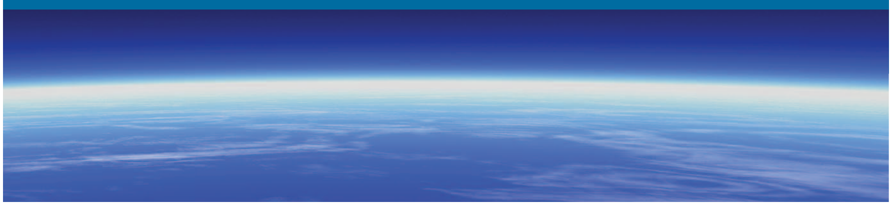
Sharing Satellite Observations with the Climate-Modeling Community: Software and Architecture

Daniel J. Crichton, Chris A. Mattmann, Luca Cinquini, Amy Braverman, Duane Waliser, Michael Gunson, Andrew F. Hart, and Cameron E. Goodale, NASA Jet Propulsion Laboratory

Peter Lean, University of Reading

Jinwon Kim, University of California, Los Angeles

// A software infrastructure for direct comparisons of climate-model outputs with satellite observation data helps tune the models and provides ground truth in understanding the Earth's climate processes. //



UNDERSTANDING GLOBAL CLIMATE change and its impact is one of the great scientific problems of our time. The scientific community is working to build models of the Earth's climate system and to acquire observational data, including remotely sensed data, that will facilitate effective decision-making and predictions.[1,2] These efforts are tightly interwoven. Scientists analyze observational data to formulate and test hypotheses about climate physics. Robust hypotheses become part of the community's knowledge base, encoded into physical process models. In turn, the models become test beds for experimentation to better understand perceived causal relationships and, eventually, find their way into global climate models (GCMs). At various points in this process, model predictions are compared with independent observations to assess whether the two agree. In most cases, this process is iterative and requires understanding how the observations were collected as much as the inner workings of the model.

At the NASA Jet Propulsion Laboratory (JPL), we've begun efforts to leverage NASA's observational datasets to assist in comparing climate-model outputs in the context of the Intergovernmental Panel on Climate Change (IPCC) and its upcoming Fifth Assessment Report (AR5).[1] NASA's data represents an enormous, geographically distributed information source with global coverage that can improve models ranging from individual physical process models on small space-time scales to regional and coupled GCMs.

Although the benefits of comparing satellite observations to climate-model output are perhaps intuitive, software that performs this functionality poses difficult software engineering challenges. For example, the heterogeneity of the underlying data and metadata formats is a significant hurdle. NASA observational data is prepared as Hierarchical Data Format (HDF) version 4/5 data with HDF-Earth Observing System (EOS) metadata.[3] The model outputs from IPCC assessments utilize network Common Data Format (NetCDF) version 3/4,[4] coupled with climate and forecast (CF) metadata.[5] (NetCDF is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of scientific data. The CF metadata conventions are designed to promote the processing and sharing of NetCDF data files for climate and forecast applications.) Developing software libraries for reading, writing, and converting between these and other formats involves significant engineering effort.

> The heterogeneity of the underlying satellite and climate-model output data and metadata formats is a significant hurdle.

There are also computationally intensive challenges, stemming from the scientific differences between remotely sensed observations and climate-model outputs. Specifically, minimally processed NASA remote-sensing products are seldom globally gridded but instead represent localized swath-based imagery that doesn't exhibit the same temporal and spatial resolution as globally gridded model outputs. This typically requires complex software engineering efforts to regrid entire datasets, which involve extracting variables and parameters from highly distributed data repositories as well as aggregating, subsetting, and reformulating new data corresponding to some uniform space-time model grid. In practice, this process can require orchestrating efforts that span organizations, datacenters, and both political and technological barriers.

The climate-modeling and satellite-observation communities have largely acquired and developed their own computing infrastructures, but efforts are now in place to link these infrastructures together in a multiagency implementation. Here, we describe current JPL efforts to create software that meets the challenges of making large-scale comparisons between remote-sensing data and climate-model outputs (for related work, see the sidebar). Our goal is to inform the broader community of what worked—and what didn't—so that the next generation of software generated in this area can benefit from the lessons learned.

Data Analysis in NASA's Earth Science Enterprise

Leveraging NASA's massive, heterogeneous, distributed observational data collection requires a new approach to access and analysis and an infrastructure to support those activities.

The NASA Earth Science Enterprise encompasses projects that combine three distinct types of data storage and processing centers. Distributed Active Archive Centers (DAACs) represent the public-facing data marts for NASA's Earth science information and provide a well-organized, efficient virtual repository for NASA's observational assets. However, access to DAAC data isn't currently optimized for analysis and comparison with climate models. The same is true of the NASA data products distributed from Science Information Processing Systems (SIPS), which often produce finer-resolution observational data but focus on algorithm development, data processing, and ultimate delivery to the DAACs. Finally, downstream from the DAACs are ad hoc, proposal-funded data systems that deal with data production and access.

