IRIS Data Services Strategic Action Plan (SAP)
Version 4.0 October 28, 2014

In June 2012 IRIS Data Services conducted a strategic planning session with 17 members of the IRIS community participating. In addition to refreshing the Mission and Vision statements, we identified the primary goals, strategic directions and obstacles and challenges for IRIS Data Services. In November of 2012 IRIS convened an External Review Panel (ERP) to review Data Services that also helped to clarify directions for DS to pursue. These two activities have resulted in this Strategic Action Plan that identifies directions that respond to the Strategic Planning session as well as the ERP.

Considering the various input we received during the Strategic Planning Process and the ERP, we have identified 10 specific areas IRIS Data Services should pursue over the next five years. Each action references the relevant portions of the Strategic Plan and/or the ERP report where appropriate. The Strategic Plan items and the recommendations from the ERP are attached at the end of this plan.

These actions are not presented in priority order.

SAP 1. Build Bridges between Data and HPC Environments. (Cycles close to data)
Issue being addressed: As the volume of data continues to increase it is no longer possible to move significant portions of the IRIS archive to individual computing environments, including HPC environments, within reasonable lengths of time. For this reason IRIS DS should find creative ways to find funding to support the bulk delivery of the entire archive to the proximity of HPC Centers.
Status: With the development of the Auxiliary Data Center (ADC) concept IRIS DS has taken the first step toward placing data close to cycles. With the selection of LLNL as the first ADC this concept can be further studied to see if it meets the needs of the broad scientific community.
Required Actions: Sources of funding to expand this approach need to be pursued as the first step was done using existing resources that previously supported the Active Backup Center in Boulder. Potential funding sources could be MRI proposals or proposals to DoE for support at additional National Labs.

- Place a duplicate of the IRIS archive in the proximity of an HPC facility.
- Encourage the IRIS community to use this system and help identify next steps.
- Develop streamlined, high performance connections between ADC and LLNL HPC environment including the identification of correct data packaging for use in the HPC environment.
- Work with the HPC community to ease access by the geosciences research community.

Status:

- IRIS Auxiliary Data Center at LLNL is in place and functioning.
    - Additional effort requires external funding.

Strategic Action Plan #	Topic	Strategic Plan element	ERP Recommendation
SAP 1	Build bridges between data and HPC computing environments	VG2, SD2	R1, R11, R12

SAP 2. Seamless Integration of Multi-disciplinary data.
Issue Being Addressed: It is currently difficult to seamlessly integrate data across domains. To some extent the IRIS DMC acts as an integrating data center for time series observations that can be converted into SEED format but this solution is not scalable. Steps need to be taken to improve the discovery, access, and utilization of data from multiple domains.
Status: This is a complex problem. IRIS DS was successful in competing for and receiving the largest EC Building Blocks award that focuses on exposing data from 14 different data sources through web services similar to what IRIS has developed for the DMC and FDSN seismological holdings.

IRIS DS has developed a Federator that will allow seamless access to data from 2 data centers in the US and 5 data centers in Europe. This system exists and the concept of federation extended beyond the US and Europe.

IRIS DS is working closely with the EC Brokering (B-Cube) and EC Registration Services (CINERGI) groups to provide unified access to data from multiple geosciences domains.

Required Actions: IRIS DS should be prepared to compete for additional EarthCube and Cyberinfrastructure for the 21st Century (CIF21) solicitations to expand efforts in this area. Progress will require additional funding. Additionally IRIS should

- Promote simple web services within seismology and build federated data access systems working within the FDSN framework. Complete the Seismological Federator system
- Extend simple web service concepts to other geosciences disciplines
- Continue to pursue funding opportunities and lead community developments in cross-disciplinary data access and models.
    - The IRIS lead WS Building Block within EarthCube is the first step in this direction.
    - Continue IRIS' leadership role in this area
    - Effort requires external funding.

Strategic Action	Topic	Strategic Plan element	ERP Recommendation

Plan #			
SAP 2	Develop interdisciplinary access to Geosciences Data	OC-6	R1, R11

SAP3. Improve Communications between IRIS DS and our User Community
Issue Being Addressed: As users become less "known" to IRIS DS through direct service connections, the DMC is in danger of losing an understanding of our users and their needs. Specifically the DMC does not require any formal login to access our services and at times we only know a user by their IP address.

Status: The DMC has extensive methods through which user interactions are tracked internally. There are two classes of user interactions, those based on email requests (primarily BreqFast) and those based on IP addresses (SeedLink, Web Services, Browser based). To this extent we know where our data are going down to the user or the specific computer. Email addresses such as gmail.com however make user identities opaque.

