FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

Technical Report Series No 2
Strengthening the Evidence Base for Resilience in the Horn of Africa

REPORT 11

Open Data Infrastructure for Resilience Analysis: Implementation, Examples, and Case Studies in Kenya

Jawoo Koo, Carlo Azzarri, Sara Signorelli, Maria Comanescu, and Zhe Guo

technical consortium
Building Resilience in the Horn of Africa

ILRI
INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE

IFPRI
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

USAID
FROM THE AMERICAN PEOPLE

Authors: Jawoo Koo[a], Carlo Azzarri[a], Sara Signorelli[a], Maria Comanescu[a] and Zhe Guo[a]
[a]*International Food Policy Research Institute (IFPRI).*

This report is prepared by experts for the Technical Consortium for Building Resilience in the Horn of Africa. For more information on the Technical Consortium contact Dr. Katie Downie - k.downie@cgiar.org.

Design and layout: Jodie Watt Media, Layout: Avocado Design Studio

Citation: Koo, J., Azzari, C., Signorelli, S., Comanescu, M. & Guo, Z. (2016). Open Data Infrastructure for Resilience Analysis: Implementation, Examples, and Case Studies in Kenya. Report prepared by the Technical Consortium, a project of the CGIAR. Technical Report Series No 2: Strengthening the Evidence Base for Resilience in the Horn of Africa. Nairobi, Kenya: A joint International LivestockResearch Institute (ILRI) and International Food Policy Research Institute (IFPRI) publication.

Table of **contents**

1 Introduction **1**

2 Platform **2**

Approach 2

Organization 2

Data Format 3

User Interface 4

Data Access via API 6

3 Examples **7**

Example 1: Human health data in Tharaka Nithi County, Kenya 7

Example 2 : Human health data in Samburu County, Kenya 10

4 Case Studies **15**

Study 1: Malnutrition and Climate Patterns in the ASALs of Kenya: A Resilience Analysis based on a Pseudo-panel Dataset 15

Study 2: Maintaining resilience in arid/semi-arid Kenya: a perspective on stocking rates in extensive livestock systems 21

List of **figures**

Figure 1	Landing page of the Catalog Datasets (http://data.technicalconsortium.org/dataset)	5
Figure 2	Example of the API search result formatted in JSON	6
Figure 3	Diseases dataset for Tharaka Nithi County, Kenya	8
Figure 4	Monthly cases of malaria in Tharaka Nthi county, by sub-county (filtered to show 2007 data only)	9
Figure 5	Diseases dataset for Samburu County, Kenya	10
Figure 6	Tuberculosis and malaria cases rankings of health facilities in Samburu County, Kenya	11
Figure 7	Location of health institutions in Samburu County, Kenya	12
Figure 8	Cases of worms, anemia, gonorrhea, eye infection, malnutrition by Samburu's divisions	13
Figure 9	Child moderate stunting rate (%)	16
Figure 10	Child moderate underweight rate (%)	17
Figure 11	Child moderate wasting rate (%)	17
Figure 12	Woman underweight rate (%)	18
Figure 13	Woman Average BMI	18
Figure 14	Normalized Difference Vegetation Index (NDVI)	19
Figure 15	Inverse Palmer drought severity index (PDSI)	19
Figure 16	Temperature	20
Figure 17	Rainfall (mm)	20
Figure 18	Livestock population in Garissa, Wajir, Marsabit, Mandera, and Turkana 2000 - 2012	22

1 Introduction

This report serves as documentation for the construction, operation and utility of the Horn of Africa Data Catalog (accessible online at http://data.technicalconsortium.org), which was developed to serve as a platform to provide indicators identified as critical to the work of the Technical Consortium for Building Resilience in the Horn of Africa (TC) and its partners. This technical report provides details on the implementation of resilience analytical framework using an open access data and knowledge management platform and presents examples on how to access and visualize them online through the web user interface and the application programming interface (API).

Building on the data catalogued and accessed within the platform, two case studies were developed to show how datasets from different temporal and spatial resolutions were harmonized and used to develop resilience analyses in the Arid and Semi-Arid Land (ASAL) region in Kenya. The first case study, **"Malnutrition and climate patterns in arid/semi-arid Kenya: a resilience analysis based on pseudo-panel"**, used the grid-based agro-ecological baseline indicators and the georeferenced demographic and health indicators to construct region-level pseudo-panel database and developed a spatial regression analysis to investigate the temporal trend of child and female malnutrition in climatological context. The second case study, "**Maintaining resilience in arid/semi-arid Kenya: a perspective on stocking rates in extensive livestock systems"**, also used the grid-based climatology data harmonized with household-level data to develop an ex-ante analysis framework to investigate better livestock management strategy in extensive systems.

2 Platform

APPROACH

The Catalog is built on the Comprehensive Knowledge Archive Network (CKAN) platform. CKAN is a web-based open-source data management system for the storage and distribution of various data types, ranging from well-organized spreadsheets and databases to the document and multimedia data files. The Open Knowledge Foundation (http://okfn.org) maintains CKAN's code base with the global network of developers. The system is widely used both as a public platform on Datahub (https://datahub.io) and in various national government-led data catalogues (e.g., United Kingdom – http://data.gov.uk, United States – http://data.gov). CKAN also provides a set of application programming interfaces (APIs) allowing other platforms and applications to connect and access the data easily. Owing to the fact global community continues with the active development of CKAN, it is widely used and that it possesses a flexible and open architecture, CKAN was chosen to be the platform for the Catalog.

ORGANIZATION

CKAN allows flexible (re-)organization of data contents, and this will support the dynamic nature of Consortium's evolving conceptual framework and possible future revision of its Theory of Change analyses. Currently, the Catalog includes datasets currently categorized hierarchically under groups, focus categories, pillars, sub-components, and counties.

Groups

A group is the unit where the data element originates, and this can expand further as more datasets are added to the catalog. Currently, four published *Groups* include:

- International Livestock Research Institute (ILRI)
- Demographic and Health Surveys (DHS)
- *Index-Based Livestock Insurance project (IBLI)*
- *Arid Lands Resource Management Project (ALRMP)*

Focus Categories

A Focus Category is a group of datasets that are identified as key aspects constituting the resilience in the Theory of Change analysis framework. Currently there are five Focus Categories are defined in the Catalog, including:

- *Enabling Conditions*
- *Household Characteristics*
- *Resilience Capacities*
- *Shocks and Stressors*
- *Wellbeing Outcomes*

Pillars

A Pillar is a thematic group of datasets, and it is defined as a part of the metadata of each dataset when added to the Catalog. Pillars span across the Groups and Focus Categories. Currently there are following four Pillars defined in the Catalog, but can be further expanded as diverse datasets are added to the Catalog:

- *Human Capital*
- *Demographics*
- *Risk Strategies*
- *Agro-ecological*

Sub-Component

A Sub-Component is a sub-category within the Pillar, further defining the characteristics of each dataset. Currently there are following four sub-components are defined in the Catalog:

- *Health*
- *Nutrition*
- *Education*
- *Ecology*

Administrative Units

When a dataset is specific to a sub-national administrative unit area, it can be facet-searched by selecting the name of the specific administrative unit (e.g., county in Kenya).

Tags

Additionally, specific set of keywords can tag each dataset to help describing its contents.

DATA FORMAT

The Catalog manages datasets originated from various sources, including well-organized and annotated formats (e.g., Microsoft Access and Excel), flat tabular format (e.g., comma-separated values (CSV)), and un-organized data values embedded in documents (e.g., Microsoft Word or PDF files). Whenever possible, all numeric data elements found in the original data source files were transformed into the most commonly-used tabular format, such as CSV. This process was first implemented programmatically when the source data was well structured, but mostly done manually to ensure the consistency across different datasets.

