行政院國家科學委員會專題研究計畫 期中進度報告

台灣西南海域大陸與海洋相互作用的特性--海陸岩圈交界與重力異常及磁力異常之小波譜值相關(2/3)

期中進度報告(精簡版)

計 畫 類 別 ：整合型
計 畫 編 號 ：NSC 95-2611-M-002-004-
執 行 期 間 ：95 年 08 月 01 日至 96 年 07 月 31 日
執 行 單 位 ：國立臺灣大學海洋研究所

計 畫 主 持 人 ：喬凌雲

報 告 附 件 ：國外研究心得報告

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 96 年 07 月 16 日

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 111, E12010, doi:10.1029/2006JE002725, 2006

Crustal magnetization equivalent source model of Mars constructed from a hierarchical multiresolution inversion of the Mars Global Surveyor data

Ling-Yun Chiao,[1] Ja-Ren Lin,[1] and Yuan-Cheng Gung[2]

Received 28 March 2006; revised 13 July 2006; accepted 21 July 2006; published 29 December 2006.

[1] Several magnetic field models of Mars have been constructed since the Mars Global Surveyor data became available. Three distinct schemes formulated through spherical harmonic functions, discrete equivalent dipoles, and the continuous magnetic field kernels have yielded results that are grossly compatible but with very different details. Models of internal potential function in terms of spherical harmonics tend to yield divergent high-degree Mauersberger-Lowes spectra, whereas crustal magnetization models exhibit flat but still significant spectra up to high degrees. To have a better fitting to the observed data seems to have dominated previous efforts that have yielded fine details with wavelengths shorter than the lateral track spacing. The variance-reduction versus model-variance tradeoff analysis is invoked in this study for the determination of the appropriate regularization. Taking advantage of the recently developed multiscale inversion, we are able to conservatively retain only the model components that are robustly constrained by the data rather than unilaterally pushing for a higher degree of fitting. With the variance reduction around 82%, we find that to reach a reasonably fair data fitting without high model variance, the high-degree power spectra of our preferred model exhibit an obvious decaying trend, implying that a lot of the short-wavelength energy embedded within established models is either not robustly resolvable or is of external origin or is simply reflecting the nonuniform distribution of sampling at short scales. The reason that models based on spherical harmonics have greater high-degree power is attributed to the spectral leakage due to the truncated representation.

Citation: Chiao, L.-Y., J.-R. Lin, and Y.-C. Gung (2006), Crustal magnetization equivalent source model of Mars constructed from a hierarchical multiresolution inversion of the Mars Global Surveyor data, *J. Geophys. Res.*, *111*, E12010, doi:10.1029/2006JE002725.

1. Introduction

[2] A magnificent magnetic field variation in the southern hemisphere of Mars has been discovered owing to the compilation of the Mars Global Surveyor (MGS) data [*Acuna et al.*, 1999]. There is, however, no significant field intensity observed for the northern lowlands although there are as many observations for the northern hemisphere as for the south. The significant magnetic signature demarcating an extensive part in the south has been attributed to ancient magnetization of the Martian crust [e.g., *Hood et al.*, 2005]. Consequently, there have been considerable efforts to construct models of the Martian crustal magnetic field, not only for delineating potential tectonic features but also to systematically summarize the robust information of the precious MGS data as completely as possible.

[1]Institute of Oceanography, National Taiwan University, Taipei, Taiwan.

[2]Department of Geosciences, National Taiwan University, Taipei, Taiwan.

0148-0227/06/2006JE002725$09.00

[3] The nature and the quality of the magnetic observations as well as features of the main phases of MGS have been documented previously [e.g., *Albee et al.*, 2001]. Essentially, the MGS satellite observed magnetic data consists in vectorial, three-component magnetic field observations at different altitudes, from below 200 km to 367–435 km, during different mission phases. The three-component data set we used in this study is the same set previously used to construct the spherical harmonics degrees 90 internal potential model [*Cain et al.*, 2003] as well as the spatially continuous magnetization model [*Whaler and Purucker*, 2005]. There are in total three-component measurements at 111,274 points, with altitudes from 102 to 426 km, composing the 333,822 field intensity data. The specific parameters and the adopted coordinate system of the data are described by *Cain et al.* [2003].

[4] The consensus established from previous works attributes the current major contributor of the Martian magnetic field to the lithospheric magnetization of a layer about 40 km thick, and that there is presently no dominant dipole field for the planet [e.g., *Voorhies et al.*, 2002; *Langlais et al.*, 2004; *Whaler and Purucker*, 2005]. One school of approach is to find the scalar internal potential

function at the surface of Mars in terms of spherical harmonics such that the gradient vectors of the model will fit the observations [*Connerney et al.*, 2001; *Arkani-Hamed*, 2001, 2002, 2004; *Cain et al.*, 2003]. Discussions about the effects of the variation of the attitudes and the lateral sampling have been raised. Interestingly, although assumptions on the internal origin of the magnetic field have been made in these studies, the divergent Mauersberger-Lowes power spectra [*Backus et al.*, 1996] toward high degrees, however, implies significant contributions from external sources. Other studies assume a continuously varying magnetization vector function $\mathbf{M(r)}$, where $\mathbf{r}$ stands for the position vector, such that the theoretical magnetic field intensity vector observed at $\mathbf{r_{obs}}$,

$$\mathbf{B}(\mathbf{r_{obs}}) = -\nabla_{\mathbf{r_{obs}}}\left\{\frac{\mu_0}{4\pi}\int_V \mathbf{M}(\mathbf{r}) \bullet \nabla_{\mathbf{r}}\left(\frac{1}{|\mathbf{r}-\mathbf{r_{obs}}|}\right)d^3\mathbf{r}\right\}, \quad (1)$$

fits the field observation. This linear data rule states that the data functional is of the form of an inner product between the model function and the data kernel. To evaluate expression (1) for a given magnetization model, a numerical scheme based on parameterizing the model function must be implemented. *Langlais et al.* [2004] use the equivalent source dipole technique that attributes the magnetic field to the contribution from 4840 dipoles with spatially varying magnetization intensity and direction, uniformly distributed across the globe and 20 km below the Martian surface. *Whaler and Purucker* [2005], on the other hand, expand the model function in terms of the data kernels. One advantage of this expansion is that it automatically avoids annihilators [*Parker*, 1994], since any component expressible in terms of the data kernels will not be orthogonal to all data kernels. That is, there will be no model components of this form that make no contribution to the data. One of the major disadvantages, however, is that the resulting Gram matrix is too sizeable and thus computationally demanding, although the matrix is usually sparse. *Whaler and Purucker* [2005] take advantage of the sparseness and indicate that an effective computation can usually be performed with only the 0.21% largest elements of the Gram matrix retained. Both these two studies obtain models that reveal power spectra similar to the former studies under degree 40. The higher-degree power spectra become considerably lower but are still significant. There have been concerns that crustal magnetic features with wavelengths shorter than the altitude of the observation might not be robustly resolvable [*Connerney et al.*, 2001; *Arkani-Hamed*, 2002]. Noticeably, since the north-south trending track spacing of the MGS has a width of $\sim 2°-5°$, that is, $\sim 100-300$ km at the equator [*Arkani-Hamed*, 2001], it has been argued that the highest harmonics degree corresponding to twice the lateral resolvable wavelength is thus about 65 [*Arkani-Hamed*, 2004]. In spite of these discussions, recent models tend to have significant power spectra contributions from much higher degrees.

2. Method

[5] We basically follow the approach that inverts for the spatial variation of the equivalent source crustal magnetization. We build the spherical tessellation initiated from a spherical icosahedron. Midpoints on the edges of each of the 20 spherical triangles are then connected to form 4 children triangles. The refinement of the spherical meshes is then executed successively until we have 10242 nodes marking the vertices of the 20480 (= 20 $\times$ 4^5) triangular faces. Summation of the integrand of equation (1) evaluated at finite Gaussian integration points [e.g., *Zienkiewicz and Taylor*, 1991] within each triangle is then computed to numerically approximate the inner product of the data rule. Let $\mathbf{m}$ be the vector with M (= 3 $\times$ 10242) magnetization model components, then the N (= 333,822) dimensional data vector $\mathbf{d}$ is constrained by

$$\mathbf{Gm} = \mathbf{d}. \quad (2)$$

[6] Notice that in the current formulation, the degrees of freedom of the model, 3 $\times$ 10242, is more than twice as much over the previous formulation based on the equivalent source dipoles, 3 $\times$ 4840 in the work of *Langlais et al.* [2004]. The parameterization of *Langlais et al.* [2004] assumes a finite amount of equivalent dipoles located on the vertices of the spherical triangular meshes. We, on the other hand, assume that the magnetization varies linearly within each of the 20480 triangles such that the magnetization is a globally continuous vector function. This further enables much better capability of resolving short-wavelength features. Elements of each row of the sensitivity matrix $\mathbf{G}$ specify the dependency of a particular datum upon the M dimensional model vector. An example of the spatial variations of selected observations reveals the localized constraints and the effects of the distinct altitudes (Figure 1). Conventionally, model estimates, $\hat{\mathbf{m}}$, can then be solved by the damped least squares (DLS) [e.g., *Lawson and Hanson*, 1974] algorithm,

$$\hat{\mathbf{m}} = \left(\mathbf{G}^T\mathbf{G}+\theta^2\mathbf{I}\right)^{-1}\mathbf{G}^T\mathbf{d}. \quad (3)$$

[7] The value of the nonnegative damping factor θ^2 controls the strictness of the imposed preference of the minimum model norm. It is also a knob for tuning the *variance reduction* (v_r) versus *model variance* (σ_m) tradeoff. Briefly, the variance reduction is defined to indicate the capability of a model ($\mathbf{m}$) to reconstruct the observed data ($\mathbf{d}$). It can be calculated by

$$v_r = \left(1 - \frac{||\mathbf{Gm}-\mathbf{d}||^2}{||\mathbf{d}||^2}\right) \times 100\%. \quad (4)$$

[8] On the other hand, the model variance is a measure of the uncertainty of a model manifested from noises contaminated to the data; it is computed [*Paige and Saunders*, 1982] by

$$\sigma_m = \sum_{l=1}^{M} s_l^2, \qquad s_l^2 = ||\mathbf{d}-\mathbf{Gm}||^2\sigma_{ll}, \quad \sigma_{ll} = diag\left[\left(\mathbf{G}^T\mathbf{G}+\theta^2\mathbf{I}\right)^{-1}\right]. \quad (5)$$

[9] It is noted that a heavier damping setup by a larger value of θ^2 usually leads to a robust model (lower σ_m), but

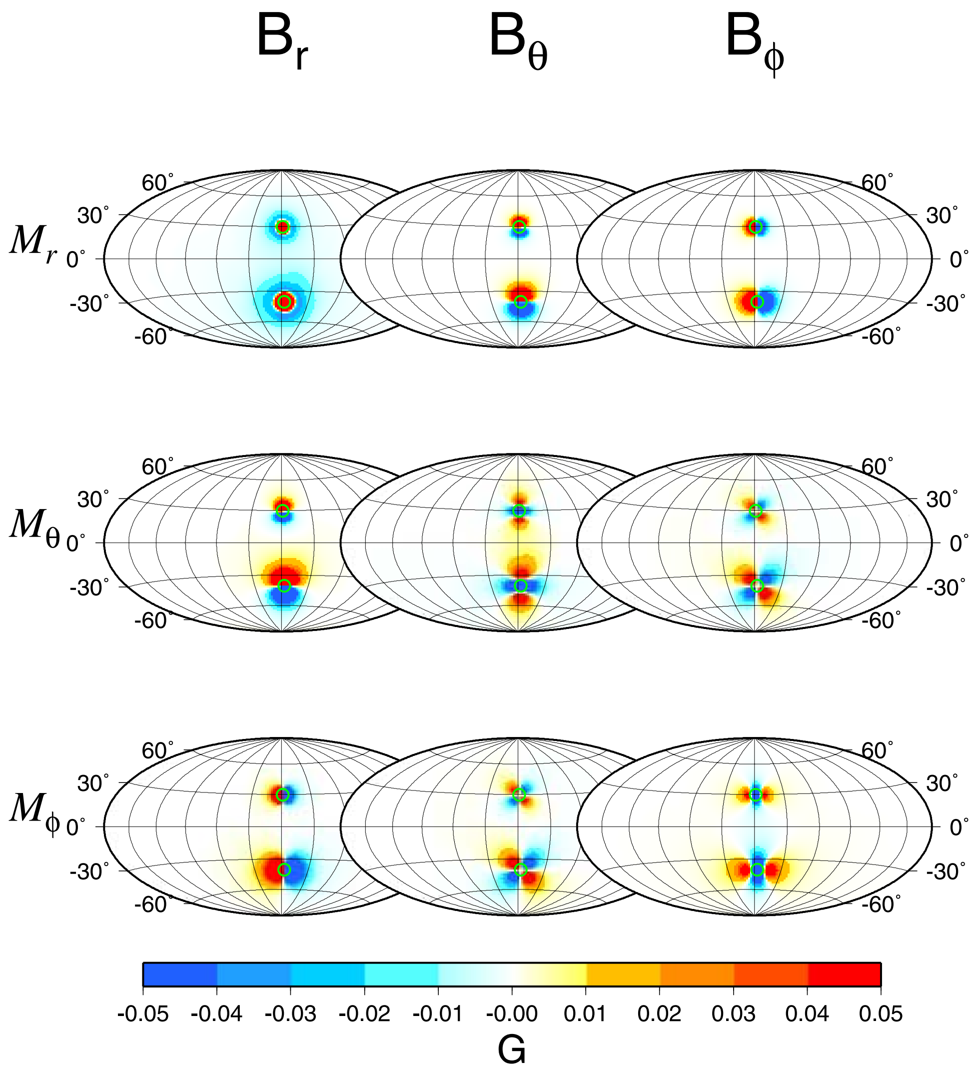

Figure 1. Spatial variation of the amplitude of the sensitivity matrix **G** (equation (2)). It can be visualized as the variation of the discrete data kernel function. In this figure a particular example is shown for the data observed at (180°E, 30°N) and an altitude of 370 km, as well as another southern hemisphere observation at (180°E, 45°S) (marked by small green open circles, respectively) but at a lower altitude of 130 km. Notice that for a lower-altitude observation, since the sources to observation distances are shorter relative to those of a higher-altitude observation, there will be a wider area such that the sensitivity of sources within it are above an effective threshold.

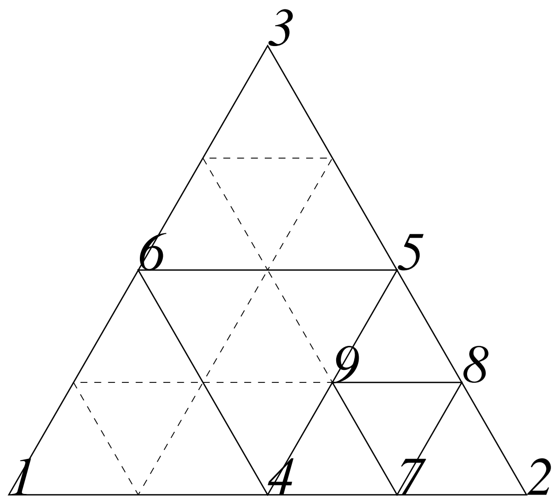

Figure 2. Triangular configuration as an example of multiresolution representation of a two-dimensional lateral variation (see the text and equation (6)).

sacrifices the data fitting (lower v_r) at the same time. We will show in the following how tradeoff between the model robustness and the data fitting helps to determine an appropriate value of θ^2 and an optimal model.

3. Multiscale Inversion Based on the Spherical Wavelet Basis

[10] It has been pointed out that minimum norm solutions obtained from DLS generally lack interpolation capabilities into sparsely sampled areas and tend to yield fragmented and fractured models [e.g., *Chiao and Liang*, 2003]. Regularizations based on enforcing model smoothness or roughness penalization have also been conventional practices in handling geophysical inverse problems [e.g., *Menke*, 1984; *Delprat-Jannaud and Lailly*, 1993]. The implementations, however, usually presume that the model smoothness [e.g., *Meyerholtz et al.*, 1989], or the intrinsic model correlation length [*Tarantola and Nercessian*, 1984], is spatially uniform or stationary. It has been shown that this is not a realistic presumption and has led to devices of multiscale regularization based on wavelet representations of models such that spatially nonstationary smoothness enhancement is automatically invoked depending on the in situ density of model constraints offered by the data [*Chiao and Kuo*, 2001; *Chiao and Liang*, 2003]. We follow the same rationale and transform the aforementioned spherical meshes into a stage to build spherical wavelet bases.

