Crop allocation to food, feed, and nonfood	41 crops	Annual: Circa 2000	Global: 5 arc min (~10 km by 10 km)	An extensive crop allocation dataset	Limited temporal coverage No distinction between seasons and systems
FAOSTAT[23]	Data portal	Crop, livestock, and fisheries	Census	Global agricultural statistics	278 crop and livestock variables	Annual: 1961 to 2021	245 countries: national level data	A compressive global agricultural dataset	Subnational data limitations National reporting and related data quality issues
GAEZ[46,81]	Gridded dataset	Crop	Census and model	Harvested area, production, and yield,	23 crop classes	Annual: Circa 2000 and 2010	Global: 5 arc min (~10 km by 10 km)	Irrigated and rainfed systems are distinguished	No distinction between seasons
GDHY[82]	Gridded dataset	Crop	Census and remote sensing	Yield	4 crops	Seasonal: 1981-2016	Global: 0.5◦ (55 km)	A clear distinction between seasons	Some locations have no data Limited crop classes no distinction between systems
GGCMI Phase I, II, and III[83–85]	Gridded dataset	Crop	Model	Yield and crop calendar	19 crops	Annual: 1901-2012	Global: 0.5◦ (55 km)	A clear distinction between systems and seasons	Coarser spatial resolution
GYGA[86]	Data portal	Crop	Census	Actual and potential yield and yield gap	13 crops	Annual: Current	70 countries: national level data	High-quality yield and yield gap data	Limited spatial coverage
M3-Crops[41]	Gridded dataset	Crop	Census	Harvested area and yield	175 crop classes	Annual: Circa 2000	Global: 5 arc min (~10 km by 10 km)	A wider range of crop classes	No distinction among seasons/system
MIRCA[18]	Gridded dataset	Crop	Census	Harvested area and crop calendar	26 crop classes	Monthly: Circa 2000	Global: 5 arc min (~10 km by 10 km)	Irrigated and rainfed systems and seasons are distinguished	Outdated Constraints in representing sub-national statistics
Mueller et al.[87]	Gridded dataset	Crop	Census	Fertilizer application rate and consumption	17 crops	Annual: Circa 2000	Global: 5 arc min (~10 km by 10 km)	An extensive national and subnational nutrient application data	Limited crop classes Limited temporal coverage
PCAM[88]	Gridded dataset	Crop	Model	Harvested area	17 crop classes	Annual: 1961 to 2014	Global: 0.5◦ (55 km)	Has a broader temporal span	No distinction between systems and seasons Model uncertainty
Ray et al.[89]	Gridded dataset	Crop	Census	Area harvested and yield	10 crops	Annual: 1961–2013	Global: 5 arc min (~10 km by 10 km)	A wider range of temporal coverage	Limited crop classes No distinction between systems and seasons
RiceAtlas[33]	Gridded dataset	Crop	Census	Harvested area, production, and crop calendar	Rice only	Annual: Circa 2010	Global: 2725 spatial units	A clear distinction between seasons	No distinction between systems Limited crop classes
SAGE[30]	Gridded dataset	Crop	Census	Crop calendar	19 crops	Annual: 1990s	Global; 0.5◦ (55 km)	A clear distinction between seasons	No distinction between systems

Accepted Manuscript
Version of Record at: https://doi.org/10.1038/s43017-024-00516-2

SPAM[90–92]	Gridded dataset	Crop	Census and model	Harvested area and yield	42 crop classes	Annual: Circa 2000, 2005, 2010, 2020 (Beta version)	Global: 5 arc min(~10 km by 10 km)	Irrigated and rainfed systems are distinguished	No distinction between seasons Model uncertainty
West et al. [93]	Gridded dataset	Crop	Census and model	Total nutrient balance	N and P balance of 140 crops	Annual: Circa 2000	Global: 5 arc min (~10 km by 10 km)	A diverse array of crop nutrient balances is addressed	Limited temporal coverage
Clawson et al.[78]	Gridded dataset	Fisheries and aquaculture	Mariculture	Farm location and taxon	6 generalized animal categories	2017	Global: 0.0083◦ (~1 km by 1 km)	Actual and modeled locations of marine aquaculture farms	Limited temporal and finer spatial resolution
FishStat[75]	Data portal	Fisheries and aquaculture	Census and reports	Global fisheries and aquaculture statistics	Datasets on production, trade, and consumption	Annual: 1950 to 2021	245 countries/territories: national-level data	An extensive dataset encompassing aquatic and fisheries data	Subnational data limitations National reporting and related data quality issues
GFW3[71]	Gridded dataset; Data portal	Fisheries and aquaculture	Model	Fishing hours by vessel, flag state, and gear type	16 gear types	Daily: 2012-2020 (gridded), 2012-present (data portal)	Global 0.1◦ (by vessel), 0.01◦ (by flag state and gear type)	Producer (vessel) specific activity at high spatial and temporal resolution	Lack of small scale producers, satellite reception, tampering
Ottinger et al.[79]	Gridded dataset	Fisheries and aquaculture	Ponds	Spatial distribution of aquaculture ponds	Pond	2019	Asia: 5-60m	Satellite (SAR) remote sensing of aquaculture ponds	Not species/taxa specific
GLPS[52,53]	Gridded dataset	Livestock	Census	Distribution and abundance of livestock species	11 to 14 livestock production system classes	Versions: 2007, 2011	Global: 0.0083◦ (~1 km by 1 km)	The 2011 version includes more accurate and higher spatial resolution inputs than the previous	Lack of detail on mixed crop-livestock systems
GLW[57,58]	Gridded dataset	Livestock	Census	Distribution and density of livestock	8 livestock classes	Annual: 2010 and 2015	Global: 5 arc min (~10 km by 10 km)	The most comprehensive global livestock dataset	The date, quality, resolutions, and timeliness of census data used are highly variable
Herrero et al.[56]	Gridded dataset	Livestock	Census and model	Biomass use, production, feed efficiencies, and greenhouse gas emissions	8 livestock production systems, 4 animal species, and 3 livestock products	2000, 2005, 2010	Global: 0.0083◦ (~1 km by 1 km)	A globally comprehensive dataset on livestock biomass use and feed efficiency	Limited temporal and spatial coverage

