



## **INTEGRATION OF CHRONOSTRATIGRAPHIC DATABASES FOR THE 21ST CENTURY**

*SUMMARIES OF*  
**NSF-SPONSORED WORKSHOP (AMHERST, MA, NOV. 8-10,  
2001)**

*AND*

*CHRONOS*      **STEERING COMMITTEE RECOMMENDATIONS FOR  
5-YEAR PROGRAM**

### **Contents**

- I. Executive Summary
- II. Mission Statement, *Chronos* Goals, and Applications
- III. Summary of *Chronos* Database System
- IV. Chronostratigraphic Data Issues– Types, Standards, Interpretation,  
Compatibility, Data Usage
- V. Database Design Issues – Public Interfaces, Toolboxes, and Information Analysis
- VI. 5-Year Plan and Budget
- VII. Appendices -- Participants, Workshop Issues, Agenda, and Selected Current  
Databases



## I. Executive Summary

### Workshop Goals and Themes of Workshop

A group of thirty quantitative stratigraphers and database specialists from diverse research teams met in a three-day workshop (Nov, 2001) to discuss methods and strategies for:

- Assembling, integrating and distributing data relevant to geologic time
- Producing a dynamic global timescale to frame Earth history events and processes for social benefit.

Three major working sessions, each consisting of focus teams and full-group discussions, considered scientific and information technology aspects of database development, including:

- Coordination of database standards and compatibility
- Designing strategies and tools for information retrieval and analysis of all types of global and regional stratigraphic data
- Discussing future directions for database integration and centralization of currently distributed depositories.

A subset of workshop participants, representing information technology and various stratigraphic shareholders, prepared a detailed 5-year plan to accomplish the Workshop *Chronos* concepts.

### Major Decisions and Recommendations

#### (1) A chronostratigraphic database system is vital for future earth science studies.

Establishing an integrated chronostratigraphic database, provisionally to be called *Chronos*, will facilitate greater efficiency in stratigraphic studies. By providing both greater ease of data gathering and allowing for metadata synergies, the *Chronos* system would be of fundamental importance in a wide variety of research, including time scale construction, paleoenvironmental analysis, paleoclimatology, paleoceanography, and paleotectonics. Beyond scientific research, *Chronos* would provide educational and societal benefits by providing a readily accessible source of information on topics of general interest (e.g., evolution, frequency of catastrophic events in earth history) and facilitating studies of issues of modern concern (e.g., atmospheric change, biodiversity).

**(2) *Chronos* will include all chronostratigraphic data that have adequate attribution.**

In order for *Chronos* to be as comprehensive as possible, the database would need to be as inclusive as possible of different data types. The system will accept all data with the potential for stratigraphic correlation, without scale limitations in time resolution or geographic scope. Quantitative and qualitative data will both be accepted and the database will serve the chronostratigraphic needs of academia (both researchers and teachers), government and industry.

Nevertheless, the data to be included in *Chronos* will also need to meet basic qualifying criteria for inclusion. These included full attribution as to source, the inclusion of the raw data upon which interpretations are based and ranking of data quality and error range estimates.

**(3) *Chronos* will have distributed nodes coordinated through a central global hub.**

The primary consideration in the design of *Chronos* is the creation of a system that incorporates the widest variety of existing databases without subjecting them to costly revisions. As such, raw data will continue to reside in the *Chronos*-affiliated databases (“distributed nodes”). A central hub will provide data search engines, data integration and manipulation capabilities (i.e., metadata), and educational information and downloads for the general public. The central hub will also act as guarantor that the distributed nodes meet qualifying criteria for affiliation, as well as provide information on data translation and format (e.g., for paleontological data, taxonomic synonymy tables and an image database). *Chronos* central hub will also provide interactive access to the major research product produced by database integration – a global timescale framing earth history events and processes in such a way as to maximize societal benefit.

The multiple-node structure of *Chronos* (figured in Section III) will allow it to fit seamlessly into the NSF-sponsored *Geoinformatics* initiative where it will, function as one of the major earth science discipline hubs.

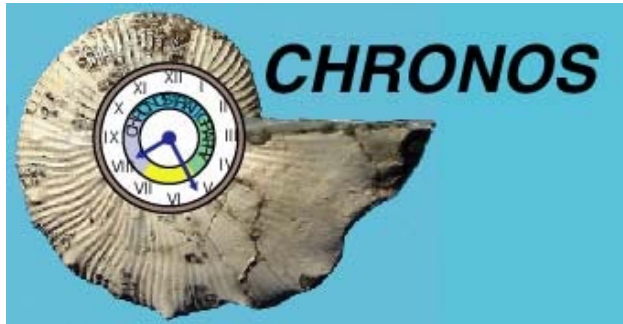
**(4) *Chronos* will be administrated under the International Commission on Stratigraphy.**

A small permanent core staff of both IT and chronostratigraphic professionals under the supervision of the International Commission on Stratigraphy (ICS, which is under IUGS) will be responsible for administering *Chronos*, especially its central hub. Participation in the core staff would likely be for limited set terms.

In order to realize this vision, a strategic plan is established. The *Chronos* administration will be modeled after that of the Ocean Drilling Program, operating under the oversight of a steering committee initially staffed by participants representing academics, industry, educational, government and international programs..

**(5) The *Chronos* system will be developed during the next 5 years, and fully functional by 2008. Progressive public access (research and outreach) will begin in 2003.**

The development of the *Chronos* central hub will proceed throughout 2003 and prototype demonstration of the biostratigraphy node will be performed at the IGC meeting in the summer of 2004. Development of radiometric, magnetostratigraphic, chemostratigraphic, cyclostratigraphic and other nodes will be completed by 2006. Simultaneously, the Central Hub will assemble the suites of search, visualization and analysis tools envisioned by the Workshop. A series of “time-slices” will be used to test the *Chronos* system and focus thematic node development.



## II. Mission Statement, *Chronos* Goals and Applications

**Stratigraphy** is the study of rocks, their records of the past (fossils, climatic proxies, magnetic polarity, etc.) and their distribution in space and time. The object of stratigraphy is to reconstruct the history of the Earth and eventually of extraterrestrial bodies.

“Stratigraphy is the heart of geology. Without it, the findings of other branches could not be knit into a single historical whole. Stratigraphy makes possible the synthesis of a unified geological science from its component parts.” [*paraphrased from J. Marvin Weller, 1947*]

**Chronostratigraphy** is stratigraphy within a time frame. Chronostratigraphy is essential for understanding the history of the Earth, because age ties everything together. The assembly of vast amounts of stratigraphic information into databases has created important tools to understand past climatic, evolutionary and geological processes and their implications for future trends on our planet.

This *Workshop on Integration of Chronostratigraphic Databases* explored how current and future stratigraphic databases can be partially unified into a powerful system for producing a dynamic global timescale and understanding the interrelationships of past geological, climatic and evolutionary trends.

### Mission Statement and Goals

#### Mission statement:

The mission for creating this integrated chronostratigraphic database is:  
***to produce a dynamic, global timescale to frame Earth history events and processes for societal benefit.***

#### Principal goals are:

- Assembly, integration and distribution of data relevant to geologic time
- Maintenance of a consensus geological timescale
- Public outreach – communicate to the public the importance of understanding rates in natural processes using the geological timescale
- Research outreach -- Provide a fundamental research tool for the broader geoscience community and a temporal framework for understanding the 4<sup>th</sup> dimension (rates and processes)

## Applications and Components

### Unified suite of thematic hubs

There is an intrinsic value in developing a unified suite of chronostratigraphy databases that seamlessly unites several distributed thematic hubs. The chronostratigraphic data to be included or linked within this system include:

- **Life** (paleontological assemblages, evolution, biodiversity, extinction, productivity proxies from carbon isotopes and other criteria, terrestrial and marine trends, etc.)
- **Climate** (orbital forcing, glaciations, temperature records from oxygen isotopes and other proxies, ecosystem changes, dust accumulation, etc.)
- **Surface processes** (carbon and other biogeochemical cycle monitors, weathering balances from strontium isotopes and other data, sediment accumulation rates, etc.)
- **Core-Mantle dynamics** (magnetic reversals and intensity trends, rates and directions of plate motions, hydrothermal fluxes of elements, volcanic ash frequency, etc.)
- **Catastrophic episodes** (iridium and other impact-related anomalies, oceanic anoxia episodes, climatic excursions, etc.)
- **Time** (*absolute* ages derived from -- radiometric decay, astronomical cycles, annual varves and other methods; and *relative* ages derived from correlations to global and regional geological stage boundaries)

### Purposes and major products of *Chronos* database network

The purpose of the chronostratigraphic database system is **improved access to chronostratigraphic information** in order to provide dynamic temporal resolution for earth science data and interpretation.

- **Standard Geological Time Scale** is the centerpiece of the database, both as a specific deliverable as well as a format for underlying data retrieval.
- *This Standard Time Scale is linked to geochronologic and biostratigraphic time scales for dynamic handling.*
- *The Standard Time Scale is based on the highest quality data. An important function of the associated database is to continuously improve the precision and accuracy of this Standard Time Scale.*
- *The Standard Time Scale is agreed upon, updated and maintained by the international community.*
- *The Standard Time Scale is transparent – the foundation data and interpretation methods are given. Indeed, a user can determine the degree that each datum is critical to the Standard.*
- Database readily ties the chronostratigraphic control to the broadest range of correlative biostratigraphic, geological and physical data.