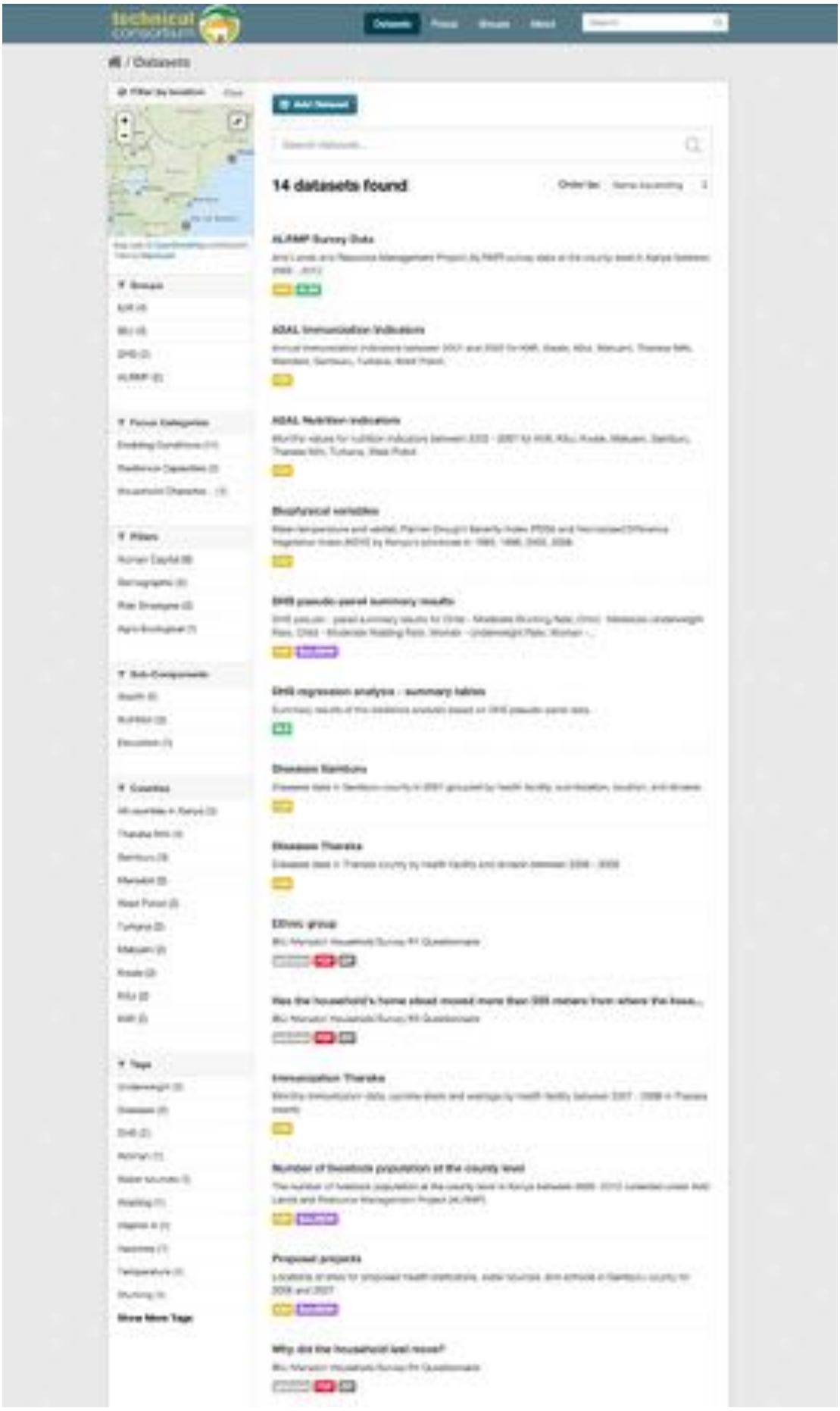
After the transformation process, the converted data files were uploaded to the HarvestChoice GitHub repository (https://github.com/harvestchoice/hoa-data) as an intermediary data access point linking with the Catalog. For example, the ASAL nutrition indicators 2002-2007 dataset accessible in the Catalog (link)

fetches converted source data file in the GitHub (link). This workflow allows update the data without directly interacting with the Catalog, thus minimizing the chances of making accidental mistakes in the data management, while CKAN facilitates the organization and classification of the datasets. Additionally, this process provides an additional benefit of tracking the historical changes in the dataset via GitHub's automatic versioning system.

USER INTERFACE

All datasets in the Catalog are accessible at http://data.technicalconsortium.org/dataset (Figure 1). The filtering in the left-hand panel narrows down the results by *Groups, Focus Categories, Pillars, Sub-Component, Administrative Units* (Counties in Kenya), and *Tags*. Click on each dataset lead the user to its landing page that includes its metadata classification and a list of data resources. Further selection of the resource will display a tabular preview of the dataset together with its metadata information. The resource can be downloaded, accessed via API, or visualized as predefined or user-defined maps, tables, and graphs.

Figure 1: Landing page of the Catalog Datasets (http://data.technicalconsortium.org/dataset)



DATA ACCESS VIA API

All data served in the Catalog are publicly accessible from the web user interface and can be programmatically retrieved using the API. Implemented in a standard protocol, this allows third-party applications to use the data directly without needing to access through the website. Each Resource page includes a button labeled as Data API, located in the top right corner, displaying technical specifications and instructions.

For example, to retrieve the first five entries within the monthly diseases by health facility dataset in Tharaka Nthi can be directly retrieved from the following URL: http://data.technicalconsortium.org/api/action/datastore_search?resource_id=b5317c1a-f685-4cb9-aabc-2e8c162e4261&limit=5

The result is a JSON formatted object that can be processed programmatically (Figure 2).

Figure 2: Example of the API search result formatted in JSON

3

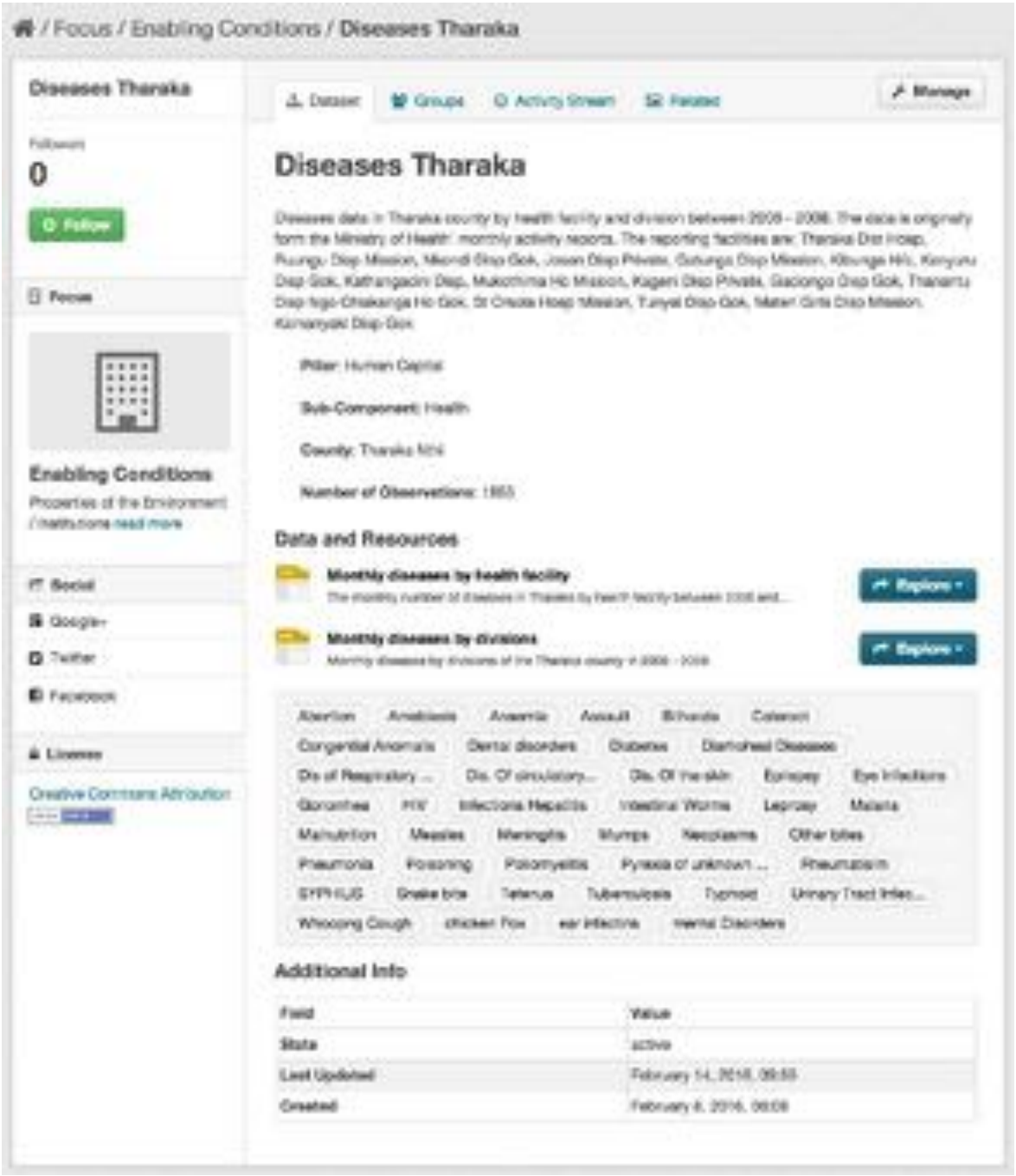
Examples

EXAMPLE 1: Human Health Data in Tharaka Nithi County, Kenya

Description

The diseases dataset for the Tharaka Nithi County in Kenya is accessible at http://data.technicalconsortium.org/dataset/diseases-tharaka (Figure 3). Its resources include monthly diseases data for three years: 2006, 2007, and 2008. The data was provided by the Ministry of Health Monthly Activity Reports, provided in Excel format. This was manually put together in a structured format (CSV) to facilitate accessibility and reporting. The resources can be downloaded and visualized as the number of diseases by reporting institution or by division. The data includes reports from 17 institutions in the county on 42 diseases.