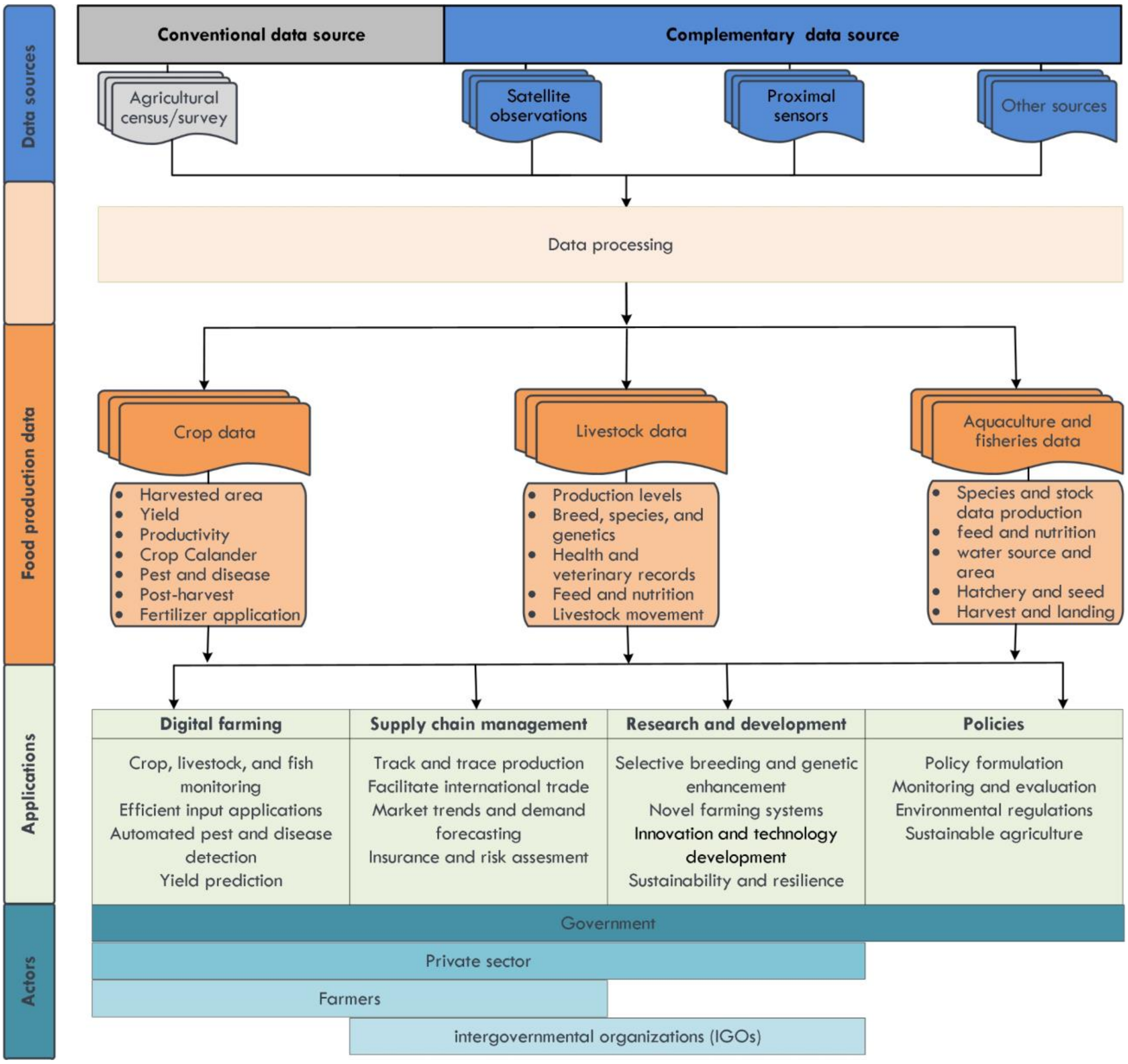

Figure 1. The centrality of food production data. Food production statistics are diversely sourced and are central to a wide array of applications and decision support for multiple actor groups.

