G E B C O

GENERAL BATHYMETRIC CHART OF THE OCEANS

INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION

COMMISSION OCEANOGRAPHIQUE INTERGOUVERNEMENTALE

Paris

CENTENARY 1903 – 2003

INTERNATIONAL HYDROGRAPHIC ORGANIZATION

ORGANISATION HYDROGRAPHIQUE INTERNATIONALE

MONACO

GEBCO Data Flow Workshop
March 9 – 11, 2011
Boulder, Colorado, USA

Executive Summary

Seafloor mapping technologies have been revolutionized over the past thirty years, and GEBCO must also evolve to accomplish its goal to produce the best possible chart of the global ocean. The GEBCO product was traditionally based on hand-drawn contours derived from lines of soundings. This work was performed by dedicated individuals that devoted their time to gather, edit, and interpret the existing soundings in a particular region for compilation into the global product. In the intervening three decades, lead lines and single beam echo sounders have been replaced by multibeam echo sounders, which are continuously being improved, the arrival of satellite radar altimetry, which can be used to predict bathymetry at low resolutions on a global scale, and vastly improved navigation accuracy, which can be used routinely by a much wider community of seafloor mappers.

Perhaps the largest change in the past thirty years has been the transformation from analog to digital methods. The current seafloor mapping enterprise is composed of digital data acquisition systems on ships and satellites that provide digital soundings or altimetry in increasingly higher volumes but varying quality. Digital soundings, observed or inferred, are then automatically edited and compiled using statistical algorithms that project the depth estimates onto a mathematical surface that is then sampled on a predefined geographic grid. It is this global grid of depths that underlies the current GEBCO global chart.

The modern GEBCO effort requires an efficient work flow to accomplish its mission to produce the best global bathymetry. Traditional efforts to produce hand-drawn contours for compilation into new "Editions" must be replaced with a federated team structure that allows continuous contributions to a constantly improving model. A streamlined architecture is required that facilitates the flow of data from multiple collection systems, through editing, and compilation, without redundancy. Common tools and standards are required to further facilitate the interoperability of the various components of this architecture. Most importantly, an organizational structure must be developed to assure the flow of

data from observation to product is robust.

On March 9-11, 2011, a GEBCO Data Flow workshop was held at the National Geophysical Data Center in Boulder, Colorado, USA. The workshop was hosted and chaired by Chris Fox, GEBCO Vice-Chair, and was attended by fourteen member of the GEBCO community (Attachment 4). The goal of the workshop was to develop a plan to allow data, grids, products, etc. to flow in a predefined manner though the GEBCO community for incorporation into the compiled GEBCO product. The first day of the workshop was dedicated to a series of presentations from each of the component organizations that responded to a predefined list of questions concerning capabilities, methodologies, and tools. Day Two focused on formulation of a data flow model, with additional discussions of copyright issues, standards, methodologies, and tools. The final half-day looked at the proposed GEBCO-100 meter grid product and how it would be integrated into the work flow.

These extensive discussions resulted in the complex overall data flow model that is shown graphically in Attachment 1. There are two primary points of integration critical to this new process:

GEBCO Data Store: The production of new bathymetric data sets, whether grids derived from altimetry or soundings, clean multibeam surveys producing GEBCO-100m grids, new compilations from GEBCO regional mapping efforts, or other sources contributed to GEBCO will be stored in a common "GEBCO Data Store," which will be hosted by the IHO Data Center for Digital Bathymetry at NGDC. Unlike the broader bathymetric archives available at NGDC, the GEBCO data store will contain only those clean, processed data used to produce the GEBCO 30-second grid. All participants will have two-way access to contribute or access data in the GEBCO Data Store and will have access to RSS (Really Simple Syndication) messaging to inform participants of new data availability.

GEBCO Editorial Review Process: Before any contributions are incorporated in the GEBCO, it will be subjected to some level of editorial review by the GEBCO group of experts. The details of this process have not been worked out, but presumably a central editor (such as the GEBCO Bathymetric Editor or the SCRUM Chair) would make and evaluation of the contribution and either approve or provide to a GEBCO expert in the particular global region being evaluated. If the contribution accepted, it will be incorporated into the GEBCO grid and permanently included in the GEBCO Data Store.

Other discussions during the workshop: Many of the tools used by the participating organization are common, but some are not and may be shared where needed. Other topics discussed in detail were the use of Digital Object Identifiers (DOIs), standardization of tools, best practices, OLEX, attribution of contributors, use of the GEBCO Gridding Cookbook, and the methods to include the GEBCO-100 products into the global GEBCO grid. All of these topics were assigned follow-up actions to maintain momentum. The issue of copyright due to the inclusion of restricted data in the GEBCO grid is addressed by the creation of the fully accessible GEBCO Data Store. An action is underway to remove copyright text from GEBCO products.

Although the goals of the workshop were met: a GEBCO data flow model was formulated and agreed upon, there are still a great many tasks to complete before the full system can become reality. A list of proposed roles is included in Attachment 2, and a list of immediate actions was compiled in Attachment 3. A review of the workshop results and progress to date will be presented at the 2011 GEBCO meetings in La Jolla.

Notes

Wednesday, March 9, 2011

Chris Fox welcomed all of the attendees and thanked them for their participation. The goal of the workshop was to develop a plan to allow data, grids, products, etc. to flow in a predefined manner though the GEBCO community and be incorporated into the compiled GEBCO product. To support this effort, standard methodologies for data cleaning and gridding need to be agreed upon and a suite of tools to support this effort need to be identified and developed. GEBCO is the international organization responsible for mapping the sea floor, so tools to do this have to be identified and shared among the organizations who are participating in this effort.

At this technical meeting, each participating organization briefly described its capabilities and processes. From these descriptions, synergies and common areas for cooperation were identified. This will help accomplish the goal of developing and improving the GEBCO 30 arc second grid (GEBCO_08). One key part of this process to improve the data is the regional mapping projects.

The GEBCO Guiding Committee approved a new pilot project to create a GEBCO 100 meter grid. The development of this product was discussed and is due in October 2011.