Required Actions: The DMC needs to develop a more sophisticated and targeted communication system. Users can self-identify themselves allowing links to be made between IP or email addresses and actual user contact information. This communications system would also allow users to register DS services they are interested in such as BreqFast, web services, or specific applications like ObsPy or jWeed.

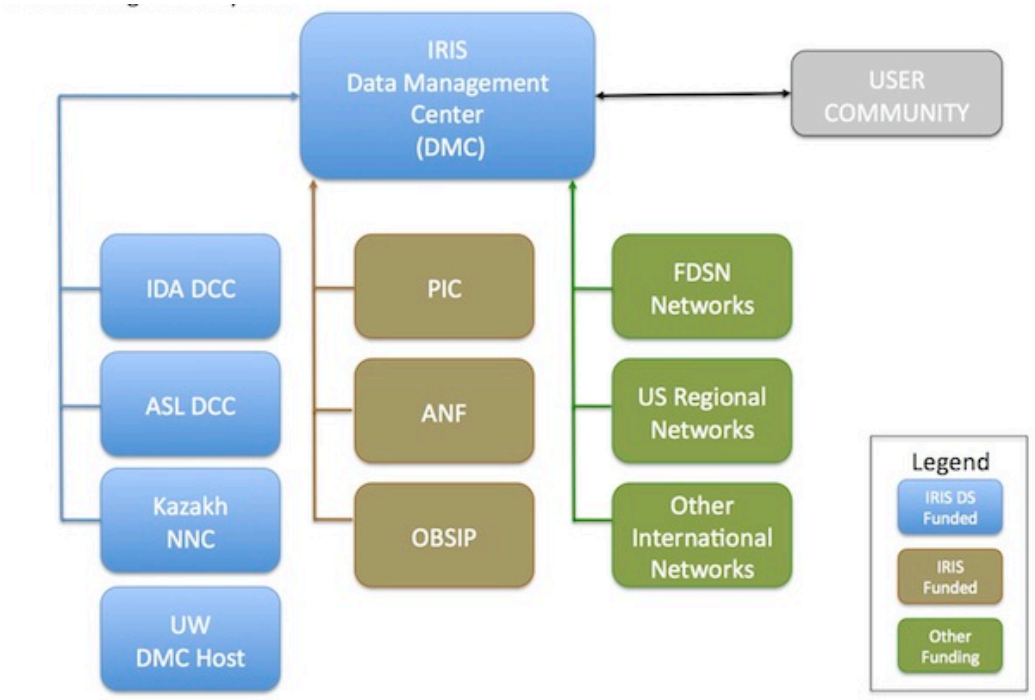
- Improve communication and education in the use of IRIS services,
- Develop an improved and more directed communications system to target channels of interest. This will allow a user to self-identify and specify which services and tools they have interest in.
- Conduct regular training courses in the use of IRIS software and access services.
- Involve early career scientists and students in beta-testing of new services and tools
- Develop on-line tutorials in the use of DS services
- Present DS services using Webinars
- Educate users more broadly in the benefits of using web services
- Work within the FDSN to enhance, promote, and improve web services more broadly
- Develop a user self help community patterned after the SAC-Help users list.

Strategic Action Plan #	Topic	Strategic Plan element	ERP Recommendation
SAP 3	Improve Communication between Data Services and its user	SD2, SD3, SD5, SD7	R2, R6

	community.		
	Develop an improved and more directed 2-way communications system to target topics of interest.		
	Better Understand the DS user community.		

SAP4. Simplified Organizational structure
Issue Being Addressed: The DS wiring diagram presented in the Briefing Book caused the ERP confusion

Status: This ERP recommendation was easily addressed and has been addressed already. The organizational requested by the ERP is shown below and also found at http://www.iris.edu/hq/programs/ds



The Components of the IRIS Data Management System

Required Actions: Redraw the DS wiring diagram.

- New diagram showing the central position of the IRIS DMC can be found at http://www.iris.edu/hq/programs/dms
-

Strategic Action Plan #	Topic	Strategic Plan element	ERP Recommendation

SAP 4	DS Organizational Structure	NA	R3

SAP5. Maintain Tight Control of quality of incoming data
Issue Being Addressed: The quality of most data entering the IRIS DMC can have marginal data quality and problems can go undetected.

