ANALYSIS AND EVALUATION OF SOCIO-TECHNICAL INNOVATION BUNDLES DASHBOARD

George Thomas Muteti[1], Cosmas Lutomia[1], Eileen Bogweh Nchanji1, Dessalegn Ketema[1], Leroy Mwanzia[1], Ranjitha Puskur[2], Everisto Mapedza[3], Hom Gartaula[2], Prama Mukhopadhyay[2]

[1]Alliance of Bioversity International and the International Center for Tropical Agriculture, Nairobi, Kenya
[2]International Research Rice Institute, Delhi, India
[3]International Water Management Institute, Pretoria, South Africa

06 September 2023

ACKNOWLEDGEMENT:
We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund.

Contact person: Nchanji Eileen Gender and Social Inclusion Expert, CIAT Regional Office for Africa, Duduville Campus, off Kasarani Road, P.O. Box 823-00621, Nairobi, Kenya

Design and layout: Denet Odhiambo, Bungu Visuals

TABLE OF CONTENTS

CITATION

George Thomas Muteti, Cosmas Lutomia, Eileen Bogweh Nchanji, Dessalegn Ketema, Leroy Mwanzia, Ranjitha Puskur, Everisto Mapedza, Hom Gartaula, Prama Mukhopadhyay. (2023). *Analysis and evaluation of socio-technical innovation bundles dashboard.*

TABLE OF FIGURES

LIST OF TABLES

ACRONYMS & ABBREVIATION

ABC	Alliance of Bioversity International and the International Center for Tropical Agriculture
AI	Artificial Intelligence
API	Application Programming Interface
CGIAR	Consultative Group for International Agricultural Research
CIAT	International Center for Tropical Agriculture
CIMMYT	International Maize and Wheat Improvement Center
CIP	International Potato Center
CRISP-DM	Cross-Industry Standard Process for Data Mining
DOI	Digital Object Identifier
GENDER	Generating Evidence and New Directions for Equitable Results
HTML	HyperText Markup Language
ICARDA	International Center for Agricultural Research in the Dry Areas
IFPRI	International Food Policy Research Institute
IITA	International Institute of Tropical Agriculture
IRRI	International Rice Research Institute
STIBs	Socio-Technical Innovation Bundles
UN	United Nations
UNICEF	The United Nations Children's Fund
URL	Uniform Resource Locator
WP	Work package
WWW	World Wide Web

EXECUTIVE SUMMARY

This executive summary provides an overview of an Inventory of available gender and socio-economic datasets, data and tools that could facilitate bundling. To achieve this output, we; (1). Collected and curated relevant information on STIBs that promote women's empowerment and resilience. (2). Designed the layout to effectively visualize the content and demonstrate the platform's functionality. (3). Present the design and layout of available STIBs datasets (4). Define and cluster different socio-technical bundles in the available gender and socio-economic datasets and visualize them. Work Package 2 (WP2) of the CGIAR Global Initiative on Gender Equality (HER+) seeks to identify and co-design socio-technical innovation bundles (STIBs) for women's empowerment and resilience. The output of this activity is led by the ABC in collaboration with (IRRI) and the International Water Management Institute (IWMI). This output report details the inventory of gender and socio-economic datasets, data, and tools that can be used to design inclusive STIBs. These bundles are categorized and clustered based on different socio-technical factors. In addition, a dashboard that includes an interactive map and repository has been developed to visualize this information. Using the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, a structured and comprehensive approach was applied to collect and analyze relevant information in a systematic and robust way. The methodology comprises five key phases: data discovery, preparation, modeling, evaluation, and deployment.

A total of 194 applicable datasets from the CGIAR GENDER (Generating Evidence and New Directions for Equitable Results) Impact Platform were examined. From these initial datasets, 84 were selected for further screening, exploring in detail associated data collection tools, data, training manuals and project documents. After additional analysis and examination, 49 datasets became the focus of our analysis of STIBs. The findings provide important insights. Based on the geographic distribution of innovation bundles by region, South America was the region with the highest number (28). Across regions and at the country level, Peru had the highest number of innovation bundles (25), followed by Uganda (6). The bundles were further categorized by CGIAR Impact Areas, with Gender Equity, Youth, and Social Inclusion having the largest number of innovation bundles (35). The quantification of the 49 socio-technical innovation bundles revealed variations in their counts based on their respective categories, classes, and clusters. The most frequently mentioned technological innovations were seeds, fertilizers, irrigation, and water management. On the other hand, the most frequently mentioned social innovations were the Internet, training, extension programs, group membership, finance, and participatory variety selection. However, this analysis was constrained by limited access to STIB resource files, by incomplete metadata, and by the absence of an Application Programming Interface (API) for exploring datasets that contain socio-technical innovation bundles.

In conclusion, output of this activity contributes to knowledge on how existing STIBs can contribute to women's empowerment and resilience by leveraging the CRISP-DM methodology. To build on this output and contribute to the toolbox under development, we will integrate data from the learning labs across Asia and Africa in the coming months, explore additional datasets from socio-economic platforms, link the toolbox with CGIAR crop observatories, incorporate Artificial Intelligence (AI) and machine learning, and develop a time series capability.

