

# Preparation of a comprehensive palynological norwegian **dataset** for digital **re-analysis** and **publication** as **Open Data**

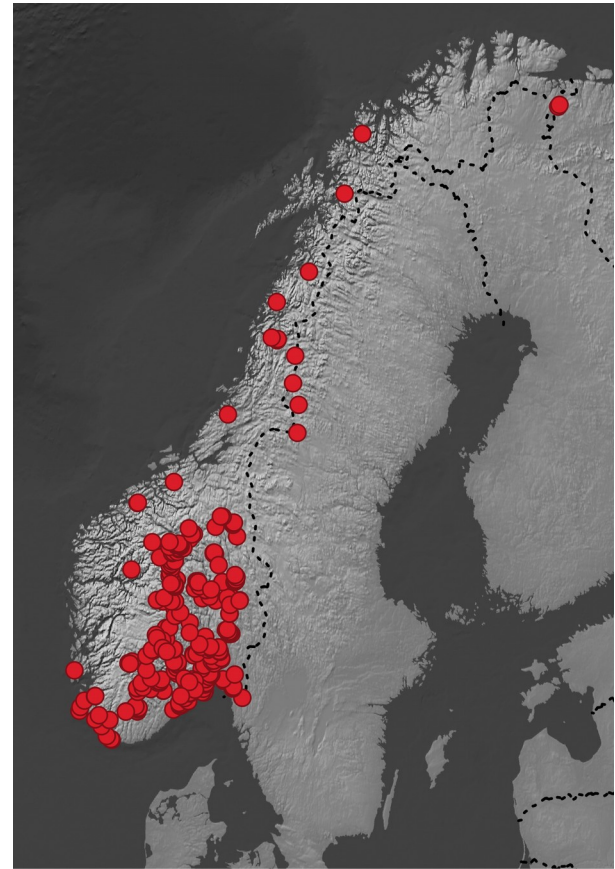
Christian Willmes, Nikolai Kirch, Espen Uleberg, Mieko Matsumoto, Helge Høeg

# Contents

- The Norwegian Pollen Dataset
- Data management practice in palynology
- Technical Workflow
- Preparation for Dataset Publication
- Re-analysis and further work

# The Norwegian Pollen Dataset in Numbers

- # Tilia files: **1039**
- # Analysed Cores: **357**
- # Core Locations: **231**
- # Cores with Age Model: **214**



# Have a look at last years CAA presentation

- Willmes, C., Uleberg, E., Matsumoto, M., Høeg, H. (2018): **Relating Norwegian palynological records and archaeological datasets**. CAA Conference 2018 Tübingen.

<https://crc806db.uni-koeln.de/dataset/show/relating-norwegian-palynological-records-and-archaeological-datasets1523871710>

The screenshot shows a web page for a dataset entry. The title is "Relating Norwegian palynological records and archaeological datasets". It is maintained by Christian Willmes and was created on 16.04.2018. The abstract describes a project to create a comprehensive dataset of pollen core data from Norway, used for quantitative analysis and GIS applications. The resources section includes a PDF file named "CAA-Tuebingen\_Willmes.pdf" accessed 50 times. The bibliography section lists the citation: Willmes, C., Uleberg, E., Matsumoto, M., Heeg, H. (2018): Relating Norwegian palynological records and archaeological datasets. CAA Conference 2018 Tübingen. A table below the bibliography lists the authors (Willmes, Christian and Uleberg, Espen and Matsumoto, Mieko and Heeg, Helge), type (presentation), title (Relating Norwegian palynological records and archaeological datasets), year (2018), and organization (CAA Conference 2018 Tübingen). The right sidebar contains information about the project group, linked authors (Christian Willmes, Espen Uleberg, Mieko Matsumoto, Helge Heeg), a presentation link, a license (CC BY), and open data status (1 star).

Start Data Publications Maps Sites Services About Imprint CRC806 Members Search resources Login or Register

## Relating Norwegian palynological records and archaeological datasets

Literature Maintained by Christian Willmes created on 16.04.2018

### Abstract

In this contribution we present the first results of a project that aims to create detailed GIS-based paleoenvironmental information from a comprehensive data set of pollen core data from Norway. This comprehensive dataset is created by the Norwegian paleobotanist Helge Heeg during more than 30 years. The pollen data comes from more than 300 drill cores and the main part of them are collected as part of archaeological excavation projects in South-Eastern Norway. Until now this data was published mostly in textual form as tables or as figures (pollen diagrams). The digital pollen count data itself has never been published and has therefore not been available for quantitative analysis before. The study will use quantitative statistics and software tools like R and GIS are used to create a comprehensive paleoenvironmental data set of high temporal and spatial resolution for the last ~10,000 years of South-East Norway. This new paleoenvironmental data set of high temporal and spatial resolution, will help to better analyze and understand the Holocene history and archeology of Norway, by relating to archaeological spatial datasets from the Museum of Cultural History at the University of Oslo. This pilot study has a large potential for more advanced paleoenvironmental studies in Norway and neighbouring countries (especially for the Stone Age), and also for technical/theoretical refinement of GIS applications to archaeological metadata.

### Resources

CAA-Tuebingen\_Willmes.pdf  
Accessed 50 times | Last updated 16.04.2018


### Bibliography

Willmes, C., Uleberg, E., Matsumoto, M., Heeg, H. (2018): Relating Norwegian palynological records and archaeological datasets. CAA Conference 2018 Tübingen

Authors	Willmes, Christian and Uleberg, Espen and Matsumoto, Mieko and Heeg, Helge
Type	presentation
Title	Relating Norwegian palynological records and archaeological datasets
Year	2018
Organization	CAA Conference 2018 Tübingen

Export BibTeX

### Project Group



### Linked CRC 806 Authors

Christian Willmes Espen Uleberg  
Mieko Matsumoto Helge Heeg

### presentation

Willmes, C., Uleberg, E., Matsumoto, M., Heeg, H. (2018): Relating Norwegian palynological records and archaeological datasets. CAA Conference 2018 Tübingen

Export BibTeX

### License

CC BY

- Attribution required
- For commercial use
- Derivations are allowed

### 1 ★ Open Data

You can look at the data  
You can change the data  
You can share the data with anyone

