Workshop on Data Products for Education and Research from the USArray

Portland, OR
October 11-12, 2004

Table of Contents

I. Introduction

EarthScope is a major new NSF initiative that is applying modern observational, analytical and telecommunications technologies to investigate the structure and evolution of the North American continent and the physical processes controlling earthquakes and volcanic eruptions. Data products from EarthScope will provide a foundation for fundamental and applied multidisciplinary research throughout the United States, which will contribute to the mitigation of risks from geological hazards, the development of natural resources, and the public's understanding of the dynamic Earth (http://www.earthscope.org/overview).

USArray, one of the three instrumentation facilities comprising EarthScope, is a continental-scale seismic observatory designed to provide a foundation for integrated studies of continental lithosphere and deep Earth structure over a wide range of scales. USArray will provide new insight and new data to address fundamental questions in earthquake physics, volcanic processes, core-mantle interactions, active deformation and tectonics, continental structure and evolution, geodynamics, and crustal fluids (magmatic, hydrothermal, and meteoric). The USArray facility will consist of three major seismic components (see http://www.earthscope.org/usarray/index.html):

1. A ***transportable array*** of 400 portable, unmanned three-component broadband seismometers deployed on a uniform grid that will systematically cover the US;
2. A ***flexible component*** of 400 portable, three-component, short-period and broadband seismographs and 2000 single-channel high frequency recorders for active and passive source studies that will augment the transportable array, permitting a range of specific targets to be addressed in a focused manner; and
3. A ***permanent array*** of high-quality, three-component seismic stations, coordinated as part of the US Geological Survey's Advanced National Seismic System (ANSS; www.anss.org), to provide a reference array spanning the contiguous United States and Alaska.

These seismic components will be complemented by an array of 30 ***magnetotelluric sensors*** embedded within the transportable and permanent arrays that will provide constraints on temperature and fluid content within the lithosphere, and by 16 permanent geodetic-grade GPS receivers, closely integrated with the PBO program, that will image continental-scale deformation. The goal of this layered design is to achieve imaging capabilities that flexibly span the continuous range of scales from whole Earth, through lithospheric and crustal, to local.

The success of EarthScope and USArray will be measured by the quality of the data products produced and the effectiveness with which they are distributed to the seismological community and to the public. Data products for USArray range from basic data that support scientific research (such as waveforms) to images and models of the 3-D structure beneath North America. These products can be classified based on the degree to which their production can be automated and on the breadth of their user communities.

I.A. Workshop Objectives and Organization

A USArray data products workshop was held in Portland, OR on October 11-12, 2004. The objectives of this workshop were to specify and prioritize standard routine and higher-order USArray data products; to establish protocols and procedures for creating, reviewing and updating these products; and to propose a framework for supporting this work. In particular we focused on when and how to define protocols that will lead to automated or semi-automated techniques for analysis of data that are currently time-consuming and require considerable scientific input. Thus we hope that some data products that are currently created by individual researchers through NSF small grants programs will become standard products that can be generated routinely at USArray facilities or at university-based "satellite" facilities. The move towards more “routine” products will greatly increase the pool of researchers who have timely and complete access to these products, will facilitate comparison of results from different parts of USArray, and will free up research time and funding for new approaches to data analysis, modeling and interpretation.

To achieve these objectives, the workshop included several talks summarizing the current and planned data management framework for USArray data and products and describing ongoing and planned higher-level data products. An afternoon plenary session on Day 1 included a number of short presentations about ongoing efforts related to USArray as well as a discussion of charges to four Working Groups. These background talks were followed by small group discussions on a variety of topics. Participants then regrouped to discuss the conclusions of the discussion groups and address several broader topics.

Contact information for the organizing committee and for all meeting participants is found in Appendices A and B. A detailed agenda and charges to the discussion groups are included as Appendices C and D. Summaries of introductory presentations provided by the speakers are given in Appendix E.

I.B. Classification of USArray Data Types

Previous EarthScope documents have outlined a simple classification scheme for data products, which range from raw data through interpretive products that require integration of seismological “products” with modeling results and other geological and geophysical observations (see: EarthScope/USArray Data Management Plan, T. Ahern, Aug. 2004). In this section, we discuss these data levels as interpreted by workshop participants. An assumption behind this discussion is that the responsibility for funding Level 0-2 products should lie with the EarthScope/USArray Operations and Management (O&M) budget. Level 3 and 4 products should be funded through NSF small grants programs based on peer reviews and evaluations by the EarthScope, Geophysics, Tectonics and other disciplinary and interdisciplinary panels. For this reason, it is important to establish a clear, but flexible, definition of the characteristics that separate Level 2 and Level 3 products and to recognize that the boundary between these levels will evolve as data processing techniques and community needs evolve over the lifetime of EarthScope/USArray.

Level 0: *Raw continuous data and event gathers of segmented raw data*: Several quality control parameters that are calculated automatically as data are transmitted to the DMC are also considered to be raw data.

Level 1: *Continuous and event-gathered quality-controlled data*: The IRIS DMC and associated satellite data handling facilities have been providing the seismological community with Level 0 and Level 1 products from the Global Seismographic Network, PASSCAL experiments and other permanent and temporary networks for 15 years. The support facilities needed to produce and deliver these products for USArray are currently in place, although there are a numbers of decisions that must be taken by the facilities and the seismological community concerning how to manage the volume of data that will be generated by USArray.

Level 2: *Products derived from waveforms*: Level 2 is interpreted to include secondary products that can be derived from waveform data using well-defined and widely accepted semi-automated procedures. At present, only a few Level 2 products are available to the seismological community on a regular basis. Examples are hypocentral information provided by ANSS and regional networks and the centroid moment tensor (CMT) source mechanism solutions produced by Harvard, by NEIC and by some regional networks. However, to fulfill the goals of EarthScope and satisfy the needs of the broader EarthScope community, which have supported establishment of the USArray based on the expectation of "products" describing critical aspects of earthquake source characteristics and Earth structure, it is essential that the variety and complexity of Level 2 products be expanded. Much of the discussion in the Working Groups centered on identifying which Level 2 products are currently possible or can be expected in the near future and which have the broadest scientific value for the EarthScope community. **It is important to note that groups of experts will be needed to evaluate proposed procedures to generate additional Level 2 products and that most Level 2 products will require continuing review by trained seismologists to ensure the quality of the results.**

Level 3: *Products that require technical analysis and interpretation of USArray products (Interpretive products)*: Level 3 products are those for which procedures and protocols are currently a topic for research and for which no established standards exist or for which there are several competing standards. Significant input by research scientists is needed in the construction of these products. The expectation is that these products will be generated through NSF grants to PIs.

Level 4: *Products that represent integration of USArray data products with other geological and geophysical data (Knowledge products).*

We expect that production of many Level 2 products will occur in parallel with Level 3 efforts to improve analysis techniques to generate these products and Level 4 efforts to integrate the results with other geophysical and geological data. We also

note that routine processing of massive data volumes following well-documented and stable protocols over a long period of time, as is required for Level 2 products, is generally not compatible with limited-term NSF investigator-driven small grants.

Because EarthScope will not fulfill its promise unless a mechanism is established for funding this work, it is important that the community achieve a consensus about criteria for distinguishing between Level 2 data products and higher-level products. **The workshop participants recommend the following procedures and criteria for determining which products are Level 2 products that should be supported by contracts included as part of the EarthScope O&M budget:**

- An oversight group appointed by the IRIS Data Management System Standing Committee (DMS-SC) should be assembled to assess procedures to produce proposed level 2 products.
- Level 2 products should be required to adhere to a timely delivery schedule.
- Level 2 products should have a clear user application.
- Level 2 products should be well-documented to optimize their usefulness to the research, education, and outreach communities.
- Contracts to produce Level 2 products should be reviewed annually. Contract renewals will not be automatic.

While some of the anticipated level 2 products will be produced at EarthScope or IRIS facilities, many may be produced at "satellite" facilities located within larger research groups

II. Workshop Summary

II.A. Background

To establish a common basis for discussion, the workshop began with overviews of several components of the existing data management framework followed by talks from university-based groups that have been, or soon will be, providing seismological "data products" to the seismological community, to the broader Earth Science community, and to the public at large. Abstracts provided by the speakers after the meeting are given in Appendix E. Many of the presentations can be found on the IRIS web site at www.iris.irs.edu/Pubications/PubFrameset.htm under "USArray Data Products Meeting."

II.B. Charges to Working Groups

Working groups covered a wide range of topics. WG1 focused on basic data products and related metadata, including discussion of whether the scale of USArray would require significant modifications to "business as usual." WG2 focused how to maximize the value of USArray for education and outreach. WG3 discussed event characteristics, and WG4 discussed Earth structure.

Charges to the Working Groups were drafted prior to the meeting and circulated to all participants for discussion (see Appendix D). These charges were revised and expanded during a plenary discussion at the workshop. Participants were also encouraged to submit one-page summaries of proposed data products to initiate discussion (Appendix E).

II.C. General Concerns Common to all Working Groups

II.C.1. Product Maintenance

Continuity and homogeneity of the data products over time remains a central concern for both the EarthScope community and for the larger public that the products will serve. EarthScope has a lifetime of at least 12 years and over that time methods are certain to change significantly. In some cases it will be difficult to compare results produced by previous techniques with those from new techniques. *It is thus imperative that when techniques to produce data products are modified or replaced, previous results are re-determined to maintain continuity of the final data product.*

II.C.2. Product Evaluation

To assure that the best possible data products are produced and assessed, contracts should be strictly limited in duration and not open to automatic renewal. An expert working group should meet a year before the Request for Proposals (RFP) to assess products, evaluate current product needs, and discuss potential new techniques that may be included as products. This group will also evaluate proposals and renewals based on well-defined criteria that include those listed in section I.B. Broader community feedback to those generating data products will occur at EarthScope and IRIS annual workshops.

II.C.3. Data Distribution

USArray data products should be archived at multiple locations, including the DMC, though this process remains to be formalized. For seismic waveform data, procedures currently in place at the IRIS DMC and at other distributed data centers provide a very successful model. For products derived from waveforms, procedures have yet to be established. Coordination is needed between EarthScope and USArray program managers, the IRIS DMC, and other groups developing web-based Earth Science data and software facilities (e.g. SCEC, CIG, GEON, ANSS) to establish efficient and effective web portals and leverage existing information technology (IT) infrastructure to service the needs of various communities. Issues that need to be discussed include definition of what constitutes a "product package," which must include the electronic publications needed to explain these products. Also needed are mechanisms to provide access to data "portals" at different levels and to regulate submission of data products.