Figure 3: Diseases dataset for Tharaka Nithi County, Kenya



Visualization

The platform offers a functionality for users to custom-create data visualizations in tables and graphs. For example, from the monthly data summary view by division (link), selecting [Graph] button in the interface will open an interface to create a temporal chart showing, for example, the temporal trend of monthly malaria cases in the county (Figure 4).

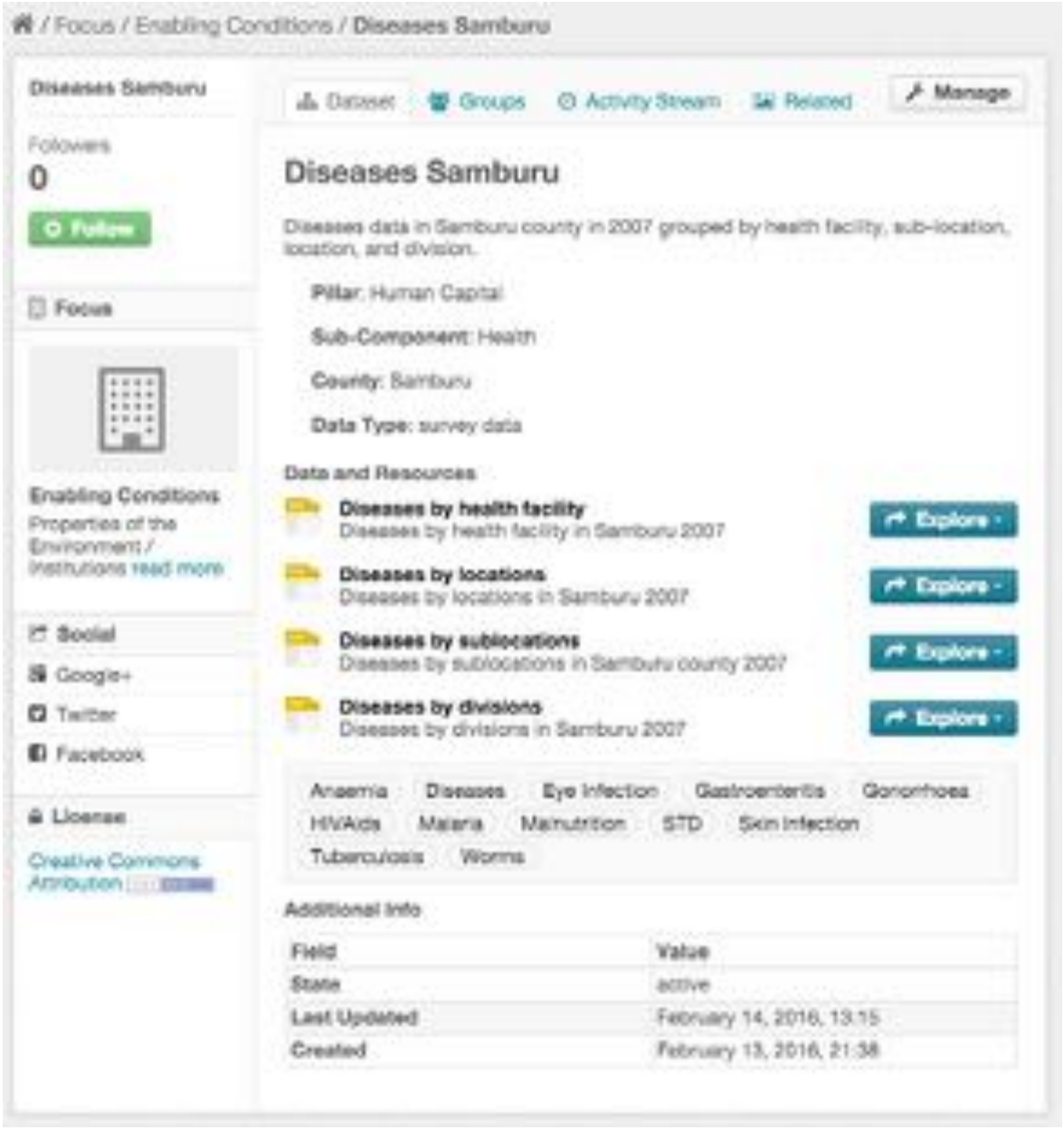
Figure 4: Monthly cases of malaria in Tharaka Nthi county, by sub-county (filtered to show 2007 data only)

EXAMPLE 2: Human Health Data in Samburu County, Kenya

Description

The diseases dataset for the Samburu County in Kenya is accessible at http://data.technicalconsortium.org/dataset/diseases-samburu (Figure 5). The dataset includes the number of diseases in Samburu in 2007 classified by health facility, sub-location, location, and division. The dataset includes reports from 53 dispensaries, health centers, and hospitals. For each of the diseases the total number is reported along with its rank within each health facility. Among the diseases are malaria, gastroenteritis, HIV, tuberculosis, worms, gonorrhoea, STD, skin infection, eye infection, anemia, and malnutrition. The data originally provided in the Microsoft Access format; it has been converted to a tabular format (CSV). The disease information extracted and aggregated at sub-location, location and division level.

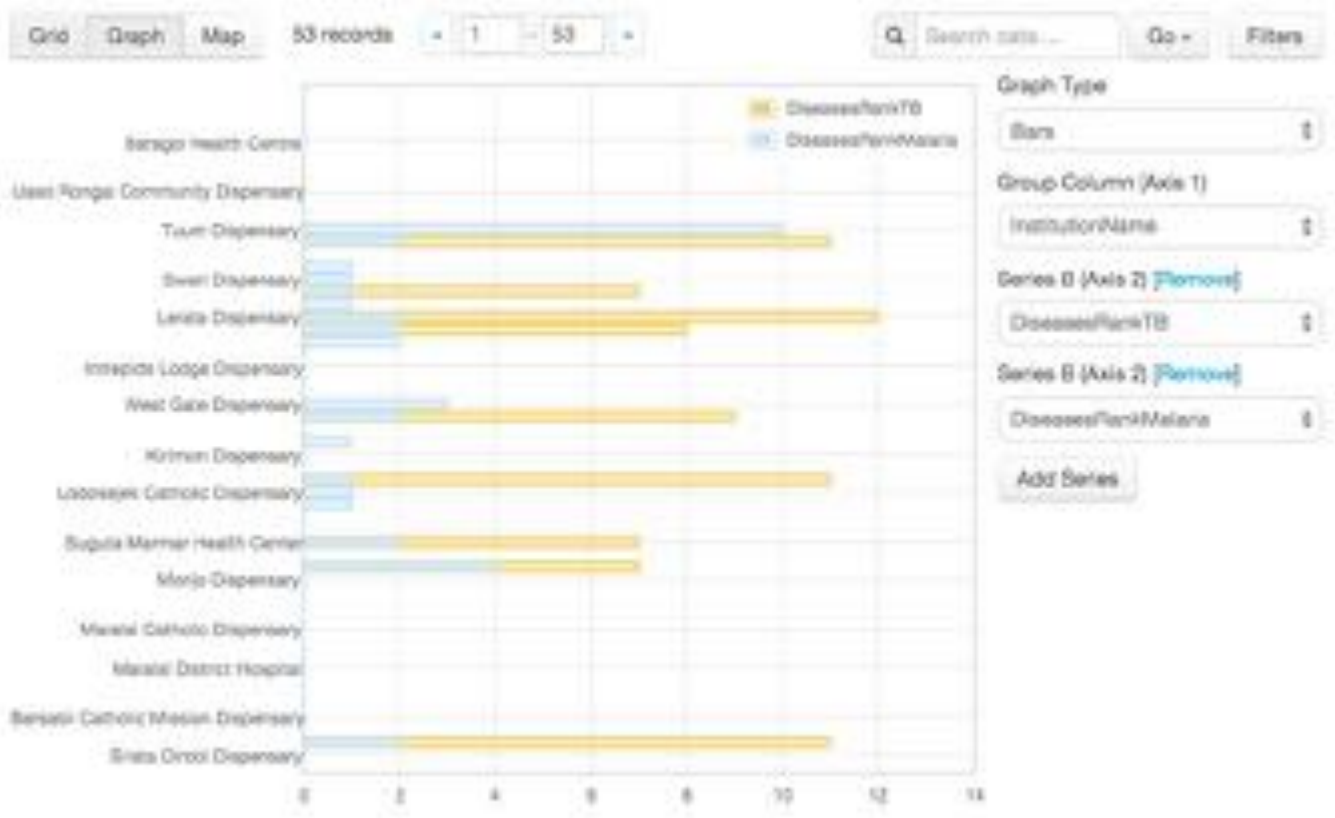
Figure 5: Diseases dataset for Samburu County, Kenya