[11] To briefly summarize the algorithm, a simplified single triangle is taken as an example (Figure 2). To discretely describe a function $f(x)$ across the interior of the triangle, we can specify the spatial variation of f at uniformly distributed nodes, such as $f_1 = f(\mathbf{r}_1)$, $f_2 = f(\mathbf{r}_2)$, $f_3, \ldots\ldots$ where $\mathbf{r}_1, \mathbf{r}_2$ are position vectors at the internal nodes 1,2 (Figure 2). These nodes are vertexes of internal triangles through successive levels of refinement of the original triangle. That is, connecting midpoints on the edges, the parent triangle $\Delta 123$ is subdivided into four children triangles $\Delta 456$, $\Delta 536$, $\Delta 146$, $\Delta 425$ (Figure 2). Each of the resulting triangles can be further subdivided accordingly. Now instead of representing $f(x)$ by $[f_1, f_2, f_3, f_4 \ldots f_9 \ldots\ldots]$ distributed uniformly throughout the triangle, there are ways to build hierarchical representations of $f(x)$. A naïve example is cast in the following sense:

$$Level\_1: \quad h_1^1 = f_1, h_2^1 = f_2, h_3^1 = f_3$$

$$Level\_2: \quad h_1^2 = h_1^1, h_2^2 = h_2^1, h_3^2 = h_3^1, h_4^2 = f_4 - \frac{h_1^1 + h_2^1}{2},$$

$$h_5^2 = f_5 - \frac{h_2^1 + h_3^1}{2}, \ldots$$

$$Level\_3: \quad h_4^3 = h_4^2, h_2^3 = h_2^2, h_5^3 = h_5^2, h_7^3 = f_7 - \frac{h_4^2 + h_2^2}{2},$$

$$h_8^3 = f_8 - \frac{h_2^2 + h_5^2}{2}, \ldots\ldots$$

$$[h_1^3, h_2^3, h_3^3, h_4^3, h_5^3, \ldots\ldots] = \mathbf{W}([f_1, f_2, f_3, f_4 \ldots f_9 \ldots\ldots]). \quad (6)$$

[12] That is, on the fundamental level, *level_1*, there are 3 degrees of freedom $h_i^1 = f_i, i = 1,2,3$ to be specified where the upper index marks the refinement level and the lower indices are for the locations of nodes. On the next refinement, there are 6 degrees of freedom, $h_i^2, i = 1..6$. As specified in equation (6), the first 3 degrees of freedom that are used to characterize the large-scale variation are inherited from the lower level of representation whereas the additional 3 degrees of freedom are obtained by the in situ deviations of $f(x)$ from the expected values predicted by linearly interpolated from larger-scale variation at each midpoint, for example, $h_4^2 = f_4 - (h_1^1 + h_2^1)/2$. That is, the original in situ variations, $f_4 = (h_1^1 + h_2^1)/2 + (h_4^2)$, are replaced by the combination of a low-passed portion (the contribution interpolated from a larger scale) and a high-passed detail. Fast wavelet transforms [e.g., *Mallat*, 1998] are efficient schemes that accomplish the transformation **W** in equation (6) that maps the strictly spatial representation f_i to a localized hierarchy representation h_i^l of this sort. In addition, lifting schemes [*Sweldens*, 1996] can be incorporated to further improve the quality of the multiresolution representation. In this study, we transform the representation based on the original spherical mesh into an expansion utilizing spherical wavelet bases [see also *Chiao and Kuo*, 2001]. That is the reason why our construction subdivides the edges of each spherical triangles of the icosahedron by 2^5 segments instead of any integer as in the work of *Langlais et al.* [2004]. In fact, starting from the formulation (1) based on the direct spatial representation; we devise a bi-orthogonal wavelet transform [*Cohen et al.*, 1992] directly on each row of the coefficient matrix **G**, that is **GW***, such that the solution model vector to be solved for is now automatically the wavelet representation of the original spatial function for the crustal magnetization. That is, equation (2) is replaced by

$$(\mathbf{GW}^*)(\mathbf{Wm}) = \mathbf{d}. \quad (7)$$

[13] The new solution becomes, instead of equation (3),

$$\hat{\mathbf{m}} = \mathbf{W}^{-1}\hat{\boldsymbol{\mu}} = \mathbf{W}^{-1}\left[(\mathbf{GW}^*)^T(\mathbf{GW}^*) + \theta^2\mathbf{I}\right)^{-1}(\mathbf{GW}^*)^T\mathbf{d}, \quad (8)$$

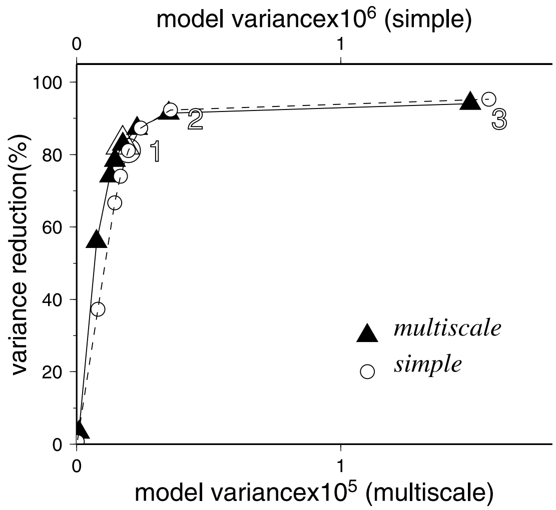

Figure 3. Curves displaying tradeoff between variance reduction of fitting (equation (4)) versus model variance (equation (5)) for different solutions. Notice that the scale of the model variance for simple damping solutions (marked by open circles and annotated on the upper horizontal axis) is almost an order of magnitude higher than the corresponding multiscale solutions (marked by solid triangles and annotations on the lower horizontal axis). The solutions marked as group 3 (*solution_3*, damping factor 10^{-4}) are apparently underdamped since the variance reductions are not much better than *solution_2* (damping factor 10^{-3}), but the model variances are considerably higher. On the other hand, *solution_1* (damping factor 3×10^{-3}) represents a relatively conservative but reliable solution without sacrificing too much variance reduction.

where $\mathbf{W}^{-1}$ is the inverse wavelet transform that reverses the operation of the forward wavelet transform $\mathbf{W}$. The advantage of solving for $\mathbf{m}$ in the wavelet domain is that with the same amount of degrees of freedom, parameters in the wavelet representation are grouped into a natural hierarchy of local scales such that the damping regularization acts to sort through successive scales depending on the local data constraints. In short, sites with dense constraints are capable of resolving more details robustly whereas robust long-wavelength components are still available for sparsely constrained area.

4. Results

[14] We execute two different groups of inversions based on the simple damping scheme (equation (3)) and the multiscale inversion (equation (8)), each with several different values for the damping factor θ^2. The variance-reduction (v_r) versus model-variance (σ_m) tradeoff curves (Figure 3) clearly indicate how an appropriate model might be selected. We first notice that with comparable variance reduction, the results obtained via the multiscale inversions (marked by solid triangles on Figure 3) have model variances that are in general an order of magnitude lower than the simple damping results (marked by open circles on Figure 3). As mentioned in the previous section, this is due to the way the model variation is assembled through the scales hierarchy from the longer wavelengths that have more accumulated constraints in the multiscale inversion. For both the simple damping results and the multiscale inversion results, high model variances are associated with the solutions that best fit the observational data (solutions marked by group 3 on Figure 3, that are located on the high-variance-reduction extreme on the tradeoff curves), implying that there are significant unreliable components poorly constrained by the data embedded in such solutions to reach high data fitting. In other words, these lightly damped solutions are overinterpreting the information content of the data. On the other hand, solutions approaching the knees of tradeoff curves (marked by group 2 on Figure 3) that exhibit almost similar variance reductions (over 92%) bear considerably lower model variances. Continuing the trend of decreasing the model variance, conservative solutions with variance reduction around 82% (solutions group 1 around the knee of the tradeoff curves) reduce the model variance even more. Further model variance decreasing (along the reversed horizontal axis on Figure 3), however, sacrifices too much variance reduction to gain just barely significant decreases of model variance, and is thus underfitting the precious observational data.

[15] For reasons discussed above, we believe that the appropriate solutions worth exploring that will reveal robust model structure without sacrificing significant amount of data information are located in between group 2 and group 1. In fact, we prefer the conservative group 1 multiscale inversion solution (referred as *solution_1* hereafter) that can be characterized as the most reliable model with a reasonably low data misfit. Simple damping group 2 solution (referred as *solution_2* hereafter), on the other hand, can be treated as a reference conventional model that might be a little bit on the overinterpreting side. The overall patterns of the crustal magnetization revealed in these two solutions are similar (Figures 4 and 5). In fact, the general features are quite similar to previous works such as those obtained by *Langlais et al.* [2004] and *Whaler and Purucker* [2005]. However, the conservative multiscale solution *solution_2* is dominated by long-wavelength structures at some places. Notice that this smoothing is not applied in a stationary sense, that is, the model is not the result of a uniform low-passing like in other conventional regularizations that enforce smoothness [*Chiao and Kuo*, 2001]. The relatively smooth model, *solution_1*, can fit the MGS data reasonably well (see also Figures 6, 7, and 8) although there are notable short-scale deviations from the observations. It is also worth pointing out that in Whaler and Purucker's model, to build the minimum RMS magnetization model, short-scale features are required to enforce null magnetization within data gaps. These short-scale features have very little effects on modifying the data misfit or to increase the variance reduction. Our solutions have considerably less and decaying high-degree power spectra but still retain reasonable data fitting. The reference simple damping solution, *solution_2*, has very similar Mauersberger-Lowes spectra up to degree 75 as compared to Whaler and Purucker's model. However, our preferred robust multiscale solution, *solution_1*, has similar power spectra to almost all previous models only up to degree 40 and then starts to dive. We will show in the next section, through inversion experiments executed on

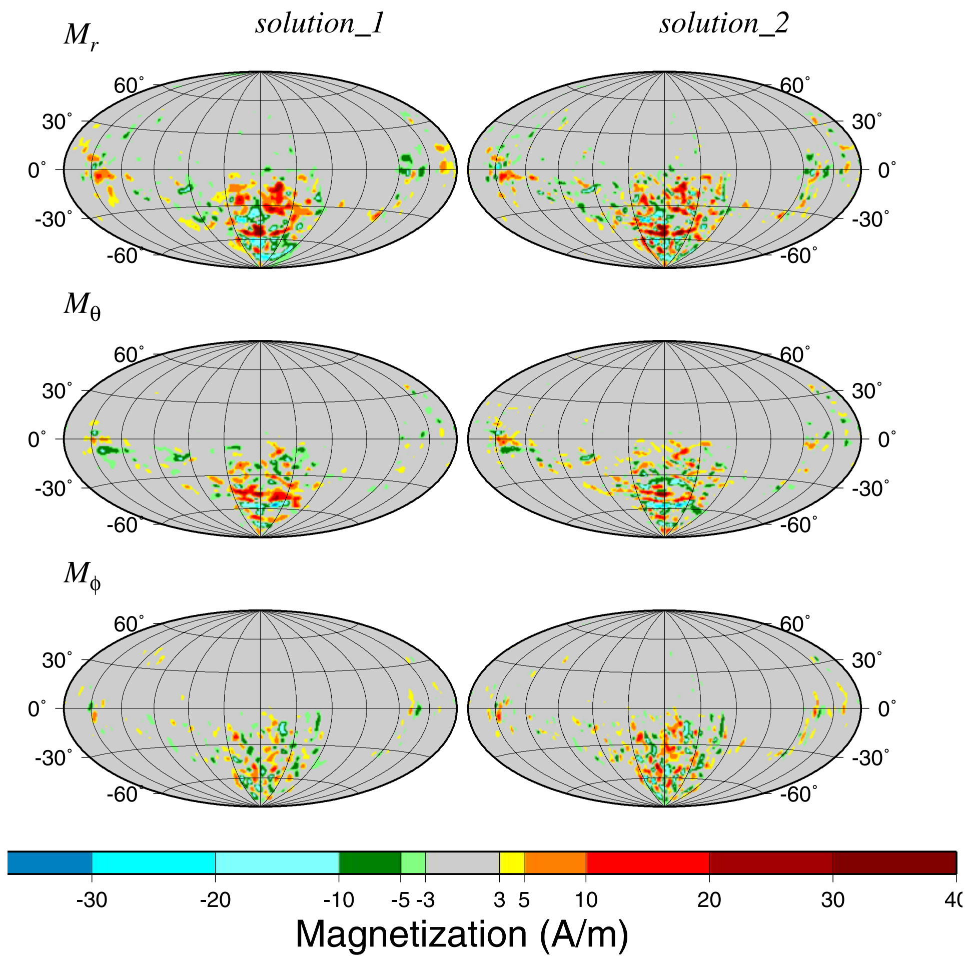

Figure 4. Crustal magnetization of Mars obtained from the multiscale inversion, *solution_1* (left column) and the conventional simple damping *solution_2* (right column). The top row is for the radial component, M_r, whereas the middle and lower rows are for lateral components, M_θ and M_ϕ.

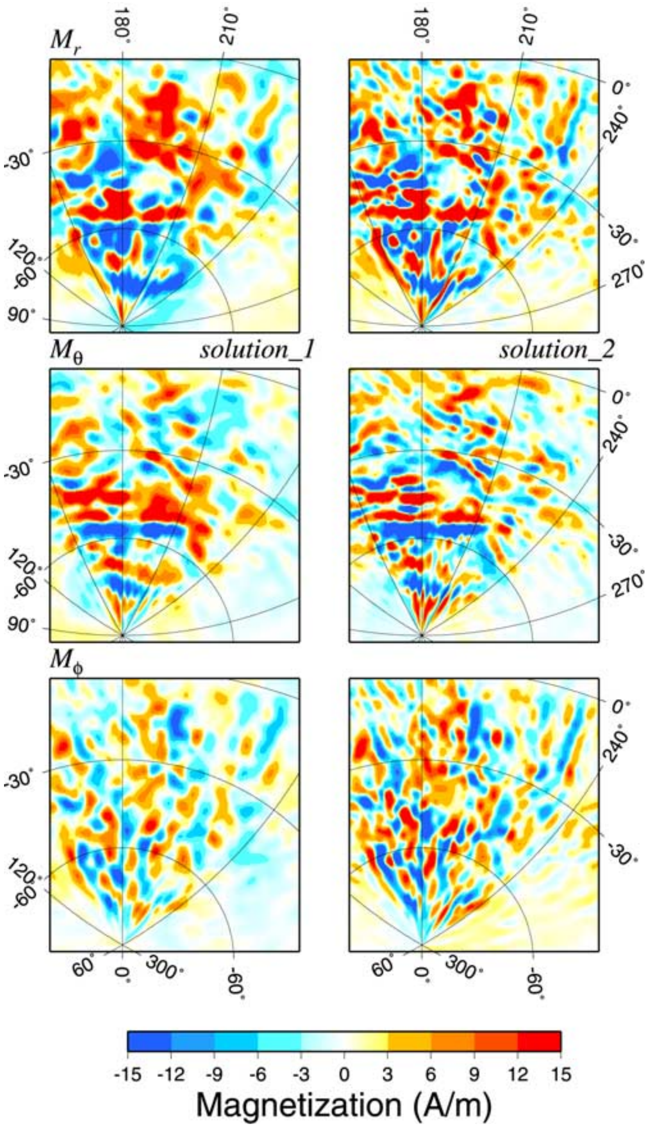

Figure 5. Comparison between different solutions showing the magnified portion of part of the southern hemisphere crustal magnetization zoomed in from Figure 4.

data generated from a synthetic model, why such a conservative choice to pick a reliable solution is important.

