



DCCD Meeting 9 & 10 April 2009

Theme: building the DCCD

Minutes of the meeting

De Amershof, Snouckaertlaan 11, 3811 MA Amersfoort

April 9th

Participants

Sarah Cremer (SC), Sjoerd van Daalen (SD), Marta Dominguez (MD), Pascale Fraiture (PF), Kristof Haneca (KH), Esther Jansma (EJ), Rowin van Lanen (RvL), Maarten Leenders (ML), Benno Ridderhof (BR), Ronald Visser (RV), Yardeni Vorst (YV), Frans van der Zande (FvdZ).

[Absent: Remi Brageu, Peter Brewer, Paul Copini, , Patrick Hoffsummer, Ute Sass-Klaassen, Menne Kosian, Catherine Lavier, Erhard Pressler, Willy Tegel, Dirk de Vries, Milco Wansleben, Tomasz Wazny, Ronald Wiemer, Ivo Zandhuis]

09.30 – 10.00 Update: activities since October 2008 (Esther Jansma)

Esther Jansma presented a short overview of the activities between October 2008 and April 2009. Among these activities were: updating data to DANS, the functional design of TRiDaS, development of a TRiDaS based Access database, developing a thesaurus, digitizing old data, website maintenance, pre-proposal RITRIS (COST), start-up development of the web application and (within RING) sorting out data and metadata.

During this presentation EJ raised the question if someone tried to access their own data delivered to DANS and if this was even possible? It should be possible to view your own submitted data. This question will be forwarded to DANS. [EJ]

The participants would like to view the DCCD functional design and the plans of DANS, this will be emailed to them. [RvL]

YV: could not find the submit button on the TRiDaS website (www.tridas.org). We should check if this was an incident or if the location of this button should be improved. [RvL: Location checked. Submit button is there.]

10.00 – 10.30 Developing TRiDaS 1.1 (Esther Jansma)

Esther Jansma presented the development of TRiDaS 1.1. including the current situation and reviews on the manuscript containing TRiDaS 1.0. The participants also briefly discussed some of the main fields within TRiDaS. This discussion was continued later in the meeting.



11.00 – 11.45

Converting tree-ring data and metadata to XML: introduction of Access TRiDaS database (Rowin van Lanen)

Rowin van Lanen presented the beta release of his TRiDaS based Access database (TRiDaS Database). This database aims at incorporating old and new tree-ring data and metadata, following the TRiDaS filing structure and exporting the data and metadata to the XML format. After a short presentation of the structure of the database and the used forms, the flexibility of the data structure itself, promoted by TRiDaS, was explained. The TRiDaS data model makes the database easily accessible for a variety of dendrochronological users. TRiDaS specifies an independent value level which is not incorporated in the current TRiDaS database. Rowin van Lanen clarifies this: To give each value an unique record identifier would lead to a database in which querying would prove sluggish and difficult. Because this option will be available in the DCCD application it is not a necessity at this moment. To still be able to view the different values, a memo field in MeasurementSeries and DerivedSeries has been added in which ASCII based values can be copied. The proposed flexibility of TRiDaS leads to unrestricted data fields, which can lead to uncontrolled data entry. Therefore Rowin van Lanen emphasized the necessity of controlled vocabularies and inputmasks in the used forms, if the datafields itself did not provide in this.

FvdZ: Regarding the 'file' fields in TRiDaS, it is possible to put one or more file(links), such as pictures, maps and documents, in a XML scheme. In the current Access database this is not an option, how is this going to work? RvL: In the future this will be possible in the DCCD application, Access does not provide in a functional data warehouse for graphical data. This would 'explode' the database. In the current TRiDaS database people will be restricted to linking to their pictures and digital files.