Status: Most of the new QA system (MUSTANG) has been implemented and has been developed in a framework that allows it to be modified and expanded more easily in the future. The length of time it will take to run metrics on all data in the archive is excessively long and so additional servers have been ordered using funds from the previous cooperative agreement to dedicate to this task.

Required Actions: One of the key activities within IRIS DS in the next five years must be directed toward the close monitoring of data quality and proactive involvement in improving global data quality.

- Retask existing resources at the DMC to focus on improving the MUSTANG system and data quality in general. This includes software developer support, involvement of a data analyst, and devoting additional data technician resources.
- Improve communication between DMC/MUSTANG and network operators
- Improve two-way communication between DMC users and QA systems
- Acquire additional resources needed to allow MUSTANG to measure metrics on all data in the DMC in an expedient manner.
- Improve the messaging system that connects the various MUSTANG components.
- Complete the system that determines when recalculations are necessary when metadata or time series data change

Strategic Action Plan #	Topic	Strategic Plan element	ERP Recommendation
SAP 5	QA systems	SD-9	R3, R13

SAP6. Data Services Technology Advisory Board (TAB)
Issue Being Addressed: Formation of a Data Services Technology Advisory Board should be considered. Such a committee could include representatives from major national and international data and information technology programs and infrastructures, and have substantial representation from the commercial IT community. The TAB would provide an advisory role that supplements the domain-specific role of the Standing Committee.

Status. The DSSC believes that the concept of external expertise being provided to IRIS DS is an excellent idea. However, rather than setting up an independent

Advisory Board for Data Services (and perhaps other elements of IRIS), the DSSC believes it makes more sense for the Advisory Council being established for the BoD to include representatives that have technical expertise in a variety of IRIS related activities including Data Services, Instrumentation Services, and Education and Public Outeach. The DSSC thinks that inviting IT experts from industry and academia to fall meetings of the DSSC makes more sense than establishing a separate Technical Advisory Board (TAB) for Data Services.

Required Actions:

- Recommendations need to be presented by DSSC to the IRIS BoD
- Establish very specific tasks and a charter for the Advisory Council to the IRIS BoD that includes input for IRIS Directorates.
- Consider having the Advisory Council involved in expanding the funding base that might have benefit for IRIS Data Services.
- Consider having key IT professionals participate in Fall DSSC meetings

Strategic Action Plan #	Topic	Strategic Plan element	ERP Recommendation
SAP 6	Establish a Technical advisory Board for DS	SD3	R5

SAP7.
Issue Being Addressed: Data Provider Agreements

Status: A Data Provider Agreement (DPA) has been developed and has been endorsed by the DSSC. Data Provider Agreements are now being requested and many have been received from data providers that submit data to the IRIS DMC. Currently IRIS DS is requesting DPAs from all current network data providers. We have about 30 of these agreements in place as of this date.

Required Actions:

- Draft should be reviewed and modified as necessary
- Data Provider Agreements should be distributed to existing data providers
- Potential new data providers should be required to sign an agreement before data are accepted.

Strategic Action Plan #	Topic	Strategic Plan element	ERP Recommendation
SAP 7	Data Provider Agreements	OC-7, SD11, SD-12	R9

SAP8.
Issue Being Addressed: Data Acceptance Policy

Status: A draft Data Acceptance Policy has been drafted and should be modified and endorsed as needed by the DSSC. This is anticipated at the Fall 2014 DSSC meeting.

Required Actions:

- Draft should be reviewed and modified as necessary
- Data Acceptance Policy will be followed for future new data sources

Strategic Action Plan #	Topic	Strategic Plan element	ERP Recommendation
SAP 8	Data Acceptance Policy	OC-7, VG-2, VG-5, SD-7, SD-12, OC-2	R10

SAP9.
Issue Being Addressed: Promote Web Services as the Global Standard

Status: IRIS has continued to lead in the adoption of web services within seismology as well as across other disciplines. We are currently working with FDSN partners to build a Federated system of data centers around the globe.