5. Discussions

[16] The sampling geometry of a particular data set such as the distribution of track spacing and the observing altitudes inevitably imposes natural limits on the shortest resolvable model wavelengths. Although the general consensus is to formulate inverse problems with enough model degrees of freedom to avoid the potential aliasing effect, the actual resolvable model components are intrinsically determined by the sampling geometry and are usually much less than those implied from the resolution presumed by the formulation. The variance reduction versus model variance tradeoff analysis helps to locate the optimal model resolution by offering the appropriate degree of strictness of regularization or damping. In principle, formulations based on data kernels [e.g., *Whaler and Purucker*, 2005] are intrinsically free from the concern of nonuniqueness since there will not be annihilators embedded. There are, however, always the problem associated with the noise contamination or observation errors. In other words, proper regularization is still essential to avoid overinterpreting the data. Unlike other previous works that pursue the best data fitting only, *Whaler and Purucker* [2005] as well as *Langlais et al.* [2004] invoke the minimization of the RMS magnetization to regularize the inverse problem. However, the model with the least data misfit still seems to be the choice for the preferred model (e.g., Table 2 of *Whaler and Purucker* [2005]).

[17] Our solutions that fit the data reasonably well have considerably less and decaying high-degree power spectra (Figure 9) although our spherical mesh, with a mean spacing of about 1.4°, is fully capable of resolving fine details beyond these higher degrees. These solutions are selected based on locating the optimal area around the knee of the variance-reduction versus model-variance tradeoff curve. That is, decaying high-degree power spectra is a consequence of having low model variance while retaining a reasonable data fitting. In other words, fine details corresponding to those high-degree power spectra are relatively less robustly constrained by the data.

[18] Notice that there are external as well as internal field contributions to the data. An inverse problem formulated following equation (1) results in an equivalent source magnetization model that extracts crustal signals as far as it is permissible. *Arkani-Hamed* [2004] used the radial component of the mapping phase data alone that are believed to be least contaminated by the external field, as well as covariance analysis and comparison between models derived from two subsampled data sets, to suppress the time-varying and noncrustal parts of the models induced by external field. Although he concluded conservatively that the degree ~62 is likely an optimum upper limit of the harmonic degrees of the crustal magnetic field that can be resolved by the high-altitude mapping phase MGS data, it is interesting to note that the resulting model, however, has high-degree power higher than *Cain et al.*'s [2003] model (Figure 9) that is based on a data set including all three components data from AB, SPO and MO phases. In other words, external field contaminations do not seem to be the main factors responsible for the differences of their high-degree power.

[19] We believe there are two major factors that result in the apparent discrepancies among models established so far. The first factor that differentiates results based on the Crustal magnetization Model (CM), might it be discrete in nature such as the GSFC model [*Langlais et al.*, 2004] or the continuous ones such as the WP model [*Whaler and Purucker*, 2005] and the model of this study, from those based directly on Spherical Harmonics (SH) can very likely be attributed to the effect of spectral leakage [*Trampert and Snieder*, 1996; *Chiao and Kuo*, 2001]. This effect is similar to the aliasing effect when truncated Fourier series is adopted to expand a function with high-degree energy. The high-degree energy that is not properly represented by their actual degrees owing to the truncated expansion will pile up near the truncated degree and distort the actual spectra especially close to the truncated degree. Instead of directly decomposing a function or a time series, the

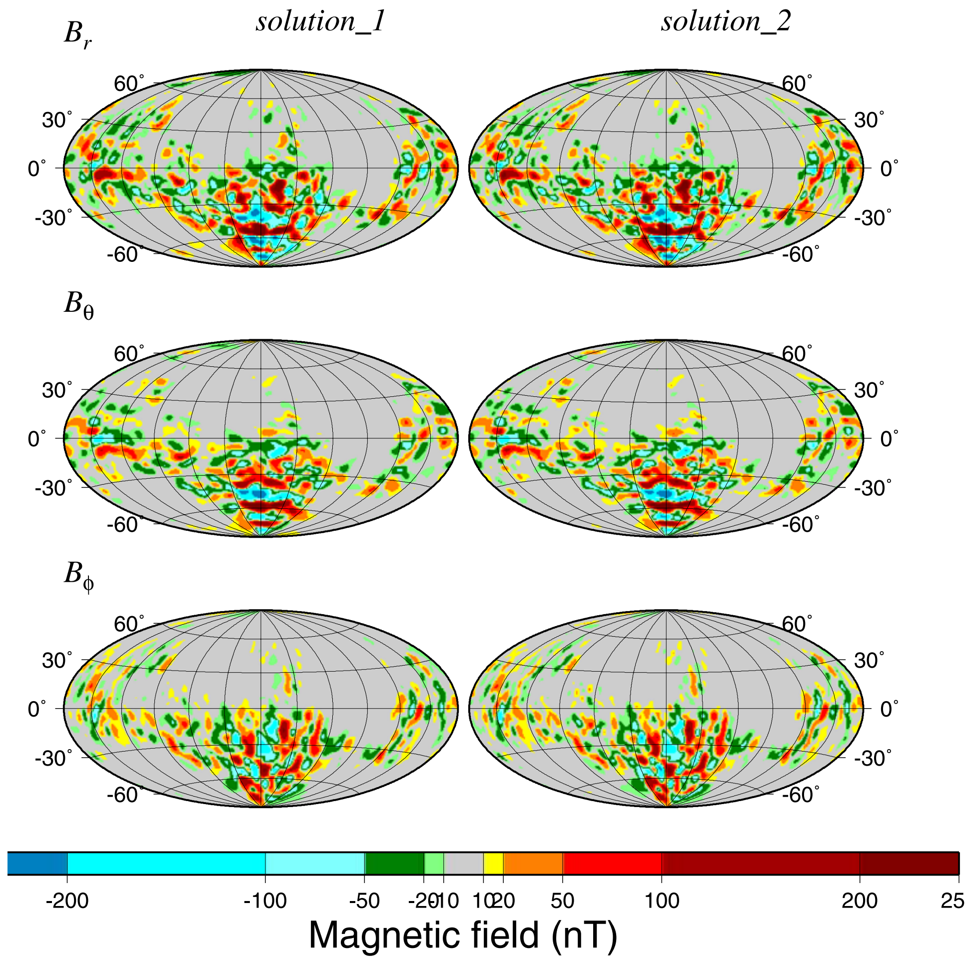

Figure 6. Comparison between different solutions similar to Figures 4 and 5 but showing the manifested magnetic field at the altitude of 200 km above Martian surface.

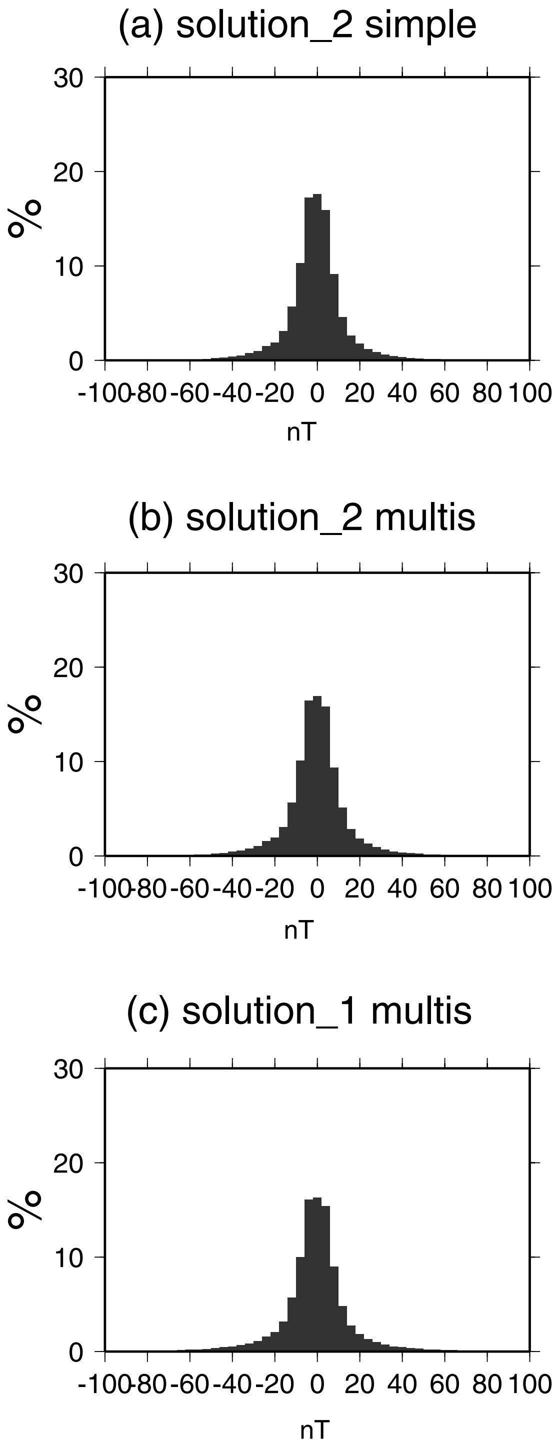

Figure 7. Histograms of fitting residuals of three solutions of this study. (a) *solution_2* using simple damping. (b) *solution_2* using multiscale inversion. (c) *solution_1* using multiscale inversion (see Figure 3 and the text).

spectral leakage is very similar in essence except that it occurs when less than enough degrees of freedom are adopted for a model parameterization of an inverse problem. It is obvious from this as well as other previous studies that it will take even higher degrees to get strictly numerically better fitting to the MGS data and that it is quite clear that degree n = 90 is just an arbitrary level of truncation. In other words, when a SH representation truncated at n = 90 is adopted to fit the MGS data, spectral leakage onto those high degrees close to the truncation degree will be inevitable. The CM models are however, truncated differently. In fact, one needs much higher degrees to completely represent these models in terms of the spherical harmonics expansion. That is, there are still significant power beyond n = 90 for CM models whereas SH models have their powers drastically annihilated reaching beyond n = 90. We believe that this is the main reason that makes the SH models have higher power around n = 90 than the CM models.

[20] The second factor that makes some models having lower high-degree power than others within the same group is regularization, the key issue that we have been discussing

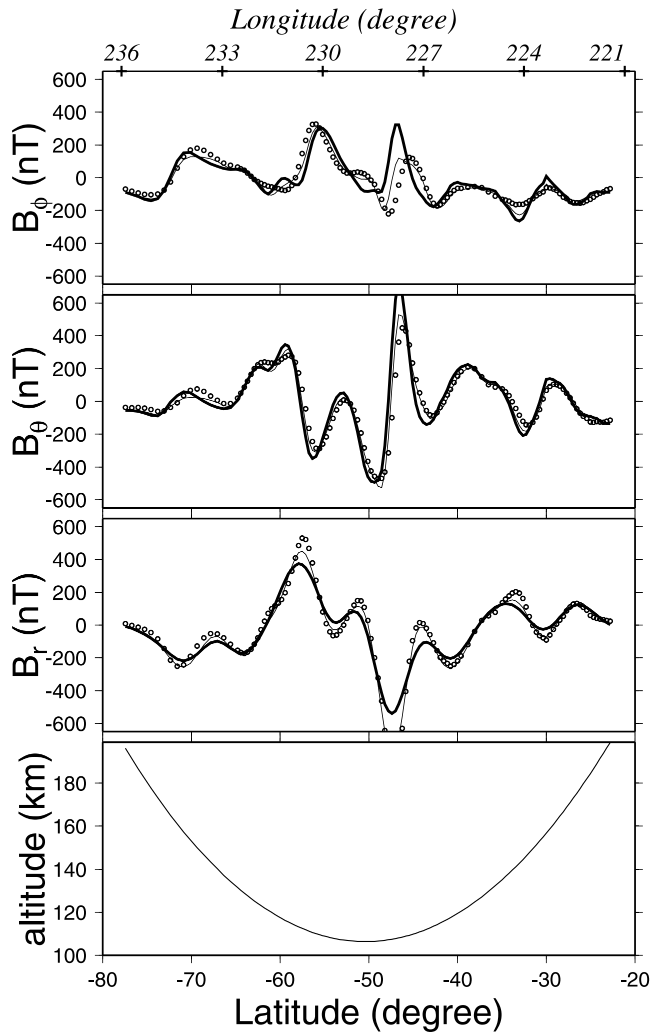

Figure 8. Comparison of calculated values of the three solutions of Figure 7 and the Mars Global Surveyor (MGS) observations for a segment of the AB2 low-altitude (shown in the lowest panel) collection period. Small open circles mark the MGS observations; thin line is the prediction from the *solution_2* model using simple damping, whereas the dark thick line is the *solution_1* using the multiscale inversion. All the variations are plotted as a function of the areodetic latitudes, but the longitudinal range is also shown on the top axis.

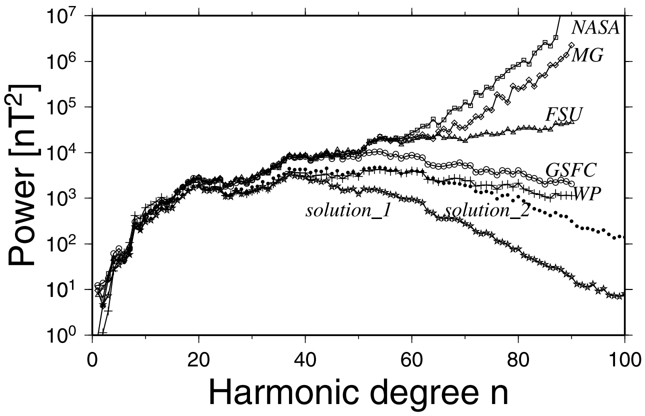

Figure 9. Comparison of the Mauersberger-Lowes power spectra at the Martian surface between selected previous models (following *Whaler and Purucker* [2005], we have NASA for the model of *Connerney et al.* [2001]; MG for the model of *Arkani-Hamed* [2004]; FSU for the model of *Cain et al.* [2003]; GSFC for the model of *Langlais et al.* [2004], and WP for the model of *Whaler and Purucker* [2005]) and the two solutions obtained in this study.

in the current study. To ensure that the solution is reliable, we suggest adjusting the strictness of the imposed regularization or damping. Enforced regularization shaves off poorly constrained model components while sacrificing some degree of goodness of fit. That is, we have reasons other than pursuing just better fitting to choose our preferred model that has high-degree power even lower than other CM models. On the other hand, we believe that the reason for the FSU model [*Cain et al.*, 2003] having much lower high-degree power than the MG model [*Arkani-Hamed*, 2004] and the NASA model [*Connerney et al.*, 2001] is that the FSU model is based on a more complete data set that reduces the degree of nonuniqueness of the inverse problem. That is, for the same amount of degrees of freedom to be modeled, more data constraints behave similarly as regularization that reduces relatively poorly constrained components and results in less high-degree power. Further comparison of spatial patterns of the surface potential among our preferred models and those established previously demonstrates the fundamental differences that might be results of the two factors mentioned above (Figure 10). Notice that the FSU model (Figure 10b) that is constrained with more data than the MG model (Figure 10a) appears to be much simpler along with much lower high-degree power (Figure 9). That is, it is very likely that a significant portion of those short-scale complexities in the MG model with high-degree power are not robustly resolvable model components. The WP model (Figure 10d) is in fact constrained by the same data set as the FSU model. So the reason why the WP model bears even less complicated structures than the FSU model is very likely due to its distinct formulation that avoids null space model components from scratch. It is interesting to note that our solution_2 model (Figure 10e) is very similar to the WP model. Whereas our solution_2 model reaches a variance reduction over 92%, the intrinsic model structure is much simpler than those previously established SH models (Figures 10a and 10b). Furthermore, we have reasons to believe that the even simpler structures in the conservative solution_1 model (Figure 10f) might be more robust. It is also worth mentioning that although the difference between the surface potential models from solution_2 and solution_1 seems to be subtle (Figures 10e and 10f), their manifestations on the crustal magnetization models are in fact significant (Figures 4 and 5).