11.45 – 12.30

Using a controlled vocabulary within a TRiDaS database (Maarten Leenders)

Maarten Leenders provided a short overview of his activities which partly are focused on creating a controlled vocabulary for the TRiDaS Database and DCCD application. Based on several controlled vocabularies available to us, like the Archaeological Basic Register (ABR) and the Art and Architectural Thesaurus (AAT) he presented a vocabulary list for object and element types.

SD: In what way do you handle object functional and type differences in a controlled vocabulary? For example: a *grain mill* is the functional description of the object, but its object type is a watermill? This is possible within the fields of TRiDaS: object.description and/or object.type.

KH: Isn't object.Type:IsPartOf the same as element. EJ agrees and suggest changing this field in Substructure or Subobject. This field, Type:IsPartOf is not a TRiDaS field!!

SD: CoverageTemporal should also include geological periods. All agree.

YV: In project.Category Maritime Archaeology should be replaced by Ship's Archaeology. Maritime Archaeology is a very broad term which incorporates a variety of sub disciplines, ship archaeology is a better term. All agree and this will changed in the vocabulary.

FvdZ: Is it possible to include terms which are not recognised by the vocabulary? We should strive to use a controlled vocabulary in TRiDaS but think about how to add new additions.

13.30 – 15.30

Hands-on: TRiDaS database (Rowin van Lanen)

In this section Rowin van Lanen presented a hands-on approach to the TRiDaS database. During this presentation the different fields within the database and TRiDaS have been discussed.

MD: Why is project.category a mandatory field? YV: We decided in previous discussions to set these mandatory fields because we wanted to ask on these fields. For example: give me all dating projects.

RvL: Can we create a picklist for these project types:

- Dating
- Provenance



- Tree age estimation
- Forest dynamics
- Forest management studies
- Climatology
- Forestry
- Woodtechnology
- Paleo-ecology
- Geomorphology
- Other

EJ: We should email these picklists to the community for additions.

YV: We should add the field " Other" to the picklist. In combination with project.description this can be used to administrate projects not entered in the list. RvL: This will be added.

MD: It should be possible to add more than one type to a single project. For example dating *and* provenance. RvL: This will be added.

KH: Living Trees should be changed to Standing Trees, because trees still standing can also be dead. All agree.

MD: It should also be possible to add more than one category to a project. For example: Archaeology and Living/Standing Trees. RvL: It will be added.

FvdZ: Period within project should have a begin and end date. Preferably through a input mask using Dublin Core.

RV: LastModifiedTimeStamp should be locked within the record.

MD and RV: Modifications in the mandatory fields should be verified by a message. For example: "Are you sure you want to alter this field".

SD: RadiusTitle and Identifier, SampleTitle and Identifier, MeasurementSeriesTitle and Identifier and DervidSeriesTitle and Identifier are basically the same. RvL: This is mandatory for TRiDaS. SD: Why is it mandatory for TRiDaS to get a title as well as a identifier? [This will be discussed with the TRiDaS group – RvL]

SD: We should also add a field which registers the observed but unmeasured rings. Rings you can see but you can not measure. [Will be added and discussed with the TRiDaS group - RvL]

15.30 – 15.50 **Digitizing paper tree-ring data: the ROB, BAAC and RDMZ archives (Benno Ridderhof)**

Benno Ridderhof presented his work on digitizing old, mainly archaeological, data and his involvement in the DCCD project at this moment: digitizing the old dendrochronological data of the ROB, BAAC and RDMZ archives.

BR and EJ: Dendrochronology should have its own official policy in data preservation and data storage. BR: Medieval Dorestad shows the need of preservation, data loss through lack of a policy, for archaeology such a policy has been created, but dendrochronological data still lacks an official policy.

SD: BAAC has a policy regarding preserving and archiving archaeological data but not regarding dendrochronological data.

15.30 – 17.00 **Workshop: DCCD thesaurus (art/furniture, architecture, archaeology, natural vegetation (paleo-ecology, living trees))**

In this section the discussion regarding the building of a thesaurus continued. Suggestions, translations and additions to the thesaurus were determined.



