DATABASE FOR REGISTERED GREENLAND BENTHIC FAUNA SPECIES

Technical Report from DCE – Danish Centre for Environment and Energy No. 166 2020

AARHUS
UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

[Blank page]

DATABASE FOR REGISTERED GREENLAND BENTHIC FAUNA SPECIES

Technical Report from DCE – Danish Centre for Environment and Energy No. 166 2020

Susse Wegeberg
Jonas Koefoed Rømer
Helle Buur Pedersen
Steffen Lundsteen†

Aarhus University, Department of Bioscience

AARHUS
UNIVERSITY
DCE – DANISH CENTRE FOR ENVIRONMENT AND ENERGY

Data sheet

Series title and no.:	Technical Report from DCE – Danish Centre for Environment and Energy No. 166
Title:	Database for registered Greenland benthic fauna
Authors:	Susse Wegeberg, Jonas Koefoed Rømer, Helle Buur Pedersen & Steffen Lundsteen[†]
Institution:	Aarhus University, Institute for Bioscience
Publisher:	Aarhus University, DCE – Danish Centre for Environment and Energy ©
URL:	http://dce.au.dk/en
Year of publication:	March 2020
Editing completed:	March 2020
Referees:	Martin Blicher, Greenland Institute of Natural Resources
Quality assurance, DCE:	Kirsten Bang
Financial support:	This study is part of the Northeast Greenland Environmental Study Program. The Northeast Greenland Environmental Study Program is a collaboration between DCE – Danish Centre for Environment and Energy at Aarhus University, the Greenland Institute of Natural Resources, and the Environmental Agency for Mineral Resource Activities of the Government of Greenland. Oil companies operating in Greenland are obliged to contribute to knowledge regarding environmental matters. The Strategic Environmental Impact Assessment and the background study program is funded under these commitments administered by the Mineral Licence and Safety Authority and the Environmental Agency for Mineral Resource Activities
Please cite as:	Wegeberg, S., Rømer, J.K., Pedersen, H.B. & Lundsteen, S.[†] 2020. Database for registered Greenland benthic fauna. Aarhus University, DCE – Danish Centre for Environment and Energy, 18 pp. Technical Report No. 166 http://dce2.au.dk/pub/TR166.pdf

Abstract: Identification of benthic fauna sampled in the Greenlandic seas still poses quite a challenge, even for skilled taxonomists. The primary result of the present project is thus a database with internal data and knowledge on Greenland benthos from Aarhus University (Bioscience) gathered in a standardised form, which can be used as a checklist. The database contains species identifications from the surveys, which totals 4998 registrations of 483 different species, and 91 photos of 30 different species. Access to the database will go through a user interface in a homepage on the internet, and which will be open with a requested logon.

Keywords:	Benthic fauna species, taxonomy, registrations, checklist, Greenland, database
Layout:	Graphic Group, AU Silkeborg
Front page photo:	Helle Buur Pedersen
ISBN:	978-87-7156- 478-5
ISSN (electronic):	2245-0203
Number of pages:	18
Internet version:	The report is available in electronic format (pdf) at http://dce2.au.dk/pub/TR166.pdf

Contents

1 Summary **5**

2 Sammenfatning **6**

3 Background **7**

3.1 Aims 7

4 Material and methods **8**

4.1 Samples and sampling 8

4.2 Database software and design 8

5 Database content **13**

6 Guideline to the Arctic benthos database **14**

Appendix **16**

[Blank page]

1 Summary

Identification of benthic fauna sampled in the Greenlandic seas still poses quite a challenge, even for skilled taxonomists. The sampling methods and processing most often follows internationally established guidelines, e.g. OSPAR. However, the detailed identification of sampled benthos to highest taxonomic level (genus, species) may be difficult due to limited availability of identification literature and also because many of the sampled species simply may be little known, or even unknown, to science.

Through a decade of work with Arctic benthos sampling and samples from surveys around Greenland in connection with oil exploration activities, the laboratories at Aarhus University (Bioscience) have gathered a comprehensive knowledge; data, photos and reference library, which is maintained on various forms internally.

As part of the Northeast Greenland Environmental Study Programme, this project aims to maintain and make this knowledge more public available as an aid to support Greenland benthos identification.

The primary result of the present project is thus a database with internal data and knowledge on Greenland benthos from Aarhus University (Bioscience) gathered in a standardised form.

The database contains species identifications from the surveys, which, after quality control, totals:

- 4998 registrations of 483 different species.
- 91 photos of 30 different species.

Access to the database will go through a user interface in a homepage on the internet, and which will be open with a requested logon. At the homepage, you can:

- Find species registered in samples from benthos surveys in Greenland in connection with oil exploration activities through species groups and lists
- Search information on these species using genus, species name or epithet
- Find information about a species; photos, author, references as well as geographic presence presented on a map and depth distribution.

The database is not complete for Greenland benthos. It will only contain information regarding species registered in connection with the latest app. 10 years of scientific identification of benthos in samples from surveys performed in connection with oil exploration activities in Greenland. Therefore, it must be emphasized, that the database only represents positive information with regard to where the species has been registered. Hence, the data do not (necessarily) represent the distribution or distributional limits of the species in Greenland waters.

Maintenance and update of the database is expected to happen in connection with future oil exploration activities in Greenland.

Steffen Lundsteen, who played a key role in this work, passed away September 23, 2018.