As an example of a product package that would make effective use of the USArray framework, we propose a series of electronic "atlases" timed to correlate with the eastward progression of the transportable array. These atlases would include maps of the parameters identified as Level 2 event characterization and Earth structure products and accompanying explanatory material presented on a level that should be interesting to students and to the geologically literate public. These atlases should be carefully reviewed and packaged in a way that makes them easy to download and print, and to make it easy for numerical values associated with plots to be extracted. They would thus represent an enduring legacy of the USArray program.

II.C.4. Development of new software tools

Some software tools are well established and will require no further development for a product to be useful. For example, if the product can be presented as a web page graphic, the task is relatively simple. This is not possible, however, for many of the data products described in this document. For example, a P wave tomography model could be presented to a general audience in the form of an image from a published paper, but a specialist might want the velocities at points in the grid and a method to interpolate the grid data to compute the P velocity at an arbitrary point in space. Developments are underway in groups like GEON to define standard objects that might prove useful for distributing some of these higher order data products. EarthScope will need to consider standardizing certain data object descriptions to facilitate the distribution of many higher level data products. For this to be useful to the community, a data product would ideally need to be connected to two software products:

- Standard "viewers" that could be used to visualize the data. An example would be a Java Applet to visualize slices of any orientation through a 3D model .
- An object "constructor" library of computer code that could be used to build an application program around a particular data product. A standard idea in object-oriented programming is abstraction of the data to the "concept" implemented through a set of data (e.g. a P wave velocity model "concept" should include a function to return the P wave velocity at a specified latitude, longitude, and depth). Standardized data objects can accomplish this if they are delivered with a set of Java and/or C++ code to construct a standard object from the external representation supplied by the EarthScope Data Portal. This would facilitate the use of the data by removing the age-old "format" issues.

Similarly, tools will be needed to construct models for comparison with data products. Specification and development of such tools should be undertaken in close collaboration with CIG. A model for development of such a facility is currently available as part of the GeoFramework project (www.geoframework.org).

II.C.5. PI Recognition

A final issue is the question of how the "routine" work of producing a USArray data product will be recognized within the scientific community. This question pertains particularly to young faculty and students as they work to gain community recognition through their scientific work. The fact that those producing the data products will have the first look at the results may help to alleviate this problem, but EarthScope will not reach its potential if researchers avoid data product efforts due to the perceived cost to their professional careers. We therefore recommend that EarthScope community spearhead a community effort to provide the recognition necessary to properly acknowledge the important efforts of providers of data products. One way to encourage such activity is to require that each product be accompanied by a peer-reviewed publication on the techniques behind the product that should be referenced by all who use it, as has been done with GMT, a widely used package for manipulating and displaying geographic and other data. Another way to engage young researchers is to award fellowships, internships, or other widely acknowledged vehicles for advancing data product efforts while supporting undergraduate, graduate, or post-graduate careers.

II.D. WG1: Waveform data products and station metadata

II.D.1. WG members: Gary Pavlis (discussion leader), Anne Trehu (recorder), Frank Vernon, Chad Trabant, Gary Egbert

II.D.2. Summary:

Over the past 15 years, the IRIS data management center (DMC) has developed an efficient database system for archiving seismic waveform data from the GSN and from temporary PASSCAL deployments. The DMC has also developed a collection of software tools to query that database and provide the waveform data packages requested by the seismological community. This discussion group explored the impact of USArray on DMC operations and discussed ways in which procedures could be improved to better serve the community. We also discussed the impact of expanding USArray to include magnetotelluric data as well as seismic data.

We agreed that the overriding goal of the DMC archive is to be inclusive, with data from all arrays operating in the US included in the database. This includes data from regional networks as well as data from the permanent, transportable and flexible USArray networks. Data do not, however, need to physically reside at the DMC. Data requests can be filled by extracting data from several different sites, provided that the process is transparent to the user. In order to achieve this, an upgrade of the existing Networked Data Center (NetDC) software will be required, which must include web-based requests and other data types in addition to the broadband waveform data it currently handles. We also agreed that arrival time picks of various phases need to become part of the Level 2 database and that standardized procedures to catalog phases should be established. Finally, we agreed that new tools must be developed for efficient visualization of station state-of-health and background noise data.

Table 1. WG1 – Waveform data products and station metadata.

Source	*Product*	**Level 0-1**	**Level 2**
ANF/DMC	Raw and quality-controlled continuous data	X	

ANF/DMC	Raw and quality-controlled event data	X	
ANF	Absolute arrival times & amplitudes		X
ANF	Polarization anomalies		X
ANF/DMC	Noise analysis		X
ANF/DMC	Waveform data corrected for instrument response		X
MTF/DMC	Continuous magnetotelluric data	X	
MTF/DMC	Magnetotelluric transfer functions		X

ANF – USArray Array Network Facility; DMC – IRIS Data Management Center; MTF – USArray Magnetotelluric Facility (not yet established)

II.D.3. Discussion of data products

II.D.3.a. Waveform data and arrival time picks:

Association of event information with time series data is a common starting point for a large fraction of seismic data processing procedures. Current practice at the IRIS DMC is to deliver only basic location information (latitude, longitude, depth, origin time) with seismograms. This will need to change with the USArray for a number of reasons. First, the Array Network Facility (ANF) will be routinely making arrival time picks as part of standard operations. This represents a significant investment in personnel time that most scientists would not wish to replicate. Secondly, the existence of a pick indicates that a real phase exists, which can be utilized for additional processing. Finally, arrival time picks are a basic starting point for many types of analysis. Consequently, delivery of as many arrival time picks as possible with event data should be a data product of the highest priority. The delivery of catalog data with waveform data has a number of practical elements that will need to be worked out to optimize the value of these data to the community.

- Because the USArray will overlap multiple regional seismic networks, many events will have multiple location estimates produced by different groups. This is a standard concept in all catalog databases, but procedures to deliver this information to the end user will need to be developed. This issue was also discussed by WG3.
- A corollary of item 1 is that IRIS DMS will need to coordinate with the ANSS to define unique event identifiers that associate independently determined location estimates to a unique "event".
- The workshop identified a strong need for one-stop shopping to acquire event-based seismograms. The archive should allow a user to extract all available seismic data for a given event even if the data are not at the IRIS DMC. A similar capability should exist for arrival time picks made by regional networks. A comprehensive set of available picks should be a metadata product available for any data request. This will require extensive coordination with the ANSS, but accomplishing this coordination will benefit both EarthScope and the ANSS. The technical problems required to solve this probably differ little from the problem of building a seamless access to waveform data from both facilities.
- When possible, arrival time measurements should come with a measure of signal-to-noise ratio. This should be a field that could be queried from a database interface to

allow users to select data based on this attribute. This could be used, for example, to improve the efficiency of one of many cross-correlation based processing procedures.

- IRIS DMS needs to develop a mechanism for flexible array and PASSCAL experiments to contribute arrival time picks to the archive. This will require a nonstandard database definition of arrival data. Universal practice in operational networks is to define one definitive pick for a given phase. This would have to be modified for the USArray archive as conceived here because the same phase at the same station might be measured by more than one person or organization. This is valuable information worth saving, and keeping everything helps resolve disagreements about which measurement is "correct".

II.D.3.b. Station metadata and noise analysis:

Conventional metadata associated with waveform data includes station and event locations, sensor orientation, timing information, and instrument response (assumed to be linear and represented by poles and zeros). The SEED data format, which has been a standard since 1991, is adequate to cover this information and should be retained for USArray. We note that this format, with some extensions to include the electrode configuration and other parameters, is also suitable for magnetotelluric data. It will be important, however, for the magnetotelluric community to come to a consensus on a common content and format to describe extensions to the SEED header to incorporate metadata required for these data.

A number of additional station-related metadata have been discussed in various USArray planning documents, including photographs of the station sites and documentation of the rock/sediment type. A photograph of the station installation should be included with the metadata from each station. Photographs of the site may, in some cases, be problematic because of privacy and station safety issues. Information on the geological characteristics of the site should also be included (e.g., sand, clay or crystalline rock). If additional geotechnical information is available from high resolution geophysical surveys (e.g., depth to basement), it should also be included in the station metadata file.

The capability to rapidly evaluate expected data quality by scanning parameters representing the background noise level and polarization anomalies at a given station would be useful to users of USArray data. This information can potentially be used to study Earth structure as well as to select low-noise data for a particular study. Summarizing background noise level, however, is complicated by the fact that the noise level is frequency- and time-dependent. At present, the power spectrum of the background noise at each site is determined regularly as part of the QUACK quality control procedure. We recommend that new tools be developed to facilitate visualization and interpretation of these noise spectra. This tool should provide long-term noise averages, short-term histories of the change in noise level with time, and long-term evolution of the station PSD characteristics (e.g., PSD movies).

Similar tools should be developed to visualize the history of station state-of-health. Station problems are logged as they are discovered, but it is not, at present, easy to retrieve this history. Users should be encouraged to report problems discovered during data analysis and this feedback should be rapidly incorporated into the station database. The existence of a simple procedure for reporting problems and for retrieving problem histories would motivate users to participate in this dynamic database. These reports, however, must be quality checked by Advanced Network Facility (ANF) or other relevant personnel to prevent inaccurate reporting from compromising this database.

II.D.3.c. Magnetotelluric (MT) transfer functions:

MT transfer functions (TFs) estimated from the raw electric and magnetic field time series form the basic data from which electrical resistivity models are derived. The TFs involve substantial processing and reduction of the raw data, and should be classed as a level 2 product. Because of the volume of data to be analyzed and the systematic nature of the data acquisition, automated procedures for quality control and routine data processing will benefit the entire community. A proposal has already been funded (S. Park and G. Egbert) through the NSF EarthScope program to define the methods and develop the computer codes that will allow TF estimation to be automated to the greatest extent possible. (Development of quality control software is also part of the proposal). For at least the first few years the PIs on this project will take responsibility for maintaining the quality of this data product. It is presently unclear how the routine processing will be conducted over the duration of EarthScope. Even with extensive automation, there will still be a need for someone with expertise in the MT method to review TF estimates before making these available to the community.

II.E. WG2: Products for Education and Outreach

II.E.1. WG members: R. Aster (recorder), L. Braile (discussion leader), D. Seber, J. Taber, M. Francissen

II.E.2. Summary

Education and outreach (E&O) efforts are essential for facilitating permitting of sites, to fulfill the scientific and educational objectives of EarthScope, and to ensure the long-term legacy of the program. USArray data product issues are of fundamental interest to the science community in direct (e.g., satisfying NSF program expectations) and indirect (e.g., attracting talented new individuals to the field) ways. Data products summarized elsewhere in this report are also the raw materials for all specifically-targeted E&O products. Indeed, the fact that E&O products are based on actual current research by professional scientists can be a powerful selling point with many audiences. However, for data products to make the transition to E&O, resources must be allocated to facilitate necessary packaging, curricular development, and strategic distribution. A key element for successful E&O is to provide context for the data, i.e., to demonstrate how the data are relevant to a specific target constituency. These needs can best be met through funding of educational specialists working closely with IRIS, the EarthScope

Office, and individual researchers. EarthScope/USArray-associated products should be integrated into IRIS E&O professional development, curricular, and other efforts, taking into full account teacher needs, state standards, and other educator priorities. USArray efforts must take full advantage of already established IRIS Education and Outreach efforts and partnerships with other organizations (e.g., USGS) to optimize content and dissemination, and to prevent redundancy.