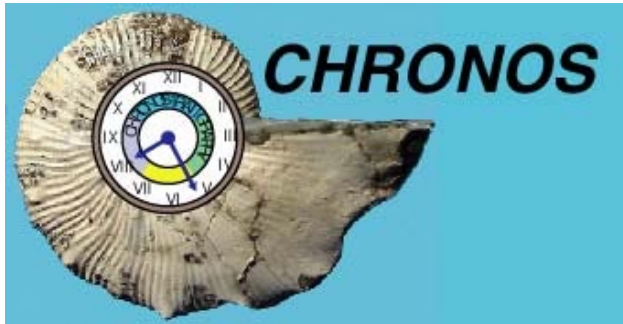
- *Database provides the capability to go to an outcrop and link it to the Standard Time Scale. The database would enable error bars to be placed on the precision of these correlations.*
- Chronostratigraphic data should be tiered to focus accuracy and precision of available datasets.
- Database structure must be dynamic as a platform for continuous improvement and revisions.
- Database should be a catalyst for defining new research directions and problems.
- Database would be linked to other Earth System and geoscience resources throughout the world, especially those which have data that bear on the time scale.

### **Benefits of *Chronos* database to Earth System research**

Major research areas that would benefit from this integrated chronostratigraphic database include:

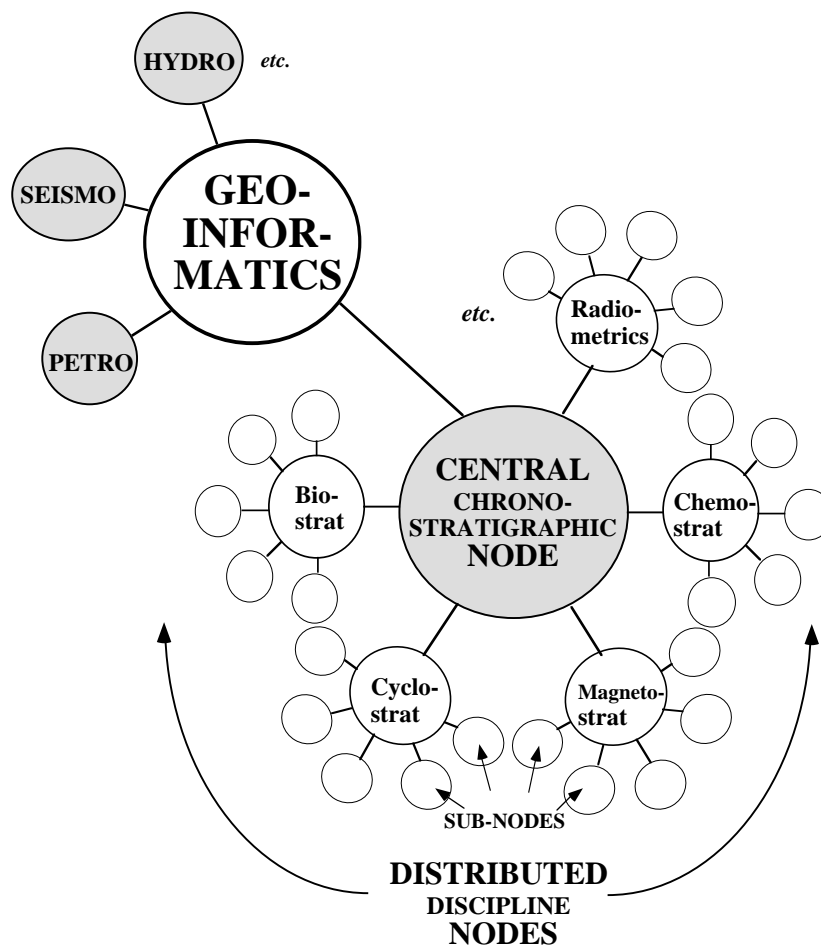
- Evolution and controls on biodiversity
- Catastrophes and abrupt climatic change (e.g., the end-Cretaceous impact, the end-Paleocene methane release, the Oligocene transition, ecologic change and recovery following critical events)
- Climate history and climate oscillations (e.g., greenhouse and icehouse trends, orbital forcing, glaciations)
- Basin modeling (sediment and elemental accumulation rates, subsidence and thermal maturity, etc.)
- Rate studies (e.g., magnetic intensity, geochemical fluxes, responses at human timescales)
- High-resolution and high-precession geochronometry (calibrated sets of decay constants, relation of tectonic and volcanic episodes to biologic events, etc.)
- Linkage of systems (e.g., terrestrial vs. marine biogeochemistry, influence of mantle processes on evolution)
- Forward projection (e.g., sea level, climate, atmosphere, and evolution trends)

All of these research areas require a time frame. In addition, the development of the *Chronos* integrated database system would advance the field of information management.



### III. Summary of *Chronos* Database System

#### Database Infrastructure



#### Central Hub and Distributed Nodes

The *Chronos* database infrastructure (schematic above) will have a central hub linked to a suite of topic-oriented chronostratigraphic databases forming Distributed Nodes (e.g., Bio, Mag, Chem, Cycles, Radiometrics, Stable Isotopes). Each of these Distributed Nodes, in turn, will coordinate the efforts of many distributed database “sub-nodes”, which are the real basis of the

entire system. An important consideration leading to this design is the efficient access and integration of data from a wide variety of sources, while requiring the least amount of costly revisions to the contributory databases.

The *Chronos* “Central Hub” will act as a data linker, integrator and manipulator, but the distributed nodes will continue to serve as the data archives. This has the additional benefit of maintaining local control over the distributed nodes, increasing the potential of data owner buy-in to the *Chronos* project. At the same time, the data archival costs also remain at the distributed node sites and do not add to the administrative costs of *Chronos* itself.

The *Chronos* “Central Hub” will connect to the overarching *Geoinformatics* network that will eventually link to “Hubs” of other major geoscience databases (e.g., seismology, petrology, etc.).

### ***Central Hub role***

The Central Node is needed to establish:

- Minimum data entry levels from expert groups
- Compatibility and formats. Ideally, interchange should be format-free.
- Establish minimum requirements for data quality (attributes, etc.)
- Contain and maintain data visualization tools, correlation tools, etc.
- Archiving center
- Home for outlying discontinued databases
- Center to handle large-volume databases
- Back-up, Trace-back functions
- Interface

Another major function of *Chronos* “Central Hub” will be to provide oversight for the distributed nodes. Such oversight entails two major responsibilities; ensuring basic data quality standards, and maintaining and promoting data format standards and translation tables. Although *Chronos* will be designed to include as wide a range of chronostratigraphic data as possible (e.g., raw and interpreted, quantitative and qualitative, academic and industrial), there will be a few basic data requirements. These include full attribution of data, ranking of data quality and error values (where possible) and a clear delineation between raw and interpreted data.

The Central Hub will also act, through its steering committee, as a facilitator of interest groups establishing common formats for their sub-disciplines. *Chronos* Central Hub and the distributed nodes will then store and update these formats as needed. An example is the development of a taxonomic synonymy and translation file for paleontology. Once format questions are largely resolved, these interest groups will then act to continually update the distributed nodes in their sub-disciplines as well as to eliminate redundant data and obsolete or superceded metadata.

From the user point-of-view, the “Central Hub” will consist of a frontispiece or front-page (possibly in the form of a stratigraphic chart) that contains the essential elements of a consensus **Standard Geological Time Scale**. From this front-page, the user can drill down to access more detailed information and link to “Distributed Nodes” with additional data specific to the major chronostratigraphic disciplines (e.g., biostratigraphy, magnetostratigraphy, etc.). There would be different portals for specialists and for public.

### ***Sub-node levels***

Sub-node levels are component databases compiled by independent investigators and groups, which are housed at various institutions. These sub-node levels will provide

recommendations and “tested” procedures for data submittal formats and other nuts and bolts. These network operation methods would be compiled and standardized (when appropriate) at node level, and shared via the central hub administration as guidelines for “best practices”. Funding, particularly at subnode level, is essential for success.

### **Administrative organization and financial support**

The organization of database components (e.g., personnel, hardware) will be modeled after that of Ocean Drilling Program (ODP), World Geologic Map (Paris), or NSIS (but with a larger, central IT advisory group). IRIS, DSDP-ODP, NOVA – examples of integrated, multi-component databases.

American Geological Institute (AGI) might serve in the role of JOI for coordinating the multi-year, multi-institutional funding from an array of international sources, and aiding in smooth collaboration with related outreach (e.g., *Tapestry of Time*, *DLESE*, Van Couvering’s time-scale information Internet site at American Museum of History, etc.) and database (industrial, government, other international) efforts.

A steering or oversight committee, with members from the ICS, academia, industry and government, will be charged with coordinating scientific community support, developing funding strategies and industry contracts, and oversight of the distributed nodes.

Aegis from the International Commission of Stratigraphy (ICS) is recommended. This global body of expert stratigraphers was founded by the International Union of Geological Sciences (IUGS) for the purpose of promoting and coordinating long-term international cooperation and establishing standards in stratigraphy.

A permanently funded, central office (although perhaps partially “virtual”) with an international, permanent core “group” of staff will run the main database system (central hub and portions of the distributed nodes). This central office will consist of both professional chronostratigraphers and IT specialists. Outside specialists in temporary project-oriented working groups would be a major component. Funding for this central office (and perhaps the distributed nodes) could be channeled through ICS (international).

The basic “sub-node” databases would rely on external grants for compilation. However, in some instances, such as the lack of any database for radiometric ages, it may be necessary for the central office to issue “development grants”.

### **User Support**

*Chronos* will maintain a user support capability through its small permanent core staff. The type of support will vary widely from helping general public inquires to aiding researchers with metadata transformations. As such, *Chronos* will establish linkages to professional volunteers in education and the various chronostratigraphic sub-disciplines in case assistance is needed for user support.