[21] To further verify the interpretation of the finer detail features discussed above, we execute recovery experiments with a known implanted synthetic magnetization model. A circular crustal model with constant 20 km depth and alternating positive and negative magnetization in the radial component, $\mathbf{M}_r$, is implanted around the equator (Figure 11a). There are no assumed lateral, $\mathbf{M}_\theta$ and $\mathbf{M}_\phi$, components. The same sampling geometry of the MGS data set is invoked as the observations. That is, the three-component magnetic field intensity data observed at

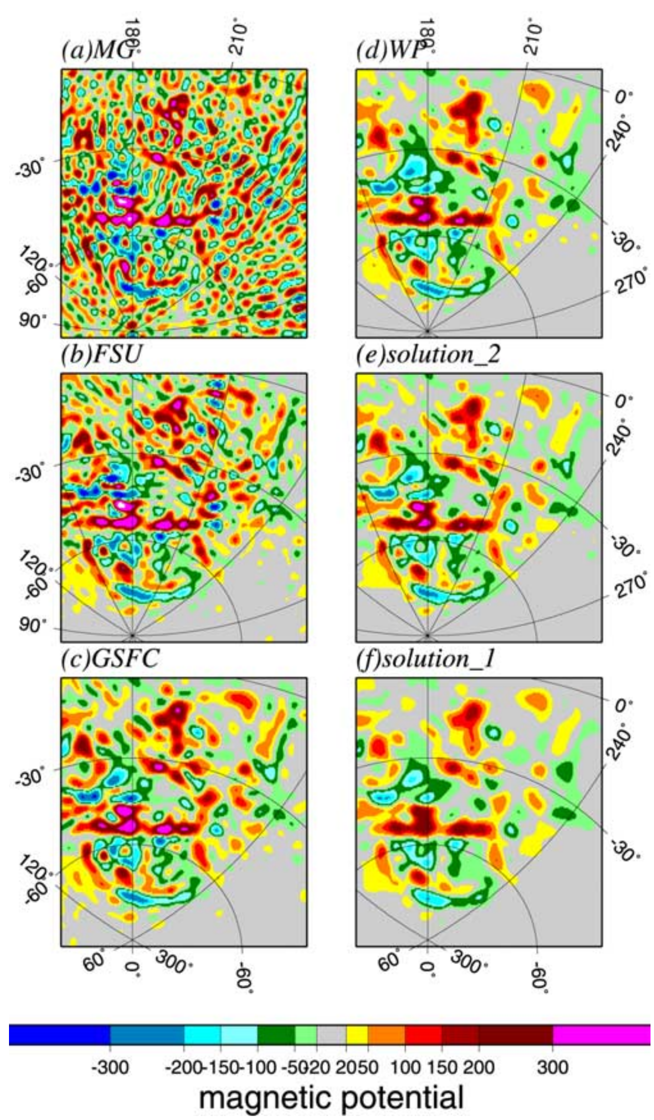

Figure 10. Comparison of the magnetic potential evaluated at the Martian surface among selected previous models and the two solutions obtained in this study: (a) MG, (b) FSU, (c) GSFC, (d) WP, (e) simple damping solution_2 model, and (f) solution_1 model obtained from multiscale regularization (see also the caption of Figure 9).

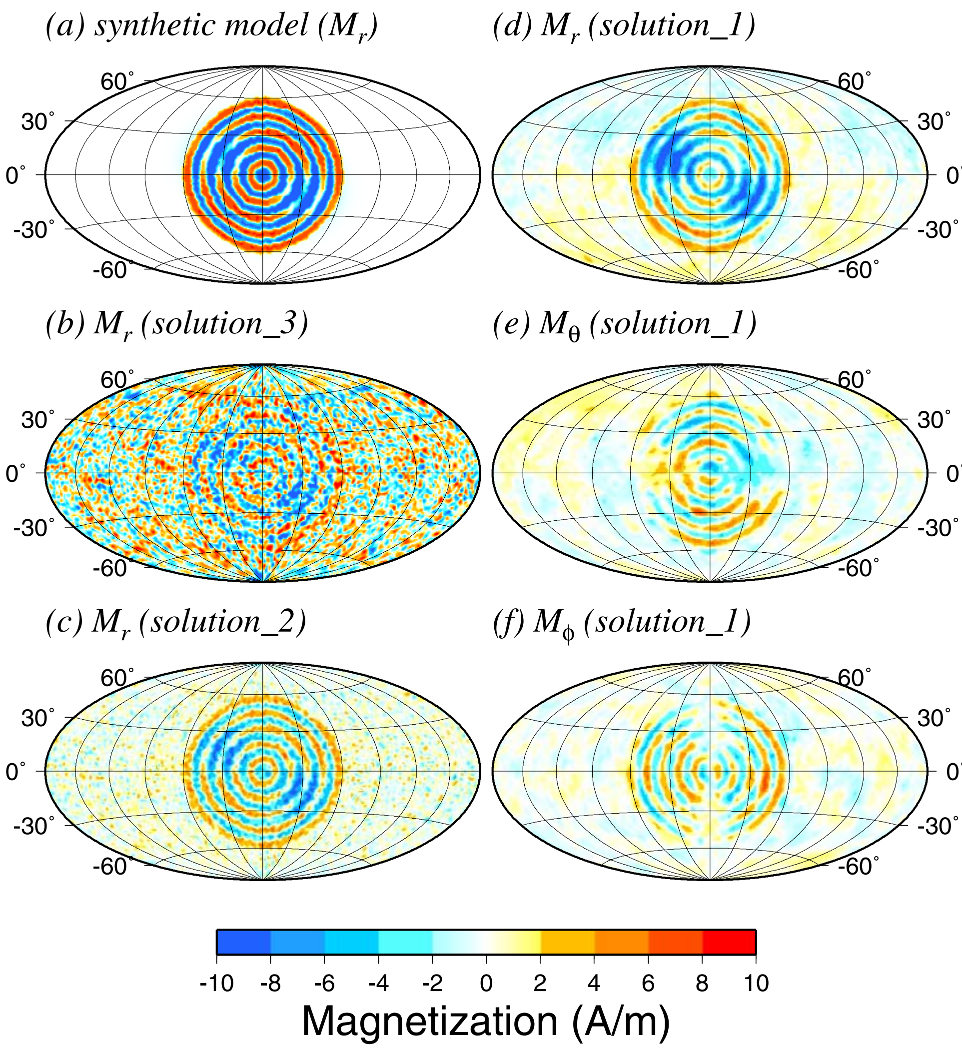

Figure 11. Inversion experiments executed using the same observational sampling geometry but with field intensity data generated from synthetic model and corrupted with noise with *N/S* about 60%. (a) The implanted synthetic magnetization model with alternating positive and negative magnetization in the radial direction centered at the equator overlaid upon two negative circular sources to the southeast and northwest quadrant. Notice that there are only implanted radial, $\mathbf{M}_r$, component. (b) Inverted result if *solution_3*, the simple damping solution of group 3 on the tradeoff curve (Figure 3), is selected. Notice the obvious corruption of the inverted model arisen from the contamination embedded within the data. (c) Similar to Figure 11b, except that it is a solution approaching *solution_2* instead of *solution_3*. Notice the improvement on reducing the corruption of uncorrelated, nonphysical structure. (d) In the preferred multiscale *solution_1* model, the effect of the multiscale regularization to annihilate unreliable model components while grouping correlated model structure is obvious. (e, f) Notice that the aliasing onto the $\mathbf{M}_\theta$ and the $\mathbf{M}_\phi$ components is inevitable (check the sensitivity matrix $\mathbf{G}$ in Figure 1).

different altitudes across Mars are replaced by synthetic data generated from the implanted magnetization structure. A considerable amount of uncorrelated noise with peak amplitude as high as 60% of the peak amplitude of the model generated data is then randomly blended in the synthetic data. Since the sampling geometry is the same, there is no need to carry out a new tradeoff analysis for the synthetic data set. Damping factors for the three groups of solutions marked on the tradeoff curves on Figure 3 are tested to obtain corresponding solutions. Not surprisingly, the recovered, underdamped simple damping solution (*solution_3* on the tradeoff diagram shown by Figure 3) is significantly corrupted by manifestation from the uncorrelated noise added to the data (Figure 11b). The corruption reduces considerably toward *solution_2*, but it is still significant and interferes with the correct interpretation of the recovered model (Figure 11c). On the other hand, the noise corruption upon the recovered model that corresponds to the multiscale inversion *solution_1* is obviously much lower (Figure 11d) and reasonable. What is worth cautioning is the significant aliasing effects onto the $\mathbf{M}_\theta$ and the $\mathbf{M}_\phi$ components that are not implanted (Figures 11e and 11f). This is, however, inevitable for any formulation based on equation (1) and is simply unresolvable by data constraints alone.

[22] In summary, the reason to carry out tradeoff analysis is to serve as an effective way of picking the right degree of regularization and thus the appropriate model components that are robustly constrained by the data. The quality of the actual MGS data is probably much better than the tested synthetic data such that the potential corruption might not be as serious as what is demonstrated in Figure 11. However, overinterpretation or overfitting data with unreliable model components is prone to misleading results that can be avoided by giving up a small fraction of the relatively less reliable data information.

[23] **Acknowledgments.** We wish to acknowledge K. A. Whaler and M. E. Purucker for unselfishly sharing all supporting material of their model on the Web: http://planetary-mag.net/jgr_mars_whaler/ [*Whaler and Purucker*, 2005]. Constructive comments from Joseph Cain and anonymous reviewers have been more than helpful in making significant improvements. All graphs have been created using the Generic Mapping Tools package [*Wessel and Smith*, 1991]. This study is supported by the National Science Council of ROC under the contracts NSC 95-2611-M-002-004 and NSC 94-2116-M-002-023. See the auxiliary material[1] for the computer files used.

References

Acuna, M. H., et al. (1999), Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment, *Science*, *284*, 790–793.

Albee, A. L., F. Palluconi, and T. Thorpe (2001), Overview of the Mars Global Surveyor Mission, *J. Geophys. Res.*, *106*, 23,291–23,316.

Arkani-Hamed, J. (2001), A 50-degree spherical harmonic model of the magnetic field of Mars, *J. Geophys. Res.*, *106*, 23,197–23,208.

Arkani-Hamed, J. (2002), An improved 50-degree spherical harmonic model of the magnetic field of Mars derived from high-altitude and low-altitude data, *J. Geophys. Res.*, *107*(E10), 5083, doi:10.1029/2001JE001835.

Arkani-Hamed, J. (2004), A coherent model of the crustal magnetic field of Mars, *J. Geophys. Res.*, *109*, E09005, doi:10.1029/2004JE002265.

Backus, G., R. Parker, and C. Constable (1996), *Foundations of Geomagnetism*, 369 pp., Cambridge Univ. Press, New York.

Cain, J. C., B. B. Ferguson, and D. Mozzoni (2003), An n = 90 internal potential function of the Martian crustal magnetic field, *J. Geophys. Res.*, *108*(E2), 5008, doi:10.1029/2000JE001487.

Chiao, L.-Y., and B.-Y. Kuo (2001), Multiscale seismic tomography, *Geophys. J. Int.*, *145*, 517–527.

Chiao, L.-Y., and W.-T. Liang (2003), Multiresolution parameterization for geophysical inverse problems, *Geophysics*, *68*(1), 1–11.

Cohen, A., I. Daubechies, and J.-C. Feauveau (1992), Biorthogonal bases of compactly supported wavelets, *Commun. Pure Appl. Math.*, *45*, 485–560.

Connerney, J. E. P., M. H. Acuna, P. J. Wasilewski, G. Kletetschka, N. F. Ness, H. Reme, R. P. Lin, and D. L. Mitchell (2001), The global magnetic field of Mars and implications for crustal evolution, *Geophys. Res. Lett.*, *28*, 4015–4018.

Delprat-Jannaud, F., and P. Lailly (1993), Ill-posed and well-posed formulations of the reflection travel time tomography problem, *J. Geophys. Res.*, *98*, 6589–6605.

Hood, L. L., C. N. Young, N. C. Richmond, and K. P. Harrison (2005), Modeling of major martian magnetic anomalies: Further evidence for polar reorientations during the Noachian, *Icarus*, *177*, 144–173, doi:10.1016/j.icarus.2005.02.008.

Langlais, B., M. E. Purucker, and M. Mandea (2004), Crustal magnetic field of Mars, *J. Geophys. Res.*, *109*, E02008, doi:10.1029/2003JE002048.

Lawson, C. L., and R. J. Hanson (1974), *Solving Least Squares Problems*, 340 pp., Prentice-Hall, Upper Saddle River, N. J.

Mallat, S. (1998), *A Wavelet Tour of Signal Processing,* 577 pp., Elsevier, New York.

Menke, W. (1984), *Geophysical Data Analysis: Discrete Inverse Theory*, 260 pp., Elsevier, New York.

Meyerholtz, K. A., G. L. Pavlis, and S. A. Szpakowski (1989), Convolutional quelling in seismic tomography, *Geophysics*, *54*, 570–580.

Paige, C. C., and M. A. Saunders (1982), LSQR: An algorithm for sparse linear equations and sparse least squares, *Trans. Math. Software*, *8*, 43–71.

Parker, R. L. (1994), *Geophysical Inverse Theory*, 386 pp., Princeton Univ. Press, Princeton, N. J.

Sweldens, W. (1996), The lifting scheme: A custom-design construction of biorthogonal wavelets, *Appl. Comput. Harmon. Anal.*, *3*, 186–200.

Tarantola, A., and A. Nercessian (1984), Three-dimensional inversion without blocks, *Geophys. J.R. Astron. Soc.*, *76*, 299–306.

Trampert, J., and R. Snieder (1996), Model estimations biased by truncated expansions: Possible artifacts in seismic tomography, *Science*, *271*, 1257–1260.

Voorhies, C. V., T. J. Sabaka, and M. E. Purucker (2002), On magnetic spectra of Earth and Mars, *J. Geophys. Res.*, *107*(E6), 5034, doi:10.1029/2001JE001534.

Wessel, J. K., and H. F. Smith (1991), Free software helps map and display data, *Eos Trans. AGU*, *72*(41), 445–446, 441.

Whaler, K. A., and M. E. Purucker (2005), A spatially continuous magnetization model for Mars, *J. Geophys. Res.*, *110*, E09001, doi:10.1029/2004JE002393.

Zienkiewicz, O. C., and R. L. Taylor (1991), *The Finite Element Method*, 4th ed., 807 pp., McGraw-Hill, New York.

L.-Y. Chiao and J.-R. Lin, Institute of Oceanography, National Taiwan University, P.O. Box 23-13, Taipei 106, Taiwan. (chiao@ntu.edu.tw)

Y.-C. Gung, Department of Geosciences, National Taiwan University, Taipei 106, Taiwan.

[1]Auxiliary materials are available in the HTML. doi:10.1029/2006JE002725.