Series of talks to describe each organization's data, tools, products, etc.

British Oceanographic Data Centre. Pauline Weatherall described the activities of the BODC in serving as the manager of the GEBCO Digital Atlas. This entails maintaining, updating and distributing the GEBCO gridded data and GEBCO's data products.

Although BODC archives data sets on behalf of GEBCO for grid re-generation purposes, it does not act as a public distribution centre for the source data sets used/submitted to generate GEBCO's grids.

Data flows into BODC for updating GEBCO's grids from a number of sources, e.g., from the GEBCO community including member country contributions, the regional International Bathymetric Chart (IBC) mapping projects, targeted data compilation activities, and commercial companies. Data received at BODC can be compiled grids, output from Electronic Navigation Charts (ENCs) and multibeam bathymetric surveys. Metadata is compiled for each data submission including the basic information describing the data. The submitted data are archived for grid re-generation purposes.

Using the software packages ESRI ArcDesktop GIS, Generic Mapping Tools (GMT) and IVS 3D's Fledermaus, BODC carries out quality control checks on submitted data sets. This includes checking for outliers and anomalies in the data sets.

Using "feather blending" routines from Global Mapper (GM) and the 'blend' algorithm from GMT, the submitted grids are merged into the existing GEBCO_08 Grid.

Martin Jakobsson commented that the GM routines should be incorporated into GMT. Weatherall thought that the incorporation of Satellite altimetry into the process would help the interpolation of the gridded data between soundings. The BODC gridding activity could be improved if there were defined/standard methodologies for updating the GEBCO_08 grid and for the metadata that documents the contributed data sets. Also, a standard method for distributing the GEBCO_08 Source Identifiers (SID) should be developed. She noted that a prototype Web Map Service (WMS) for this has been developed. Developing more WMS applications and eliminating the overlap of activities of GEBCO members should be considered.

Discussion. Barry Eakins asked how the regional updates are incorporated into the GEBCO grid. He

is especially interested in how higher resolution grids are used to update the lower resolution GEBCO grid. Chris Fox remarked that this is why you have an editorial process and that standard procedures need to be developed and formalized. This will help with the duplication of effort. Jakobsson said that there needs to be a tighter link between the IBCs and the GEBCO grid, especially in shallower areas and the Arctic and other ice covered areas. Various PIs have different ways of incorporating gridded data into their grids. David Sandwell noted that there are different "levels" of data (as defined by NASA for example) and one needs to learn to handle these types of data. Care should taken in handling the various datums used in the compilations. There should be a standard process for submitting data to the regional mapping projects. Walter Smith noted that the GEBCO Cookbook project will do just that. Martin Jakobsson recommended that this proposed process be kept simple or it may not be used. Colin Jacobs said the time and effort to implement this could also affect its use. Fox noted one outstanding issue is how to identify source data within the compiled grids (ACTION?)

UK's National Oceanography Centre at Southampton. Colin Jacobs gave an overview of his current duties as GEBCO Editor. He noted that his organization is undergoing budget and resource reallocations and that his time for GEBCO activities is becoming more limited. His current duties focus mainly on maintaining a close liaison with BODC (Weatherall), organizing the regional mapping projects, soliciting data from industry and updating the GEBCO grid areas. The Southeast Pacific regional mapping project is progressing well. Other regional bathymetric mapping projects need a sizable amount of assistance. For example, many involved in these projects have not heard of the tool GMT. Walter Smith recommended that the Nippon Scholars participate in these projects; it would be good training and would bring new data into GEBCO. Data at Southampton are received from various organizations affiliated with the UK's Natural Environment Research Council (including his home organization, the National Oceanography Centre at Southampton), and other European countries' surveys and industry as appropriate and as the opportunity arises. Jacobs pays special attention to areas where the GEBCO grid needs improvement and investigates how to achieve this with available data. Jacobs manually assesses the quality of the data and uses the tools CARIS, Fledermaus, ARCGIS and GMT to grid and check the data. He admitted that he can do a better job of documenting his data and relies on BODC to document the gridding process.

Discussion. Jacobs feels the biggest problem for the future is funding and defining the GEBCO activities at NOC. Having clear roles with no overlap in the GEBCO activities would go a long way in defining the GEBCO roles at the National Oceanography Centre (NOC) including the BODC. The quality control of the data is left to the PIs, so the compilation and the gridding of the data are the roles to be defined. Acquisition and use of available data is a GEBCO role, but who within GEBCO makes a request for the data, and what GEBCO component incorporates the data? This needs to be sorted out. Jacobs suggests using the regional mapping projects, and the upcoming regional Arctic/Antarctic mapping meeting is a good example of how to potentially do this.

NOAA/National Geophysical Data Center. Lisa Taylor described the data compilation activities of the IHO Data Center for Digital Bathymetry. By hosting the DCDB, NGDC receives data from a wide variety of organizations internationally including the GEBCO partners. Also, NGDC's involvement in national and other international programs results in data flowing into the data center. Data are assimilated into the DCDB data bases, documented, checked for gross data errors and outliers and placed into a long term archive. NGDC does not edit the data. Documentation includes FGDC, ISO and GCMD standard metadata descriptions. Barry Eakins noted that the data coming into NGDC is diverse and includes marine trackline geophysics and multibeam swath sonar data. NGDC distributes raw and processed data and integrated data products. NGDC has developed a number of state-of-the-art tools for distributing the data as Web Map and Feature Services, and interactive ArcGIS web maps which can identify the source within a compiled data set. There are no restrictions or copyrights on the data and the data are publicly available unless the data contributor requests restrictions on the data. In

this case, the data are used for specific projects such as compiling US coastal digital elevation models for tsunami inundation modeling.

Eakins described the Coastal DEM compilation process, noting that working with the data set's datums is a critical aspect of this process. He uses the GMT Spline Algorithm and MBSystem to grid the data. He reiterated his question about the techniques to consider in updating lower resolution gridded data using higher resolution gridded data. Eakins says he regrids the data when new data are received; Weatherall noted she edits the existing grid.