2 Sammenfatning

Identifikation af marine bunddyr indsamlet i prøver fra arktiske farvande er en udfordring for selv erfarne og specialiserede taksonomer. Prøvetagnings- og analysemetoder er standardiserede internationalt, men bestemmelsen af de indsamlede bunddyr til taksonomisk slægts- eller artsniveau kan være svær: til dels fordi der er begrænset og svært tilgængeligt litteratur til at støtte bestemmelsen, men også fordi mange af de indsamlede arter ganske enkelt er sjældent sete, eller måske endda helt ukendte, af videnskaben.

Igennem årtiers arbejde med arktiske bundfaunaprøver og artsbestemmelse i laboratorierne på Aarhus Universitet (Bioscience) er der opbygget betragtelig og unik viden, data, fotografier og referencebiblioteker, som nu eksisterer i forskellig form internt.

Som en del af *Northeast Greenland Environmental Study Programme* er målet med dette projekt at bevare denne viden for eftertiden, og samtidig gøre den mere tilgængelig som hjælp til artsbestemmelse af arktisk bundfauna.

Det primære resultat af projektet er en database, hvor interne data og viden om arktisk bundfauna fra Aarhus Universitet (Bioscience) samles i standardiseret form.

Database indeholder registreringer af artsbestemte arter fra undersøgelser, som, efter kvalitetssikring, i alt udgør:

- 4998 registreringer af 483 forskellige arter
- 91 fotos af 30 forskellige arter.

Databasen gøres tilgængelig via en hjemmeside, som kræver en adgangskode, der fås ved henvendelse til Aarhus Universitet. På hjemmesiden kan man:

- Klikke sig frem til arter, der tidligere er fundet ved prøvetagning i Grønland via artsgrupper og artslister
- Søge information om disse arter frem ved at angive slægt, fuldt artsnavn eller delvist artsnavn.
- Se information om en art: fotografier, autor, kildehenvisninger og geografisk forekomst af arten angivet på kort og med dybdeudbredelse.

Databasen er ikke et komplet værk over arktisk bundfauna. Den indeholder kun den positive information om, hvilke arter der er blevet fundet hvor i de seneste ca. 10 års videnskabelige artsbestemmelser fra prøver indsamlet i forbindelse med olieselskabers aktiviteter i Grønland.

Vedligehold og opdatering af databasen forventes at ske i forbindelse med fremtidige olieefterforskningsrelaterede aktiviteter i Grønland.

Steffen Lundsteen, hvis identifikationsarbejde databasen er baseret på, deltog i arbejdet frem til sin død d. 23. september 2018.

3 Background

Identification of benthic fauna in the Arctic is still a challenge because of the limited supply of identification literature. The sampling methods and post analysis most often follows international established guidelines, e.g. OSPAR, whereas the detailed identification of sampled animals to preferred taxonomic level (genus, species) is moreover hampered because sampled animals may be unknown to science.

This project therefore has designed and programmed a website housing benthos registrations from Greenland surveys, obtained species photos and identification sources to support identification of benthos species.

The database includes species identified in seabed samples obtained in surveys in Greenland, and which were performed as part of establishing baseline in connection with oil exploration activities.

The database is part of the Greenland Environmental Data Center maintained by DCE – National Centre for Environmental and Energy, Aarhus University, and Greenland Institute of Natural Resources.

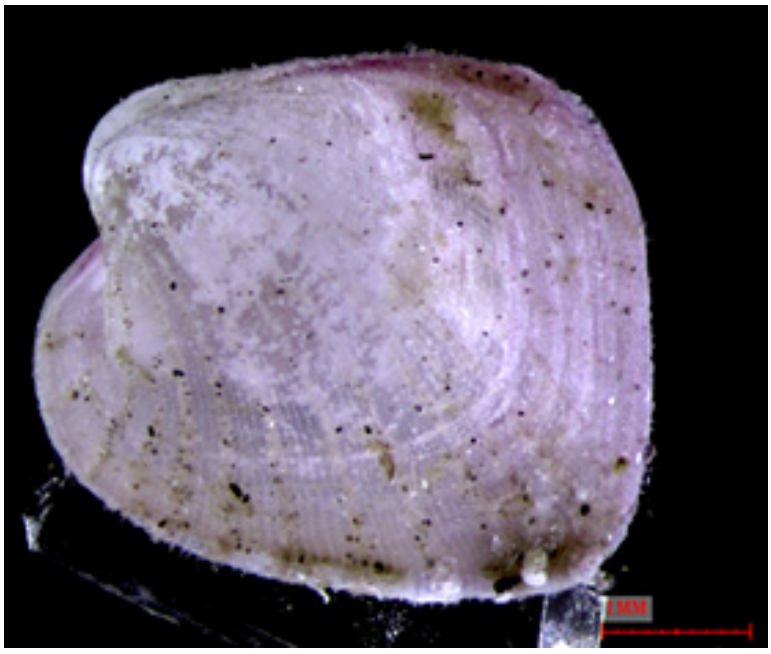
3.1 Aims

A web based macro-fauna identification support system is developed to share visual and spatial species information for aid in identification of Arctic macrofauna.

Thus, the system is developed to present species photos, depth and distribution maps to support species identification supplemented with further photos of species characteristics, and relevant literature references.

The resulting database framework is hence a macro-fauna identification support system including user interface.

