Info Note

Accelerating understanding of resilient agriculture

Using a data sprint to catalyze new insights from existing evidence

Kristal Jones, Andreea Nowak, and Todd Rosenstock

NOVEMBER 2021

Key messages

- The Evidence for Resilient Agriculture (ERA) dataset contains more than 100,000 observations describing which agricultural practices work where.
- A Data Sprint was held for six months to leverage ERA to generate new insights that maximize the impact of the dataset and build new scientific collaborations.
- A sample analysis that compared expert elicitation and ERA data of benefits of CSA identified little convergence and highlighted the need for integration of diverse evidence.
- A sample analysis that generated predictive scenarios to improve food and nutrition security in Ethiopia leveraged ERA data to identify pathways that account for local diversity.
- A sample analysis that assessed data deserts and research gaps using ERA data found a lack of evidence in regions with pastoral systems.

The Evidence for Resilient Agriculture (ERA) dataset contains over 100,000 observations from over 2,000 studies from sub-Saharan Africa (Nowak and Rosenstock 2020). ERA is structured to provide evidence on the observed impacts of agricultural management practices with potential to increase productivity, build resilience, and reduce greenhouse gas emissions as well as contextual factors that condition adoption and impact. The size and scope of the dataset provides opportunity to ask new questions, apply novel analytical approaches, and explore the role that data can play in innovative development practice.

To accelerate collaboration and insights that utilize ERA, a Data Sprint ('the Sprint') was organized in mid-2021. The Sprint was virtual and brought together 50+ scientists to jumpstart analyses through conversation and collaboration. Scientists worked in small teams to explore in a highly targeted way new analyses, data structures, and implementation opportunities that are possible with ERA. Over a four-month period, each collaborative team produced a proof-of-concept output that can be extended over time and that provide a blueprint to expand and refine the data and workflows that comprise ERA.

Data Sprint

A data sprint is a focused and defined time period in which individuals work in collaborative fashion to generate new analyses from existing data. A data sprint is different than a hack-a-thon, in that participants are not competing with one another but are instead working together to identify strengths and opportunities within the team and the data that can catalyze new insights.

The three main goals of the ERA Sprint were to leverage the dataset to generate new insight and information for development practice, receive feedback from technical and topical experts to improve ERA's structure and content, and to build new relationships and collaborations centered around ERA as a data source and an approach to evidence synthesis. In total, almost 40 scientists and development practitioners participated in the Sprint. Most participants worked across some aspects of ideas and data, with almost two-thirds identifying datasets and/or statistics as one of their skillsets. Half claimed a focus on big science ideas and one-third brought a focus on development applications to the table. The Sprint took place over the course of six months, from May to October 2021, and consisted of several interconnected phases.

Start of the Sprint: A two-hour virtual meeting was held to kick off the sprint process. The kick-off first focused on learning the functionality and content of the ERA dataset, and equally on getting to know who was at the table and what they would bring to collaborative relationships. The

Funded by the European Union

CGIAR

RESEARCH PROGRAM ON Climate Change, Agriculture and Food Security

CCAFS

remainder of the meeting focused on brainstorming new questions and analyses, identifying possible challenges to these ideas, and beginning to scope mechanisms for collaboration. These conversations bounced between plenaries with interactive tools (like Miro boards) to quickly share ideas among the whole group, and small discussion sessions based on shared interests.

Ideas in Action labs: Over the month following the Sprint kick-off, a series of labs were held once a week to share ideas, troubleshoot challenges (with code or other technical aspects of the dataset), and discuss progress on emerging analyses. These sessions were meant to maintain the momentum and accelerate through challenges that often slow down new analytical teams.