INTRODUCTION

Agri-food systems in developing countries face numerous challenges, of which climate change has recently become a prominent one. In this context, inclusive solutions are critical for ensuring that products and benefits created by the agricultural innovation system are accessible to both women and marginalized groups (Beuchelt & Badstue, 2013). The Gender Equality (HER+) Initiative addresses the four main dimensions of gender inequality in agri-food systems to build climate change resilience and empower women. These dimensions include voice, agency, gender norms, and policy.

In the context of the Initiative on Gender Equality, socio-technical innovation bundles (STIBs) are a combination of technological and social innovations that are co-designed to address women empowerment and resilience. STIBs are demand-driven and context-specific, meaning that they are tailored to the specific needs and constraints of women farmers in a particular setting. Therefore, inclusive and equitable STIBs play a critical role in addressing these challenges by enabling resilient agri-food systems (Nchanji et al., 2023). To achieve healthy, equitable, resilient, and sustainable outcomes, STIBs should be co-created and designed to fit the specific context of the agri-food system (Barrett et al., 2022). As the development and implementation of STIBs progress, they hold the capacity to substantially contribute to both global gender equality and food security.

The goal of this effort is to systematically gather evidence on the effectiveness of STIBs in enhancing women's empowerment and resilience from existing socio-economic and gender data. To support the Gender Equality (HER+) initiative's socio-technical bundling goal, relevant datasets were obtained from the CGIAR GENDER Platform.

The specific objectives of this activity are:

1. Collect and curate relevant data on STIBs that promote women's empowerment and resilience.
2. Develop an inventory of available gender and socio-economic datasets and tools that could facilitate aggregation.
3. Define, cluster and visualize the different socio-technical bundles among the available gender and socio-economic datasets.
4. Develop an interactive map showing the distribution of socio-technical innovation bundles across various countries and regions.

METHODOLOGY

We followed a structured and comprehensive methodology called the Cross-Industry Standard Process for Data Mining (CRISP-DM). The CRISP-DM methodology encompasses five key phases: data exploration, preparation, modeling, evaluation, and deployment. Informed by scientific concepts and methodologies, CRISP-DM is based on industry best practices. It draws on established principles from statistics, machine learning, and data analysis, incorporating them into a structured framework that can be applied in real-world scenarios. The CRISP-DM methodology is widely recognized and highly regarded for its ability to effectively guide data mining projects, combining scientific principles with practical considerations. This methodology was specifically tailored to the HER+'s objectives, allowing for the identification and assessment of innovation bundles that have the potential to empower women and enhance their resilience. It guided the analysis and evaluation of the identified STIBs in a well-defined manner. By using CRISP-DM, we ensured that findings were based on a systematic and rigorous approach. The methodology consists of five steps tailored to the objectives:

1. **Data Mining:** This initial step involves acquiring and extracting relevant data and studies from the CGIAR GENDER platform.
2. **Interrogation and Discovery:** In this stage, the resource files are thoroughly examined and analyzed to better understand the content. This process includes filtering and refining the data to focus on the studies of particular interest.
3. **Identification:** The identification phase entails recognizing and pinpointing the selected studies' technology practices and social innovations. This step identifies key components and factors contributing to the objectives.
4. **Clustering:** This stage involves categorizing datasets into three types of agri-food system innovations: technological[1], technical[2], and social[3], (Barrett, 2022)[4]. Grouping the data into categories enables further analysis to better understand the different types of innovations in diverse contexts.
5. **Socio-Technical Innovation Bundling:** Is the process of combining technological innovations with technical and social innovations to address complex challenges and achieve more sustainable and inclusive outcomes in agri-food systems.
6. **Bundle:** A specific combination of technological, technical, and social innovations designed to deliver context-specific targets. For instance, a technological innovation (e.g., biofortified

1 **Technological innovations:** Products, tools, or practices that can be used to achieve a specific objective, such as increased productivity, food security, and nutrition. Examples include crop varieties, animal breeds, crop and livestock management practices, machines and equipment, processing technologies, and digital applications.

2 **Technical innovations:** Innovations that strengthen capacity, including farmer, extension, or investor decisionsupport services. Examples include training programs, manuals, guides, animations, and incubator programs. Technical innovations enable the necessary knowledge and information to optimize existing technological innovations' adoption, usage and benefits.

3 Social innovations: Innovations that create enabling conditions for technological and technical innovation use. Examples of social innovations include farmer memberships to organizations, policy and institutional arrangements, business models, strategies, concepts, finance, partnership models, and public/private delivery strategies.

4 Sxxxxxxxxx.

drought tolerant bean variety) might be combined with gender sensitive nutrition training programs (technical innovation) and group-based aggregation and marketing of grain (social innovation) to improve resilience to climate change, nutrition, and women empowerment.

The given illustration below shows a context diagram outlining the five key steps in implementing the methodology.

