MB-System Planning Meeting Summary
(and a Resulting Work Plan)
January 17-18, 2012
Lamont-Doherty Earth Observatory
Organized and Convened by Dale Chayes and David Caress
(Version 1.0, 27 January 2012)

Goals of this meeting:

1. Define priorities and tasks for the next five-year MB-System project to be proposed to NSF-OCE in February by Caress and Chayes.
2. Define priorities and tasks for MB-System related development efforts by other groups.

Participant Composition:

The intent was to have a small group composed of the core MB-System developers, users that have substantially contributed to the project over the past few years, and representatives of organizations that have expressed interest in parallel development efforts contributing to MB-System

In-the-room Participants:

Bob Arko (LDEO)
Brian Calder (CCOM/UNH)
David Caress (MBARI)
Dale Chayes (LDEO)
Vicki Ferrini (LDEO)
David Fischman (NGDC/NOAA)
Paul Johnson (CCOM/UNH)
Gordon Keith (CSIRO
Peter Lemmond (WHOI)
Suzanne O'Hara (LDEO)
Val Schmidt (CCOM/UNH)
Christian dos Santos Ferreira (MARUM/U. Bremen)

Remotely (via Google Hangout and/or email)

Kurt Schwehr (Google)
Hartmut Piertrek (BSH)
Bob Covill (Tekmap)

Invited but Unable to Participate:

Dana Yoerger (WHOI)
Bill Ryan (LDEO)
Christopher Fox (NGDC/NOAA)
Steve Roberts (UCAR)
Larry Mayer (CCOM/UNH)

LDEO:	Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA
MBARI:	Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA
CCOM/UNH:	Center for Coastal and Ocean Mapping (CCOM)/ Joint Hydrographic Center (JHC), University of New Hampshire, Durham, NH, USA
NGDC/NOAA:	National Geophysical Data Center, NOAA, Boulder, CO, USA
CSIRO:	Commonwealth Scientific and Industrial Research Organisation, Marine and Atmospheric Research, Hobart, Tasmania, Australia
WHOI:	Woods Hole Oceanographic Institution, Woods Hole, MA, USA
MARUM/U. Bremen:	Center for Marine Environmental Sciences University of Bremen, Bremen, Germany
Google:	Google Earth, Google, Mountain View, CA, USA
BSH:	Bundesamt für Seeschifffahrt und Hydrographie (Federal Maritime and Hydrographic Agency of Germany), Hamburg, Germany
Tekmap:	Tekmap Consulting, Fall River, Nova Scotia, Canada

Agenda:
This meeting involved two days of informal discussions. All participants were encouraged to identify and elaborate on:

- Problems – what needs improvement?
- Priorities for new development – what needs to be added?

The first day the discussion was intended to be comprehensive, with input from everyone and discussion of everything. Overnight the participants were asked to identify their highest priority issues. The second days discussion focused on the high priority issues, ending with a ranking by vote (each participant having three votes) of the priorities.

Some Implicit Priorities:
David Caress identified a few issues as having priority regardless of the discussion. The other participants therefore didn't include these in their rankings:

- Fixing the problem calculating beam angles in current generation Kongsberg multibeam data, which makes the initial conversion from raw.all files to format 59 very slow and degrades recalculation of the bathymetry by raytracing.
- Re-architecting the MBIO API to reflect additions to the "standard" information required for processing swath data.
- Recasting the coordinate systems used internally.
- Implementing full handling and application of sensor offsets, allowing straightforward handling all cases:
    - Data are initially uncorrected for offsets and biases
    - Data are initially corrected for offsets and biases, but with errors

    - Data are initially corrected for offsets and biases with no errors

List of High Priority Issues:
One or more participants identified the following issues as being of high priority. The number of votes for each issue is shown.

- Water column data (6)
- Metadata (3)
- Improved patch test calibration (3)
- Improved performance (3)
- Data formats (3)
- Long-term persistence of MB-System (3)
- Documentation (2)
- GIS object output (2)
- Support for compressed data files (2)
- Integration with CHRT (new version of CUBE) (2)
- Easier public access to the code archive (1)
- Quality indicators (1)
- Generic file format (1)
- Easier software distribution (1)
- BAG files (1)
- Arbitrary Weighting (1)
- Very low altitude, very high resolution (1)
- Realtime display (1)
- Overall, project manager GUI (1)
- Switch graphical tools from X11/Motif to Qt (1)

Explanation of High Priority Issues:

- Water column data (6)
    - Most current multibeam sonars now allow the collection of beamformed time series, or "water column", data. There is tremendous interest in developing capabilities to display these data, identify features and targets, and make use of the features and targets. Adding water column capabilities to MB-System, starting with a tool to simply display "movies" of the beamformed wedges, is the highest priority identified in this workshop.
- Metadata (3)
    - Better integration of MB-System with data archiving systems requires better capabilities for embedding and extracting metadata. The requirements for this necessarily must be specified by the archives (e.g. NGDC, MGDS/LDEO).
    - The most immediate needs are for mbinfo to output its statistics in a form that can be easily parsed, and adding sonar model, specifications, and settings to those statistics. The initial step will be to incorporate the MGDS version of mbinfo that outputs an inf file in a JSON format.