Required Actions:

- Promote web services within seismology
- Develop a federated system for seismology using the FDSN standards for web services
- Extend web services to other domains through EarthCube and other new funding sources
    - Collaborate with brokering and registration systems

Strategic Action Plan #	Topic	Strategic Plan element	ERP Recommendation
SAP 9	Promote Web Services	VG-1, VG-2, VG-3, VG-4, VG-5, SD-5, SD-9. SD-10	R13

SAP 10.
Issue Being Addressed: Broaden representation within DS Governance

Status: Needs further discussion within current IRIS governance

Required Actions:

- Further discussion within DSSC
- Possible expansion of liaisons attending at least Fall DSSC
- Make sure that experts from other geosciences domains have an opportunity to provide input into IRIS and IRIS DS operations

Strategic Action Plan #	**Topic**	**Strategic Plan element**	**ERP Recommendation**
SAP 10	Expand standing committee representation to include other geo-expertise	SD4	R14

2012 Strategic Plan Items

Visionary Goals (VG)
- VG1- Simplified Data Management
- VG2- Sustainable Reliable Archive
- VG3- Intuitive Data Access
- VG4- Stellar Science Facilitation
- VG5- Diverse Data and Integration
- VG7- Transparent Data Quality
- VG8- Innovative Enabling Data Products
- VG9- Diversified Funding
- VG10- Connected Generations & Disciplines

Strategic Directions (SD)
- SD1-Engage and Enhance the Community
  - SD-2 Engage the Science Community
  - SD-3 Expand Outreach to the Scientific User Community
  - SD-4 Broaden the User Base
  - SD-5 Promote the Capabilities Broadly
  - SD-6 Influence Governance
  - SD-7 Incentivize International Data Access
- SD-8 Transform Data Management
  - SD-9 Improve Data Quality
  - SD-10 Revolutionize Data Handling
- SD-11 Diversify & Expand Funding
- SD-12 Understand Priorities

Obstacles & Challenges (OC)
- OC1- Disparate User Expectations
- OC2- Unprioritized Scope
- OC3- Debilitating Inertia
- OC4- Neglected Fund Raising
- OC5- Unrealized Quality Control System
- OC6- Restricted Customer Base
- OC7- Protectionist Data Policies
- OC8- Undervalued Core Maintenance

External Review Panel (ERP)

Recommendation 1: [R1]

1. a) IRIS Data Services should be commissioned to construct a next generation data and computational environment to ensure the realization of high-end scientific developments (large-scale data-intensive science), driven by scientific priorities and expertise, with an enhanced data products program.
2. b) There should be a focus on the seamless integration of multi-disciplinary data, data-intensive applications, placing “data close to cycles”, massive data-mining tools, as well as real-time assimilation of data and simulations.
3. c) The selection of the higher level tools and services to be developed should be based on the priorities identified in the Seismology Grand Challenges document, using a competitive process.

IRIS Data Services has correctly invested considerable effort to create modern web-service delivery, but needs to overcome community apathy for migration to these more efficient modes of working.

Recommendation 2: [R2]
IRIS Data Services need to make more effort on communication and education to improve use of modern data access methods.

The current organizational structure of IRIS DS carries with it a considerable historical component, and the Panel suggests a clearer structure based on the nature of data flow.

Recommendation 3: [R3]
A simplified organizational structure for IRIS Data Services should be considered, centered on the DMC as the dominant node, with all Data Collection Centers treated in a comparable way.

Recommendation 4: [R4]
IRIS Data Services should maintain a tight control on the quality of incoming data, and the homogeneity of the QA procedures applied in the different Data Collection Centers.

The range of issues that need to be dealt with by IRIS Data Services increases with the broader range of data being handled. We therefore recommend the creation of a new Advisory Board to allow the DS to draw on expertise beyond that available to the Standing Committee.

Recommendation 5: [R5]
A Data Services Technology Advisory Board should be established, including representatives from major national and international data and information technology programs and infrastructures, and should include substantial

representation from the commercial IT community, to provide periodic advisory role supplementing the domain-specific role of the Standing Committee.

IRIS Data Services provide excellent support to a diverse range of users, but will need to know more about their users to facilitate migration to more modern request tools, and, given the new emphasis on data products for the public, educators and non-seismologist scientists, to establish an environment in which the broader user community can provide self-help rather than the full burden falling on DS staff.

Recommendation 6: [R6]

1. a) Investigate mechanisms to develop a resilient self-help community for new data flows and services to lessen the impact on DMC resources. This will become more of an issue when such services become more widely used in interdisciplinary science.
2. b) IRIS DS needs better understanding of user community for their web services and data products, not only to ensure the DMC is providing features and support, but also to create a vibrant user community. This understanding will also allow for better metrics describing the users, the nature of the utilization of the services and data, and possibly the impact on science.

The range of data being handled by IRIS DS has moved beyond the different flavors of seismology. To improve capacity for multi-disciplinary interactions it is desirable that IRIS DS be aware of emerging International data standards and able to translate meta-data into appropriate forms

Recommendation 7: [R7]

To serve a broader community and new multi-disciplinary applications, web services that conform to community defined standards will be needed (e.g., OGC, KML etc.); these can likely be deployed with relatively modest additional effort.

