Data Quality Screening Service Work Plan

Project Description

Introduction

NASA's Remote Sensing data have a rich set of quality information. These range from simple quality flags to complicated bitfields to external criteria such as cloud cover. Additional quality information is often included in external documentation or peer-reviewed literature and may include latitude dependencies, physiographic criteria (e.g., not valid over deserts), temporal trends, etc. The explanations of these quality factors also appear in a variety of forms, such as data product documentation and journal articles. Thus, it is a laborious process for data users to locate the quality explanations, read the quality indicators and code an interpretation of the indicators. Worse, each individual user must repeat the process, resulting in a huge duplication of effort. If it is merely difficult for science users, it is virtually impossible for machine applications to incorporate and apply the quality information; they must either reject the data product or ignore the quality information.

We propose to solve this problem through the construction of a data quality screening service that will allow the user to apply quality screening on the fly. The service will be deployable as a simple REST (Representational State Transfer) Web Service at the data provider, taking as input the original data product and producing as output the same product, but with quality screening applied (Fig. 1) according to the cognizant science team's recommendations as stated in the documentation.

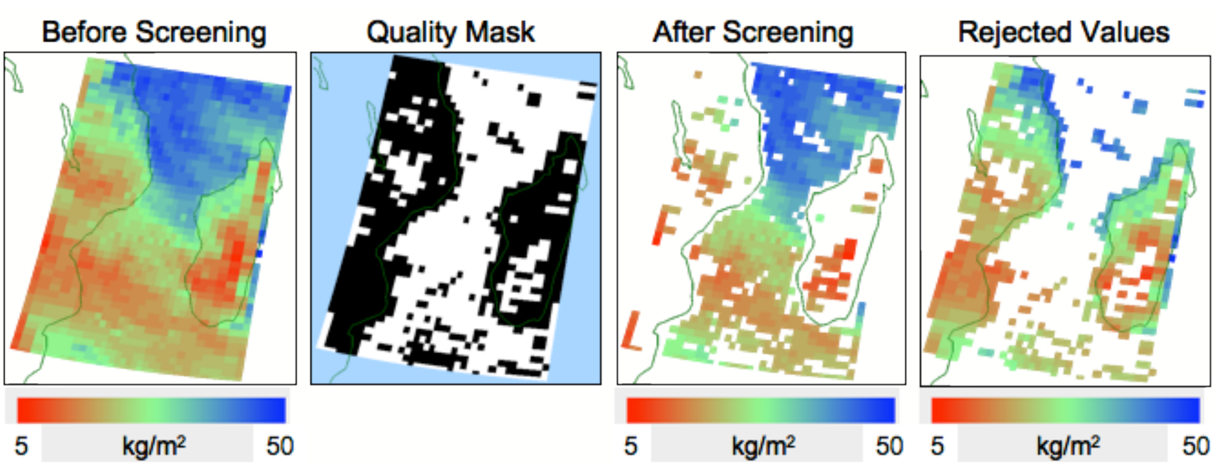


Fig 1. Depiction of an unscreened data field, along with quality mask (black showing rejected points), screened data field, and data field with rejected points only. The data shown are AIRS Total Column Precipitable Water near Madagascar on 4 Jun 2009.

The quality screening recommendations will be encoded in a community-based data quality ontology, together with the representation, location, and meaning of the quality information.

This ontology-based approach will enable the service to be reused, expanded and maintained at low cost.

This will be accomplished by leveraging existing technologies: the Data Quality Screening Service will be hosted in existing Web Service infrastructures at the GES DISC and MODAPS, and be offered through the operational Mirador search tool, whereas the quality screening ontology and modules will be extensions of a GES DISC prototype.

Technical Approach

We will implement a web based quality screening service to make Earth science data quality information easy to use. for users of all stripes. The service will be implemented at two data centers, the Goddard Earth Sciences Data and Information Services Center (GES DISC) and the MODIS Adaptive Processing System (MODAPS), demonstrating the generality of the solution for heterogeneous data sets and distributed data centers.

The service will have the following features:

Server-side Screening: provides the data files on-demand, already pre-screened.

User-Side Screening Support Masks: alternatively, provides a basic quality mask in a standard format to be applied on the client side using simple, reusable modules.

Science Team Recommendations: the above services will include preset criteria based on the science team's recommendations

User-tunable Criteria: alternatively, allows the user to customize their quality screening.