Discussion. In the future, Taylor would like to develop an enhanced IHO DCDB interface for data submission, display, search and retrieval; increase collaboration/coordination with IHO Member States for data submission to IHO DCDB, and develop ISO standard metadata for all data in the DCDB Eakins noted that NGDC is receiving large data submissions and better documentation is necessary for it to be processed efficiently. He observed that gridded multibeam and lidar data are more useful than raw data, but having the source data is necessary for editing out gross errors. David Sandwell asked what is done with the bad data. Eakins said that data are flagged and not used, but remain in the source data as flagged.

U.S Naval Oceanographic Office. David Fabre described the sources of data NAVO receives which include data from Navy ships, data associated with international agreements and other sources such as data mined from the internet. Data are stored in variable resolutions in their data warehouse in an Oracle database with spatial and temporal tags. Data are referenced to the WGS-84 ellipsoid and are in GSM (Generic Sensor), PSM and BAG (Bathymetric Attributed Grid) formats. They are experimenting with a PFMWDB (Worldwide DataBase). This would be a variable resolution database of grid layers. The BAG format optimizes data exchange and includes metadata, elevation data values, uncertainty, a tracking list and digital certification. The BAG format is open source and can be downloaded from University of New Hampshire. Processing is done with NAVO legacy programs, PFM (Pure File Magic), SABER and Fledermaus. Some of their problems are too much and too little data depending on the area and who collects it. Current development efforts are researching bathymetry fusion, developing DUES implementation, looking at merging DBDB with the ESRI products and multi-solution gridding. Some issues for them include subjective area based editing ("dot killing"), extreme volumes of raw data to be processed and best resolution grid database development.

Discussion. The Digital Bathymetric Data Base is mostly made up of Mutibeam surveys. Sandwell asked about the dot killing techniques; how does one edit out the outliers if you don't use the dot killing technique.

University of Stockholm. Martin Jakobsson gave an overview of the processing flow of the data for the International Bathymetric Chart of the Arctic Ocean (IBCAO). Much of the data used in this regional compilation are from sources associated with the University of Stockholm. Other sources are from the region, international sources (e.g., NGDC) and private companies. All the raw data are processed and placed into an Oracle database. One problem that they have is storing hundreds of millions of data values including the associated metadata. Data are stored in one degree units. A GIS is used for data analysis using block median filtering and continuous curvature spline. Gridding is done using a new GMT functionality, the Spline Pyramid. Not much single beam data are used anymore, mostly multibeam data. Data from the University of Stockholm's Oden icebreaker using the multibeam echo sounder and chirp sonar EM 122 will greatly contribute to the IBCAO regional mapping project. The data will be incorporated into the Oden Mapping Data Repository.

Discussion. Jakobsson emphasized the coordination with GEBCO as a key part of his data activities. Acquiring new data sets and incorporating them into IBCAO and then into GEBCO grids is the logical sequence. How is the best way to do this? There are politics and a lot of work involved in the acquisition of the data but the technical aspects are known. How do we evaluate grids that go into the GEBCO grid? Walter Smith observed that we don't want to reinvent the process and that GEBCO should use good gridded data, but should not lose sight of the source data which can be critical for cases like sea mounts, and other special areas. In areas where there are "black holes", i.e. no data, it is best to use satellite altimeter data. Fox asked how much of University of Stockholm's methodologies can be used by others. Perhaps GEBCO might be able to convince ESRI, Fledermaus, etc., to incorporate these systems into their software. In closing, Jakobsson announced the upcoming meeting *IHO/IOC GEBCO Arctic – Antarctic Seafloor Mapping Meeting 2011*, May 3-5, in Stockholm.

Lamont Doherty Earth Observatory. Suzanne Carbotte introduced the Global Multi-Resolution Topography (GMRT) which has 11 resolutions down to 25 meters, uses gridded tiles for global coverage, is dynamically maintained and includes attribution to the source data. The user experiences a seamless transition between resolutions based on chosen zoom level, going from 100m swath data and grids to global coverage. Attribution to source data includes comprehensive imbedded documentation and ping files which trace back to the sources. Sources of the data are terrestrial and oceanic and include satellite altimetry from SRTM, ASTER, BEDMAP, Smith and Sandwell v12, regional map data from IBCAO v2, high resolution multibeam swath data from ~500 cruises, and other contributed grids from US and foreign scientists and data centers. GMRT data can be disseminated in many formats for client applications including NetCDF, ArcGIS, and Fledermaus. LDEO has web tools for the user to make custom products. GMRT is also implemented as an OGC Web Map Service and is compatible with Google KML. All source data is documented in the Marine Geoscience Data System. Swath and gridded data QA is done using MBSystem. For gridding the data in-house tools as well as MBSystem are used.

Discussion. Using the Haxby gridding process, each cruise is gridded at 100 m and then is decimated to a lower resolution. Carbotte noted that if no data exists in a 100 m cell, then this cell is not used in the gridding process. How is down scaling done and how is it noted in the gridding documentation? Updating grids is dependent on the resources available. It was stated that Google will publish the 100 meter grid on Google Earth. Jakobsson asked about periodic updates versus continuous updating of the grids. Obviously, if you have continuous updating you have more versions than if there are periodic updates. Carbotte would like see improved methodologies for feathering with regional grids. The pilot GEBCO 100m is due in September 2011. How does it interface with GEBCO 30"?