Products specific to E&O are summarized in Table 2. Most are packages of data products assembled from data products discussed by the other working groups. Resources needed to develop these products are discussed in section II.E.7.

Table 2. E&O products

Source	*Product*	*Level 2*	*Level 3/4*
IRIS E&O	Station Blog/Station Gallery	X	
IRIS E&O/others	Seismogram packages with teaching notes	X	X
IRIS E&O/others	Source and structure packages with teaching notes	X	X
IRIS E&O/others	Wave propagation and ground motion packages		X
IRIS E&O/others	Earthquake prediction lottery		X

II.E.3. Discussion of data products

II.E.3.a. Broad USArray Station Information

A key component of USArray's on-the-ground and outreach success includes communicating science results to individuals and organizations participating in the siting of the facility. This information will be communicated to siting partners both on the web and in a planned quarterly siting outreach newsletter. Besides including feature items on specific scientific topics for the general participating public, the newsletter could include links to on-the-ground information to help build the siting community. Examples include a "Stationblog", a weblog describing how each USArray station was used (e.g., for locations, data downloads). Much of the Stationblog content could be automatically generated and it could provide links to seismograms, events, and other products (see below). A similar product would be a "Stationgallery" that provides a photo-record of stations, sites, and communities that includes information about sites and sponsors (if desired; a site host could opt out if privacy is an issue). This could be a subset of a general station database that includes photographs already being assembled by USArray. Both of these products could be produced at the 0-1 Data Products level.

II.E.3.b. Seismogram Products

Raw or reviewed/assembled seismogram products will be built rapidly and easily in USArray, and will be widely disseminated on the web. Education and Outreach will require reformatting and contextualizing these products for a broad audience. Examples of resulting products include raw seismograms displayed and interpreted for E&O purposes (e.g., as 24-hour pseudohelicorder record which could be done at Data Products

Level 0-1. This type of display might emulate commonly used formats from the widely distributed AS-1 educational seismometer (e.g., http://www.jclahr.com/science/psn/as1/), regional networks, and other Educational Seismology Network (ESN) components, and should provide quick interpreted looks at seismic activity for a wide range of users. It would be preferable to have these products properly scaled to true ground motion. Properly scaled seismograms can be used to promote understanding of orders of magnitude, metric units, and other basic math and science concepts as well as highlighting the high-tech instrumentation aspects of USArray. At a higher product level (2-3), special Event seismograms (similar to interpreted and annotated FARM products from the IRIS DMC) would be especially valuable. The near-real-time aspect of seismology is a fascinating aspect of the science (for example, it is a compelling feature of the IRIS/USGS Museum Display Program). Real-time interest provides links to media, geography, politics, and other wide-ranging areas of general interest. Linking to or augmenting USGS/NEIC types of reports and showing noteworthy locally recorded earthquakes or earthquakes of global interest should be done for all such events to take advantage of "teachable moments" that arise following such events.

II.E.3.c. Source and Structure Products

Epicenters and hypocenters are of wide interest to the general public and to educators. Partnering with the USGS and regional networks, and including EarthScope-specific tie-ins for source information is thus of fundamental importance. Hypocenter and other source information should be tightly coupled to the dissemination of seismogram products (see above). The deep structure of the Earth (and how seismologists infer it) is also a very powerful vehicle for clarifying seismology's unique and important role in Earth Sciences. Source and structure can be linked through tectonic context, as is already done in many cases in USGS mini-posters and other materials disseminated following notable events. Hypocenter maps and associated interpretations (levels 2-3) are core products for the broad community. These efforts should be extended to include not just global events, but also local events with basic interpretation recorded near a particular community's stations to take advantage of local interests and "sense of place". Teaching the concept of magnitudes can be made timely by integrating recent global earthquakes into, for example, an earthquake location/magnitude "Exercise of the Week" that incorporates local or regional station data; this could be automatically generated to a certain extent, but it would still require expert review and annotation. As such it is probably a *Level 3* product. Links to background information such as tectonic context and previous events in the region would also be desirable. Another educationally interesting opportunity might be to demonstrate reciprocity where sources and receivers are fortuitously arranged. Higher-level source information data products such as source and radiation patterns, or rupture models should be packaged into educationally useful web products and/or posters (*Level 3 and 4*). An engaging way to display different types of faulting might be flash animations of faulting keyed to CMT solutions.

A readily available data product that could tie into basic Earth Structure interests would be "How thick is my crust?" In this case, an engaging web interface keyed to participant's zip codes could lead into other materials about continental evolution, for example. This provides a way to link the continent-spanning aspect of USArray with local interest.

Seismic imaging is mysterious to general Education and Outreach audiences. A high-level "How do they do it?" module linked to imaging products and example data from USArray would seek to explain via straightforward graphics and movies. Such a module would also provide a link to wave propagation, and seek to convey fundamental elastodynamic concepts in approachable terms. USArray will provide exceptionally dense record sections that can provide clear examples of fundamental concepts such as shadow zones and general refraction/reflection of waves. An associated product could be based on interpreted record sections to answer such basic questions as "How do we know the Earth has a core?", and would include ray-tracing diagrams, and data examples (*Level 2 and 3*). IRIS experience has shown that cutting-edge discovery is very exciting to audiences. A "Discovery of the Earth" project consisting of interpreted animations of imaging data products such as body-wave tomography could highlight the exploratory aspect of USArray throughout North America for many target audiences.

II.E.3.d. Wave Propagation/Ground Motion Products

Wave propagation appears in many state science standards, and USArray will provide compelling empirical examples that can be used to tie to this and can augment more traditional ways of showing wave propagation (e.g., wave tanks). The general response of the Earth and falloff of wave amplitude with distance could be excitingly demonstrated in a "weak motion" map across the transportable array, ANSS, and other stations from earthquakes of interest. Such a product could be merged with strong motion ShakeMaps produced by USGS (http://pasadena.wr.usgs.gov/shake/), the California Integrated Seismic Network (CISN) and other groups. "How much did it shake where I live?" could provide a local link to global or national events of special interest. In a variant, "Bring the Earthquake Here" would be a tool for generating synthetic ShakeMaps for timely or historic events brought to a local region (with interpretive comments on just how realistic this is). A key purpose of this exercise might be to show how attenuation varies across the US level and discuss links to earthquake hazard (*Levels 2 and 3*).

II.E.3.e. Earthquake Prediction Lottery

Once a year (e.g., Earth Science Week) USArray and partners could host a "Guess Next Week's Biggest Earthquake" contest (global and U.S. category). This exercise should include pre-event overview and follow-up describing global earthquake distribution and USArray recordings, activities, and then provide an interpretation of the winning earthquakes. "!WIN AN AS-1!" (educational seismometer) from IRIS is one hook that could be used to engage classrooms (*Level 2 and 3*). Such an exercise illuminates core concepts of probability as well as Earth Science.

II.E.4. Education and Outreach Audiences

Audience classifications have been well delineated by IRIS E&O and other programs in this area. In keeping with the wide-ranging goals of EarthScope, we would expect to enfranchise all relevant groups into the USArray data products effort. IRIS E&O generally classifies audiences as follows: K-4 (primary school), Middle/High School, Community colleges, Undergraduate institutions, Graduate Programs, Public and Informal Education, Media Relations, and Congress and other Official Audiences. The degree of interpretation varies with level; products for K-4 will be limited to specially packaged examples, while many Graduate level products could be quite similar to standard USArray products. Congressional and media-specific products will be developed and disseminated primarily as raw material for the use of the EarthScope Office.

II.E.5. Key Partners

Education and Outreach is an inherently highly collaborative enterprise as we seek to reach broad and extensive audiences. Clear partners in all efforts are: the EarthScope Office, the USGS, the Digital Library for Earth System Education (DLESE), State Geological Surveys and their organization, the Association of American State Geologists (AASG) , IRIS-affiliated and other Universities, the Southern California Earthquake Center (SCEC), the Geosciences Network (GEON), and the George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES).

II.E.6. Dissemination

A Siting Newsletter would be produced to engage and inform (and generally thank) the siting community on USArray/EarthScope progress, and also serve as a general interest EarthScope publication for the general public. Posters and 1-page information sheets should illuminate what we know and DON'T know in the context of EarthScope discovery. Focusing on the unknown aspects of scientific research is particularly relevant to USArray and EarthScope and is not emphasized in many education and outreach materials. Reaching teachers in the classroom is a fundamentally important aspect of realizing EarthScope goals of altering the public perception and appreciation of Earth Science. We envision USArray involvement with IRIS and other Professional development specialists/programs (e.g., National Science Teachers Association workshops). Direct involvement with the community of science teachers and their organizations is essential for getting teachers to use new curricular materials. Major media and multimedia are key opportunities for reaching a broad audience. Animations and other products developed for E&O purposes should be strategically used to spur major media interest and in some cases may be used directly in such productions. Another major opportunity for dissemination comes through associations with museums. We envision a flexible museum program incorporating USArray data products coordinated with IRIS and other partners. Much of the infrastructure already exists; both the Museum Lite (a simple customizable web-based museum display) and Major

Museum (special state-of-the art efforts with high-profile partners) programs of IRIS/USGS can be updated/augmented to include EarthScope/USArray materials. A final key aspect of disseminating educational and outreach materials is to partner with existing web portals, particularly the EarthScope web portal, but also those developed by key partners such as GEON, CIG, USGS/ANSS, IRIS, PBO, SAFOD, and DLESE.

II.E.7. Resource Needs of USArray Education and Outreach

The overall needs of E&O require providing core dedicated Operations and Maintenance human resources for interpreting, contextualizing, and disseminating data products. Simply implementing the core recommendations above to level 2 and incorporating an assessment of activities would require at least of order 2 FTE's, supplemented with additional funding for visiting teachers and interns. The rotation of interns and teachers through such an operation is viewed as essential to keep it effective and dynamic. IRIS Education and Outreach experience suggests that the following expertise be included at some level to realize the potential of USArray Education and Outreach interaction with data products: 1) specialists for reviewing/casting data products for E&O purposes, 2) a teaching specialist, 3) interns, sabbaticals, and other visitors bringing fresh ideas and experiences, 4) curriculum standards specialists, 5) assessment capabilities.