Accepted Manuscript
Version of Record at: https://doi.org/10.1038/s43017-024-00516-2

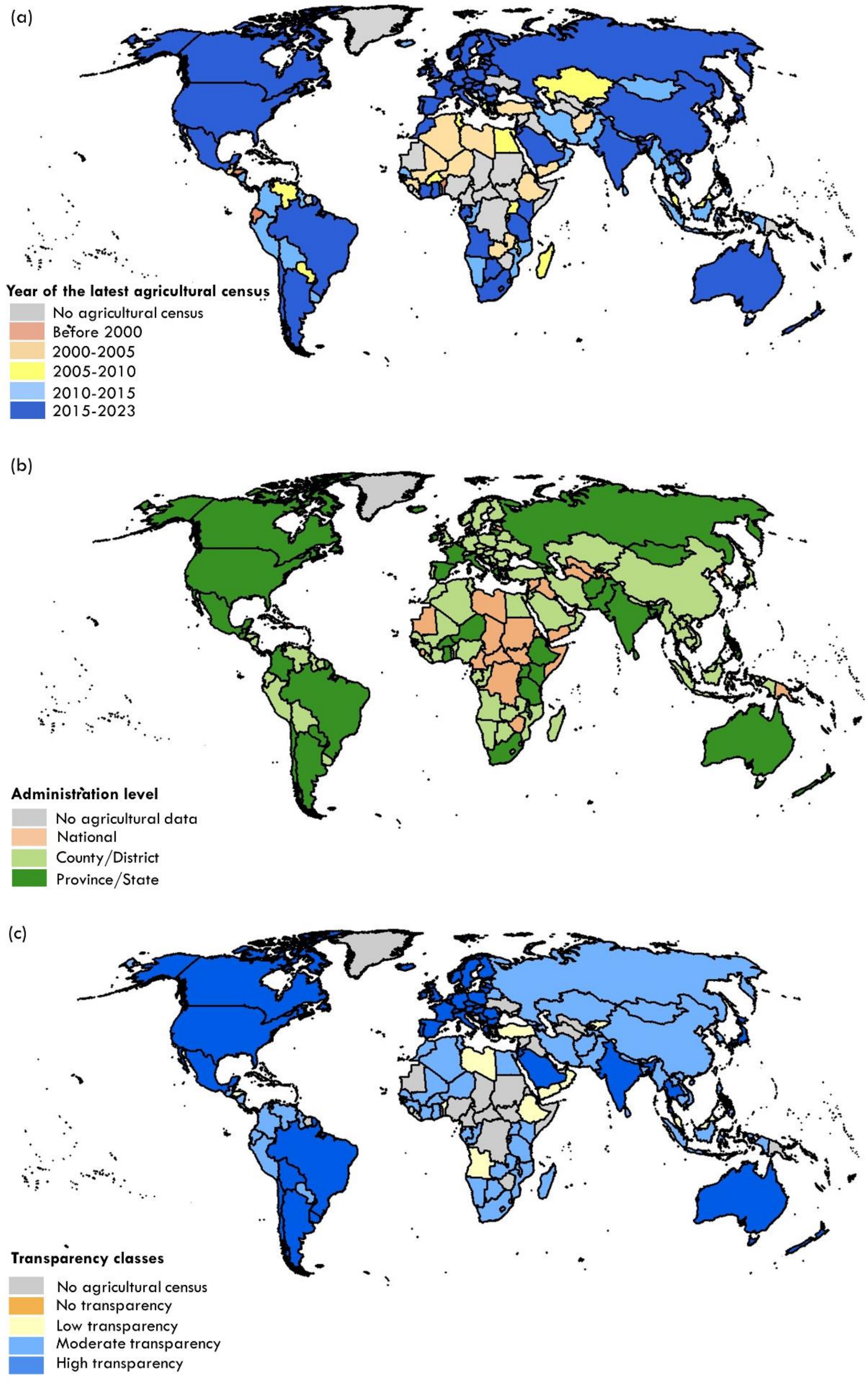

Figure 2. The current state of agricultural census information. Panels show (a) the year of the latest agricultural census, (b) the finest administration level of publicly and easily accessible statistics, and (c) the transparency of the agricultural census as evaluated based on FAIR criteria (see Supplementary Information for description of transparency rating). The year of the latest publicly available agricultural census was assessed primarily through the WCA portal; we did not consider agricultural census reports that are not publicly available, even if they are known to have been conducted. To be considered available, a minimum requirement is the presence of metadata that provides an overview of the census reports.

Figure 3. Identified challenges and potential solutions to address global food production data scarcity.