In this enterprise, users are responsible for discovering content as well as understanding layout, format, sampling characteristics, and other important properties of observational data products. Furthermore, observational datasets are generally not aligned with the common model-output formats and often require side-by-side examination for conversion. Converting between data (and metadata) formats is a resource-intensive engineering activity that involves existing software libraries, glue code, and a firm understanding of the details of the underlying data representations.

Comparison analysis is both data- and resource-intensive. A simple analysis task might require subsetting data from two different instruments by space and time, filtering out missing data and outliers, and regridding to a common resolution. Ordinarily, users obtain a superset of the data they actually need and subsequently filter and regrid it themselves. In practice, this requires moving large volumes of data over a network link, much of which might ultimately be discarded.

The burden of moving data can be substantially eased by performing the computation itself—for example, filtering and regridding—at the site where the data resides. Moving the computation to the data (instead of the other way around) requires new thinking about distributed processing environments. In the context of our work, three components of these environments have repeatedly emerged:

RELATED WORK IN ACCESS TO EARTH OBSERVATION DATA

Several organizations have contributed to making Earth observation data available to interdisciplinary researchers, especially in climate modeling.

GIOVANNI

NASA Goddard Earth Sciences (GES) division designed Giovanni to enable exploration and visualization of its Earth observation data without requiring specialized tools or software.[1] (Giovanni is an abbreviation for GES Interactive Online Visualization and Analysis Infrastructure.) The Giovanni team has built a Web application that lets users select an instance or specific study field that corresponds to identified data of interest. Once the down-selection process is complete, users can leverage the Web browser to generate visualizations, perform analysis, and download the generated products from the dataset.

Giovanni services a broad range of research interests in general, whereas Climate Data eXchange (CDX) and Regional Climate Model Evaluation System (RCMES) specifically support the advancement of climate research. Giovanni has developed an accessible, intuitive user interface but hasn't yet focused on model-to-data comparisons. Our work has focused on modularity and extensibility, allowing researchers to use existing tools (like Giovanni) or to develop their own tools that communicate with RCMES and CDX via Web service APIs.

PYNGL/PYNIO

The Computer and Information Systems Laboratory at the National Center for Atmospheric Research (NCAR) designed the PyNGL and PyNIO tools as Python modules that let users access the NCAR Command Language (NCL) from within Python (www.pyngl.ucar.edu). From this interface into NCL, users can open various scientific file formats, such as Hierarchical Data Format (HDF), Network Common Data Format (NetCDF), Grib, and Coordinate Measuring Machine. PyNGL also incorporates a visualization component that can generate scientific and statistical plots.

PyNGL and PyNIO are powerful technologies that require the user to have a working knowledge of Python and the target dataset's format and structure. Both tools have been instrumental in the development of CDX and RCMES. When designing CDX and RCMES, we provided abstraction over these core technologies, enabling users to work with CDX data and services without requiring Python experience.

The evolving ecosystem of both mature and developing software for scientific data access and analysis makes the CDX framework and RCMES toolkits well-timed mechanisms for composing scalable data systems to meet the software engineering challenges of comparing observational data to model output.

EARTH SYSTEM GRID

Earth System Grid (ESG)[2] and the Earth System Grid Federation (ESGF) represent an international partnership of institutions focused on making the next generation of climate-model outputs available to scientists and other users throughout the world. ESG is built on a peer-to-peer software platform with two canonical peer types:

- a *data node* is a software stack for preparing, transforming, and delivering climate-model output data (and more recently, through our work) to ESG gateways; and
- a *gateway* is a data portal providing search, discovery, access, and dissemination of data.

Our efforts have led us to develop strong partnerships with the ESG and ESGF principals. Prior to the most recent Intergovernmental Panel on Climate Change proposed assessment,[3] the ESG platform focused solely on climate-model outputs. However, with our assessment efforts with CDX, ESG can expand to bring the wealth of NASA observations to bear on the available climate-model outputs.

EARTH SCIENCE COLLABORATORY

The Earth Science Collaboratory (ESC) project is an emerging effort from the Earth Science Information Partners (ESIP) Federation and its community to provide a platform for international scientific collaboration. The ESC will offer scientific "work benches" that let researchers dynamically execute workflows and other tools, store the results, and ultimately annotate, share, and disseminate scientific data and information.

The ESC relates to our efforts to bring observations and climate models together in a unified form. ESC infrastructure could leverage the CDX tooling framework for comparing observations to models and for providing climate-related services. ESC could also leverage the ESG platform.