II.F. WG3: Event Characteristics

II.F.1. WG members: Ray Willemann (discussion leader), Stuart Sipkin (recorder), Harley Benz, Lind Gee, Linus Kamb, Paul Richards, Mitch Withers, Göran Ekström (left part way through), David Simpson (arrived part way through)

II.F.2. Summary

The motivation for USArray is primarily as a facility for studying the structure and dynamics of North America, but it will also provide data that can be used to characterize seismicity with unprecedented accuracy and detail. Earthquake parameters, including locations, magnitudes and moment tensors, play a major role in resolving seismogenic structures, the stress fields responsible for those structures, and elastic strain at seismogenic depth. In addition, inferences about Earth structure from seismic wave propagation often benefit from more accurate and complete information about the earthquakes that generate the seismic waves. Thus, making measurements to characterize seismic events is not just an added benefit of USArray, it is intrinsic to studying structure and dynamics.

II.F.3. ANSS/USArray relationship and priority levels

USArray data products related to event characteristics are unique in that they will be produced in a cooperative effort with the Advanced National Seismic System (ANSS).

Part of the cooperative effort undoubtedly will involve the ingestion of USArray waveform and phase data by the appropriate components of the ANSS (either the NEIC or one of the regional seismic networks) and the computation of event parameters by the ANSS. In addition, USArray facilities, such as the IRIS DMC, may make additional measurements that are more important for studying Earth structure and dynamics than for responding to earthquake disasters or studying seismic hazard. A central tenet of the cooperation between USArray and ANSS is that each will share all of its measurements promptly with the other and that USArray portals will distribute the ANSS seismicity catalog rather than preparing an independent catalog. The following prioritization of products reflects the requirements of the USArray mission to study structure and dynamics, not the requirements of the ANSS mission.

II.F.4. Table of event characterization products

The table of event-related products is divided into four priority levels, reflecting the importance of each product to the Earth structure goals of USArray. The priority levels supplement the EarthScope classification of products by level of complexity. Earthquake location is the fundamental product, which is required for development of any other products, and therefore has the highest priority. Priority B products are those that are essential to understanding the physics of earthquakes and/or using earthquakes as sources in studying Earth structure. Priority C products are those which could further improve understanding of the physics of earthquakes or for performing seismic hazard analysis, but only to the extent that priority A products are accurate. Priority D products are those for which USArray instrumentation or network configuration is not well suited, or that are less urgently required.

Event-related products from USArray must be coordinated with the catalog of events produced by the ANSS. Thus the table of event-related products includes a column, headed “ANSS,” which is checked where the primary role of a USArray facility would be to provide measurements that could be used to improve the ANSS catalog. Ongoing research continues to result in improved procedures for even some of the most fundamental event characterization products, and new procedures are often not yet tested widely to enough to be generally accepted. But for some measurements there is already a consensus on reasonable procedures, and the work required to provide the measurements consists of relatively straightforward software development by a USArray facility. These measurements are checked in column “Level 2”. Some event-related products are currently computed far more accurately, or with far greater precision, in the context of individual research projects, and with the high density of broadband stations of the TA and the FA it should be possible to develop procedures for routinely achieving a comparable accuracy and resolution in USArray products. For these products there is a check in the column “Level 3” to indicate that a principal-investigator project should be able to develop such procedures for implementation during the early part of the USArray project.

Table 3. Event Characterization Products.

Source	*Product*	*Level 2*	*Level 3*
Priority A			
ANSS	Prompt earthquake location	X	
ANSS/USArray	Precise earthquake location	X	X
Priority B			
ANSS/USArray	Phase picks	X	X
ANSS/USArray	Conventional magnitudes and scalar seismic moment	X	X
ANSS/USArray	Moment tensors	X	X
ANSS/USArray	Finite Fault models	X	X
ANSS	Amplitude decay with distance	X	
ANSS	Peak ground velocity and acceleration	X	
ANSS	Response spectra (5% damped)	X	
ANSS	ShakeMap	X	X
ANSS	Macroseismic effects (DYFI)	X	
ANSS/USArray	Event type classification	X	X
ANSS/USArray	First motion fault mechanisms	X	X
Priority C			
ANSS/USArray	Higher order moments		X
ANSS/USArray	Moment rate functions		X
ANSS/USArray	Static and dynamic stress drop		X
ANSS/USArray	Radiated Energy		X
ANSS/USArray	Rupture velocity and directivity		X

note: Directivity is currently reported in California as part of the "ShakeMap" effort and as a component of finite fault models

Many of the event-related products have checks in both the "Level 2" column and the "Level 3" column. These are products for which existing practices for routine computation are good enough that they should continue but where there is a clear path forward to significant and important improvement. Among the Priority A and B products, prompt locations and conventional magnitudes are unlikely to be significantly improved upon by a research effort related to the USArray mission. As detailed below, however, many other high priority products should be improved in the near future to support USArray studies of Earth structure and dynamics. Examples include phase picking by cross-correlation and relative locations, which have been enormously effective in improving results that included tomography in the context of research projects involving many thousands of events, and centroid–moment tensor solutions for numerous smaller events, which, for example, provide initial phase information required for surface wave phase velocity studies, as well as information on stress.

II.F.5. Priority A Products:

II.F.5.a. Earthquake locations

For those scientists using USArray data for its principal objectives of studying Earth structure and dynamics, it will be essential to have accurate characterizations of the earthquakes and explosions recorded by the transportable and flexible arrays. This is critical for improving the accuracy of the information derived from tomographic studies. In addition, the USArray project represents a major opportunity to improve other aspects of Earth structure and dynamics that depend critically on accurate earthquake locations, including fault dynamics, crustal stress and strain, and the relative importance of strength in the crust and mantle in lithospheric dynamics.

At present, seismic events are generally located one at a time by measuring the arrival times of particular seismic signals and then interpreting these observations in terms of the travel times predicted for a standard depth-dependent Earth model. Uncertainty in the location is estimated from the sensitivity of arrival times to perturbations in the accepted hypocenter. When teleseismic arrival times are used, the resulting location estimates may be in error by several km for events detected at hundreds of stations, and by a few tens of km for events detected at tens of stations. When regional arrival times are used, it is common for different institutions to report different location estimates for the same event. The locations differ typically by two or three km in areas with high station density, and by on the order of five km in areas with fewer stations. There are two independent reasons why these location estimates are so poor: the relatively crude velocity models for shallow-focus (<200 km) earthquakes and errors in arrival time picking.

Two types of major improvement can be anticipated in estimates of the location of seismic events. First, an improvement in *absolute* location estimates can be achieved if systematic station corrections are used to interpret the measured arrival times when events are located one at a time. Such corrections can be employed to achieve the equivalent of using travel-time models that are determined separately for each station in the network. Second, very substantial improvement in *relative* locations can be expected using waveform comparison methods to pick relative arrival times together with an algorithm for locating numerous events at the same time. Events whose ground truth locations are known can serve as reference events that will enable relative locations to be turned into absolute locations. Numerous studies of particular seismic regions have demonstrated that relative event locations can be estimated with precision on the order of 100–200 m, representing an improvement over current standard locations by one to two orders of magnitude.

In order best to achieve its objectives of advancing knowledge of Earth structure and dynamics, USArray and ANSS must cooperate to take full advantage of these demonstrably achievable advances in earthquake location accuracy. This essential cooperation could include the following activities of USArray facilities:

i) Providing the NEIC with source-dependent travel time corrections for TA and FA stations, based on the best available velocity models available when each station is deployed, and updating the corrections when the velocity model can be significantly improved.

ii) Routinely and promptly providing NEIC with relative arrival times measured by comparing waveforms at TA, FA and nearby stations for closely separated pairs of events that are recognized as part of routine quality analysis of the waveform data.

iii) Computing relative locations of earthquakes within clusters in support of fault-dynamics research, such as studies of the interaction between seismic events in which each event in a sequence affects stress near its neighbors, and thus acts to generate or inhibit future events.

iv) Searching waveforms for signals from events that are not otherwise recognized in routine quality analysis by comparison with waveform archives of the station when it recorded past events, and measuring relative arrival times by waveform comparison.

v) Compiling "ground-truth" information about explosions and other events with very accurately known absolute locations, especially within or near FA deployments, with the intention of using ground truth locations and relative locations of other events to compute very accurate absolute locations of the other events.

Activities (*i*) and (*ii*) will be undertaken with the expectation that NEIC could take advantage of USArray products to enhance ANSS products. Activities (*iii*), (*iv*) and (*v*) will result in highly specialized USArray products that will be of interest primarily USArray investigators, but essential for the work of that community. Activities (*i*), (*iii*) and (*iv*) are level 2 products, which might require short NSF-supported projects to establish reliable procedures that would subsequently be implemented as routine procedures at USArray facilities, even while the products continued to be reviewed by experts in the USArray community. Activity (*ii*) is discussed in more detail in the following section. Activity (*v*) depends partly on better identifying explosions, as discussed further below.

II.F.6. Priority B Products:

II.F.6.a. Phase picks

Phase picking is both a level 2 and level 3 activity strongly coupled to the earthquake location problem. There are currently available methods for picking phases, both manual and automatic, that are implemented at the NEIC and at the regional seismic networks. These produce level 2 products. There is a need, however, for more sophisticated, better automated methods such as those using waveform comparison techniques. A research effort is required to refine these methods for operational use, and to carefully monitor their early use. Issues related to archiving and distributing picks were discussed by WG1.

II.F.6.b. Conventional magnitudes and scalar seismic moment

We encourage the ANSS to continue calculation of conventional magnitudes (*e.g.* m_b, M_S, m_{bLg}) and consider this a top priority for inclusion in a USArray "event" object. For lower magnitude events these will be the only available characterizations of event size. For larger events where a moment estimate is available, overlap with conventional scales is necessary for regional and local tuning of the calculation. Furthermore, by investigating the difference between conventional scales and moment, studies can be performed relating to the source, path, and site effects. Conventional measures of magnitude also provide a tie to older data and are key parameters for source scaling, which becomes particularly important in less seismically active areas of the continent.

Scalar seismic moment, unlike empirical magnitudes, is a physical measure of event size. Scalar moment can be estimated for events that are too small for the complete moment tensor to be computed, and done so automatically, but procedures must be calibrated by comparison with scalar moment for events that have been studied in detail by comparison of observed and synthetic seismograms. Routinely computation of scalar seismic moment will therefore initially require first a principal investigator effort to establish calibration parameters. It will be important that the PI participate in subsequent evaluation of the automated calculations. Scalar moments should be distributed by adding them to existing catalogs, such as the ANSS catalog. Flexible Array deployments could be calibrated to compute scalar moment by comparison of results to events for which scalar moment is independently computed from the Transportable Array.