More Information

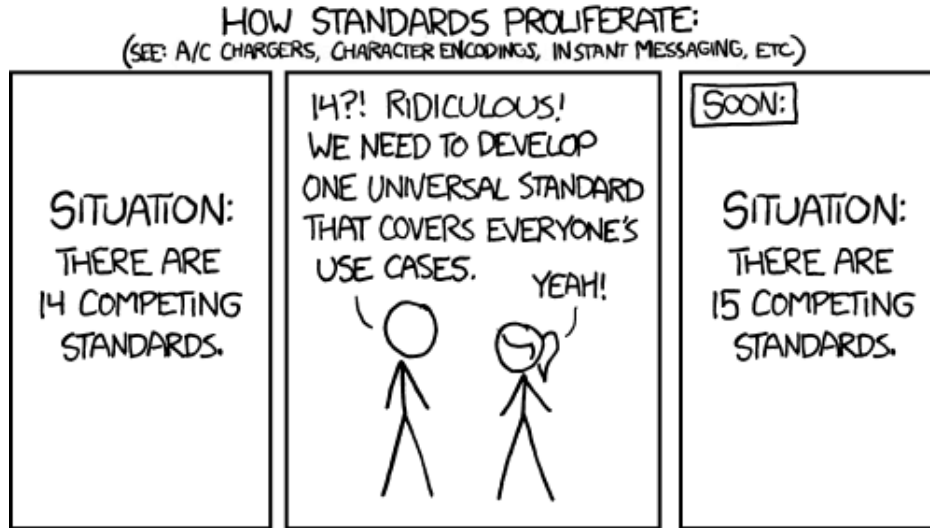


# Data management practice in palynology?

# Is there a ‚Standard‘ for storing Palynology data?



# Why Standard!?



<https://xkcd.com/927/>

## PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS **THE** CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13  
20130227 2013.02.27 27.02.13 27-02-13  
27.2.13 2013.II.27. 27½-13 2013.158904109  
MMXIII-II-XXVII MMXIII <sup>LVII</sup>/<sub>CCCLXV</sub> 1330300800  
((3+3)×(111+1)-1)×3/3-1/3<sup>3</sup> 2013  
10/11011/1101 02/27/20/13 0<sup>2</sup>1<sup>3</sup>2<sup>3</sup>3<sup>7</sup> 2-27-13  
5 67 8

<https://xkcd.com/1179/>

# Let's have a look at some pollen databases and repositories



# European Pollen Database



## EUROPEAN POLLEN DATABASE

[Home](#) | [EPD Board](#) | [Data Policy](#) | [Get Data](#) | [Contribute Data](#) | [Activities & News](#) | [Reports](#) | [Partners/Links](#)

The European Pollen Database (EPD) aims at storing and developing a relational database containing data and metadata of fossil and modern pollen records from natural archives (e.g. lacustrine sediments, peat bogs, marine sediments) collected on or near the Eurasian continent. The goal of the EPD is to develop an open platform to foster the scientific study of long-term palaeoecological records to address various themes such as biogeography, vegetation history, ecosystem conservation. In addition, the EPD community strives at making the knowledge gained to the wider public (schools, non-scientific public).

The EPD is a non-profit structure that is made available free of charge to the scientific community and is managed on a volunteer basis by the [EPD Board](#) to provide scientific guidance and oversight for the database. The activities are mainly carried by [open working groups](#), which play a significant role in developing and in sharing work to maintain the database. Should you be interested in taking part in this process, feel free to contact the working-groups coordinator.

Prior to using the database, you are urged to carefully read the [Data Policy of the EPD](#).

The EPD is managed by the [IMBE laboratory](#) (Mediterranean Institute of marine and terrestrial Biodiversity and Ecology) and the [Pytheas Institute](#) (Earth Sciences and Astronomy Observatory) in Aix-en-Provence, France. The EPD has been financially supported by the [Aix-Marseille University](#) since 2007.

[home](#) • [EPD Board](#) • [Data Policy](#) • [Get Data](#) • [Contribute Data](#) • [Activity & News](#) • [Reports](#) • [Partners/Links](#)  
Copyright © 2007-2019 The European Pollen Database, funded by Aix-Marseille University, Pytheas Institute, and IMBE



<http://www.europeanpollendatabase.net>

# European Pollen Database

From the EPD Manual....

History: „The need for a European pollen database was discussed during the closing session of the International Geological Correlation Programme (IGCP) 158B project in **Krakow**, Poland, in June 1988.“ (EPD Manual, p. 1)

A hint on existing data models: „*Tilia* is able to read files that are in any of the following formats: Brown, Cambridge (POLLDAT<sub>A</sub> output), Marseille, Minnesota, Wisconsin, and WK1 or WKS.“ (EPD Manual, p.9)

The EPD is based on a complex Entity Relationship Model (more about this later).

# Neotoma Paleocology DB



## NEOTOMA PALEOECOLOGY DATABASE

About Data Education Workgroups News Events Publications Documents People [Login](#)

Home



### What is Neotoma?

*Neotoma Paleocology Database and Community is an online hub for data, research, education, and discussion about paleoenvironments. Anyone with an Internet connection can access Neotoma.*



*Neotoma cinerea. Photo by Roger W. Barbour*

The primary philosophy behind Neotoma is data sharing so that users can easily:

- Discover: find information efficiently by searching the database on spatial, temporal, and metadata criteria
- Explore: interactively browse and visualize live data and metadata
- Share: get data and information in a variety of useful formats (e.g., downloads, reports, graphics)