Photo 1: *Lyonsiella abyssicola.*



4 Material and methods

4.1 Samples and sampling

Registrations included in the database are species identified from samples taken by box corer, van Veen grab or similar for quantitative and qualitative sampling. All samples were sieved using 1 mm mesh size and conserved in low concentration formaldehyde prior to fauna analysis in accordance with OSPAR (2004).

The database contains information regarding species registered in connection with the latest app. 10 years of scientific identification of benthos in samples from surveys performed in connection with oil exploration activities in Greenland: Kanumas 2008, Store Hellefisk Banke - Disko West, South Greenland 2010, Pamiut 2011, Pitu 2013, Umimmack 2013, Sisimiut oa 2014, NE Greenland 2016 and 2017. Samples not collected by oil companies, are samples collected in projects by Jørgen L.S. Hansen, Mikael Sejr, Aarhus University, and Martin Blicher, Greenland Institute of Natural Resources.

4.1.1 Registrations

The identifications with following photos and references to taxonomic sources have been provided by the benthic fauna expert, Steffen Lundsteen. Sadly, Steffen passed away during the project and thus did not have the opportunity to follow the work to its accomplishment. Therefore, the data in the database have two quality levels; species identifications that are documented by photos and references, and registrations of species with location and abundance, only.

The species have been identified and registered over a 10 years period, and names may have changed during the period for nomenclatorial reasons or changes in taxa delimitations. For potential need to update taxa/species names, most species, with their synonyms, if any, are registered in WoRMS (World Register of Marine Species). In WoRMS, additional information about the species will often also be accessible.

4.1.2 QA of species entries to the database

The taxonomic resolution in the database is at the species level. Fauna registered on a higher taxonomical level (class, order or family) are not included in the database. Any species samples tagged “sp” or “indet” were thus excluded from the database.

4.2 Database software and design

A relational sql database was chosen for the project, since this is well suited to represent the type of structured data in question. With a normalized relational sql database, data consistency is ensured and redundancy is avoided. Using sql queries, data can also easily be extracted from the database for display on the website as well as various other purposes. The relational database diagram shown below (Figure 1) constitutes the structure that enables storing of the benthos species data in a standardized manner.