- Improved patch test calibration (3)
    - MBeditviz provides an interactive patch test capability, but this requires improvement. The additional features should include an ability to color the soundings in the 3D cloud display according to a variety of parameters, particularly including identifying soundings from different swathes or files. An automated, objective solution for bias parameters using selected soundings is also strongly desirable.
- Improved performance (3)
    - Making MB-System run faster is a high priority because of the increasing size and complexity of the seafloor mapping datastreams. There are many possible approaches to improving performance; the focus should be on those with broad applicability, significant impact, and ease of implementation. Some of the likely improvements identified include:
        - Augmenting the datalist scheme so that datalists have inf files, and thus data trees outside bounds of interest can be skipped.
        - Recasting the ancillary files (particularly fbt) for improved read performance.
        - Allowing raw files to be located on different filesystems than the processed files so that processing does not involve reading from and writing to the same filesystem.
- Data formats (3)
    - Because MB-System is designed on the premise of supporting seafloor mapping data in the native format (or as close to that as possible), supporting new and evolving data formats is as much a necessity as a priority.
    - The formats identified by participants as development priorities include:
        - Kongsberg – need to keep up to date with incessant changes
        - Reson – need to keep up to date with incessant changes
        - GSF – need to stay in sync and insure MB-System writes GSF correctly
        - Hypack/Hysweep – increasingly common as Hysweep dataloggers are being installed on many blue water survey ships.
        - XTF - difficult because nearly every swath/mb data stream is unique, but important to MARUM and many others.
        - Geoswath - important for many, including BSH and CCOM)
        - SEA SwathPlus – this sonar in use at CCOM
        - Edgetech 4600 – this sonar in use at CCOM
        - Klein Hydrochart - this sonar in use at CCOM
    - Some additional formats that might become priorities:
        - Humminbird

        - ME-70 - NGDC will be getting data from NOAA fisheries, and they may be extracting bathymetry from the water column data.
        - MB1 – new sonar from Odom/Teledyne.
- Long-term persistence of MB-System (3)
    - Participants (repeatedly) noted that the core MB-System developers are "not young", and that there is no discernible plan on their part to insure the viability of the MB-System project beyond the professional lives of Caress and Chayes.
    - The suggestion is to work towards broadening the numbers of people that actively contribute to the MB-System codebase, with the hope that future core developers can be identified (or will identify themselves).
    - One specific suggestion is to hold periodic developer workshops in which the focus is on how to add format i/o modules and how to build new applications on top of the MBIO API.
- Documentation (2)
    - Participants view the current documentation as lacking in several important respects:
        - The cookbook manual has not been updated since 2006, and so does not cover the use of MBeditviz, MBnavadjust, the processing of subbottom data, and a variety of other new capabilities.
        - The basic "how do I process this particular data type" documents all pertain to very old data, and aren't available for the current installations on UNOLS vessels.
        - There is inadequate documentation of the API and how to write i/o modules and new applications.
        - Training materials, including current data samples and examples, are not readily available
    - Participants noted that a variety of training materials, including videos, have been created by various users. Some effort should go towards collecting such materials in a centralized web site for easy access and contribution.
    - Documentation priorities for the future include:
        - User workshops – teaching how to process data with MB-System
        - Developer workshops – teaching how to write i/o modules and new applications in MB-System.
        - Write a new cookbook manual
        - Develop a new set of data samples and processing examples (including those used in the cookbook).
        - Develop a test suite to demonstrate correct functioning of an installation (overlapping with the cookbook examples).

        - Develop how-to-process-this-particular-data documents for common datastreams, particularly those from prominent research vessels.
        - Collect and make easily accessible MB-System training materials developed by users.
- GIS object output (2)
    - The data product outputs for GIS and other end user tools needs to be augmented. Some high priority forms desired include:
        - Shapefiles for ESRI and other GIS
        - KML file for Google Earth
- Support for compressed data files (2)
    - NGDC and MGDS desire to reduce their data storage requirements by supporting reading of swath files in a compressed form, most likely gzip.
- Integration with CHRT (new version of CUBE) (2)
    - CUBE (Combined Uncertainty and Bathymetric Estimator), is an error-model based, direct DTM generator that estimates the depth plus a confidence interval directly on each node point. CUBE was developed by Brian Calder of CCOM several years ago, and has become a standard component of multibeam data processing for hydrographic (chart-making) processing.
    - Enabling MB-System users to use CUBE for gridding and data cleaning is a priority.
    - Developing an MB-System tool incorporating CUBE was a priority in the previous MB-System NSF proposal in 2006.
        - The proprietary nature of the CUBE source code (restricted by CCOM licensing agreements with industry partners) meant that we could not use the CUBE code directly in MB-System.
        - CCOM was willing to let us distribute executable applications built from CUBE source, but this was too complicated for our limited resources.
        - We enabled IVS to integrate their product Fledermaus with MB-System, in turn enabling use of the Fledermaus implementation of CUBE within an MB-System processing environment.
    - Brian Calder of CCOM is developing a new version to be called CHRT, which will be distributed using a new client-server model.
        - CCOM plans to make executable CHRT servers available freely for several platforms, and to make the API for interacting with the server openly available.
        - MB-System can therefore include an open source, separately developed CHRT client that integrates with the CHRT servers that can be obtained from CCOM.
        - To accomplish this, the MB-System core developers will become CCOM partners by “sweat equity”.