### Recent Data Uploads

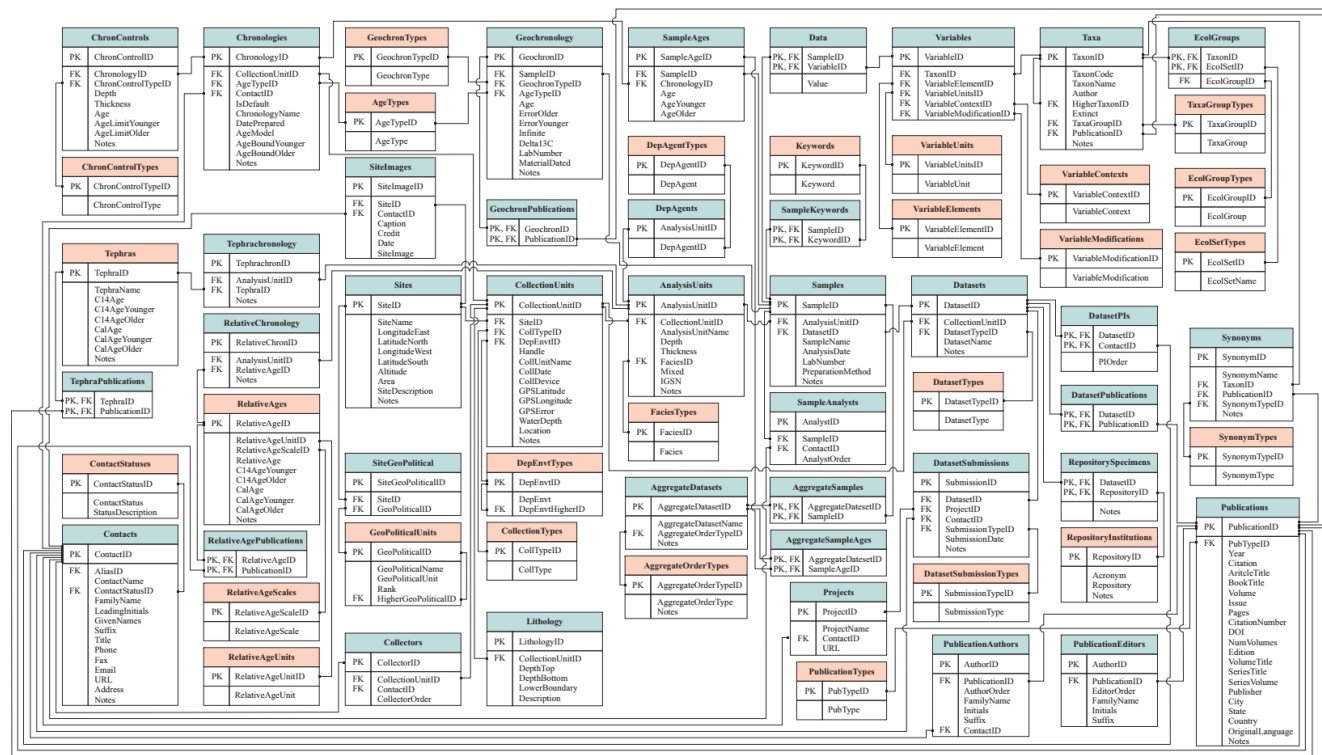
- Lake WA01**  
North American Pollen Database  
Investigator(s): R. Ravindra  
Location: Canada | Yukon Territory  
Uploaded: 2019-04-20
- Lake WA01**  
North American Pollen Database  
Investigator(s): R. Ravindra  
Location: Canada | Yukon Territory  
Uploaded: 2019-04-20
- Site 31, Georgian Bay, Canadian Hydrographic Service Database**  
Investigator(s): T.W. Anderson, J. Terasmae  
Location: Canada | Ontario  
Uploaded: 2019-04-19
- Site 23, Georgian Bay, Canadian Hydrographic Service Database**

### News Highlights

**Neotoma adopts official Code of Conduct for the database**

The Leadership Council has officially adopted a Code of Conduct... [Read more>>](#)

## NEOTOMA: ENTITY RELATIONSHIP DIAGRAM



Source:  
E. Grimm: Neotoma  
Database Structure.  
[www.neotomadb.org](http://www.neotomadb.org)



The screenshot shows the Pangaea Data Publisher website. At the top left is the Pangaea logo, a globe with the text 'PANGAEA Data Publisher for Earth & Environmental Science'. To the right, there is a navigation bar with 'SEARCH', 'SUBMIT', 'ABOUT', and 'CONTACT'. A 'Not logged in' status is visible in the top right corner. The main content area is divided into several sections:

- Submit Data:** A green button with a globe icon and the text 'Submit Data'.
- Welcome to PANGAEA® Data Publisher:** A section with the text: 'Our services are generally open for archiving, publishing, and re-usage of data. The World Data Center PANGAEA is member of the ICSU World Data System.'
- ALL TOPICS:** A search bar with the text 'Search for measurement type, author name, project, taxa...'. Below it is a grid of 12 topic categories, each with a representative image and a label: CHEMISTRY (14136), OCEANS (24260), LITHOSPHERE (40388), BIOLOGICAL CLASSIFICATION (30003), ATMOSPHERE (39032), PALEONTOLOGY (241120), ECOLOGY (17323), BIOSPHERE (17323), LAND SURFACE (24274), GEOPHYSICS (21376), CRYOSPHERE (17323), LAKES & RIVERS (226), HUMAN DIMENSIONS (226), FISHERIES (226), and AGRICULTURE (17376).
- Latest News:** A section with a date '2019-04-17' and the title 'FUNDING INFORMATION'. The text reads: 'PANGAEA data sets now contain information about funding awards as part of the project metadata. Please, check (click on project in the metadata)'. Below this is another news item dated '2019-03-07' titled '10 YEARS OF DATACITE', with the text: 'The year 2019 sees the 10 year anniversary of DataCite, Jan Brase, former DataCite Director, tells us the story how it all began.' There is a 'DataCite' logo and a link to 'Show all 29 news items...'. There is also a 'Featured Data' section with several entries, including 'Wörmer, L., Hoshino, T., Bowles, MW et al. (2018): Dipicolinic acid concentration in sediment samples', 'Nitz, I., Grosse, G., Jones, BM et al. (2018): Remote sensing quantifies widespread abundance of permafrost region disturbances across the Arctic and Subarctic Datasets', 'Braun, MM, Mall, P., Sommer, C et al. (2018): Annual glacier elevation change rate raster dataset, South American Andes 2000 and 2011-2015', 'Cornils, A., Schnack-Schiel, SB (2018): Abundance and distribution of planktonic Copepoda in the Southern Ocean and other regions from 1980 to 2005', and 'Costas, S., Ramirez, M., Mendes, MMP et al. (2017): Surficial sediment texture from the Iberian Atlantic Margin Sediments Database (ISAM-Sea)'. There is a link to 'Show all 34 featured data sets...'. There is also a 'MAP' button on the left side of the grid.

<https://pangaea.de/>



# Pangaea data submission guidelines

PANGAEA is an archive for any kind of data from earth system research and thus has **no special format requirements** for submissions. Data might be submitted in the authors format and will be converted to the final import and publication format by the PANGAEA editors. The data provider is kindly requested to keep the following points in mind to minimize the preparatory work prior to upload.

- For samples taken or measurements made somewhere on earth, the provision of position(s) is mandatory (latitude/longitude in decimal degree is preferred).
- If data are supplementary to a publication, the (preliminary) citation with journal title and abstract must be added.
- If data are related to a project (where PANGAEA is the designated archive) add the project acronym as label.
- Date/Time must be provided in ISO-format (e.g. 1954-04-07T13:34:11).
- Parameters are always accompanied by a unit.
- Abbreviations should be explained.
- Extended documentations may be added as plain text or pdf-file.
- Submit data tables as excel or tab-delimited text files; specific formats (e.g. shape, netCDF, segy ...) may be added in zip-archive.