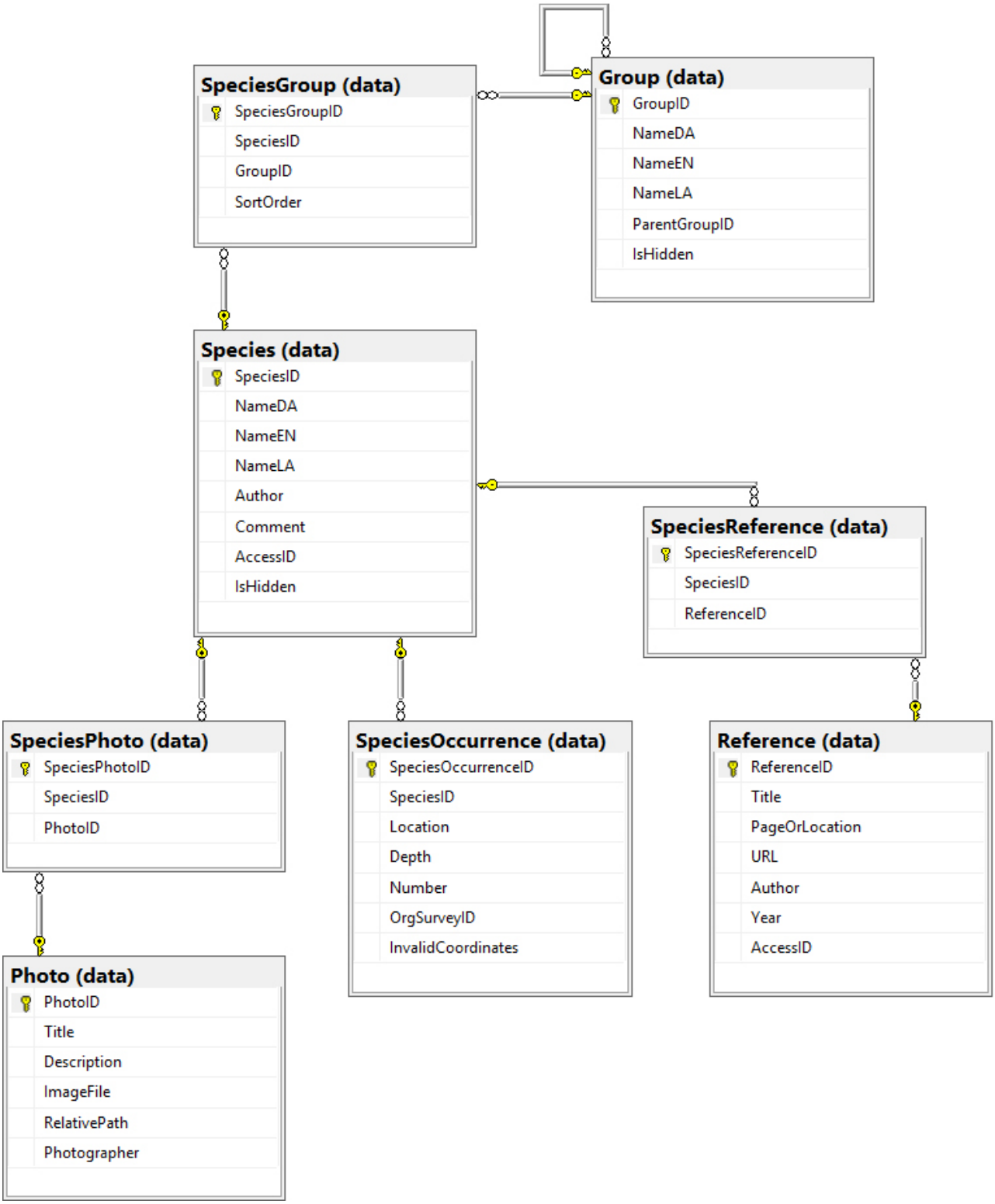


Figure 1. Relational database diagram. The central entity is Species. A Species has a taxonomic group relation. Species may also have 0 to many References - a book or work in which it is documented. In the same way, it may also have 0 to many SpeciesPhotos as well as 0 to many known SpeciesOccurrences – sample locations.

4.2.1 Database programming

Content Management System, CMS

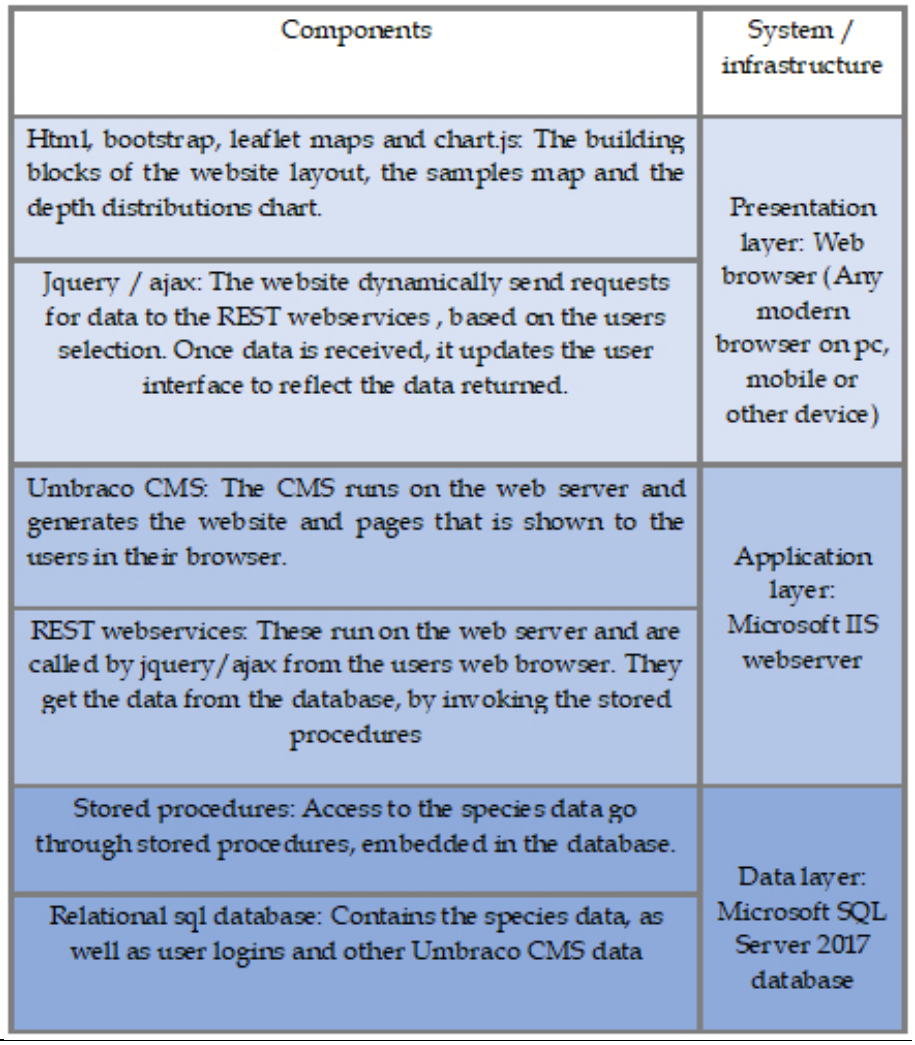
When developing a web based system, it is beneficial to base it on a content management system (CMS). A CMS makes it easier to maintain layout, pages, text content and users via an administrative user interface. For this purpose the widely used open source CMS Umbraco was chosen. The website will be made available on an AU server.

Webservices / server-side programming

As interface between the database and the web pages shown in the users browser, REST (Representational State Transfer) web services were built using Microsoft Visual studio, the C# (C-Sharp) programming language and the WCF (Windows Communication Foundation) web service framework.

Complete architectural overview of the system

Figure 2 presents a diagram showing the components of the entire database system and how they interact. It can be seen as a three tiered architecture with presentation layer, application layer and data layer - a common foundation of almost all dynamic websites. It should be read and interpreted bottom-up.



Components	System / infrastructure
Html, bootstrap, leaflet maps and chart.js: The building blocks of the website layout, the samples map and the depth distributions chart.	Presentation layer: Web browser (Any modern browser on pc, mobile or other device)
Jquery / ajax: The website dynamically send requests for data to the REST webservices , based on the users selection. Once data is received, it updates the user interface to reflect the data returned.	
Umbraco CMS: The CMS runs on the web server and generates the website and pages that is shown to the users in their browser.	Application layer: Microsoft IIS webserver
REST webservices: These run on the web server and are called by jquery/ajax from the users web browser. They get the data from the database, by invoking the stored procedures	
Stored procedures: Access to the species data go through stored procedures, embedded in the database.	Data layer: Microsoft SQL Server 2017 database
Relational sql database: Contains the species data, as well as user logins and other Umbraco CMS data	

Figure 2. Diagram of the database system components and their interaction. The diagram should be read and interpreted bottom-up.