        - The details of this arrangement remain to be formalized.
- Easier public access to the code archive (1)
    - Accomplished during the meeting by making the subversion code archive available for anonymous downloads.
- Quality indicators (1)
    - BSH and MARUM would like MB-System to handle and calculate the sounding quality information that is required to enable determination of IHO compliance or non-compliance of survey data.
    - CCOM has relevant knowledge of this issue.
    - Significant extensions to the MBIO API and new data products will be required to accomplish this.
- Generic file format (1)
    - As always, there is a desire for a "generic file format" for swath sonar data.
    - Any format that can be defined can be supported.
- Easier software distribution (1)
    - The install_makefiles mechansism for generating makefiles and enabling building the software has never been elegant and has always complicated efforts to package MB-System for standard package managers (e.g. Fink, debian, rpm).
    - Considerable recent effort by Covill, Schwehr, and Chayes has made progress towards implementing a configure-based build system. This will be completed.
- BAG files (1)
    - A recent, but now standard data product in the hydrographic community is the Open Navigation Surface, or Bathymetry Attributed Grid (BAG) file format.
    - BAG files incorporate estimated uncertainties and other metadata in addition to the gridded bathymetry.
    - Supporting BAG output requires a facility for estimating uncertainties, which ties to the "Quality indicators" priority discussed above.
    - Supporting BAG as inputs to gridding and plotting will enable users to easily incorporate much of the data available from Hydrographic agencies.
- Arbitrary Weighting (1)
    - To date, users can arbitrarily set priorities at a file level for gridding and mosaicking.
    - Data weighting should be extended to allow users to apply quality values or other parameters (e.g. beam angle, bottom detect type) to the calculation of data weights.
- Very low altitude, very high resolution (1)
    - The existing code handles lateral resolutions of beams and navigation to scales of about 0.1 m.

    - In order to enable surveys from altitudes of 1-2 m with associated resolutions of ~0.01 m, the codebase needs to be augmented to handle resolutions to about 0.001 m without errors.
    - Low altitude surveys also potentially include bathymetry derived from photogrammetry of stereo camera images and from structured light laser systems.
- Realtime display (1)
    - MB-System does not include a realtime display tool.
    - The desired tool would be based on the MBgrdviz interface, and would allow both incorporation of previously existing data and on-the-fly gridding and display of realtime survey data.
- Overall, project manager GUI (1)
    - MB-System remains primarily a command line package, in which even the graphical tools must be started from the command line.
    - Some users, notably Bob Covill and Paul Rybinski, have built GUIs for MB-System that they and a few others use.
    - More than one GUI can exist to meet diverse needs.
    - The target application within core MB-System is a GUI that enables project management with automated updating of data products following processing activities.
- Switch graphical tools from X11/Motif to Qt (1)
    - The existing graphical and visualization tools are all based on Motif/X11, which is ancient and problematic from a variety of perspectives.
    - The consensus of the smart and knowledgeable people asked is that MB-System should be migrated to Qt, which is both open source and enabling of true cross-platform portability.

A Work Plan

Over the next five years, the MB-System software package will be developed by an increasingly diverse group with different priorities and multiple sources of funding. David Caress and Dale Chayes will continue to be the project leads taking primary responsibility for the source code archive, the primary source code distribution, the project websites, communication and project coordination. David Caress will continue to be the lead programmer and architect, and will likely continue to spend more time writing MB-System code than anyone else. However, based on the discussions leading up to and including this meeting, we anticipate a considerable amount of MB-System development effort beyond Caress and Chayes.

Below is a summary of the possible institutional participants and the likely priorities for efforts:

Caress (MBARI) and Chayes (LDEO) – Five Year Renewal Proposal to NSF-OCE

- Development and support related to the NSF-supported marine geology and geophysics community:
- Proposed 15 February 2012 for five years of support
- Development priorities
    - Data formats
        - Keep up to date with data from Kongsberg multibeam installations on UNOLS vessels
        - Keep up to date with data from Reson multibeam installations on WHOI NSDF platforms (Sentry, Jason, Alvin)
        - Support input from BAG files
        - Support other data formats associated with NSF-supported projects as they become relevant.
    - Re-architect underlying MBIO library to support additional information needed for processing current generation swath data
        - Particularly including beam quality information
    - Recast the internal coordinate systems used in MB-System
    - Augment MB-System to fully handle sensor offsets and biases and allow for corrections to errors as needed.
    - Allow data weighting in gridding or mosaicking according to quality measures or arbitrary schemes.
    - Implement straightforward improvements that substantially improve performance of key tasks.
        - Augmenting the datalist scheme so that datalists have inf files, and thus data trees outside bounds of interest can be skipped.
        - Recasting the ancillary files (particularly fbt) for improved read performance.
        - Allow raw files to be located on different filesystems than the processed files so that processing does not require reading from and writing to the same filesystem.
    - Improve patch test calibration capability of mbeditviz by allowing the soundings to be colored according to a variety of parameters.
    - Add capability to generate additional data products:
        - Shapefiles for ESRI and other GIS
        - KML file for Google Earth
        - BAG files
    - Support reading from files compressed using gzip.
    - Redo the graphical utilities to use Qt graphics rather than Motif/X11.
        - mbedit, mbvelocitytool, mbnavedit, mbgrdviz, mbeditviz, mbnavadjust
    - Develop tools for working with multibeam water column data
        - Develop tool to display water column data
            - This will be a platform for implementing various algorithms for identifying and extracting features of interest

        - Augment MBgrdviz to display 3D renderings of features extracted from water column data
        - Likely to involve collaboration with German and NGDC groups
    - Develop MB-System CUBE client, with mbgrdviz style interface.
        - Necessarily involves collaboration with Brian Calder of CCOM
    - Develop realtime display tool(s)
- Documentation
    - Manual pages
    - Cookbook Manual
        - To be written jointly by Dale Chayes and Krystal Andersen (an MBARI technician)
    - Workflow documents and examples for specific data streams
    - Data example suite used for cookbook and for automated testing
- Dissemination
    - Websites
    - Code repository access
    - Source code distributions by ftp
    - Training
        - User workshops and developer workshops
        - Probably one each year, alternating between user and developer focused