Visualization

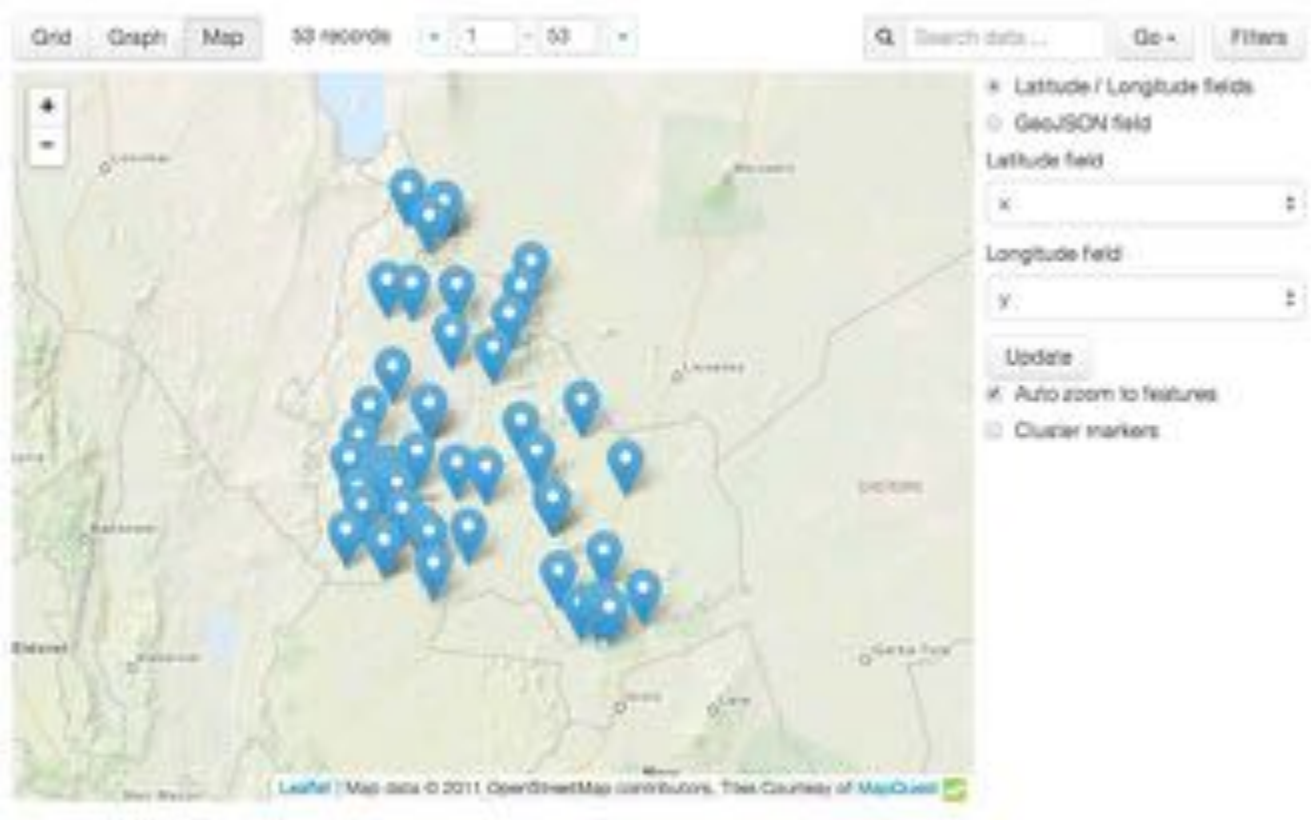
The number of diseases in Samburu in 2007 dataset grouped by health institution is accessible at (link), and the data can be visualized in tabular, chart, or map format. For example, a bar chart can be created online to compare the rank of tuberculosis and malaria cases in each facility (Figure 6).

Figure 6: Tuberculosis and malaria cases rankings of health facilities in Samburu County, Kenya



In addition, as the dataset includes the locations (coordinates in X and Y) of the health facilities, a map view can be created to locate them as a marker (Figure 7).

Figure 7: Location of health institutions in Samburu County, Kenya



Since the underlying dataset is also grouped at the division level, a further disaggregated visualization is possible. For example, as shown in Figure 8, a comparison of the number of cases of worms, anemia, gonorrhea, eye infection, and malnutrition at the division level can be visualized in a chart (link).

Figure 8: Cases of worms, anemia, gonorrhea, eye infection, malnutrition by Samburu's divisions

Recommendations for Future Utility of the Data Catalog

This data catalog was originally conceived for several reasons; first, to organize existing secondary data into a resilience-focused infrastructure, second, to be able to determine whether there was adequate and appropriate secondary data available in the region to conduct baseline studies for projects aimed at enhancing resilience, and third, to provide a central point for future data collection to be uploaded and subsequently used. As the catalog is structured in accordance with the IGAD member states' investment plans to "end drought emergencies", it would be logical for the agencies responsible for implementation of these plans to both curate and maintain these databases. For example, in Kenya, the NDMA is the government body responsible both for the development and implementation of the Kenya Ending Drought Emergencies investment plan and therefore should be the recipient of and the gateway to data to inform project baselines and future impact assessment studies.

In order to be able to fulfill this function, however, an assessment of capacity, both technical and operational, needs to be carried out to ascertain the potential for these agencies to undertake this role. In addition, countries must derive a benefit to maintaining this database and for this to be evident, not only is it necessary for the data within to be easy to use as an output to enhance the monitoring and evaluation of their investments, but the uploading of new datasets must also be feasible and simple. Countries must invest in training and capacity building of staff, in addition to server space. What would be useful as well would be a consensus amongst all actors on protocols for data collection and cu ration going forward, in the form of easily applicable manuals outlining standards and guidelines.

It should also be noted that this catalog will not only be of use to those agencies and individuals working on projects designed to enhance resilience, but will in fact have a much broader potential application. As the data is catalogued using a framework relatively analogous with sustainable development, the datasets can be used in any scenario requiring prediction models, baselines and impact assessment and evaluation. The aim going forward would be to broaden the geographical parameters of the data beyond Kenya and the Horn of Africa to include West Africa and other regions. The catalog can function as the data needed for many different types of platforms satisfying multiple agendas.

4

Case Studies

STUDY 1: Malnutrition and Climate Patterns in the ASALs of Kenya: A Resilience Analysis based on a Pseudo-panel Dataset[1]

Authors
Sara Signorelli, Carlo Azzarri, Cleo Roberts, and Zhe Guo

Summary
In this case study, authors examine resilience of food security in arid and semi-arid lowlands (ASAL) in Kenya, using repeated cross-sectional data (collected in 1993, 1998, 2003, and 2008). Short- and long-term food security in response to changing agro-climatic conditions were measured, using indicators of children's and women's nutritional status. Since measurements over time on the same units of analysis were not available, synthetic cohorts of individuals (children and women) having similar characteristics were reconstructed from the harmonization of available datasets, creating a pseudo-panel dataset. The dataset is used to observe resilience to agro-climatic shocks, including temperature increases, drought, and changes in vegetation cover, geographically harmonized from the grid-based datasets from other sources.

In order to study the link between individual nutrition status and climatic conditions, authors constructed a database combining household survey data with remote sensing information on biophysical variables. At the micro-level, individual modules of the last four Demographic and Health Surveys (DHS) waves of survey data available for Kenya (1993, 1998, 2003, and 2008) were used. The DHS data are nationally-representative household surveys that provide data on a wide range of indicators on population, health, and nutrition. In particular, each wave contains anthropometric measures for both children below five years of age and women of reproductive age, which were used as outcome variables in the analysis. Biophysical variables measured through satellites sourced from the HarvestChoice Database (http://harvestchoice.org/data) were then associated to each DHS household at the district level in each wave. In particular, climate-related characteristics, such as mean temperature and rainfall, Palmer Drought Severity Index (PDSI) and Normalized Difference Vegetation Index (NDVI) were used. Overall, the study showed how arid and semi-arid areas of Kenya are particularly affected by child and female under-nutrition.