References

1. S.W. Berrick et al., "Giovanni: A Web Service Workflow-Based Data Visualization and Analysis System," *IEEE Trans. Geoscience and Remote Sensing*, vol. 47, no. 1, 2009, pp.106–113.
2. D.N. Williams et al., "The Earth System Grid: Enabling Access to Multi-Model Climate Simulation Data," *Bulletin Am. Meteorological Soc.*, Feb. 2009; http://dx.doi.org/10.1175/2008BAMS2459.1.
3. Intergovernmental Panel on Climate Change, "Scoping the IPCC 5th Assessment Report, Proposal for an IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections," IPCCXXX/Doc. 11, IPCC, Apr. 2009; www.ipcc.ch/scoping_meeting_ar5/doc11.pdf.

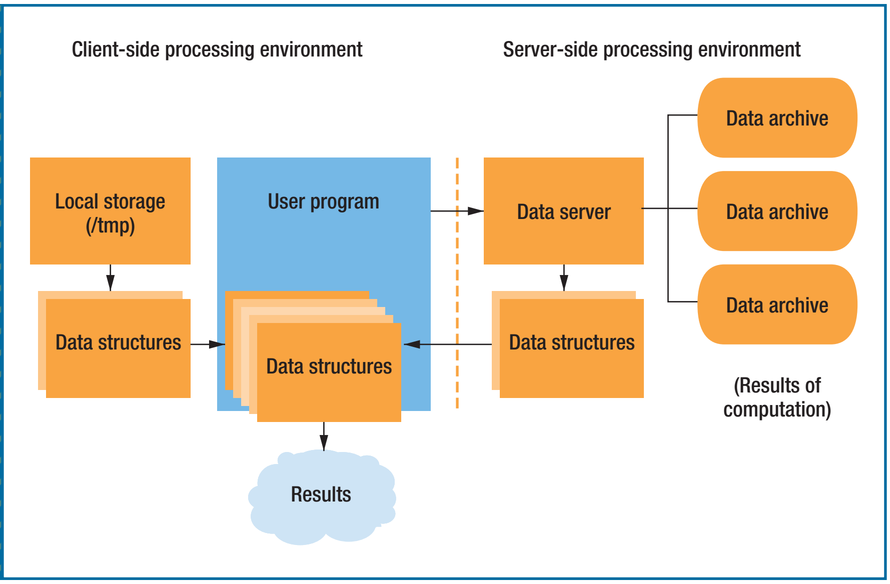


FIGURE 1. The Climate Data eXchange (CDX) architecture and philosophy. CDX aims to move computation to the data in the server-side processing environment and send only the computational results (data structures) back to the requesting client. Both server- and client-side data structures are inputs to a user program, which is agnostic about where the results originated.

- *mission-specific climate services* that provide purpose-driven data access and manipulation, such as transformation to NetCDF or CF and regridding to new space-time scales and resolutions;
- *flexible, interactive user interfaces* to data-access services and computational workflows; and
- *client toolkits* that encapsulate the service and user interface capabilities and facilitate data access within user programs or from the command line.

Software frameworks for distributed environments that collocate processing with the data can mitigate several challenges of dealing with massive datasets.

CDX: An Architectural Framework for Climate Data

In 2008, JPL began development of one such framework, Climate Data eXchange (see Figure 1).[6] CDX is a software analysis environment to deal with climate data challenges. Its core architectural principles include

- separation of services and data,
- remote processing of large datasets,
- metadata-driven discovery, and
- dynamic packaging of computational services for climate data.

CDX supports analysis of model outputs in comparison with distributed observational data. Many CDX functions are unique to climate research, but the services are domain-agnostic. CDX aims to provide both client and server components (see Figure 2). The client component provides software libraries that users can integrate into analysis routines in languages, such as Matlab, IDL, and R, and application environments such as Web portals. On the server side (below the gateway layer in Figure 2), CDX enables users to remotely view available data, retrieve a product from a NASA archive or mission datacenter, perform processing to subset, regrid, or mask directly on the data, and retrieve a remote model. CDX minimizes the data that must move over the network by providing services to perform as much computation as possible where the data natively resides.

The client toolkit is an easily installable client command set and an API for interacting with data made available through the servers. Client tools include operations for virtual listings of data from all servers and for transparently searching, manipulating, transforming, and accessing data as if it were local.

Access to data collections is through commonly agreed-upon protocols and services, which enforces established policies for distributed access control. Consequently, CDX promotes a common interface for connecting distributed data repositories. Protocols are agreed upon by working with the climate-modeling community primarily through the Program for Climate Model Diagnosis and Intercomparison (PCMDI; www-pcmdi.llnl.gov) and the Earth Systems Grid (ESG; www.earthsystemgrid.org) Foundation and with the observational data community, primarily through our role at NASA.

