○ **Earth Perspectives**
a SpringerOpen Journal

RESEARCH ARTICLE

Open Access

IRI Data Library: enhancing accessibility of climate knowledge

M Benno Blumenthal[*], Michael Bell, John del Corral, Rémi Cousin and Igor Khomyakov

Abstract

Background: Climate variability affects a broad swath of socio-economic sectors, and if it increases or the sector becomes overly-tuned to past or present climate conditions, climate variability becomes of increasing concern to a wide range of non-climate specialists. The significant challenges to building the capacity of non-climate specialists to use climate information in research and decision-making include the difficulties in accessing relevant and timely quality-controlled data and information in formats that can be readily incorporated into specific analysis and reporting.

Methods: The IRI Data Library is a facility designed to cope with these issues of information dissemination. Methods developed include Map Rooms which are designed for rapid access to needed information for particular user groups, analysis tools useful for a wide range of users (especially while training), and a metadata framework that uses semantic technologies to transform metadata from a variety of sources into a variety of standards.

Results: The results are tools to merge standard climate products with GIS information (e.g. averaging climate data over the political boundaries used to geolocate health and socio-economic data), as well as simplified access/transformation of large datasets only available as collections of many files or service points elsewhere.

Conclusions: The IRI Data Library is thus a key platform that makes climate and other data products more widely accessible through tool development, data organization and transformation, and data/technology transfer.

Background

Climate variability affects a broad swath of socio-economic sectors, and if climate variability increases or the sector becomes more efficient and thus more precisely tuned to past or present climate conditions, it becomes of increasing concern to a wide range of non-climate specialists.

For example, public health professionals are increasingly concerned about the potential impact that climate can have on health outcomes. In the absence of effective disease control, climate determines the spatial and seasonal distribution of many infectious diseases and is a key determinant of inter-annual variability in disease incidence, including epidemics and longer-term changes in endemicity (Kelly-Hope and Thomson 2008).

Protecting public health from the vagaries of climate will require new working relationships between the public health sector and providers of climate data and information. It will also demand a wide variety of strategies

occurring at multiple levels. One of these strategies is to increase the public health community's capacity to understand, use, and demand appropriate climate data and information to mitigate the public health impacts of the climate. However, good information is not enough. The public health community must also be able to distinguish between different kinds of data and information products to determine what is relevant for their specific needs, how it can be readily accessed, and what methodologies and tools can best serve their purpose. Health practitioners and researchers concerned with climate-sensitive decisions are not routinely trained to consider these issues.

Significant challenges to building the capacity of health professionals to use climate information in research and decision-making include the difficulties in accessing relevant and timely quality-controlled data and information in formats that can be readily incorporated into specific analysis with other data sources (Thomson et al. 2011). While initiatives to improve health communities access to relevant quality controlled climate data are underway (Dinku et al. 2013) many barriers remain in terms of

* Correspondence: benno@iri.columbia.edu
International Research Institute for Climate and Society, Columbia University, Palisades, NY, USA

data, services, practice and policy (IRI 2006) that will need to be overcome for climate and environmental information to play a significant part in reducing climate-related risks with regard to health (Connor et al. 2010).

These barriers include (but are not limited to) a lack of:

- access to relevant local and globally accessible data that may be used to create policy-relevant evidence for local, national, and regional decision-making;
- ability to generate new knowledge because there is insufficient capacity to understand, assess, and use climate information (as well as other environmental and demographic information), in analyses designed to support a specific research question;
- effective and available tools to enable the analysis of relevant data in space and time and which communicate easily with other software used for research or knowledge sharing;
- policies for data sharing as well as technological constraints to knowledge and data sharing that could facilitate networks of researchers to engage with each other around common research agendas; and
- a policy and practice environment that is responsive to new information concerning changes in disease risk.

The capacities of the IRI Climate Data Library can be used to build an integrated knowledge system to support the use of climate and environmental information in climate-sensitive decision-making. Initially funded as an aid to climate scientists for exploratory data analysis, it has now expanded to provide a platform for interdisciplinary researchers focused on topics related to climate impacts on society (del Corral et al. 2012).

The IRI Climate Data Library

As its name suggests, it represents a collection of datasets, both locally and remotely held, designed to make information more accessible for the library's users. Datasets in the library come from many different sources, "data cultures", and formats. By "dataset" we mean a collection of data organized as multidimensional dependent variables, independent variables, and sub-datasets, along with the metadata (particularly on purpose and use) that makes it possible to interpret the data in a meaningful manner.

The Ingrid programming language, which facilitates the infrastructure for the Data Library, is an environment that allows working with datasets (e.g. to read, write, request, serve, view, select, calculate, and transform). It hides an extraordinary amount of technical detail, letting the user think in terms of manipulations of datasets rather than manipulations of files of numbers. Among other things, this hidden technical detail makes

it possible to access data on servers in other places, doing a calculation only on the small necessary portion of a dataset, or translating to and from a variety of formats and between "data cultures". Thus, the Data Library is a very powerful, open-source, computational engine that offers, at no cost to the user, the opportunity to:

1. access, manage and manipulate any number of datasets from a variety of earth science and climate-related topics, including public health;
2. create analyses of data (including climate and health data) ranging from simple averaging to more advanced Empirical Orthogonal Function (EOF) analyses using the Ingrid programming language;
3. monitor current and review past climate/environmental conditions with maps and analyses;
4. create multi-dimensional visual representations of climate and public health data, including animations over time; and
5. customize and download data plots and maps in a variety of image and data formats, including those compatible with geographical information systems (GIS) or other software for data visualization.

Traditional GIS platforms are now widely used by planners and decision makers in society. However, they are highly-focused on geospatial capabilities and have limited functionality for temporal analysis. Without information on the latter, meaningful inference about the causation of disease outbreaks is impossible (Jacquez 2000). Furthermore, many tools are unable to readily process the vast quantities of space-time data associated with, for example, the outputs of a global climate model. The IRI Climate Data Library overcomes the limitations imposed by GIS platforms by being based on a much more general multi-dimensional data model that includes both space and time dimensions. All datasets, including GIS features (such as points, lines, and polygons) are geo-located and temporally referenced in a uniform framework. Functions and operators in the Data Library use this framework to perform a wide range of analyses that integrate climate/environmental datasets and public health-related datasets. Large datasets, such as 100-year climate change model intercomparison results, are available through the Climate Data Library's cataloging and data transfer protocol support. In addition, the Data Library's interface and functions can be used to access shared repositories in different parts of the world.