\*\*\*\*\*

## **Chronos Steering Committee**

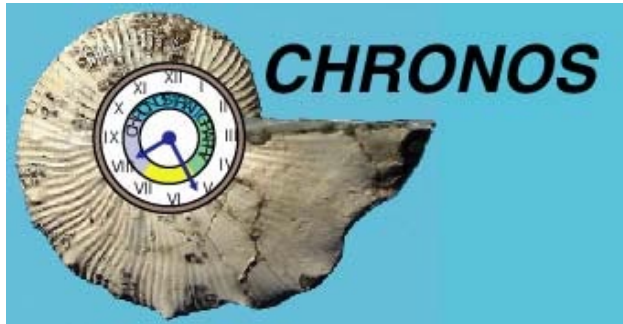
### Tasks:

1. Information compilation and distribution:
  - A. **Brochure** and Web version (new CHRONOS website at [www.eas.purdue.edu/chronos/](http://www.eas.purdue.edu/chronos/))
  - B. **White Paper** (and associated detailed “proceedings” of the Workshop) –. This will be circulated to all participants, and the final version will also be a formal report to NSF.
  - C. **Publicity** of the Workshop and *Chronos* initiative.
2. Prepare **Strategic Plan**
3. Acquire seed funding for tactical needs
4. Establish widespread collaboration and support
5. Establish *Chronos* development structure and funding coordination

### **Initial Membership of Chronos Steering Committee**

- Academic – Charles Marshall, Jim Ogg
- Government – Bruce Wardlaw
- ICS – Felix Gradstein and Jim Ogg
- Industry – Paul Sikora, Jeff Stein, Tony Gary
- Public institutions – Brian Huber (Smithsonian)
- Education – Cinzia Cervato
- Adviser (NSF) – Rich Lane

There is a 1-year “sunset” on each member’s term, so that new blood can be continuously involved in the planning, development and coordination.



## IV. Chronostratigraphic Data Issues – Types, Standards, Interpretation, Compatibility, Data Usage

### Scope of Chronostratigraphic Database

#### What data?

- Data useful for correlation (all potential correlation tools). Also assemblages.
- Use all available data, including gray literature, government reports, theses, etc.
- No time-filter –from annual varves to multi-million-year sequences
- No scale constrictions – use global, regional, and local
- Need parallel databases (geochronology, isotopes, “biology”, etc.) that are inter-linked
- Encourage funding to institutions that contribute databases for community use

#### Data types

- Two basic types:
  - Points (e.g., geochronology)
  - Sequential or stratigraphic (e.g., biostrat, chemostrat)
- Two basic forms:
  - Quantitative (e.g., physical values)
  - Qualitative (e.g., taxa assignments, facies)
- Marine versus Terrestrial
- Core versus Outcrop
- Biostratigraphy (and assemblages for paleoecology)
- Stable isotopes (C, O, S, Sr)
- Radiogenic (U-Pb, K-Ar, Sr, etc.)
- Magnetostratigraphy
- Cyclostratigraphy
- Event stratigraphy (tephra, etc.)
- Sequence stratigraphy
- Dendrochronology
- Varves
- Ice cores

### Data Issues

#### Data characteristics

- Must include access to RAW data.
  - But, what is the “Raw” data? All published “observations” are interpreted to some degree.

- Compiling Raw data:
  - *Retrospective acquisition (data behind previous interpretations)*
  - *Forward acquisition (promote standards for reporting and publishing results) from now on*
- Acceptance of qualitative observations – This is also important.
- Digital image atlas for species, and where these samples are stored. Links.
- Language not a limitation: Include a program (software) of translation from other languages for inclusion in database.
- Sequencing (e.g., Sadler techniques) is important.

### **Data quality**

- Difficult to assign quality, but minimum requirements established by Central Hub.
- Full attribution of data and source
  - Data collection
    - *Where, when, how collected, by who, and why.*
    - *Need to state what the reason was for sampling, and what are the implications and how this might limit its use for other purposes.*
  - Data storage – Where and how
  - Details of lithology and geologic context of samples
  - Processing of data
    - *Methods, instruments, sample preparation*
    - *Uncertainties, measurement precision, potential errors*
  - Reference to published data
- Published versus un-published
- Everything would have “dates of entry”
- Quality marking
  - By originator (who best knows the limitations) – why done, nature of database
  - By later users (e.g., reviews, as in some on-line book sellers).
    - *However, even though such critiques would be very useful, these might cause legal/political problems.*
    - *After debating pros and cons (including legal aspects), it was decided not to include this feature of outside ratings, even if it would be useful for future investigators. Instead, links or pointers could be included to published re-evaluations or discussions of the data.*

### **Data interpretation**

- One applies Reference Standards and Models to the basic observations (databases) to derive Interpreted Datasets
  - *Taxonomic concepts, taxonomic dictionaries*
  - *Radiometric decay constants*
  - *Magnetic stripes on ocean floor*
- The methods that raw data is processed to get interpreted datasets should be transparent.
- Filters – user can select subsets based on different criteria.
  - Interpretations can be peeled off

- Enforcement of information access (who does this?) behind age assignments applied to any chronostratigraphic data, and to the compiled geological time scale.

#### **Other data issues**

- Data structure
  - Time scale = underlying master control
  - Both qualitative and quantitative elements
  - Definitions, dictionaries, metadata styles must be agreed upon (role of central management)
- Heritage data (past publications, collections, etc.) is an interesting issue. The majority of the current biostratigraphic time scale rests on partially qualitative observations (e.g., fossils assemblages in outcrops) that are not yet in databases. National Research council study (report expected soon) deals with the problems of “heritage”, such as core storage.
- Flexibility must be maintained.
- Maintenance of data – particularly data migration – is difficult. Probably requires national and international effort. Similar problems with core repositories.
- Compatibility depends on method and users.
- Instead of trying to tie geochronology into biostratigraphy, use “good” radiometric dates in the field and do the biostratigraphy around them. “It is always easier to find fossils than a bentonite.” Such a bio-radiometric database component would start modestly and could be built up from scratch.



## V. Database Design Issues – Public Interfaces, Toolboxes, and Information Analysis

The *Chronos* system will enable users to access and retrieve relevant data, to perform visualization and simple analysis within a user workspace, and to download data (and a subset of analysis tools) into other software applications.

### Public Interfaces

*Chronos* is envisioned as an Internet-linked suite of databases that are accessed through thematic nodes and a central hub. Users would locate and display data through different forms of point-and-click Web interfaces.

### Types of User Interfaces

- Chronologic Front-page
  - Clickable charts
    - *Time scale(s)*
    - *Paleomag*
  - Linkage to "Tapestry of Time" -- paleontological information
    - *Tool for K-12 education*
    - *Chronostrat should be included in this effort from early stage*
- Geographic Front-page
  - GIS / Map retrieval (touch maps)
  - Interfaces (or partial mirrors) with::
    - *USGS Geologic Map of US*
    - *Ocean chart of ODP*
    - *UNESCO World Map*
- Thematic Front-page
  - *Biostrat, Radiometric ages, Paleomag, etc.*
- "Open" interface -- databases open to everyone on distributed nodes
  - Connecting peer-to-peer, enhancing the utility
    - *Security issues*
  - Owner of datasets would have control until finished with it, then allow access
    - *Flags in databases to control accessibility*

- User Personal Workspace
  - Work with downloaded *Chronos* or uploaded personal data in temporary "workspace"
    - *Have control of it until research is complete*
    - *Input data to workspace for using tools on central server*
    - *Example is ODP -- data is withheld until public release*

### Ways to retrieve data (Search Engines)

Search engines will be both comprehensive and user friendly (i.e., accessible to the general public). Likely possibilities include geographic retrieval through GIS map capability and graphic retrieval via mouse clicks on various chronostratigraphic displays such as time scales and paleomagnetic columns. Index and keyword searches will also be included, with user options for binary or fuzzy search parameters. Distinction will also be made between searches for descriptive or qualitative data (i.e., for basic education and general public inquiries) and raw data for scientific research.

- Drilling starting with metadata
  - Example -- Logic of the 200 Ma age assignment for base-Jurassic
  - General "drill-down" to specific data
  - Specific "drill out" across nodes and disciplines
  - "Drill up" to implications of data item
- Index or keyword searches
  - Meta-database to contain keywords, indices
  - Taxonomy -- synonymies
- Images or photographs
  - Key index fossils
  - Cores
- Queries
  - Stored or canned queries
    - *Capture frequently used queries*
    - *Options to edit parameters of saved queries*
    - *User can create canned queries*
    - *Cookies*
  - Forms to provide selections
- Concept search tools
  - San Diego Supercomputer Center has been working on such tools
  - Knowledge-based system

### Centralized and Distributed Data

- Central server
  - Contains metadata for routing to other nodes
  - Contains some "Core" data
    - *Time scales*
    - *Essential data behind "Standard Time Scale"*
    - Hosts some "orphaned" databases

- Distributed thematic nodes
  - Contain or link to raw data from individual databases
  - Contain or link to data processed through interpretation models
- One issue is the archiving of data

## Results of Searches

Search engines will enable data integration and capability for downloading. Downloads of the search results will be provided in formats (e.g., Excel, ASCII, comma-delimited files) most applicable for importation to a wide variety of applications.

- Three main pathways for users:
  - User is curious to know an answer => Return "answer" to query with supporting data
  - User wants the data for own purpose => Return basic or raw data for screen listing or output
  - User wants to examine or analyze data on-line => Return data into personal workspace, which has set of tools to do "complete job"
- Scope of data availability
  - Temporary display of holdings and relevant expanded sets for further selection by user
  - Simple visualization is coupled with data retrieval to aid user in more sophisticated selections
- Fuzzy query or fuzzy logic
  - Bring back data within bounds
  - Fuzzy logic applied (from here to here with some uncertainty applied, score applied, and retrievals are graded)
- Data compiler
  - Retrieves raw data from distributed databases
  - Aggregates data
  - Sorts data by different categories
- Data interpreter
  - Distinguishes between raw data and interpreted data
  - Ability to apply simple interpretation models (e.g., a time scale)
  - Manipulates and displays (e.g., chronogram, seafloor spreading rates)
  - Limited ability to "do the science"
- Data is exported in common "portable" forms
  - Low-level tables (ASCII, Tab-delimited)
  - Plots in PDR
  - Options for integration into GIS -- Spatial information analysis
  - For time-event data, one can choose which time scale is applied (e.g., the current *Chronos* "standard", a previous "standard", or other common published scales)

## Toolboxes, Workspace, and Other On-Line Capabilities

*Chronos* will have a smooth, sophisticated suite toolbox of applications for all users. The visualization and research tools will be flexible, and designed so that a high-school student can apply them without requiring advanced training.