# Pangaea data submission guidelines

## Additional recommendations:

- Preferred format for data tables is TAB-delimited TEXT-files (ASCII), submitted as ZIP-archive, or excel-format.
- Several tables with different format should be provided on different sheets.
- Parameter name with unit must appear in the header line (or PANGAEA parameter ID).
- Use proper event/site/sample labels, e.g. as defined during an expedition (if appropriate).
- Provide references by its DOI or (even better) as pdf; documentations should also be provided as pdf (documents will be stored in ePIC and linked via a handle).
- Numeric parameter columns must contain numbers only; exception see quality flags.
- Text parameter columns can be filled to a maximum of 255 characters per field.
- If the result of a scientific analysis is zero, the corresponding field in the data table must be filled with 0 (and not left empty).
- Fields without data should be left empty (and NOT filled with '-', 'n/a', -9999 or '\*' etc).
- Remove empty lines and columns; those will not be imported.
- Avoid abbreviations.
- Avoid redundant information.
- **Use standards as far as available**
- Care for proper geocodes
- Use the decimal point.
- Use the English language

# NOAA WDC Paleo

**NOAA** NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION  
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Formerly the National Climatic Data Center (NCDC)... [more about NCEI](#) »

Home Climate Information Data Access Customer Support Contact About

Home > Data Access > Paleoclimatology > Paleoclimatology Datasets













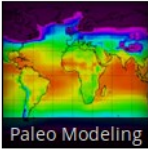





**Quick Links**

- Land-Based Station
- Satellite
- Radar
- Model
- Weather Balloon
- Marine / Ocean
- Paleoclimatology

**Datasets**

- Borehole
- Climate Forcing
- Climate Reconstruction
- Coral and Sclerosponge

## Paleoclimatology Datasets

 Borehole	 Cave	 Coral	 Fauna	 Fire History	 Forcing
 Historical	 Ice Cores	 Insect	 Lake	 Lake Levels	 Loess
 Paleo Modeling	 Paleocean	 Plant Macros	 Pollen	 Reconstructions	 Tree-ring

<https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets>



# NOAA Paleoclimatology

- No specific instructions for pollen data
- Tree species codes
- <https://www1.ncdc.noaa.gov/pub/data/paleo/templates/tree-species-code.txt>
- Stems from „Tree Ring“ Data



NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION  
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION



Formerly the National Climatic Data Center (NCDC)... [more about NCEI](#) »

Home Climate Information Data Access Customer Support Contact About

Search



Home > Data Access > Paleoclimatology > Contributing Data

Quick Links

Land-Based Station

Satellite

Radar

Model

Weather Balloon

Marine / Ocean

Paleoclimatology

Datasets

Search

Products

Perspectives

Contributing Data

PaST Thesaurus

Education and Outreach

About the Program

Severe Weather

Blended & Global

## Contributing Data



500 Frequent Contributors

You can contribute your data to the World Data Service for Paleoclimatology (WDS Paleo), and NOAA will permanently archive it. Contributing your data will allow you to satisfy obligations to publishers and funding agencies, make your data freely and permanently available without restrictions, have your data discovered and used by colleagues and others, and may increase the citation rate of your associated articles.

### Instructions for Contributing Data

Data contributions should consist of metadata and one or more data tables including a chronology table, if applicable, as well as raw data and reconstructions. Contribute data and metadata using one of the contribution templates listed below.

**Data Synthesis Projects:** WDS-Paleo requests that leaders of synthesis projects have their designated liaison contact WDS-Paleo staff ([paleo@noaa.gov](mailto:paleo@noaa.gov)) early during research and development to speak about topics including appropriate data formats and the use of standard controlled vocabularies for defining terms. There is a new Synthesis template available below to allow contributors to provide us the metadata and data from the records used in a synthesis project. The final product data from a synthesis project can be contributed using our standard contribution template.

<https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/contributing>



# It all comes down to a software called Tilia



HOME STORE DOWNLOAD CONTACT FAQ MAILING LIST HELP

## TiliaIT

### Tilia

Tilia is software designed for managing and graphing paleontological data and metadata, especially stratigraphic data. The program is especially suitable for graphing microfossil data such as pollen, ostracodes, and diatoms, although it may be used for graphing virtually any kind of stratigraphic data.

Tilia runs on the Microsoft® [Windows®](#) platform, version XP or later, either 32-bit or 64-bit.

### Features

- Spreadsheet for entering data, samples and variables (e.g. pollen taxa, loss-on-ignition, geochemical variables, charcoal).
- Download data directly from the [Neotoma Paleocology Database](#).
- Automated calculation of sums for calculation of percentages.
- CONISS cluster analysis, both stratigraphically constrained and unconstrained.

### Versions

All versions except the Free version have a Product Key, which unlocks functionality not available in the Free version.

- **Free.** Includes spreadsheet, all spreadsheet calculations, all metadata forms, and all functionality with the Neotoma Paleocology Database. Does not include graphics.
- **Professional.** Includes all functionality, including graphics. License to an individual user only. Price: €100.

<https://www.tiliait.com/>



Windows window titled "Tilia" showing a spreadsheet of pollen data. The spreadsheet has columns for pollen types (A-Y) and rows for samples (1-29). The data represents pollen counts for various species across different samples.