4.2.2 Data Loading and data QA

The primary source of the data loaded in the system, was an internal Aarhus University Access database created and maintained by Steffen Lundsteen.

From this internal database, tables with the subset of data and attributes chosen as relevant for the arctic benthos database system was extracted. Following that, a lot of data exploration was carried out to gain understanding of the data. Then sql queries were written to combine, relate and transform the data to fit into the relational database model described in the previous section.

Data for the web based database include benthos species from Greenland studies identified in 2008-2018 and include 483 species (examples of Excel files in the database are listed in Table 1).

Table 1. Examples of Excel files with benthos species lists from localities in Greenland waters.

Loading of photos for species

Files with photos of Bryozoa, Gastropoda, including Bivalvia, Crustacea and Echinodermata as well as Polychaeta and Priapulida were located and uploaded (Table 2).

This comprehensive library of detailed images of species and species characteristic existed alongside the internal Access database. Most of them were in very high resolution, which is good as source, but not suitable for displaying on a web page.

The images had not been consistently named or indexed in relation to the species database. Therefore, at first the folders containing the photos were reviewed and classified as relevant / not relevant. The result of this was a new set of folders, as well as a document with a list of species, for which useable photos existed.

Then a Python script (see Appendix) was constructed to search the folders, species by species, and find all images with the species name as part of the filename. The located photos were rescaled to a suitable size by the script, written to a new folder and a csv (comma separated values) index file linking species and file names was generated.

Finally, the image index file was used to insert the correct entries in the relational sql database.

Table 2. List of folders with photo material of Greenland species from the AU network drive to be transferred to the web based database.

Name	Date modified	Type
09 Animalia indet_TOM	05-12-2013 13:34	File folder
09 Porifera_Undlades vers1	02-07-2015 10:16	File folder
12 Hydrozoa_Undlades vers 1	15-12-2017 12:51	File folder
13 Scyphozoa_Undlades vers 1	11-02-2015 11:25	File folder
14 Anthozoa_undlades vers 1	03-07-2015 10:54	File folder
15 Bryozoa_8 arter vers 1	14-12-2010 10:41	File folder
21 Priapulida_1art vers 1	23-04-2012 15:04	File folder
22 Polychaeta_få billeder ikke artsbestemte	21-02-2013 10:36	File folder
25 Echiura_få arter_undlades vers 1	03-12-2013 13:30	File folder
27 Sipuncula_ikke artsbestem_undlades vers 1	14-12-2010 10:41	File folder
28 Solenogastre_1 art dårlig billede_undlades vers 1	20-01-2015 12:38	File folder
29 Gastropoda_21 arter vers 1	11-12-2018 12:04	File folder
32 Bivalvia	02-02-2015 09:50	File folder
33 Scaphopoda_undlades vers 1	20-01-2014 13:15	File folder
34 Brachiopoda_2 arter undlades vers 1	30-01-2013 12:23	File folder
35 Crustacea	14-12-2010 10:40	File folder
39 Echinodermata	29-05-2015 15:42	File folder
39 Echinodermata kopi Marin Basis	28-03-2011 12:34	File folder
40 Ascidiacea	14-12-2010 10:36	File folder
98 Mamoralik_2005_divers billeder_kan måske bruges	14-12-2010 10:36	File folder
DIVERSE 11122018	11-12-2018 12:09	File folder

Coordinates

Coordinates of samples were selected from the Access database in prioritized order, pending availability:

1. Coordinates registered for the exact place of the sample.
2. Coordinates registered for the station of the sample.
3. Coordinates registered for the center of the location / area of the sample.
4. Coordinates registered for the center of the survey of the sample.

Once coordinates were loaded to the database, a spatial quality assurance was carried out, marking coordinates that were clearly incorrectly registered as invalid, e.g., on land. Those were excluded from the website.

5 Database content

The MS SQL server database structure behind the website has been designed and loaded with as much data as could be extracted from the original archives.

The contents of the database is currently drawn from the surveys listed in Table 3.

Table 3. List of the benthic surveys from which benthos samples have been obtained for fauna analyses.

Survey	Year	Latitude (WGS84)	Longitude (WGS84)
Kanumas	2008	73,5	-62
Store Hellefisk Banke - Disko West	2009	69,05	-55,23
South Greenland	2010	61,75	-50,4
Pamiut	2011	62	-49,25
Pitu	2013	74,3	-60,71
Umimmack	2013	77,5	-11,3
Sisimiut area	2014	66,85	-54
NE Greenland	2016	78,55	-9,3
NE Greenland	2017	78,68	-5,08

The extract from these surveys, following quality control, totals:

- 4998 registrations of 483 different species.
- 91 photos of 30 different species.

It must be emphasized, that the database only represents positive information with regard to where the species has been registered. Hence, the data do not (necessarily) represent the distribution or distributional limits of the species in Greenland waters.

Identifications from future surveys will be incorporated in the database when a database update is appropriate and funding in place.