Beyond basic visualization tools, *Chronos* will also have more sophisticated capabilities for data manipulation and metadata transformations. Providing licensing issues can be resolved, these will include statistical applications (e.g., SAS plots) and programs for ranking and scaling and graphic correlation.

Most tools can be smaller versions of larger commercial packages. Some tools will be available for download into the user's host computers, or they can apply them within a temporary personal Workspace. Tool development would be in public domain. In addition, output options will include format conversions to other common tools.

## On-Line Data Visualization and Manipulation

Output from searches can be put into other tools (on-line & commercial) for visualization and simple analyses. The ability to "do the science" can be accomplished in a personal temporary workspace.

- Visualization tools will be available to user
  - Two options for users:
    - *Work with tools on data in central server workspace*
    - *Download selected tools into one's personal computer*
  - Use what's available ("Buy versus Build")
    - *E.g., SAS/SAS Plot*
    - *Export in format for uploading into Excel or other personal software*
    - *License issues*
  - Make available unique tools
    - *RASC, GraphCor*
  - Develop tools that are not already available
    - *Public domain*
- Examples of visualization and analysis tools
  - Graphic presentation of integrated stratigraphies
  - Fossil ranges
  - Creation of zonations using available algorithms (Unitary Associations, RASC, Graphic correlation, General operations)

## Other Database Issues

### Dataset Submission (Direct or Indirect)

- Encourage contributions
  - Prioritization of what data should be captured
  - Standard forms
  - Thematic nodes should capture data
    - *Distributed nodes handle their own databases*
    - *Responsibilities of nodes include security*

- Availability of new data
  - Newsletter or broadcast to users of the data
    - *User-set keywords*
    - *Updates to previous users*
- Cross-correlation between databases
- Protocol for data submission
  - Minimum standards
  - Validation
  - Primary versus derived data flags
  - Maybe all formats are not totally acceptable
    - *Publicize "best practices" on methods, data forms, etc.*
    - *Excel templates*
    - *Data structure ("standard data model") on central server*
- Gatekeepers
  - Steering committees (central, and thematic)
  - Multiple levels -- central hub, distributed nodes, associated databases
- Who owns the data?
  - ODP model of 1-year monopoly
- Can you delete data after it has been referenced?
  - How to distinguish duplicate data
    - *Harland '82 time scale had Jur/Cret age partially fixed by same data point that was referenced by one study as latest Jurassic, and another as earliest Cretaceous*
    - *Intelligent software or gatekeeper for duplication/redundancy? ["Déjà vu" committee]*
  - Superseded age (e.g., assigned biostratigraphy on a formation used for paleomagnetism may be re-interpreted)
  - Erroneous data (mis-coded samples, wrong taxonomy, etc.)

## Standard Geological Time Scale

The research capabilities provided by *Chronos* will result in major new chronostratigraphic products such as a new “standard” time scale. Such a time scale will be dynamic, maintained and continually improved by ongoing research at *Chronos*, while also being interactive by providing direct linkages to the raw data and metadata interpretations that form the foundations of the time scale.

\*\*\*\*\*



## VI. *Chronos* 5-Year Plan and Budget (*summary*)

### Goal and components of the 5-year plan

The goal of the initial 5-year development program is a fully functional *Chronos* system that will allow searching and advanced manipulation of data obtained from a wide array of distributed databases spanning the entire spectrum of chronostratigraphy. The *Chronos* program consists of several simultaneous multi-institution projects working in concert (*see attached timetable*).

- (1) Central Hub functions
- (2) Development of methodology and toolboxes for interactive multiple databases.
- (3) Compilation and integration of new databases.
- (4) Time-slice projects that will serve as “test-beds” for *Chronos* interfaces, new database procedures, types of toolboxes and public access.
- (5) Public access (research and outreach) to key stratigraphic information.

#### A. Central Hub functions

The Central Hub, in addition to coordinating the thematic nodes and associated distributed databases and developing visualization and analysis tools (see details in Parts III and IV), has a major role in developing the central interface, including the following:

##### ***Central Hub information technology functions and concept***

Hosts the development and distribution of the Database Service Provider (interface to the individual databases)

Provides a means to interface to local databases

Publishes services (database interface) to central hub in a standard open method

Provides a means to dynamically create a client interface

Hosts White/Yellow/Green pages for details on the services and groups involved

Service Index Repository (UDDI interface) hosted at Central Hub provides an index of services (database interfaces)

Distribute the WSDL files that describe the interface(s) to the services registered

Support the efforts of the Thematic nodes

Hosts the development and distribution of the test clients

Reference client for general use

Provides a template for others to create clients

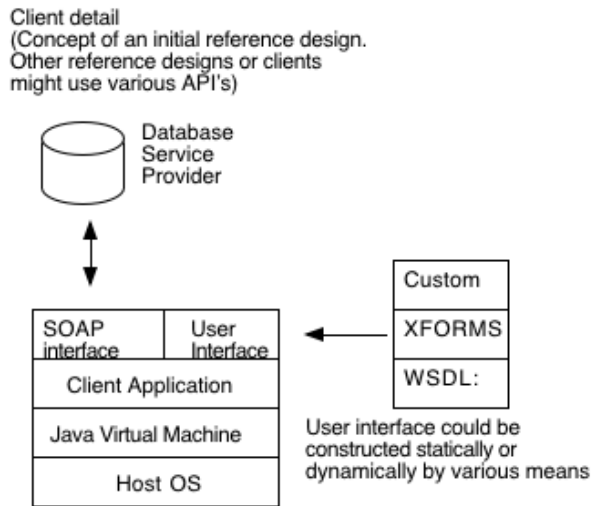
Provides a template for future rich client development

**Central Hub Standards**

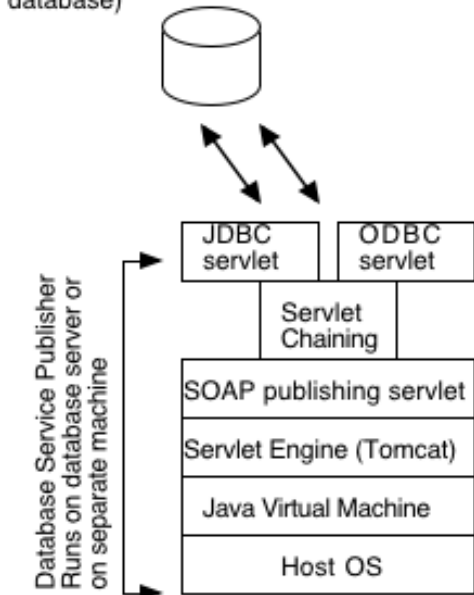
All open source / free software – Tomcat, Java (JDK), Apache, etc.

All open standards and protocols – SOAP, XML (Xforms, XSLT, etc.), UDDI, WSDL

No locking of data on certain platforms or operating systems



Database  
( Could be any OS  
and any JDBC/ODBC reachable  
database)

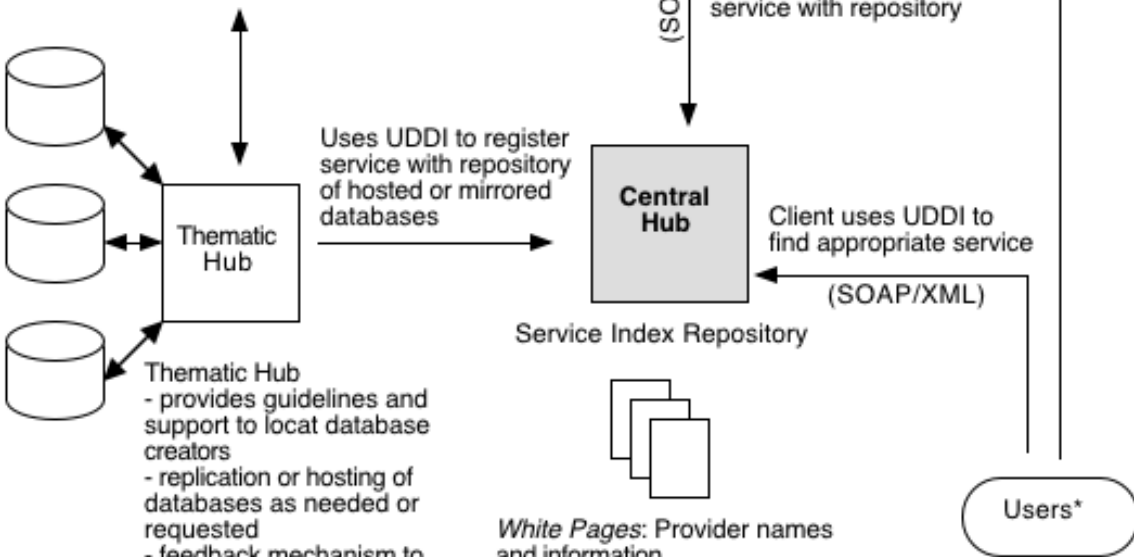


**DRAFT concept for creation of Database Service Publisher with Central Hub to coordinate services and interface methods along with initial reference client for users.**

Some Terms:

- UDDI (Universal Description Discovery and Integration)
- WSDL (Web Services Description Language)
- SOAP (Simple Object Access Protocol)
- XML (eXtensible Markup Language)
- XForms (an XML UI application)

After locating appropriate service a client talks directly to the the SOAP publishing servlet. The structure of this communication will be defined by their WSDL for the service



**Thematic Hub**  
- provides guidelines and support to locat database creators  
- replication or hosting of databases as needed or requested  
- feedback mechanism to Central Hub on development and usage issues.