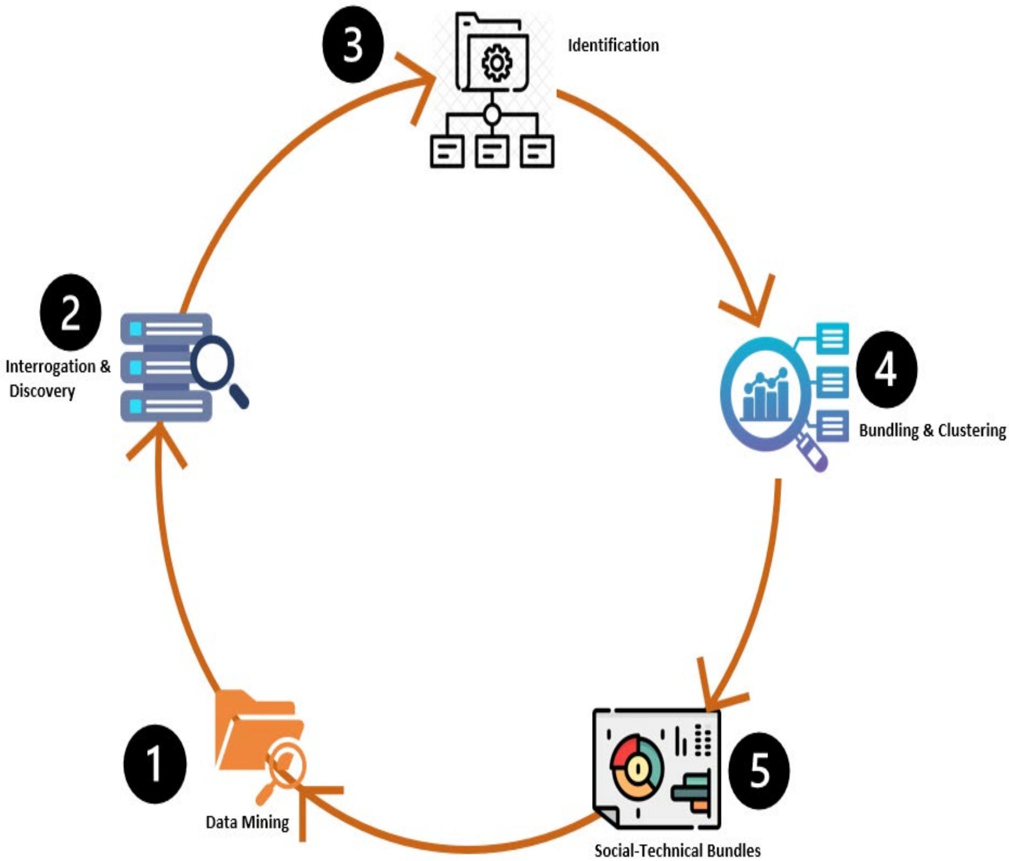

Figure 1: Context diagram

Data Sources

The CGIAR GENDER Platform was the primary source of the collected datasets. GENDER is a CGIAR impact platform that synthesizes and amplifies research, fills gaps, builds capacity, and sets directions to enable the CGIAR to have maximum impact on gender equality, opportunities for youth, and social inclusion in agriculture and food systems (CGIAR Gender Platform, 2023). The GENDER Platform contained valuable information pertinent to HER+ WP2 saving time and resources for other outputs and providing baselines and benchmarks for new datasets. Additionally, the platform's data quality was well known and documented, facilitating data mining and inventory creation.

Additional datasets were collected from collected from other platforms. We used web scraping, or web extraction or harvesting, to extract data from the World Wide Web (WWW) and save it to a file system or database for later retrieval or analysis (Zhao, 2017). Web scraping involves the programmatic extraction of data from websites or web pages, using language code to automate the data collection process. The first step involved identifying the GENDER platform as the target URL for scraping. Next, the structure of the GENDER platform website was inspected to gain a comprehensive understanding of its configurations, with a specific focus on APIs. Subsequently, Python code was written to facilitate the scraping process. Once the code was prepared, it was deployed to initiate the scrape request and gain access to the target Uniform Resource Locator (URL). Regular consultations were held with HER+ work package 2 team to ensure the extracted data's format met their requirements. The cleaned data was then stored for further analysis. The process was iterated upon, incorporating additional input from the team. Finally, the stored data was programmatically queried to extract valuable insights.

Data Exploration

Apache Superset – a user-friendly interface for creating interactive dashboards, exploring data, and generating visualizations – was used to implement the first stage of the methodology (Foundation Apache Software, 2021). The CCGIAR GENDER Impact platform had an extensive collection of 195 datasets. Upon closer inspection, we found that 85 of these datasets had data collection tools, data, training manuals and project documents included. Out of 85 datasets, 49 of them had different socio-technical innovation bundles.

The remaining 36 datasets were unsuitable for due to the following reasons:

1. Some datasets contained irrelevant resource files from international development agencies like the UN and UNICEF in Excel format.
2. Some datasets were duplicates.
3. Some dataset documentation did not mention innovations or technology.
4. HTML resource files were unusable in some datasets.
5. Some which were not relevant to the socio-technical innovation bundling objectives.

Constraints and Assumptions

Constraints

One of the main constraints was that half of the studies had resource files with limited access. Additionally, many studies either lacked resource files or had incomplete metadata, omitting critical details useful for identifying STIBs. Lastly, lack of APIs in the CGIAR's GENDER platform meant that data could not be easily queried, a major constraint to the collection of the datasets.