Scripps Institution of Oceanography. David Sandwell outlined the methods used to construct the Smith and Sandwell predicted depths (1 minute Mercator projected grid) as well as the SRTM30_PLUS grid (30 arc second latitude longitude projected grid). Walter Smith added significant clarifications and additions to the presentation. The SRTM30_PLUS grid forms the basis for the GEBCO08 (V5.0) grid as well as the Google Earth grid (V4.0). The SIO group is working on V7.0 of the SRTM30_PLUS grid which contains a significant amount of new data. The main focus of the SIO effort is to construct global bathymetry at 1-10 km resolution for scientific research. To do this, they acquire data from a wide variety of sources with a focus on single beam soundings in the deep ocean but would like to incorporate more multibeam data in the future. SIO has approximately 7,600 unique data sources (ship tracks, point soundings, subgrids). The data are stored in CM format (an internal format) and a CM Editor is used to flag bad data through a comparison with predicted depth. Each cruise has its own Source Identifier (SID) as originally developed by the SIO/NOAA/US_NAVY/NGA group for the

construction of the first MOU grid. While constructing the global bathymetry/topography grid, a complimentary grid of the SID number used to constrain each grid cell is also constructed. This SID grid serves two functions. When anomalies are visually identified in the bathymetry grid, the number in the matching SID grid can be used to identify the CM-file containing the bad soundings. The raw sounding data for this cruise can then be inspected and outliers flagged either using automated or visual editing tools. In addition the SID number is linked to the source of the data as well as any metadata. There was discussion of how the SID grid could be used to attribute data sources at the pixel level, although this was not a popular idea.

The gridding process is done by using GMT, ER Mapper, and in house programs. Approximately 7 student years have been spent flagging bad data points in the 7,600 CM-files so this effort should not be repeated. When updating the grid, a global re-gridding is done which takes approximately 20 hours on a desktop computer. Not all of the tools used are transportable. A Mac version of ER Mapper would be welcomed.

SIO publishes a Seamount Discovery Tool as an overlay to Google Earth. It has been used on two multibeam transits to survey interesting features while avoiding already-surveyed seafloor. Currently it lacks information from many of the US multibeam cruises so needs to be updated (on the SIO to-do list). In addition, it will require periodic updates to include swath lines from new publicly-available sources.

Discussion. In the future, Sandwell and Smith will construct a new marine gravity grid based on new altimeter data being acquired by CryoSat and Jason-1. The raw CryoSat profiles are 1.4 times better than Geosat and if the duration of the CryoSat/Jason-1 data exceeds 3 years, the new gravity field will be 2 times better than the gravity available today. The new gravity will be used to update Smith and Sandwell predicted depths and extend coverage to 88 degrees latitude. Fox asked how much better the global bathymetry grid would be. Sandwell estimates that this will improve the global grid by a factor of two. He says that as many as 20,000 new seamounts may be found. Over the next 3-5 years as the gravity field is being improved, these improvements will be folded into the 1-minute predicted depths and then into the SRTM30_PLUS. The question is how and when these updates are folded into the GEBCO grid and how to update the voids in the regional GEBCO contributions that may have been previously based on outdated predicted depth. (Note IBCAO does not currently use predicted depth so this is not an issue in the Arctic.) Sandwell noted that there will be other innovations in the construction of global grids such as adding uncertainties to the source data and developing anisotropic gridding based on the paleo-spreading direction. He believes convergence on a single global grid will stifle these innovations and that it is healthy for the community to develop multiple products. This seemed to be a minority view among the other participants at the meeting who think multiple global grids will confuse the user.

One issue that was discussed is that both BODC and SIO are increasing their data holdings and each group creates new SID numbers independently. This will likely result in duplication of SID numbers. Carbotte noted that it is really the cruise name that is unique so this should not be an issue. The SID numbers are really a programming issue and overlapping numbers can be corrected later.

NOAA Center for Satellite Applications and Research. Walter Smith reviewed the Laboratory for Satellite Altimetry data activities. They acquire data from a wide variety of sources including government, academia, the Web and GEBCO members. Some data are easier to incorporate than others, especially noting that data from JAMSTEC is very easy to incorporate into his database. The data are stored in simple file structures with ASCI text file for documentation. They rely mostly on the

data providers (PIs) for ensuring the general integrity of the data. They do not usually redistribute the source data. They quality check the data by comparing input xyz and output grids against new multibeam data not previously ingested. Data documentation is by research papers and technical reports. For gridding the data in dense or full-coverage areas, GMT is used. In areas with sparse data, satellite altimetry is used. Updates to the grid are done on an ad hoc basis. The limitations of the methodologies and tools for satellite altimetry are that the spatial resolution is limited and equal weights are given to all constraints. Also, new algorithms have been written but have not been tested with the best data. In the meantime, products are being developed using older versions of the algorithms.

Discussion. In the future, Smith would like to see ways to incorporate and estimate uncertainty and an improvement of the satellite altimetry inversion by determining the best software coding. Having portable software code is also desirable. He would also like to see source attribution and track line depictions for the Google Earth's bathymetry. He noted that lack of resources will be a major constraint in the future. He needs expertise in database management and GIS. This way new databases could be brought into their gridding process. He then recounted a general history of the development of satellite altimetry.

Wrap-up Discussion. Fox said that methods for satellite altimetry are still evolving and getting better. Where there is excellent bathymetry, it should be used and not regrid globally. Sandwell asked why not? The good bathymetry grids should be used as input to the satellite altimetry grids. Fox said that the good bathymetric grids should not be replaced. Jakobsson does not want several different versions of the same global grid. All of the good data should be used to make one good product. Fox said this data flow and integration is the next topic for Day 2.

Thursday, March 10, 2011

Chris Fox opened the workshop by summarizing the previous day. He noted that all of the GEBCO organizations that presented have much capability in all aspects of seafloor mapping. There are many tools used and varied processes involved. He proposes that a well defined and agreed upon process for generating GEBCO products would be beneficial by minimizing overlaps and efficiently optimizing resources. High on the list is developing standards and defining a process for updating the GEBCO products.

A list of discussion points collected from the previous day was distributed to be used for much of the Thursday discussion.

Fundamental question: What is the overall data flow, who does what and how does it all fit together?

This discussion includes the development of a Data Flow Model. A draft Data Flow Model was distributed. (NOTE: the final version of the Data Flow Model is Attachment 1). Jakobsson stated that the data are transferred to BODC after it is collected. Fox commented that the GEBCO community needs to know that this is the case. This is important to determine if the data are being archived.

Archive discussion. In any archive discussion, we need to think for the long term. There are multiple archives so it is difficult to determine if the data are being properly archived for long term. Fox asked how this works for regional mapping project data. An agreed upon process needs to be developed.