The station density of USArray should permit routine calculation of moment tensors down to M 3.5. If the threshold for moment tensor calculations could be lowered, the number of events for which scalar moments are needed would decrease.

II.F.6.c. Moment tensors

At present, moment tensors are routinely generated for global earthquakes of ~M5.5 and larger by the NEIC, Harvard, and other groups using data from the IRIS global network. The ANSS, Berkeley Seismic Lab (BSL) and others routinely produce moment tensors for events of ~M3.8 or larger which occur within specified regions of the United States using data from the ANSS and regional networks. Moment tensors for smaller events are sometimes reported, but the catalogs are not complete at those levels. Automated teleseismic moment tensor retrieval algorithms, while producing good estimates of seismic moment (and moment magnitude), sometimes lead to incorrect source orientations and focal depths, and seismologist review is needed to ensure reliability. Because rapid accurate estimates of earthquake size and focal depth are a vital tool for planning emergency response, it is important to improve the reliability of automated techniques. Establishment of USArray provides an opportunity to compare and evaluate the various algorithms being used to determine moment tensors for both teleseismic and regional events and to improve automation of procedures for obtaining this important information.

II.F.6.d. Finite fault models

Finite fault models characterize rupture propagation and total slip using a grid of patches on the rupture surface. These models can be computed automatically for many earthquakes within an array, although the threshold is generally larger than the thresholds for computing the moment tensor and integrated source parameters, such as directivity and rupture duration. The calculation of a finite fault model can be carried out automatically, as is currently done in northern California, but procedures for doing so for as many events as possible and for taking advantage of all useful transportable and flexible array data must be developed. Expert review of individual finite fault models will be necessary to ensure quality, at least for the immediate future.

Because finite fault models are comprised of a large number of parameter values, these models should not be included in catalogs, although some integrated parameters of the model, such as source time function and directivity, can be included. Instead, the complete models should be distributed as independent products that are linked to the catalog entry. The data distribution portal should include software to aid visualization, for example by displaying a video of the rupture process and an image of the final rupture distribution. Data from the flexible array will be especially useful for computing finite fault models for events that occur within a deployment because of the higher resolution possible with dense station spacing. However, it must be noted that strong motion instruments, which will be part of the ANSS but are not included in USArray, provide critical data for models of larger events.

The need for new model visualization tools is not unique to this product, and applies to products discussed by all groups.

II.F.6.e. Amplitude decay with distance and frequency

Knowledge of attenuation with distance is critical for a number of applications, from assigning M_L to prediction of peak ground acceleration for ANSS products such as ShakeMap. Routinely measuring amplitude in different frequency bands for a large subset of earthquakes will lead to a data set that can be used in a wide variety of applications.

II.F.6.f. Peak ground velocity (PGV) & peak ground acceleration (PGA)

Observations of PGA and PGV are critical for studies of strong ground shaking. As part of the ANSS, these parameters are routinely generated for use in products such as ShakeMap for earthquake response. Although these parameters are critical for ShakeMap as well as for studies of attenuation and other engineering seismology applications, USArray/EarthScope is not the optimal experiment for determining these parameters, both in terms of the spacing (70 kilometers is large for most applications of these observations) and the instrumentation (generally these observations are made on strong-motion time series since the broadband sensors will clip/respond non-linearly in the near source region). However, USArray/EarthScope data, when used in conjunction with ANSS data, will contribute to determination of these parameters at larger distances.

II.F.6.g. Response spectra (5% damped)

Response spectra are critical for engineering applications. As part of the ANSS, spectral acceleration is estimated at 0.3, 1.0, and 3.0 s period for use in ShakeMap. As with PGA and PGV, these products are typically derived from strong-motion records. The design of USArray is not likely to provide significant support for the production of response spectra.

II.F.6.h. ShakeMap

ShakeMap is and will continue to be a very important ANSS product. The USArray will contribute data to the production of the ShakeMaps where appropriate but, in general, the USArray experiment is not particularly suited to providing data for this product due to instrument spacing and type. Linking ShakeMap-type USGS/ANSS products to the observed continent-scale measurement of USArray, however, may be of particular interest in a research (e.g., attenuation) or education/outreach context.

II.F.6.i. Macroseismic effects (DYFI)

Non-seismic products like NEIC's "Did You Feel It" maps are important for outreach and community-involvement.

II.F.6.j. Event type classification

Seismic signals originate from many different sources – from earthquakes, nuclear explosions, chemical explosions of different types, mine collapses and mine "bumps", roackbursts, and from exotic sources such as bolides and sonic booms from space shuttles. While event identification may be used to exclude events other than earthquakes from seismic hazard estimates, structure studies also benefit from identification of man-made events, which can often be located very accurately. Source types can distinguished in many cases by comparing amplitudes of different regional phases in different frequency bands. The procedures to measure these amplitudes can be automated if a principal investigator effort is undertaken to determine the group velocity windows for each regional phase at stations of the transportable and flexible arrays. The phase amplitude measurements should be distributed by adding them to existing catalogs that include phase information, such as the ANSS catalog.

II.F.6.k. First-motion focal mechanism

Analysts of the California Integrated Seismic Network (CISN), which is a component of the ANSS, currently compute first-motion focal mechanisms for all events that are recorded with sufficiently good azimuthal coverage, typically down to a magnitude of 1.5 for events within the network. For earthquakes with magnitudes below approximately 3.8, currently the lower threshold for computing moment tensors with regional waveform data, these are the only indication of source mechanism that we have.

Hopefully, the denser station coverage provided by USArray will allow the computation of moment tensor solutions for smaller magnitude earthquakes, perhaps as low as 3.5. The remaining earthquakes, with sizes below this lower threshold, will be those with less reliable first-motion determinations that produce first-motion mechanisms of limited value. It should be noted, however, that algorithms for automatically computing first-motion focal mechanisms exist, and can be implemented at USArray data centers for little to no cost.

II.F.7. Priority C

II.F.7.a. Higher order moments and directivity

Higher-order moments can be computed for events that are too small to allow complete inversions for rupture. Defining procedures to estimate these parameters semi-automatically will require a significant research effort. In part, the procedures must be calibrated by comparing their results with complete models of rupture for events that are larger or were surrounded by many seismic stations at very close distances. The integrative measures of event properties typically are represented by one or a few parameter values, which should be distributed by adding them to an existing catalog, such as the ANSS catalog.

II.F.7.b. Moment rate functions

Source time functions are the raw material for more sophisticated derived quantities such as slip inversions, directivity measurements, etc. A catalog like the CMT catalog of the moment rate functions would provide a robust benchmark for seismologists to investigate and compare source physics.

II.F.7.c. Static & dynamic stress drop

Static and dynamic stress drop, while important, are not necessary for analysis of large events where moment tensors and finite fault models are available. We thus consider these parameters to be of secondary priority as a USArray data product. It is also unlikely that they will be included in the ANSS catalog. The value in these parameters is for comparison to large, previous events that are poorly recorded and for which there are a paucity of examples, particularly in the central and eastern U.S. In some cases (e.g. magnitude 3.5 to 5), it may be necessary to perform aftershock surveys to more accurately estimate stress drop.

II.F.7.d. Radiated energy

Radiated seismic energy from an earthquake is an important physical quantity that reflects the underlying dynamics of source rupture. Radiated energy may scale better with some types of seismic magnitude that with seismic moment. How radiated energy scales has implications for whether or not small and large earthquakes are fundamentally the same physical process. Widespread measurements of radiated energy on a large data set would be a step towards untangling this problem.

II.F.7.e. Rupture velocity and directivity

Most measurements of earthquake rupture velocities derived from seismograms are near the theoretical limit for dynamically propagating cracks and are difficult to replicate in the laboratory. This observational constraint on rupture propagation again can tell us something about how a small percentage of earthquakes grow from inconsequential small events to societally relevant damaging events. The variation of rupture velocity with magnitude, geology and tectonic regime will be possible with the new data set.

II.G. WG4: Earth Structure Data Products

II.G.1. WG members: M. Wysession, E. Hauksson, K. Dueker, A. Dziewonski, D. James, M. Fouch (recorder), B. Romanowicz, A. Meltzer, N. Shapiro, P. Maechling, J. Ritsema

II.G.2. Summary

Earth Structure data products include models for a variety of parameters to characterize Earth materials beneath the North American continent. Knowledge of these parameters is needed by the broader EarthScope science community to constrain geologic and geodynamic models and are therefore key to the success of USArray and of EarthScope as a whole. Development and dissemination of these products will benefit from collaboration and coordination between USArray and other Earth Science information technology programs such as CIG and GEON.

II.G.3. Table of Data Products on Earth Structure

In the following table we list level 2 and level 3 products that will deliver useful Earth structure information both to the seismological community and to the wider Earth Science community and interested public. This listing of products should not be viewed as exclusive. In particular, the list of level 2 products will evolve continuously. In many cases, products currently classified as level 3 may be classified as level 2 products as methodologies mature. Similarly, new products currently undefined in this document that become level 3 products may also eventually become level 2 products.

In the table and discussion below, several data products are listed under both levels 2 and 3. This reflects the fact that in a number of disciplines both routine and advanced methodologies exist for obtaining the desired product. The routine methods are grouped as level 2 whereas the advanced research level methodologies (still in development and not generally established) are classified as level 3.

Table 4. Table of Earth structure products

Source	*Product*	*Level 2*	*Level 3*
Priority A			
ESF/NSF[2,3]	Crustal thickness maps	X	X
ESF/NSF[2,3]	Tomographic Images	X	X
ESF/NSF[2,3]	Shear wave splitting maps	X	X
ESF[2]	Differential travel times (interstation)	X	X
ESF/NSF[2,3]	Surface wave phase and group velocity measurements	X	X
Priority B			
ESF/NSF[2,3]	Normal mode frequencies/mode splitting	X	X
ESF/NSF[2,3]	Waveform synthetics (1D and 3-D)	X	X
ESF/NSF[2,3]	Upper mantle discontinuity maps	X	X
ESF[2]	Differential travel times (intrastation)	X	X
NSF[3]	Crust and mantle attenuation		X
NSF[3]	Conductivity structure		X
NSF[3]	Wavefield imaging		X
NSF[3]	Detailed crustal structure (3-D)		X
NSF[3]	Mantle Anisotropy		X
NSF[3]	Mantle/core structure		X

[1] Produced by ANF or DMC (level 2)
(ANF = USArray Array Network Facility; DMC = IRIS Data Management Center)

[2] Produced by scientific community through EarthScope Facility (ESF) (level 2) or through a "satellite" facility residing in an academic research group but supported by EarthScope O&M funds.