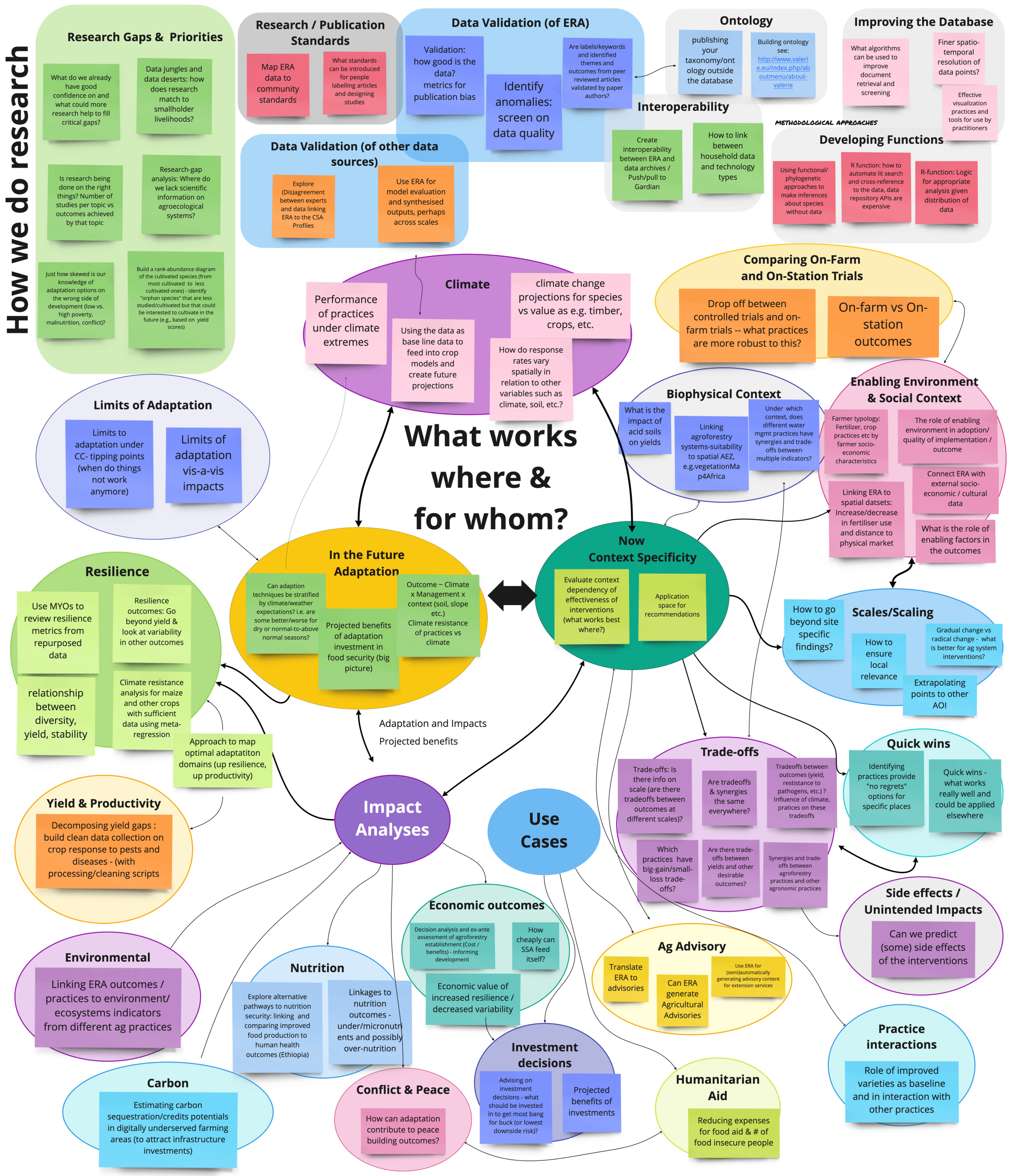

Figure 1. Collective mapping of potential questions about how to use and improve ERA data and workflows. The ideas were identified through group discussions during the Sprint kick-off meeting.