	A	B	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Pollen			20	40	60	80	100	120	140	160	180	200	220	240	260	280	300	320	340	360
2	Code	Name	Group																		
3	#Chron1	Sample age		0	178,4615	356,9231	535,3846	713,8461	892,3077	1070,769	1235,484	1386,452	1537,419	1688,387	1839,355	1990,323	2141,29	2292,258	2688,75	3167,083	3645,417
4	#Chron2	Sample age		0	173,0769	346,1538	519,2308	692,3077	865,3846	1038,462	1208,871	1376,613	1544,355	1712,097	1879,839	2047,581	2215,323	2383,064	2879,375	3485,208	4091,042
5	aa	Betula, bjerk	A	39	37	23	11	26	36	40	124	87	132	96	100	87	118	101	93	115	72
6	ab	Pinus, funu	A	165	200	207	161	135	214	192	103	116	65	110	132	111	117	102	101	120	90
7	ac	Corylus, hassael	A	6	5	7	1	9	3	11	47	52	55	41	31	45	54	31	39	35	49
8	ad	Alnus, or	A	7	16	27	4	8	11	14	16	43	78	56	51	77	90	76	74	85	90
9	ae	Picea, gran	A	91	108	138	179	161	70	90	42	23	20	11	6	6	3	1	3	0	1
10	af	Fagus, bek	A	8	2	2	0	8	7	2	3	1	3	1	1	4	0	1	0	0	1
11	ag	Carpinus, agnbak	A	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
12	ah	Populus, osp	A	0	0	0	0	0	0	1	0	0	0	2	3	1	3	1	1	2	3
13	ai	Ulmus, alm	A	1	0	0	0	0	1	0	1	1	7	4	4	1	4	6	4	3	1
14	aj	Tilia, lind	A	0	0	1	0	1	0	0	1	2	4	5	4	2	4	4	4	0	4
15	ak	Quercus, eik	A	6	4	1	2	2	5	5	14	12	36	20	23	17	28	20	16	16	21
16	al	Fraxinus, ask	A	0	0	0	0	0	0	0	0	0	1	2	0	0	1	1	2	1	2
17	ba	Salix, selje, vier	B	0	0	0	0	0	0	0	1	0	1	1	2	1	0	0	1	1	1
18	bb	Juniperus, ener	B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	bc	Viburnum, krossved	B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	bd	Viscum, misteltein	B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	be	Hedera, efay	B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	ca	Calluna, ressling	C	92	54	32	45	21	27	10	17	10	25	12	17	9	2	6	7	1	1
23	cb	Pempetrum, krekling	C	3	5	1	9	2	0	0	2	0	1	1	0	0	0	0	0	0	0
24	cc	Ericales, andre lyngarter	C	9	6	1	11	5	1	1	1	1	0	2	1	0	1	0	0	1	0
25	cd	Myrica, pors	C	27	17	16	7	9	10	24	18	24	16	6	6	0	1	0	0	0	0
26	da	Cyperaceae, halvgressfamilien	D	6	3	7	5	4	3	1	122	18	1	8	17	4	12	18	18	15	2
27	db	Poaceae, gressfamilien	D	19	13	13	1	4	2	2	37	7	22	20	17	11	14	20	23	26	10
28	dc	Cladium, storak	D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	ea	Cerealia, dyrket kom	E	0	5	1	2	2	1	1	0	3	1	1	2	1	0	1	0	0	0



## Asking Dr. Eric Grimm himself

Q: I am desperately searching for a list (or better a table) of plant taxonomy for my pollen data.

Is there any kind of schema contributors use for their data at NeotomaDB?

A: [...] Neotoma has a **preferred taxonomy for pollen taxa** based on modern taxonomic concepts. The easiest way to access it is through the Tilia software [www.tiliait.com](http://www.tiliait.com). After entering your taxa names in the spreadsheet, you validate against the Neotoma Taxa table with the menu entries Neotoma > Variables > Validate Spreadsheet Taxa. Synonymies will be suggested. Any names that don't match will be listed, and you can search for alternatives with wildcard searches – **misspellings are very common**. You can actually do this with the free version of Tilia, or if you're a student you can apply for a free student license for full capability.

(... to be continued on next slide ->)

# Asking Dr. Eric Grimm himself

[...] Otherwise, you can access the entire Taxa list here:  
<https://tilia.neotomadb.org/retrieve/doc2/>.

A big list of services will be listed. Click on GetTaxaByTaxaGroupID. In the page that appears enter VPL for the TAXAGROUPID (VPL = vascular plants). A table with all the vascular plant taxa will appear. You can copy and paste this to Excel.

Tada! This seems to be the current **defacto standard**.