[3] Produced by scientific community through standard research-level proposal (level 3) (funding from various sources, including NSF-EarthScope, NSF-geophysics, USGS-NEHRP.)

II.G.4. PRIORITY A.

II.G.4.a. Crustal thickness maps

Crustal thickness is a parameter of great interest to a wide range of geologists and geophysicists, both as primary information on which to base geologic interpretations and as a source of "noise" in the data that must be removed for studies focused on sub-crustal processes. Precise estimates of crustal thickness can be determined through several seismological data acquisition and analysis techniques, including large-aperture, active-source profiles and receiver function analysis of broadband data from earthquakes. First order estimates of crustal structure based on receiver functions will soon be a routine, level 2 process. Receiver function analysis also continues to be an area of active research and level 3 improvements in the methodology appear regularly in the scientific literature. The feasibility of incorporating these new methodologies into level 2 processing should be examined.

II.G.4.b. Tomographic images

Level 2: Tomographic images are among the most fundamental products that will be produced by USArray and of great interest to the broader Earth Science community and the public. We recognize that several different tomographic methods exist, using body waves and/or surface waves, waveforms and/or travel time picks. Some assessment will have to be made by a panel of experts to determine which methods are sufficiently "standard and routine" to be considered for Facility support. Results will also need to be evaluated by a scientist to ensure quality control before being released as part of the Level 2 database. *Level 3*: Tomography is an active area of research and new or improved techniques and methodologies appear regularly in the scientific literature.

II.G.4.c. Shear wave splitting

Level 2: Shear wave splitting measurements provide information on crust and mantle deformation. They will be produced as a level 2 data product using standard analysis techniques. In addition, initial evaluation of simplicity or complexity of splitting results should be included in the product. *Level 3*: Methods for obtaining shear wave splitting measurements continue to be the subject of active research and improvements and entirely new techniques are currently being explored. As they mature, the feasibility of incorporating these new methodologies into level 2 processing should be examined.

II.G.4.d. Differential travel times

Relative intra-station travel time measurements for P and S will be determined for all stations as input for velocity perturbation tomography computations. The inter-station measurements are typically performed by multi-channel cross-correlation of the same phase from the same event for stations in the array. These times should be made available as a product.

II.G.4.e. Surface wave phase and group velocity measurements

Surface waves provide information about the shear velocity structure of the crust and upper mantle and are particularly sensitive to temperature and composition. They are also sensitive to azimuthal anisotropy and possess unique sensitivity to radial anisotropy (i.e. anisotropy with a vertical axis of symmetry that can be detected by simultaneously inverting Raleigh and Love wave dispersion). These measurements are commonly obtained by first measuring phase and group velocity dispersion curves along specific source-station paths and then inverting the entire data set at a given frequency to obtain maps of phase and group velocity distribution over the region of interest. *Level 2*: group and phase velocity dispersion curves and associated error estimates for all possible source-station and interstation paths. It should be possible to agree on a standard method to invert these data for map distribution of phase and group velocity. *Level 3*: New array methodologies for measuring phase and group velocities in 2-D are active areas of research and appear regularly in the scientific literature. Improved inversion algorithms for the determination of phase and group velocity maps are also in this category.

II.G.5. PRIORITY B.

IIGF.5.a. Normal mode frequencies/mode splitting

Normal mode measurements are usually performed using globally distributed records from large earthquakes, and constrain estimates of properties averaged over the great circle path containing the source and the receiver. Although USArray data will not be collected over the entire globe, normal mode frequency shift measurements can be routinely made, at least for the fundamental mode branch, at USArray stations for earthquakes of M>7, and contributed to a global database, providing increased spatial sampling for great circle paths that cross North America. This should be coordinated with the GSN, since the global coverage is needed to obtain higher level products such as splitting functions. *Level 2:* fundamental mode frequency shifts for M>7 earthquakes. *Level 3:* Several methods to compute splitting functions of low frequency, high Q modes exist and are still under development. The products would be developed in combination with data from the Global Seismographic Network. Measurements of normal mode attenuation are still the subject of active research. The dense USArray sampling will provide an opportunity to investigate the short wavelength variability of normal mode amplitudes and the respective contribution of elastic versus anelastic effects.

II.G.5.b. Waveform synthetics (1-D and 3-D)

Synthetic seismograms should be produced for associated events recorded by the array (as defined by the ANF) in standard data formats as part of a FARM-type data product. In particular, 1-D synthetics obtained by normal mode summation can be routinely computed at periods above the microseismic peak, provided agreement is reached on the appropriate Earth model. Several standard methods exist for the computation of shorter period 1D seismograms (e.g. reflectivity, generalized rays). 3-D synthetics can be computed using a variety of techniques that are still at the cutting edge of development. This category is more controversial than those above, and will require significant expert input to determine the type of 3D synthetic seismograms to be produced. We anticipate close coordination with CIG in this effort.

II.G.5.c. Mantle discontinuities

Images of mantle discontinuities provide important insights into mantle processes. Receiver function analysis is a powerful tool for generating these images. Because the signals are generally weaker than for imaging the base of the crust, techniques to develop these images are still in a level 3 research phase. As they mature, the possibility of routine generation of these products (Level 2) should be periodically reexamined.

II.G.5.d. Crust and mantle attenuation

Models of attenuation in the crustal and mantle derived from absolute and differential body wave attenuation measurements as well as surface wave and normal mode measurements are important for understanding crustal heterogeneity, the composition of pore fluids, and other geodynamic processes. Such models are currently in a level 3 research phase.

II.G.5.e. Models of electrical resistivity

The next step beyond magnetotelluric transfer functions would be models of electrical resistivity. There are still significant variations on approaches to interpretation of MT profile data, and 3D interpretation of more widely spaced sites is not well developed. Thus any products beyond TFs will involve a significant research component. A goal of several MT groups is to develop some sort of gross continental scale model of crustal and upper mantle resistivity. This would be a valuable product for help in planning higher resolution MT surveys of specific targets, by allowing researchers to provide improved boundary conditions for modeling and inversion.

II.G.5.f. Differential travel times (intrastation)

Relative intra-station travel time measurements should be determined for significant events. These measurements are made between different phases recorded at the same station. An example of intra-station measurement is correlation of first-arriving P-waveforms with a later-arriving phase such as PP.

II.G.5.g. Wavefield imaging

The density of USArray recording will provide unprecedented opportunities for the development of 3D wavefield imaging techniques (as level 3 products), involving progress on both theoretical and data processing aspects.

II.G.5.h. 3-D Crustal and upper mantle structure.

This product will involve the integration and reconciliation of tomographic models and receiver functions analyses of Flexible Array results with results from active source crustal studies. Developing mechanisms to achieve this is a level 3 research effort and will depend on the application.

II.G.5.i. Mantle anisotropy

Multiple methods for obtaining mantle anisotropy exist, including shear wave splitting, surface wave measurements of polarization and azimuthal anisotropy, radial/transverse receiver function comparisons, azimuthal variations in Pn travel times, and waveform inversion combining surface waves and body waves. Combinations of these methods should be evaluated and potentially included as level 3 products.

II.G.5.j. Lower mantle/core structure

While the main focus of USArray is on investigation the structure of the crust and mantle immediately beneath the North American continent, the array will provide the opportunity to study deeper mantle and core structure and anisotropy in some parts of the Earth with unprecedented resolution. At level 2, travel times of core phases should be routinely measured (as part of the event characteristic data product). Measurements of travel times, splitting and amplitudes of Sdiff and ScS and possibly other phases that interact with the core/mantle boundary (CMB) and the inner core still involve development of methodology and should be considered to be at level 3, together with development of models of structure and anisotropy and fine-scale discontinuity structure in regions sampled by USArray data.

II.F.6. Level 4 Products (Knowledge Products)

While these integrated knowledge products were not a primary focus of this workshop, a number of expected EarthScope products depend on integration of level 2 and 3 USArray products with geologic data, laboratory information on rock properties and geodynamic models. Examples of several such higher level EarthScope products are listed here.

- ***3-D crust/mantle fabric***. *Combines all known measurements of seismic anisotropy and, where available, lab measurements on crust and mantle xenoliths for fabric strength.*
- ***3-D crust/mantle rheology (viscosity)***. *Based chiefly on seismic measurements of mantle Q and shear wave structure, possibly buttressed by gravity or rebound estimates.*
- ***3-D crust/mantle temperature and density***. *Derived from tomography (both velocity and Q), discontinuity topography, and gravity and guided by laboratory results from mineral physics.*
- ***3-D crust/mantle chemistry***. *Predicted from tomography (both velocity and Q), discontinuity structure, and guided by constraints from geochemistry and mineral physics.*
- ***History and evolution (4-D maps)***. *Time varying crust and mantle structure, guided by geology and plate reconstructions.*
- ***Stress maps***. *Based largely on earthquake focal mechanisms, and supplemented by borehole stress measurements, volcanic lineaments, and well breakouts.*
- ***Tectonic maps (rifts, plate boundaries, orogenic belts, volcanic belts...)***. *Compiled from an integration of all of the above.*
- ***Models of tectonic processes***. *Crafted to explain tectonic maps.*
- ***Paleoplate boundaries***.

III. Summary and Recommendations

A USArray data products workshop was held in Portland, OR on October 11-12, 2004. The objectives of this workshop were to specify and prioritize standard routine and higher-order USArray data products; to establish protocols and procedures for creating, reviewing and updating these products; and to propose a framework for supporting this work. Four working groups covering the following topics developed prioritized lists of products, which are discussed in this report:

- Waveform products and metadata
- Outreach and Education
- Event characterization
- Earth structure

Products were classified into the 5 levels established for EarthScope products. These range from raw data (Level 0) through analyses that integrate a wide range of geologic and geophysical data (Level 4). It is expected that Level 0-2 products, which include earthquake source parameters and lithospheric structure information that can be obtained routinely using well-established and semi-automated procedures, will be funded through contracts that specify what will be delivered and are supported as part of the USArray operations and management budget. Level 3 and level 4 products will continue to be funded through individual grants from NSF and other science programs. Groups discussions therefore focused on classifying products and defining an approach for distinguishing between Level 2 and Level 3 products.

It is critical to remember that this boundary is expected to be dynamic – as procedures mature, products will migrate from Level 3 to Level 2. It is also important to note that many Level 2 products will require continuing scientific oversight to ensure quality. **We recommend the following procedures and criteria for defining Level 2 data products:**

- An oversight group recommended by the IRIS Data Management Systems Standing Committee should be assembled to assess the qualifications of proposed level 2 products.
- Level 2 products should be required to adhere to a timely delivery schedule.
- Level 2 products should have a clear user application.
- Level 2 products should be thoroughly documented.
- Level 2 products should be reviewed annually. Renewal of contracts to produce level 2 products should not be automatic.