Quality Ontology: brings order and inter-relations to the diversity of quality information, enabling the construction of a small(er) number of general-purpose quality-screening routines and providing a way for machine-based services to interpret the quality information.

Simple HTTP (Hypertext Transfer Protocol) REST interface: makes the services callable by machine-level code or easy to incorporate in search interfaces, such as Mirador.

Metrics Instrumentation: shows how quality screening is being used by the community.

Quality Screening Impact Views: provides the users pictorial and statistical views of how a given quality set of screening criteria will affect the data.

2.4.1 Server-Side Quality Screening

The main feature (i.e., with the most users) of the service is Server-Side Quality Screening: this applies quality screening criteria to the input data file, stored in Hierarchical Data Format (HDF). The output file transmitted to the user holds fill values in the data arrays where the data have failed to pass the quality screening criteria. (These criteria may be either preset, based on science team recommendations, or user-tunable.) The rejected data values are stored in a companion array (Fig. 1) in the output file. By applying internal HDF compression, the size of the output file is roughly the same as the size of the input file, as fill values do not add to the size. This process occurs on the fly as the user invokes the service Uniform Resource Locator (URL).

Server-Side Quality Screening is implemented by two main processes, the Masker and the Screener (Fig. 2). In the first step, the Masker takes as input the data file to be screened and user selection of quality criteria. It queries the ontology for information on how to locate and interpret the quality information for the data product in question and level of quality screening. Based on this information, it retrieves the appropriate Java module from a library repository and invokes it with arguments also derived from the information in the ontology (described in a later subsection). The Java module will in turn acquire any needed information from the data file or ancillary files, outputting a simple bitmask for each data field to be screened, matching the dimensions of the field exactly, with 1 indicating "keep" and 0 indicating "reject". This Screener then applies this mask to the fields in the data file to obtain a screened data file.

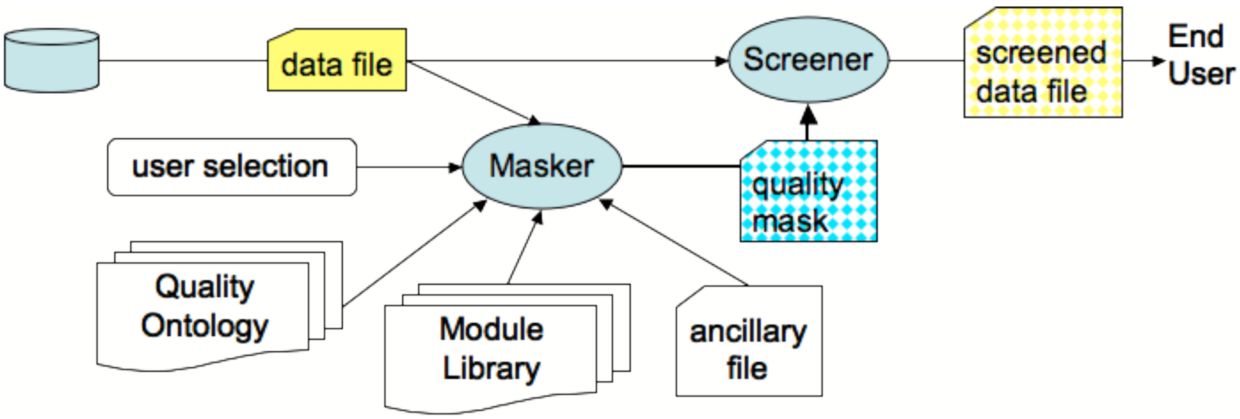


Fig 2 The Server-Side Quality Screening occurs in two steps: the Masker forms a bitmask using user inputs, the quality ontology, data file and ancillary files (if necessary). The mask is then applied by a generalized Screener module. This simple, modular architecture maximizes extensibility and reusability.

2.4.2 User-Side Quality Support Masks

Expert users may wish to experiment with different quality criteria settings. Rather than repeatedly transfer the data via the Server-Side Quality Screening, this service can return *just a bitmask* file based on the user's screening criteria and matching the dimension of the data array being screened. The user can iterate on this, retrieving or visualizing just the bitmask until he/she finds the optimal set of criteria for his/her use. This will be accompanied by simple utility modules to apply the mask to the data, available in a variety of programming and tool languages, including C, Java, Python, IDL, Matlab, and Integrated Data Viewer.