6 Guideline to the Arctic benthos database

The website is developed with layout, pages, login functionality and user interface to dynamically interact with the Arctic benthos database, hosted and maintained by Aarhus University, AU. Users must request a license from AU to gain access.

The website is available at this url:

https://benthos.bios-rks-web.bios.au.dk

Screen shots of the user interface of the system is shown in Figure 3-4.

The following functionality is implemented, and can be achieved easily with few clicks.

For logged in users, it is possible to find species either by querying for taxonomic name or browsing the taxonomical groups of species in the database. A list of matching species will be shown (Figure 3):

1. Select a species group from the list OR search by full or partial species name.
2. See the list of matching species (Figure 3).

After selecting the requested species, the user is directed to the page with the contents of the species details (Figure 4).

3. View details of a species on the match list (Figure 4). The details includes:
    a. A list of literature describing the species as well as potential comments. This information may not be available for all species, but a direct link to look up the species in the WoRMS registry will always be available.
    b. A section concerning spatial records of samples presented by an interactive map, which also include the number of registrations per sample. The user can click each sample location to see the coordinates, exact depth, and the number of specimens of the species counted in the sample.
    c. Depth distribution of the species registrations presented in a frequency diagram showing at what depths the samples were taken, in which the species was found.
    d. Multiple images of the species can be viewed in an image-slider.

These information items are only shown when such data is available in the database for the species

The user interface has been prepared to make it very easy to switch back and forth between the list of results and the details window.

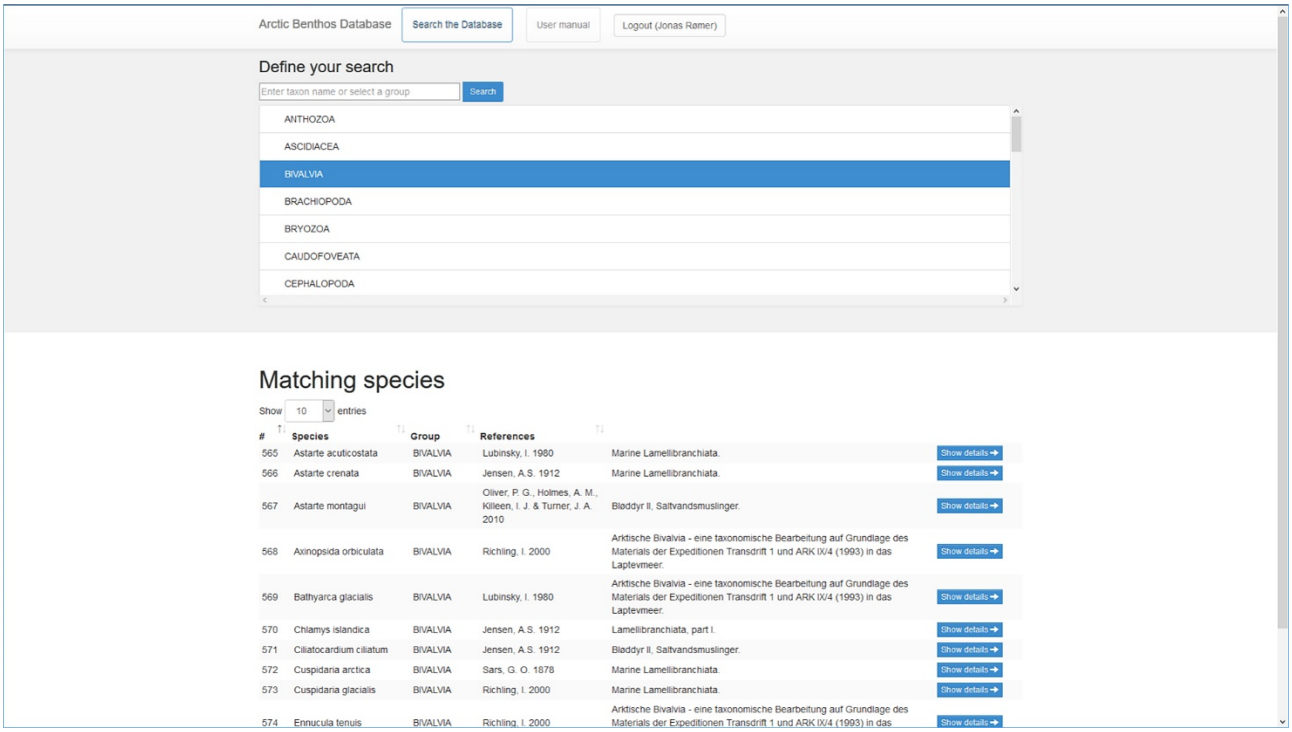


Figure 3. Search function for requested species. It is possible to find species either by querying for taxonomic name or browsing the taxonomical groups of species in the database. A list of matching species is shown.