While some Level 2 products will be created at existing USArray data facilities (DMC, ANF), others should be generated at "satellite" facilities located within university-based research groups. **We recommend that the IRIS community work with NSF and EarthScope to establish mechanisms to support establishment of such "satellite" facilities.**

With a distributed system of facilities supplying both waveform data and products derived from the data, it is important that interfaces be established to simulate a "virtual" data management center, which will be managed through EarthScope and DMC portals. For example, all freely and openly available seismic data from networks operating in the US should be accessible through a single center even though the data may not all physically reside at the DMC. This includes data from permanent regional networks that are part of the ANSS as well as data from the permanent, transportable and flexible USArray networks**.** As USArray develops, current differences between the major priorities of IRIS and of the ANSS will become less distinct. It is essential that close collaboration between these different facilities is maintained in order to ensure optimal use of resources. **We recommend that "virtual data centers" be established through cooperation among the different agencies currently generating seismological data products.**

At present, the DMC distributes primarily waveform products of different types with limited station metadata. Examples of high priority additions to the Level 2 product base are arrival time picks of various phases, background noise parameters, earthquake locations, maps of crustal thickness and shear wave splitting, and tomographic models of the crust and upper mantle. New software tools will be needed for efficient visualization of these products. To facilitate use of this information for higher order research studies, new seismological and geodynamic modeling tools will also be needed. Efforts to develop these tools should be coordinated with related efforts by other Earth Science groups such as SCEC, GEON and CIG. **We recommend that a representative of USArray be included in ongoing and planned efforts to coordinate development of cyberinfrastructure for the Earth Science community.**

Although education and outreach products should be based on what professional scientists are using, resources must be allocated to facilitate specific education and outreach packaging, curricular development, and strategic distribution of USArray data products. These needs can best be met via the funding of educational experts working closely with IRIS, the EarthScope Office, individual researchers, and other facility and science specialists. **We recommend that USArray take full advantage of existing IRIS Education and Outreach efforts as well as partnerships with USGS and other organizations to optimize content and prevent redundancy of products designed for education and outreach.**

Appendix A: Organizing committee

Rick Aster	aster@dutchman.nmt.edu
Matt Fouch	fouch@asu.edu
David James	james@dtm.ciw.edu
Anne Meltzer	ameltzer@lehigh.edu
Stuart Sipkin	sipkin@usgs.gov
Anne Trehu (chair)	trehu@coas.oregonstate.edu

Appendix B: Other attendees

Tim Ahern	tim@iris.edu
Harley Benz	benz@usgs.gov
Larry Braile	braile@purdue.edu
Ken Dueker	dueker@uwyo.edu
Adam Dziewonski	dziewons@seismology.harvard.edu
Gary Egbert	egbert@coas.oregonstate.edu
Goran Ekstrom	ekstrom@seismology.harvard.edu
Mari Francissen	mari@iris.washington.edu
Lind Gee	lind@seismo.berkeley.edu
Peter Goldstein	peterg@llnl.gov
Egill Hauksson	hauksson@gps.caltech.edu
Christel Hennet	chennet@earthscope.org
Shane Ingate	shane@iris.edu
Phil Maechlin	maechlin@usc.edu
Linus Kamb	linus@iris.washington.edu
Tom Owens	owens@sc.edu
Gary Pavlis	pavlis@indiana.edu
Paul Richards	richards@ldeo.columbia.edu
Jeroen Ritsema	jeroen@gps.caltech.edu
Barbara Romanowicz	barbara@seismo.berkeley.edu
Dogan Seber	seber@sdsc.edu
Nikolai Shapiro	nshapiro@fignon.colorado.edu
David Simpson	simpson@iris.edu
John Taber	taber@iris.edu
Chad Trabant	chad@iris.washington.edu
Frank Vernon	flvernon@ucsd.edu
Raymond Willemann	ray@iris.edu
Mitch Withers	mwithers@memphis.edu
Michael Wysession	michael@wucore.wustl.edu

Appendix C: Agenda

USArray Data Products Workshop
Portland OR, October 11-12, 2004

Reminder: The objectives of this meeting are to specify and prioritize standard routine and higher-order USArray data products; to establish protocols and procedures for creating, reviewing and updating these products; and to propose a framework for supporting this work.

We plan to leave the meeting with a well-developed draft of a document that discusses these procedures and protocols in detail, for initial presentation to the IRIS Executive Committee on November 3. This document will be used as input to broader EarthScope discussions on integrated EarthScope products and to help make the case to NSF for expanded support of EarthScope-related research.

Our discussions should build on the concepts outlined in "The Need to Identify USArray Products" and should be defined within the larger framework of EarthScope data products (EarthScope Data and Sample Products for Science and Education), which defines characteristics of products levels 0-4. Both of these documents are attached.

October 10

8:00p-10:00p Informal meeting of Organizing Committee and Discussion Leaders

October 11

8:30a-9:00a Introduction: *(Trehu)*
- 8:30-8:35 Welcome and logistics (Anne Trehu)
- 8:35-8:45 The IRIS perspective (David Simpson
- 8:45-8:55 The need for seismological data products (Anne Meltzer)
- 8:55-9:00 Questions

9:00a-10:00a Data Management Framework: *(Meltzer)*
- 9:00-9:10 The EarthScope Data Portal (Christel Hennet)
- 9:10-9:30 USArray and the IRIS DMC (Tim Ahern)
- 9:30-9:45 The Advanced National Seismic System (Lind Gee)
- 9:45-10:00 Discussion

10:00a-10:15a Break

10:15a-11:35a Examples of Ongoing and Planned Higher Level Data Products *(Sipkin)*

10:15-10:35 The Harvard CMT experience (Goran Ekstrom)
10:35-10:55 Data products from SCEC (Phil Maechling)
10:55-11:15 Receiver Reference Models (Tom Owens)
11:15-11:35 The Array Network Facility (Frank Vernon)

11:35a-12:00a Introduction of Working Group charges and general discussion of meeting objectives *(Sipkin)*

12:00p-1:00p Lunch

1:00p-2:30p Plenary Session – Chair *(James)*

1:00-1:45 SHORT (5 minute max) presentations on a variety of topics.

Dogan Seber: GEON
Greg Anderson: PBO
Rick Aster: IRIS E&O
Michael Wysession: CIG
Harley Benz: NEIC
Gary Egbert: The EM component of USArray
Larry Braille: E&O
Paul Richards: the new Chinese seismic array
"open mike" session

1:45-2:30 Continue whole group discussion on Working Group charges and meeting objectives..

2:30p-2:45p Break

2:45p-6:00p Working Groups meet.

7:00p-8:00p Dinner

After dinner: Working Groups begin writing assignments.

October 12

8:30a-10:30a Compile write-ups from Working Groups

10:30a-10:45a Break

10:45a-12:30p Plenary session: Presentation of results from Working Groups *(Fouch)*

12:30p-1:30p Lunch

1:30p-3:00p Plenary session: Future products, funding frameworks, user-friendly products for the broader Earth Science community and E&O, etc. *(Aster)*

Formal end of workshop at 3 pm

3:00p + Work on report – Organizing Committee, Working Group Leaders, and any other volunteers,

6:30p-8:00p Working dinner of Organizing Committee to summarize where we are and what still needs to be done to have a draft of the report ready to distribute at the DMS-SC and EXECOM meetings in late October.

Appendix D: Draft charges to working groups (*Discussion Leader/Recorder*)

(note: These charges were modified somewhat at the workshop, and planned discussion groups 4a,b,c met as a single, larger group.)

1) **Waveform data products and station metadata:** What are the waveform data products of greatest interest to the community that can reasonably be accommodated by quasi-automatic processing? Will the scale or nature of USArray waveform products require modifications to DMC procedures? (e.g., SPYDER threshold magnitudes). What associated metadata or ancillary products (e.g., site noise characterization) need to be produced?

Gary Pavlis/Anne Trehu

2) **USArray data products for Education and Outreach:** Who are the essential E&O partners for IRIS/USArray? What are the foreseeable unique needs of the education and outreach community that should be addressed in data products? What are the audience groups? What portals and other communication strategies should be used for dissemination or E&O materials?

Larry Braile/Rick Aster

3) **Event characteristics:** This includes hypocenters and moment tensor solutions. At present, hypocenters and moment tensors are routinely generated for global earthquakes by the NEIC and its partners using data from the IRIS global network and for U.S. earthquakes by the ANSS. The partnership between USArray and ANSS presents a unique opportunity to rethink the way earthquake monitoring in the US is done. This will include addressing the following questions: How will the ANSS use USArray data to better fulfill its mission? How will USArray use ANSS products to better fulfill its mission? What products should be jointly produced? What worthwhile data products are not going to be produced by ANSS? Additional critical issues include location accuracy criteria, automatic vs. reviewed locations, magnitudes, and moment tensors, definition of authoritative solutions/catalogs, attribution, and public access portals.

Ray Willemann/Stuart Sipkin

4) **Earth structure:** This category potentially includes a wide range of possible products:
   - 4A - tomographic models of Vp, Vs and attenuation on a global, lithospheric, or crustal scale;
   - 4B - models of interfaces in the Earth derived from seismic waves scattered at these discontinuities;
   - 4C - analysis of seismic anisotropy.

Products in each of these three sub-groups will be the focus of separate discussion groups. At present, these are Level 3 or 4 products.

Questions to be addressed include:
What subset of these products can be standardized for automated "production"?
How should procedures be defined for these "production products"?
How should software be verified and maintained?
How should the products be distributed to the community?
Should/can we distribute interim steps in the data product processing sequence to the community?
For data products that would benefit from broader-scale databases, should USArray be responsible for collecting, parsing and distributing these additional data as part of a given data product?
Should USArray data products be produced and distributed for flexible array deployments?

4A Ken Dueker/David James

4B Tom Owens/Anne Meltzer

4C Barbara Romanowicz/Matthew Fouch

Appendix E: Summaries of introductory presentations

1. Data management framework:

1.a. The EarthScope Data Portal (Christel Hennet)

The promise of EarthScope is to take a multidisciplinary approach to studying the structure and evolution of the North American continent and the physical properties that control earthquakes and volcanoes. To be successful in this task requires making different data types and related information available to a broad range of scientists, educators, government agencies, the media, and the public. The EarthScope portal will provide seamless, single-point access to all EarthScope related data, data products, and tools and can be viewed as the most important legacy of the National Science Foundation's largest investment in solid-Earth Science, and a fundamental database for the next generation of Geoscientists. Efforts are now underway to define the structure of the portal, define general requirements for its content, explore how we can leverage already existing capabilities at the DMC and UNAVCO data centers, and work with groups such as IAGT and GEON to provide state of the art integrated access, visualization tools and capabilities.