Assumptions

A key premise of the methodology is that the resource files related to the identified innovation bundles would not be stored in an inaccessible toolbox. Instead, these files can be easily accessed using a Digital Object Identifier (DOI).

RESULTS AND SUMMARY

This section summarizes the results and highlights key insights and trends discovered during the data exploration and analytics phase for the socio-technical innovation bundles.

Results

This first section of results and findings describes when and where the datasets were studied, and how they were classified and categorized. It's worth keeping in mind that the number of datasets identified in the documentation of existing datasets may differ due to various reasons. These include instances where the same technology and innovation bundle[5] dataset is found in different regions or countries, datasets that fall under multiple CGIAR impact areas, instances where providers collaborate to create a single dataset, and clustering of datasets across multiple factors.

Distribution of datasets per release year

Based on the data, the distribution of datasets span nine years, from 2012 to 2021. In 2017 there were 22 datasets which is the highest number in any given year, because for this particular year there were CGIAR Research Programs (CRPs) introduced. The graph below depicts the yearly distribution of datasets.

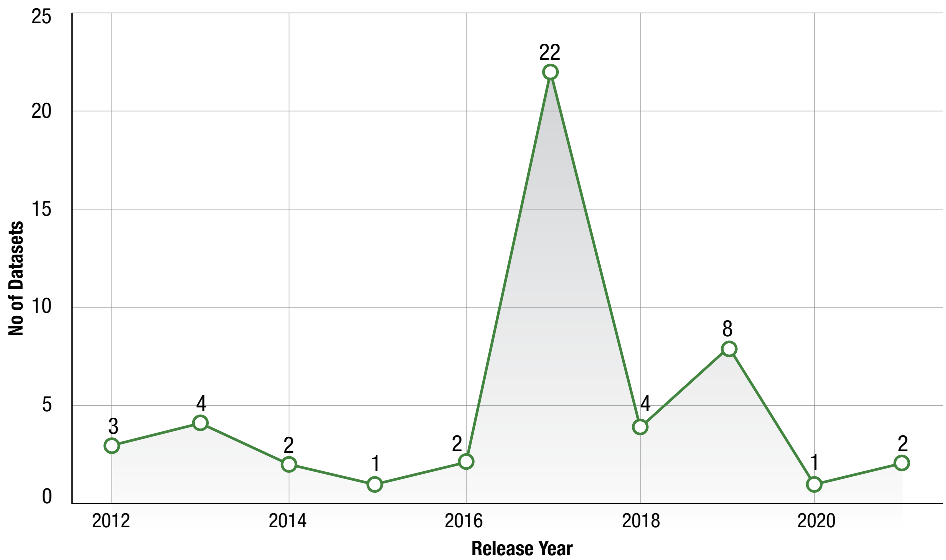

Figure 2: Distribution of datasets per release year

5 Technology and innovation bundles are comprehensive combination of specific inputs/tools, offered to farmers along with the necessary knowledge, information, and processes to facilitate their adoption, usage, and ultimate benefits. (Nchanji, et al., 2023)

Distribution of datasets per region

Regarding distribution per region, most datasets were found in the South American region, with Eastern Africa coming in second. Southern Asia took the third spot, while the remaining regions only had one dataset each. However, only one dataset fell under four different regions. The table below shows the distribution of datasets across regions.

Table 1: Distribution of datasets per region

REGIONS	NUMBER OF DATASETS
South America	28
Eastern Africa	10
Southern Asia	4
Middle Africa	1
Western Africa	2
Southern Africa	1
Southern Asia; Northern Europe; Australia and New Zealand	1
Central America	1
Southern Asia; Eastern Africa	1
Grand Total	**49**

Dataset distribution map.

According to the distribution map, Peru emerged as the country with the largest number of datasets, establishing South America as the region with the highest dataset count. Importantly, several datasets were attributed to multiple countries, contributing to variations in the total dataset count. While the map visualization primarily focuses on a country-level view, it's worth noting that the data is available in detail up to administrative level 3 within each country. The figure below illustrates this dataset distribution.

COUNTRY	NO. OF DATASETS
Peru	25
Uganda	6
Ethiopia	2
K enya	3
Bangladesh	5
N epal	2
Ecuador	1
DRC Congo	1
G uatem ala	2
Bolivia	1
Mali	1
Malawi	1
Tanzania	1
Columbia	1
India	3
Senegal	1

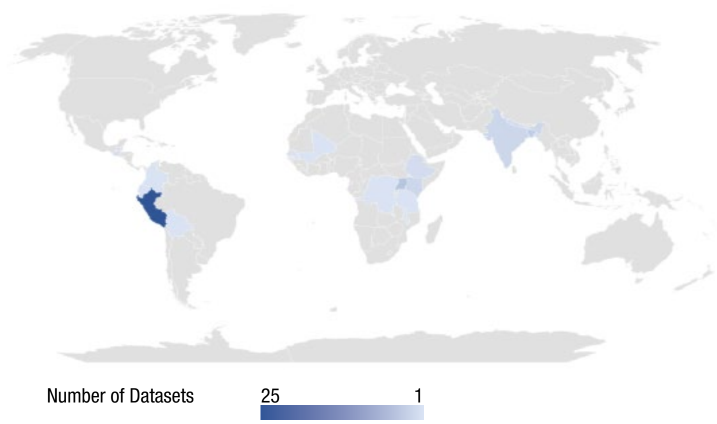

Figure 3: Dataset distribution map

Classification of datasets per CGIAR impact areas

The datasets were systematically classified according to CGIAR's impact areas, with the most substantial representation (35 datasets) observed within the categories of Gender Equity, Youth, and Social Inclusion. The following graph delineates the distribution of these datasets across CGIAR's five impact areas.