# The Neotoma Taxa Table

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Valid	Notes	Author	TaxonID	Extinct	TaxonName	TaxonCode	ValidatorID	TaxaGroupID	ValidateDate	PublicationID	HigherTaxonID	
2	TRUE		Miller, 1754	1	FALSE	Abies	Abi	44	VPL	2017-04-21	299	329	
3	TRUE		Linnaeus, 1753	2	FALSE	Acalypha	Acy	44	VPL	2012-03-21	300	694	
4	TRUE		Marshall, 1785	3	FALSE	Acer saccharum-type	Ace.sa-t	44	VPL	2012-03-21	300	338	
5	TRUE		Linnaeus, 1753	4	FALSE	Acer negundo	Ace.ne	44	VPL	2016-03-13	300	338	
6	TRUE		Linnaeus, 1753	5	FALSE	Acer pensylvanicum	Ace.pe	44	VPL	2012-03-21	300	338	
7	TRUE		Linnaeus, 1753	6	FALSE	Acer rubrum	Ace.ru	44	VPL	2016-03-13	300	338	
8	TRUE		Linnaeus, 1753	7	FALSE	Acer saccharinum-type	Ace.sc-t	44	VPL	2012-03-21	300	338	
9	TRUE		Lamarck, 1786	8	FALSE	Acer spicatum	Ace.sm	44	VPL	2012-03-21	300	338	
10	TRUE		Linnaeus, 1753	9	FALSE	Acer undiff.	Ace.ud	44	VPL	2012-03-21	300	338	
11	TRUE		Linnaeus, 1753	10	FALSE	Adiantum	Adi	44	VPL	2017-06-11	9777	1625	
12	TRUE		Linnaeus, 1753	11	FALSE	Aesculus	Aes	44	VPL	2017-04-18	300	3555	
13	TRUE		Linnaeus, 1753	12	FALSE	Allium	All	44	VPL	2017-04-23	946	5221	
14	TRUE		(Linnaeus) Moench, 1794	13	FALSE	Alnus incana-type	Aln.in-t	44	VPL	2014-06-03	613	330	
15	TRUE		(Chaix) de Candolle, 1805	14	FALSE	Alnus viridis-type	Aln.vi-t	44	VPL	2015-04-12	613	24464	
16	TRUE		Miller, 1754	15	FALSE	Alnus undiff.	Aln.ud	44	VPL	2014-06-03	613	330	
17	TRUE		Linnaeus	16	FALSE	Ambrosia-type	Amb-t	44	VPL	2012-03-21	3717	3668	
18	TRUE		Linnaeus, 1753	17	FALSE	Amorpha-type	Amo-t	44	VPL	2014-02-09	300	9690	
19	TRUE		R. Brown, 1818	18	FALSE	Anacardiaceae undiff.	Anaeae.ud	44	VPL	2014-02-09	1490	347	
20	TRUE		Linnaeus, 1753	19	FALSE	Apocynum	Apo	44	VPL	2017-06-07	300	30410	
21	TRUE		Britton & Brown, 1896 [P.D.]	20	FALSE	Spermatophyta undiff. (aquatics)	Speyta.ud(aqu)	44	VPL	2017-06-25	3711	5481	
22	TRUE		Berchtold & J. Presl, 1825	21	FALSE	Aquifoliaceae	Aqueae	44	VPL	2014-12-07	1490	9101	
23	TRUE		Jussieu, 1789	22	FALSE	Araliaceae	Araeae	44	VPL	2018-11-06	1490	9099	
24	TRUE		M. Bieberstein, 1819	23	FALSE	Arceuthobium	Arc	44	VPL	2015-03-16	300	24094	
25	TRUE		Willdenow, 1809	24	FALSE	Armeria	Arm	44	VPL	2017-07-27	3702	1298	
26	TRUE		Linnaeus, 1753	25	FALSE	Artemisia	Art	44	VPL	2017-02-07	1491	22590	
27	TRUE		Roth, 1800	26	FALSE	Athyrium	Ath	44	VPL	2017-04-11	9777	29342	
28	TRUE		Lamarck, 1783	27	FALSE	Azolla	Azo	44	VPL	2017-10-23	9777	9547	
29	TRUE		de Candolle	28	FALSE	Berlandiera-type	Ber-t	44	VPL	2012-03-21	314	418	
30	TRUE		Linnaeus, 1753	29	FALSE	Betula	Bet	44	VPL	2017-09-13	10034	32627	
31	TRUE		Linnaeus	31	FALSE	Bidens-type	Bid-t	44	VPL	2012-03-21	300	418	
32	TRUE		Jussieu, 1789	32	FALSE	Boraginaceae	Boreae	44	VPL	2017-10-15	10130	33447	
33	TRUE		Swartz, 1800[1801]	33	FALSE	Botrychium	Bot	44	VPL	2017-03-25	299	1207	
34	TRUE		J.F. Gmelin, 1791	34	FALSE	Brasenia schreberi	Brs.sc	44	VPL	2015-01-10	613	9506	
35	TRUE		Swartz, 1788	36	FALSE	Bumelia	Bum	44	VPL	2017-03-28	300	1847	
36	TRUE		Jacquin ex Linnaeus, 1762	37	FALSE	Bursera	Bur	44	VPL	2013-08-01	3721	2036	
37	TRUE		Aublet, 1775	38	FALSE	Cabomba	Cab	44	VPL	2015-01-10	613	3123	
38	TRUE		Linnaeus, 1753	39	FALSE	Callitriche	Cll	44	VPL	2012-03-21	300	698	
39	TRUE		Linnaeus, 1753	40	FALSE	Caltha	Cal	44	VPL	2016-03-13	613	350	
40	TRUE		Jussieu, 1789	41	FALSE	Campanulaceae undiff.	Cameae.ud	44	VPL	2012-03-21	1490	351	

# Preparation of the Dataset for publication in a data journal

**Earth System Science Data**  
The data publishing journal

119 | Imprint | Contact | Data protection

Submit a manuscript | Manuscript tracking

**Earth System Science Data**  
OPEN ACCESS | INTERACTIVE PUBLIC PEER REVIEW | ARTICLE LEVEL METRICS

Chief editors: Thomas Blunier, David Carlson, Kirsten Elger, Jens Klump, Yasuhiro Murayama, Ge Peng & Hans Pfeiffenberger

Earth System Science Data (ESSD) is an international, interdisciplinary journal for the publication of articles on original research data (sets), furthering the reuse of high-quality data of benefit to Earth system sciences. The editors encourage submissions on original data or data collections which are of sufficient quality and have potential to contribute to these aims. The journal maintains sections for regular-length articles, brief communications (e.g. on additions to data sets) and commentaries, as well as review articles and special issues. This short commentary draws on ESSD author, reviewer and editor experiences over its first 10 years of operation to define guidelines, requirements and benefits of the ESSD processes.

**Search articles**  
Search:   
Author:

**Search web pages**  
Search:

**News**

18 Apr 2019  
**Researchers examine historic levels of greenhouse gases in the atmosphere**  
A study has examined a 2000-year history of greenhouse gases in the world's atmosphere, providing key data for taking action to tackle climate change. The article, led by Dr. Mauro Rubino, a lecturer in Environmental Science at Keele University, has been published in ESSD. ▶

18 Dec 2018  
**ESSD editorial**  
Guidelines, requirements and benefits of the ESSD processes are defined in an editorial recently published in ESSD. ▶

05 Dec 2018  
**Global Carbon Budget 2018**  
Detailed data of the Global Carbon Budget 2018 were made available in Earth System Science Data (ESSD) today. ▶

**News archive** ▶

**Recent papers**

17 Apr 2019  
**Global whole-rock geochemical database compilation**  
Matthew Gard, Derrick Hasterok, and Jacqueline Halpin  
Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2019-50>, 2019  
Manuscript under review for ESSD (discussion: open, 0 comments)

**Journal metrics**

Clarivate analytics	IF 8.792
Clarivate analytics	IF 5-year 8.414
Scopus	CiteScore 8.18
CWTS	SNIP 2.620

**Follow @ESSD\_journal**

**Earth System Science Data**

**Earth System Science Data**

**Earth System Science Data**



# Preparation of the Dataset for publication in a data journal

## First steps

Before the submission of your manuscript to the Editorial Support for peer review, you are kindly requested to do the following:

- to submit the data set(s) referenced in the manuscript to a repository that meets the preconditioned [repository criteria](#),
- to read the [general terms](#) and the [article processing charges](#) for this journal;
- to read the [licence and copyright](#);
- to read the [manuscript preparation](#) for this journal;
- to agree and comply with the [general obligations for authors](#);
- to register the manuscript in order to receive a link to upload the manuscript files into the Copernicus Office Editor. The registration also defines the manuscript subject areas and index terms as the basis for the [editor assignment](#).

## File submission for review process

After the manuscript registration, you are kindly asked to upload those files which are necessary for the peer-review process. The following files are required:

- the **abstract** as plain text to be pasted into the upload form where requested;
- the complete **manuscript** (title, authors, affiliations, abstract, text, tables, figures) as a \*.pdf file. Data sets, model code, video supplements, or other assets to your manuscript should be submitted to a reliable repository receiving a DOI, cited in your manuscript, and included in your reference list. Reviewers can then access these relevant sources.