Ongoing and multifaceted communication: At the end of the kick-off meeting, participants began to self-identify into teams with shared interests and complementary skill sets. The kick-off meeting and all subsequent lab and team meetings were virtual and the ERA team created a dedicated Slack channel and invited all Sprint participants to join, as a way to facilitate momentum and quick progress through ongoing conversation. Code sharing and collaborative workflows happened through GitHub.

The ideas

The core the Sprint kick-off meeting was a wide-ranging discussion about two overarching ideas: what works where and for whom, and how to leverage the ERA data and infrastructure to answer those questions. Figure 1 shows a mind map of all possible questions, topics and themes identified through group discussion.

- Research gaps and priorities: What high-priority gaps in understanding of relationships between production practices and development outcomes can be addressed by analyses of ERA?
- Research standards: What standards are needed to ensure that ERA data are appropriately used and that new data can be included in ERA in the future?
- Socioeconomic outcomes: What are the impacts on nutrition and human health, livelihoods and conflict of production practices meant to build resilience?
- Data validation: How can ERA analyses improve the dataset to ensure that the evidence base included in the dataset is clean, well-structured and unbiased?
- Transparent data architecture: How can reproducible workflows and open-source tools be used to improve the database and support collaborative analyses?
- Interoperability: How can the structure and documentation of the ERA dataset and underlying data ontology be shared to maximize interoperability?
- Environmental outcomes: What are the impacts on emissions, sequestration, water, and soil health of production practices meant to build resilience?
- Projected and observed benefits: Which production practices positively impact yield, emissions and resilience? Are there trade-offs among these impacts or places for quick wins?
- Context specificity: How do the impacts of production practices meant to build resilience vary by social and environmental context?
- Use cases: How could findings from Sprint analyses be used in development practices?
- Scales and scaling: How can ERA data be scaled up to look beyond specific sites and inform policy and practice and scaled down to link to household data and ensure local relevance?

Sample analyses

After the Sprint kick-off meeting, teams were formed around questions inspired by the ideas generated through group discussion. Each team spent the remainder of the Sprint (about five months) working on analyses to drive answers to one or more specific question. Here we briefly outline a few selections of results.

Disagreement between experts and data: Expert opinion is often used to meet the demand for information on the performance of agricultural technologies, despite known bias and lack of scalability with this approach. This team sought to answer questions about how similar or different expert assessments are when compared to data from farm trials. Linking ERA to other data sources allowed for a detailed comparison to assess whether there more agreement for some types of practices than others.

A Sprint team combined ERA with the Climate Smart Agriculture Profiles ("CSA Profiles") database (Nyakundi et al. 2020), which uses expert elicitation to estimate CSA benefits and contains more than 1600 datapoints and 15 outcome indicators across three continents. To perform the analysis, data were subset and merged based on countries, practices, commodities, and outcomes that matched across the two datasets. The team built a linear regression model to understand whether there is a tendency to over- or under-estimate benefits across certain practices or commodity groups.

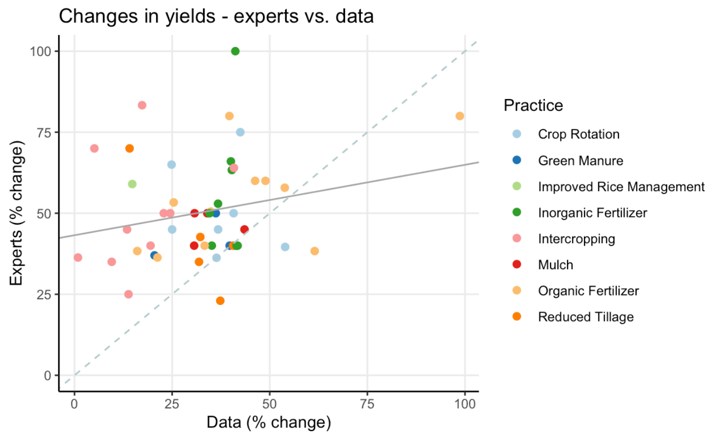

Figure 2. Consensus on yield benefits across expert assessments and data from trials, by practice and commodity groups. The dotted 1:1 diagonal line represents the scenario in which expert assessment and trials data match perfectly.