Caress (MBARI)– Ongoing Projects at MBARI

- Development and support related to the MBARI Mapping AUV operations and to the MBARI Ocean Imaging Project that involves ~2 m altitude surveys with high frequency multibeam plus stereo cameras
- All processing software development included in MB-System
- Development priorities
    - Navigation post-processing and adjustment for submerged platforms such as ROVs and AUVs
        - Collaborative with Dana Yoerger at WHOI/DSL/NDSF
    - Realtime terrain and vision aided navigation for ROVs and AUVs
        - Collaborative with Steve Rock at Stanford
        - Requires re-architecting of MBIO
    - Combining low altitude multibeam bathymetry with photogrammetric bathymetry from stereo camera imagery
        - Requires MBIO re-architecting planned by Caress/Chayes above

Ferreira (MARUM), Piertrek (BSH) and others in Germany: Multiyear proposal for software development focused on MB-system

- Proposal to be submitted in April 2012 to German government
- Led by MARUM, but incorporating priorities from multiple agencies and institutions in Germany, particularly BSH
- Development priorities

        - Adding capabilities for calculating IHO compliance/noncompliance
            - Requires MBIO re-architecting planned by Caress/Chayes above
        - Water column imaging
            - Develop tools for imaging and extracting information from beamformed water column data
            - Collaborative with Caress/Chayes efforts
        - New data formats
            - Hypack/Hysweep
            - XTF
            - Geoswath
- Documentation priorities
    - Multimedia
    - Training materials

Calder (CCOM/UNH): Development of CHRT server

- Brian Calder is leading development of a new version of CUBE to be called CHRT that will be made available as a server to enable client-server implementations by CCOM industrial and "sweat-equity" partners
- Development priorities
    - CHRT server with standard testing cases for validating implementations
    - Calder is willing to work with Caress/Chayes to enable development and dissimenation of a CHRT client application in MB-System

Fischman (NGDC/NOAA): Possible software effort at NGDC to speed development of NGDC priorities.

- David Fischman plans to explore the possibility of internally proposing some development effort involving MB-System
- Development priorities
    - Water column imaging
        - Develop tools for imaging and extracting information from beamformed water column data
        - Collaborative with Caress/Chayes efforts
    - Metadata insertion and extraction

Detailed Discussion Notes:

17 January 2012
Seminar Room in Lamont Hall

0910: Welcome (Dale)
Introductions around the room

Welcome and intro remarks (Dave Caress.)
The goal of this informal event is to gather input for the next proposal to continue to supporting and growing MB-System. The threshold is high for supporting proposals this year because resources are tight at NSF.

Christian: GEOMAR and the German Hydrographic office intend to seek funding to work on MB-System

Dave: Brief history of MB-System (see slides)
We are looking for a list of what we aren't doing very well, priorities for development, good ideas, where is the technology going?
There will be some restructuring (refactoring) of MB to clean up

Dale: Dave added features for others, we need to think about strategy for the next MB proposal.
Christian: Hans Werner is retiring (there is a job open), AWI has been using CARIS, GEOMAR and MARUM are using MB - focused on open source,

Data format support:

- Kongsberg (keeping up to date)
- Reson (keeping up to date)
- GSF (stay in sync, write GSF correctly)
- Hypack/Hysweep (2 formats, ascii (HSX), binary/proprietary (HS2) - HS2 uses a wrapper around native sonar files. Consider asking for format doc and/or API. Perhaps leverage from their big customers. Could investigate the "stripping" stragety for HSX files. There is an HS2 reader (.dll) that comes w/ Hypack and works under Windows. Hysweep works in a projected coordinate system relative to a chosen location which is not included in the data file.
- XTF (swath/mb data contained in) (nearly) every stream is unique. MARUM coded sidescan C3D 1624 (read only) support so far. Will create writing support
- Humminbird: new, inexpensive (NZ) sonar, support for? Someone wrote an FBT creator
- Geoswath (important for many, including BSH)
- SEA SwathPlus:
- Edgetech 4600:

- Klein Hydrochart (the bathy version of the 5000?
- MB1 (Odom/Teledyne)?
- Delta-Ts are fading in the NDSL, migrating to Reson. SM2000 has been repalced with a Reson. In power consumption, the DeltaTs are desireable.
- ME-70? - extracting bathy. NGDC will be getting data from NOAA fisheries. Tom Weber is working on generating bathy. Requires beamforming and bottom detection. One can buy a bottom/bathy add-on (hardware.)

- How to cope with "low useage" (e.g. on-off research efforts or new systems) data types? e.g. a "simple" (easy to understand/create/read/write) format. Perhaps FBT (or an "improved" FBT)

- Do we have a sense/breakdown of which constituencies (academic researchers, NOAA, hydrographers, etc.) use which formats? Growing use of GSF (exported from CARIS.) Many packages have trouble writing correct (self readable) GSF. SAIC is aware of the problems with poorly written GSF files.

- Potential for establishing a "clearing house" for data format documents, changes?