*White Pages:* Provider names and information

*Yellow Pages:* Providers arranged by location, type, and protocols in use

*Green Pages:* Specific API and interface information and specific information about individual services

\*The clients used to access may be either the reference client or a custom client developed by other groups

### ***Time line for Central Hub (Information technology functions)***

#### *Year 2003:*

- Staffing and equipment integration
- Creation of initial Database Service Publisher Software Development Kit (SDK) with Java Database Connectivity (JDBC) interface to allow for publishing of data services on the net
- Draft initial XML scheme for Database Service Publisher Simple Object Access Protocol (SOAP) publisher and initial Extensible Stylesheet Language transformation (XSLT) document
- Draft of the Web Services Description Language (WSDL) spec for initial Database Service Provider SDK. This allows for a means for the automatic user interface creation on the client.
- Simple test client for Database Service Publisher SOAP publisher
- Assess the integration servlets the Database Service Publisher will need in order to interface to the databases and pass the information along via the SOAP servlet

#### *Year 2004:*

- Add Universal Description Discovery and Integration (UDDI) ability to the initial Database Service Provider SDK. UDDI will allow the databases to register their services with the central hub for any to search and access a service.
- Bring Universal Description Discovery and Integration (UDDI) repository on-line to allow the initial databases to begin publishing their services to the repository
- Creation of initial draft reference client with XForms integration
- Creation of additional Database Service Publisher database interface servlets to access a greater range of database configurations
- Draft WSDL, XForms and UDDI specs published allowing other groups to begin creating clients. In addition the source code for the initial client would be opened for all to use as a baseline.

#### *Year 2005:*

- Final WSDL, XFORMS and UDDI specs published allowing based on feedback from users of the year 2 draft
- White/Yellow/Green Pages are created and published to allow for more detailed information about the groups, organizations, databases and published services
- Continue interface servlet development as needed for Database Service Publisher
- Aggressive deployment of Database Service Publisher to the various Databases

#### *Year 2006:*

- Full deployment mode
- Rich client development (more advanced client interface mechanisms)

#### *Year 2007:*

- Full network systems support
- Publish final API specs for Database Service Publisher, Central Node systems, schemes and both the reference and rich clients

## **B. Development of methodology and toolboxes for interactive multiple databases**

There are currently several on-line databases that are primarily biostratigraphic-oriented or contain a vast amount of biostratigraphic information. The first major task of *Chronos* will be to coordinate several of these into an initial Biostratigraphy thematic node. The next steps will be to simultaneously expand into multiple thematic nodes (see section C) and construct a useful toolbox for exploring the integrated data.

2003-2005 = 7 Biostrat databases (trial thematic node)  
(EGI, Paleobiology, Neptune, USGS, ODP, ODSN, Palynodata)  
2004-2006 = Multiple thematic nodes as hub  
2005-2007 = Visualization & analytical tools

## **C. Compilation and integration of new databases**

Several major types of chronostratigraphy lack on-line databases. It is anticipated that the easiest “new” thematic hubs to construct will be key radiometric ages (estimated to be less than 1000 entries, but with an elaborate array of supporting information for each entry) and magnetostratigraphy (in conjunction with the PMAG database initiative at Scripps Institution of Oceanography). Cycle stratigraphy and geochemical stratigraphy will require workshops to plan the methodology for coding this type of information and the participating institutions. Additional types of chronostratigraphic information include ice core records, Quaternary high-resolution stratigraphy, eustatic sea-level oscillation interpretations, chemical mass accumulation rates, plate motions, volcanic episodes, and other processes. The program of expanding into all types of chronostratigraphic information will occupy *Chronos* beyond the initial 5-year plan.

2003-2005 = Radiometric, Magneto-stratigraphy  
2004-2006 = Cycle-stratigraphy, Geochem-stratigraphy  
2005-2007 = Additional (eustatic, ice cores, ...)

## **D. Time-Slice projects**

Time-slice projects will serve as “test-beds” for *Chronos* interfaces, new database procedures, types of toolboxes and public access. The following are designed to incorporate terrestrial, shallow-marine and deep-marine data spanning paleontology, geochemistry, magnetic stratigraphy, lithostratigraphy, cyclostratigraphy, eustatic variations, and other types of stratigraphic changes. The time-slices will also serve as the initial focus for data compilation and procedural testing by new nodes (radiometrics, magstrat, chemostrat, cyclostrat, etc.).

*Latest Permian-earliest Triassic* – 5 myr slice spanning the major ecosystem catastrophe, Siberian trap eruptive event, and post-extinction recovery. The main emphases are to code information in different disciplines covering a short interval and test how to set up databases to cover different stratigraphic methods. Types of information include marine stratigraphy, terrestrial stratigraphy, magnetics, radiometrics (including Siberia), possible cycles (Italy), sediment

lithologies (Meishan condensation, Italy eustatic records), geochemistry (C-13, S-34 major shifts, etc.). The simultaneous scientific question is to establish the order, relative spacing of events and absolute spacing (rates of evolution, etc.) via graphic correlation and other methods. This is the Beta-test of how to code and access different types of stratigraphic data.

**mid-Miocene** (3 full stages of Burdigalian through Serravallian) – 10 myr slice incorporating major climatic changes and eustatic sea-level swings. This is a test where there exist an immense amount of stratigraphic coverage, including industrial emphasis, yet a poorly developed biostratigraphy for many regions. Major sea-level shifts result in industrial emphasis, and possible contribution (linkage) of their datasets. This will be the Beta-test of the ability to access and manipulate a large number of datasets. A secondary feature is involvement of the petroleum industry, which has discovered several major reservoirs associated with these eustatic episodes.

**mid-Cretaceous** – 20 myr slice incorporating oceanic anoxic events that are associated with major Large Igneous Provinces (LIPs) and global warming. The emphasis is on establishing rates in an interval where radiometric control is limited. Cycle stratigraphy will be an important component (add Italian team?).

**Cambrian** – the Cambrian explosion is the most challenging, because of the lack of a well-established stratigraphy, hence the ability of the **Chronos** system to make a major impact on our understanding. Other possible “4<sup>th</sup> time-slice” intervals are the Mio-Plio boundary and events during the late Devonian.

## **E. Public access (research and outreach) to key stratigraphic information**

From the end of the initial year of **Chronos** development, the geological research community and general public will be able to access portions of the database information and other stratigraphic details. It is important that **Chronos** maintain high-visibility and useful sites as soon as possible, and progressively enhance the level of detail and data visualization/manipulation. The public outreach program will include education modules that illustrate stratigraphic and geological time, plus utilize the progressive search and visualization capabilities of **Chronos** as these become available.

2003 = Geological time scale explanation (*in association with the International Commission on Stratigraphy and site to be developed at the American Museum of Natural History*)

2004 = Biostrat node (initial 7 databases) with simple search; Geological time scale linked to pictures of key fauna-flora and paleogeography.

2005 = Complex biostrat search with analysis tools; “Mining” into geological time scale.

2006 = Temporal, geographic & thematic search into all databases with visualization tools; Instructional modules using **Chronos** system.

2007 = Visualization and analysis of all databases in one’s own workspace; Nice time-slice instructional modules.

**F. Budget Estimate** (*see attached Excel table for details by project and institution*)

<i>Task</i>	2003	2004	2005	2006	2007	Total
Central hub	400K	450K	450K	450K	450K	\$ 2.2 M
Outreach	50K	100K	100K	100K	100K	\$ 0.5 M
Biostrat node	800K	800K	400K	350K	350K	\$ 2.7 M
Other nodes	200K	500K	700K	550K	300K	\$ 2.3 M
Time-slice	200K	300K	300K	300K	100K	\$ 1.2 M
<b>TOTAL</b>	\$ 1.6 M	\$ 2.1 M	\$ 1.9 M	\$ 1.7 M	\$ 1.3 M	<b>\$ 8.8 M</b>

**F. Initial Chronos key institutions** (*preliminary suggestions*)

**Central Hub** (tasks and portals are possibly shared among more than one institutions)

*Iowa State University* (Cinzia Cervato)

**Biostratigraphy Node**

*EGI at University of Utah* (Paul Sikora, Tony Gary) – Database integration and tool development

*Harvard University and Smithsonian* (Charles Marshall, Brian Huber) – Node & *Paelobiology* database

*USGS* (Bruce Wardlaw) – *Paleodata* database

*Iowa State University* (Cinzia Cervato) – *Neptune* database

*American Assoc. Strat. Palynology* (Martin Farley) – *Palynodata* database

*Ocean Drilling Program* – *Janus* database

*Universities of Bremen and Kiel* – Ocean Drilling Stratigraphic Network (ODSN) databases

**Other Stratigraphic Nodes**

*MIT or Berkeley* (initial workshop) – **Radiometric** thematic node

*Purdue University* (Jim Ogg) and *Scripps Institution of Oceanography* (Cathy Constable) – **Magnetostratigraphy** node

*Johns Hopkins* (Linda Hinnov, initial workshop) – **Cycle stratigraphy** node

*University College London* (John McArthur, initial workshop) – **Geochemical stratigraphy** node

Other institutions for eustatic, ice core or other thematic node workshops are yet to be suggested

## **Time-Slice Projects**

*Boise State University (Permian Research Institute) and Nanjing University (China) for initial workshop – **Permian-Triassic** database assembly and trial*

*Rutgers University (?) for initial workshop – **Mid-Miocene** database assembly and trial*

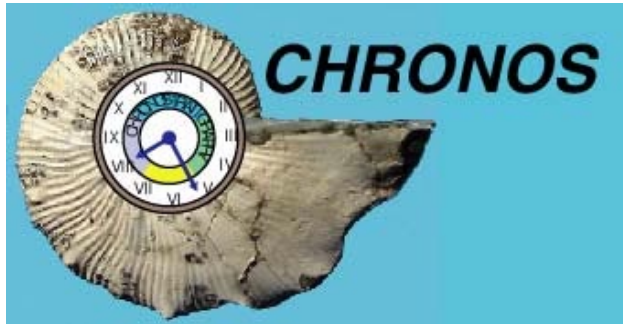
*University of Utah (Paul Sikora) for initial workshop – **Mid-Cretaceous** database assembly and trial*

*Oxford University (?) – **Cambrian** database assembly and trial*

\*\*\*\*\*

## **Acknowledgements**

The Conference Center at the University of Amherst arranged meeting facilities, audio-visual and Internet systems, hotel accommodations, coffee breaks, and luncheon and dinner banquets. Funding for conference travel, web site logistics, and facilities at University of Amherst was provided by a workshop grant from the National Science Foundation. Purdue University provided secretarial and business office support and website hosting at no cost.