A further challenge to spatio-temporal analysis used in agriculture, hydrology, and public health is the integration of climate/environmental data with the sector data. There are normally important differences in the spatial and temporal scales of the datasets. Within the Data Library, an environmental dataset can be temporally
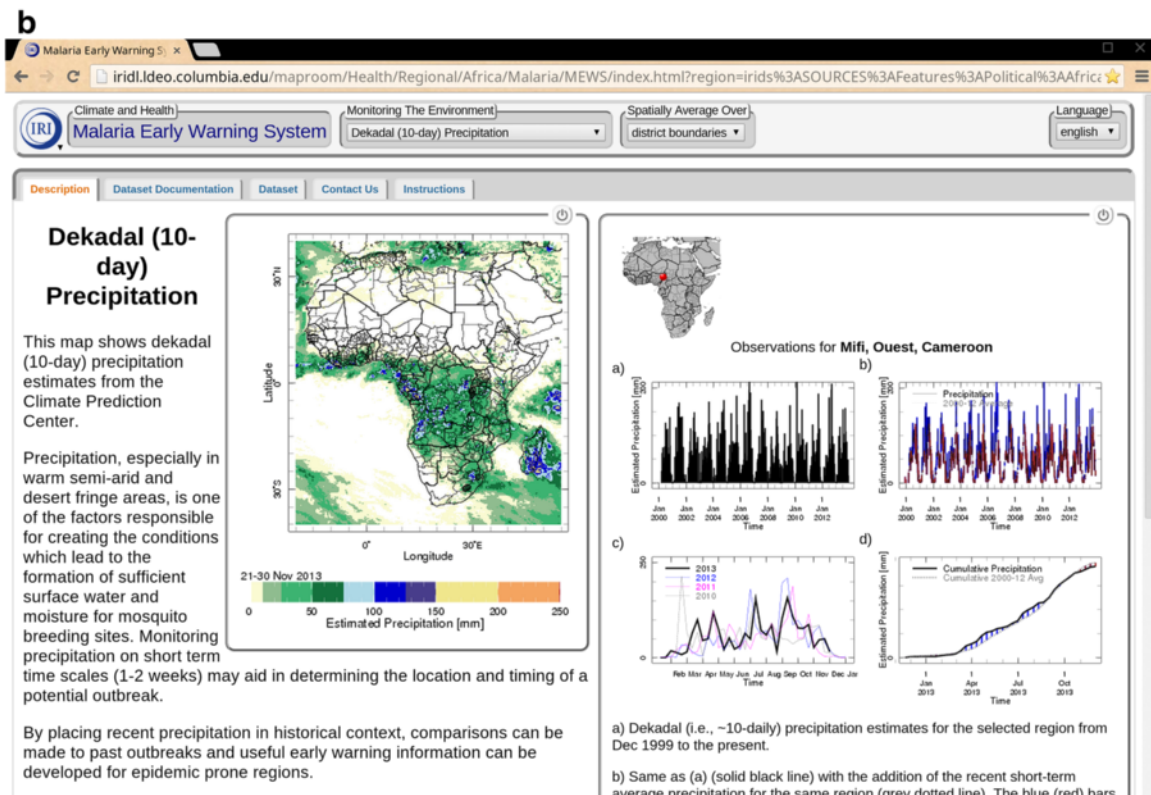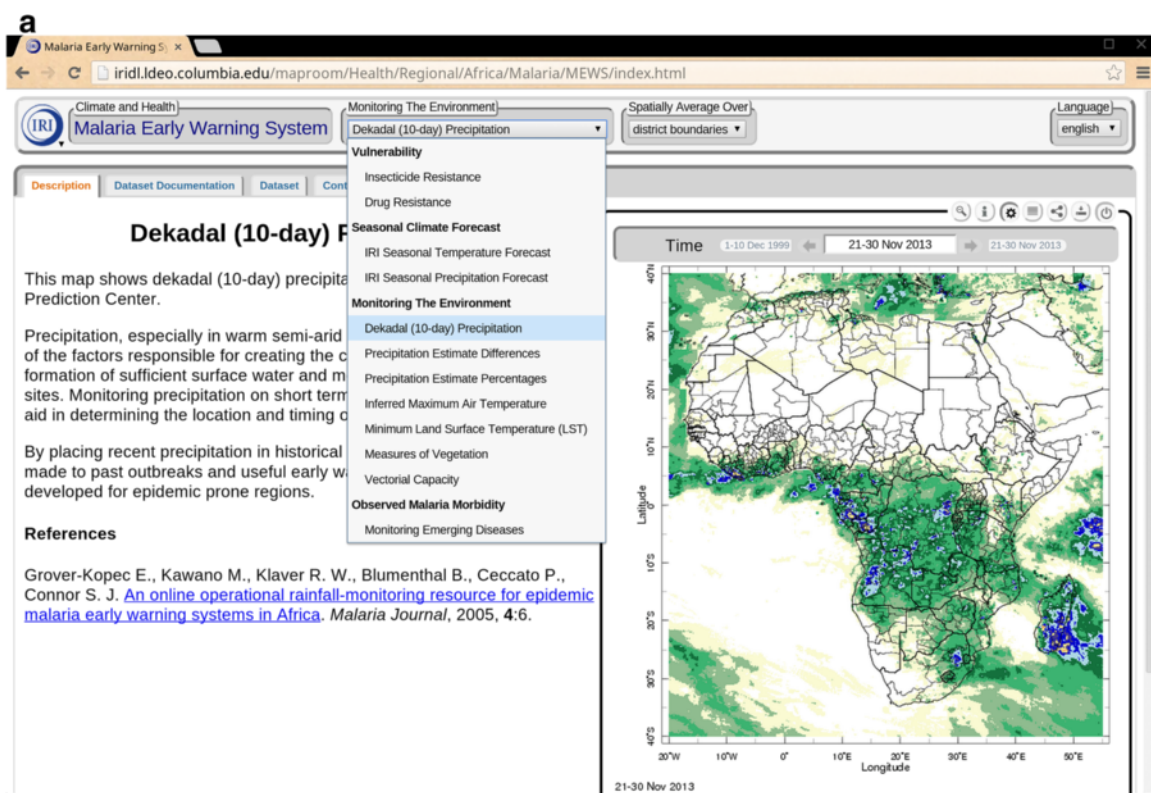
Figure 1 (See legend on next page.)

(See figure on previous page.)
Figure 1 a. Malaria Early Warning System (MEWS) Map Room is utilized because where malaria is not adequately controlled, its distribution and seasonality are driven by various climate factors such as temperature, humidity and rainfall. By knowing when conditions are suitable for transmission of malaria, health officials are granted several weeks, sometimes months of warning to apply insecticides, stockpile medicines and alert hospitals. This figure shows the dekadal (10-day) precipitation map from the Monitoring the Environment section of the MEWS Map Room: other facets of MEWS are displayed in the pull-down menu (visible): other MEWS sections focus on Vulnerability, Seasonal Climate Forecast, and Observed Malaria Morbidity. **b**. MEWS Dekadal Precipitation with time series selected.

averaged to match the time frequency of the sector data. If the sector data is based on a geographic points or administrative polygons, the environmental dataset can be sampled with the same geographic constraints.

The IRI Climate Data Library can be used via two distinct mechanisms that are designed to serve different communities. Expert Mode serves the needs of operational practitioners and researchers that have an indepth knowledge of the functionality of the system and are able to customize it to their own specific needs. Advanced users may develop custom functions and perform tailored analyses using Ingrid, the Data Library's programming language. This functionality is widely used around the world by climate researchers as Expert Mode allows users with programming skills a very extensive level of personalized functionality. Online tutorials, examples, and function definitions are part of the Data Library.

Methods
Map Rooms
Map Rooms are web-accessible tools targeted at particular user-groups, the end result of a process which evaluates user-group needs and builds tools that helps address those needs. These tools preselect data and analyses suitable for the task, building an easy-to-use framework for addressing the users' immediate needs, as well as providing links that allow the user to quickly download the data into the user group's standard tools for further analysis. While there are now many map rooms and hundreds of map room pages, several stand out in their operational use by their user groups.

The Malaria Early Warning System (MEWS) Map Room (Figures 1a,b, (Grover-Kopec et al. 2005) is utilized because where malaria is not adequately controlled, its distribution and seasonality are driven by various climate factors such as temperature, humidity and rainfall. By knowing when conditions are suitable for transmission of malaria, health officials are granted several weeks, sometimes months of warning to apply insecticides, stockpile medicines and alert hospitals. The MEWS maps illustrate models of climate suitability for seasonal endemic malaria, and recent climate conditions, such as rainfall anomalies, which may be associated with epidemic malaria in warm semi-arid regions of Africa. It is used by national malaria-control program personnel in Africa. Data used include CPC/Famine Early Warning System

Dekadal Estimates, NASA MODIS vegetation, and analyses based on NOAA NCEP/NCAR CDAS-1 Reanalysis and CPC Merged Analysis of Precipitation.

The Desert Locusts Map Room (Figures 2a,b, (Ceccato et al. 2007a, Ceccato et al. 2006) is utilized because swarms of desert locusts can travel thousands of miles and can threaten the food security and livelihoods of up to one fifth of the world's population. Recent plagues caused an estimated $400 million in damages and affected 8.4 million people. Knowing when and where environmental conditions are right for these insects to multiply helps authorities control their numbers. The map room shows maps and analysis products illustrating recent climate conditions, such as rainfall and vegetation growth, which provide ideal breeding conditions for the locusts. It is used by the U.N. FAO and regional locust-control workers. Data used include NOAA CPC CMORPH precipitation and NOAA MODIS vegetation.