- Consider a "master" processing format. Might create a faster and/or more maintainable code base. FDR (retain capability to read and re-write native formats)

- It is mandatory to archive (~forever) the originally logged data! (Dale's opinion)

- Water column support: e.g. from EM302, 122, ME70/EK60 (same format), some Resons (7x series). Many new sonars are generating wc data. There is emerging science interest in wc data for methane seeps, oil seeps, water column structure, calibrating/ed backscatter from wc data. A student of Steve Holbrokes? Mashkor/Tom - turned into FM mid-water tool. Echoview derived from Echo (EK500) from CSIRO. Dave is planning to do some wc work at MBARI. ME60 would/could add a big (fisheries) constituency. FDR

------- break ------

Restructuring: historically: amplitude, sidescan, bathy. Considering expanding: bathy, per-beam backscatter, water column, sidescan (not laid out on the seafloor.) Some sonars are now supporting various kinds of "quality" factor. Fits with desires to incorporate IHO/hydro info.

- Convert to modern build system. Bob Covill's contribution is moving into a branch in the SVN repo will eventually become the encouraged approach.

- Recast the MBIO API to improve ease of use and to handle all desired information (as part of the code restructure.)

- Recast internal coordinate systems – make self-consistent throughout the codebase and possibly change to different conventions. Platform orientation and geo-orientation, attitude corrections, ray-tracing. Current coordinate system is a historical relic. What is the right coordinate system to use?

- Improve the way offsets are handled internally to properly accommodate modern nav, attitude and sonar systems. Can you use the ones that are in the data files?

Documentation:

- Documentation (man pages just aren't enough any more, they are necessary)
  - there needs to be a formal document set beyond man pages

- Bringing and maintaining the Cookbook up to date - consider moving to a wiki for distributed contributions, or both and/or something else. There doesn't have to be a single strategy but probably only one bit can be supported in an NSF proposal.

  - Should "we" assemble a course syllabus and materials?

- Multimedia based documentation (e.g. videos), add some other languages to Christian's? What is the right collection of levels of detail.

  - The data examples used in the documentation should be the same set that is used in the "make test" as part of the build system.

  - Consider an annual "training/education" session - using the doc set.
  - collect the existing content into new docs
  - add to man page "how to start processing EM122s"

- Definition of and use of the *.fbt files

- Graphical tools – what to use instead of Motif/X11?
  Support for 3D things (after choosing the next gui thing). Currently considering QT. Any opinions about others? Is Qt good for (another) decade?

- Support for remote (near the data) processing (move to performance)

- Potential to implement CUBE as open source?
Cube will never be available as open source (or very difficult, due to existing license agreements.) However, the new version (which won't be called Cube) currently under development will have a code development model. They (CCOM) will provide test suite based on the reference implementation of the algorithm. Algorithm is designed to run in a client/server model with a defined ICD across an (IP?) network. Access to industrial associates is available in return for services/loaners, etc. For non-profits and research entities can contribute sweat equity which would allow

access to the source but not permission to redistribute as source code. If MB does not distribute compiled pieces of the server, would CCOM do that? Perhaps, depends on someone building them.
Slides from Brian describing the client/server model. Seems like a viable model for MB. How do we reference that in our upcoming proposal? At least, we'd need a letter from CCOM/Brian to include in our proposal.

- Embedding IHO standard-relevant information and calculations
We need to understand (from NOAA, CCOM, MARUM, BSH, ?) what, exactly, is required, what else would be good, nice...
The surveyor, not a software package, provides “IHO-quality” survey. NOAA hydro production folks are fairly tightly tied to CARIS. Other parts of NOAA might be interested, e.g. for rt monitoring, prototyping
An example of uncertainty model calculations would be very helpful.

- Open Navigation Surface (ONS) Bathymetric Attributed Grid (BAG) capability, encouraged by many, organized by CCOM. Of interest to NGDC, it would be great if MB could read a BAG. NGDC is currently exporting BAGS to xyz and reading as MB163 (for tsunami and storm surge efforts). How would MB read a BAG? An HDF5 regular array (grid) and metadata (ISO19115) depth, uncertainty (meaning of is defined in the metadata), hydrographer's modifications, optional extensions (Navy's nominal depth, aux info), certification (digital signatures). Code is available from opensurface.org requires the HDF5 library, zlib, szlib, xerces. IHO is finalizing S100. BAG will be included as S102. How important is this to the NSF-supported community?

- Real-time display capability: particularly to make a background grid and plot real-time data "on top" of it. If it's added to MB it would be through MBGRDVIZ, with similar capabilities of SeaSurvey. Roger Davis (UH/HMRG) has a package that does some of this.

- Improved calibration (patch test) capability
  - MBeditviz: improved usability (discrete colors for swaths) for patch test.
  - Consider objectively solving for bias parameters, using MBeditviz to select the data to use.
  - Use the solution characteristics to inform quality/error estimates

- Additional metadata scaffolding (or simply use existing comments)? w/ ONS? What exists right now, (nearly) all data formats support comments and MB can embed comments and read those comments. MBprocess adds comments about what it is doing. MBProcess can add metadata (about 6-8 years ago.)
Can MB support ISO19115, or FGDC (in xml.) Perhaps the right strategy is to migrate Suzanne's mbinfor2r (now emits json/xml) back into mbinfo implementing some of the common metadata standards. FDR

- mbinfo: Consider changing date format to ISO standard date YEAR-MONTH-DAY

FDR

- Data search capabilities: search via datalist for parameters. Gordon has a java app (w/ GUI) that searches inf files and returns a new datalist. Could search on a matching string (regex) in a path name.