CDX heavily leverages open source software, particularly the Apache Object-Oriented Data Technology (OODT) framework,[3] and an open source development approach. Open source and shared development have been key to the framework's rapid development.

The CDX framework has helped enable two key applications: the publishing of NASA observational data to the ESG platform, and the comparison of regional climate models and satellite observations via an analysis environment established at JPL called the Regional Climate Model Evaluation System (RCMES), which we describe later.

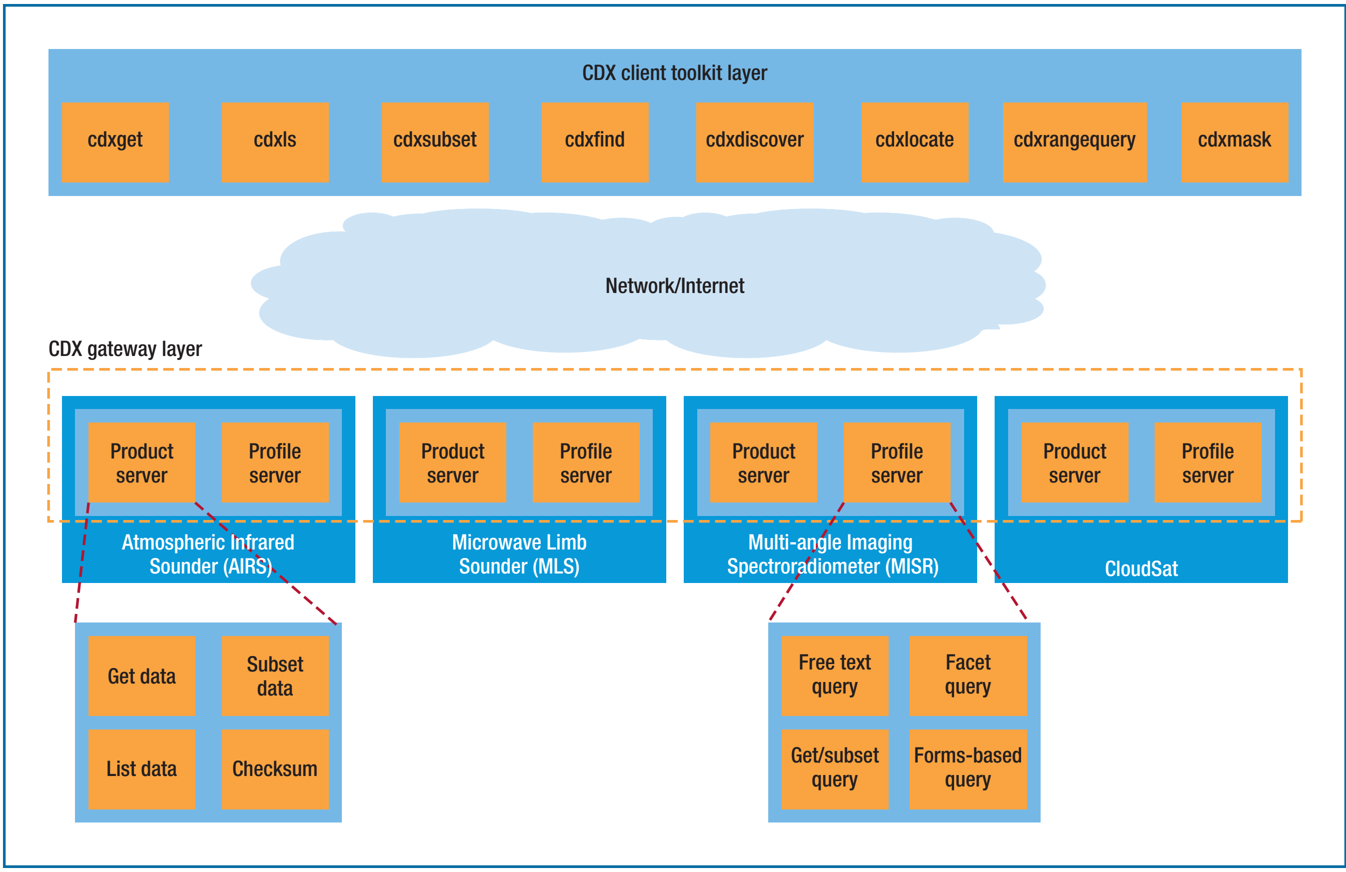


FIGURE 2. The CDX framework. A client toolkit layer hides the distributed nature of the underlying data and services, easing the barrier for scientists to compare model outputs with observations available from NASA missions, such as the Atmospheric Infrared Sounder (AIRS). The toolkit provides a set of functions—cdxget, cdxls, and so on—and the server side of the framework—that is, the gateways to NASA data products—implement the server sides of the distributed functions requested by client tools.