User-Side screening has another valuable use: it enables us to offer quality screening for data distributed by other data providers. This is the case for datasets from CALIPSO and Cloudsat. The GES DISC has local copies of these data as part of its A-Train Data Depot [1], which allows us to compute the quality masks. However, the official distributor for Cloudsat and CALIPSO data are the Cloudsat science team and the Atmospheric Sciences Data Center, respectively.

2.4.3 Science Team Recommendations

The default quality screening ruleset will be based on the recommendations of the science team for that particular product. These recommendations essentially amount to "best practices" for applying quality criteria to screen out unusable data. While these are occasionally simply "use/don't-use" recommendations, they may also be gradational ("bad/good/best") and may even depend on the projected use of the data. An example of this is the difference between AIRS Total Precipitable Water pixels tagged with a quality level of 0 which are deemed suitable for individual assimilation, and the less conservative quality level of 1 which gives a higher yield, but is recommended for statistical climate studies [1].

2.4.4 User Tunable Criteria

Aside from screening based on the science team recommendations, users will be able to apply their own criteria. For example, where the science team may recommend rejecting retrievals when the cloud cover was greater than 30%, some users may consciously choose to override this default, selecting a higher or lower rejection threshold. This capability will be paired with the User-Side Quality Support Masks to avoid excessive repeated data transfers during the user's iterative exploration of the effects of different quality criteria. The masks will include metadata on what criteria were applied, along with optional user annotations. These annotations can provide useful feedback to the science team as well as suggest improvements to the ontology.

Users will be able to save user-tunable criteria to their workstations as an XML file, which can later be uploaded through Mirador to apply to other data files from the same dataset in bulk. This XML file can also be shared with other researchers, so that those researchers can reapply them to a dataset thus enabling true reproducibility of results.

However, as the technical evaluations questioned the usefulness of this, it will be deferred to the second year, so that we can gauge its utility based on user outreach.

2.4.5 Quality Ontology

Quality information, representation and location vary considerably from data field to data field and dataset to dataset. Nonetheless, there are many commonalities: quality flags, quality levels, bitfields inside integers, statistical measures and cloud cover thresholds can all be found in multiple datasets. There are also common threads in the reasons for rejecting data, such as calibration errors, cloud contamination and glint contamination.

By organizing the different attributes of data quality into an ontology, we can implement a system that can be easily extended to cover the diversity of data fields and products. An ontology is a shared "formal representation of a set of concepts within a domain and the relationships between those concepts" [2]. Ontologies and the technologies that support them are designed for extensibility to accommodate novel classes and properties.

Ontologies also lend themselves to reuse. We will evaluate existing ontologies in this realm and extend such ontologies around aspects of data quality, including concepts related to:

- *Location*: where the quality information is located
- *Structural Representation*: how quality information is stored (e.g., bitfields, flags, levels, thresholds)
- *Meaning*: what the quality information indicates (i.e., the kind of error, or usability for different purposes)

The Location and Structural Representation parts of the ontology will enable us to construct table-driven routines for quality screening. Thus, a single routine can be used to interpret any integer bitmask, given sufficient information in the ontology about how the bitmask is constructed (i.e., which bits are where).

The Meaning part of the ontology is essential for expressing to the user in a succinct form the various choices for data screening and thus goes to the very heart of the ACCESS call. Not only does it supply a framework within which to interpret the much more detailed quality information, but it is particularly helpful in comparing quality across datasets that have different quality representations. It is absolutely indispensable for machine-level applications attempting to assess and apply quality criteria automatically. Also, the quality meaning is critical to recording the overall provenance when the quality screening service is used in a workflow and ultimately explained to the end-user.

2.4.6 Simple HTTP REST Interface

In order to make the service easy to use and easy to integrate with other programs, we will make it available through a simple REST service, similar to the ones currently offered by the GES DISC. In other words, the service can be invoked through a simple URL, e.g.,

> http://airspar1u.ecs.nasa.gov/daac-bin/OTF/HTTP_services.cgi?
> FILENAME=AIRS.2009.05.07.150.L2.RetStd.v5.2.2.0.G09128131425.hdf&
> SHORTNAME=AIRX2RET&
> SERVICE=QualityScreen&CRITERIA=SciTeamRecBest

This simple interface makes it easy to integrate into other applications, such as search tools. These URLs can even be embedded in web pages and documents. Thus a researcher could include in an article or report not only the data files used, but also the quality screening that was applied to the data, thus enhancing the reproducibility of the results.