The Indonesian Fire Map Room (Figure 3, (Someshwar et al. 2010) is based on research on peatland fires in the Indonesian province of Central Kalimantan that has uncovered a close correlation between satellite rainfall data and fire hotspot activity. In particular, rainfall during the dry season from June to October is critical in determining fire incidence. This finding means such data can help indicate whether an upcoming fire season will be more or less intense than usual, and can help authorities take preventive measures to avoid impacts to biodiversity, public health and global greenhouse gas emissions. The fire map room shows ten-day precipitation estimates for Indonesia; graphs that show the relationship between the number of fires and the NINO4 index in the previous month for the four Kalimantan provinces. It is available in English and Indonesian Bahasa. It is used by provincial environment, forestry and meteorological agencies. Data used include NOAA CPC CMORPH.
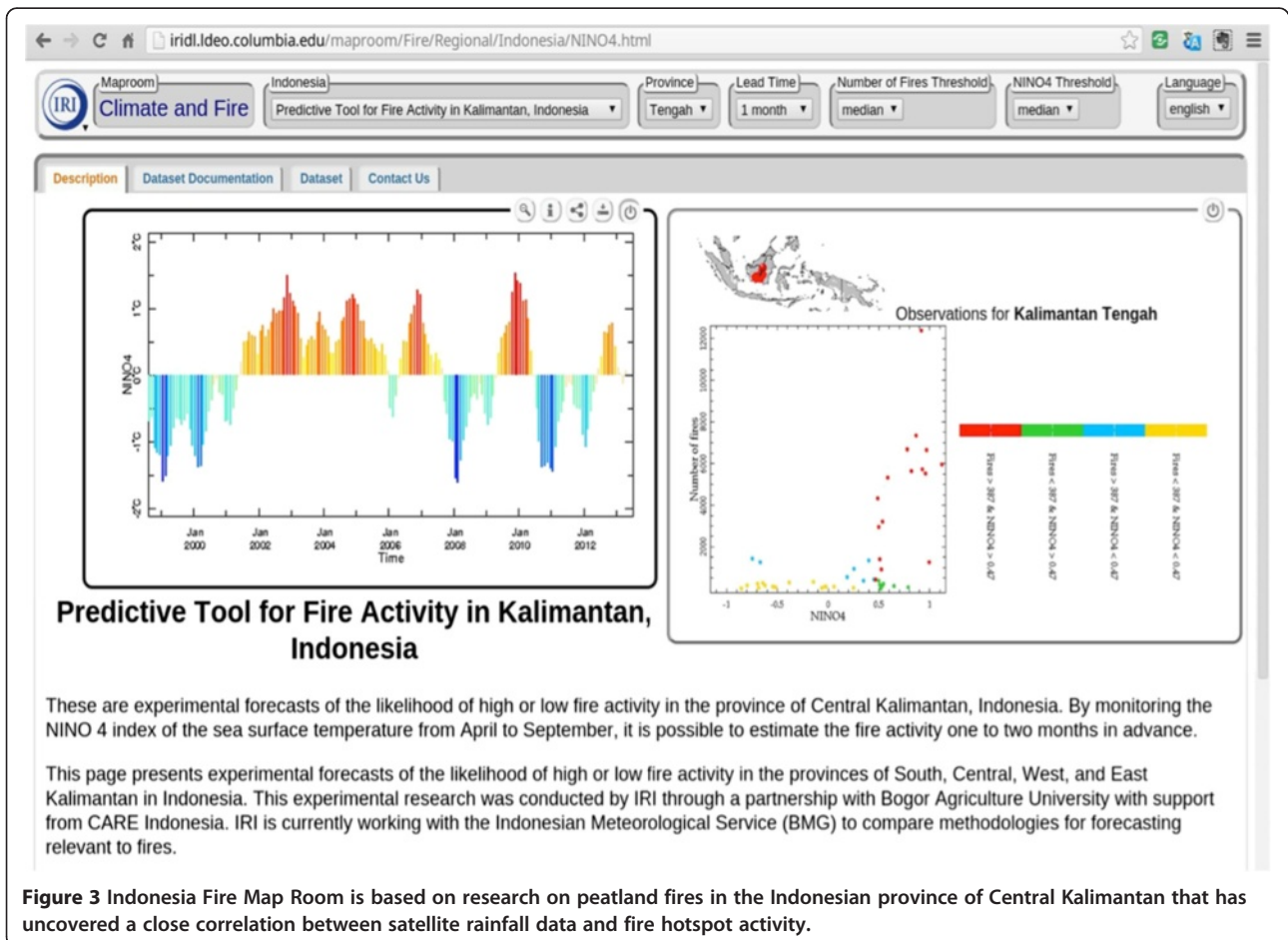
The IFRC Map Room (Figure 4) addresses the problem where in responding to disasters such as cyclones, floods and other weather-related events, humanitarian organizations must decide when and where to send aid. Determining which areas are likely to be hit first or hardest by an event can mean the difference between life and death. Also critical is the prediction of disaster "hotspots", or areas at high-risk because of their location and the vulnerability of their populations (e.g., a densely populated flood plain.) It shows the relative severity of
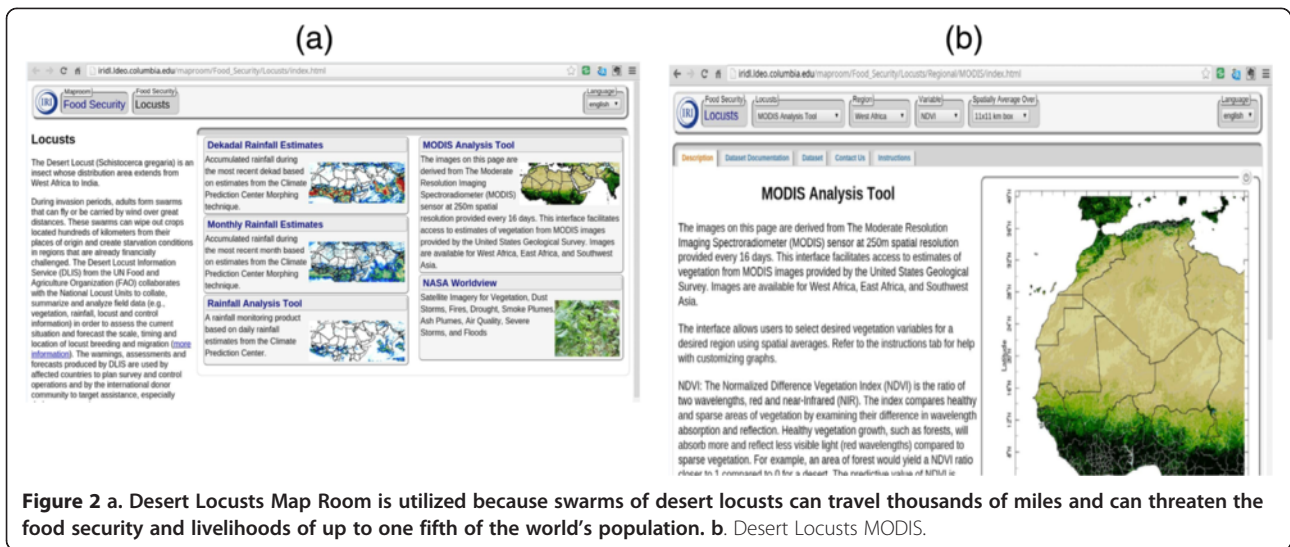
Figure 2 a. Desert Locusts Map Room is utilized because swarms of desert locusts can travel thousands of miles and can threaten the food security and livelihoods of up to one fifth of the world's population. b. Desert Locusts MODIS.

forecast rainfall events, 1–6 days in advance; "predictions in context" maps showing where seasonal forecasts indicate enhanced chance for continuation/reversal of previously observed rainfall; and population and poverty maps. It is used by the International Federation of Red Cross and Red Crescent Societies' operations-support department. Data used include the NOAA ESRL PSD Reforecast, NOAA CPC Merged Analysis of Precipitation, IRI Seasonal Forecasts, and CIESIN gridded population.