Better support for GIS data products:
- Currently support for geoTiff
- Export to GIS vectors (Shapefiles). The program Steve put up on his site is a simple example of this using the OGR library. The Shapefiles are tagged with filename, year, and julian day so they can be queried from a GIS.
- kml/kmz
- outputs should be OGC and/or ISO standard

- Consider alternate debugging options (gdb DDD, etc. vs printf()) FDR

- Performance improvements
Finer grain parallelization (take advantage of multicores)
Computing closely coupled to the data store
Distributing data to be close to computers
Perhaps indexing files?
Faster mbdatalist (an inf file at the datalist level)
Consider capability to separate (and re-merge) water column data to/from separate files
Water column is large volume data and user interest in it seems like it will be a growth area
interested users are always encouraged to contribute to MB
Consider the ability for processed and "raw" files are in different file systems (can currently be done w/ softlinks), perhaps through mbdatalinks? perhaps mbdatalist can make/manage softlinks? Different processing systems/groups use different strategies for staging (logged, "raw", processed, etc.). Consider setting up a processing environment that points to files elsewhere...
Document a representative collection of strategies and examples about performance optimization. Consider some example code for testing processing environments.
Performance on different types of storage, e.g. SSD, conventional disk, different disk interfaces, e.g. iSCSI, Thunderbolt, strategies (parallel vs. sequential)
Consider utility of the linear algebra packages (because they have built-in parallelization capability.) Balance w/ minimizing dependencies.
List of file extensions

- File locking: to prevent multiple folks from editing the same file. Comes with some additional complexity - orphan lock files. Hardly anyone has complained. The idea is good. Enabled Christian's multiprocessing. mbdatalist has options to find lock files and to delete them.

- MB-System build validation (test suite)
- Testing and validation would one of my big requests for improvements. e.g. Kurt's code that checks a SEGY rev 2 and shows all the fields that are outside of the spec.

- Licensing issues (GSF and Oregon tides package Oregon State Tide Prediction)

- Read gzip files in list (i.e. mbgrid)
MB-System was originally designed to use piping, some aspects remain in the code
Converting to using gzlib might (will) break the ability to pipe
Need to figure out if this is really a performance improvement: FAR
Should ask the list if the loss of piping capability is important

- Language bindings calling the libraries so we don't have to code the command line commands
What languages? ruby, python, perl, java, matlab....

- Archive data format: discussion of what to archive (data files (as raw as possible), plus working versions, reading and writing code, format docs,...)

- mblevitus: need to enable newer capabilities - Dale talk to Steve

- Incorporate tiled data as a background data set: coverage service vs. image, integrating w/ a web service, etc. What is GMT planning on?

- mbm_gridplot: add -K option to add file

- Capabilities that don't exist in other packages
  Sound absorption correction
  Robust bias (calibration) estimator

- Water column display (Dave has explored this using Simrad data and MB)

- Better use of multicore. Anything that improves performance is a good thing with the increasing size of MB data. [move to performance]

- Improved (faster) mbgrid [move to performance]

- GDAL to make more portable grids for export

- Lack of high latitude support (around the pole)

- Dateline bugs

- Can't apply multiple sound speed profiles to a single mb file

- Sound speed profiles from a Moving Vessel Profiler (MVP) and applying based on times. How do you decide when to switch... General case of 4D sound speed (and absorption) profiles and tides

- Sound speed corrections for mb58 (raw.all) format files still not correct? Kongsberg xx2s, 710, 2040. Format changed so they do not provide the ray tracing corrected beam angles. So that if you ray trace you have to calculate those angles using attitude data. MB does not currently handle this correctly: it is trying to calculate from the SVP and angles and iteratively estimating angles. This is slow.

- JHC/UNB's Swathed has the notion of an incremental grid - maintain grids of the soundings and their weight so you can add soundings or grids without recomputing the whole grid. Dave has been working on a similar approach with an "update" capability in mbgrid - to be efficient when "joining" large grids. Limits algorithms - they have to be reversible. Alternately maintain a spatial index to what was done to each "chunk" so you can "back out" small pieces.

- MB currently supports a couple of grid weighting functions (exponential & beam footprint). Consider adding an the capability to apply a user-generated weighting function, a flexible method, perhaps built-in "other" methods... This is distinct from existing capability to apply a weight to a whole file. Once we are carrying quality values, then they would be important for gridding. FDR

- Quality values (weights):
1) As derived in/by the sonar (for those that do)
2) Integrated sonar data (incorporating latency, attitude/position references, system uncertainty...)
3) User's opinion (objective computed or subjective)

- Selective use of data. E.g. data collected in very deep mode of EM300, are very difficult to process for backscatter. Would like to be able to exclude some modes from some kinds of processing. Since different sonars do very different things it might be hard to generalize. What about ping length?

- Survey's referenced to the ellipsoid (instead of sea surface and tides....) any sense in MB incorporating this strategy, perhaps for other hydro office.

--- Break for lunch (~1215 - 1445)---

Review Christian's work flow diagram. Discussion

Vals' slides from phase measuring swath sidescan systems: Geoswath (Geoacoustics) uses a Vernier detector, Klein 5000, and Edgetech 455 kHz(multiple staves, John Bird, CAATI algorithm - over determined, least squares process.

Elaborate many items above

--- break --

Dave's slide of data processing flow. Once set up, the order isn't particularly important except that MBnavadjust creates a project on it's own so you really need to have the editing done first.

18 January 2012
Seminar Room in Lamont Hall

Bob Covill's GUI project manager (URL) to Bob's blog

Hartmut understands that NSF is not likely to fund the IHO development effort and he will seek other support

Discussion (Val and Dave C.) about the internal structure of MB and GSF. Merits of being able to use MB (mbcopy) as a tool to translate other formats into GSF. Data loss exporting from CARIS into GSF often looses data - this isn't an issue with the GSF format, but rather with the exporter(s).