Sharing Satellite Observations for the IPCC Assessment

The upcoming IPCC AR5 report has also motivated efforts to investigate how NASA's observational data holdings can contribute to the analysis of multimodel ensembles. Traditionally, the variability of model predictions across ensemble members have characterized uncertainties about future climate predictions. However, it's increasingly recognized that model predictions differ in their reliability and that component predictions and their uncertainties should be weighted to compensate.[1]

ESG represents an infrastructure for sharing the output of climate-model experiments in this manner.[2] It's a distributed, federated network of data nodes, accessed via gateways, with each node responsible for providing services to access the data it serves. The ESG is designed specifically for model output, which means it benefits from the relative standardization that exists in the modeling community. ESG data conforms to the CF metadata convention[5] and packages all its data in NetCDF[4] with common names and meaning.

As one of the principal organizations that collects remote-sensing data for Earth and climate science, NASA has a natural role to play in extending ESG to serve satellite observational data for use in climate research. Indeed, it has an imperative to develop an independent infrastructure that has architectural notions similar to ESG and is interconnected with it to deliver NASA data to the scientific research community.

In 2008, we received funding from NASA's Innovate Partnership Program (IPP) to prototype the infrastructure necessary to integrate NASA observational data with the ESG platform.[10] We demonstrated ESG access to and sharing of NASA observations and interoperability between the two communities to support future research. We deployed CDX at several SIPS sites and leveraged it to bring observational data to a locally deployed data node before