Jakobsson said they put a copy of their source data in the DCDB and have a working copy in house. By doing this, there is an official GEBCO copy available. It was discussed that this could lead to many versions of the source data (and even grids) being available from multiple organizations. Jakobsson noted that for his regional mapping project he only gets gridded data externally; the multibeam data used is collected by his institution. Sandwell recommended building up an archive of the GEBCO data. He has an in-house archive of the data he uses to make his products. He does not think it is possible to have "one" grid. Fox asked if more than one definitive grid is needed. Sandwell said that it was not bad to have more than one. Jakobsson observed that GEBCO must make the data readily available or lose the trust of the data contributors. But he cautioned that it is imperative not to let restricted data be disseminated or lose the trust of the contributors also. He supports "one" GEBCO grid. Smith also supports a GEBCO archive of data. The technical aspects are solvable, it needs to be done. Sandwell notes that there are many copies of the GEBCO database. Taylor said that there is one at NGDC, BODC and Scripps. Sandwell observed that there are possibly more than three and this may cause the "GEBCO" name to be tainted. Fox pointed out that GEBCO is the only real option for international cooperation for global bathymetry. Jakobsson thinks there should only be one GEBCO 30" global grid. Carbotte strongly recommended preserving the source attribution especially if GEBCO is to get international contributions. Fox stated the GEBCO mission should be to produce the definitive global gridded bathymetry database. GEBCO needs to decide if we are going to work on the process of getting a definitive 30" grid or not.

Development of Data Flow Chart (Attachment 1) – Data are collected or come from PIs, next are archived at the DCDB; a working copy is maintained at BODC, LDEO, SIO and elsewhere as appropriate; the regional mapping projects (IBCs) get the data needed. The regional mapping projects will get data directly sometimes, but will need to get the data usually from DCDB and/or BODC. Jakobsson asked if regional mapping projects can send restricted data to DCDB. The answer is yes. Smith commented that he does not want to go to lots of places to get data. Sandwell was concerned that diagram would leave out organizations; he suggests to make it simple and to make it easy to join the GEBCO federation. NGDC agreed to host one GEBCO archive center. It was noted that the regional mapping projects get a variety of data to compile the grids, including grids of satellite altimetry. This then becomes that part of the GEBCO grid. The data flow sequence is raw data go to the regional mapping projects and then the regional grids get incorporated into the GEBCO grid. Raw data are archived at DCDB, the data needs to go to the GEBCO editorial board then produce the definitive 100m and 30" grids.

Jakobsson asked how to submit data to archive? Fox responded that there are established processes for doing this. NGDC has established data submission agreements which set standards for sending data to the archive. Raw satellite altimetry is archived by producers (PIs). A need for an archive for the satellite data grids was stated. Carbotte noted a need to solicit data for regional mapping projects, but some PIs won't want the data to be archived at the IHO DCDB. These data are needed for the development of the regional maps. There was an expressed desire for being very clear about how this archiving process works; i.e., restricted data can be retained by the regional mapping project to develop its products but care needs to be taken not to have these data publicly available. Smith said that there is a need to give attribution to the derived gridded data like what is done in Google Earth. Jakobsson said it should also track back to source data.

What do we do with contributed grids? The data goes to iSCRUM and BODC, and then goes through the editorial process. BODC then contacts appropriate GEBCO partners and sends data to the appropriate place for incorporation.

What is the process for updating (e.g., blending data) the GEBCO Grid? How do we integrate satellite data into the update process?

The keys are the regional mapping projects. They should use common technologies, e.g. the GEBCO Cookbook, and standard processes. Questions to discuss include 1) How will this be used in regional mapping?; 2) How will the data get to the GEBCO archive?; 3) How are these procedures standardized in the regional mapping?; and 4) How are the Nippon scholars involved in this process? Recommendation to have Jacobs review (visit) the regional mapping projects and assist in implementing standard procedures. Fox suggested that GEBCO could fund standard hardware and software for this. Jacobs said using the Nippon Scholars may not be possible as they may not be available. Recommendations: since regional mapping projects are key in the development of the GEBCO grids, GEBCO needs to 1) aggressively work to implement the IBCs; 2) use Nippon Foundation funds to implement the GEBCO Cookbook in the regional mapping projects; 3) have a GEBCO person take the lead.

What metadata should be required and what are the evaluation criteria for the data that gets merged into GEBCO product? --- Need simple standards and requirements for the grid contributors.

Collection level metadata is needed to document the GEBCO grids. Recommendation to build on what NGDC has developed. Carbotte noted that there is a minimum and a maximum of information that can/should be included in metadata. Post the metadata schema to web site and coordinate – Action, Eakins. Carbotte noted that much of the data has documentation/metadata so having it visible in an inventory should be our goal.

What is purpose of a GEBCO SID? Is it needed or not needed? Should there be a unique SID? Should there be a common metadata catalog link? How do we identify the source data in the GEBCO grids?

Source Identifiers (SID) are useful for helping to sort out periodic updates of the data. Sandwell said that each of his sources (cruises) has a unique SID. Discussion on how to implement and deal with the gridded data. Recommendation to make a database of SIDs robust for future applications. It is a big task to put SIDs in the GEBCO grid, but since we already know what the underlying source data is, much of this information is already there. (Post meeting comment by Sandwell: "I think it is a trivial task to put the SID grid into the GEBCO grid since we know all the source data that went into every pixel of the grid.")

How do we share tools, best practices, methodologies, etc. (e.g., Spline Pyramid). What is the process/mechanism for sharing?

Regional mapping projects are the key to the GEBCO grids so sharing common tools for these projects is a good way to do this. National hydrographic offices are too cautious and bureaucratic to readily share these tools. Jakobsson suggested that academia is more flexible to sharing. A partnership is needed. This workshop, the GEBCO Cookbook, and the NAVO-NGDC GSF are good examples of sharing. Recommendation to share the Spline Pyramid with GMT and GEBCO.

Is it better for a continuous updating of the grid or it is better to have periodic updates?

There will be a distributed archive which will have versions of the same data.