1.b. USArray and the IRIS DMC (Tim Ahern, Mari Francissen, Linus Kamb, Chad Trabant)

The IRIS DMC will ingest most seismic data from USArray in real time. Data from the Transportable Array (TA) will normally flow simultaneously to the ANF and to the IRIS DMC. Data from the backbone (BB) will flow simultaneously to the USGS DCCs at NEIC and ASL as well as the IRIS DMC. Flexible array (FA) data will flow either in real time to the ANF and DMC simultaneously or to the AOF where it will be QA'd and forwarded to the DMC. The IRIS DMC is the only facility where all USArray data will be available.

We anticipate that data from the seismic components of PBO and SAFOD will be archived both at the Northern California Earthquake Data Center (NCEDC) and at the IRIS DMC. We estimate that there will be roughly 7 terabytes per year from USArray, 5 terabytes per year from PBO borehole seismic sensors, and possibly 2 terabytes per year from the SAFOD downhole seismic sensors.

USArray Data Products will be classified into 5 levels, as discussed in section I.B. The IRIS DMC is currently producing all anticipated Level 0 data products for the USArray data. This includes continuous waveforms in the BUD and event gathers in SPYDER®. The IRIS Quality Assurance Framework is producing quality assurance estimates for a variety of parameters on the raw waveforms including Power Spectral Density estimates of the noise at all USArray stations for which we receive data in real time. The IRIS DMC is also producing Level 1 data products in the form of quality assured waveforms in the Archive and event segmented products in the FARM.

Level 2 Data Products exist to some extent at the present time, but considerably more work is still required. We have the ability to produce record sections on demand but not automatically. We also produce maps showing station locations and enhanced metadata for stations. We still need to develop the ability to store and distribute event related information such as hypocenters, phase readings, moment tensors and similar information.

Several techniques currently support the access to USArray data by the research community interested in building products. Full copies of all waveforms can be supplied using SEEDlink, LISS, autoDRM or the Data Handling Interface. Those wishing to develop systems to generate products should have easy and simple access to all waveform data and related metadata. We are also in the process of wrapping our DHI services and presenting them as a Web service.

The DMC is focusing its current efforts towards the development of a distributed system that can manage USArray data products that are developed in a distributed environment as well as being archived in a system composed of multiple archives. We are considering a system where data products are wrapped in layers of XML that provide information about 1) the product specific XML, 2) generic product creation information, and 3) information needed by the Uniform Product Distribution System. By wrapping products with layers of XML we believe a flexible, distributed system can be developed. This approach will also make products available using Web Services. We recognize the need to work closely with the Advanced National Seismic System (ANSS) in this effort.

The long-term goal is to provide efficient and effective management of USArray data products that allows products to be created at a variety of nodes and yet managed in an efficient manner. We anticipate making USArray data products available either through the IRIS DMC as well as making them available through the EarthScope Data Portal.

1.c. The Advanced National Seismic System (Lind Gee)

The Advanced National Seismic System (ANSS) is an USGS-led plan to modernize the US seismic monitoring infrastructure. It was authorized at $170.3M in 2001, but only $16.6M has been appropriated over the last 4 years.

In these 4 years, the ANSS has primarily focused on instrumentation - from the installation of collaborative ANSS/USArray backbone stations to the deployment of reference strong motion stations in numerous urban communities. In the last year, however, there has been an increasing focus on the "system" and a number of working groups are addressing fundamental issues related to the performance standards, data archiving, and product development.

The ANSS is organized on a regional basis, with each region coordinating with neighboring regions and with NEIC to provide national monitoring capabilities. The ANSS is the authoritative source for rapid earthquake information and examples of ANSS products include waveforms, earthquake locations and magnitudes, phase and amplitude readings, ShakeMaps, source mechanisms (first motion and moment tensor), rupture models, aftershock probabilities, tectonic summaries, and earthquake posters. Not all regions currently have same capabilities - and while there is some standardization in software, it is not uniform across all regions.

While the ANSS is still working on issues related to data archiving, a composite earthquake catalog is available now. The catalog effort was started in 1995, with the goal of creating a master catalog from the ANSS contributing networks (then the CNSS - Council of the National Seismic System). The current catalog contains hypocenters and magnitudes, with preferred solutions selected on the basis of regional authority. As of the time of this meeting, the catalog spanned earthquakes from 1898 to present – with nearly 2 million events. The catalog is available in a Web-based search (http://quake.geo.berkeley.edu/anss/) - and the entire catalog may be downloaded via FTP. The Web-based interface also allows users to generate maps.

There are plans under discussion to improve the ANSS catalog, including adding other parametric data (phases, amplitudes, mechanisms, etc) and providing views of multiple hypocenters (not just the authoritative solution). The ANSS catalog is one obvious opportunity for collaboration with EarthScope. For example, including EarthScope data in ANSS products, incorporating EarthScope earthquake products in the ANSS catalog, and providing access to the ANSS catalog through multiple points such as the EarthScope data portal.

1.d. The Array Network Facility (Frank Vernon)

Additional information on the array network facility can be found at http://anf.ucsd.edu

2. Examples of Ongoing and Planned Higher Level Data Products:

2.a. The Harvard CMT Experience (Göran Ekström):

For more than 20 years, NSF has funded a research effort at Harvard University aimed at systematic analysis of global and regional seismicity. This research has resulted in a widely referenced catalog of earthquake moment tensors, the so-called Harvard CMT catalog. The CMT catalog contains more than 21,000 earthquakes, and approximately 100 solutions are now added each month. The results of the analysis are routinely published in Physics of the Earth and Planetary Interiors, and the catalog is distributed in many other ways as well, by IRIS, USGS, ISC, and on our own web site.

Three appreciated attributes of the CMT catalog are that it is (1) comprehensive, (2) continuous and, (3) homogeneous. These qualities are not easy to maintain in a research project, which usually strives for innovation and change. However, they enhance the value of our catalog to its users. Consideration of how these and related qualities can be ensured will be an important part in the development of new seismological data-product projects.

The CMT Project does not provide "extra" funds for the seismology research group at Harvard -- it is not a "cash cow". Our continued interest in devoting time and energy to the CMT project is motivated by two factors: (1) our research in seismic tomography and seismotectonics depends on the availability of CMT solutions, and (2) the CMT project generates opportunities of discovery, such as of anomalous classes of earthquakes and of unexpected patterns of seismicity. A continued interest on part of a PI in the quality of the routine analysis is an important component in a successful data-product project. The responsibility for generating a high-quality and useful product has to rest with a committed researcher.

2.b. Receiver Reference Models (Tom Owens):

An overview of a funded prototype EarthScope products proposal was presented. The project will lead to the development of standardized "Receiver Reference Models" (RRM), an approach analogous to Centroid Moment Tensors (CMT). The RRM could serve as a means of quickly identifying anomalous regions; as a starting point for more detailed analysis of USArray data, such as more rigorous imaging methods; and as an easily-accessible resource for other research geosciences who need to correlate their results with seismologically-determined Earth structure. By undertaking a historical data mining effort through application of the technique to the existing IRIS data archive for seismograph stations in North America, we can both validate the method by comparison with published studies and average models (such as CRUST2.0) as well as develop a database of historical data processed and available in a manner identical to future USArray stations. General aspects of the project that may be extensible to other products include:

1. An automated processing framework for USArray data that is continuously transmitted to the IRIS Data Management Center (DMC) as well as the mining of the existing DMC archive using a new application, SOD (Standing Order for Data).
2. Methods for integrating the results of seismological analysis with other geoscientific data through standards evolving in the GEON project.
3. A hierarchical internet-based product delivery system that to make USArray products available to users with interests and experiences ranging from research scientists to K-12 students and teachers.

2.c. The SCEC Experience With Providing Products to the Seismological Community and the Public (Phil Maechling)

The Southern California Earthquake Center (SCEC) is a consortium of research organizations that produces a wide range of geophysical data products relating to earthquakes in Southern California. SCEC data products span the complete spectrum of U.S. Array data levels from level 0 through level 4. SCEC researchers also represent important U.S. Array data product consumers. In addition, SCEC operates a Community Modeling Environment that provides a collection of geophysical modeling programs and access to high performance computational environment for performing geophysical simulations.

As part of the southern California Data Management Center for the California Integrated Seismic Network (CISN), the SCEC Data Center contributes Layer 0, 1, and 2 data products. These products include continuous and triggered broadband waveforms, strong motion data, station metadata, event peak amplitudes, phase picks, earthquake catalogs, alternative earthquake catalogs, and source descriptions.

USArray Layer 3 and 4 data products are also created by SCEC. Important SCEC data products in these categories include geological and geophysical models including the

SCEC Community Velocity Model, the SCEC Community Fault Model, the SCEC Community Crustal Motion Model, and the SCEC Community Block model. Because these products are oriented towards both to the research community and to the public, SCEC's process and experience producing these products may be extrapolated onto a national scale by USArray.

SCEC is also developing significant data processing and data management capabilities through its Information Technology Research Project the Community Modeling Environment (CME). This geosciences and IT collaboration provides a library of geophysical modeling codes and access to significant computing and data storage resources. Simulation-based data collections, such as collection of synthetic seismograms from earthquake simulations are an important data product produced by this Project. This environment for performing geophysical simulations is available to researchers within SCEC and may also be useful to U.S. Array researchers.

SCEC researchers are developing new scientific and computing approaches to performing Seismic Hazard Analysis. This SHA research requires a combination of observational data, geophysical models, high performance computing, and data management and analysis capabilities. This type of research, if performed on a regional or national scale, would carry U.S. Array research into socially relevant areas.

SCEC has a strong Education and Outreach product development effort. Outreach products include Earthquake Hazard Awareness print material, 3D visualization data products, and broadcast quality animations of earthquake information after significant earthquakes.

Appendix F. Summary of acronyms used in this report.

ANF – Advanced Network Facility
AOF – Advanced Operating Facility
ANSS – Advanced National Seismic System
CIG – Computational Infrastructure for Geodynamics
CMT – Centroid Moment Tensor
DLESE – Digital Library for Earth System Education
E&O – Education and Outreach
ESN – Educational Seismic Network
GSN – Global Seismographic Network
DMC – Data Management Center
DMS-SC – Data Management System Standing Committee
GEON – Cyberinfrastructure for the Geosciences
GMT – Generic Mapping Tool
IRIS – Incorporated Institutions for Seismology
MT - MagnetoTelluric
NEIC – National Earthquake Information Center
O&M – Operations and Maintenance
PBO – Plate Boundary Observatory
PSD – Power Spectral Density
SAFOD – San Andreas Fault
TF – Transfer Function
USGS – United States Geological Survey