Predictive Tool for Fire Activity in Kalimantan, Indonesia

These are experimental forecasts of the likelihood of high or low fire activity in the province of Central Kalimantan, Indonesia. By monitoring the NINO 4 index of the sea surface temperature from April to September, it is possible to estimate the fire activity one to two months in advance.

This page presents experimental forecasts of the likelihood of high or low fire activity in the provinces of South, Central, West, and East Kalimantan in Indonesia. This experimental research was conducted by IRI through a partnership with Bogor Agriculture University with support from CARE Indonesia. IRI is currently working with the Indonesian Meteorological Service (BMG) to compare methodologies for forecasting relevant to fires.

Figure 3 Indonesia Fire Map Room is based on research on peatland fires in the Indonesian province of Central Kalimantan that has uncovered a close correlation between satellite rainfall data and fire hotspot activity.
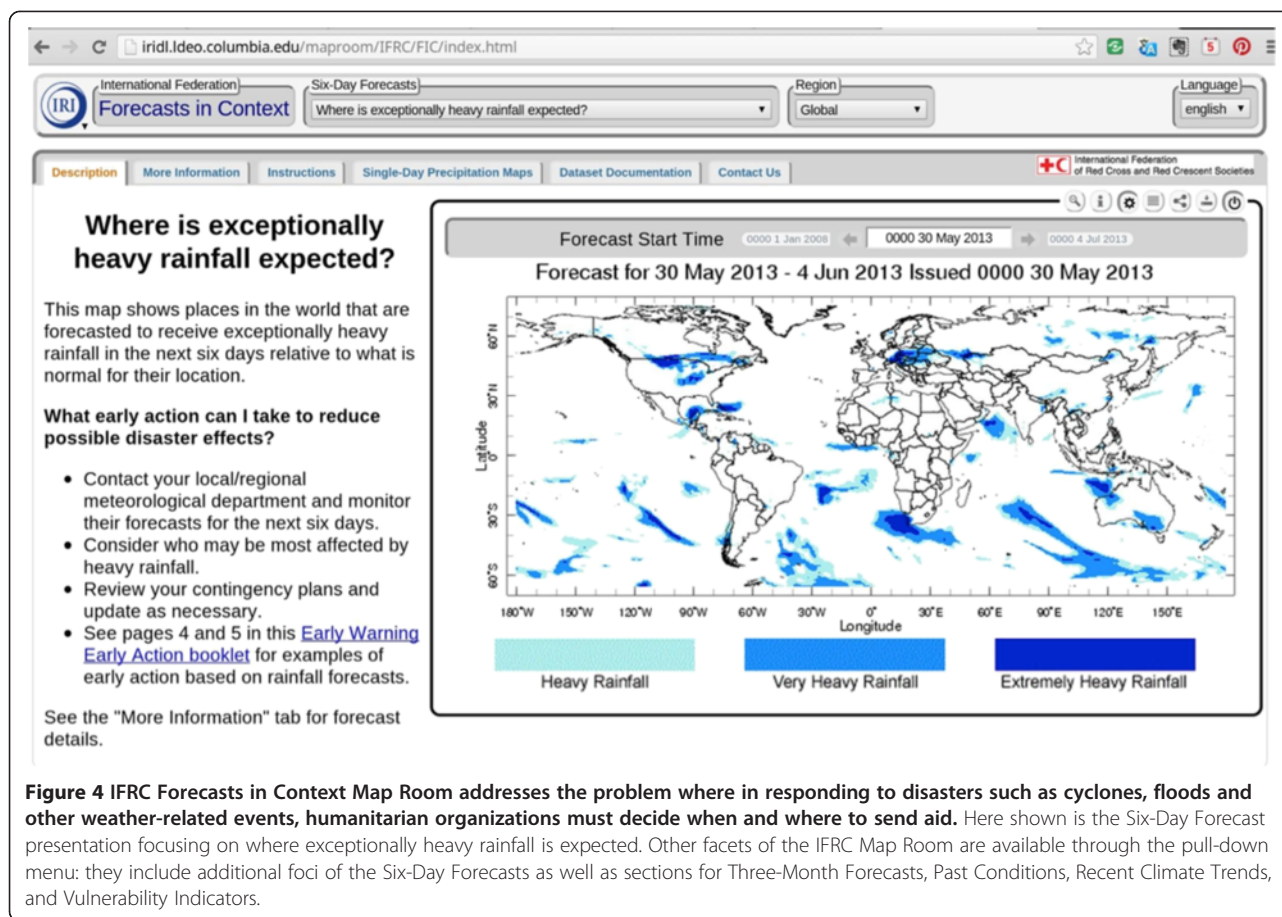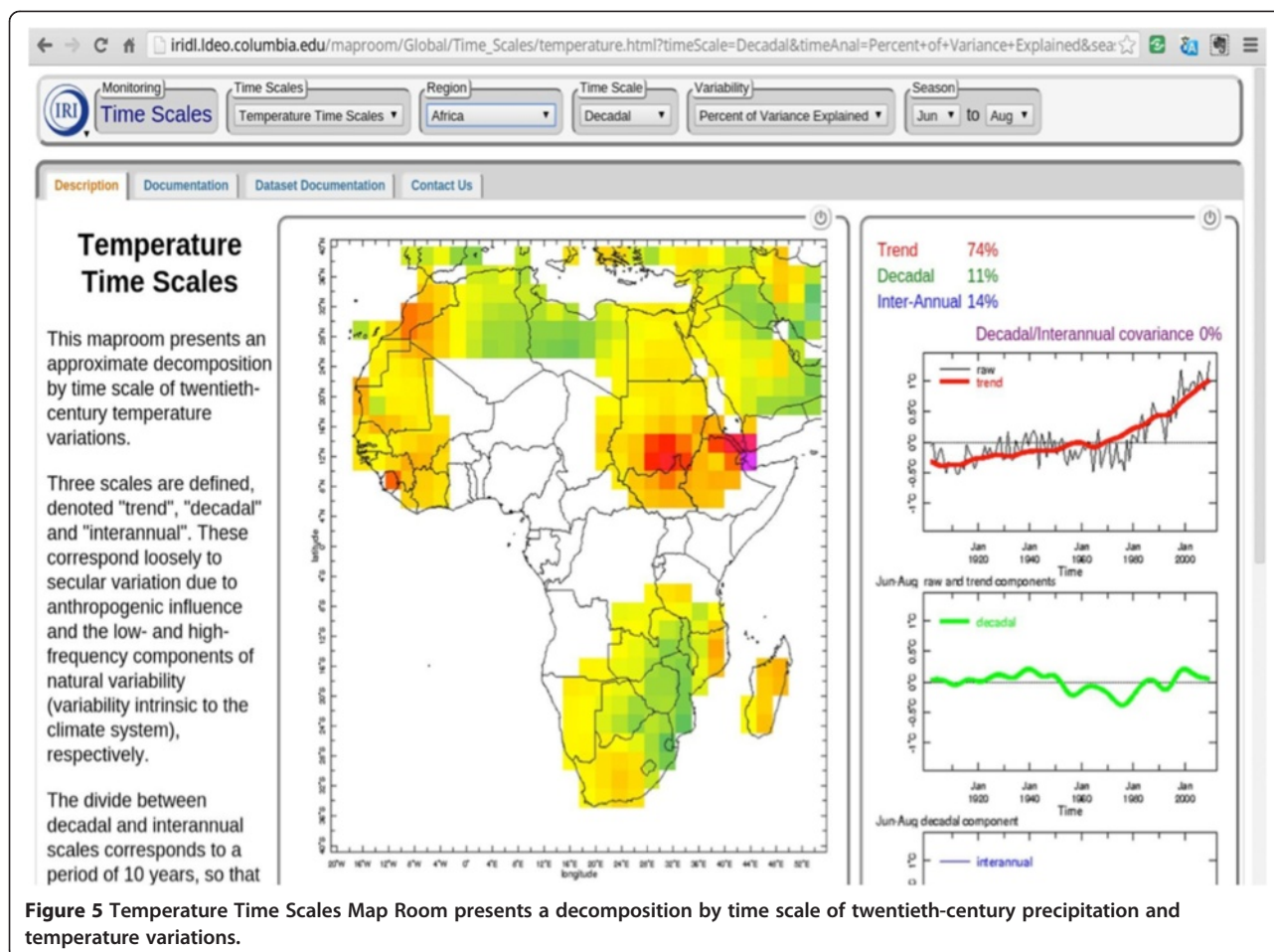
Figure 4 IFRC Forecasts in Context Map Room addresses the problem where in responding to disasters such as cyclones, floods and other weather-related events, humanitarian organizations must decide when and where to send aid. Here shown is the Six-Day Forecast presentation focusing on where exceptionally heavy rainfall is expected. Other facets of the IFRC Map Room are available through the pull-down menu: they include additional foci of the Six-Day Forecasts as well as sections for Three-Month Forecasts, Past Conditions, Recent Climate Trends, and Vulnerability Indicators.