A Simple Object Access Protocol (SOAP) interface may also be added in the later stages in consultation with the Program Manager, based on demand by machine-level clients.

2.4.7 Metrics

Offering this screening capability as a web service encoded in the URL provides a unique opportunity for metrics collection. For the first time, we can obtain metrics on what quality screening the community is using, simply by examining the URLs in the web access log. These metrics can be easily categorized and organized using the ontology. Furthermore, with each invocation of the service, we will log the number of rejected points and the reason for that rejection, giving us an assessment of the impact of quality screening on the data chosen by the users. This usage and impact information will be invaluable feedback to science teams as they prepare the next version of the processing algorithms. These metrics will be supplemented by feedback on the impact of the service on science results, which will be obtained from users through outreach vehicles such as the AIRS mailing list.

The REST service invocations will be recorded in the Web logs at MODAPS and the GES DISC for transmission to the EOSDIS Metrics System (EMS). Labels will be agreed upon with EMS to enable the service to be reported on, as well as fractionated into the data products screened. Logs will also be saved locally for more detailed metrics analysis.

The initial User Metrics to be collected are:

- Number of DQSS invocations
    - Break down by product
    - Break down by quality level (e.g., good, best, etc.)
- Yield (ratio of used data points to total points)

These will likely suggest further metrics and breakdowns as usage progresses.

After the implementation stabilizes, cost metrics will also be captured to determine how much it costs to extend the framework to include new datasets.
Code metrics will also be captured to estimate long-term lifecycle costs.

2.4.8 Quality Screening Guidance: Case Studies

One of the most important aspects of guidance needed is to impress on the users the importance of quality screening. To this end, we will document several case studies where applying the proper quality screening procedures can dramatically change the interpretation of the data. These will be prominently featured in the Mirador search interface [3] and accompanied by concise tutorials on how to use the screening service. The combination of the case study, tutorials and easy-to-use service will be less intimidating to the non-expert user than the current data disclaimers, thus helping to expand the potential user community.

2.4.9 Datasets

The datasets to be addressed in this study will be primarily Level 2 (swath) data addressing cloud properties and water vapor. Both User-Side and Client-side screening will be offered for data that are served by either MODAPS or the GES DISC: MODIS L2 Atmospheric Profile, Precipitable Water and Cloud products; AIRS L2 Standard Retrieval and Support products; and MLS Water Vapor, Ice Water Content and Relative Humidity with respect to Ice products. In addition, User-Side screening will be available for data products that are supported for value-added services at the GES DISC: CALIPSO Vertical Feature Mask; and Cloudsat Cloud Classification and Radar-Only Liquid/Ice Water Content.

This does not cover all the desired datasets: for instance, both swath data for both TES and Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) would be useful as well [4], but they are served by other data centers. The extensible nature of the Data Quality Screening Service architecture would make it relatively easy and inexpensive to accommodate these datasets in possible follow-on work, however. Additional datasets may be added to demonstrate scalability and flexibility, particularly if the User-Tunable Criteria is deemed to have limited applicability. This decision will be made in the second year.

Implementation: Reuse and Enhancement

The implementation of the quality screening service leverages existing components at both the GES DISC and MODAPS (Fig. 3). The software to be developed comprises: the Screener, in both server and client-plug-in forms; and the Masker, including masking modules for different quality representations and locations. The rest of the service reuses or extends pre-existing components. The key components are described below, along with their current Software Technology Readiness Level [5]. The target TRL for the overall Data Quality Screening Service at the end of the project is 9, fully operational.