The Gridding Cookbook should be used for regional mapping compilations. How do we make sure it is used?

The regional mapping projects should be made aware of the GEBCO Cookbook at the beginning of the project. It will be available on the GEBCO website. Fox suggested that it be part of the Nippon Scholars' training program.

All workshop attendees commended Karen Marks for a job well done in developing the GEBCO Cookbook.

Should OLEX Fishing map data be used by GEBCO for incorporation into its models? How to do this economically?

This was endorsed enthusiastically. Jakobsson suggested that the OLEX data could be used as backdrop/fill-in for areas in the GEBCO grid where the data are missing. He will investigate its availability for GEBCO. Fox will send inquires to the Antarctic cruise ships about their data also.

Discuss the issues with using different altimetry models for GEBCO.

Sandwell and Smith will collaborate on the details on how to input satellite altimetry into the GEBCO grid. But they also need to look at other compilations. The objective is to have one product from Smith and Sandwell. Smith asked if there should be different versions as they work through the process. Jakobsson suggested a test with the North Atlantic data to see if satellite altimetry can improve it. – Action: Jakobsson, Smith, Sandwell.

Should there be one definitive GEBCO grid, combining all methodologies to create one grid? Proposal for distributed archive of cleaned up ship sounding data.

There was a discussion about attribution for gridded data in the GEBCO grid. It is doable, but has resource implications. Other issues include what type of attribution and are the bad data attributed? Recommendation: Gridded data should be attributed using the GEBCO Viewer with polygons identified as to source. Use this for the GEBCO 100m and the GEBCO 30" products. There could be another layer which identifies the source data. Weatherall will review the source inventory using their Web Map Service. Conclusion: 1) BODC will continue to produce the GEBCO Grid. They know where data are constrained. 2) The Viewer will have polygons showing the information for the project data that went into developing the GEBCO Grid. Carbotte suggested that it also show the National contributions. There could be one layer for grids and one layer for track lines. Eakins handed out an example of source attribution being used for United Nations Convention on the Law of the Sea Extended Continental Shelf mapping.

GEBCO Publications for the various GEBCO products - the use of Digital Object Identifiers?

Each GEBCO grid and input grids would be assigned a DOI. Therefore the next version of the GEBCO grids will have a North. Weatherall will lead the effort. There will also be a group authorship of the GEBCO members. Carbotte asked if there will be different versions of the GEBCO grid. Weatherall said it would stay the same as it is now, a version every six months (approximately). It will still be called GEBCO_08, but different version. i.e., month and year is the version.

Open Data Policy/Copyrights

Sandwell said that the biggest problem he has is the classified and propriety restriction on the availability of the data. He proposed setting up a distributed archive of cleaned ship soundings. The data in this archive would be available to all without registration or agreements. All of the members of the GEBCO Data Flow meeting supported the depositing of their data in a common archive and make it publicly available. In addition, Jacobs and Sandwell agreed to look into removing all copyright text from their products and web pages.

Friday, March 11, 2011

Chris Fox opened the day by distributing a new version of the Data Flow Model (Attachment 1). He stated that restricted grids are not in the GEBCO grid. Also noted was that ships' track lines should be a major part of the Regional Mapping projects. Taylor said that SCUFN input to the GEBCO grid was through the DCDB. Fox will present the results of this workshop to the GEBCO Guiding Committee in October.

Data Flow Model Question was how to ensure that the data gets to the right places as noted on the model (chart). This is a challenge, but this is a process which is solvable. Primary to this is to put the data in the GEBCO Data Store (GDS). Sandwell recommended that the store be open and data readily available. There will be an RSS feed announcing new data in the store. Would there be restricted data in the store? Yes, but only accessible to GEBCO members. This is a design issue. Fox asked how the Data Store would work. Would Jacobs and Weatherall grid raw data? Data would be gridded only if there is no gridded data for that area. One possibility is maintaining the GDS at the BODC with a back-up at NGDC. Including lake data in the Data Store is possible, but there is nothing official currently.

The Roles in developing and distributing GEBCO Grids were distributed (see attachment 2)

Developing GEBCO-100m prototype before next GEBCO meeting

Vicki Ferrini described a LDEO proposal to develop the GEBCO 100m global gridded bathymetry database. This would give GEBCO two core products, the GEBCO 100m and the GEBCO 30" global gridded bathymetry. LDEO proposes to contribute its 100m database to the GEBCO 100m grid. It would act as the International Data Assembly Center and provide tools to the GEBCO Editorial Board to assimilate and evaluate the data as it is sent to them from LDEO. This will be an iterative process. After the GEBCO 100m grid is developed, the data would be available to be assimilated into the GEBCO 30" database. LDEO would develop the GEBCO 100 m database using swath bathymetry and existing gridded bathymetry. LDEO has a large data base of swath data including the NSF Ping files and the GMRT. Characteristics of the GEBCO 100m are that it is dynamically maintained (new versions will be periodically distributed as updates are incorporated). Also the 100m database will have all the data attributed to the sources. The goal is to develop a working prototype by September 2011. Ferrini noted that this project will be a technical proof of concept.

Discussion. John Cartwright asked about the gridding routines. Jakobsson said that they should not be a problem especially in deep water and that at 100 m resolution multibeam data is less of a problem. Fox asked about the method to take the 100 m data and merge it with the 30" grid. There are issues including the projections used, e.g., Mercator, Polar Stereographic, etc. The data developed for the 100m projects need to get into the Regional mapping projects as well as to Sandwell and Smith for incorporation into the satellite altimetry data base. Perhaps it might be better for LDEO to send the