While in the rest of the country under-nutrition has been improving, trends appear negative for the ASAL areas, especially in the case of child wasting and woman underweight. Temperature shocks emerge as the most detrimental factor for nutrition, again especially in ASAL areas. Droughts, on the other hand, seem to play a significant role only in affecting stunting, while NDVI plays a mixed role,

[1]Full paper of this case study is available at http://www.technicalconsortium.org/publications (See Technical Report Series No 2).

with some cases where more vegetation is associated with higher levels of under-nutrition. Availability of a non-agricultural job within the household is positively associated with nutritional outcomes and so is woman education, especially in ASAL counties. However, they are also associated with bigger losses in the event of temperature shocks, which question the role of non–agricultural activities in increasing resilience.

Corresponding Data in the Catalog

All the source data and the analysis results from this case study is available in the Catalog at http://data.technicalconsortium.org/dataset/dhs-pseudo-panel-summary

In addition, a set of data visualizations on the analysis results were made as the Supplementary Information (e.g., analysis results in Figures 9, 10, 11, 12, and 13; biophysical input data in Figures 14, 15, 16, and 17). These are accessible online at (link).

Figure 9: Child moderate stunting rate (%)

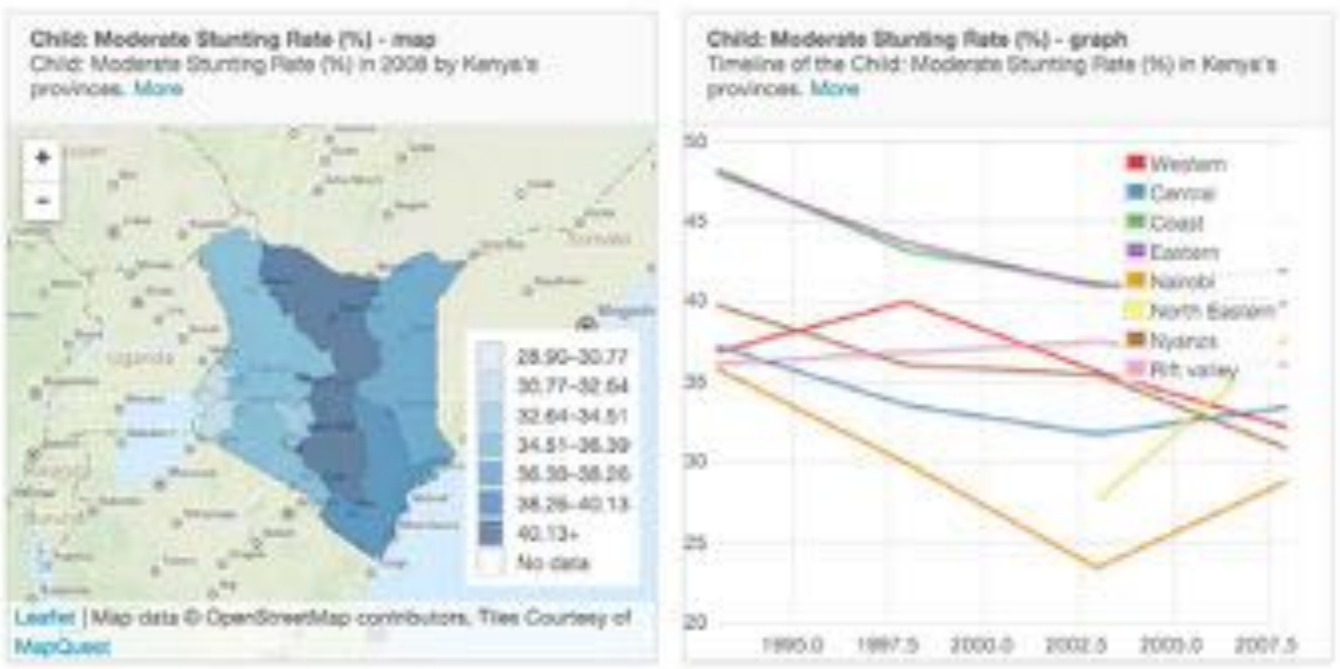


Figure 10: Child moderate underweight rate (%)

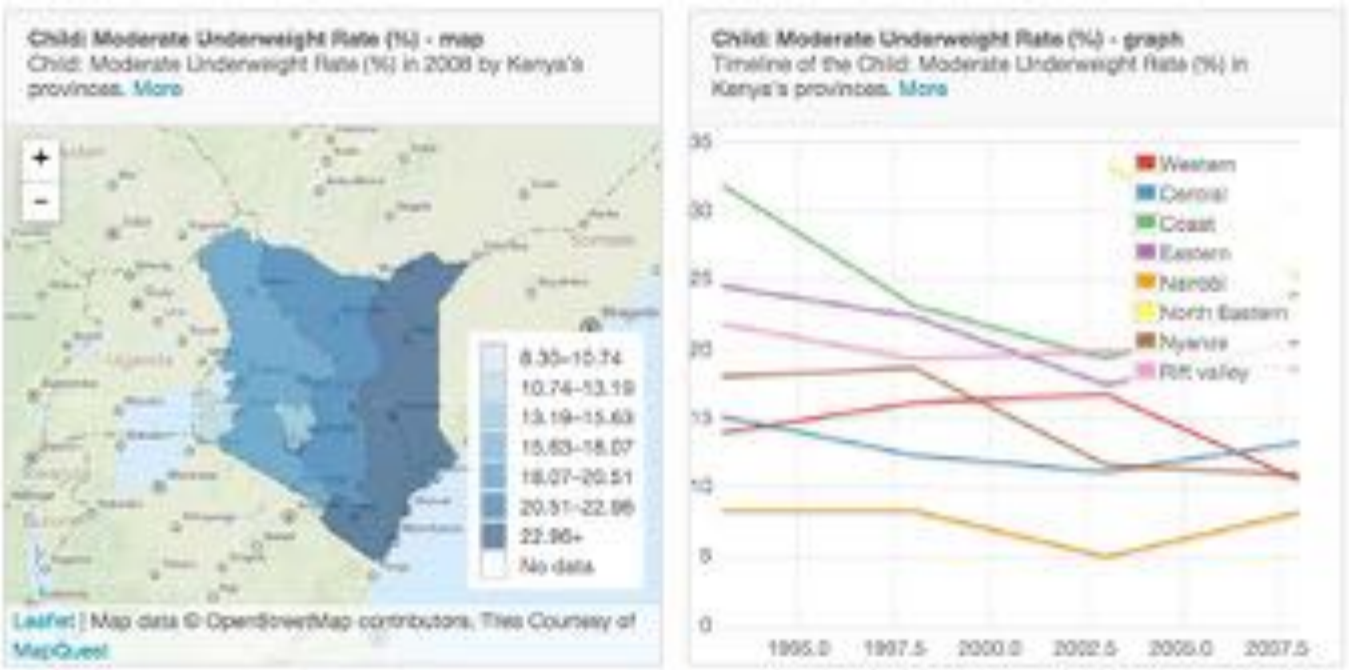


Figure 11: Child moderate wasting rate (%)

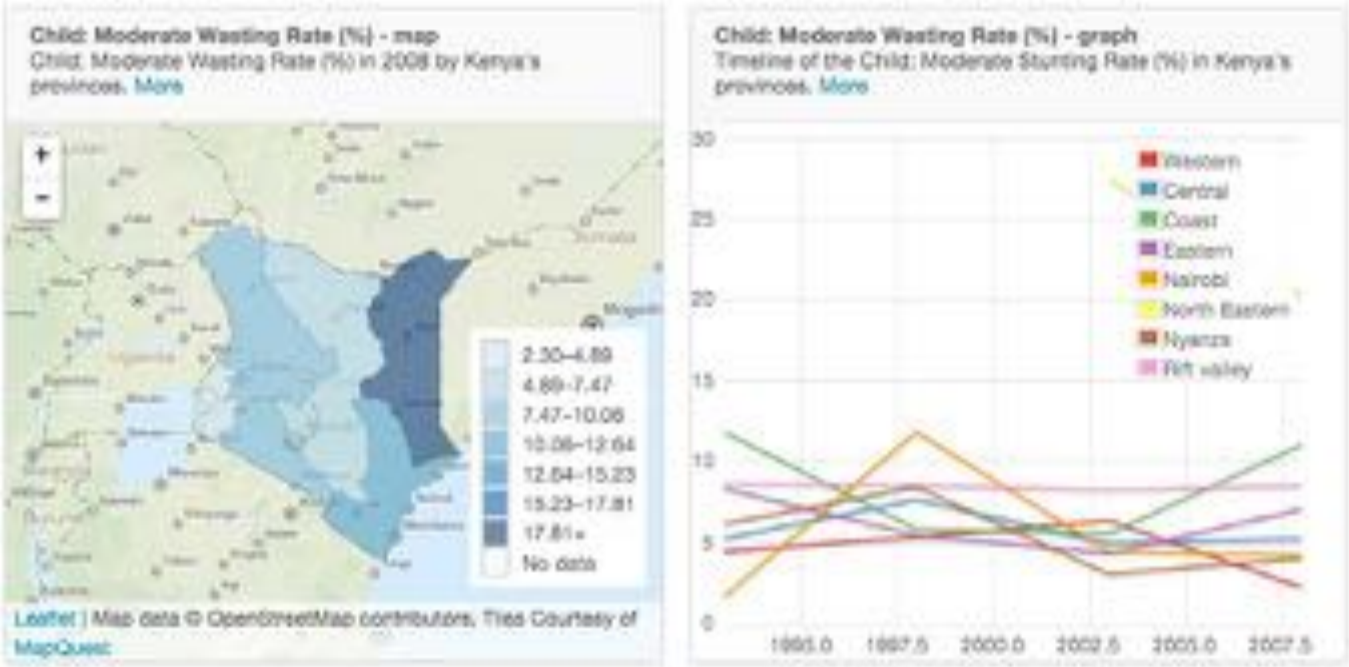


Figure 12: Woman underweight rate (%)