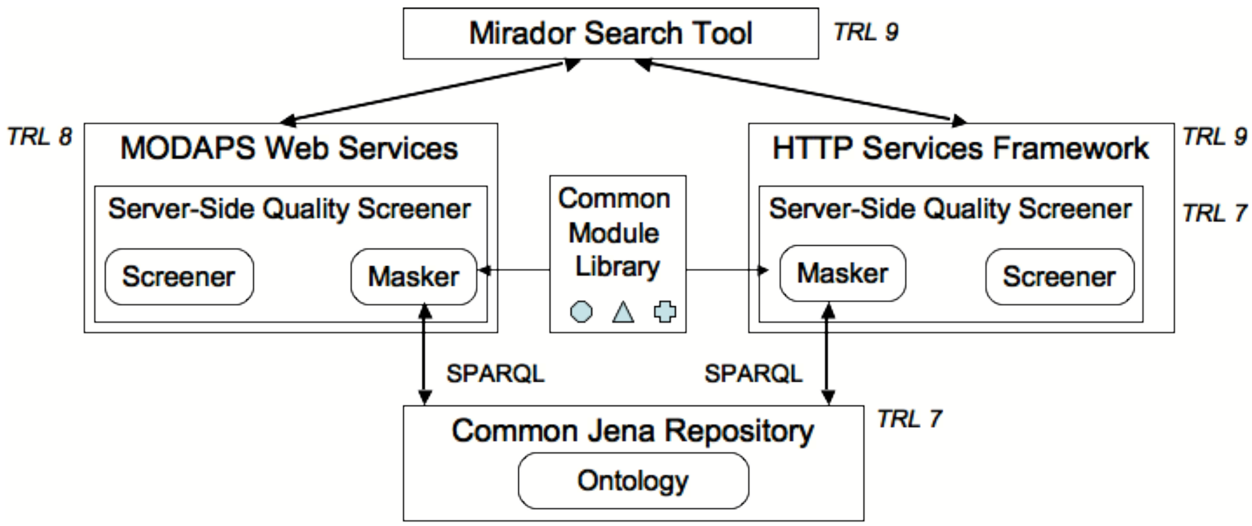


Fig 3. Technical implementation of the Server-Side Data Quality Screening Service: the outer infrastructure is operational or nearly so, while the core screening and semantic web components are currently demonstration prototypes in the GES DISC test baseline, which simulates the operational environment. The reuse of existing frameworks and working prototypes ensures operational deployment by project completion.

2.6.1 Mirador Search Tool

The Mirador Search tool will be an easy-to-use portal for users to request the quality screening service. Mirador has search capabilities for all of the GES DISC's data, while MODIS data will be searched via an OpenSearch request to the MODAPS Web Services (currently prototyped and scheduled to be complete by the end of 2009). Mirador also provides access to on-the-fly services such as NetCDF conversion and subsetting, with new services added simply through configuration files. Mirador is fully operational, with a Technology Readiness Level (TRL) of 9. Using Mirador to access both data centers will allow the users to do integrated searches and quality impact comparisons, a valuable tool for multi-sensor science problems.

2.6.2 Web Services Frameworks

Both the GES DISC and MODAPS have already deployed Web Services frameworks, albeit in very different implementations. The GES DISC framework is based on Common Gateway Inteface and offers both format conversion and subsetting services operationally (i.e., TRL=9). New services, such as the quality screening, are added through the HTTP Services configuration file, without the need for any software changes to the framework. The MODAPS framework is based on Java servlets, and offers a variety of subsetting, mosaicking, reprojection and reformatting services, as well as a search service that will be utilized by the Mirador search tool. Both the GES DISC and the MODAPS web services frameworks have been audited for security vulnerabilities.

2.6.3 Quality Screening Software

A prototype of the Quality Screening software has been developed for select fields in the AIRS Level 2 Standard Retrievals product. This implements quality screening for two different kinds of quality representations, with the proper selection guided by a query to an ontology repository.

The prototype has been integrated with Mirador and the HTTP Services framework within the GES DISC test baseline (which mirrors the configuration of the public baseline) making it a “simulated operational environment” [13]; hence this component is at TRL 7. To progress to TRL 9, the prototype will be generalized and expanded to account for other quality representations and meanings. After extensive testing, it will be promoted to the publicly accessible baseline for general use by the user community. The service will then be expanded to additional datasets.

2.6.4 Semantic Web Technologies

The ontology will be represented in the Web Ontology Language (OWL), and stored as Resource Description Framework (RDF) triples in a repository. The repository will use Jena[6], a framework that supports RDF and OWL and queries using the SPARQL Protocol and RDF Query Language (SPARQL) language[7]. These technologies are widely used operationally and are being used at the GES DISC for Mirador, together with an ontology for data products and their parameters.

2.6.5 Quality Ontology

The Quality Ontology will include descriptions of data quality representations, file structures and science team usage recommendations. A “micro-ontology” has been prototyped to support the Quality Screening software prototype above, making this component also TRL 7. In order to progress to TRL 9, this ontology will be expanded to account for other types of quality representation, location and meaning, by adding RDF triples to the Jena repository. Indeed, one of the advantages of an ontology-based approach over a traditional relational database approach is that the data model can be easily extended without modifying an existing database schema. Many of these elements will actually be drawn from ontologies that already exist or are in progress and will be complete by the project start date (Table 1). Quality representation elements will be largely derived from data product documentation.