Findings showed only a weak correlation between expert and evidence-driven estimates of relationships between practices and yield benefits and experts tend to overestimate benefits (r=0.2). Implications of the analysis include the highlight the pitfalls of relying on expert judgement and the need to calibrate expert opinions when evidence is lacking. The harmonized dataset was

relatively small. There is an opportunity to increase data interoperability, particularly through harmonized taxonomies and data structures, which would also for more breadth and depth in comparisons across different types of evidence.

Modeling pathways to food and nutrition security: In Ethiopia and many other sub-Saharan African countries, food production has increased over the past several decades but undernourishment remains high and is projected to increase, with climate variability being a key driver of undernourishment (FAO et al., 2020). Ethiopia's national agricultural plans and investments have focused on building adaptive capacity in specific cropping systems or with specific practices and technologies. However, the wide variation in agroecological zones and farming systems in the country makes it difficult to design comprehensive food and nutrition security pathways using only the local evidence base.

To address gaps in the local evidence base, Damerau et al. (in preparation) combined ERA data with other national and global data sources to build predictive models about the potential impacts on food production and undernourishment across agroecological zones and farming systems in Ethiopia. ERA data provided information on relative change in crop yield, biomass yield, rainfall use efficiency and total production costs for 15 practices and 22 crops. The team built a linear optimization model to generate scenarios that maximize caloric production under various natural resource and social constraints that reflect minimum targets by outcome.

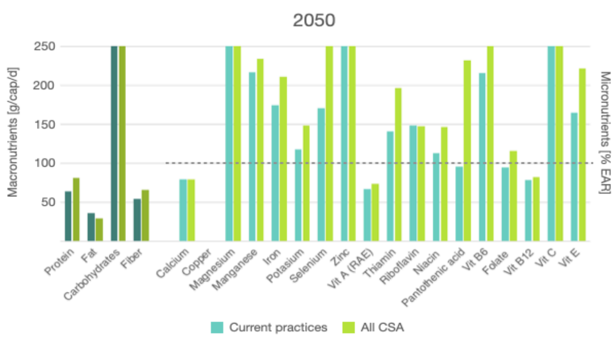

Figure 3. The combined CSA scenario provides sufficient food energy in 2050, while the current practice scenario does not. In both scenarios nutrient supply declines relatively to 2018 but the combined CSA scenario provides a more adequate nutrient supply than the current practice scenario by 2050.

The resulting predictions from two different scenarios, presented in Figure 3, show that current practices do not and will not meet micronutrient requirements now nor in the future, and will fail to meet basic energy requirements by 2050. In contrast, the combined CSA scenario (which includes six CSA practices shown to have the strongest impact on desired food security and environmental outcomes) will meet basic energy requirements and more micronutrient targets now and in 2050. By drawing on ERA data, the research team was able to build these and other predictive scenarios that can estimate not only comparative impacts but also tradeoffs over time and space by drawing on a much larger evidence base than what is available in any specific location.

Data deserts and research gaps: The thousands of agricultural trials and research studies conducted across sub-Saharan Africa since the 1970s have not been evenly distributed by geography or type of cropping system. Instead, some areas and farming systems have a large and dense evidence base on what types of agricultural practices can improve different priority outcomes, while other areas have little to no scientific evidence base. Given that agricultural research for development funding is not unlimited and prioritization is a necessity, it is important to identify specific geographies, cropping systems and communities that could benefit from increased research activity to improve outcomes.

Using the ERA dataset, Lamanna et al. (unpublished) mapped the existing evidence base by geography and major cropping system to characterize overall data density and identify data deserts. The team also characterized research effort (number of studies) by farming system and compared the distribution of research effort across farming systems to the total area represented by a given system and the population density of each farming system region.