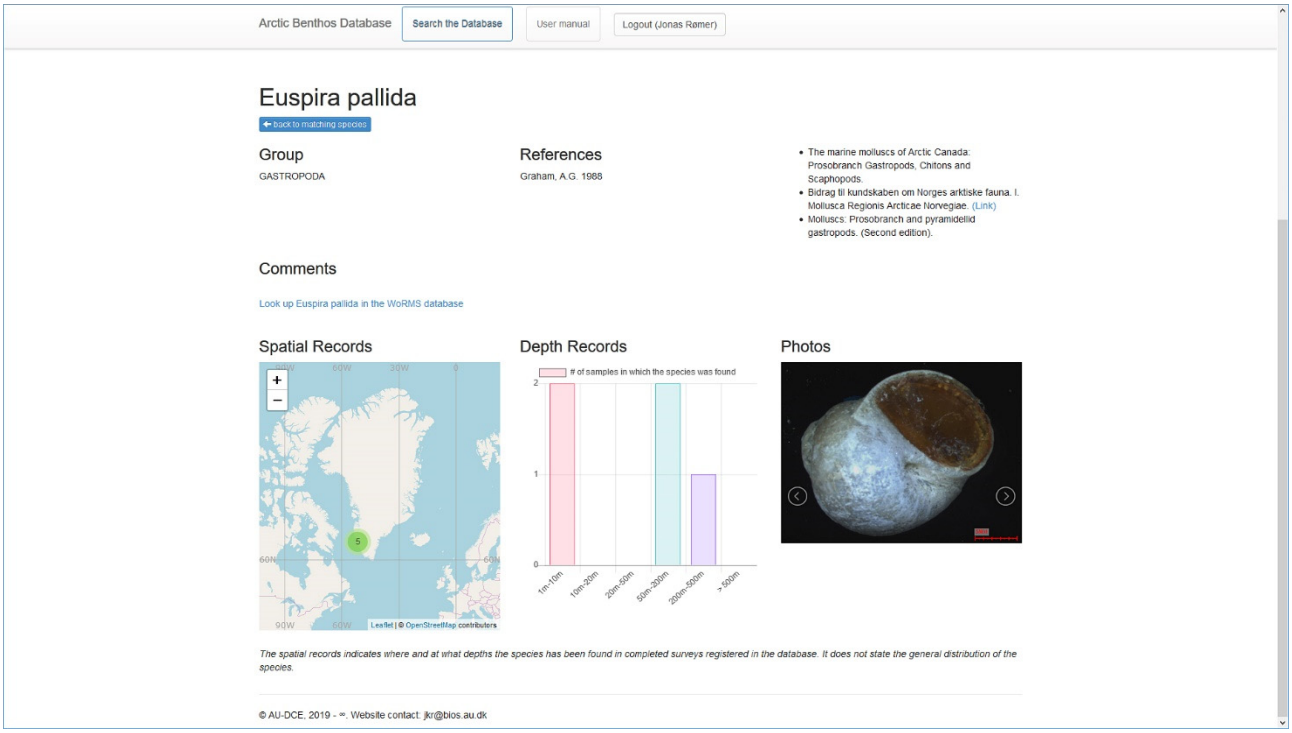


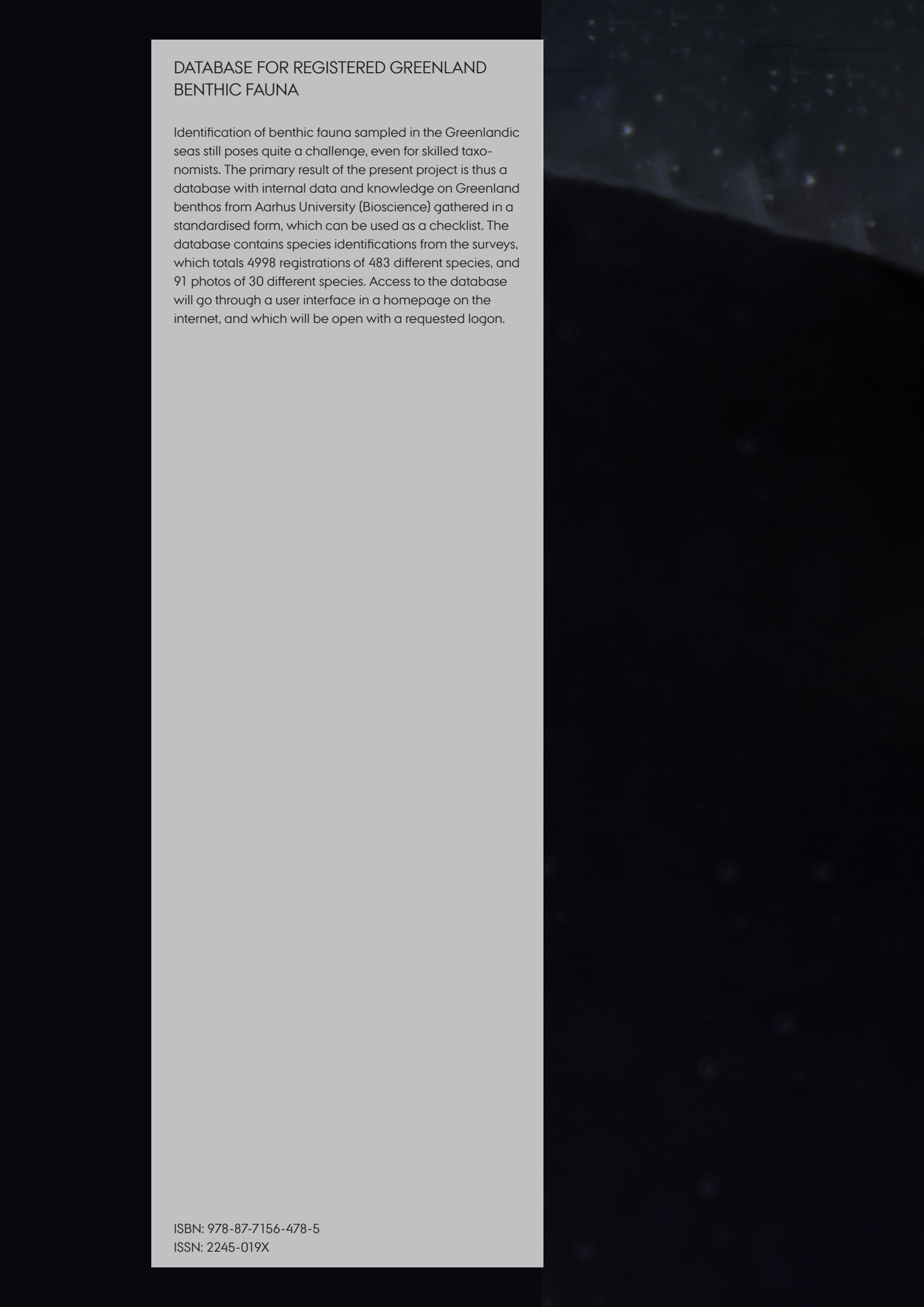
Figure 4. The contents of the species details page: a list of literature for identification and descriptions of the species together with potential comments and a direct link for looking up the species in the WoRMS registry; spatial presentation of registrations in an interactive map including number of registrations per sample; depth distribution of registrations; and photos with species characteristics.