Ontology	**Domain reuse**
Semantic Web for Earth and Environmental Terminology (SWEET)[]	High-level description of phenomena and realms
Multi-sensor Data Synergy Advisor	Data lineage
Earth Science Information Partners (ESIP) Data-type and Services	Data product and service type descriptions
CF-1 (Climate and Forecast)	Internal file and data structures
Mirador Datasets and Parameters	Datasets and parameters at the GES DISC

Table 1. Ontology reuse from a variety of domains.

In addition, elements will be drawn where appropriate from ontologies in the broader community, such as the Statistical Core Vocabulary [8].

2.6.6 Visualization for Quality Impact Views

The quality impact views provide a way for users to preview the effects of a given set of quality screening criteria. They will consist of basic statistics, such as points kept vs. rejected and the mean value of rejected and kept points. In addition, images similar to Fig. 1 will provide a graphic view of the effect of the screening. These images will be generated by repurposing existing browse generation software for AIRS and MODIS as well as visualization software in the A-Train Data Depot for MLS, CALIPSO and Cloudsat data.

2.6.7 Common Module Library

The Common Module Library is simply a web site from which the specific Java masking modules will be served to the GES DISC or MODAPS to construct quality masks. (While they can be obtained at runtime, the latest versions will likely be cached locally at each site.)

Long-term Continuance

Operational Deployment

The Data Quality Screening Service is designed from the start to be integrated into the operational systems at MODAPS and the GES DISC. This integration will occur early and often, as enhancements are made to the service, following the standard promotion procedures at the respective data centers. Since the DQSS code will be specifically implemented for inexpensive integration into the existing systems, cost of operational deployment should be negligible.

Maintenance

Rapidly moving prototypes can produce code that is relatively expensive to maintain. To mitigate this, a refactoring phase is planned at the end of the development cycle. This will allow us to attain a code-base that is both robust (and less susceptible to outyear bugs) and easier / cheaper to modify.

The chief outyear maintenance activity will stem from modifications to the quality screening schemes for the data, which typically happens on reprocessing boundaries. This will entail modifying the ontology, and potentially the DQSS code for some ontology changes. Note that much of the effort here is simply understanding the quality screening revisions, an activity necessitated as part of standard product data support in any case. Also, the extensibility of the ontology-based approach makes it relatively inexpensive to add even novel aspects of quality representation or meaning, which limits the total life-cycle costs.

Plan of Work

Tasks

The tasks needed to accomplish the end goals fall into three categories: Infrastructure; Data-Specific Ontology and Masker Modules; and End-to-End Integration.

Infrastructure is the development of components needed for all of the datasets being supported. This begins by specifying Use Cases, followed by a detailed design that describes how all of the pieces fit together and specifies the interfaces among them. Also included is implementation of modules that are not dataset-specific. The current quality screener prototype will be refactored and generalized to include a generic screener, a wrapper for masker modules, an ontology-driven interface for tunable screening and language-specific modules for User-Side application of quality masks to data. Existing visualization software will be repurposed for quality impact views.

Data-Specific Extensions consists of extending the prototype ontology of the representation, location and meaning of quality information to accommodate successive data products. This is followed in each case by construction of specific masker modules to account for each new way of screening data encountered as the ontology is extended. Data-specific visualization and extensions to the user interface are also included in this task.

End-to-end Integration is the process of inserting all developed components into the existing infrastructure at the GES DISC and/or MODAPS site. This will be done first in a test mode, hidden from the user community, for the team to run end-to-end testing, from search tool to service invocation. The scientists on the team will verify the quality screening results. The milestone will be completed by promoting into the operational mode, resulting in a capability ready for use by the science community.