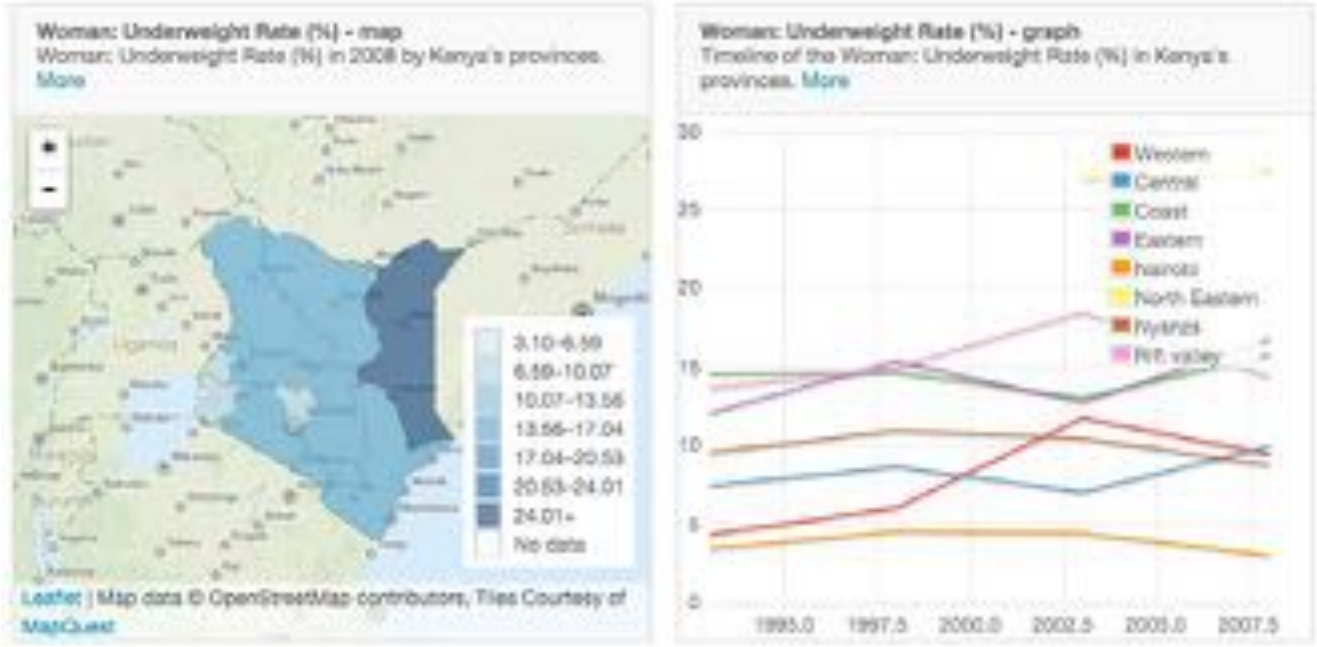


Figure 13: Woman Average BMI

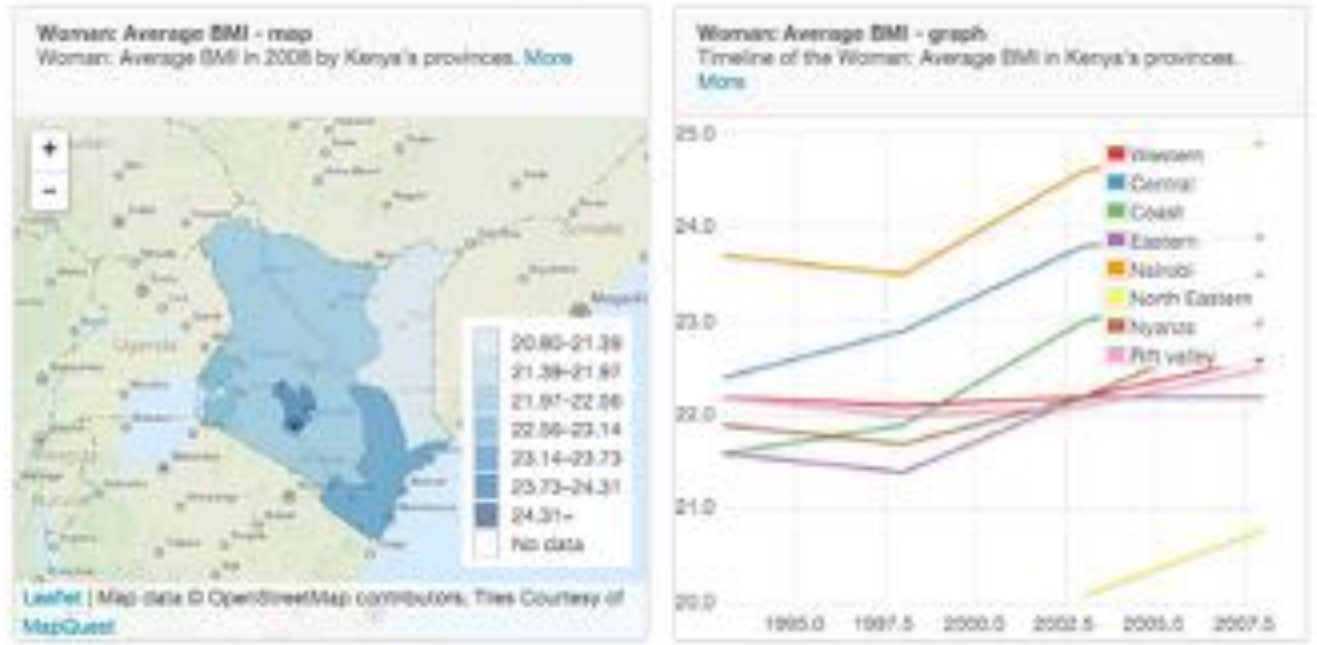


Figure 14: Normalized Difference Vegetation Index (NDVI)

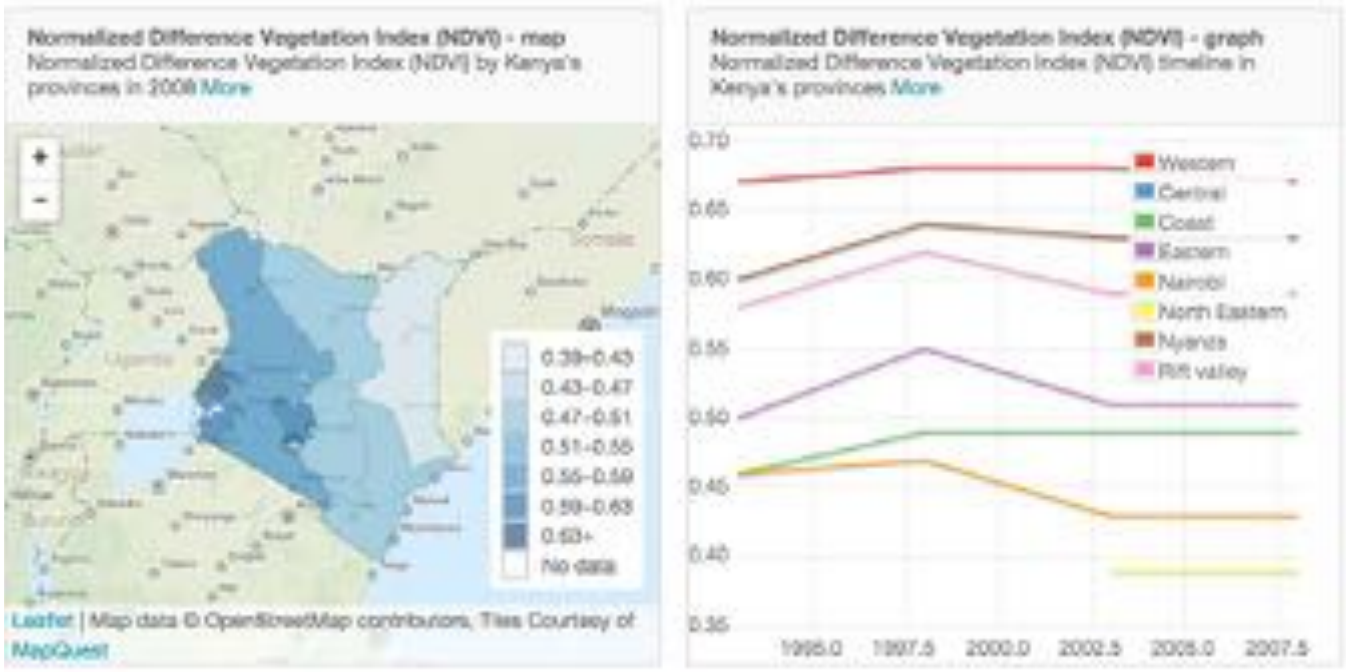


Figure 15: Inverse Palmer drought severity index (PDSI)