## VII. Appendices

### A. Workshop Participants

<i>Name, e-mail, web</i>	<i>Mail address, Phone</i>	<i>Expertise, Role</i>
<b>BARNOSKY, Anthony</b> (Tony) <a href="mailto:barnosky@socrates.berkeley.edu">barnosky@socrates.berkeley.edu</a> <a href="http://ib.berkeley.edu/labs/barnosky/adbprofile.htm">http://ib.berkeley.edu/labs/barnosky/adbprofile.htm</a>	Museum of Paleontology University of California 1101 Valley Life Sciences Building Berkeley, CA 94720-4780 Tel: (510)-643-6275 FAX: (510)-642-1822	PALEO, coordinating on-line availability of many paleo databases (Paleontology Database Network)
<b>BOBE, Rene</b> <a href="mailto:Bobbe.Rene@nmnh.si.edu">Bobbe.Rene@nmnh.si.edu</a>	Evolution of Terrestrial Ecosystems Program Dept. of Paleobiology, NHB MRC 121 Smithsonian Institution, Washington, DC 20560 (202) 357-3033 office (messages with Paleocology lab) Geraldine E. McBrinn ETE database administrator): (202)-357-1974 (703) 750-3443 home (Bobe) Fax: (202)-786-2832	PALEO, Evolution of Terrestrial Ecosystems project
<b>CERVATO, Cinzia</b> <a href="mailto:cinzia_spencer@hotmail.com">cinzia_spencer@hotmail.com</a>	Dept. Geological and Atmospheric Sciences 253 Science Hall I, Iowa State University Ames, Iowa 50011 Tel: (515) 294-4477 FAX: (515) 294-6049	PALEO, ISOTOPE, strat databases (NEPTUNE and NORGES) with global correlation
<b>CONSTABLE, Cathy</b> <a href="mailto:cconstable@ucsd.edu">cconstable@ucsd.edu</a> <a href="http://mahi.ucsd.edu/tauxe/cathy/">http://mahi.ucsd.edu/tauxe/cathy/</a>	Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093-0225, USA Tel: (858) 534-3183 Fax: (858) 534-8090	MAGNETICS, databases and geomagnetic analysis
<b>DIVER, Patrick</b> <a href="mailto:diverpl@bp.com">diverpl@bp.com</a>	BP Exploration 501 Westlake Park Blvd., P.O. Box 3092 Houston TX 77253-3092	SYSTEMS, Stratigraphic database design
<b>DAVYDOV, Vladimir</b> <a href="mailto:C/Owsnyder@boisestate.edu">C/Owsnyder@boisestate.edu</a> <a href="http://pri.boisestate.edu/">http://pri.boisestate.edu/</a>	Permian Research Institute Boise State University MG205, 1910 University Dr. Boise, ID 83725 Tel: 208-426-???? Fax: 208-385-4061	PALEO, PaleoStrat system (specialty is Permian)
<b>FIRTH, John</b>	Ocean Drilling Program, Texas A&M	INTEGRATED, SYSTEMS,

<p>firth @odpemail.tamu.edu  <a href="http://www-odp.tamu.edu/sciops/staff/firth/">http:// www-odp.tamu.edu/sciops/staff/firth/</a></p>	<p>University            1000 Discovery Drive            College Station TX 77845-9547            Main Tel: (979) 845-2673            Office Tel: (979) 845-0507            Fax: (979) 845-0876</p>	<p>Ocean Drilling database, global view</p>
<p>GARY, Tony  <a href="mailto:tgary@egi.utah.edu">tgary @egi.utah.edu</a></p>	<p>Energy &amp; Geoscience Institute            University of Utah            423 Wakara Way, Ste. 300            Salt Lake City, UT 84108</p>	<p>SYSTEMS, strat database system, knowledge management</p>
<p>GRADSTEIN, Felix  <a href="mailto:felix.gradstein@geolgi.uio.no">felix.gradstein @geolgi.uio.no</a></p>	<p>Museum for Geology and Paleontology            University of Oslo            Sars' gate 1            N-0562 Oslo, NORWAY            Tel. +47-67-540966 (home)            Fax +47-22-854215</p>	<p>INTEGRATED, Chair of International Commission on Stratigraphy (ICS), database uses, stratigraphic correlation methods.</p>
<p>GRAHAM, Russ  <a href="mailto:rgraham@dmns.org">rgraham @dmns.org</a></p>	<p>Dept of Earth &amp; Space Sciences            Denver Museum of Nature and Science            2001 Colorado Blvd            Denver, CO 80205-5798            Tel: 303-370-8363            FAX: 303-331-6492</p>	<p>PALEO, head of FAUNMAP project</p>
<p>HINNOV, Linda  <a href="mailto:hinnov@ekman.eps.jhu.edu">hinnov @ekman.eps.jhu.edu</a></p>	<p>Department of Earth and Planetary Sciences            JOHNS HOPKINS UNIVERSITY            Baltimore, Maryland 21218, U.S.A.            Tel (410) 516-7034            Fax (410) 516-7933</p>	<p>CYCLE, Cycle stratigraphy</p>
<p>HUBER, Brian  <a href="mailto:huber.brian@nmnh.si.edu">huber.brian @nmnh.si.edu</a></p>	<p>Department of Paleobiology, MRC NHB-121            National Museum of Natural History            Smithsonian Institution,            10th and Constitution Ave., NW            Washington, DC 20560            tel.: 202-786-2658            fax: 202-786-2832</p>	<p>PALEO, ISOTOPE, foraminifers, Micropaleo centers</p>
<p>LANE, H. Richard (<i>NSF observer</i>)  <a href="mailto:hlane@nsf.gov">hlane @nsf.gov</a></p>	<p>National Science Foundation, Room 785            4201 Wilson Blvd.            Arlington, VA 22230, USA            Tel: (703)-306-1551            Home: (703)-294-6631            Fax: (703)-306-0382</p>	<p>NATIONAL PERSPECTIVE, but Lane coordinated the former Amoco strat database system and is co-chair of ICS. NSF Observer at Workshop.</p>
<p>LECKIE, R. Mark  <a href="mailto:mleckie@geo.umass.edu">mleckie @geo.umass.edu</a>  <a href="http://www.geo.umass.edu/faculty/leckie/">http://www.geo.umass.edu/faculty/leckie/</a></p>	<p>Department of Geosciences            Morrill Science Center            University of Massachusetts            Amherst, MA 01003-5820            Dept: (413)-545-2286            Off:            FAX: (413)-545-1200</p>	<p>PALEO, foraminifers, Cretaceous. Workshop Host</p>
<p>MARSHALL, Charles  <a href="mailto:marshall@eps.harvard.edu">marshall @eps.harvard.edu</a></p>	<p>Dept. Earth and Planetary Sciences            Harvard University, 20 Oxford Street,            Cambridge, MA 02138            Dept: (617) 495-2351            Tel: (617) 495-2572            FAX: (617) 495-8839</p>	<p>PALEO, Paleo database program</p>
<p>McARTHUR, John</p>	<p>Professor of Geochemistry,</p>	<p>ISOTOPES, integration of Sr &amp; C</p>

<p><a href="mailto:j.mcarthur@ucl.ac.uk">j.mcarthur@ucl.ac.uk</a>  <a href="http://www.ucl.ac.uk/GeolSci/people/mcarthur/">http://www.ucl.ac.uk/GeolSci/people/mcarthur/</a></p>	<p>University College London,  Gower Street,  London WC1E 6BT.  Off; 020 7679 2376  Dept: 020 7679 7822  FAX: 020 7387 1612</p>	<p>with paleo data, global correlation, statistics</p>
<p><b>MITHAL, Rakesh</b>  <a href="mailto:C/O_firth@odpemail.tamu.edu">C/O_firth@odpemail.tamu.edu</a></p>	<p>Ocean Drilling Program, Texas A&amp;M University  1000 Discovery Drive  College Station TX 77845-9547  Main Tel: (979) 845-2673  Office Tel: (979) 845-0507  Fax: (979) 845-0876</p>	<p>SYSTEMS, Ocean Drilling (ODP), major database design and development</p>
<p><b>MUNDIL, Roland</b>  <a href="mailto:rmundil@bgc.org">rmundil@bgc.org</a></p>	<p>Berkeley Geochronology Center  2455 Ridge Rd.  Berkeley, CA 94709  tel: +1 510 644 1350  fax: +1 510 644 9201</p>	<p>CHRONO, radiometric dating applied to stratigraphy</p>
<p><b>OGG, James</b>  <a href="mailto:jogg@purdue.edu">jogg@purdue.edu</a>  <a href="http://www.eas.purdue.edu/people/facpage/ogg.html">http://www.eas.purdue.edu/people/facpage/ogg.html</a></p>	<p>Dept. Earth &amp; Atmospheric Sciences  maincode Civil Bldg 1397, Purdue University  West Lafayette, IN 47907-1397  off = (765)-494-8681  lab = -494-0257  fax = -496-1210</p>	<p>MAGNETICS, INTEGRATED, Mesozoic time scale, cycles, magstrat, past 'digital time scale' for Phanerozoic, Secretary of ICS. Workshop Co-Coordinator</p>
<p><b>PÁLFY, Jozsef</b>  <a href="mailto:palfy@zoo.zoo.nhmus.hu">palfy@zoo.zoo.nhmus.hu</a>  <a href="http://www.pal.nhmus.hu/~palfy/">http://www.pal.nhmus.hu/~palfy/</a></p>	<p>Department of Geology and Paleontology  Hungarian Natural History Museum  PO Box 137, Budapest  H-1431 Hungary  Tel: +36 1 338-3905  Fax: +36 1 317-1669</p>	<p>CHRONO, radiometric ages on Mesozoic zones</p>
<p><b>SADLER, Pete</b>  <a href="mailto:sadler@ucr.ac1.ucr.edu">sadler@ucr.ac1.ucr.edu</a>, OR  <a href="mailto:peter.sadler@ucr.edu">peter.sadler@ucr.edu</a></p>	<p>Dept of Earth Sciences  1432 Geology Building  Riverside, CA 92521  Dept: (909) 787-3434  Office: (909) 787-5616</p>	<p>INTEGRATED, Stratigraphic statistical methods, applications to Paleozoic time scale</p>
<p><b>SCOTT, Robert W.</b>  <a href="mailto:rwscott@ix.netcom.com">rwscott@ix.netcom.com</a></p>	<p>RR3 Box 103-3  Cleveland OK 74020  Phone/fax: 918-243-7871</p>	<p>INTEGRATED, statistical analysis of stratigraphic databases</p>
<p><b>SIKORA, Paul</b>  <a href="mailto:psikora@egi.utah.edu">psikora@egi.utah.edu</a></p>	<p>Energy &amp; Geoscience Institute  University of Utah  423 Wakara Way, Ste. 300  Salt Lake City, UT 84108  tel: 801-581-4122  fax: 801-585-3540</p>	<p>PALEO, micropaleo database system. Workshop Co-Coordinator</p>
<p><b>SÖDING, Emanuel</b>  <a href="mailto:esoeding@geomar.de">esoeding@geomar.de</a></p>	<p>GEOMAR  Kiel, Germany</p>	<p>SYSTEMS, ODSN (Ocean stratigraphy network), on-line access and tools</p>
<p><b>SPIESS, Volkhard</b></p>	<p>Dept. Earth Science, University of Bremen</p>	<p>SYSTEMS, ODSN (Ocean</p>