Other maprooms permit access to sophisticated analyses. The Time Scales Map Room (Figure 5, (Greene et al. 2011) presents a decomposition by time scale of twentieth-century precipitation and temperature variations. It sheds light on the characteristics of historical temperature and precipitation variability, in the process clarifying the potential utility of different types of climate information in the context of anticipated climate-related risks that will tend to vary as well, with slower variations modulating the likelihood of adverse or beneficial events that play out on shorter timescales. Three scales are defined and correspond to i) secular variation due to anthropogenic influence, ii) an interannual component of natural variability, and iii) a decadal component of natural variability. Natural variability is variability intrinsic to the climate system, and interannual and decadal are separated by a cutting period of 10 years. Consequently the variability due to the El Niño-Southern Oscillation is classified as interannual, while variability on timescales of 10 years or longer is classified as decadal. The user may define a season of interest and results display as a map of variance explained or standard deviation, and as time series at a given location. Data processing consists of linear regression in order to extract slow, trend-like

changes and low-pass filtering (Butterworth), to separate high and low frequency components in the detrended data. Another version is available in the IFRC Map Room where the statistical values are categorized by degree of importance of variability to provide less technical tailored information for planning purposes on different timescales. It uses data from CMIP3 multi-model ensemble mean representing the secular variation due to anthropogenic influence and the scale decomposition is applied to monthly mean precipitation and temperature from CRU TS3.1.

The Flexible Forecast Map Room (Figure 6) consists of probabilistic temperature or precipitation seasonal forecasts based on the full estimate of the probability distribution, an extension to the more traditional three tercile forecast. Probabilistic seasonal forecasts from multi-model ensembles through the use of statistical calibration, and, based on the historical performance of those models, provide reliable information to a wide range of climate risk and decision making communities, as well as to the forecast community. The flexibility of the full probability distributions allows delivery of interactive maps and point-wise distributions that become relevant to user-determined needs, since probability of exceeding

Figure 5 Temperature Time Scales Map Room presents a decomposition by time scale of twentieth-century precipitation and temperature variations.
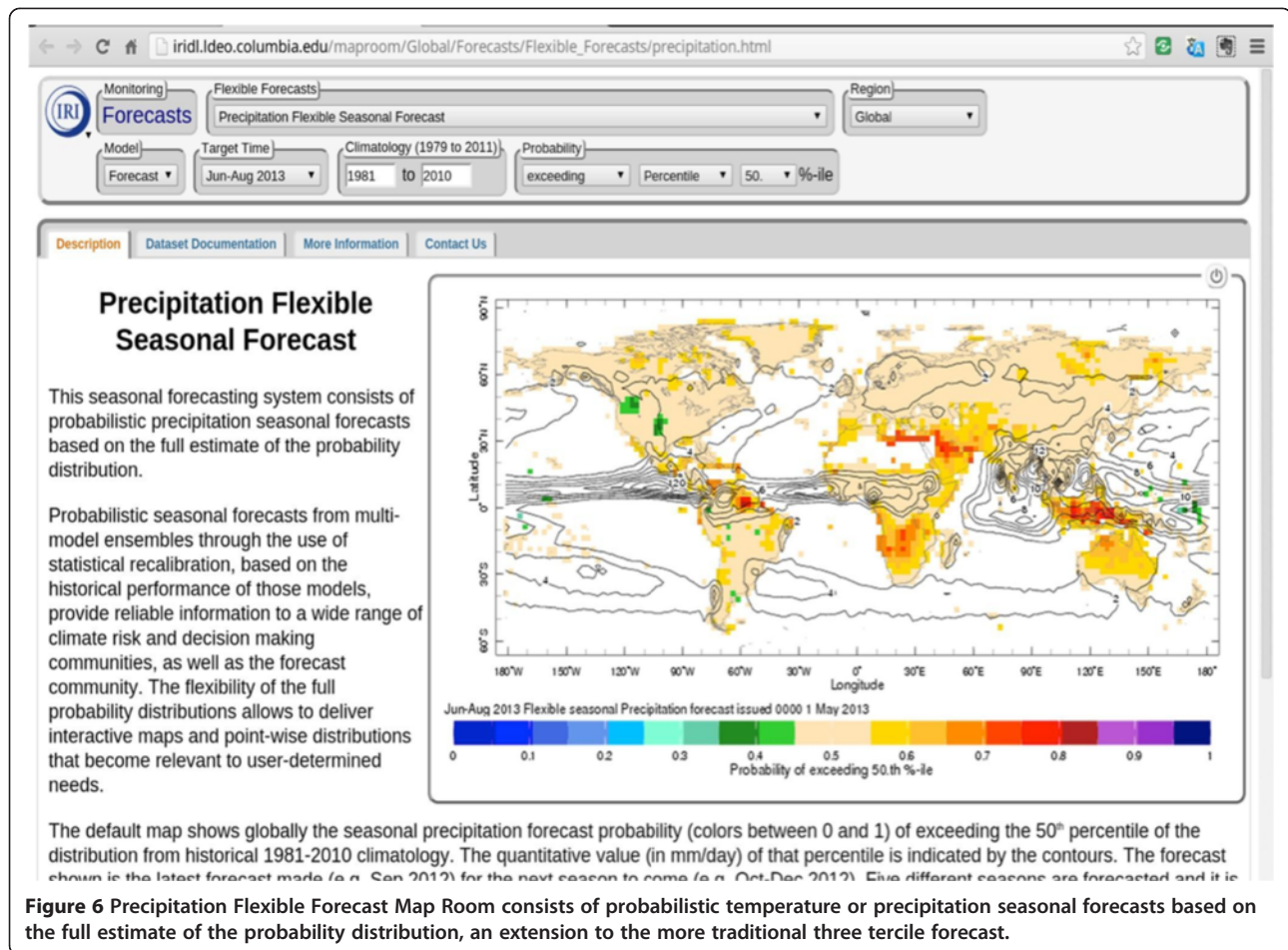
a user-defined historical percentile is actionable. It allows the users to tailor the forecast to real-world problems that may vary from malaria control planning to disaster risk management to hydropower management, to name just a few. It uses historical observations of monthly temperature from CAMS and precipitation from CMAP combined with IRI forecast data.

The Drought Map Room (Figure 7) was developed with funding from the NOAA Climate Test Bed and in collaboration with partners from the NOAA Earth System Research Laboratory and Climate Prediction Center (CPC) to produce quantifiable, probabilistic forecasts of drought over the U.S. and Mexico a few months in advance using the standardized precipitation index (SPI, (McKee et al. 1993) as an indicator of precipitation deficits. The maproom includes analyses of past and current drought using the CPC Unified and U.S. Climate Divisions precipitation datasets as observational inputs at SPI accumulation periods of 3, 6, 9, and 12 months. The user can view maps of the SPI analysis at each of the accumulation periods and click on the map to view time series of the SPI at the selected location over recent years, with the D0-D4 drought severity thresholds from

the North American Drought Monitor indicated. Two methodologies are used to produce probabilistic drought forecast maps. The first forecast map tool uses an "optimal persistence" method (Lyon et al. 2012) based upon the correlation between SPI calculated for recently-observed precipitation and SPI calculated for a future month using recently-observed and historically-observed mean precipitation. The SPI Multi-Model Ensemble Forecast Tool builds upon the Persistence Tool, but uses forecast SPI values based upon precipitation from the IRI Multi-Model Ensemble (MME) at locations, starting months, and leads where hindcast correlation skill from the IRI MME improves over the Persistence method. Both tools display maps of the probability of SPI falling below a user-selected threshold and the forecast SPI in a future month for a user-selected marginal probability.