Other possible review files include the following:

- If you have **supplementary material** to your manuscript which does not meet the above-mentioned criteria to be hosted by a reliable repository, you can submit your supplement as a \*.zip archive or single \*.pdf file. The overall file size of such a supplement is limited to 50 MB. Larger supplements have to be submitted to a reliable data repository in any case receiving a DOI, cited in your manuscript, and included in your reference list.
- The **author's response** (also final author comment in the public discussion) in case of "minor" or "major" revisions must be submitted as one separate \*.pdf file (indicating page and line numbers), structured in a clear and easy-to-follow sequence: (1) comments from referees/public, (2) author's response, and (3) author's changes in manuscript. Regarding author's changes, a marked-up manuscript version (track changes in Word, [latexdiff](#) in LaTeX) converted into \*.pdf and combined with the author's response should be provided.
- **Former manuscript versions** in case the present submission is an update of an already published paper must be indicated. Please prepare a marked-up manuscript version (track changes in Word, [latexdiff](#) in LaTeX) showing the differences between the former and present version, converted into \*.pdf and uploaded as a supplement.

[https://www.earth-system-science-data.net/for\\_authors/submit\\_your\\_manuscript.html](https://www.earth-system-science-data.net/for_authors/submit_your_manuscript.html)



# Preparation of the Data set as a Data publication

The screenshot displays the MDPI Data journal homepage. At the top, navigation links include MDPI, Journals A-Z, Information & Guidelines, Initiatives, and About. On the right, there are links for Login, Register, and Submit. A search bar is located below the navigation, with fields for Title / Keyword, Author / Affiliation, Journal (set to Data), and Article Type (set to all). A Search button is also present.

The left sidebar contains a Journal Menu with links to Data Home, Aims & Scope, Editorial Board, Instructions for Authors, Guidelines for Reviewers, Special Issues, Article Processing Charge, Indexing & Archiving, Most Cited & Viewed, Journal History, and Editorial Office. Below the menu are social media icons for Twitter and Facebook.

The main content area is titled "Data — Open Access Journal" and includes the journal's ISSN (2306-5729) and a description of its focus on data in science. It lists key features: Open Access (free for readers), High visibility (indexed in ESCI and IET), Rapid publication (16.5 days to decision), and Recognition of reviewers (discount on APC). A diagram illustrates the cycle: DEPOSIT DATA leads to PUBLISH ARTICLE, which leads to PUBLISH DATA DESCRIPTOR, which then leads back to DEPOSIT DATA.

On the right side, there are buttons for "Submit to Data" and "Review for Data", the MDPI logo, a QR code, and information about the journal's focus on "Data Sensing and Analysis in Design, Construction, Operation, Monitoring, and Maintenance of Built Environments". Guest Editors listed are Dr. Changbum R. Ahn, Dr. Eric Jing Du, and Dr. Youngjib Ham.



# Preparation of the Data set as a Data publication

## Data Descriptor Information

For a Data Descriptor, the following additional information is required (see the [template](#)).

- **Dataset:** DOI number or link to the deposited dataset in cases where the dataset held at a third party repository. If the dataset is submitted and will be published as a supplement, this field will be filled by the editors of the journal. In this case, the data should be uploaded at the time of submission.
- **Dataset License:** License under which the dataset is made available (CC0 is recommended, but other licenses are permissible).

Data Descriptors comprise the following sections:

- **Summary:** A short summary of the dataset, methods, background information on why and how the dataset was collected, short description of funded or unfunded research projects that are or will eventually be based on the dataset, and list of publications based on the dataset that were possibly already published. Optionally, authors may wish to describe potential benefits of publicly releasing and describing the dataset. In general, the Summary section is similar to an introduction section in a research article.
- **Data Description:** What data is contained? Which format? How can it be read and interpreted? For example, in tabular data give a full description of each column heading.
- **Methods:** Main methods applied to collect and treat, as well as to use and reuse the data. Notes on validation and curation techniques applied. Notes on data quality, noise, etc.
- **Usage notes (optional):** Further notes on the usage of the dataset that will help other researchers to access and further understand practical aspects of working with the data. If there are ethical or compelling commercial reasons that the data cannot be made available, either in part or in full, these should be described in as much detail as possible. You should make clear how the data can be accessed and if there are circumstances in which access would be denied (e.g. if complying with the request would compromise anonymity of human participants or if an embargo applies); we recommend full access wherever possible.

<https://www.mdpi.com/journal/data/instructions>

Geographisches Institut | Universität zu Köln | Dr. Christian Willmes | 23.04.2019



# Preparation of the Data set as a Data publication

- A meta Table:
  - Filename
  - Sitename
  - Location (Lat, Lon)
  - Samples #
  - Depth (cm)
  - Age model (oldest sample, years BP)
  - If published → reference and citation

# Preparation of the Data set as a Data publication

- Data files (for each core)
  - Depth
  - Sample Age
  - Taxon (according to Neotoma Taxon taxonomy)
  - Absolute count