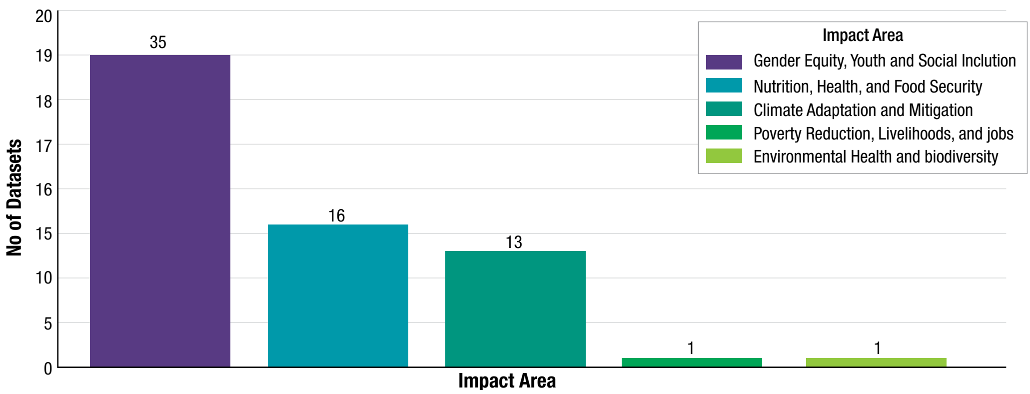

Figure 4: Distribution of datasets per CGIAR impact areas.

Categorization of datasets per providers

Multiple dataset providers contributed to the CGIAR impact platform. Among them, CIP stands out as the largest contributor, followed by IFPRI, with CIAT coming in last. The remaining providers each contributed only one dataset. Interestingly, there's a unique case where one dataset had contributions from more than one provider. Below, you'll find a figure categorizing the datasets based on their respective providers dataset.

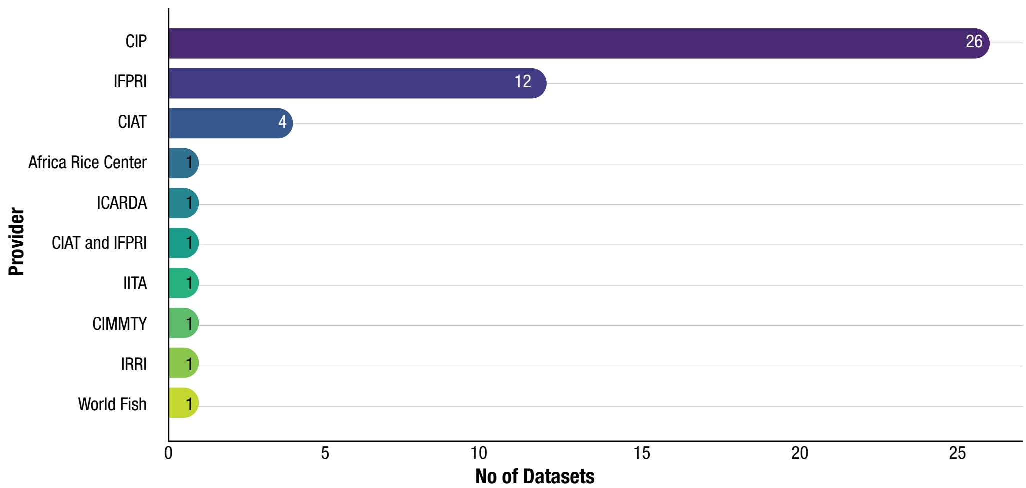

Figure 5: Distribution of datasets per provider

Clustering of datasets into innovation categories.

The datasets were clustered into three distinct innovation[6] categories: social, technical, and technological. Interestingly, a majority of the datasets belonged to multiple categories, with social and technical categories topping the charts with 24 datasets. The three categories combined had 12 datasets, while social and technological categories had 10. On the other hand, technical and technological categories had only 1 dataset. The graph offers a visual representation of the distribution of datasets based on the categories.