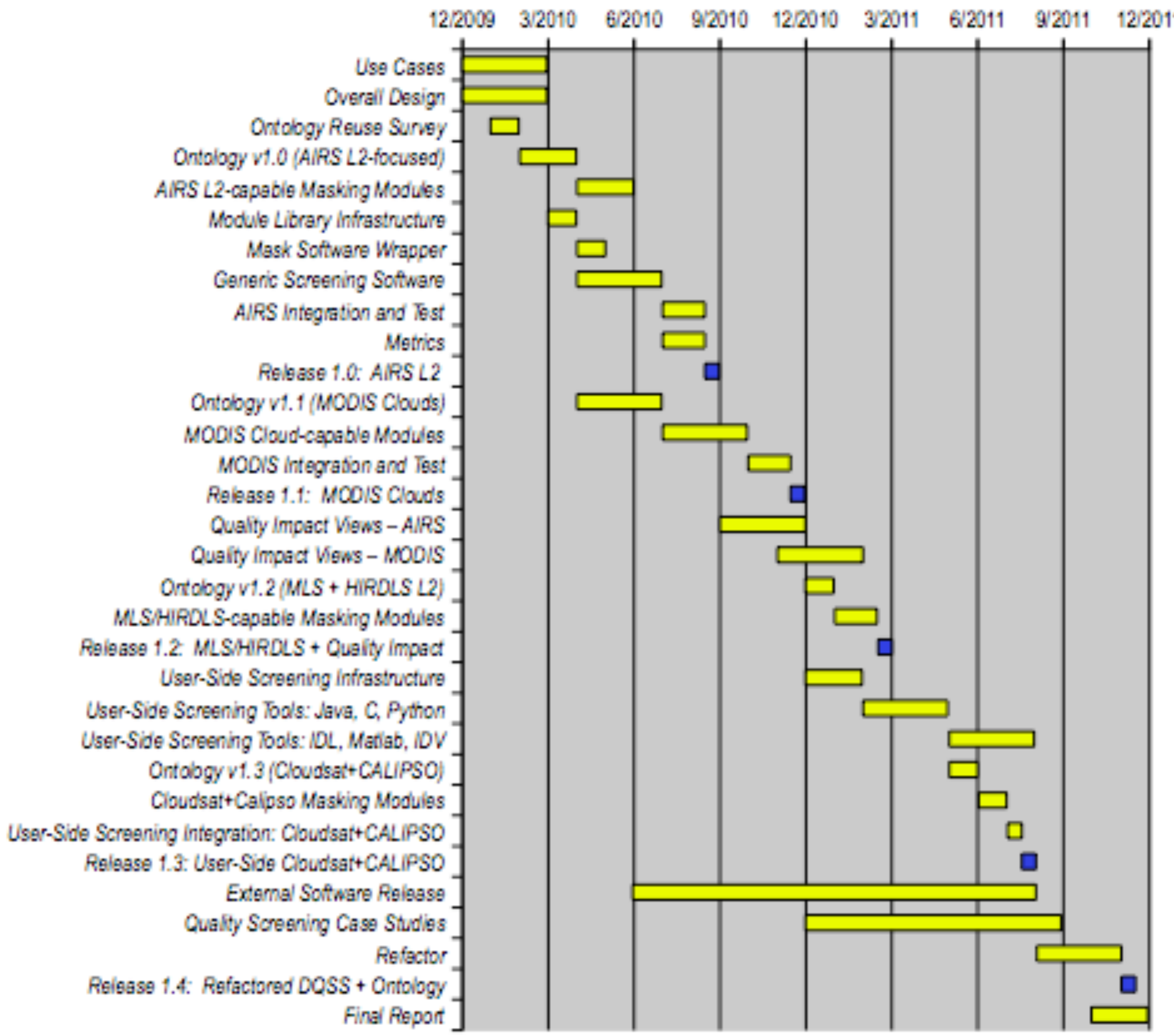
Key Milestones

Table 2 below shows the completion of milestones for the project by quarter.

	Infrastructure	**Data-Specific Extension**	**End-to-End Capability**
		Year 1	
Q1	Use Cases and Overall Design (0.11)	Ontology reuse survey (0.01)	
Q2	Server-side screening: mask module wrapper and generic Screener (0.08)	AIRS Level 2 retrievals (0.05)	
Q3	Metrics Collection (0.05)	MODIS cloud products (0.03)	Server-side screening for AIRS Level 2 (0.03)
Q4			Server-side screening for MODIS cloud products(0.03)
		Year 2	
Q1	Quality impact views and statistics	MLS + HIRDLS (0.01)	Quality Impact Views + Server-side screening for MLS and HIRDLS (
Q2	User-Side screening plugins - Java, C, Python		Interactive screening for all products
Q3	User-Side screening plugins - IDL, Matlab, IDV	Cloudsat + CALIPSO (0.02)	User-Side screening for CALIPSO and Cloudsat; Software Release
Q4	Refactor for long-term maintenance	Publish ontology (0.01)	Quality Screening Case Studies Final Report

Table 2. Key milestones.