vspiess @uni-bremen.de

Klagenfurter Strasse, P.O.Box 330440  
D-28359 Bremen, Germany  
Tel: (49) 421-218 3387  
Fax: (49) 421-218 7179  
Home: (49) 421-382 751

stratigraphy network), on-line  
access and tools

STEIN, Jeff  
steinja @bp.com

BP Exploration  
501 Westlake Park Blvd.  
P.O. Box 3092  
Houston TX 77253-3092

SYSTEMS, chronostratigraphic  
integration

TAUXE, Lisa  
ltauxe @ucsd.edu  
http://  
sorcerer.ucsd.edu/tauxe/tauxe.html

Scripps Institution of Oceanography  
Univ. Calif. San Diego, La Jolla, CA 92093-  
0220  
Tel: (858) 534-6084  
Fax: (858) 534-0784

MAGNETICS, stratigraphic and  
other applications

VEIZER, Jan  
veizer @science.uottawa.ca  
http:// www.science.uottawa.ca/  
~users/jveizer/  
http://  
www.science.uottawa.ca/geology/  
isotope\_data/

Department of Earth Sciences, University of  
Ottawa  
365 Nicholas St., P.O. Box 450, Stn. A  
Ottawa, Canada K1N 6N5  
Tel.: (613) 562-5800 ext.6461  
Fax: (613) 562-5192

ISOTOPE, carbon-oxygen array  
through Phanerozoic

and  
Institut fuer Geologie, Mineralogie und  
Geophysik, Ruhr Universitaet  
44780 Bochum, Germany  
Tel.: +49-(234) 32-23250  
Fax: (234) 32-14571

WARDLAW, Bruce  
bwardlaw @usgs.gov

U.S.Geological Survey,  
MS 926A National Center  
12201 Sunrise Valley Drive  
Reston, VA 20192  
Tel: (703) 648-5288  
FAX: (703) 648-6953

INTEGRATED, statistics applied  
to stratigraphy

**Plus, logistical help by:**

Chunfu ZHANG (Purdue University) – Web site for Workshop  
czhang6@purdue.edu, Tel: 1-765-494-0257

Steve NATHAN (Univ. Massachusetts) – Logistics and computer assistance at Workshop  
snathan@geo.umass.edu, Tel: 1-413-545-2593

**Sent Regrets:**

Kay Berensmeyer (Smithsonian), Samuel Bowring (Univ. Harvard), Lucy Edwards (USGS), Jason Hicks (Denver Museum), Steve Holland (Univ. Georgia), Susan Kidwell (Univ. Chicago), Ken Miller (Rutgers Univ.), A. Krishna Sinha (Virginia Tech.), Walter Snyder (Boise State Univ.), and Hermann Zimmerman (NSF)

## B. Workshop -- Selected Objectives and Issues

*(Initial document distributed to participants at the Workshop)*

### **Introduction -- Selected Examples of Chronostratigraphic Databases and Approaches, and Future Needs.**

Representatives of selected academic and industrial database groups will present brief overviews and demonstrations of compilations, software capabilities, application tools, and planned objectives. These include outcrop and well records, stratigraphic and paleogeographic overlays, and global temporal compilations of isotopic and biological data. Considering these current and ongoing efforts, which major types of stratigraphic data (quantitative, qualitative, temporal, and spatial) is “falling between the cracks” or requires special emphasis in the future? What components of the distributed analytical toolbox are absent? How can future growth of present databases be most efficiently handled to prevent duplication of effort? These presentations will set the stage for the working groups on Friday and Saturday.

**Knowledge Management Strategy.** For each type of database and for the entire distributed suite, a knowledge management strategy can be proposed based on desired input and output by stakeholders and an evaluation of current and future technology feasibility. The relevant task models are a key component. It is important that database contributors and users share their ‘perfect world’ desires with the developers, and that developers share their visions of future technological and analytical opportunities to encourage more creative utilization.

### Theme #1 **Chronostratigraphic Data -- Types, Standards and Compatibility**

**Stratigraphic Database Applications and Future Directions.** What would be the ideal stratigraphic database system (contents and tools) for various types of Earth system history problems? Who are the current and future users? Which problems require focused and specialized databases, and which require a broad integrated system? How can we mine the stratigraphic information of the past century and the forthcoming information of the next decades?

This major topic is part of all themes and will be a major goal of the Workshop.

**Chronostratigraphic Data Types and Standards.** Chronostratigraphic data comes in many flavors that are intimately associated and inter-calibrated with each other. How can the values and inter-relations of different data types be included into a database system? Each type of chronostratigraphic data (fossils, radiometric ages, sedimentary records of sea level or climate oscillations, paleomagnetic polarity, stable isotopic ratios, etc.) has different methods of interpretation, reliability control, and associated quality. What background information should be included, how are obsolete values removed, etc.?

**Industry – Data Sharing, Database Usage.** Research teams in the petroleum industry and government surveys have been leaders in assembling extensive and comprehensive stratigraphic databases. Examples include the former Amoco well and outcrop database (now a repository at the University of Utah maintained by the Energy & Geoscience Institute), the ‘BIOLOG’ program at ELF Exploration-Production, the Late Paleozoic biostratigraphy database for Eurasia at BP Amoco (under John R. Groves), and the Mesozoic-Cenozoic sequence stratigraphy and chronostratigraphy compilation coordinated by Exxon under Jan Hardenbol and Martin Farley (published as a chart series in SEPM Special Volume 60 in 1998).

Providing public access to portions of these confidential databases and stratigraphic tools in the industrial and government domains would be beneficial to all geoscience, and industry may gain substantially from integration of their subsets into a larger global compilation. It is only fair that industry and government surveys, which are potential major users of the stratigraphic database network, provide a share of the funding for the continued development and accessibility.

<b>Scientific Focus Group Suggested Discussion Issues</b>	<b>IT Focus Group Suggested Discussion Issues</b>
Data Quality What to include? What type (e.g., raw or interpreted)? How, who decides? What disciplines? Compatibility	Hardware Central server vs. distributed Best archival and retrieval methods Software How data stored (e.g., standard vs. multiple formats)? If standardized – best format?; import/ export issues

Paleontology – e.g., taxonomic consistency Radiometric – e.g., standardization of methodology Stable Isotope, Cycles, Paleomag – e.g., interpretation, methods Administrative Who makes these decisions?; For how long?	If multiple – integration, linkage issues, HTML vs. XML Data Entry and Type Quality control Binary vs. fuzzy Compatibility issues
---	---

## Theme #2 Database Design – Data Retrieval and Analysis

**Data Exchange and Format Compatibility.** Some systems, such as Amoco’s former in-house stratigraphic database and ODP’s site-core archive, were designed for flexible inclusion and manipulation of all types of lithologic, paleontological, geochemical, isotopic and chronostratigraphic data. Other systems have specialized tools for particular data types. Regardless of the internal database formats of each site, it would be ideal to establish a common format for efficient interchange among systems and applications.

**Public Interfaces and Toolbox.** Ease of public access, including downloading, by all geoscience researchers must be a major objective of any public-funded database program. Currently, very few systems offer this capability. Possible types of queries would include global and regional event lists, range charts and population dynamics data and paleoenvironmental data for both modern or paleogeographic reconstructions. Other possible interfaces include the abilities for researchers to upload their own data either for inclusion in the main database (after quality controls) or for trial comparisons and to manipulate stratigraphic data from different repositories.

A common Web directory with pointers to all databases and tools is essential. This Web page requires a dedicated web page administrator and communication of updates from participants. Placement or mirroring of this central Web site, selected databases and tools within the National Geophysical Data Center (NGDC) or World Data Center System is an option.