# Table structure (example)

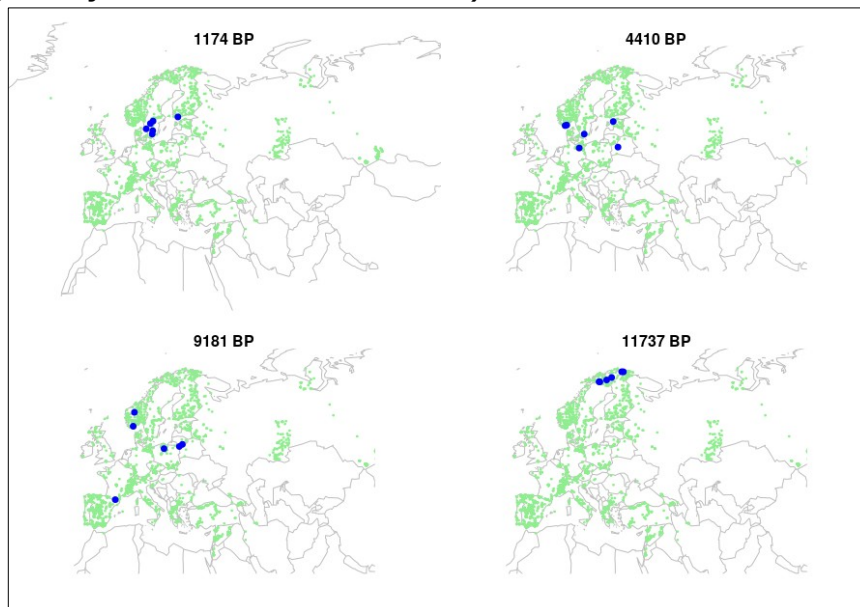
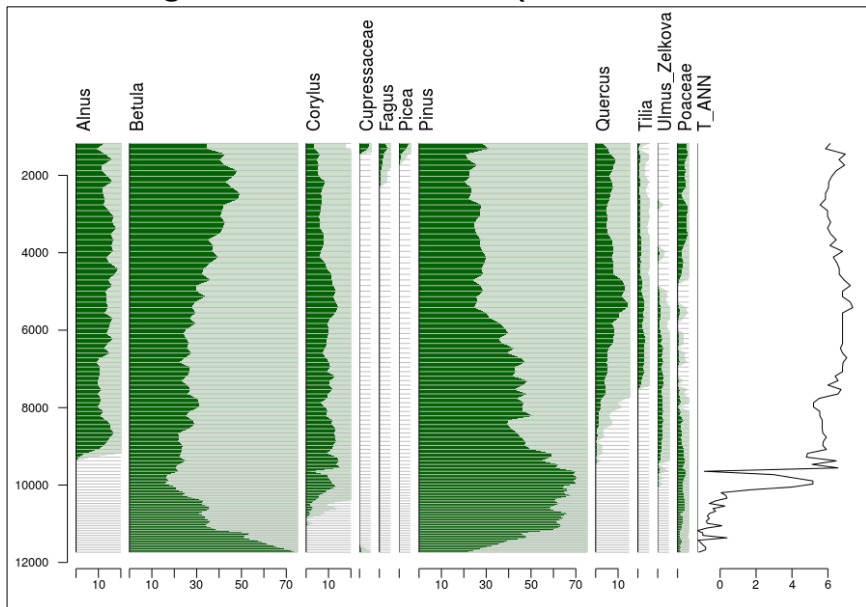
Depth	0	5	10	15	20
Sample Age	0	125	230	350	490
Taxon 1	0	20	25	40	48
Taxon 2	23	60	51	55	72
Taxon ...					
Taxon n	211	33	54	148	124

# Preparation of the Data set as a Data publication

- File format of tables: UTF-8, .csv
- All files will be bundled in a .zip Archive

# Re-analysis and preparation of GIS paleoenvironmental records

- Function MAT takes a training dataset of biological data (species abundances)  $y$  and a single associated environmental variable  $x$ , and generates a model of closest analogues, or matches. (Cited from “Package rioja” R documentation).



# Have a look at last years CAA presentation

- Willmes, C., Uleberg, E., Matsumoto, M., Høeg, H. (2018): **Relating Norwegian palynological records and archaeological datasets**. CAA Conference 2018 Tübingen.

<https://crc806db.uni-koeln.de/dataset/show/relating-norwegian-palynological-records-and-archaeological-datasets1523871710>

The screenshot shows a web page for a dataset entry. The title is "Relating Norwegian palynological records and archaeological datasets". It is maintained by Christian Willmes and was created on 16.04.2018. The abstract describes a project to create a comprehensive dataset of pollen core data from Norway, used for quantitative analysis and GIS applications. The resources section includes a PDF file named "CAA-Tuebingen\_Willmes.pdf" accessed 50 times. The bibliography section lists the citation: Willmes, C., Uleberg, E., Matsumoto, M., Heeg, H. (2018): Relating Norwegian palynological records and archaeological datasets. CAA Conference 2018 Tübingen. A table below the bibliography provides details: Authors (Willmes, Christian and Uleberg, Espen and Matsumoto, Mieko and Heeg, Helge), Type (presentation), Title (Relating Norwegian palynological records and archaeological datasets), Year (2018), and Organization (CAA Conference 2018 Tübingen). The right sidebar contains information about the Project Group, Linked CRC 806 Authors (Christian Willmes, Espen Uleberg, Mieko Matsumoto, Helge Heeg), a presentation link, a license (CC BY), and 1 Open Data item.

Start Data Publications Maps Sites Services About Imprint CRC806 Members Search resources Login or Register

## Relating Norwegian palynological records and archaeological datasets

Literature Maintained by Christian Willmes created on 16.04.2018

### Abstract

In this contribution we present the first results of a project that aims to create detailed GIS-based paleoenvironmental information from a comprehensive data set of pollen core data from Norway. This comprehensive dataset is created by the Norwegian paleobotanist Helge Heeg during more than 30 years. The pollen data comes from more than 300 drill cores and the main part of them are collected as part of archaeological excavation projects in South-Eastern Norway. Until now this data was published mostly in textual form as tables or as figures (pollen diagrams). The digital pollen count data itself has never been published and has therefore not been available for quantitative analysis before. The study will use quantitative statistics and software tools like R and GIS are used to create a comprehensive paleoenvironmental data set of high temporal and spatial resolution for the last ~10,000 years of South-East Norway. This new paleoenvironmental data set of high temporal and spatial resolution, will help to better analyze and understand the Holocene history and archeology of Norway, by relating to archaeological spatial datasets from the Museum of Cultural History at the University of Oslo. This pilot study has a large potential for more advanced paleoenvironmental studies in Norway and neighbouring countries (especially for the Stone Age), and also for technical/theoretical refinement of GIS applications to archaeological metadata.

### Resources

CAA-Tuebingen\_Willmes.pdf  
Accessed 50 times | Last updated 16.04.2018

### Bibliography

Willmes, C., Uleberg, E., Matsumoto, M., Heeg, H. (2018): Relating Norwegian palynological records and archaeological datasets. CAA Conference 2018 Tübingen

Authors	Willmes, Christian and Uleberg, Espen and Matsumoto, Mieko and Heeg, Helge
Type	presentation
Title	Relating Norwegian palynological records and archaeological datasets
Year	2018
Organization	CAA Conference 2018 Tübingen

Export BibTeX

### Project Group

### Linked CRC 806 Authors

Christian Willmes, Espen Uleberg, Mieko Matsumoto, Helge Heeg

### presentation

Willmes, C., Uleberg, E., Matsumoto, M., Heeg, H. (2018): Relating Norwegian palynological records and archaeological datasets. CAA Conference 2018 Tübingen

Export BibTeX

### License

CC BY

- Attribution required
- For commercial use
- Derivations are allowed

### 1 ★ Open Data

You can look at the data  
You can change the data  
You can share the data with anyone

More Information





# Thank you!

Contact:

Dr. Christian Willmes

[c.willmes@uni-koeln.de](mailto:c.willmes@uni-koeln.de)

Twitter: @cwillmes