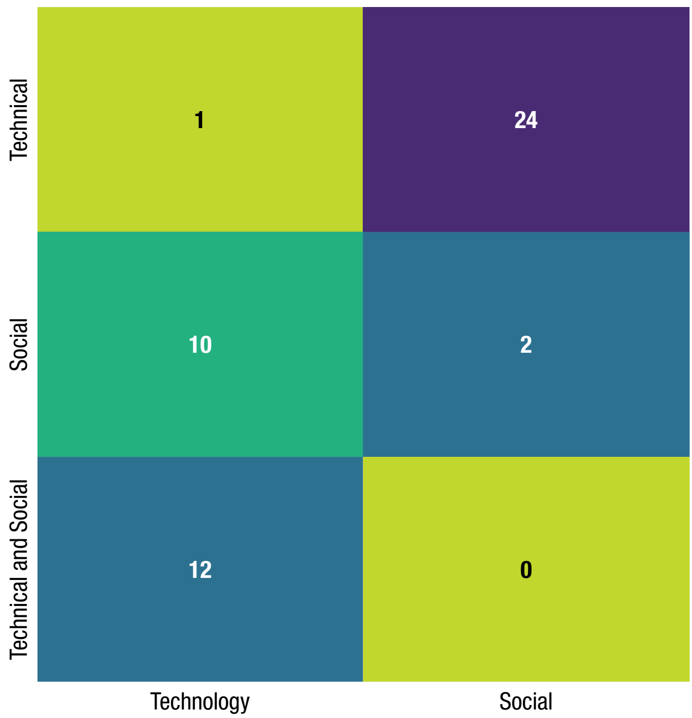

Figure 6: Cluster of datasets into Social, Technical and Technology categories

Clustering of datasets into innovation categories with regions.

The datasets were organized into clusters based on a combination of innovation categories and regions. The innovation categories included social, technical, and technology, while the regions were Sub-Saharan Africa, Americas, Southern Asia, and Australia and Europe. The results showed that the combination of social and technical categories had the highest number of datasets (20) in the Americas region. The graph below was used to visually represent the distribution of datasets based on innovation categories and regions.

6 An Innovation are products (Seed, breed, machines), agronomic practices, processes, methods and models that are designed to improved agricultural productivity.

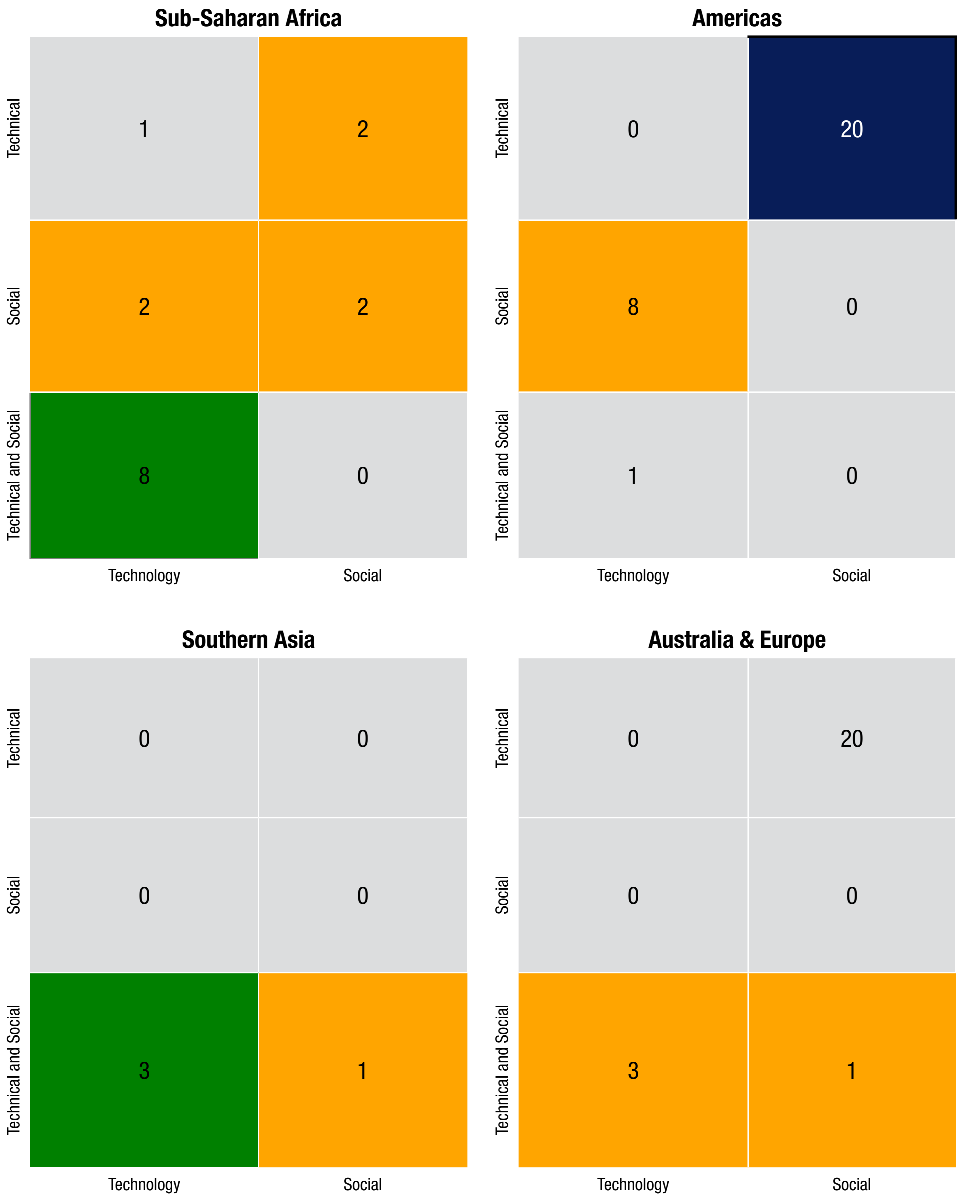

Figure 7: Clustering of datasets into innovation categories and regions

Innovations identified.