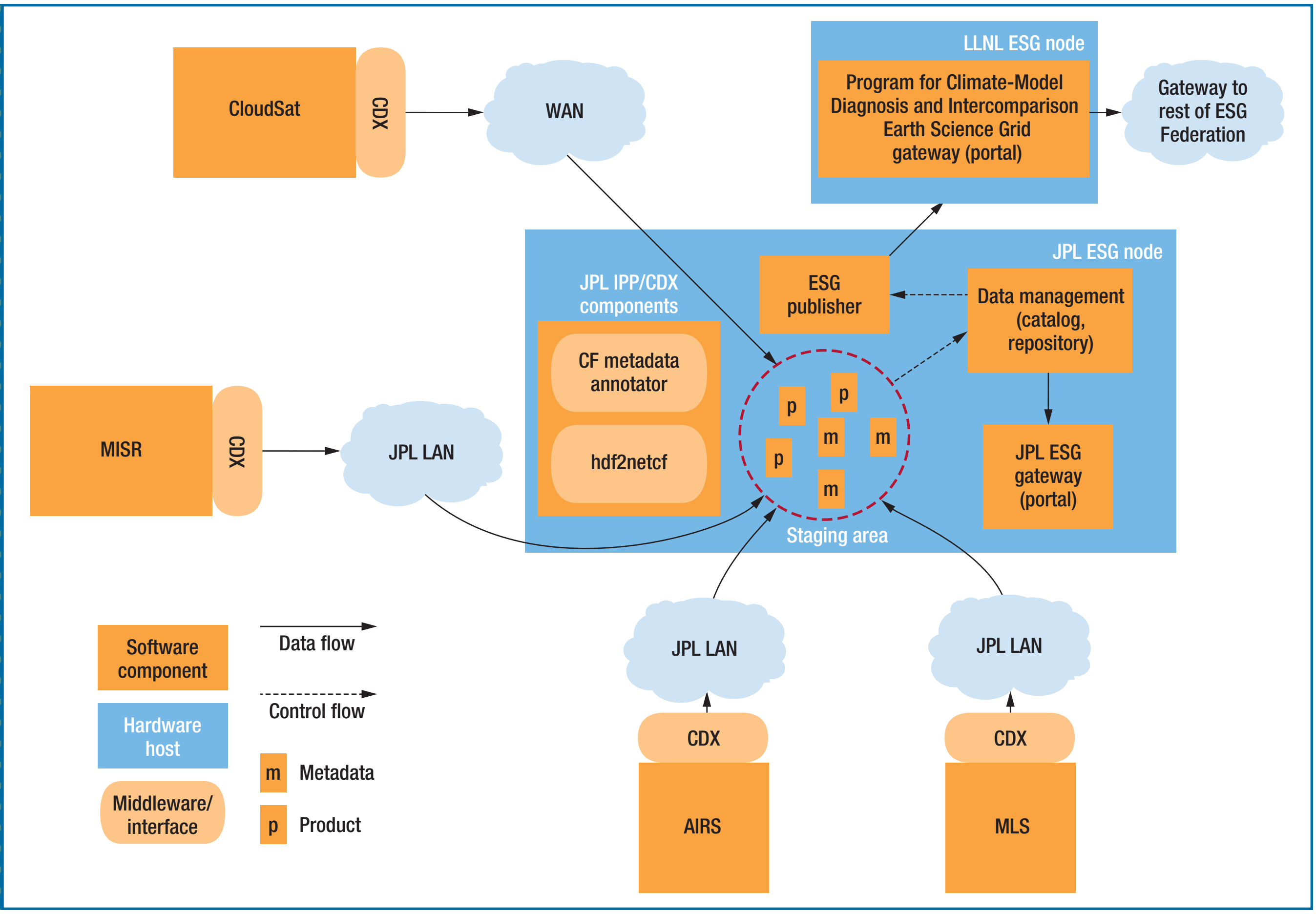


FIGURE 3. Integrating NASA observations with the Earth System Grid (ESG) platform. Distributed remote-sensing data made available by CDX is collected at the Jet Propulsion Laboratory (JPL) ESG node, where data products (labeled "p") and metadata (labeled "m") are aggregated into a staging area. The collected products and metadata are then reformatted in the climate and forecasting metadata standard as NetCDF formatted files and published to the Lawrence Livermore National Laboratory (LLNL) ESG node.

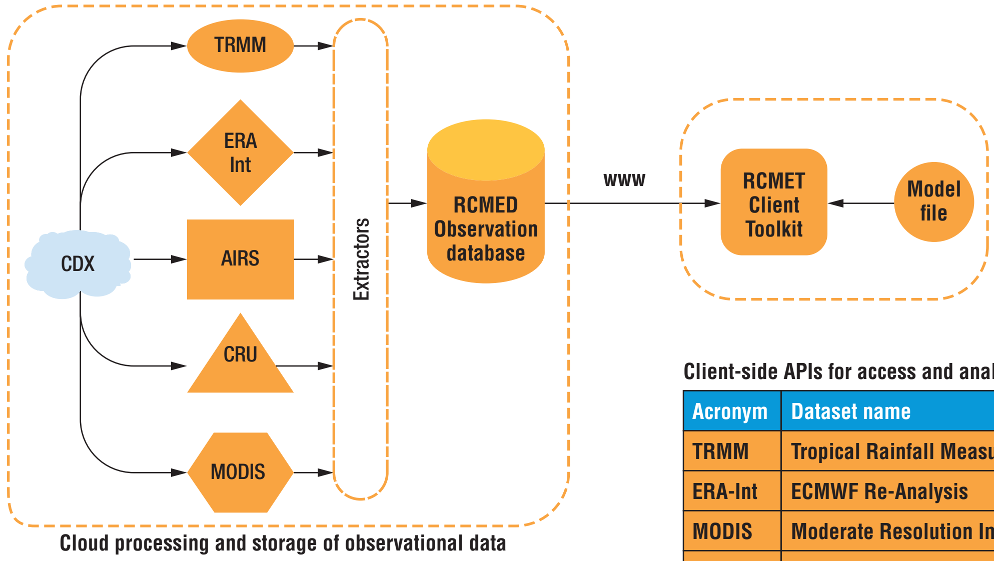
publishing it to the NASA gateway (see Figure 3). This end-to-end effort required addressing all the software engineering challenges we've described: data extraction, mapping, rectification, metadata management, and—once the data is prepared—publishing the observational data to ESG for direct comparisons with a model.

A key tenet of our approach was to leave NASA's computing infrastructure in place. As Figure 3 shows, each participating SIPS deployed a set of CDX services for retrieving, preparing, and publishing observational data to a staging area on the JPL ESG server. From there, the data was cataloged and deployed to the Apache OODT-based data management infrastructure that published its results to ESG.

Enabling Regional Climate Modeling with Observations

The development and validation of regional, near-term climate models presents another timely and important use case for NASA observational data. The IPCC AR5 experiments seek improvements in simulating the behavior of dozens of climate-related variables on time scales measured in decades and at regional resolutions. Model accuracy depends on researchers' ability to use existing observational datasets efficiently as calibration inputs for determining model bias. The vast body of NASA observational data, together with relevant datasets from other organizations, forms a corpus of available data for evaluating climate models.

This evaluation task is complicated by the different purposes and organizations for which the datasets are collected, their annotation and storage in various formats, and their physical residence in geographically and organizationally heterogeneous reposi-



Client-side APIs for access and analysis

Acronym	Dataset name
TRMM	Tropical Rainfall Measuring Mission
ERA-Int	ECMWF Re-Analysis
MODIS	Moderate Resolution Imaging Spectroradiometer
CRU	Clinate Research Unit
AIRS	Atmospheric Infrared Sounder

FIGURE 4. Regional Climate Model Evaluation System (RCMES) architecture. RCMES separates the Regional Climate Model Evaluation Database (RCMED) from the Regional Climate Model Analysis Tool (RCMET). Remote-sensing data enters the RCMED via CDX as well as extractor programs, which deconstruct the underlying file formats and metadata structures and provide the information in a uniform data model to RCMED. RCMED exports a Web service (labeled "www") to RCMET, which uses the collected remote-sensing data and the model outputs to perform a series of regridding operations (aligning the model output and remote-sensing data temporally and spatially), and then a set of metric calculations, producing plots including mean error, bias, and time-series comparisons.

tories. Identifying, interpreting, and transforming these datasets for comparison against model output represents a significant challenge for climate researchers. Although all NASA observational data products conform to the HDF format and the HDF/EOS metadata standard, they must also be accessible through additional formats and standards to increase their value to other communities.