Demo of mbeditviz. discussion of enhancing cal mode, how to pick/color segments, color by files, there are situations where more than two colors are desirable. Colormap for the data, bathy, perhaps quality. Color by distance (displacement) from a gridded surface. It would be nice to see/show a color bar (scale.)

Lock files: demo (Dave C.) - showing the warning for locked files still in use by leaving editviz running and trying to run mbedit on one of the files. Add ability to selectively remove lock files by username, host IP, etc. Right now the removal is all or nothing. A running app won't break if it's lock file is removed, but defeats the intent. Consider improved parsing of the lock file content. Need to explore remote locking and lock file content - e.g. does it contain the path to the file.

Would be nice to have a geo context (location map inset) in mbedit. In the queue after the transition to QT.

Format of (newest) FBT files: FM (IVS) does not read the "new" (~year old) fbt files. Need to update the documentation. FBT is supposed to contain bathy (and nothing else), keep the file small, structured to read fast. Capable of centimeter resolution. Header is doubles, content is shorts w/ scaling. New version carries attitude. No travel times or angles. Not intended to be processing format. Consider adding weighting - it has to be there if we are going to allow weighting in MB. Consider making it extensible: probably FBT is not the right place for that. Format 71 can contain amplitude and sidescan - fbt files are an instance of 71s in which the amplitude and sidescan have been stripped out. Footprint size: perhaps - current effort contains beam angles.

Coverage map: can extract a crude estimate from the coverage mask in the inf file. Create a grid from the data and count the boxes that have data in them.

--- break --

There are two settings in mbdefaults related to file locking and fbt files. For your environment, you can elect to use the old fbt file format so creating new fbt files will

be in the old format. So you can continue to make/use files that are compatible w/ FM/IVS. (see the man page for mbdefaults)

You can disable use of lockfiles via mbdefaults for your environment. Multiple folks working together have to use the same defaults or chaos might ensue.

Searching through an mb datalist: does it make sense to allow addition of user-supplied key/value pairs? Other mb apps would ignore pairs that they don't recognize. First: extend .inf files so that they exist for whole datalists. Then: ignoring unknowns safely. inf files don't include all of the comments, only the meta comments.

Review the "votes" from last night.

Documentation:

Maintain formal docs: man pages and cookbook which need to be up to date "how do I get started" Probably also encourage more informal docs

Is it possible to have a 2-5 page description of how to create your own file format? Perhaps in 15-20 pages. Does that mean that the architecture is too complex. Perhaps can outline the structure and make a check list and what needs to be done in a few pages. The subtlety of what has to be addressed in those functions is more complex. The breadth of input files is broad (xdr, netcdf, multiple files, etc.) which is part of what makes it complex.

Think about how to maintain long term development and support for mb. How can we bring in younger programmers? Developer's manual? How to write code. Consider a developer's workshop (different from user training.) Consider discussion with other folks (e.g. Paul Wessel and Walter Smith) about this issue - lessons learned, common issues.

User facilitation - consider somewhat more formal user training (for using mb.) Annually? But not around fall AGU. We might find someone qualified to run a user training session associated w/ AGU.

Quality indicators:

The re-factoring/architecting will probably be done before quality indicators are incorporated. One set, two sets, more than two sets of quality values?

General sources: uncertainty modeling (CUBE) (set of parameters) for each model of echosounder (which isn't quite right) which are quasi-static, real-time approaches based on signal strength - vendors may (or may not) implement them the same way, RMS error around phase detects, standard deviation of weighted mean time of

arrival for amplitude detects. Different vendors (and different types of sonars) do/will use different methods. Marketing potentially confuses the problem. Published (well documented) methods are desirable and can be done using observables with no proprietary knowledge of a particular sonar implementation. They only address the acoustic system, not the whole survey system.

Could potentially be many quality assessments/factors/indicators for a single data set (file/ping/swath.)

Historically, MB has only had per-beam flags.

Potentially add "a couple" of quantitative values that come from somewhere and possibly manipulated by users. At least some sonars will include values in their data. mb ("the API") has to be able to deal with them across different sonars and vendors so that they can be operated on say when constructing a grid.

Perhaps we should also discuss this in a broader context - other folks. Will propose something for further discussion. (Current thinking: keep beam flags, add 3 values: appropriately calculated uncertainty estimates (vert and horiz); a weighting value perhaps derived from user-settable characteristics (e.g. beam angle, detect type...)

Could use Lurton/Augstin definitions for acoustic part. Got reference. The Rob Hare et. al, paper on the whole system.

What do you intend to do with QFs?
1) quantitative statment(s) about this sounding (can be used computationally)
2) good:not-good, etc. (probably not computational), might belong elsewhere (metadata?)

Could look at the IHO standards which contain some defined parameters. Yes and no....
S44 (current QC surveys) talks about TPU, categorization of surveys (orders) and max limits for them. Does not specify how they must be computed.

CHRT (CUBE V2):

How about GeoCoder also? (requires real acoustic values.) It has awkward licensing terms (like original CUBE) so it's unlikely that it would be GeoCoder but it's not out of the question to re-implement some of the ideas in GeoCoder. Approach by re-coding algorithms.
Open to other/additional implementations
Adding seafloor characterizations to MB
Need to be sure that the new architecture allows high quality seafloor characterization. Val will provide a list of needs.