赴國外研究心得報告

計畫編號	95-2622-M-002-004
計畫名稱	台灣西南海域大陸與海洋相互作用的特性-海陸岩圈交界與重力異常及磁力異常之小波譜值相關(2/3)
出國人員姓名 服務機關及職稱	喬凌雲 (Chiao, Ling-Yun) 臺灣大學理學院海洋研究所教授
出國時間地點	May 9-11 , 2007 ; Hotel Birke, Kiel, Germany
國外研究機構	International Data Exchange Workshop 2007 :Building a Global Data Network for Studies of Earth Processes at the World's Plate Boundaries MARGINS, NSF (美國國家科學基金會大陸邊緣計劃辦公室)

工作記要：

本計劃核定之出國差旅經費原訂參加 2006AGU. 雖然以共同作者提出論文三篇(見附件一);並擬以通訊作者口頭發表一篇。後接獲美國 MARGINS, NSF 邀請參加 International Data Exchange Workshop 2007 :Building a Global Data Network for Studies of Earth Processes at the World's Plate Boundarie. 由於該工作小組將研商各國學界研究用海域板塊邊緣之觀測資料的分享共用。一方面涉及我國學者使用他國(尤指美、日)觀測資料;另一方面則由於我國科會海科中心海洋資料庫已頗有成效,在未來海洋學門指定海研一號貴重儀器中心統籌運作,如何參與國際地科學界合作將是重要的課題。而且本次會議受邀者俱為全球重要海洋科學中心, 亞洲地區僅邀請我國與日本, 機會難得, 因此決定參加 2007 年 5 月 9-11 日在德國的研討會。該研討會參與者均為大型跨國地科計劃之 PI 或資料庫負責人(見附件二), 會期雖然不長, 但目標明確討論集中, 對於發展國際性板塊邊緣之觀測資料分享共用的文化頗有重大推展。而且我國學界海域觀測資料之發展與管理也受到一定之矚目。對於嗣後發展改進方向很有啟發。研討會初步共同工作報告如附件三。

由於研討會對於受邀參與者提供部分補助,其不足部分由海科中心支援。故本年度並未使用計劃核定之出國差旅經費。

附件一

HR: 0800h
AN: **S51C-1283**
TI: **Multi-scale upper mantle tomography of the Eurasia using surface waveform data**
AU: * **Gung, Y**
EM: ycgung@ntu.edu.tw
AF: Department of Geosciences, National Taiwan university, P.O. BOX 13-318, Taipei, 106 Taiwan
AU: **Chiao, L**
EM: chiao@ntu.edu.tw
AF: Institute of Oceanography, National Taiwan university, P.O. BOX 23-13, Taipei, 106 Taiwan
AB: We invert long period seismograms in the time domain in the framework of normal-mode-based nonlinear asymptotic coupling theory (NACT) [Li and Romanowicz, 1995] for the seismic structure underneath Eurasia. While only Eurasia region, where the sensitivity is the highest in the selected data set, is inverted for its radial anisotropic structure, the effects from global 3D heterogeneous structure are taken into account in the forward stage. The implementation of multi-scale inversion is achieved by converting the partial derivative matrices of the initial model parameters, either spherical harmonics or globally distributed spherical triangle meshes, into a multi-resolution wavelet representation in our interested region. In the scheme of multi-resolution wavelet representation, model parameters are grouped into natural hierarchy of local scales such that the damping regularization acts to sort through successive scales depending on the local data constraints. In other word, the spatially nonstationary smoothness enhancement is automatically invoked depending on the in-situ rigors of the model constraints offered from the data. As a result, sites with strong constraints are capable of resolving more details robustly whereas stable long wavelength components are still available for sparsely constrained area [Chiao and Liang, 2003].
DE: 7200 SEISMOLOGY
DE: 7208 Mantle (1212, 1213, 8124)
DE: 7255 Surface waves and free oscillations
SC: Seismology [S]
MN: 2006 Fall Meeting

HR: 14:55h
AN: **T53F-06**
TI: **Possible eastern edge of the Meso-Tethyan slab in the lower mantle beneath southern Tibet**
AU: * **Kuo, B**
EM: byk@earth.sinica.edu.tw
AF: Institute of Earth Sciences, Academia Sinica, POB 1-55 Nankang, Taipei, 10000 Taiwan
AU: **Lin, P**
EM: r90241303@ntu.edu.tw
AF: Institute of Oceanography, National Taiwan University, 1 Roosevelt Road, Taipei, 10000 Taiwan
AU: **Chiao, L**
EM: chiao@ntu.edu.tw
AF: Institute of Oceanography, National Taiwan University, 1 Roosevelt Road, Taipei, 10000 Taiwan
AB: We analyzed slowness of P, ScP, and PcP recorded by Indepth III and HIMNT temporary arrays in Tibet from events in the Sumatra-Sunda subduction zones. Raypaths of P, ScP, and PcP constitute a wide range of incidence angles that could help sorting the position of an anomaly. Unlike the HIMNT data, the Indepth slowness residuals show significant discrepancy between P and the group of ScP and PcP. The mean of the former is 0.10 s/deg, while the mean of the latter is -0.46 s/deg. The residual vs. incidence angle pattern is robust and excludes systematic velocity variations in the upper mantle as the source of the discrepancy. The contribution from D” can be ruled out by the similarity between ScP and PcP and the extreme magnitude of the residuals. The favored depth range to suit the anomaly is the upper part of the lower mantle where ScP and PcP still remain close but together far enough from the P paths. Models that are consistent with our observations are (1) a horizontal gradient over a distance of 300 km with 1 percent increase in Vp per 100 km towards northwest; and (2) a volumetric anomaly of 2-3 percent with a 90-deg or obtuse corner that is sampled by the bundle of rays to the array. The combination of the two also gives the observed negative residuals. Both (1) and (2) represent a boundary of a subducted slab exposed sideways to the mantle. The broken Meso-Tethyan oceanic lithosphere that was last consumed ~150 Ma is the most likely candidate for this structure.
DE: 1212 Earth's interior: composition and state (7207, 7208, 8105, 8124)

DE: 1213 Earth's interior: dynamics (1507, 7207, 7208, 8115, 8120)
DE: 7203 Body waves
SC: Tectonophysics [T]
MN: 2006 Fall Meeting

HR: 15:10h
AN: **T53F-07**
TI: **Imaging Upper Mantle Structure Beneath the Tibetan Plateau and the Himalaya by Multiscale Finite-Frequency Tomography**
AU: * **Hung, S**
EM: shung@ntu.edu.tw
AF: Department of Geosciences, National Taiwan University, Taipei, 106 Taiwan
AU: **Wang, C**
EM: r92241318@ntu.edu.tw
AF: Institute of Oceanography, National Taiwan University, Taipei, 106 Taiwan
AU: **Chiao, L**
EM: chiao@ntu.edu.tw
AF: Institute of Oceanography, National Taiwan University, Taipei, 106 Taiwan
AB: The Tibetan plateau and the Himalaya, created by the Indo-Asian collision started 50 million years ago, are the classic sites for studies of the evolution of continental orogeny. Determining seismic velocity structure of the underlying crust and mantle is essential for understanding how plate tectonics and mantle dynamics shape the towering Himalaya mountains and flat topography of Tibet. Recent development in finite-frequency tomography has been proven useful in imaging 3-D velocity variations on the scale comparable to the characteristic wavelength of the waves. Using available data from the INDEPTH and HIMNT experiments, we conduct finite-frequency traveltime tomography for compressional wavespeed heterogeneity of the upper mantle beneath Tibet. We measure relative delay times of P-wave arrivals between stations using multichannel cross-correlation of bandpass-filtered waveforms in different frequency ranges. The measured traveltime delays of the same phase arrival at different frequencies are actually sensitive to individual unique volume of structural heterogeneity surrounding the ray path. Such frequency-dependent sensitivity is naturally represented by 3-D banana-doughnut shaped Fréchet kernels for tomographic imaging. Moreover, multiscale wavelet-adaptive parameterization is invoked in the inversion and the resulting velocity models have spatially-varying resolutions subject to the quality of data sampling. The preliminary model reveals a region of relatively high P-wave velocity extending continuously from the uppermost mantle to the depth of 350 km beneath central Tibet (30°N--34°N). At depths above 200 km, there is a strong lateral gradient of 3--4% in P wavespeed from high velocity structure beneath the Himalaya to low velocity beneath the Tibetan plateau.
DE: 3260 Inverse theory
DE: 7203 Body waves
DE: 7270 Tomography (6982, 8180)
DE: 8102 Continental contractional orogenic belts and inversion tectonics
DE: 8120 Dynamics of lithosphere and mantle: general (1213)
SC: Tectonophysics [T]
MN: 2006 Fall Meeting

附件二

Last Name	First Name	Institution
Abers	Geoff	MARGINS Office, Boston University (USA)
Ahern	Tim	IRIS (USA)
Al-Habsi	Harib	Sultan Qaboos University (Oman)
Arko	Bob	LDEO, Columbia University (USA)
Bach	Wolfgang	University of Bremen (Germany)
Baker	Maria	National Oceanography Centre, Southhampton (UK)
Barckhausen	Udo	BGR-German Geological Survey
Baumann	Peter	Jacobs University Bremen (Germany)
Beaudoin	Yannick	UNEP-Continental Shelf Programme (Norway)
Blackman	Donna	Ridge 2000 Office, Scripps institution of Oceanography (USA)
Blower	Jon	University of Reading (UK)
Briones	Katiusca	Oceanographic Insitute of Ecuador
Cannat	Mathilde	Institut de Physique du Globe de Paris (France)
Carbotte	Suzanne	LDEO, Columbia University (USA)
Chen	Bob	CIESIN, Columbia University (USA)
Chiao	Ling-Yun	National Taiwan University (Taiwan, ROC)
Clark	Dru	University of California, San Diego (USA)
Condit	Christopher	University of California, San Diego (USA)
Cogan	Christopher	Alfred Wegener Institute for Polar and Marine Research (Germany)
Damm	Timo	University of Kiel (Germany)
Devey	Colin	IFM-GEOMAR (Germany)
Diviacco	Paolo	Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy)
Dransch	Doris	GeoForschungsZentrum Potsdam (Germany)
Fox	Christopher	National Oceanic and Atmospheric Administration (USA)
Galkin	Anastasia	GeoForschungsZentrum Potsdam (Germany)
Goldstein	Steven	LDEO, Columbia University (USA)
Goodwillie	Andrew	LDEO, Columbia University (USA)
Graybeal	John	Monterey Bay Aquarium Research Institute (USA)
Haak	Katherine	Ridge 2000 Office, Scripps Institute of Oceanography (USA)
Halvorsen	Oystein	UNEP-Continental Shelf Programme (Norway)
Hanafusa	Yasunori	JAMSTEC (Japan)
Haq	Bilal	National Science Foundation (USA)
Hosseini	Keivan	Ferdowsi University of Mashhad (Iran)
Huettmann	Falk	University of Alaska, Fairbanks (USA)
Javidpour	Jamileh	IFM-GEOMAR (Germany)
Jones	Craig	GNS Science (New Zealand)
Kandel	Cary	MARGINS Office, Boston University (USA)
Khan	Shuhab	University of Houston (USA)
Klump	Jens	GeoForschungsZentrum Potsdam (Germany)
Le Bas	Tim	National Oceanography Centre (UK)
Lehnert	Kerstin	LDEO, Columbia University (USA)
Lezaeta	Pamela	MARGINS Office, Boston University (USA)
Lowry	Roy	British Oceanographic Data Centre (UK)
Matsuda	Shigemi	The Center for Deep Earth Exploration (Japan)
Meier	Thomas	Bochum University (Germany)
Miller	Stephen	University of California, San Diego (USA)
Miville	Bernard	International Ocean Drilling Program (Japan)
Moussat	Eric	IFREMER (France)
Neben	Soenke	BGR-German Geological Survey
Nygard	Atle	University of Bergen (Norway)

Pirenne	Benoát	University of Victoria (Canada)
Ramirez-Llodra	Eva	CMIMA-CSIC (Spain)
Ranero	Cesar	CMIMA-CSIC (Spain)
Ryan	William	LDEO, Columbia University (USA)
Sadeghi	Hossein	Ferdowsi University of Mashhad (Iran)
Salters	Vincent	Florida State University (USA)
Sarbas	Baerbel	Max-Planck-Institute for Chemistry (Germany)
Schaap	Dick	SeaDataNet (Netherlands)
Schaefer	Angela	Jacobs University Bremen (Germany)
Schirnick	Carsten	IFM-GEOMAR (Germany)
Schoolmeester	Tina	UNEP-Shelf Programme (Norway)
Shiomi	Katsuhiko	National Research Institute for Earth Science and Disaster Prevention (Japan)
Shipley	Thomas	University of Texas (USA)
Stransky	Julia	IFM-GEOMAR (Germany)
Trueger	Mickael	IFREMER (France)
Tsuboi	Seiji	JAMSTEC-IFREE (Japan)
Unnithan	Vikram	Jacobs University Bremen (Germany)
Venuti	Fabio	National Oceanography Centre (UK)
Wallrabe-Adams	Hans-Joachim	University of Bremen (Germany)
Weatherall	Pauline	British Oceanographic Data Centre (UK)
Weinrebe	Wilhelm	IFM-GEOMAR (Germany)

附件三

Building a Global Data Network for Studies of Earth Processes at the World's Plate Boundaries

Draft Workshop Report

Executive Summary

An international group of scientists, data managers, and information technologists convened for a 2.5 day meeting in Kiel, Germany, to explore the opportunities for international data exchange and to address the cultural and political challenges to building a global data network that facilitates mid-ocean ridge and continental margin related research internationally. Workshop participants discussed technical, procedural, and organizational issues of global data sharing, and reached agreement on the following set of recommendations grouped broadly under the working group themes:

Science User Needs

- *Open access to data is fundamental to verifiable scientific progress. All data that are necessary to reproduce published scientific results, including field data, processed data, and laboratory (derived) data products, need to be published and archived in accepted archives. We need to advance a culture among scientists that is more open to data sharing.* (T1-R1[1]; T2-R5; T4-R4)
- *Scientists require access to multidisciplinary data and data integrated from both the marine and terrestrial world.* (T1-R2; T1-R3)

Data Documentation and Publication

- *Uniform best practices and standards need to be developed and used routinely within the international community for data acquisition, data submission to data centers, and data publication. Best practices should include formal submission agreements between individual institutions and respective inter/national data centers and the use of globally unique identifiers for data. Scientific societies should take on an active role in formulating best practice guidelines for the publications of data. In addition, new mechanisms are needed to track the use of data sets both to ensure academic recognition and to support scientific collaborations.* (T1-R4; T2-R2; T2-R4; T2-R6; T2-R7; T2-R8; T4-R1)
- *The ultimate responsibility for ensuring adequate documentation of a field program lies with scientists. Metadata creation and data submission should be made as easy as possible for ship*

---

[1] Theme 1 – Reccommendation1

operators and scientistswith development of new tools for automation to support and further the implementation of best practices and standards. Funding agencies must be involved in enforcing standard practices for data documentation and submission to data centers. (T2-R1; T2-R3; T2-R4; T2-R6)

Data and Metadata Interoperability

- *The community must minimize the proliferation of metadata standards and work toward a uniform approach for scientific metadata. Processes need to be defined regarding how to develop community-based standards, guidance, and profiles. New efforts to develop standards and protocols to support interoperability should build upon and take advantage of existing community-based projects.* (T3-R1; T3-R2; T3-R3; T3-R4)
- *Development of a data discovery service across distributed marine geoscience data resources within the international community is an achievable initial goal. Data centers should work to expose their data resources via web services using e.g. OGC or OAI protocols.* (T3-R5; T3-R6)

Opportunities and Obstacles for International Data Sharing

- *International programmes and bodies such as GEOSS, the eGY and ICSU that stimulate the development of global data sharing systems should be leveraged to promote an initiative for a global data network for marine and terrestrial geoscience data.* (T4-R5)
- *A dedicated task group should be established to advance the implementation of a global data network. In addition, special interest groups to share experience and solutions on issues concerning metadata and interfaces should be formed with tools to facilitate collaboration.* (T4-R6; T3-R7)

Based on these recommendations, the following next steps are identified; 1. Develop test bed sites for a data discovery service across globally distributed data resources; 2. Establish forums for guidance and development of best practises in the areas of data acquisition, metadata, vocabularies, and interfaces; 3. Formulate a dedicated task group to advance international alliances; 4. Convene a follow-up meeting in one year.

1. Motivation for the Workshop

Over the past decade, rapid advances have occurred in database technology for scientific research providing new access to data and new tools for data visualization and integration. Along with these advances in Information Technology has come the growth of digital data collections for a broad suite of data related to the marine geosciences. Developments in database connectivity provide new opportunities for open exchange of data across distributed data collections, greatly expanding the volume and diversity of data available to the scientist to address a particular scientific problem of interest. These advances hold great promise for the solid earth sciences, an inherently multi-national and multi-disciplinary field, which involves the collection of typically unique data sets during oceanic and terrestrial expeditions conducted by research institutions around the globe.