Schedule



Information Technology Plan

In general, the IT tools needed to carry out the project are already in place at one or both of the data centers. These include development and test machines, ontology editors (Cmap Ontology Editor and Protégé) and triple stores (Jena). One additional element that would be useful is a long-term visual collaboration tool for ontology development. This could range from video conferencing technology to screen-sharing to specific collaborative ontology editors. A solution thereto will be explored, though we will fall back to face-to-face meetings if necessary, piggybacking off the AIRS Sounder Science meetings in particular.

Deliverables

Use Cases: Documentation of likely use cases
Overall Design: Event trace diagrams, Class diagram, Interface Specifications, and Narrative.

Ontologies: OWL files, published to ESIP ontology repository
Quality Screening Case Studies: Short web pages with examples
Java Classes: Source code released to the community
Operational Web Services: Web services visible through Mirador, MODAPS and ECHO

Data Sharing Plan

This project will produce two important products to be shared with the community. The first is the new quality ontology elements (those not simply reused from other ontologies). The quality ontology has potential use in other systems, such as science workflow systems. These will be contributed to the ESIP Data-type and Service Ontology [9], a project of the ESIP Semantic Web Cluster, to which Dr. Lynnes and Prof. Fox are already key contributors.

The second product is a set of Java modules to create quality masks and screen HDF files. These will be released as Open Source through the GSFC Software Release office and GSFC Open Source site. The submission process will commence with the initial coding in Q2 of the first year in order to give enough time for the software release process.

ESDSWG Participation

Dr. Lynnes will contribute 0.15 FTE to the Technical Infusion Working Group (TIWG), with an emphasis on Services Interoperability and Orchestration and Semantic Web subgroups. He has been an active participant in all TIWG subgroups over the past three years, serving as co-chair of the overall TIWG in 2008. He has contributed to the Earth Science Information Partners (ESIP) Data-type and Service Ontology.

Prof. Fox, supported by his Systems Programmer, will contribute 0.10 FTE to the TIWG, particularly the Semantic Web subgroup. Prof. Fox is an internationally recognized expert in semantic web technology, as well as the current chair of the TIWG Semantic Web subgroup. Prof. Fox also coordinates the effort to develop the ESIP Data-type and Service Ontology.

5.3 Outreach

Dr. Olsen will lead the outreach to AIRS users, through the AIRS mailing list and presentations at Sounder Science meetings, with assistance from Mr. Vollmer. As editor of the AIRS data product documentation, he will add information about using the service. Mr. Wolfe will use similar MODIS vehicles, such as the MODIS science team meetings. Dr. Lynnes will also present the work at meetings of the Earth Science Information Partners.

Budget and Cost Plan

Budget Summary

Budget Narrative

[1] Savtchenko, A., Kummerer, R., Smith, P., Gopalan, A., Kempler, S., and G. Leptoukh, 2008. A-Train Data Depot: bringing atmospheric measurements together, IEEE Trans. Geosci. Remote Sensing, **46**, p. 2788-2795.
[2] Ontolology entry, Wikipedia, http://en.wikipedia.org/wiki/Ontology.
[3] Lynnes, C., Strub, R., Seiler, E., Joshi, T. and P. MacHarrie, 2009. Mirador: A simple fast search interface for global remote sensing data sets, *IEEE Trans. Geosci. Remote Sensing*, **47**, 92-96.
[4] Fetzer, E., B. Lambrigtsen, A. Eldering, H. Aumann, and M. Chahine, 2006. Biases in total precipitable water vapor climatologies from Atmospheric Infrared Sounder and Advanced Microwave Scanning Radiometer, *J. Geophys. Res.*, **111**, D09S16, doi:10.1029/2005JD006598.
[5] Department of Defense, 2005. *Technology Readiness Assessment Deskbook*, http://www.dod.mil/ddre/doc/May2005_TRA_2005_DoD.pdf.
[6] Jena- A Semantic Web Framework for Java, http://jena.sourceforge.net/
[7] World Wide Web Consortium, 2008, SPARQL Query Language for RDF, W3C Recommendation, http://www.w3.org/TR/rdf-sparql-query/.
[8] Ayers, D., L. Feigenbaum, W. Halb, M. Housenblas, T. Heath, and Y. Raimond, The Statistical Core Vocbulary, http://sw.joanneum.at/scovo/schema.html
[9] http://wiki.esipfed.org/index.php/Data_Service_Ontologies