All: The need of a controlled vocabulary is evident.

FvdZ: We should think about how to incorporate new additions or exceptions into DCCD and eventually TRiDaS.

FvdZ, RV and RvL: Because of the flexibility in the TRiDaS fields and especially in the different levels of Project, Object and Element, we should strive to create one thesaurus "cloud" in which the complete thesaurus and related terms are combined.

The discussion was cut short by the clock striking 17.00. The discussion was planned to continue on the second day of the meeting: April 10th.

April 10th

Participants

Paul Copini (PC), Sarah Cremer (SC), Sjoerd van Daalen (SD), Marta Dominguez (MD), Pascale Fraiture (PF), Esther Jansma (EJ), Menne Kosian (MK), Rutger Kramer (RK), Rowin van Lanen (RvL), Maarten Leenders (ML), Ronald Visser (RV), Ronald Wiemer (RW), Frans van der Zande (FvdZ).

[Absent: Remi Brageu, Peter Brewer, , Kristof Haneca, Patrick Hoffsummer, Ute Sass-Klaassen, Catherine Lavier, Erhard Pressler, Benno Ridderhof, Willy Tegel, Yardeni Vorst, Dirk de Vries, Milco Wansleben, Tomasz Wazny, Ivo Zandhuis]

09.30 – 10.30 **Building the DCCD website application (Rutger Kramer)**

Rutger Kramer introduced DANS to the participants. He explained the aims of DANS as an organisation and its role within the DCCD project. DCCD will be a Trusted Digital Repository which eventually (well in the future) will be able to communicate with other Trusted Digital Repositories. The DCCD application will be based on Fedora (Flexible Extensible Digital Object Repository Architecture) Common. Fedora is open source and can store any kind of data. DANS will build the DCCD application through several iterations, which will each result in a functioning application. To inform the DCCD community of the progress and receive feedback, a special DCCD website with discussion forum will be created. DANS is planning to finish the application in July 2010 after which a follow up on the application will continue until October 2010. The DCCD application will be open source and available for improvements throughout the dendrochronological community.

11.00 – 12.30 **Workshops: what questions will we ask to the application, which answers do we expect (files (content, format), maps etc.))**

This section was dedicated to the discussion about the required possibilities of the DCCD application. Questions were raised about the future data storage possibilities.

MD: Will DCCD be our online database? EJ, MK, RV: This would not be wise, local databases will still be used. RK: With the possibility of frequent update to the online DCCD application.

EJ: This would mean that a DCCD application for local offline data storage and periodic, easy update possibilities to DCCD online has to be developed by DANS. RK: Agrees.



SD: Is it possible for the researcher to determine the access level of specific users to his or her dataset? Is it possible to set restriction to queries, for example you can only retrieve the data regarding the Roman period [All: Yes].

SD: It should be possible to share created queries with the community (data sharing within DCCD).

MK: DCCD will need a dynamic query builder for formulating queries on variety of fields.

SD: DCCD will need export possibilities focused on the current popular text-based formats (FH, Tuscon etc.) EJ: The DCCD application will need a translation module for import and export of the common dendrochronological formats.

14.00 – 15.00 **Workshops: (a) what questions will we ask to the DCCD application, which answers do we expect (files (content, format), maps etc.); (b) thesaurus**

In this section the discussion continued on the future DCCD application. The following subjects were discussed and approved by the participants:

1. Repository
 - a. Allows sustainable management of data
 - b. Allows changes (by authors of the data and later users)
 - i. Versioning; part of TRiDaS!
 - ii. New versions include descriptions of recent changes
 - iii. Queries only on latest version
 - iv. Definition of roles and responsibility in labs
2. Webapplication
 - a. General
 - i. Membership, to assign rights
 - ii. Back ups of content
 - iii. Gives overview of data in repository (even if part is invisible)
 1. General and descriptive introduction of DCCD: goal, participants, content. Statistics (of data, use..)
 2. Show fields of (meta) data that in principle can be retrieved
 3. WARNING: AVOID POSSIBLE EXPLOITS
 - iv. Possibility to link DCCD to external XML tree-ring archives
 1. Contact DANS-Cornell (Corina)! NB. Peter Brewer will visit us on 8-9 June 2009.
 - v. Possibility to link parts of DCCD to non-dendro web applications (e.g. MACHU)
 1. Identify other websites and parts we want to link [All, communication to EJ]
 2. Contact [EJ and RvL]
 - vi. Possibility to store associated documents (pictures etc.)
 - b. Database functionality
 - i. Possibility to use archiving services of DCCD to store own data and use it as local TRiDaS database)
 1. Take this in consideration in project definition and while programming; this will imply extra costs after completion (maintenance etc)
 2. Local and internet version of DCCD (also compare alternative ii below)
 3. Periodic synchronization of local database with DCCD



- ii. Alternative: offer stand-alone TRiDaS-conform database. In this case the developed TRiDaS Access Database of RvL can function as a free introduction to the TRiDaS structure, after which members of the DCCD would be able to use a stand-alone local DCCD application with a periodic update to the DCCD web application.
- c. Content (TRiDaS)
 - i. The system (query engine) should be flexible regarding the data model
 - ii. Possibility to edit existing data
 - iii. MEMO Consensus about within-ring information is next step for TRiDaS (fire scars, frost rings, early and latewood etc.)
 - iv. MEMO pith and sapwood information (now under radius) should be under measurementSeries NB. Is now solved in TRiDaS v1.2.
 - v. MEMO field with code of representative series (MEMO: often in DCCD for dating purposes) under Sample. NB Is now solved in TRiDaS v1.2.
- d. Access/queries
 - i. Reusable queries and possibility to share query formats
 - ii. Ability to apply report format (MEMO: report builder) and share it with others MEMO big work!
 - iii. Receive requests for data uses through fixed channels within the system
 - iv. Output after query: full TRiDaS XML plus user defined measurement file
 - v. MEMO: a new project always contains new data (series or derivedSeries)
 - vi. Re-use of old data: DCCD recognizes this and makes a link
 - vii. Possibility to query on the value level
 - viii. Possibility to ask for values associated with specific years
- e. Access restrictions:
 - i. MEMO Keep restrictions when looking for protected data. MEMO give list of protected-data owners (see avoid exploits)
 - ii. Work with several pre-defined levels of access (maintenance level, data contributor, non-dendro scientist, student, general public)
 - iii. Data owner gives permission per person, specified fields and values, using default settings and deviations from these settings
 - iv. Possibility to specify licence agreement per data owner and/or data set, Word document or pdf
 - v. Restriction not only on level, but also on content (e.g. period and object type)
- f. Uploading and downloading metadata and data
 - i. Possibility to upload and download data in different formats; download: metadata in TRiDaS and data in TRiDaS or other format
 - ii. Data uploading will be possible in three ways: (a) TRiDaS, (b) metadata in TRiDaS and measurement data in other format, (c) metadata input through form + measurement data in other format
 - iii. Possibility to upload and download data (series) in at least seven common formats (fh, V-format, CNRS/Besancon, Dendro32, Tuscon, Catras, TRiDaS)
 - iv. User needs to accept licence agreement before downloads
- g. Conversion of measurementSeries and derivedSeries
 - i. System should do data conversion



EJ: DCCD should give the opportunity of geographical searches. This is a strong wish of the community. DCCD should have the possibility to add GIS functions later on and give a basic spatial support from the start.

**15.15 – 17.00 Presentation of workshop results, discussion, planning of next steps
(e.g., thesaurus, data deliveries, COST)**

Because of the duration of the discussion on the DCCD application the planned thesaurus discussion will be postponed. This discussion will continue mainly through email contact.