The international marine geoscience community is actively engaged in scientifically aligned goals through the InterRidge-Ridge2000 programs and InterMARGINS-MARGINS programs. These programs represent broad multi-disciplinary initiatives focused on understanding the fundamental processes of crustal formation, modification and destruction at the world's plate boundaries. InterRidge and InterMARGINS are international programs, which aim to coordinate efforts and priorities in mid-ocean ridge and continental margin research, respectively, across nations. Ridge2000 and MARGINS are aligned US-funded programs which conduct focused investigations in a few geographic locations, most of which involve international partners. At present there are no formal agreements for data sharing within these international communities, and data exchange occurs primarily by informal agreements between scientists directly involved in specific projects. However significant benefits to these linked marine-terrestrial geoscience research efforts internationally could be achieved if data collections maintained as national efforts could be better linked and if broader access were initiated. New database technologies are available that enable independent globally distributed sites to share, link, and integrate their data holdings and services while maintaining full ownership and credit for these holdings.

To explore current opportunities and challenges for international data exchange to support continental margin and mid-ocean ridge research, the workshop entitled "Building a global data network for studies of earth processes at the world's plate boundaries" was convened with two primary goals:

1. *To explore current data management efforts relevant to continental margin and mid-ocean ridge science goals within partner countries.*
2. *To devise a strategy for building a global data network to support the sharing and exchange of data of greatest scientific interest for continental margin and mid-ocean ridge studies.*

The primary hoped for outcome of this meeting was the development of new partnerships between marine geoscientists and data centers within the international community to establish greater access and exchange of data sets of broad interest for continental margins and mid-ocean ridge research.

2. Workshop Structure

The workshop was convened by four scientists from Germany, Japan, and the United States from within the InterMARGINS and InterRidge communities, and jointly sponsored by InterMARGINS, MARGINS, InterRidge and Ridge2000. The US National Science Foundation and the German project "The Future Ocean" provided additional financial support. About 70 people from 14 countries attended the workshop, including scientists from the InterRidge and

InterMARGINS communities, data managers representing data centers and data systems across a spectrum of geoscience data, and information technologists involved in various aspects of interoperability development.

The workshop was held at the meeting facilities of the Hotel Birke in Kiel, Germany. The official program started on May 9 in the morning and lasted for 2.5 days. Interested participants were invited to continue discussions on May 11 in the afternoon. The workshop ended with an informal field trip to the historical town of Lübeck on May 12.

The first 1.5 days of the workshop were devoted to presentations within three general areas:

(a) *Science Needs:* Scientists outlined their needs for data access and defined data sets of broad interest for continental margin and ridge-related science.

(b) *Data Resources*: Representatives of data centers presented existing data systems available for academic research. These presentations were complemented by poster presentations and live demonstrations of the systems.

(c) *Technologies:* Information technologists reported about emerging technologies for interoperability and data sharing.

The afternoon of Day 2 and morning of Day 3 were devoted to working group sessions to discuss technological as well as organizational and cultural issues of global data exchange. The working group discussions were structured into four themes, each of which (except for the 'Science User Needs' group) had two sessions:

1. Science User Needs & Concerns
2. Data Documentation and Publication
    - *a.* Standards for Data Documentation
    - *b.* Data Publication
3. Data and Metadata Interoperability
    - *a.* Standards & Technologies for Metadata and Interfaces
    - *b.* The Low-Hanging Fruit for Data Exchange
4. Opportunities and Obstacles for International Data Sharing
    - *a.* Archives and Data Contributions
    - *b.* Implementing an International Data Network

Each working group addressed a range of questions provided to the session leaders by the workshop conveners, and was charged to generate a set of recommendations working group leaders presented in plenary sessions. Questions and recommendations are outlined in the following section.

3. Working Group Discussions

3.1. Theme 1: Science User Needs & Concerns

Scientists engaged in plate boundary research study the wide variety of active processes associated with the formation, modification, and destruction of the crustal layer of the earth, which supports life on the planet. Plate boundaries transect the oceans, hug the continental margins, and penetrate into continental interiors. They are the locus of most earthquake and volcanic activity on earth and of the pervasive fluid-chemical-thermal interactions associated with the development of unique ecosystems and the formation of economical metal deposits. Increasingly, these active environments are studied as integrated complex physical, chemical and biological systems, subject to a variety of influences, rather than as primarily geological structures. To address these interdisciplinary goals, scientists increasingly require access to multidisciplinary data sets and from both the terrestrial and marine setting. These needs represent unique requirements and challenges for scientific data access and exchange.

The science user working group considered the following questions;

- *What are science user needs and concerns with regard to data sharing?*
- *What are the key data sets needed for international exchange?*
- *What links exist and are desired between the marine and terrestrial world?*
- *What capabilities are desired that are currently lacking? What technologies are promising to scientists?*

There is strong endorsement within the science user community of the principle of fully open access to data. Scientists desire access to all existing data relevant for the problem they wish to address. For programs conducted in the open ocean, scientists desire access to everything collected in a geographic area of study. Closer to shore, along the continental margins, there may be economic or national security concerns that affect access to some kinds of data, but much data of value to basic science should be available. Easy access to a diverse suite of data is necessary for many studies (Table1). However, many of the data resources currently available represent disciplinary databases. More focus is needed on building data systems to support integrative science, providing access to multi-disciplinary data. Although the fundamental science questions associated with continental margin studies transect the shoreline, the shoreline represents a major boundary in how data are collected, organized and later archived. This disparity is a significant obstacle to scientific data access.

Recommendations

T1-R1 ***Open data access is fundamental to verifiable scientific progress.*** Full open access to data is needed, first and foremost to support scientific progress but also very importantly, to enable the verification of research results. Geosciences in general relies on a unique set of field observations, so differs from most experimental sciences in that most measurements are difficult to repeat. With the typically unique data sets used to support plate boundary studies, research results are often impossible to verify without open access to field observations and measurements.

Scientists want unrestricted access to all data as feasible within the framework of national requirements and proprietary periods of data collectors. National needs may require limitations for some data types and in some environments (eg. Ultra-high resolution bathymetry in shallow coastal waters, on-land gravity, reflection seismics in petroleum-rich basins), but every reasonable effort should be made to release such data in a reasonable time frame. For research data subject to proprietary hold periods, scientists would like access to metadata describing the existence and location of the data at an early stage with mechanisms that support interactions between data collectors and other scientists wishing to form collaborations.

T1-R2. ***Scientists require access to multidisciplinary data.*** The integrative science programs that characterize modern studies at mid-ocean ridges and continental margins drive the need for integrated access to multidisciplinary data. More and more, scientists seek to work across traditional disciplinary boundaries either through developing collaborations or by acquiring interdisciplinary expertise. Data systems, which support and facilitate collaborations and multidisciplinary access are required. Scientists need access to multi-disciplinary databases of geographically referenced data as well as of physical property measurements such as experimentally-derived material properties. Derived data sets including images, and data-based models have tremendous value for interdisciplinary studies and these need to be preserved.

T1-R3. ***Integration of data resources from both the marine and terrestrial world is needed.*** Research along the continental margins requires access to both terrestrial and marine data. However, available data resources typically stop at the shoreline with different agencies and organizations involved in terrestrial and marine studies. Significant obstacles to obtaining access to data across the shoreline relates to differences in how data are collected and organized. Whereas offshore work is usually defined and organized by cruise, onshore field studies are characterized in a variety of ways, by networks of instruments, by investigating group, by national or other geographic boundaries, or otherwise. Also, onshore and near-shore data sets tend to be spread through a wide array of national agencies with varying standards and missions. Data systems are needed which support the ability to search for and find related data objects in a variety of different frameworks that make sense for the problem at hand, not always dependent on the platform or group collecting the data. While geographic data access makes sense for many problems, time-series data inherently require the need for searches at a wide variety of time scales, and the wide variety of characteristics of different data sets indicate other primary search categories may be valuable as well.

T1-R4: ***Mechanisms are needed to track the use of data sets both to ensure academic recognition and to support scientific collaborations.*** While the existence of open data collections representing the accumulation of data from many individual studies provides important resources for scientists, an ongoing concern is how to ensure that credit to original data collectors is preserved.

Within the current framework of citation supported by scientific journals, it is often not possible to cite the large number of original data sources used for a new synthesis (see also chapter 3.2.2 on Data Publication).

3.2. Theme 2: Data Documentation and Publication

The development of digital data resources for marine geoscience data, along with new technologies for data visualization and analysis, is changing the way marine geoscience research is conducted. More and more scientists make use of digital data collections as primary resources for data in an area of interest, for conducting global syntheses, and to facilitate new multi-disciplinary studies. The utility of digital data resources fundamentally depends on the comprehensiveness and the quality of the data they provide and therefore requires that data are (a) openly and fully accessible, and (b) documented properly at all stages of the data life cycle, from initial acquisition, through processing, to primary and later secondary publication to ensure evaluation of data quality. These requirements deeply impact the scientific data culture, imposing new obligations on scientists such as metadata compilation and full disclosure of data, and changing the way data is referenced and cited. This Theme focused on issues of Data Documentation and Publication.

3.2.1. Session I. Data Documentation

The breakout group on Standards for Data Documentation addressed the following questions:

- *Review current practices for different sub-domains.*
- *How can we achieve standardized data documentation during acquisition in the field/at sea? For derived data?*
- *How do we ensure the highest level of data quality? What metadata requirements are necessary?*
- *What roles can and should agencies, ship operating institutions, and publishers play?*

Working group discussions focused primarily on field data acquisition during marine surveys. Current practices for data acquisition and documentation at sea are highly heterogeneous across the global marine geoscience community. In many cases, data documentation is the exclusive domain of the science party. While scientist must ensure that adequate documentation of their data of interest is obtained for their own use, this documentation is typically recorded only in the scientists own workbooks or spreadsheets and is seldom captured for later incorporation into data systems. In addition, the documentation that a scientist may provide for their own data reduction purposes is often insufficient to facilitate later use of the data by others. The ROSCOP cruise form, widely used to report cruises within the European community, captures only a minimal documentation of cruise operations. Furthermore, on many modern expeditions routine collection of data types other than those of primary interest to the science party may occur which can go largely undocumented. The challenge then is to move toward more thorough and complete data documentation for all marine programs carried out within the international research community.

The consensus is that while the collection of cruise metadata is often incomplete and that this is a global issue, improving data documentation at sea is a tractable problem. The needed information is collected in some form during a field program. The problems are to find relatively easy ways to get this information out of the notebook or personalized electronic file of the scientist or technician, and into a standardized format, and to formalize the transfer of this record-keeping to the relevant database system.

Procedures for capturing this information need to be of obvious benefit to the scientists themselves and must minimally impact their existing responsibilities. The current bureaucratic overhead of research for scientists is high and it is important to design documentation procedures that add minimum extra burden to their responsibilities.

To facilitate more complete documentation of data acquisition at sea, standardized metadata forms have been developed within some communities (e.g. the MGDS forms developed for the US MARGINS and Ridge2000 programs; www.marine-geo.org/metadata_forms.html). IFREMER has established a data quality plan that outlines procedures for standard data acquisition aboard their ships **(**ask Eric Moussat to describe further**).** The sample registry SESAR provides unique identifiers (the International Geo Sample Number) for samples to ensure that all sample analyses can be ultimately tied to a unique sample. The existing standardized forms of the MGDS were examined during breakout group discussions as possible working models for basic data documentation at sea. The information requested is generic and should not be considered an extra burden for scientists to provide. Marine expeditions involve a wide array of data collection activities in addition to the standard underway geophysical data streams (e.g such as multibeam, gravity, magnetics) and all of these must be documented (e.g cores and dredges, biology samples from dives, OBS deployments, etc.). Ideally, standard forms should be designed so that they can replace scientist's personal records. An "open format" where scientists can add columns to the standardized format according to their requirements would be needed.

Recommendations

T2-R1. ***The ultimate responsibility for ensuring adequate documentation of a field program lies with scientists.*** On many ships and for many data types, the shipboard science support staff will produce the needed data documentation as part of their routine operations. But the shipboard support staff is unlikely to have access to all information on the full suite of data acquired during a program. Scientists bring their own sensors on board, and are typically in charge of station operations associated with sampling or instrument deployment. As the primary interest and responsibility for the scientific data acquired during an expedition reside with the science party, the ultimate responsibility for ensuring comprehensive documentation for all data should also lie with the scientists. For some ships, (e.g. within the UK) a data/metadata specialist who is responsible for generating complete documentation of survey operations sails on each cruise (eg. **Ask Roy Lowry to confirm and for more info here on BODC operations**). As an alternative, standard practice should include the identification of a "data liaison" from within the science party, who works with the ship's support staff to ensure capture of all needed information.

T2-R2. ***Routine use of standardized data documentation procedures should be adopted by ship operators and scientists.*** Comprehensive and standardized data documentation at sea is a tractable goal. The standardized electronic metadata forms provided by the MGDS, the data quality plan of IFREMER, and assignment of IGSNs to samples are steps in the right direction and provide models for wider adoption. While ships are operated by different agencies in different countries, each with its own procedures and requirements for survey operations, the concept of standard metadata forms should be generally applicable. Metadata forms need to be developed in close collaboration with users with easy mechanisms for users to customize forms for specialized use. Required basic cruise level information should include listings of the science party, roles and affiliations, an inventory of all projects associated with an expedition and of all data types collected. Minimum required metadata for all kinds of data acquisition are ***date, time, latitude-longitude, and depth.*** All rock and sediment samples should be assigned International Geo Sample Numbers (IGSNs). For sensor data, other required basic information includes:

- Information on all ship sensors operated during program, including manufacturer, make, model, and if possible serial number.
- Basic sensor information for any sensors brought on board by the science party.
- Calibration information for all equipment.

T2-R3. ***Automated tools for metadata creation at sea are needed.*** Metadata creation suitable to support long-term preservation of data is time consuming for scientists to produce and they lack sufficient incentive. New automated methods to tag data with required metadata at the time of data acquisition are needed. The long-term future vision to support marine geoscience data acquisition is a web-based shipboard event logging system that pulls in the required information such as navigation, person, sampling event or operation, sample type, etc. The shipboard event logging system should include pull down menus of controlled vocabularies to describe operations. A comprehensive shipboard data acquisition system is in use for IODP cruises and is a model for wider application.

T2-R4. ***Funding agencies must be involved in enforcing standard practices for data documentation and submission to data centers.*** Requirements for the standard documentation and submission of data acquired during all field programs will need to be enforcable through funding agency actions.

3.2.2. Session II: Data Publication

Discussions in this working group were concerned with issues relating to policies and procedures for data publication:

- *What data need to be accessible (raw vs. derived, published vs. unpublished)?*

- *How should data be identified? (Use & granularity of unique identifiers for data)*
- *How can new requirements for data publication be implemented, what are the special disciplinary issues?*

Issues concerning data publication are a key concern to both individual scientists and to data system providers. Scientists publish the data that they acquire through analytical, experimental, or computational procedures as a major product of their research, 'marketing' them to gain credit and reputation that ultimately form the currency of their careers (Edwards et al. 2007). In many scientific cultures, data have traditionally been treated as private intellectual property and have typically been shielded carefully, often even after publication. Journal articles frequently contain only fragments of a 'published' dataset (tables with 'representative analyses'). Publication of raw data has been a rare exception and data documentation in general is poor and quite heterogeneous. Edwards et al. (2007) state that the *"private-ownership practice has led to a plethora of data collection practices and data formats, many of them idiosyncratic, as well as an absence of the metadata needed by other scientists to understand how the data was originally produced."*

While many scientists now recognize the benefits of digital data collections and support their existence, they are rightfully concerned that access to their data via digital data resources will circumvent the original journal publication of the data and leave them without being properly cited and receiving credit for their data. Policies and procedures for data publication as well as the design of a global data network need to address these concerns. The appropriate use of globally unique identifiers for data that allow dataset to be identified and cited independent of a journal publication, but also allow to link data in digital data collections to the original publication in the scientific literature can contribute to a satisfactory solution.[2]

Scientific data come in many different types. The main differences relate to their origin (e.g. sensors, observation, experiment, modeling), their nature (digital data, physical specimens, numerical, images, video), and the level of processing (raw data, corrected, reduced, or 'derived' data). Data related to oceanic expeditions ranges from geophysical to geochemical to biological data. Data acquired shipboard range from raw to processed data, among them underway geophysical data streams (e.g. multibeam, gravity, magnetics), CTD casts, and rock, fluid, or biological samples. 'Derived' data are mostly generated on-shore in laboratories, with application of a wide range of processing procedures to raw geophysical data or by analyzing samples collected during a cruise. Guidelines are necessary to define criteria for identifying data that should be preserved, data that should be published, and data that should be 'discarded' after use. An example for such guidelines

---

[2] *For example, the German project "Publication and Citation of Scientific Primary Data" (http://www.std-doi.de) has prototypically implemented a system for the publication of scientific data, which is open to the scientific community in any scientific field. This project uses persistent identifiers (DOI, handle.net and URN) to identify datasets available in a digital format.*

are the "Rules of Good Scientific Practice" adopted by the Max-Planck-Society that take a general perspective on the data preservation issue:

"Scientific examinations, experiments and numerical calculations can only be reproduced or reconstructed if all the important steps are comprehensible. For this reason, full and adequate reports are necessary, and these reports must be kept for a minimum period of ten years, not least as a source of reference, should the published results be called into question by others."