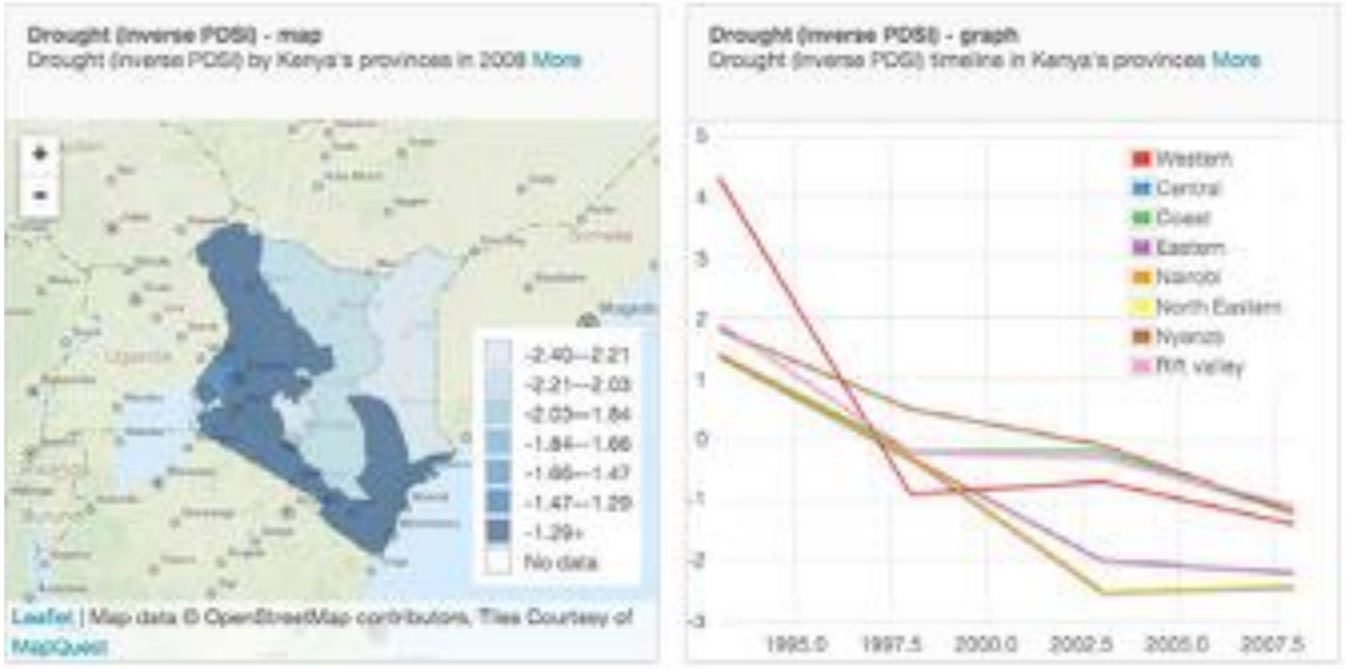


Figure 16: Temperature

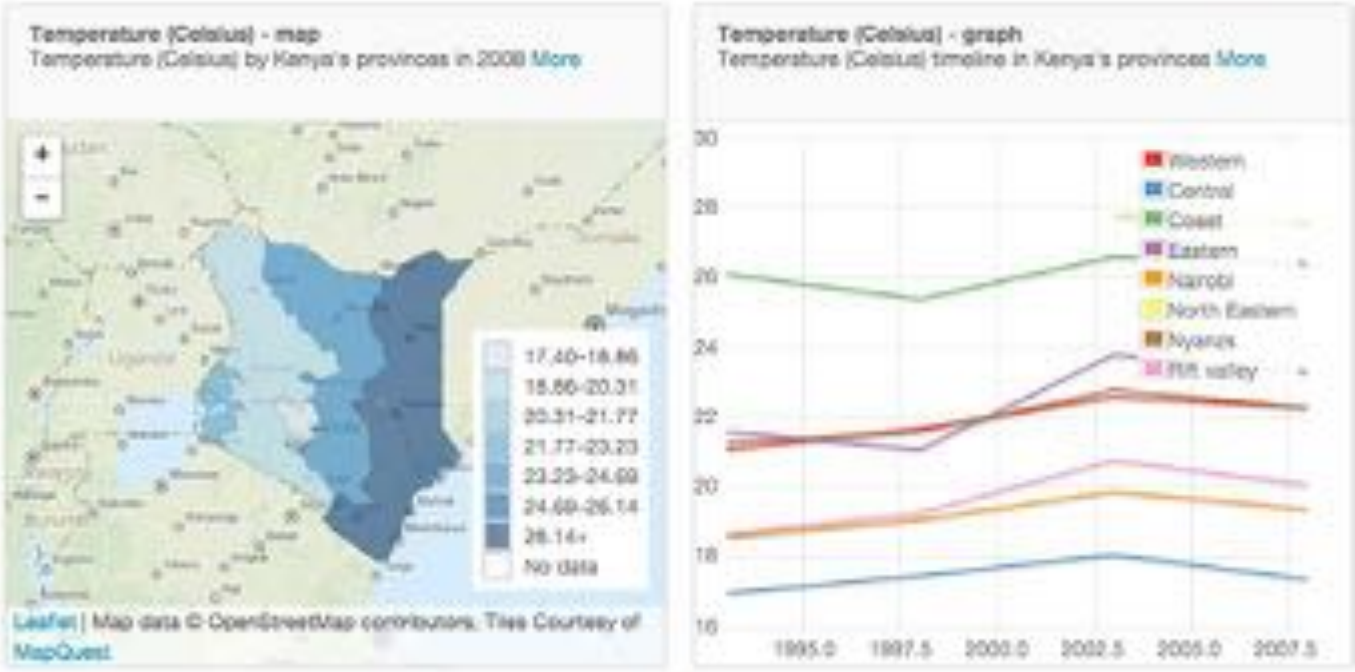
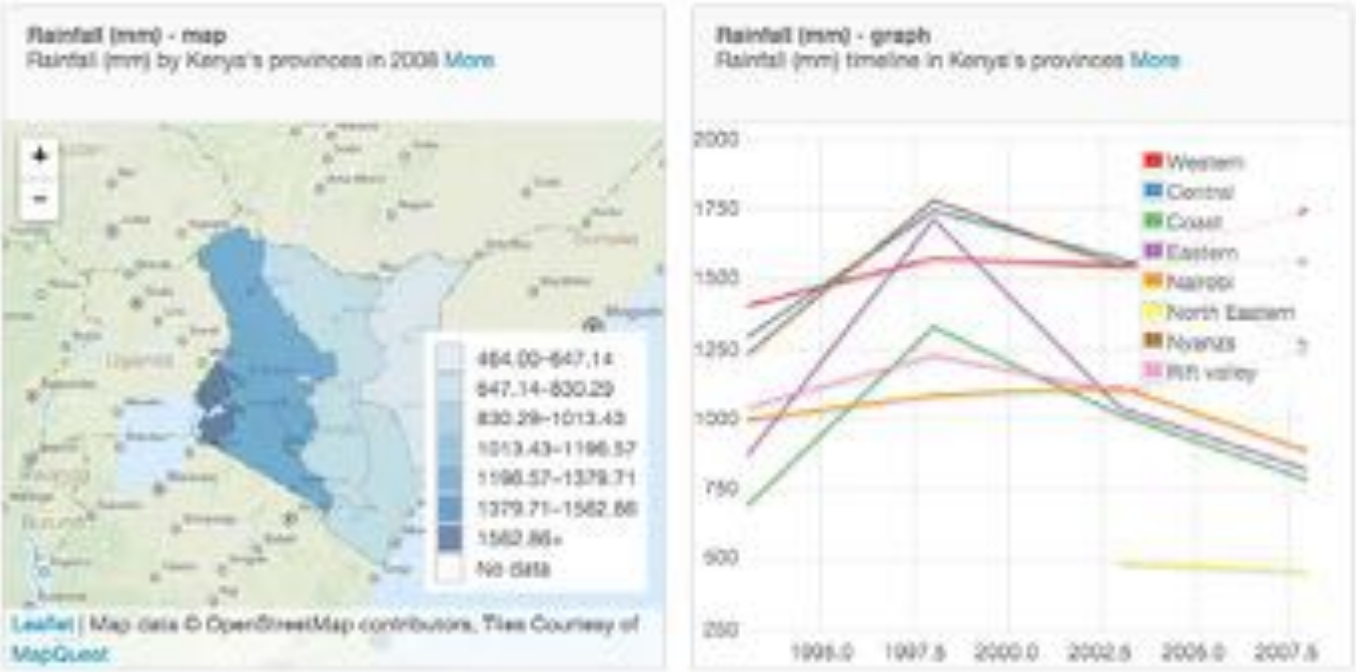


