OC3 Meeting Bern Notes from breakout group 02.10.2014

1   Data included

1.1   Core top data?
Potential problems with core tops: age models, non-analogue situations, mixing of sediment surface due to bioturbation
→ Only core tops with age control will be used for database

1.2   Down core data
Concentrate on deglacial but also include older time series
Data resolution: Anything that is available or specific resolution needed?
→As long as data is well documented it can be included in database

2   Database product/structure

2.1   Structure of database
Ideal: Link all data with core name and core metadata to get easy access to different data files (Age models, depth models, stable isotope data etc.)
two different structure ideas
1. "Master database" with core metadata that is linked to age models and datasets (Eric Galbraith, Olivier Cana, Tim Bolliet)
2. Spread sheet (e.g. EXEL) with divers pages

Problem with all databases: due to changing age models and increasing datasets a flexible permanently changing database is needed

2.2   End product

2.2.1   *Database that compiles the existing data and is "finished/closed"*
WOCE like database with update every X years?
→ Associated existing project (Stefan Mulitza): Project to program ODV-like software ISOMAP for paleodata (proposal for funding for software engineer submitted), with this funding project will be maintained for 4 years
Open Question concerning WOCE like database*: will all existing data be included in Paleo ODV database or will low quality data be erased during updates?*

2.2.2   *"open" database that is permanently updated*
•   Crowed sourced wiki- like framework: flexible database data can be inserted into database forever
•   Pangea/NOAA- like

Questions concerning an "open" database:
How are users enabled to upload new data in the future?
Who maintains this database in the future?

3   Funding

Can we get funding for one person to assemble database?

Pages: only supports supporting meetings and infrastructure, no funds for salaries

NSF: Individual grant proposal to NSF etc. if someone had project related to OC3→ important to point out group interest

NSF: Joint proposal: Several PI and a lot of collaborators in proposal for funding

US Earth cube initiative: leverage from non US citizen to get somebody to do exchange to US

→ Multiple funding sources might work

→ use good public relation of Pages with NSF

4   Spreadsheet

Product database should include:

1. Raw data (minimum amount of columns,)
2. derived products (mandatory)
3. metadata (see 4.1.)
4. age models (see 4.2.1.)
5. depth models (see 4.2.1.)

4.1   Metadata obligatory

- Latitude
- Longitude
- Depth (original, composite, corrected)
- Coring technique
- Species
- analysed material

4.2   Metadata additional

- kind of core
- kind of treatment
- kind of storage (e.g. cool storage with temperature)
- link every core to original publications; Citation to core descriptions including core photos etc.

4.2.1   *Age models/ depth models*

Original depth in the core barrels, often corrected to

- account for missing core tops
- stack different parallel cores together
- account for core stretching that occurs in piston cores

In about 20% of the published data (Pangea, NOAA) only age scales are given and age tie points are missing or it is unclear whether depth refers to original depth within core barrel or depth corrections have been applied

→ Different age/depth models have to be included in database and easily be accessible

Quality control age models

Obligatory:

raw data: AMS original lab data, lab reference, type of species used for AMS,

processed data: calender age, reservoir age

tuned age model   depth/age tie points and with explanations (e.g. aim used for tuning)

Quality control depth models

Original archive depth needed, core depth, depth in sampling barrel the user must be able to find each sample back in the archive

If different depth scales are used, age model should include tie points to both depth scales (depth in core barrel mandatory!)

Open Question: Who will put metadata into database? Who checks evaluates data/metadata for potential errors

5   Quality control

How do we evaluate datasets? Do we need a quality control?

- all data MUST have a depth column that allows the user to find the sample back in the achieve (depth in core barrel), datasets that only include composite depth or age models have to be excluded
- Column for comments behind data that should be used to comment on uncertainties in data production e.g. quality of the core (e.g. hidden events like turbidates, wholes, hiatus), quality of data (e.g. lab problems, pot. contamination)
- Control by data flagging
    - Data flagging in/out option (0/1) → analyses program excludes potentially biased data from use, but those data does not have to be erased from dataset
    - Qualitative data flagging
      Problem: how to be consistent?
      Subjective view of person in charge that even might change over process of evaluation
      → set of criteria for data evaluation would be needed (e.g. concept of MARGO)

6   Existing databases compilations

Tim Bolliet and Oliver Cartapanis already compiled datasets from existing databases (Pangea, NOAA) with following problems

- ~ 20% of published data can not be linked to original depth in achieve as published with age model or composite depth, tie points not included,
- incorrect data (not necessarily controlled by pangea data manager) ("rubbish in- rubbish out")
- missing data quality control in most datasets no data flags

→ how identify problems? Which records shall we put aside?

7   Approach

- Start with limited amount of cores
- existing databases could be used as spin off

→ Join existing data compilations of Olivier, Tim, Lorraine

→ Add unpublished data of Alan, Nick

→ Get people to search for unpublished data within home institutes

→ Search for unpublished $\delta^{13}C$ data of published $\delta^{18}O$ data records

Question

How do we get people to submit data?

→ Co-authorship for all data contributors