Recommendation 8: [R8]

IRIS DMC must constantly track emerging international standards, such as ISO 19115 for metadata or the Open Geospatial Consortium and ISO 19119 standards for web services, and implement these standards for data exchange as appropriate.

Recommendation 9: [R9]

Utilize the OAIS (Open Archival Information System) Reference Model as appropriate to examine the overall archive functionality in relation to ingest and provision of seismic information (http://public.ccsds.org/publications/archive/650x0m2.pdf). In particular, look into establishing appropriate Submission Agreements between the archive and existing/future data producers.

A wide range of data from a huge number of sources are currently archived at the Data Management Center, we feel that a more transparent procedure is needed for the acceptance of data and the establishment of appropriate working arrangements with the data providers.

Recommendation 10: [R10]
Develop a clear and publically available decision tree for "what to archive?" to formalize the approval process, with a clear definition of what level of commitment can be made from the DMC level, IRIS level, or NSF level. The roles and responsibilities of the contributing parties and perhaps funding alternatives need to be negotiated.

With the recent move of IRIS to take on OBSIP, a range of opportunities needs to be considered in the ocean realm.

Recommendation 11: [R11]
The addition of OBSIP (Ocean Bottom Seismograph Instrument Pool) data streams to DMC brings more ocean seismic observations and investigators into the IRIS network. Other seismology-relevant data sets, such as ocean passive acoustics, geodetic measurements, tsunami buoys and tide gauges, should be assessed as either candidates for direct archive at the DMC (if not properly preserved and accessible elsewhere), or made easily available to the IRIS community through conduits to the designated archive sites.

The Panel was asked to consider the role that IRIS Data Services might play for future High Performance Computational resources (HPC) dedicated to seismology. We do not consider that the establishment of a hardware facility linked to Data Services would be an appropriate model, But, the expertise in Data Services could be applied to support and facilitate such a development.

Recommendation 12: **[R12]**
IRIS DS should support both the development of the computational environment and tools, as well as the access to the required computational resources needed for individual developments (via a modest on-site cluster and as a broker for access to national HPC capabilities)

The Panel further considered a range of issues related to the international nature of seismology, where the efforts of IRIS Data Services have been a major contributor to the present state.

Recommendation 13: [R13]
1. The DS is encouraged to continue its efforts for the establishment of the future architecture of a true global platform for the collection and distribution of high-quality data, and to develop and disseminate the appropriate standards and QA procedures to facilitate the improvement of data quality worldwide.

2. DS should expand efforts already initiated for the establishment of web-services as a global standard for data access and distribution.

Data entering the Data Management center comes from networks established for a variety of purposes, the Panel consider that some attention should be given to the broader context and the way in which data may be used outside the traditional IRIS goals

Recommendation 14: [R14]

1. a) DS should produce a detailed map and a strategic plan for all the current data and services recognizing broader applications than seismic source and structural studies.
2. b) DS should ensure that the composition of its SC reflects the different fields covered by its data and services, as well as the different non-research applications such as monitoring and hazard research

ERP Recommendations to NSF are included here for completeness

We recognize that the implementation of these recommendations in full would require additional resources for IRIS Data Services.
We also make a number of recommendations directed to NSF rather than IRIS
Recommendation N1:
We encourage NSF's continued support and development of EarthCube and/or similar initiatives, but caution that the close involvement and funding of the IRIS DMC and other relevant domain data centers will be key to ensuring that broadly useful cyberinfrasture for the geosciences is developed. IRIS DMC staff are already fully committed and EarthCube will need to provide new resources to enable their full and essential participation and avoid the shortcomings of past NSF-funded initiatives in data integration.

Recommendation N2:
NSF is encouraged to coordinate with the European Community and other key national agencies (such as Japan, China) the establishment of global funding and cooperation frameworks targeted to the improvement of global availability of data and understanding of earthquakes.

Recommendation N3:

1. a) NSF and IRIS should make a concerted effort to reposition the DS with the agencies responsible for the tasks presently covered, such as Department of Energy for all their applications related to nuclear monitoring research and hazard for critical infrastructures and energy systems and the Department of Interior for applications connected to earthquake monitoring and hazard assessment.
2. b) NSF, IRIS and USGS should develop a common understanding on the role occupied by the DS data and services in the hazard and monitoring fields.