Analysis tools

The IRI Data Library is a framework that allows easy application of analysis filters to a wide variety of data. There are several important factors that distinguish it as an analysis framework.

Figure 6 Precipitation Flexible Forecast Map Room consists of probabilistic temperature or precipitation seasonal forecasts based on the full estimate of the probability distribution, an extension to the more traditional three tercile forecast.
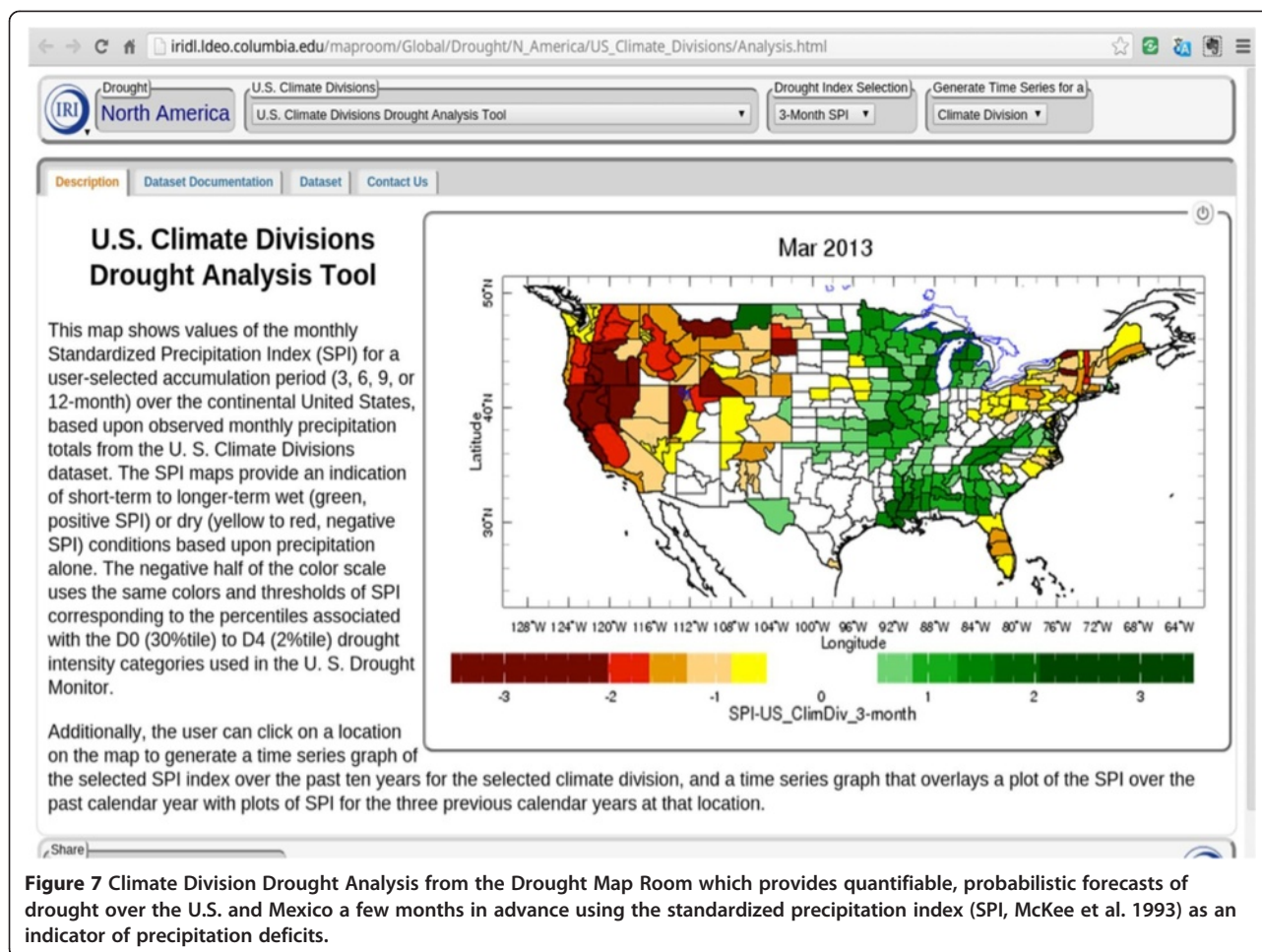
1. Data are organized into datasets comprised of sub-datasets and multi-dimensional variables with use metadata: these variables can be quite large (terabytes) with many dimensions, so that a single variable can conceptually unify what in practice may be many files spread across many directories, details the user can ignore (or be blissfully unaware).

2. Analysis filters usually return variables (sometimes datasets), i.e. data with associated use metadata. This means filters can be chained together, any analysis result behaves as if it were a named dataset. In fact a number of variables named in the dataset collection are analyses based on other datasets.

3. Specifying a calculation is separate from actually executing it, so that chains of calculations processing large amounts of data can be specified and manipulated while the actual execution of the data flow (or portions thereof) is delayed until it is actually required. This allows one to think in the abstract about manipulating the entire dataset, yet actually access it one portion at a time. It also allows shifting the responsibility of efficiently arranging the

calculation away from the user, who can then focus on the actual scientific and statistical analysis.

This easy access to analysis filters is particularly useful in training. The Climate Information for Public Health course (Cibrelus and Mantilla 2010), for example, is intended to engage decision makers directly, not just through expert lectures, but also through focused discussions and practical training sessions. These sessions introduce the participants to geographical information system (GIS)-based computational tools for analyzing epidemiological data with climate, population and environmental data. To allow the students to focus on the course content and still be able to analyze their own data in the context of available climate information, we have built services that allowed them to access and analyze their own data within the Climate Data Library, as well as adding analysis functions particularly useful in health analyses, ranging from k-means clustering to disease epidemic threshold calculations. This course and its tools have been taught in an annual Summer Institute and in sessions around the world, some using the Standalone Data Library (see below).

Figure 7 Climate Division Drought Analysis from the Drought Map Room which provides quantifiable, probabilistic forecasts of drought over the U.S. and Mexico a few months in advance using the standardized precipitation index (SPI, McKee et al. 1993) as an indicator of precipitation deficits.

Another example of advanced analysis using the IRI Data Library is to create spatial-temporal maps of malaria incidence using health surveillance data. Monthly data on clinical malaria cases from 242 health facilities in 58 *subzobas* (district boundaries from the National Statistics and Evaluation Office) in Eritrea from 1996 to 2003 were used in a novel stratification process to guide future interventions and development of an epidemic early warning system. The process used principal component analysis and nonhierarchical clustering to define five areas with distinct malaria intensity and seasonality patterns and has been used by the Eritrean Malaria Control program in its planning process (Ceccato et al. 2007b).

Semantic Technology
Often in a research community there are several different metadata standards used to describe the same object. Associated with each metadata standard is a conceptual model, frequently not explicit, which describes the object in its own way. We are using an RDF/XML (Resource Description Framework) framework to address this issue, and create a flexible, reusable solution that can adapt to a variety of new metadata standards. It implements a semantic framework for explicitly writing down multiple metadata schema and conceptual models as *ontologies*; the ontologies identify metadata elements and concepts and characterize the relationships between them. We also use the framework to write crosswalks, i.e. explicit characterizations of the relationships between concepts and metadata elements belonging to different systems, including the connections between the metadata objects and the concepts they represent. Not only does this framework allow translation between alternate systems, it also facilitates building a more complete description of data objects out of a number of narrowly-focused standard systems. Going beyond standards, it can explicitly describe the data models implicit in programs that display and manipulate data. Writing Models, Crosswalks, and Objects all with RDF/SemanticWeb means that these data models and metadata standards can be combined into a single framework, leading to an interoperable metadata standard (Blumenthal et al. 2011).