RCMES is a comprehensive effort to address these challenges directly by developing both a scalable data store for observational data, the Regional Climate Model Evaluation Database (RCMED), and a client toolkit, the Regional Climate Model Evaluation Toolkit (RCMET). Figure 4 shows the RCMES architecture for performing rapid model-to-data comparisons and analysis. These products work in concert to provide an environment that gives users ready access to large volumes of observational data as well as a standard, extensible software toolset for accessing and transforming the data for research.

Here again, the volume of available relevant observational data makes it unwieldy for researchers to download and process datasets en masse on a local machine. This is particularly true for regional climate scientists who are often interested only in a highly constrained geographical area. RCMET provides RESTful[14] query interfaces to the RCMED data (middle of Figure 4), facilitating programmatic integration and allowing highly selective querying to minimize data transfer. Furthermore, all RCMET observational data is unified against a common model whose core concept is *dataPoint*—a minimum atomic unit of observation in space and time for a specific parameter. In effect, this eliminates the need for researchers to learn multiple data formats (HDF, NetCDF, Grib, and other scientific formats). Rather, client-side calls to the preprocessed data on the server return homogeneous results from multiple observational datasets for the region of interest.

RCMES relieves scientists of the need to also become data management experts. By providing an infrastructure to support acquisition and transformation of the body of available observational climate datasets, RCMED dramatically simplifies the task of obtaining sufficient data for model comparison. Furthermore,

ABOUT THE AUTHORS

DANIEL J. CRICHTON is a principal computer scientist, program manager, and principal investigator at the NASA Jet Propulsion Laboratory. His research interests include data-intensive systems for planetary, Earth, and biomedical sciences. Crichton has an MS in computer science from the University of Southern California. Contact him at crichton@jpl.nasa.gov.

CHRIS A. MATTMANN is a senior computer scientist at the NASA Jet Propulsion Laboratory, working on instrument and science data systems on Earth science missions and informatics tasks. His research interests are primarily software architecture and large-scale data-intensive systems. Mattmann received his PhD in computer science from the University of Southern California. He's a senior member of IEEE. Contact him at mattmann@jpl.nasa.gov.

LUCA CINQUINI is a senior software engineer at the NASA Jet Propulsion Laboratory, where he's involved in architecting and developing software components for management and analysis of Earth system data. His major activity is working at the Earth System Grid Foundation. He also collaborates with the National Oceanic and Atmospheric Administration to promote shared development and common infrastructure between the two agencies. Cinquini has a PhD in high-energy physics from the University of Colorado. Contact him at luca.cinquini@jpl.nasa.gov.

AMY BRAVERMAN is a senior statistician at the NASA Jet Propulsion Laboratory. Her research interests include information-theoretic approaches for the analysis of massive datasets; data fusion methods for combining heterogeneous, spatial, and spatio-temporal data; and statistical methods for the evaluation and diagnosis of climate models, particularly by comparison to observational data. Braverman has a PhD in statistics from the University of California, Los Angeles. Contact her at ajb@jpl.nasa.gov.

DUANE WALISER is chief scientist of the Earth Science and Technology Directorate and a senior research Scientist at the NASA Jet Propulsion Laboratory as well as a visiting associate in the Geological and Planetary Sciences Division at the California Institute of Technology and an adjunct professor in the Atmospheric and Oceanic Sciences Department at the University of California, Los Angeles. His research interests are in climate dynamics and in global atmosphere-ocean modeling, prediction, and predictability. Waliser has a PhD in physical oceanography from the Scripps Institution of Oceanography at the University of California, San Diego. Contact him at duane.waliser@jpl.nasa.gov.

MICHAEL GUNSON is the project scientist for NASA's Orbiting Carbon Observatory (OCO-2) designed to measure the global distribution of atmospheric carbon dioxide. His research interests include satellite instrument projects for measuring the chemical composition and state of the Earth's atmosphere, remote sensing, and system design for improving the scientific exploitation of increasingly large and complex datasets returned by satellite instruments. Contact him at mgunson@jpl.nasa.gov.