A large part of the discussion was related to who should submit the data to the archive (database), revealing differences in culture between countries on how the ships are operated. It also brought to the forefront that the data submission requires standard data input, like cruise name, dates, participants etc, that are already available in some form to the ship operator. This standard data should be pre-loaded, or be easily available without re-entering.

Recommendations

T2-R5: ***All data necessary to reproduce published scientific results needs to be published and archived in an accepted data archive.*** Raw data from sensors should be archived along with the appropriate metadata that allow processing and interpretation of the data. In addition, standard (routine) corrections should be applied to the "raw" data to make the data more easily usable to a larger community. These corrected data should be archived as well. Physical samples are considered 'raw' data for analytical data such as geochemical measurements and should be archived to ensure that analytical data is reproducible and can be complemented by new measurements. Sample repositories barely exist for samples from ocean going expeditions, and are virtually absent for land-based expeditions. It is critical that samples carry globally unique identifiers to ensure unambiguous identification and allow tracking their analytical history.

During a cruise, some data types may be processed. Files with processed data should be submitted to the relevant databases, accompanied by adequate documentation about the processing method. For post-cruise processed data, the situation can be very different. While it is unclear how to proceed, there was consensus that PIs should notify collecting institution database groups when they submit processed data to relevant data banks.

T2-R6: ***Data submission should be streamlined and standardized.*** Procedures are needed to seamlessly integrate data into databases, and make the process of data submission as easy as possible for scientists, while ensuring comprehensive and consistent data documentation. Data submission requires standard data input, like cruise name, dates, participants, etc. that are already available in some form to the ship operator. This standard data should be easily available so that researchers submitting their data do not have to re-enter this information.

Data types such as geochemical measurements need a standard set of parameters (sample and analytical metadata) at the time of publication to accompany the sample information before a paper is accepted. Editors need to link acceptance of a manuscript to the submission of the data and accompanying metadata to a public "accepted" archive. Whenever possible, published derived data should be in a re-usable format (e.g. electronic data table).

T2-R7: ***Unique identifiers for data should be used at the level of a study or publication***. The working group reached consensus that unique identifiers for data should be applied at the level of a "study" or "publication", and not at finer granularity such as a single analysis. This recommendation pertains to raw data as well as peer-reviewed published data, which is often derived data. Modern publications already have unique identifiers (DOI). Older publications might not, and incorporation of that data in databases might require "new" unique identifiers.

T2-R8: ***Scientific societies should take on an active role in formulating best practice guidelines for the publications of data.*** There is the general recognition that the existence of databases has improved the quality and documentation of the published data. Societies should take on the role of formulating best practice guidelines for data publication. These best practice guidelines need to be "enforced" by funding agencies through policies established based on the guidelines, and journals for which editors should enforce the guidelines.

3.3. Theme 3: Data and Metadata Interoperability

The goal of data interoperability requires solutions to the challenges of data exchange in a world of heterogeneous data formats, metadata formats, diverse vocabularies, and varying interfaces or protocols for metadata and data transport. Adoption of standards in these areas will be required to enable true interoperability, and tools and services will need to be available that support the chosen standards. To the extent each community or sub-community has its own data management needs and practices, those groups must collaborate to establish agreements on the common approaches they will follow to enable interoperability.

3.3.1. Session I: Standards and Technologies for Metadata and Interfaces

The Session I working group was asked to consider the following questions:

- *Review existing standards for metadata & interfaces, their current use, success, advantages & disadvantages*
- *Review existing registries for data resources, their use, success etc.*
- *Are new technologies or standards needed?*

Metadata: Group discussions began with metadata and the requirements for metadata. The need for and range of metadata required vary depending on the intended application and it is important to define what the use applications are before consideration of what metadata standards should or could be adopted. Motivations for metadata include; to describe data (who, what when, where, how, data quality); to facilitate data discovery and new scientific collaborations; to reprocess and synthesize data; to exchange data including harvest it to one location for specialized use; and to generate user interfaces.

The development of a consistent community practice with respect to metadata is hindered by a wide range of problems including: (list as sidebar)

- Benefits of metadata may not be adequately understood by those who originally document a dataset, leading to minimal and inadequate metadata for most reuse.
- Interpretation of standards differ and there is little guidance on how to fill them out.
- Some required information for the intended use is not provided. This is an inevitable outcome of different users of data having different interests and needs from those who originally documented the data.
- To make metadata fully discoverable and usable by users from other scientific domains, it may be necessary to satisfy a number of sophisticated standards and vocabularies, even for a single data set. This level of sophistication is not supported by current tools and data models, and not expected by users.
- For legacy data, it may be impossible to recover the needed metadata after the observations have taken place.
- The desire to control what information is exposed sometimes constrains the metadata that is provided (for example, the location of a ship working in an ecologically or financially sensitive area).
- Initial creation of metadata by users can be time-consuming, confusing, and unrewarding (due to the amount of metadata requested, poor tools and user interfaces, and limited infrastructure supporting metadata creation).

Common practice for how metadata is provided also varies greatly between disciplines and data types. For some data types, metadata may be embedded in formatted data (e.g. GeoTIFF, HDF, NetCDF, NITFS, SEGY, MGD77, ESRIgrid (ARCASCII), GRIB). For embedded metadata, additional challenges include inconsistent metadata formats in file headers and the often inadequate models and structure for information (metadata/data) adopted in the file format. For other data types, metadata are provided external to data. Currently used standards include FGDC, DIF, Dublin Core, and ISO 19115 (following the implementation approach of ISO 19139).

Most data and metadata centers are moving to work with ISO 19115, but it is a somewhat general-purpose standard. To become more useful for a particular community, a *profile* or *extension* (see sidebar) must be developed that meets the community needs. Of course, such tailored enhancements of the standard will not work with the ones developed for other communities, unless specific measures are taken to assure interoperability. In addition, ISO standards are not freely available (and in fact are somewhat costly). The workshop participants expressed concern that these issues might inhibit widespread adoption of ISO 19115.

Include as side bar

“Extensions, Profiles, and Vocabularies

Extensions are additions to a metadata standard that allow users to provide information in additional fields that were not mentioned in the original standard. In standards such as ISO 19115, extensions include[1]:

- addition of a new metadata section

  alteration of the domain of a metadata element (for example, assigning a code list to specify what responses are allowed for that metadata element)
- addition of terms in a code list
- addition of a new metadata element to an existing metadata element
- addition of a new metadata entity
- changing the obligation of a metadata element from optional to mandatory (but not the reverse, which would break the core standard)

Constraints are considered a specialized subset of extensions, in which additional restrictions are placed on the standard. (In the above list items 2 and 6 are constraints.) In this case the term 'extension' is describing the addition of information to the standard, even though the metadata instances that follow the standard are restricted.

Profiles are the community-specific application of the metadata standard. In a sense,
profile = metadata content standard + extensions.
Profiles must meet the core requirements of the metadata content standard (that is, provide the mandatory elements that the standard requires) but can include extensions (described above). Since we also know a metadata content standard is composed of the core metadat set, a profile also can be thought of as
profile = core metadata set + optional elements + extensions.

The developers of most content standards expect and encourage the development of extensions and profiles, and may direct how they are to be specified and/or registered. A community that adopts a profile increases the interoperability of its metadata internally. It even increases its interoperability with communities that use other profiles, because the use of the core metadata elements is shared.

An important way that content standards may be constrained is through the use of vocabularies. Vocabularies can be used to fill out particular fields within the standard. The vocabulary used may be specified within the standard itself (for example, some fields in ISO 19115 define possible entries); or the standard may describe how to specify the vocabulary or vocabularies used (netCDF COARDS/CF allows users to specify the "standard vocabulary"); or the standard may be silent about vocabularies (the CSDGM is fairly open about how many fields are filled out). As noted above, extensions are a common way to narrow the options for filling out fields requiring textual responses. " From MMI sensor workshop report (http://marinemetadata.org/smireportdraftpdf).

Interfaces*:* To develop an interoperable system requires more than standardization on data and metadata formats. It also requires consideration of the interfaces to data catalogs or data servers that facilitate data transport between distributed repositories, and of the interfaces to services such as vocabulary list servers, unique reference systems (that generate unique identification numbers or strings for objects and data sets), and universal resource name resolvers (that can translate a URN to a web site, or to other information as appropriate). The specification for these interfaces includes the *transport protocols*, which describe how the connection is made between systems, and is likely to include a specification of the content that is transferred using the protocol. That content specification is analogous, and in some cases the same as, the content specifications described above.

Just as there are a wide variety of data and metadata formats currently in use, there are also a wide range of protocols in common use for interfaces (e.g. SOAP, REST, OAI-PMH, UDDI, WSDL, OPeNDAP, THREDDS).

General needs with respect to interfaces are for a well defined overarching architecture that is open for neighboring communities to access; consistent ways to discover data; coherent, consistent and complete standards with respect to a science domain; better tools to work with standards, and better collaborative tools that gracefully integrate appropriate interfaces, or can be used to develop new ones. Interfaces must be chosen and implemented appropriate to use requirements.

Registries*:* Registries provide searchable lists of ‘objects’, which are typically computation resources but may range from websites, to metadata, to data sets, to data systems. An overview of some existing registries relevant for marine, and more broadly geoscience, data are listed in Table 2. Registries for a variety of other kinds of ‘objects’ are currently lacking. For example, registries of Web Map Services, online KML resources, or of sensor information are all needed.

Registry	Objects	Services	Interface protocol	Metadata
GCMD	Datasets	WxS		DIF
STD-DOI	Datasets		SOAP	
OceanPortal	websites			
SESAR	samples		WSDL/SOAP	
Pangaea	Datasets		OAI-PMH	DIF, DC, ISO
WDC	Datasets			
GeoNetwork	Datasets		Z39.50	ISO, FGDC, DC
GeoConnections	Datasets			FGDC, ISO
SEDIS	Datasets	WMS	OAI-PMH	ISO
NDG	Datasets		OAI-PMH SOAP,REST	MOLES , FGDC, ISO, DIF, DC, CSML
OAIster	DOIs		OAI-PMH	
GEON	All	WxS	WSDL/SOAP	

Principles for selection. When selecting the protocol, content, and vocabulary specifications and tools for a community, consideration should be given to the needs of the community and characteristics of the available resources (specifications and tools). Factors to consider include the degree of adoption of each resource (within the community, and as a whole); the degree to which the resource describes or satisfies the characteristics of interest to the community, or can be extended to do so; and degree to which the resource will be used in automated systems. Another important consideration is whether the agreement is intended to come up with a working solution as quickly as possible, or is able to develop a solution that can support future growth of both the community and the larger environmental cyberinfrastructure. More capability is possible, and required, for systems to support anticipated advances in cyberinfrastructure.

There are several existing community-based efforts relevant to the selection and development of standards and protocols to support data exchange within the marine geoscience community. These include the SeaVox project (http://www.bodc.ac.uk/data/codes_and_formats/seavox/) and the Marine Metadata Initiative (MMI, www.marinemetadata.org). SeaVoc is a Vocabulary Content Governance Group, moderated by BODC (****Roy Lowry- would you like to add further description here?***). The MMI hosts a wide range of information on specifications and tools and encourages contribution of information developed by the community for others (in that and other communities) to use. They also encourage community projects, which are developing their own standards to consider using the MMI site to host their materials and publish their deliberations.

Recommendations

T3-R1. ***The community must minimize the proliferation of metadata standards and work toward a uniform approach for scientific metadata.*** There are two basic approaches to the problem of proliferating metadata standards; 1. develop a single uniform specification for scientific metadata, and 2. facilitate mediation or crosswalks between what is hopefully a limited number of different metadata standards. A single universal specification is unattainable, but a coherent, consistent, science-focused approach, ideally focused on building a minimum subset of profiles around a single standard, will limit the proliferation of profiles and ensure that the concept of developing crosswalks is viable.

T3-R2. ***The community must create agreed processes for community development of standards, guidance, and profiles.*** Governing structures are needed to enable the development of a community consensus about the overall standard(s) and approaches, and to establish processes to develop "official" extensions as needed for different specialized fields.

T3-R3. ***Community-based best practices for adoption of the ISO 19115 standard are required.*** As many groups within the global geoscience community are moving to adopt the ISO 19115 standard, there is a strong desire to avoid fragmentation and adopt a common solution to the problems of interpretation associated with this standard. To address these issues, a sub-committee of scientific data-metadata users needs to be established to come up with a best practice document with clear examples for application of the ISO 19115 standard (and ISO 19139). These guidelines would provide recommendations developed by the scientific community to resolve the interpretation ambiguities of the ISO standard, and make the current standard more portable between data and metadata centers.

T3-R4. ***New efforts within the marine geoscience community to develop standards and protocols to support interoperability should build upon and take advantage of existing efforts.*** Community-based efforts including the SeaVox project (www.bodc.ac.uk/data/codes_and_formats/seavox/) and the Marine Metadata Initiative (MMI, www.marinemetadata.org) offer relevant services, as well as forums for participation and contribution.

3.3.2. Session II: The 'Low-Hanging Fruit' for Data Exchange

The working group for session II focused their discussions on identifying opportunities for interoperability in the near future given the existing data resources within the global marine geoscience community. This group was asked to;

- *Explore realistic opportunities for the implementation of international data exchange*
- *Define a plan for easy start*

A growing variety of data resources relevant for marine geoscience research now exist within the international community. Each provides varying levels of data discovery and data delivery through their own custom search interfaces. At present, to find data of interest across these distributed data centers, a user must first be aware of all relevant data resources, visit each site, learn how to use the particular search interfaces provided (often in a language other than their own) just to determine whether data of interest exists at that data center. In contrast to the current scenario, what users desire is the ability to discover (and then access) data of interest seamlessly across distributed centers without the need for pre-existing knowledge of each resource and how to use their individual search tools.

The general consensus was that an achievable initial goal is to develop a data discovery resource across possibly a subset of the distributed and heterogeneous data resources now available within the international community. Discussions regarding how to implement a resource discovery interface focussed on the potential scope, as well as organizational and technical issues.

Scope: One approach for building a resource discovery-only interface would be to gather metadata from distributed resources across the marine geoscience community into a central repository (e.g. through the World Data Center system), which would build the discovery interface. Metadata could be gathered by harvesting from distributed data centers or through centers contributing to the central repository. The European Union has adopted the model of a central metadata resource through the SeaDataNet project (see http://www.seadatanet.org/**).** However, a centralized metadata repository for the broader global community is unlikely to be an optimal solution in the short term. Working group participants agreed that a more practical approach would be to identify a few select focus sites for building a discovery-only interface as a proof-of-concept (e.g. the MoMAR site on the Mid-Atlantic Ridge or the Nankai subduction zone). Existing international programs such as InterRidge, InterMargins, or IODP could be used to host the data discovery service.

Organizational Considerations: An organization structure for the discovery of marine data across the European community already exists with SeaDataNet. There are currently several marine geoscience data providers within North America, Asia, and the UK with significant data holdings that could participate to bring in a larger suite of resources across the global community. There was a general consensus that an on-line forum or process to support group collaboration would be valuable.