<b>Scientific Focus Group Suggested Discussion Issues</b>	<b>IT Focus Group Suggested Discussion Issues</b>
Archival vs. Application What do we expect of our database? Data Table Structure Integrated vs. discipline-segregated Quantitative, qualitative or both? Output Type Who are expected users? What are their expectations? Raw data vs. interpreted? User Issues User interface – functions? Direct or indirect? Format – how friendly? Search Engines, Cross-indexes	Data manipulation/retrieval Gathering data - distributed databases (e.g., NASA) vs. central database Output retrieval – data compiler vs. data interpreter User Issues User interface – type? Direct or indirect? Format – how friendly? Search Engines Cross-indexes

## Theme #3 Integrated Chronostratigraphic Databases – Future Directions, Needs, Frameworks

**Database Administration Mechanisms and Issues.** How are the quality and reliability of database entries determined? Some paleomagnetic databases contain subjective ‘quality’ factors, and perhaps similar cautions are required for certain types of geochemical data and radiometric ages. A taxonomic dictionary and equivalence table is important for paleontological databases, and it is probable that other standardizations are required for chronostratigraphic data (e.g., monitor standards and decay constants) and other stratigraphic information. Temporary confidentiality may be a requirement for industrial well data or ongoing individual research efforts. Another aspect is intellectual property issues for data and tools.

**Visions, Strategic Goals and Timetables.** Where do we wish to be with stratigraphic databases at the end of this decade and the next decade? What are the required innovations, data acquisition, user interfaces, manipulative tools, infrastructure, and other components for these information systems? Should stratigraphic data be available via a

distributed network or by a centralized repository? An itemized task list, a timetable for actions, and a series of benchmarks to measure accomplishments will be proposed. At this planning stage, the strategic goals of an ideal program should not be hindered by potential financial and logistical constraints.

**Challenges and Issues.** Within the framework of an idealized development program for stratigraphic databases, what are the potential bottlenecks, software and hardware hurdles, new data types, financial and logistical shortcomings, and needed innovations? What is required to overcome these technical problems?

For the inter-compatibility and accessibility of distributed databases, what are the major issues concerning cooperation, merger and redundancy of information, open data contribution, intellectual property, confidentiality, review and certification of data reliability, infrastructure, and other community concerns? Can we recommend structural procedures and safeguards to solve such issues?

**Roles of NSF and International Stratigraphic Commission (IUGS) for Infrastructure and Standards.** Both common standards and funding are required to achieve the dual goals of a smoothly integrated, yet distributed, comprehensive suite of stratigraphic database repositories and efficiency in access and manipulation of data by international research groups. NSF and other agencies have provided partial funding for several databases, and these should be integrated into a common infrastructure when any temporary requirements for confidential usage have passed. Perhaps NSF and other agencies should request that such public-funded databases include options for search engines and downloading information in a common format. The International Commission on Stratigraphy (ICS) has proposed a Working Group on Stratigraphic Databases to coordinate international access and utilize the progressive compilations in updating the integrated geological time scale of stratigraphic events. Under the IUGS, the Commission on Management and Application of Geological Information (COGEOINFO) ‘*establishes earth-science data standards and facilities and stimulates the full and wise use of geoscience information in the definition and solution of regional and global problems*’ (IUGS Strategic Planning Committee, 2000); therefore, presumably, future stratigraphic databases should strive to adhere to their recommended standards, and we should help guide the development of these IUGS standards.

These and other infrastructure concerns will be some of the concluding topics for working teams. This Workshop will play a vital role in providing future visions for an effective distributed database network and recommendations for achieving these goals to NSF, ICS and IUGS.

<p align="center"><b>Scientific Focus Group Suggested Discussion Issues</b></p>	<p align="center"><b>IT Focus Group Suggested Discussion Issues</b></p>
<ul style="list-style-type: none"> <li>Administration               <ul style="list-style-type: none"> <li>Central vs. distributed</li> <li>Who? (e.g., coordinator, committee?)</li> <li>Who chooses administrator(s)?</li> </ul> </li> <li>Permanent vs. rotating</li> <li>Funding               <ul style="list-style-type: none"> <li>Cost estimate (within order of magnitude)</li> <li>Sources – short term                   <ul style="list-style-type: none"> <li>High costs</li> </ul> </li> <li>Sources – long term                   <ul style="list-style-type: none"> <li>Minimal costs?</li> <li>Ensure database permanence and quality</li> </ul> </li> </ul> </li> <li>Continued development               <ul style="list-style-type: none"> <li>Growth and applying new technology</li> </ul> </li> <li>Prioritization               <ul style="list-style-type: none"> <li>What’s the plan?</li> <li>Benchmarks, definitions for success.</li> <li>Timetable; when is transition from short- to long-term growth?</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Administration               <ul style="list-style-type: none"> <li>Central vs. distributed</li> <li>Who? Where? How?</li> <li>Short- vs. long-term stages</li> </ul> </li> <li>Funding               <ul style="list-style-type: none"> <li>Sources                   <ul style="list-style-type: none"> <li>Short- vs. long-term</li> <li>Basic vs. developmental</li> </ul> </li> <li>Long-term issues                   <ul style="list-style-type: none"> <li>Obsolescence prevention – periodic inventories?</li> <li>Process for continuing data entry.</li> <li>Insurance of data preservation</li> </ul> </li> </ul> </li> <li>Prioritization               <ul style="list-style-type: none"> <li>IT benchmarks</li> </ul> </li> </ul>

## **C. Workshop Schedule and Agenda** *(Logistical details have been deleted)*

### **INTEGRATION OF CHRONOSTRATIGRAPHIC DATABASES FOR THE 21<sup>ST</sup> CENTURY**

**Conference Center, University of Massachusetts, Amherst, MA  
November 8-10, 2001**

#### **THURSDAY, NOVEMBER 8 – INTRODUCTION AND GOALS**

- 12:30 – Welcome, introductions, logistics – Jim Ogg & Paul Sikora
- 12:45 – Perspective of the NSF and ICS – Rich Lane
- 1:15 – Review of agenda and workshop goals – Group
- 1:30 – Current and future trends in knowledge management – Anthony Gary

#### **SELECTED EXAMPLES OF CHRONOSTRATIGRAPHIC DATABASES/APPROACHES**

[Moderator – Jim Ogg]

- 2:00 – Database presentation 1 (ODSN, Söding & Speiss)
- 2:30 – Database presentation 2 (EGI, Sikora)
- 3:15 – Database presentation 3 (ODP, John Firth & Rakesh Mirthal)
- 3:45 – Database presentation 4 (Paleontology Database Network, Tony Barnosky)
- 4:10 – Database presentation 5 (PaleoData, Charles Marshall)
- 4:35 – Database presentation 6 (Neptune and Norges, Cinzia Cervato)
- 5:00 – Database presentation 7 (PaleoStrat, Vladimir Davydov)
- 5:25 – Formation of working panels for Friday
- 5:30 – *Reception*

#### **FRIDAY, NOVEMBER 9 – FOCUS WORKING GROUPS**

##### **THEME 1: CHRONOSTRATIGRAPHIC DATA -- TYPES, STANDARDS, COMPATIBILITY**

- 8:00 – Review of goals; Introduction to chronostratigraphic data topics  
Main moderator = Mark Leckie, Main recorder = Jan Veizer
- 8:15 – Five panels (overlapping) in two focus groups – Discussion of common issues, possible solutions
  - Scientific #1 (moderated by Roland Mundil)
  - Scientific #2 (moderated by Brian Huber)
  - Scientific #3 (moderated by John McArthur)
  - Info-Tech #1 (moderated by Cathy Constable)
  - Info-Tech #2 (moderated by Pete Sadler)
- 9:30 – Summarization of interim conclusions within each panel
- 10:15 – Convening of the two focus groups (Scientific, Info-Tech) for preparing “position statements” that summarize issues and associated recommended solutions.
- 12:00 – *Workshop Luncheon*
- 1:00 – Presentations of issues and recommendations by focus groups on chronostratigraphic data topics  
Discussion and partial consensus by entire Workshop

##### **THEME 2: DATABASE DESIGN – INFORMATION ANALYSIS, RETRIEVAL, EXCHANGE**

- 1:30 – Introduction to database design topics  
Main moderator = Paul Sikora, Main recorder = Lisa Tauxe
- 1:45 – Five panels (overlapping) in two focus groups – Discussion of common issues, possible solutions.
  - Scientific #1 (moderated by John Firth)
  - Scientific #2 (moderated by Jozsef Palfy)
  - Scientific #3 (moderated by Robert Scott)

- Info-Tech #1 (moderated by Patrick Diver)
- Info-Tech #2 (moderated by Rakesh Mithal)
- 3:00 – Summarization of interim conclusions within each group
- 3:45 – Convening of the two focus groups (Scientific, Info-Tech) for preparing “position statements” that summarize issues and associated recommended solutions

### **SATURDAY, NOVEMBER 10 – INTEGRATION AND WORKSHOP RECOMMENDATIONS**

- 8:00 – Presentations of ideas and recommendations by focus groups on database design topics
- Discussion and partial consensus by entire Workshop

### **THEME 3: INTEGRATED CHRONOSTRATIGRAPHIC DATABASES – FUTURE DIRECTIONS, NEEDS, FRAMEWORKS**

- 9:00 – Introduction to integrated database topics
  - Main moderator = Jim Ogg, Main recorder = Linda Hinnov
  - Scientific #1 (moderated by Felix Gradstein)
  - Scientific #2 (moderated by Bruce Wardlaw)
  - Scientific #3 (moderated by Jeff Stein)
  - Info-Tech #1 (moderated by Russ Graham)
  - Info-Tech #2 (moderated by Emanuel Söding)
- 10:30 – Convening of the two focus groups (Scientific, Info-Tech) for preparing “position statements” that summarize issues and associated recommended solutions
- 11:00 – Presentations of ideas and recommendations by focus groups on integrated database topics, and items of consensus by entire Workshop
- 11:30 – Workshop Lunch*
- 12:00 – Synthesis discussion and finalization of Workshop recommendations and future plans
- 1:00 – Formal portion of workshop adjourns**
- 1:30-4:30 pm Meeting rooms will remain available for members of different focus groups and conference conveners to discuss and prepare the Workshop report
- 5:30 – Informal gathering for those staying through Saturday night*