Crosswalking between different standards can be as simple as two different names for the same quantity, but

sooner or later the mapping gets more complicated. Frequently, different objects are related conceptually but are very different structurally. Our framework thus has both structure and conceptual models. Structure models describe how dataset metadata is written, e.g. cfatt which describe the attributes of a Climate and Forecast Metadata (CF) Convention netcdf file. Conceptual models describe the conceptual objects represented in the convention, e.g. cf-obj which describes the more abstract objects (like geo-located data) that are being described in the CF convention, objects that are also described in other systems, but are not explicitly written in any given CF netcdf file. XML Schema is a common way to represent structure models for XML files, and we have a translation of XML Schema to RDF/OWL which allows us to create conforming XML files from RDF information. We have applied this to the WCS Schema, for example, to extract the needed information for an OPeNDAP WCS service based on RDF extracted from CF/netcdf files. We also have included controlled vocabularies such as CF standard names or GCMD scientific parameters. Controlled vocabularies are a common way to structure classifications, and important for us to build a faceted search that works across diverse datasets.

The framework is established by creating ontologies for each metadata representation of these objects, and rule-based crosswalks between them so that each object is expressed in all representations, thus all objects can be viewed in multiple systems. This technology has been encapsulated in a Java based persistence/inferencing framework for OPeNDAP (Cornillon et al. 2009) as part of a NOAA/IOOS project (Holloway et al. 2010). This work combines custom innovations, the use of ontologies, and leading Semantic Web technologies, such as, Sesame and OWLIM. Because this framework was developed on Java technology, the system is highly portable between various platforms.

We also developed an XML element extraction system based on Java, which allows the extraction of information from the framework into an XML format that is based on data description and delivery standards (WMS, SERF, etc.). With these tools we can further develop technologies of delivering climate data and analysis to partner systems.

Results
Merging standard climate products with GIS information
A central part of the IRI Data Library's functionality is that it brings a wide variety of data together into a framework that allows that data to be analyzed together. The framework is sufficiently general that it overcomes the differences that disparate domains can have, while able to represent the results of the analysis so that it can be used for further analysis.

For example, consider different ways that data can be characterized geospatially. Atmospheric and oceanic scientists tend to have multi-dimensional data, a simple case being where temperature or precipitation is characterized as being a function of latitude, longitude, and time. Health or economics sciences, on the other hand, tend to have data as a function of time and geospatial entity, which might be a district or state or census tract. GIS data tends to have two structures, either a raster image with associated projection information, or vector descriptions of shapes, describing how to draw geospatial entities as sets of polygons or points. The IRI Data Library combines all three of these geospatial frameworks, allowing interoperability.

A manifestation of that geospatial interoperability is how the Data Library brings together three-dimensional (longitude, latitude, time) climate data and GIS spatial entity descriptions in the MEWS Malaria Map Room mentioned earlier, including a tool which not only displays a zoomable map of precipitation, but also allows selection of a district and displays downloadable time series computed by averaging the three dimensional data over the district (Figure 1b). This combination of data to get time series for a particular geographic entity is an important enhancement to the accessibility of the original data, for users whose other information is geolocated by entity. In this case, the user requests the analysis simply by clicking, but the underlying analysis functions can be used to combine many data sets and geographic entities.

Simplified access/transformation of large datasets
An essential role of the Data Library has been to enhance the value of publicly available disparate datasets by bringing them into a single framework that allows them to be analyzed together. One way we have done this is by transforming large datasets from their reference format (frequently large numbers of files in specialized formats with either purely descriptive access documentation or highly specialized metadata) to a more conceptual structure that allows the user to make selections and analyses in space, time, and physical variable without mastering the original multipart structure, the requests being made in the same way whether the variable represents megabytes or terabytes. The data can then be directly analyzed, partially analyzed to reduce its size and/or make the data more suitable for the users' needs, and displayed and/or downloaded in a wide range of commonly used formats and for many commonly used tools.

Transforming from a structure that is appropriate for a provider to one that is appropriate for a user is a critically enabling step, preventing technical barriers from keeping users from accessing the information. It is also important to remember that both the provider and the user have critically important needs. While the user needs to analyze the data as a coherent whole, the provider needs to characterize the data with clean provenance: which parts of the dataset were created when,
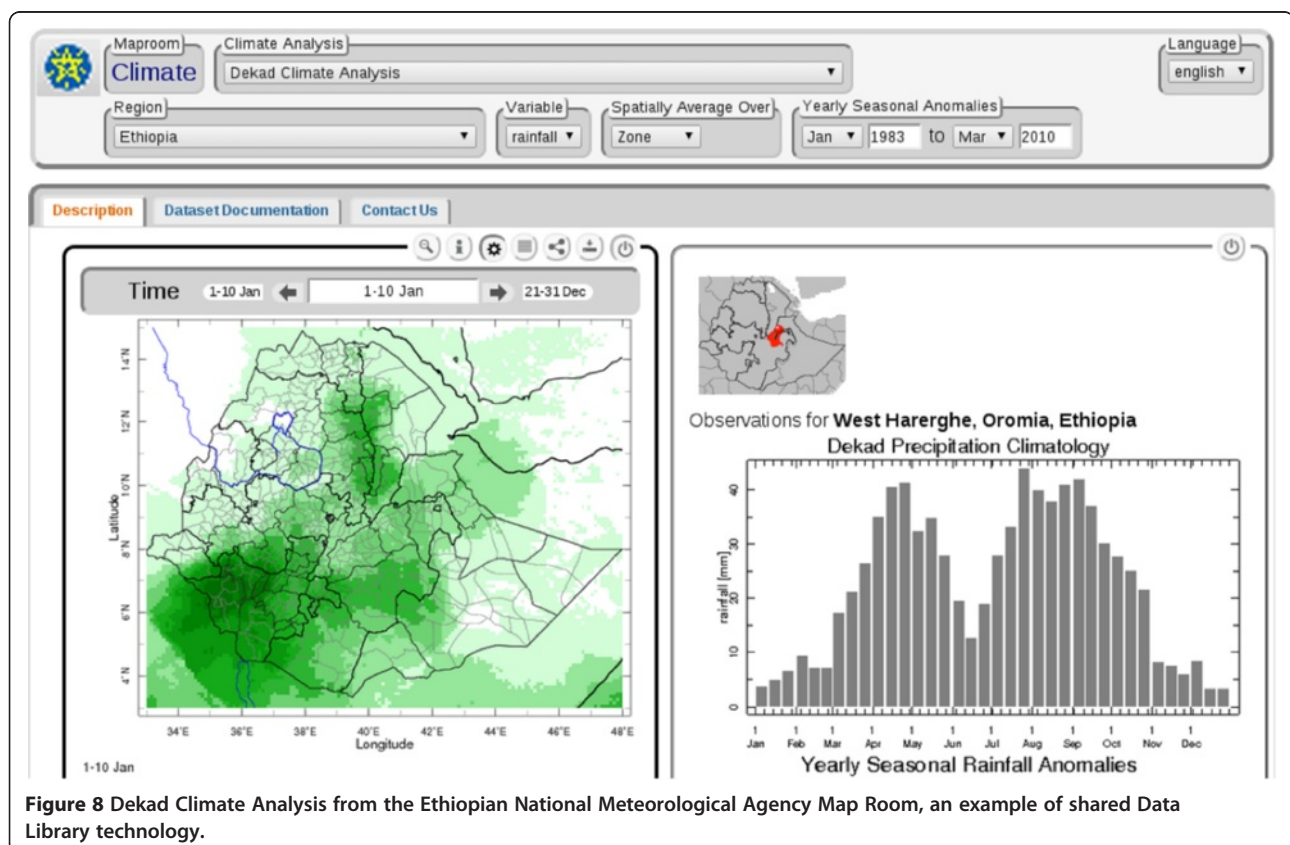
were any parts of the dataset revised, and as the dataset is extended in time what new segments have been added. Simply stating that each extension to the dataset creates an entirely new one, for example, means that anyone trying to track the changes in the dataset would falsely think there is an enormous volume of data in keeping all the versions.