Technical issues include how to obtain the needed metadata from distributed resources. Metadata could be harvested by a central portal in an agreed upon standard format on a regular basis (e.g. like the standardized collection level metadata provided via the Cruise Summary Report within SeaDataNet). Harvesting is preferred over the submission of data by providers as it encourages them to invest in themselves and develop web services for their data resources. Some data centers have deployed OGC web services for serving some elements of their data holdings (e.g. PANGAEA, NGDC, IRIS, MGDS, PetDB). An alternative approach would be to serve metadata through the OAI-PMH. SeaDataNet will be using ISO metadata standards (19139). The broader marine geoscience community could move to adopt this standard to facilitate interchange with the EU community. As part of developing a common metadata standard, there is a clear need within the community to harmonize and map vocabularies for key parameters including platforms, devices/sensors and data types. In this context, the MMI initiative or SeaDataNet itself could play a role. IODE members usually offer services as well.

The advent of Google Earth/Google Maps as a tool for locating data is an attractive option for a community of distributed data providers to enable quick visualizations of location of their data resources. Data providers could provide a KML service with their collection metadata to show locations of their data, purely for discovery. Serving a visualization of the actual data through KML is also low cost ('this is what the data looks like') as existing images can be readily wrapped in KML (e.g. using PHP). However, the value of this service would depend on the data type and quality.

Recommendations

T3-R5. *Development of a data discovery service across globally distributed marine geoscience data resources is an achievable initial goal.* First steps should focus on metadata starting with the collection (cruise) level metadata (e.g. geographic extent, expedition info, list of parameters/data types, instrument, temporal extent). Sample, station, and track locations should be provided to enable data resources to be discovered in map-based searches. A few select mid-ocean ridge and continental margin test-bed sites could be adopted for building a proof-of-concept discovery-only interface. For example, the MoMAR site on the Mid-Atlantic Ridge, Nankai subduction zone, Costa

Rica or New Zealand margin would all be excellent candidate sites given the current interest in these areas within different groups of the international community. The existing relevant data resources for these sites can be readily identified. A simple search interface could be built which perhaps could be hosted for the international community at InterRidge, InterMargins or through IODP. An initial low-cost–of-entry data discovery service could take the form of a repository of existing KML resources with encouragement for others to offer their resources in this format.

T3-R6. ***Data centers should work to expose their data resources via web services e.g. OGC or OAI protocols.*** Workshop discussions pointed in the direction of the creation of Virtual Organizations (VO) in which VO members provide independent ways to expose their resources to generic portals. Web services enable data resources to be readily harvested by other services and provide scientists with the flexibility they desire to discover and access data in the front-end analysis and visualization tool of their choosing. A large and increasing number of geographical information systems can interface with OGC-compliant web services so that data from many different sources can be discoverd, visualized, inter-compared and analysed. Developing a "critical mass" of OGC-compliant services is an important strategic goal for achieving the vision of truly interoperable federated systems.

T3-R7. ***Development of special interest groups with tools to facilitate collaboration is needed.*** The range of experience and level of expertise/resources available to different segments of the global marine geoscience data management community varies widely. At the same time, technologies for information management are rapidly evolving. Development of special interest groups to share experiences and solutions and provide guidance would be very valuable for this community. An on-line forum or process to support group collaboration would be needed (e.g. Google, Elgg.).

T3-R8. **A dedicated task group is needed to harmonize and map vocabularies for key parameters including platforms, devices/sensors and data types.** There are existing processes that could be used, but harmonization is not a trivial task. In addition to facilitating interoperability between existing data centers, harmonization of vocabularies and development of a publicly accessible vocabulary service would be very valuable as new data resources are being built. Where possible, existing community efforts should be leveraged to advance this goal (MMI, SeaVox, SeaDataNet).

3.4. Theme 4: Opportunities and Obstacles for International Data Sharing

3.4.1. Session I: Archives and Contributions

As a first step toward identifying opportunities and obstacles for international data sharing, the working group assessed the range of policies and practices for data archiving and data access on an international scale, addressing the following questions:

- *What are the national, institutional, agency, and society policies with regard to data contributions & enforcement?*
- *Are available archives adequate? Are there orphan data types?*
- *What is the status of contributions to archives in practice? Do they need to be improved? If yes, how can that be achieved?*

Countries represented in the working group included Spain, France, Norway, the U.K., Japan, New Zealand, Taiwan, the U.S., Oman, Canada, and Germany. During a round table, working group members described – to their best knowledge - data policies of their country for the data types and data centers relevant to the workshop topic. Several issues of general note are summarized here:

- Data policies, where they exist, vary widely among and within countries and on all levels.
- Many countries still do not enforce contributions from individual investigators at private/academic institutions, even if official policies require it. One notable exception offered was the U.K., where a NERC-funded investigator was penalized for non-compliance.
- Government agencies typically have stronger policies and better enforcement than private/academic institutions, even in cases where investigators at such institutions receive government funding.
- Several countries have comparatively stricter policies for data within their Exclusive Economic Zone (EEZ), particularly while UNCLOS[3] mapping and claims are underway. In some cases, a country may require any research vessel traversing its EEZ, foreign or domestic, to submit a copy of all data collected.
- Overall, the situation has improved from 5-10 years ago. Improvements in technology such as faster network connections and larger storage systems have made it easier for investigators to post their data online and/or contribute it to data centers.
- In some countries, it is still often necessary to "know the right person" in order to find and obtain data sets.

Encouragingly, the overall trend in the last decade is toward greater openness in data sharing. Some countries still "guard" valuable data sets by imposing process fees and intellectual property claims, but there is growing consensus to build interoperable systems and to adopt data standards. An example is the recent series of E.U. initiatives including SeaSearch (2002-2006) and SeaDataNet (2006-2010). Also, recent natural disasters have caused some countries to more fully acknowledge the need for broad and open access to data. In the academic community, data management systems that are developed and operated through science initiatives such as the NSF Ridge 2000 program are recognized by their target community to provide a highly useful service.

[3] United Nations Convention on the Law of the Sea

Appreciation of such systems substantially contributes to a culture change in the science community toward more open data sharing.

While the number and variety of data centers all over the globe is continuously growing, anecdotal evidence suggests that many countries lack data centers for particular data types. Examples offered include paleoclimate data in the U.K.; ocean bottom seismometer data in France; undersea acoustics, marine seismic reflection, hydrology, and volcanology in the U.S., wildlife observations in New Zealand. Many countries also lack facilities for curation of physical specimens. Further, some data centers (or networks of centers) exist but are incomplete, such as sea level (tide gauge) data in the U.S.

With the growing number of data centers it becomes increasingly harder for scientists to easily find all the data in their area of interest. Perhaps the most significant, and universal problem with existing data centers (and networks) is a lack of standard registration and discovery. No mechanism is known to exist for truly comprehensive, interoperable, international search across global data holdings.

Recommendations

T4-R1: ***Uniform best practices and standards for data acquisition and data submission should be adopted on a global scale***. In order to achieve a higher level of data contributions to data archives and to facilitate the enforcement of data policies, ship operators and scientists world-wide are encouraged to adopt consistent best practices for data acquisition and submission. Metadata should be collected in a standardized way, and automated wherever possible. For example, where possible metadata should be encoded directly into data streams from sensors and other data acquiring devices. As part of best practices, formal submission agreements should be established between individual institutions and respective inter/national data centers in order to effectively aid regular, timely, and standardized contributions to data centers.

T4-R2: ***Real-time (field) data, processed data, and laboratory data products should all be archived.*** Experience has shown that raw data need to be archived because they become useful for applications that were not anticipated during the original acquisition. At the same time, it is important to also archive the processed data such as edited multibeam data, because in this form they are most useful to the broadest range of users beyond the specialists who are experienced with handling the raw field data types.

Archives for derived data (data products) are glaringly missing, especially for products, which are never formally published or are only published in print journals (unavailable online). Solutions should be explored to parallel efforts in other science fields (e.g. astronomy) to archive derived data products in collaboration with university libraries or journals.

T4-R3: ***The strong and continued interest in data from coastal waters should be leveraged to attract funding for data systems and standards.*** Strong interest in shelf mapping and benthic habitats within EEZs in particular can help increase resources to operate data centers and advance interoperability technologies and standards.

3.4.2. Session II: Implementing an International Data Network

A common vision, broad community support, an organizational framework, and resources are required to implement the envisioned international data network. In order to address these issues, the working group discussed the following questions

- *What levels of data access are desired? How can we achieve this? What technologies, agreements are needed?*
- *What are the appropriate organizations to advance the goal of data sharing (e.g. eGY, GEOSS)? Do we need separate ones?*
- *What funding is needed? What are the funding opportunities?*

Users ultimately want to have open access to all data for their field of research, easily discovered and accessed on a global scale via a central portal and downloadable for free in a common format that is supported by many standard applications. Modern internet and database technologies now allow to construct an overarching infrastructure/ architecture that can act as an umbrella network over all relevant data centers in the geological, geophysical, and geochemical community and that supports interoperability or uniform communication by applying technical and content standards to give users insight into its inventories and access to its data sets. The use of metadata, which document the data, give information on the originators and location of the data sets, and also inform about its quality, is fundamental, and the inclusion of metadata in any system should be safeguarded. Metadata is not only needed to ensure that data originators and original data centers receive their credits, so that both future users and funding agencies are made aware of their value, but are also essential to ensure the expected quality level of available data sets. Web services have started to be used for setting up overarching systems. This technology allows providing a standard programming interface (API), so that other users can build their own applications on top of the databases that are managed by the Web service. This is a development, which facilitates sharing and interoperability. However, it needs to be safeguarded that the references to data originators and original data centers are guaranteed.

The challenge at this point it to identify sensible ways to organize the development of a global data network and establish the right culture that allows such endeavor to move forward with support from both scientists and decision-makers. The scientific culture needs to become more open toward data sharing. Global and regional issues, developments in technology, more multidisciplinary research, and understanding of the benefits will be drivers for this culture change. National and international governments are stimulating more open data sharing and exchange by directives and guidelines, and in fact, an ongoing trend worldwide is apparent that institutes, organizations, and agencies are more open to data exchange than ever before. For example, in Europe the INSPIRE Directive calls for easy access to public domain data. At a global level, the GEO initiative by the

G8 countries promotes the development of GEOSS[4], considered as a System of Systems. Backed by policy makers and decision makers, these initiatives also help significantly to get appropriate funding for data facilities. Various reports on access to research data have been published recently, e.g. a report by the Organization for Economic Co-operation and Development (OECD) 'OECD Principles and Guidelines for Access to Research Data from Public Funding'.

However, data policies still exist and will remain in place at institutional or national levels that restrict access to specific data types and specific data domains, e.g. data sets from hydrographic surveys in the EU or military circles. At international level and in specific domains data policies have been formulated, which aim for openness, at the same time respecting local policies. Another way to overcome restrictions would be to create data products with specific resolution that make use of the original data sets for their production. These data products then can be made free and open and in most cases already satisfy the needs of the users. For scientific purposes, a moratorium period of 2 years is normal practice to ensure that the originating scientists can make exclusive use of their data for their scientific work. However also in those situations all data sets should be known by entries in the metadata systems.

Programmes and decisions of the large international bodies such as GEOSS, ICSU, and CODATA can serve as umbrellas and references for motivating an initiative for a global data network for geological, geophysical and geochemical data, because they help to make decision makers and policy makers understand the motivation and position. But programmes such as GEOSS cannot steer the development of the network itself. This has to be done by the players (data centers, science communities) themselves, who need to prepare a plan for an overarching network that will connect existing data centers and include a critical mass of players. This network will not only provide access to metadata and data, but also establish uniform vocabularies and standard protocols.

In order to start building this network, two approaches are possible: (1) Start building a network among a small number of World Data Centers to demonstrate its feasibility and thereby creating a snowball effect. (2) Start with a wider group of data centers, including the World Data Centers, to get more foundation and involvement at the institutional level, e.g. get all marine geoscience surveys and relevant institutes in the EU joining the plan, which seems feasible as part of the upcoming EU Calls for Proposals in the Framework Programmes.

Recommendations

T4-R4: *Advance a culture among scientists that is more open to data sharing.* Scientists need to agree that sharing of data is beneficial rather than harmful to research and knowledge circulation, so that they actively support and contribute to a data infrastructure that is based on open data sharing. This culture change can be advanced for example through practical examples and cases. For example, participation in international and multidisciplinary projects such as the EU Research Framework Programme or the International Ocean Drilling Program requires researchers to be more

[4] GEOSS = Global Earth Observation System of Systems

open towards data sharing. Participation in these projects is attractive to researchers because they provide funding opportunities at a time where many have to ‘fight’ for research funding. A critical aspect of advancing an open data exchange culture is that data infrastructure guarantees appropriate credit to data authors and data providers.

T4-R5: ***International programmes and bodies such as GEOSS, the eGY, and ICSU that stimulate the development of data sharing systems should be leveraged to promote our initiative for a global data network for marine and terrestrial geoscience.*** There are ongoing international research programmes such as CODATA, the eGY and the EU Frameworks Programmes as well as a number of international policies, adopted and driven by governments, that are intended to encourage and support international cooperation toward a global data infrastructure. International bodies such as IOC, ICSU, WMO, and UNEP, with membership at the country level, adopt and support these programmes and plans. Reference to these programmes and bodies is important to make decision makers and policy makers understand the motivation and position of an initiative for a global data network that is emerging from this workshop.

T4-R6: ***A dedicated task group should be established to advance the implementation of a global data network.*** In many regions of the world, the time seems ripe for starting to construct a network of data resources on an international level. A task group composed of the organizers of the workshop and some of the workshop attendees should prepare a plan for broad access to metadata and data by means of an overarching network that will connect existing data centers. In particular, the task group should

- formulate a precise definition of the aims and scope of the overarching system;
- prepare a matrix of relevant organizations from all over the globe that should be invited and engaged in its further planning and proposal making;
- explore options of funding programmes/agencies, e.g. NSF and EU Framework, and seek to formulate an overarching proposal, which is submitted to funding programmes by a range of proposals, which demonstrate clear interaction and complimentarity.

It will be helpful to study how ongoing international programmes, such as e.g. ODP or eGY, have done their preparations and have achieved their goals of getting international recognition and funding from various complimentary resources.

4. Next Steps

From the working group recommendations, the following actionable next steps are identified:

<u>1. Adopt test bed sites for development of a data discovery service across distributed data resources within the international community (T3-R5)</u>

Form alliances focussed on a few select mid-ocean ridge and continental margin test-bed sites (e.g. the MoMAR site on the Mid-Atlantic Ridge, Nankai or Costa Rica subduction zone or New

Zealand margin) where relevant data centers work to expose available data within these sites. A KML repository for these sites could be developed, perhaps hosted through InterRidge, InterMargins or IODP. A centralize resource or registry of relevant web services is needed.

2. Establish Forums for Guidance and Development of Best Practices

Special interest groups should be established to share expertise and solutions for development of interfaces and for metadata standards (T3-R7). As many groups within the global geoscience community are moving to adopt the ISO 19115 standard, there is a strong desire to avoid fragmentation and adopt a common solution to the problems of interpretation associated with this standard (T3-R3). A task group to establish best practices for implementation of this standard is needed. A dedicated working group is also needed to develop best practices for data documentation at sea (T2-R2; T4-R1). Existing procedures need to be assessed in light of data documentation requirements to establish guidance for routine shipboard operations across the global marine geoscience community, There is an immediate need and opportunity to harmonize and map vocabularies for key parameters including platforms, sensors/devices and data types (T3-R8). Interested data centers should form alliances building upon existing efforts to move forward on joint development of publicly accessible vocabulary services to facilitate interoperability.

3. Create higher level task force (**T4-R6)**

Assembly of a high level task force focused on forming international alliances among data centers within the marine geoscience world is needed. National and institutional marine data centers should work to align in parallel with current efforts within the terrestrial world involving National Geologic Surveys and efforts such as the FDSN, and join at high level.

4. Follow-up workshop in one year.

Building a global data network will require future and regular forums for the international marine and terrestrial geoscience community to meet, assess progress, evaluate new opportunities and define future directions

Appendices

1. Agenda
2. URLs table
3. Participant list
4. List of acronyms
5. One-pagers on data centers (online only)