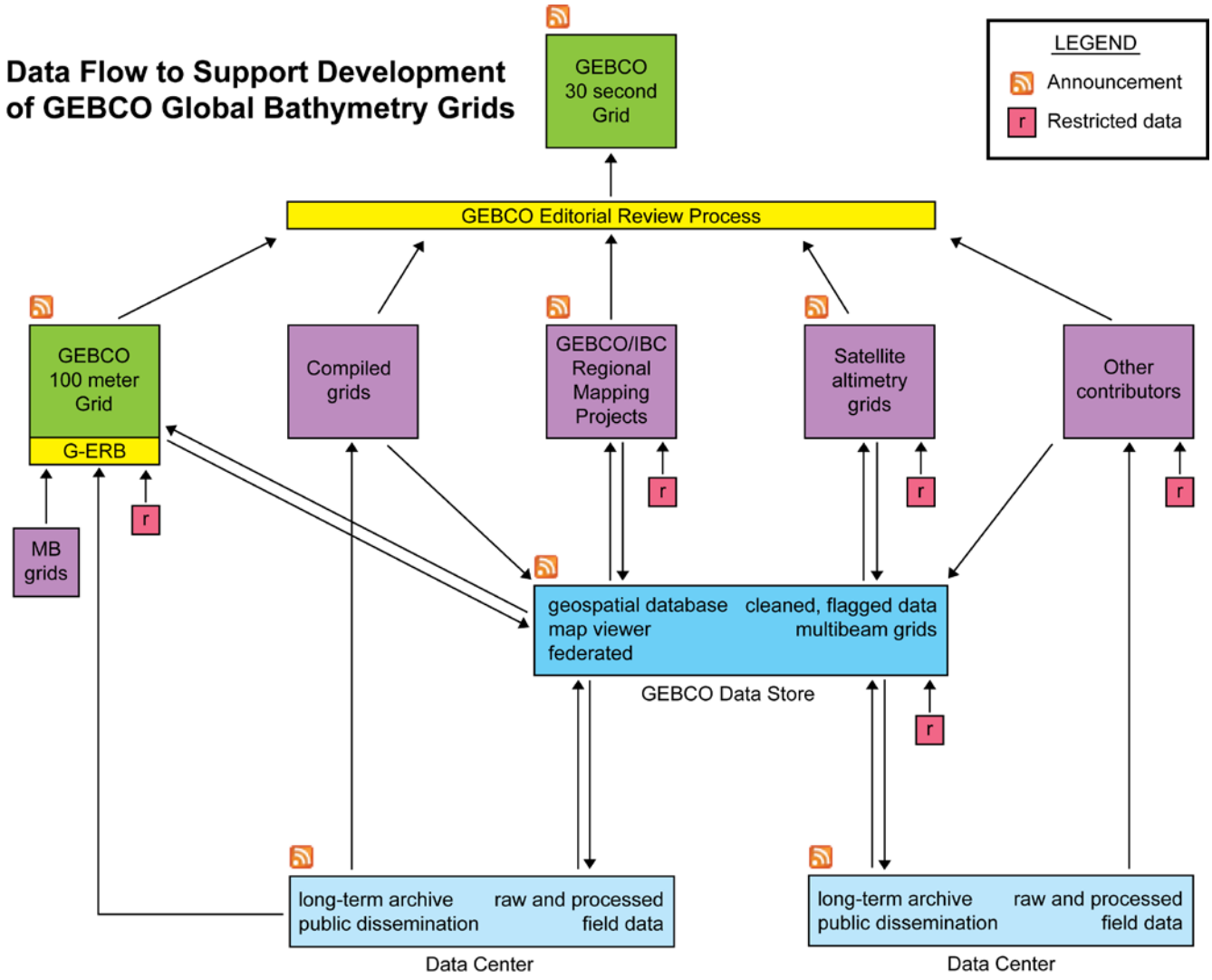
100m data to BODC and have them distribute. Ferrini and Weatherall will test this during the prototype stage. (Action). Cartwright asked if there are standard grid tiling schemes. Jakobsson noted that there are many schemes. It was recommended that LDEO and NGDC should have the same scheme (Action - Cartwright). Having a standard tiling scheme might encourage more data contributions to the project. How to get data into the 100m project was discussed. There needs to be an editorial process and perhaps a Web Map Service needs to be developed to support this. Jakobsson urged that this project is linked to the regional mapping projects. Fox recommended that a good presentation and demonstration of the prototype be developed for the customer, i.e., the GEBCO Guiding Committee. It should be something impressive, such as something using a Web Map Service and/or Google Earth. Sandwell commented that there might be better participation in the project if required user web site registrations were eliminated.

The Action items from the meeting were reviewed and accepted (Attachment 3).

In closing, all agreed that the meeting was very useful and informative. Smith commented that the technical aspects were more detailed than can usually be realized at the annual TSCOM meeting which is held in conjunction with the GEBCO Guiding Committee meeting. Fox noted that this meeting is good planning for the annual GEBCO meeting. We will try to have more of these in the future and will try to get some funding for them. Fox thanked all for their participation. Participants thanked NGDC for hosting the meeting.

Attachment 1

Data Flow Model



Roles in developing and distributing GEBCO Grids

- Maintain web site [BODC]
- Integrate compiled products and approved, cleaned data [BODC]
- Review contributed products and evaluate and approve each new GEBCO grid release [GEBCO Editorial Review Board]
- Develop and support contributor attribution [BODC, NGDC, LDEO, Smith/Sandwell, SRTM30_PLUS]
- Build global 100-m cleaned multibeam grids [LDEO]
- Build global altimetry 30 arc second grids using blockmedianed data in GEBCO Data Store [Smith/Sandwell]
- Build regional compilation grids (e.g., IBCAO) using “working store” pulled from GEBCO Data Store [ISCRUM lead]
- Contribute other compiled products [e.g., NGDC (Coastal Relief Model)]
- Build GEBCO Data Store [NGDC, using rsync]
- Manage GEBCO Data Store [NOC]
- Select data in public data centers for use in GEBCO grid, and clean and ingest that data into GEBCO Data Store [NOC]
- Manage, archive and distribute field data (raw and processed) [IHO DCDB, BODC, etc.]
- Solicit other agencies, organizations for bathymetric data and products to contribute [IHO DCDB, “GEBCO”, GEBCO Regional Teams]
- Manage restricted data [each agency responsible for the data]