--- lunch break --

Metatdata:

A new mbinfo output type that is:
  selectable: text, json, xml....
  extensible
  sufficient to bootstrap an ISO metadata record

easier to parse output format
look at some samples of current file level
Could encourage vendors to incorporate more metatdata details into raw files
Switch to ISO standard time format in new output formats. Consider including both time formats

Improved calibration:

See earlier discussion, add robust estimator(s) in mbeditviz
needs documentation

Native support for reading gzip files:

It's complex to read through the gzip libraries because the I/O is so diverse in many of the different multibeam formats that MB supports. For simple byte-stream formats it's not particularly hard, but for NetCDF and GSF, their libraries are actually doing the open and reading so they would have to be modified.

Discussion of unzipping into a temporary store (ramdisk, memory, scratch diskdir, $MBTEMPDIR).... perhaps need a mechanism (mbset) to save or destroy unzipped files. Possibly implement in a way that the source file can be on a read-only file system.

gzip library offers support for a user space file system

Additional (including GIS) outputs:

There are additional data produces that would be valuable to lots of folks. Need to develop a clear list of output types:
- well known text (wkt) (search Wikipedia) for track line (NGDC)
- shapefile (ARC)
- kml/kmz
- gml
- gpx
- BAGs

- route (plans) kml, ascii plan (Kongsberg's format), Endeavor, NMEA
- point data (sites, way points, etc.), NMEA
consider capability to convert to/from formats via OGR
mblist can already output to NetCDF, possibly others?

Water column data:

A growing field of interest
When signal level is low, the (a)TVG ramp at range makes them look big - interferes w/ amplitude thresholding.
Construct voxels
Consider impact of compression methods on small features...
SIS drops water column data where there is no valid bottom detect
Shallow water systems (Reson) can produce a gigabyte of data per minute in shallow water....

Generic File Format (GFF)

The goal is to have a low threshold (aka easy) way to get some new data (type, source) in MB for exploration, either for a new sonar or some lab-gadget/experiment. Generalize the concept of a sonar, reference frame, mounting, "beams", etc.

How is this different from (the goal of) GSF? more generic, does not have sensors specific records, fields are more precisely defined. The cycle time for adding things to GSF can be long (to very long.)

If you define a format that meets your needs, MB can support that. Probably faster if you contribute the code. A generic one could become mb72.

Arbitrary weighting:

A weighting scheme to fit in fbt files, it would be good to have a method for user-generated weights to be entered.

Most existing formats do have the capability to store weights so ancillary files might be the most likely approach. Hadn't thought about pushing it into existing formats (that don't have a place for it.)

Calculate on per-sounding basis, may depend on real-time stuff.. "It just wasn't as simple as I originally thought it would be" but surely there is a reasonable way to do it.

FAR

Long term persistence of MB (dev and support):

Sustainability plan
Talk w/ Paul, Russ Rew (NetCDF)
Can't do the long term thing by just hiring a programmer (need the fire in the belly)
Consider: summer interns, a Google summer student, (above) developer's forum, what else, developers mailing list
Discussion of pros and cons of a single list vs separate lists.
Perhaps find someone with interest in QT? Jens at MARUM has significant experience w/ QT/C++/OO. There is other QT expertise at MARUM.

Improved Performance (including i/o and disk:
---
much detail from yesterday
more robust code (valgrind)
trackdown the malloc opportunities for software quality improvement
extinguish the compile time warnings
FUSE (via libgz and perhaps independently for low overhead "file" access)
threading might make sense in some contexts (e.g. gridding, ray tracing)
coarse parallelization mbm_multiwhatever
processing close to the data (large, fast disks with local computing)
document some clues about how to assess performance and optimize in a few typical processing environments
Learn (perhaps) from HADOOP (Apache's version of very large scale paralleled computing, Google's mapreduce) HADOOP's distributed file system might be a feature

Perhaps an mbsystem web site that would allow you to upload a file and get the output of (say) mblist, back. Perhaps Google would host it?

Apple multithreading library libdispatch grand central dispatch (becoming available for Ubuntu and FreeBSD.
Google also has something out there.

Improved GUI (project mgr):

Bob Covill and Hartmut have done some work on various versions.
They do different things, coded differently
How do you tradeoff simplicity vs completeness
It would be possible to create a data flow (graphical) strategy (e.g. built in LabView)
GRASS has a gui interface that reaches their command line tools. There is an mbio reader for GRASS. Could be Covill.
How do you capture the file system structure of the data store organization in (to use with/in) such a gui? projects, data lists, data products, etc.

Art Lerner-Lam stopped by for a visit.

public repo - anonymous read:
---
done

----- break ------

Very shallow water:

Address limitations of scaling in the current fbt implementation to enable working at centimeter scales by handling data a millimeter level.

They will be required for much of the IHO areas. Mostly they work at decimeter scales.

See earlier notes.

Existing, old fbt files get handled.
New ones are created in the new format.
There is a flag in mbdefaults to force writing of old-style fbt files

==========
In the current instance, mbinfo for mb58 files (and presumably for Reson) it tells you if the file contains water column data. Also more info (w/ -n for notice)

Reviewed the whole set of notes starting yesterday morning.

Add relative complexity and duration for each of the "voted" tasks. Added one more vote.

Discussion of priorities/prioritization.
Most of the documentation efforts could be pursued immediately. The developer's guide part might change: API will get extended. Extract/insert functions will get extended. The functions will remain but arguments will be added.
Probably internal changes before starting QT. Will do the internal/underlying changes as soon as possible.

The change to QT would enable a native port to Windows for the GUIs but it will remain a command-line tool. It's a consequence, but not a driver or reason. On a Mac it would be Cocoa-native and not require X11.