ANDREW F. HART is a staff software engineer at the NASA Jet Propulsion Laboratory. His research interests lie in user interfaces for Web applications, frameworks for rapid Web application development, and data modeling. Hart has an MS in computer science from the University of Southern California. Contact him at andrew.f.hart@jpl.nasa.gov.

CAMERON E. GOODALE is a software engineer at the NASA Jet Propulsion Laboratory. His research interests are in user experience, using Web standards for geographic information systems (GISs), and open source data management. Cameron has an MBA with an emphasis in GISs from the University of Redlands. Contact him at cameron.e.goodale@jpl.nasa.gov.

PETER LEAN is a California Institute of Technology postdoctoral scholar at the NASA Jet Propulsion Laboratory. His research interests are in meteorology and the evaluation of weather and climate models. Lean has a PhD in atmospheric predictability from the University of Reading, UK. Contact him at p.w.lean@reading.ac.uk.

JINWON KIM is a visiting scholar in the Joint Institute for Regional Earth System Science and Engineering at the University of California, Los Angeles, and a professor at Yonsei University, South Korea. His research interests center on climate modeling, including work in multimodel ensembles and regional climate comparisons. Contact him at jkim@atmos.ucla.edu.

RCMET's routines and service wrappers help reduce ad hoc programming and move model evaluation to a more automated, repeatable process that can take hours to days instead of weeks to months because RCMED and its tight interactions with RCMET obviate the need for data collection, aggregation, and preparation tasks that normally add significant time to conducting comparisons.

Our work has brought the climate-modeling and remote-sensing communities closer together by developing multicenter collaborative teams to integrate the diverse data representations, standards, software tools, and infrastructure on both sides. Our specific focus on using existing tools and infrastructure to integrate NASA-based Earth science data systems with those of the climate-modeling community via ESG has yielded several important lessons, including

- the importance of defining standards and understanding how they apply to both climate models and observations;
- the difficulties in understanding the substantial differences between observations and model output in format, meaning, and representation; and
- how to establish cross-disciplinary teams that include scientists from the climate and satellite communities as well as computer scientists with experience building systems for these communities. The science, systems, and standards are equally important and must be well defined.

By demonstrating feasibility and identifying likely challenges at larger scales, our recent efforts represent a first step toward formulating a blueprint for larger national and international efforts between the satellite and climate communities. Our pilot activities, both in publishing observations to the ESG platform and with RCMES, have shown the value of providing services that compare satellite data with climate models. We plan to continue working with the climate-modeling and satellite communities by defining data, software, and architectural standards that promote access, transformation, and use of satellite data in support of climate research.

Acknowledgments

We acknowledge the contributions of several individuals, including Dean Williams, Gerald Potter, Joao Teixeira, Robert Ferraro, Seungwon Lee, Alex Guillaume, Jinwon Kim, Jason Hyon, Ken Wolfenbarger, and Diane Evans. This work was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

References

1. Intergovernmental Panel on Climate Change, "Scoping the IPCC 5th Assessment Report, Proposal for an IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections," IPCCXXX/Doc. 11, IPCC, Apr. 2009; www.ipcc.ch/scoping_meeting_ar5/doc11.pdf.
2. F. Giorgi, C. Jones, and G. Asrar, "Addressing Climate Information Needs at the Regional Level: The CORDEX Framework," *WMO Bulletin*, vol. 58, no. 3, 2009, pp. 175–183.
3. B Fortner, "HDF: The Hierarchical Data Format," *Dr Dobb's J. Software Tools and Professional Programming*, 1998; www.drdobbs.com/database/184410553.
4. R.K. Rew and G.P. Davis, "NetCDF: An Interface for Scientific Data Access," *IEEE Computer Graphics and Applications*, vol. 10, no. 4, 1990, pp. 76–82.
5. R.K. Rew et al., "The CF Conventions: Governance and Community Issues in Establishing Standards for Representing Climate, Forecast, and Observational Data," slide presentation, Am. Geophysical Union Fall Meeting, abstract #IN52A-07, 2007.
6. C. Mattmann, A. Braverman, and D. Crichton, "Understanding Architectural Tradeoffs Necessary to Increase Climate Model Intercomparison Efficiency," *ACM SIGSOFT Software Eng. Notes*, vol. 35, no. 3, 2010, pp. 1–6.
7. C. Mattmann et al., "A Software Architecture-Based Framework for Highly Distributed and Data Intensive Scientific Applications," *Proc. Int'l Conf. Software Eng.* (ICSE 06), IEEE CS, 2006, pp. 721–730.
8. D. Crichton et al., "Facilitating Climate Modeling Research by Integrating NASA and the Earth System Grid," 2011; http://bit.ly/f7akgL, 2011.
9. R. Fielding, and R.N. Taylor, "Principled Design of the Modern Web Architecture," *ACM Trans. Internet Technology*, vol. 2, no 2, 2002, pp. 115–150.