Appendix

The python script used to index images in relation to species

```
import numpy
import pandas
import os
import PIL
from PIL import Image
import ntpath
from shutil import copyfile

# load and filter the csv
csvframe = pandas.read_csv('species_photos.txt', sep='\t')
csvframe_with_images = csvframe[csvframe.Images == 1] # marked as having rele-
vant images
#print(csvframe_with_images)

# list the chosen groups/folders
imagefolders = []
# r=root, d=directories, f = files
for r, d, f in os.walk(u'O:\ST_ElementaerDatabase\Fotos\Grønland fotos, system, MA-
STER'):
    for folder in d:
        imagefolders.append(os.path.join(r, folder))

def find_imagefiles(basedir, speciesstring, alternativespeciesstring):
    files = []
    for p, d, f in os.walk(basedir):
        for file in f:
            if speciesstring.lower() in file.lower():
                files.append(os.path.join(p, d if isinstance(d,str) else '',file))
            if(isinstance(alternativespeciesstring, str)):
                if(alternativespecies-
string.lower() in file.lower()):
                    files.append(os.path.join(p, d if isin-
stance(d,str) else '',str(file)))
    return files

# for each species, find the matching files
# and store results in new image files and a data frame to write as csv for db up-
date
results = pandas.DataFrame(columns=('Species', 'Group', 'Title', 'ImageFile'))
for index, row in csvframe_with_images.iterrows():
    print(row['Taxon'], row['Group'])
    for imagefolder in imagefolders:
        if(imagefolder.lower().endswith('vers 1')): # skip base group folder
            continue
```

```
        #if(row['Taxon'].lower() in image-
folder.lower()): # this is the folder with species images
        imagefiles = find_imagefiles(imagefolder, row['Taxon'], row['AlternativeT-
axon'])
        for imagefile in imagefiles:
            if(imagefile.lower().endswith('.tif') or image-
file.lower().endswith('.tiff') or imagefile.lower().endswith('.jpg') or image-
file.lower().endswith('.bmp')):
                # this is a good image, process it
                try:
                    print(imagefile)
                    if(imagefile.lower().endswith('.tif')):
                        copyfile(imagefile, imagefile+'.tiff') # pil-
low will only open file if it has correct extension
                    image = Image.open(imagefile)
                    if(imagefile.lower().endswith('.tif')):
                        os.remove(imagefile + '.tiff') # pil-
low will only open file if it has correct extension
                    # resize
                    wpercent = (800/float(image.size[0]))
                    hsize = int((float(image.size[1])*float(wpercent)))
                    image.resize((800, hsize), PIL.Image.BICUBIC)
                    # save
                    outputfile = ntpath.basename(imagefile)
                    outputfile = os.path.splitext(os.path.basename(outputfile))[0]
                    title = outputfile
                    outputfile = row['Taxon'] + '_' + row['Group'] + '_' + output-
file + '.jpg'
                    print(outputfile)
                    fullpath = os.path.join('c:\\temp\\benthosimages\\', output-
file)
                    image.save(fullpath)
                    results = results.append({'Species' : row['Tax-
on'], 'Group' : row['Group'], 'Title' : title, 'ImageFile' : outputfile } , ig-
nore_index=True)

                except Exception as e:
                    print(str(e))
                    pass

    row = next(csvframe_with_images.iterrows())

# save the results dataframe to csv
results.to_csv('c:\\temp\\benthosimages\\imageindex.txt', sep='\t')
```



DATABASE FOR REGISTERED GREENLAND BENTHIC FAUNA

Identification of benthic fauna sampled in the Greenlandic seas still poses quite a challenge, even for skilled taxonomists. The primary result of the present project is thus a database with internal data and knowledge on Greenland benthos from Aarhus University (Bioscience) gathered in a standardised form, which can be used as a checklist. The database contains species identifications from the surveys, which totals 4998 registrations of 483 different species, and 91 photos of 30 different species. Access to the database will go through a user interface in a homepage on the internet, and which will be open with a requested logon.

ISBN: 978-87-7156-478-5
ISSN: 2245-019X