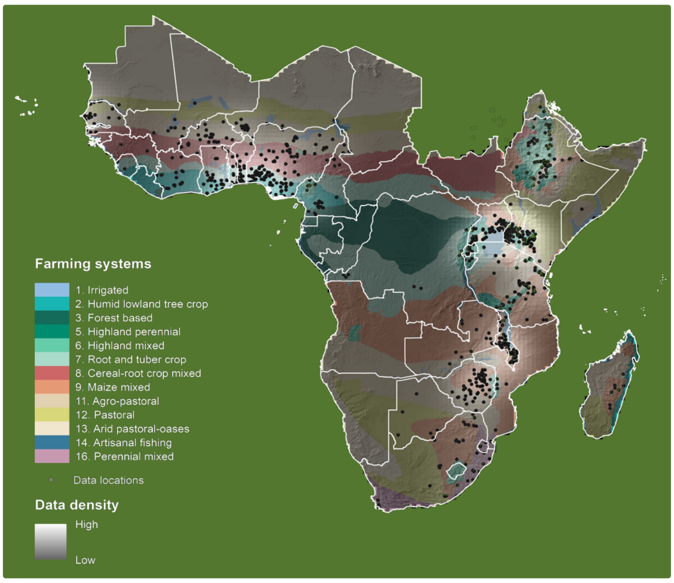

Figure 4. Location and density of studies in sub-Saharan Africa that focus on adaptive potential and CSA pillars (figure adapted from Hickson et al. unpublished)

Figure 4 shows distribution of the current evidence base for CSA impacts by geography and farming system type. Maize-mixed farming systems have the largest research effort, focused on specific geographies, although these

systems are the largest type of farming system by total area. Each type of pastoral system (agro-pastoral, pastoral and arid pastoral-oases) covers almost as large an area as maize-mixed systems. However, research effort on these systems is smaller than that on almost all other farming systems. Analyzing the ERA data to produce findings like those in Figure 4 can highlight equity issues in the lack of an evidence base from which to set adaptation plans and priorities, as well as identify research gaps and opportunities.

Way forward

The ERA dataset and workflows continue to be refined with the goal of creating a living dataset and evidence synthesis that can be updated and expanded as the evidence base on resilient agriculture grows. The Data Sprint focused on generating analyses that provided a 'proof of concept' of different possible uses of the ERA data alone and in combination with other data sources to ask a variety of development and methodological questions. Results presented in this InfoNote highlight the potential outputs and uses of analyses that can look broader over space and time using ERA data than is often possible when drawing on only the local evidence base.

Going forward, the ERA team will build on lessons learned from the Sprint in terms of the types of experts and types of questions best suited to the ERA data, as well as the challenges of integrating ERA data with other sources of information. As workflows for adding new evidence, analyzing the data and identifying intersection points with other data sources are improved, the ERA data will become increasingly assessable to researchers and practitioners.

Further reading

- FAO, IFAD, UNICEF, WFP and WHO. 2020. The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets. Rome, FAO. DOI: 10.4060/ca9692en.
- Hickson K, Jarvis A, Rosenstock T, et al. unpublished. Digital Atlas of Adaptation in Africa: A Brief Compendium of Challenges and Opportunities. International Center for Tropical Agriculture (CIAT).
- Nowak A, Rosenstock T. 2020. Evidence for Resilient Agriculture (ERA): What is it?. CCAFS.
- Nyakundi F, Osiemo J, Grosjean G, et al. 2020. Climate Smart Agriculture Country Profile Data. DOI: 10.7910/DVN/FO42EP.

This publication and the Data Sprint were developed with the financial support of the European Union. Its contents are the sole responsibility of the authors and do not necessarily reflect the views of the European Union.

Kristal Jones *(kristal@jgresearch.org) is a Principal Researcher, JG Research and Evaluation.* ***Andreea Nowak*** *(a.nowak@cgiar.org) is Research Associate with World Agroforestry Centre (ICRAF).****Todd Rosenstock*** *(t.rosenstock@cgiar.org) is Principal Scientist, the Alliance of Bioversity International CIAT.*