Diverse innovations emerged from the datasets, and upon careful analysis, it became evident that seed technology was a predominant theme across the majority of these datasets. Additionally, we identified several other notable innovations, such as Participatory variety selection, irrigation techniques, fertilizer usage, water management practices, training sessions, extension services, group memberships, financing methods, and the utilization of the Internet. The accompanying graph below showcases the top technologies that were identified.

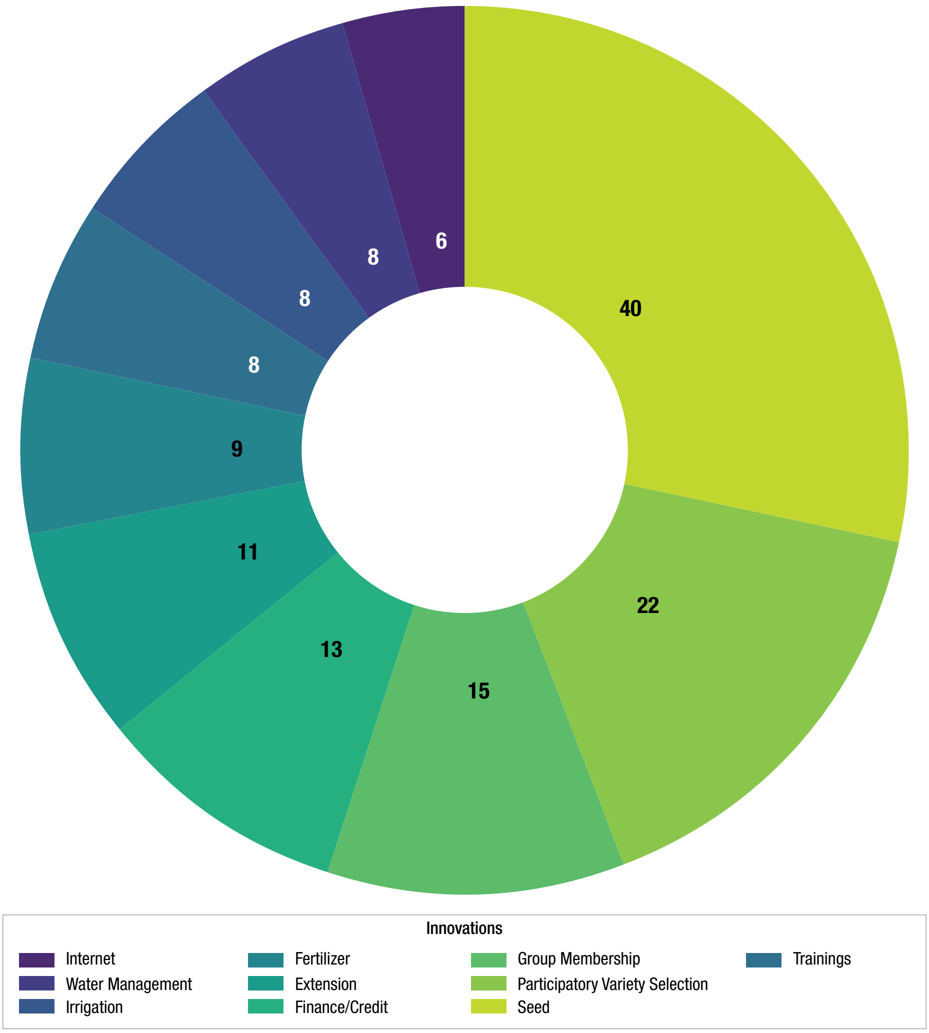

Figure 8: Top innovations identified.

Analytics Dashboard (Toolkit)

An analytics dashboard is a graphical representation of data that helps you quickly monitor and comprehend information. The output includes an analytic dashboard that dynamically displays the results and findings discussed earlier. The Analytics Dashboard has several features, including:

- An interactive map showing the distribution of socio-technical innovation bundles datasets across various countries and regions.
- Socio-technical Innovations bundles are added as data is being collected.
- Visualization of socio-technical innovation bundle results and findings to show their effectiveness.
- Clustering of datasets into innovation categories.
- An API is now available and accessible, allowing seamless integration and data transfer with other platforms in CGIAR.