Attachment 3
Action Items

Number	Action	Name	Due Date
DF1 – 01	GMT surface needs to incorporate Martin/Benjamin's fix (Spline Pyramid)	Smith & Jakobsson	October 2011
DF1 – 02	Source identifiers need to be synchronized and eventually replaced	Weatherall and Sandwell	In progress
DF1 – 03	Use of high-resolution grids to update low-resolution grids	Eakins will test	In progress
DF1 – 04	Publications and data citations, use of DOIs	Jakobsson and Weatherall	In progress
DF1 – 05	Plan for the design of distributed archive	NGDC	In progress – testing rsync
DF1 – 06	OLEX subscription update	Jakobsson	In progress
DF1 – 07	Discuss OLEX with Antarctic cruise ships	Fox	Complete
DF1 - 08	Facilitate regional mapping; Send Gridding Cookbook to Dave Monahan	Marks	In progress
DF1 – 09	Metadata schema for grids	Eakins and others	In progress
DF1 – 10	North Atlantic test altimetry algorithm	Smith, Sandwell & Jakobsson	In progress
DF1 – 11	Update flow chart	Taylor and Eakins	Completed 3/11/11
DF1 – 12	Make sure NGDC GDAL grid-data comparison and SIO remove/restore (grid update) scripts get into GEBCO cookbook	Marks and Eakins	in progress
DF1 – 13	Design/understand tiling scheme and build map service	Cartwright	pending
DF1 – 14	Design prototype webpage that includes the viewer for GEBCO 100m grid	Cartwright	pending
DF1 – 15	GEBCO webpage should include links to regional maps, SRTM, etc.	Weatherall	Partially done? – see GEBCO's web site
DF1 – 16	Write a short workshop report to circulate to group	Clark	Completed
DF1 – 17	Look into alternatives to current registration requirement	Weatherall	In progress
DF1 – 18	Develop workshop presentation	Fox	10/1/11
DF1 – 19	Send altimetry grids to IHO/DCDB	Sandwell	In progress
DF1 – 20	Send clean Scripps data to IHO/DCDB	Sandwell	Completed 3/11/11
DF1 – 21	Post wish list/requirements for submitting grids to GEBCO on the GEBCO website	Weatherall and Taylor	pending
DF1 – 22	Test compilation and distribution of GEBCO 100m data, including dissemination of data to GEBCO partners	Ferrini and Weatherall	pending
DF1 – 23	Reciprocal removal of copyrights at UCSD and BODC	Sandwell and Jacobs	In progress

Attachment 4

Data Flow Workshop

AGENDA

Wednesday, March 9, 2011

8:30 Welcome, Opening comments Chris Fox

Series of talks to describe each organization's data, tools, products, etc.

Guidelines for Workshop Presentations

Please focus your presentation on data, data flow, methodologies and tools that can be used to support the GEBCO product. Each presentation should be no more than half an hour, leaving 15 minutes for detailed discussions and should address the following questions:

- Where does your organization obtain data?
- How do you manage your data?
  - What tools and methods do you use to clean and disseminate your data?
  - How do you assess the quality of your data?
  - How do you describe and document your data?
- How do you grid your data?
  - What is your gridding process?
  - What tools do you use?
  - How do you evaluate your grid?
  - How do you update your grid?
  - How do you document your grid development?
- What are the limitations of the methodologies and tools that you currently use?
- What capabilities and functionalities are on your wish list?
- Where do you see your efforts going in the future?
- What issues do you face that complicate your effort?

9:00	BODC (GEBCO Compilation)	Pauline Weatherall
9:45	SOC (GEBCO bathymetric editor)	Colin Jacobs
10:30	NGDC (Data ingest, archive, gridding)	Lisa Taylor/Barry Eakins
11:15	NAVOCEANO	David Fabre
12:00	Lunch	
1:00	U. Stockholm (Regional Mapping)	Martin Jakobsson
1:45	Lamont-Doherty (Data ingest, 100 m grids)	Suzanne Carbotte
2:30	Scripps (Global edited soundings for gravity/bathymetry calibration)	Dave Sandwell
3:15	NOAA/STAR (Satellite Altimetry)	Walter Smith
4:00	Discussion/Wrap-up	

Thursday, March 10, 2011

8:30	Summary and Discussion of Previous Day	Chris Fox (lead)
9:30	Formulation of a Data Flow Model	Chris Fox (lead)
12:00	Lunch	
1:00	Open Data Policy/Copyrights	(suggested by Dave Sandwell)
	Other issues identified on the first day	
	Other Topics?	

Discuss and agree on:

- Standards for data/grid contributions (e.g., data access, file formats, notifications)
- Define the standard methodologies/processes for updating/gridding GEBCO grid
- What tools do we have and what tools need to be revised or developed to meet GEBCO product needs?

4:00 Finalize plans, assign actions

6:00 Group Dinner, Free to all, except NOAA/NESDIS employees
Boulder Cork, 3295 30th Street. Boulder, CO 80301, (303) 443-9505
http://www.bouldercork.com

Friday, March 11, 2011

8:30 Developing GEBCO-100m prototype before next GEBCO meeting

Background from GEBCO Guiding Committee meeting	Chris Fox
Presentation	Suzanne Carbotte
Discussion	

12:00 Adjourn

Attendees

Suzanne Carbotte	Lamont Doherty Earth Observatory (Day 1 and 2 only)
John Cartwright	National Geophysical Data Center (Day 3 only)
Dave Clark	GEBCO Permanent Secretary
Justin Coplan	Lamont Doherty Earth Observatory
Barry Eakins	NOAA/NGDC
David Fabre	U.S. Naval Oceanographic Office
Vicki Ferrini	Lamont Doherty Earth Observatory
Chris Fox	NOAA/NGDC and GEBCO Guiding Committee
Colin Jacobs	Southampton Oceanography Centre and GEBCO Editor
Martin Jakobsson	University of Stockholm and GEBCO iSCRUM
David Sandwell	Scripps Institution of Oceanography
Walter Smith	NOAA/STAR and GEBCO TSCOM
Lisa Taylor	NOAA/NGDC and IHO Data Center for Digital Bathymetry
Pauline Weatherall	British Oceanographic Data Center and GEBCO Digital Atlas