Shared Data Library Technology

While the map rooms and the IRI Data Library Server represent using the IRI Climate Data Library as a service, additionally the technology has been directly shared with IRI partners in two ways: a laptop configuration and a bootable USB disk drive configuration. The laptop configuration has been used by the Data Library team and IRI scientists to conduct training and instruction in parts of Ethiopia, Niger, Madagascar, and Indonesia where internet connectivity is intermittent, slow, or non-existent. The bootable USB disk drive configuration has been deployed to Niger (ACMAD and AGRHYMET), Ethiopia (National Meteorological Agency, Figure 8, (Dinku et al. 2011), India (IIT-Delhi and IMD), Tanzania (Tanzania Meteorological Agency), Chile (CEAZA and CAZALAC) and Indonesia (CCROM). The USB disk drive is connected to an in-country computer and runs the Data Library software in a standalone or mirror configuration.

The standalone Data Library configuration for partners is used in situations where targeted data services are needed. This may be in a region where only regional data are to be analyzed and delivered over the internet. The mirror Data Library configuration is used where the partner would like to see parts of or all the IRI Data Library datasets in their local configuration. This configuration allows the partner to view, analyze, and deliver both local regional data and global data, as well as allow the partner to share their local data with the IRI Data Library in a seamless data catalog available to both the partner and the IRI. This technology can be used to create a federation of remotely deployed Data Library sites and the IRI Data Library. This means that locally stored data can be shared globally over the internet among federation members.

The Data Library technology incorporates a content delivery service over local area and wide area networks. The portable Data Library can be used in a classroom setting (local area network) where all the visualization, analysis, and data delivery capabilities of the software are accessible to each student sitting in front of a computer or tablet that is running a browser. In a wide area network setting, the portable Data Library can be used as a self-contained website, or as part of an existing website. The start-up costs for a partner to implement a ready-

Figure 8 Dekad Climate Analysis from the Ethiopian National Meteorological Agency Map Room, an example of shared Data Library technology.

made website for climate data can be minimized by using a portable Data Library.

When partners are evaluating the risk factors of climate change and variability on various sectors they encounter two impediments. One is the inability to bring sectoral and climate data together in a unified framework for comprehensive analysis. The other impediment is that often, government ministries (or departments within a single ministry) are reluctant to share data. A portable Data Library brought into a region can help remove some of these barriers. It is a neutral platform that can be installed in almost any (neutral) location. The functions within the Data Library software can be used to align climate data with the spatial and temporal resolutions of the sector data of interest. Once this alignment is performed, the correlation and statistical functions within the Data Library can be used to determine the climate risk factors affecting a sector or sectors in the partner's region.

Conclusions

The IRI Data Library is a key platform that makes climate and other data products more widely accessible through tool development, data organization and transformation, and data/technology transfer. Tools developed include Map Rooms which are designed for rapid access to needed information for particular user groups, analysis tools useful for a wide range of users, tools to merge standard climate products with GIS information (e.g. boundaries of political entities used to geolocate health and socio-economic data), and simplified access/transformation of large datasets only available as collections of large numbers of files or service points elsewhere. We have developed a metadata framework that uses semantic technologies to transform metadata from a variety of sources into a variety of standards. We have also shared Data Library technology with partners to assist them in their own data sharing and access.

Authors' contributions
MBB is lead author and designer/builder of the Data Library, MB authored the Drought Analysis Tool among others, JdC leads in Shared Data Library Technology, RC authored the Time Scales and Flexible Forecast Maprooms, IK facilitates Shared Data Library Technology. All authors read and approved the final manuscript.

Responsible editor: Xiubin Li.

References
Blumenthal MB, del Corral JC, Liu H, Holloway D, Potter N (2011) Semantic Framework for climate metadata interoperability (T248A). In: WCRP Open Science Conference; 24–28 Oct., Denver CO, USA

Ceccato P, Bell MA, Blumenthal MB, Connor SJ, Dinku T, Grover-Kopec EK, Ropelewski CF, Thomson MC (2006) Use of Remote Sensing for Monitoring Climate Variability for Integrated Early Warning Systems: Applications for Human Diseases and desert Locust Management. In: IGARSS Denver: IEEE International Conference on Geoscience and Remote Sensing Symposium

Ceccato P, Cressman K, Giannini AS, Trzaska S (2007a) The desert locust upsurge in West Africa (2003–2005): Information on the desert locust early warning system and the prospects for seasonal climate forecasting. Int J Pest Manag 53:7–13

Ceccato P, Ghebremeskel T, Jaiteh M, Graves PM, Levy M, Ghebreselassie S, Ogbamariam A, Barnston AG, Bell MA, del Corral JC, Connor SJ, Fesseha I, Brantly EP, Thomson MC (2007b) Malaria stratification, climate, and epidemic early warning in Eritrea. Am J Trop Med Hyg 77(6):61–68

Cibrelus L, Mantilla G (2010) Climate Information for Public Health: A Curriculum for Best Practices - Putting Principles to Work. International Research Institute for Climate and Society Report, Palisades, New York

Connor SJ, Omumbo J, DaSilva J, Green C, Mantilla G, Delacollette C, Hales S, Rogers D, Thomson MC (2010) Health and Climate - Needs. Procedia Environmental Sci 1:27–36

Cornillon P, Adams J, Blumenthal MB, Chassignet E, Davis E, Hankin S, Kinter J, Mendelssohn R, Potemra JT, Srinivasan A, Sirott J (2009) NVODS and the development of OPeNDAP. Oceanography 22(2):116–127

del Corral JC, Blumenthal MB, Mantilla G, Ceccato P, Connor SJ, Madeleine C, Thomson MC (2012) Climate Information for Public Health: the role of the IRI Climate Data Library in an Integrated Knowledge System. Geospat Health 6(3):S15–S24

Dinku T, Hailemariam K, Maidment R, Tarnavsky E, Connor S (2013) Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. Int J Climatol doi:10.1002/joc.3855

Dinku T, Hilemariam K, Grimes D, Kidane A, Connor SJ (2011) Improving availability, access and use of climate information. World Meteorological Bulletin 60:2

Greene AM, Goddard L, Cousin R (2011) Web tool deconstructs variability in twentieth-century climate. Eos Trans AGU 92(45):397

Grover-Kopec EK, Kawano M, Klaver RW, Blumenthal MB, Ceccato P, Connor SJ (2005) An online operational rainfall-monitoring resource for epidemic malaria early warning systems in Africa. Malar J 4:6

Holloway D, Blumenthal MB, Liu H, Potter N (2010) Using Semantic Web Technologies with OPeNDAP. In: 2010 Fall Meeting, AGU, 13–17 Dec, 2010; San Francisco, Calif. IN41C-1369

IRI (2006) A Gap Analysis for the Implementation of the Global Climate Observing System Programme in Africa. In: International Research Institute for Climate and Society. Palisades, NY

Jacquez GM (2000) Spatial analysis in epidemiology: Nascent science or a failure of GIS? J Geogr Systems 2:91–97

Kelly-Hope L, Thomson MC (2008) Climate and infectious diseases. *Seasonal Forecasts*. In: Climatic Change and Human Health, pp 31–70

Lyon B, Bell MA, Tippett MK, Kumar A, Hoerling MP, Quan X, Wang H (2012) Baseline probabilities for the seasonal prediction of meteorological drought. J Appl Meteorol Climatol 51:1222–1237

McKee TB, Doesken NJ, Kliest J (1993) The relationship of drought frequency and duration to time scales. In: Eighth Conf. of Applied Climatology, Amer. Meteor. Soc.; Anaheim, CA, pp 179–184

Someshwar S, Boer R, Conrad E (2010) Managing Peatland Fire Risk in Central Kalimantan, Indonesia. Washington DC, World Resources Report

Thomson MC, Connor SJ, Zebiak SE, Jancloes M, Mihretie A (2011) Africa needs climate data to fight disease. Nature 471:440–442

doi:10.1186/2194-6434-1-19
Cite this article as: Blumenthal *et al.*: IRI Data Library: enhancing accessibility of climate knowledge. *Earth Perspectives* 2014 **1**:19.