CONCLUSION AND RECOMMENDATIONS

Summary

Based on the results and findings presented above, the following key insights and trends can be observed:

- The CCGIAR GENDER Impact platform boasts a vast collection of 195 datasets, with 85 containing resource files. Of these 85, 49 have led to socio-technical innovation bundles.
- Datasets were released between 2012 and 2021, with 2017 being the highest with 22 bundles and 2020 being the lowest with only one dataset. It's interesting to note the variation in releases over the years.
- Regarding datasets, Peru has the highest number with 25, followed by Uganda, with 6. A few countries have just one dataset, such as Ecuador, DRC Congo, Bolivia, Mali, Malawi, Tanzania, Columbia, and Senegal.
- The datasets have been categorized into their respective CGIAR impact area. The largest number of 35 datasets fall under the category of Gender Equity, Youth, and Social Inclusion. The second largest category is Nutrition, Health, and Food Security, with 16 datasets, followed by Climate Adaptation and Mitigation, with 13 datasets. The remaining categories have one dataset each.
- According to the findings, the most mentioned innovations are Participatory variety selection, irrigation techniques, fertilizer usage, water management practices, training sessions, extension services, group memberships, financing methods, and the utilization of the Internet. Seed is the most frequently mentioned, with 40 mentions.
- After clustering the datasets, it was found that Social has the largest number of bundles with 48, followed by Technical with 37 and Technology with 23.
- The CGIAR GENDER Platform lacks an API built for accessing the data resources, thus making data interoperability and accessibility a constraint.
- 110 datasets were eliminated because resource files had no DOI or Dataverse metadata available, limiting further exploration and enables innovation bundling.

Recommendations

The following recommendations are made based on the insights generated from the analysis and evaluation of the datasets:

1. To establish a robust knowledge base, the initiative's researchers should integrate the data gathered from the learning labs[7] into the existing toolbox, which also contains innovation bundles from various studies. A comparative analysis should then be conducted to determine

7 Learning labs can be understood as a multi-sectoral, multi-stakeholder and multi-disciplinary spaces where practice and research interact, and learning occurs through the intervention of STIBs.

the effectiveness of these bundles in enhancing women's empowerment.

2. We recommend adding all the resource materials from the learning labs to the analytics dashboard (toolkit). These include data collection tools, animations and graphics, datasets, reports, guidelines, and manuals.
3. We recommend the creation of a Dataverse entry for every CGIAR initiative or project that generates data. Each Dataverse dataset should have an accompanying Digital Object Identifier (DOI) for the purposes of data accessibility and transparency.
4. To facilitate data interoperability, the CGIAR GENDER platform should build an API.
5. Consider expanding the scope by integrating additional datasets from platforms like GARDIAN into the collection.
6. The toolkit should be linked with the CGIAR crop observatories to enhance its functionality further.
7. We recommend improving the toolkit by adding more functionalities, incorporating Artificial Intelligence and machine learning to streamline data generation.

Conclusion

The CRISP-DM methodology provided a structured and comprehensive approach for analyzing and evaluating socio-technical innovation bundles based on data obtained from the CGIAR GENDER Platform. The findings revealed the distribution of socio-technical innovations across regions and countries. South America led the regional distribution and Peru emerged as the country with the most bundles. In terms of temporal distribution of socio-technical innovation bundles datasets, the year 2017 was particularly noteworthy among all the years. The International Potato Center (CIP) stood out as the primary provider of datasets. The identified innovations within the datasets include seed, Participatory variety selection, irrigation techniques, fertilizer usage, water management practices, training sessions, extension services, group memberships, financing methods, and the utilization of the Internet. Participatory variety selection and seed technology was the most cited innovations. However, various limitations and challenges were encountered while collecting and analyzing the STIBs datasets. These included limited access to resource files, incomplete metadata, the absence of an API, and assumptions about the availability of resource files through DOIs. As a result, the analysis and evaluation process were constrained.

Overall, the findings provide valuable insights for designing context-specific, gender-responsive, and effective STIBs to enhance women's access, resilience, and empowerment. The CRISP-DM methodology ensures a systematic and rigorous approach, contributing to the broader understanding of women's empowerment and resilience. The findings and recommendations will be a building block for creating inclusive and equitable STIBs, potentially driving positive change locally and globally.

REFERENCES

Apache Software Foundation. (2023, July 27). What is Apache Superset? Retrieved September 12, 2023, from https://superset.apache.org/docs/intro/

Barrett, C. B., Benton, T., Fanzo, J., Herrero, M., Nelson, R. J., Bageant, E., ... & Wood, S. (2022). Socio-technical innovation bundles for agri-food systems transformation. Springer Nature. https://doi.org/10.1007/978-3-030-88802-2

Beuchelt, T. D., & Badstue, L. (2013). Gender, nutrition-and climate-smart food production: Opportunities and trade-offs. Food Security, 5, 709-721. https://doi.org/10.1007/s12571-013-0290-8

CGIAR Gender Platform. (2023). About us. Retrieved September 12, 2023, from https://gender.cgiar.org/about-us

Nchanji, E., Puskur, R., Lutomia, C. K., Dessalegn, K., Mapedza, E., Gartaula, H., ... & Barrett, S. (2023). Socio-technical innovation bundles (STIBs) for women's empowerment and resilience in the agrifood system. Alliance of Bioversity International and CIAT. https://alliancebioversityciat.org/stories/socio-technical-innovation-bundles-stibs-women-empowerment-resilience-agrifood-system

Zhao, B. (2017). Web scraping. In L. A. Schintler & C. L. McNeely (Eds.), Encyclopedia of big data (pp. 951–953). Springer. https://doi.org/10.1007/978-3-319-32010-6_483