Figure 17: Rainfall (mm)



STUDY 2: Maintaining Resilience in Arid/Semi-Arid Kenya: A Perspective on Stocking Rates in Extensive Livestock Systems[2]

Authors
Siwa Msangi and Sara Signorelli

Summary
The extensive systems of northern Kenya are subject to the same dimensions of vulnerability as seen in others across the drylands of Eastern and Sahelian sub-Saharan Africa, namely the availability of feed and water for animals. The grazing lands used by the herders are typically unrestricted in access and are an "open-access" resource to others who wish to graze their animals. This kind of situation is common to many common-pool resources that provide benefits for the users, but whose use is not restricted or regulated in anyway. Grazing commons can be over-stocked with animals, beyond the biophysical capacity of the herbaceous grasslands to regenerate, thereby constituting a mining of the resource beyond the level that would otherwise be able to provide a sustainable source of feed. The itinerant behavior of livestock herders, i.e. transhumance, is a response to this situation, and provides an important livelihood strategy for livestock keepers. Rather than solving the problem of the grazing commons, the transhumance merely spreads out the issue over space, and creates a complex web of resource exploitation patterns that have both a temporal and spatial dimension. This results in a situation in which everyone is worse off than under the case where cooperation is undertaken, and the joint benefit of all users is considered.

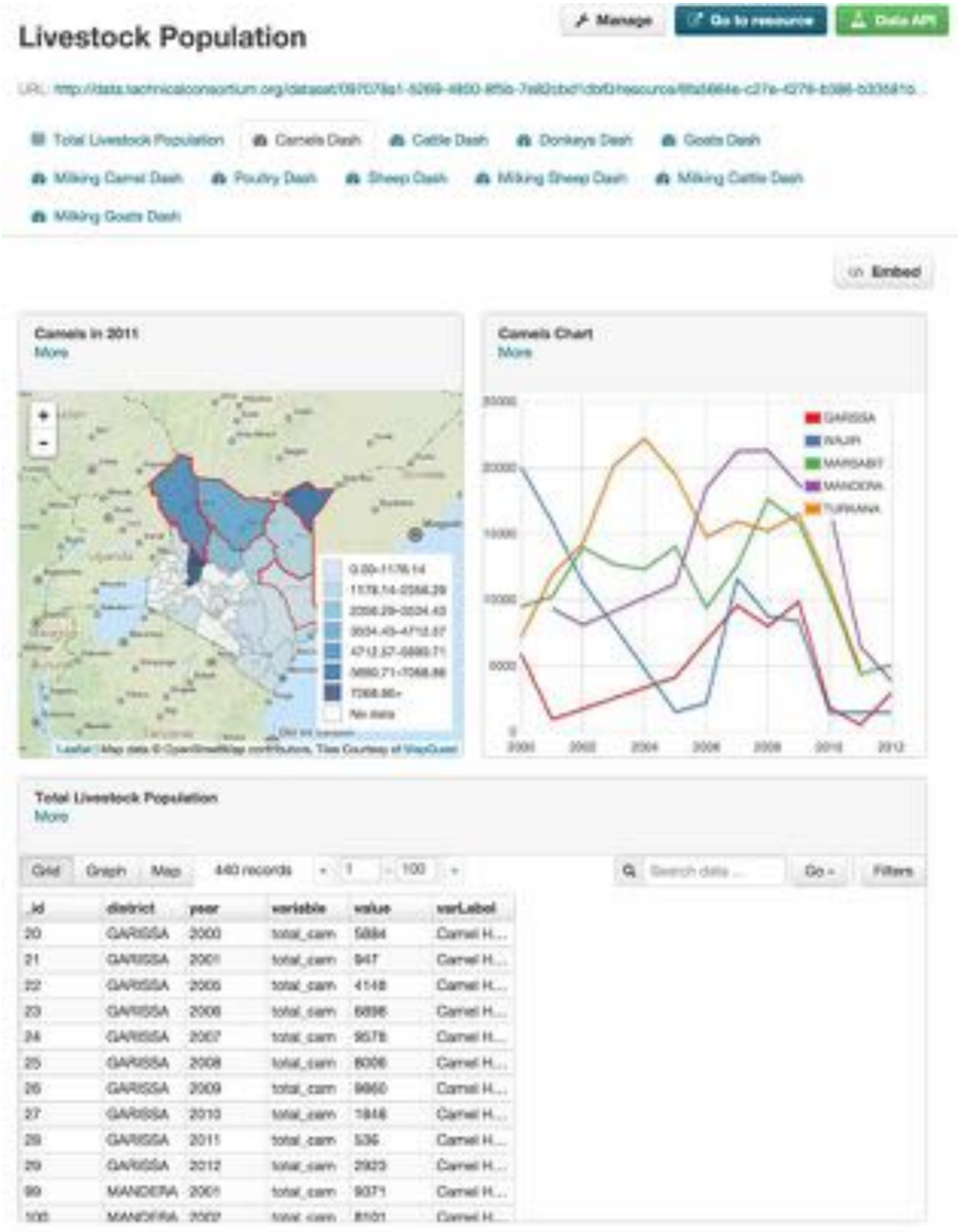
In this analysis, authors make use of household level data to characterize some very basic characteristics of pastoral livestock keepers, representing their behavior and adjustment to environmental shocks. In order to come up with realistic values for milk consumption and production at the household-level, for pastoralists, authors made use of the data contained within the micro-level individual modules of the latest four Demographic and Health Surveys (DHS) waves of survey data available for Kenya (1993, 1998, 2003, and 2008) were used. To characterize the optimal behavior of the far-sighted pastoralist, a dynamic programming framework was developed, in which the decision maker makes an explicit tradeoff between the benefits that occur in the present period and those that will occur in the future. Scenarios applied in the framework illustrated the benefits that can be realized from better management of livestock in extensive systems. Overall results illustrated the long-run benefits of limiting stocking rates to better match with the feed availability are clear and significant. Future extension of the study includes the inclusion of the climate effect (e.g., increased frequency of dry years in succession).

Corresponding Data in the Catalog
The underlying livestock data used in this study is available in the Catalog at http://data.technicalconsortium.org/dataset/number-of-livestock-population-at-the-county-level-1

Additionally, a data visualization dashboard was created as a Supplementary Information at (link) (Figure 18).

[2] *Full paper of this case study is available at http://www.technicalconsortium.org/publications (See Technical Report Series No 2).*

Figure 18: Livestock population in Garissa, Wajir, Marsabit, Mandera, and Turkana 2000 - 2012



The International Livestock Research Institute (ILRI) works to improve food security and reduce poverty in developing countries through research for better and more sustainable use of livestock. ILRI is a member of the CGIAR Consortium, a global research partnership of 15 centres working with many partners for a food-secure future. ILRI has two main campuses in East Africa and other hubs in East, West and Southern Africa and South, Southeast and East Asia. www.ilri.org

CGIAR is a global agricultural research partnership for a food-secure future. Its science is carried out by 15 research centres that are members of the CGIAR Consortium in collaboration with hundreds of partner organizations. www.cgiar.org

The Technical Consortium for Building Resilience in the Horn of Africa provides technical support to IGAD and member states in the Horn of Africa on evidence-based planning and regional and national investment programs, for the long-term resilience of communities living in arid and semi-arid lands. It harnesses CGIAR research and other knowledge on interventions in order to inform sustainable development in the Horn of Africa. www.technicalconsortium.org