Manual- EBP-Nor database

Document version: 20230915 Database page: <u>https://ebp-nor.sfb.uit.no</u>

The function of the EBP-Nor database is to organize the full set of data connected to targeted species in the Earth BioGenome Project Norway project. It also acts as the contextual source for brokering sequence data to ENA through NeLS. This requires adhering to ENA requirements while allowing for a wider scope of descriptors using attributes of the <u>ERGA Sample Manifest</u>.

Purpose of this manual

Part 1 – Database overview.

1.1 Overview of the EBP-Nor database and ENA data model1.2 The EBP-Nor database structures data into 5 internally linked tables

Part 2 – Adding and modifying data.

- 2.1 Adding data to a new record by writing or copy-pasting from any text source.
- 2.2 Modifying a record using the interface directly.
- 2.3 Adding data by uploading a tab-separated file.
- 2.4 Modifying data by uploading a tab-separated file.
- 2.5 Template values.

Part 3 – Linking database records.

- 3.1 Purpose of linking records.
- 3.2 Example linking an Isolate and Specimen record.

Part 4 – Additional functionality of the database.

- 4.1 User roles in the EBP-Nor database.
- 4.2 Applying for a role.
- 4.3 Functionalities as writer.
- 4.4 Quick search and adding filters.
- 4.5 Field types.
- 4.6 Values, Added values, Added sub-values, Delete button.

Part 1 – Database overview.

1.1 Overview of the EBP-Nor database and ENA data model.

In ENA (part of INSDC) the EBP project (<u>PRJNA533106</u>) is structured into multiple project layers (Figure 1.). Like its global counterpart, the EBP-Norway project (<u>PRJEB6531</u>) has been created and will only exist as formal placeholders at project level 2. Working with the EBP-Nor database will interact with levels 3 (species) and 4 (species isolates) in the umbrella structure. The EBP-Nor database is further linked to sample information (Sample), the sequencing method and related files (raw reads), and the genome assembly (Analysis) (Figure 1 and Figure 2). Note that ENA coins the 4th project level as "Study", this can be thought of as an anchor for connecting data.



Figure 1 - How the projects in the EBP umbrella project is structured in ENA/INSDC. The figure shows examples of the content at the various project levels and emphasizes the study and connected data. Figure 2 is superimposed in Figure 1 and represented by the orange, green, teal, and yellow boxes.



Figure 2 - ENA sequence data model.

1.2 The EBP-Nor database is structured as five internally linked tables (Figure 3).



Figure 3 - Internal contextual data relationships in the EBP-Nor database (purple boxes and gray arrows) and how these relate to the ENA data model (purple arrows).

- Species represents the formalized list of targeted species intended for genome sequencing. This
 relates to project level 3 in the EBP umbrella structure of ENA. A record in Species describes the
 formal sampling coordination, compliances, permits, red-list status and level 3 project
 requirements by ENA.
- **Isolates** (not to be confused with specimen) describes the isolation information regarding the spatiotemporal data of the collection event. Also crucial to the **Isolate** is formalizing the Study description in which the connected data can be anchored.
- **Specimen** (more towards the <u>medical/microbial definition</u>) describes what part, organ or tissue is selected for sequencing (and/or biobanking, and/or barcoding). However, only the specimen's name and organ/tissue need to be given due to mandatory requirements in ENA.

- **Sequencing** is where the sequencing method and filename related data are recorded. Note that the experiment (sequencing information) and Run (files) must exist in pairs. One genome assembly may be the result of multiple sequencing runs with different platforms. E.g., mRanTar1 has 4 records (corresponding to experiment-run pairs) where one is illumina paired-end reads in FASTQ, and three are PacBio HiFi reads in BAM formats.
- **Assembly** represents information regarding the assembled genome and will contain the information required by the manifest file during submission.

Part 2 – Adding and modifying data.

Perquisites for adding or modifying data in the database:

- \Rightarrow Need to be logged inn. (FIDE and LS Login can be used)
- \Rightarrow Need to have permission role as writer. (See Part 4 for how to get permission.)

Given the perquisites, data can be added the EBP-Nor database either of by directly writing or copypasting data into the interface, or by uploading a tab-separated file.

2.1 Adding data to a new record by writing or copy-pasting from any text source.

This example will show how a set of prepares data can be added. Data like this can come from text files, spreadsheets, pre-made forms etc. First, we choose the database to add data into (Figure 4). See the description in Part 1.2 for a summary of data covered by the databases.



Figure 4 - Database selection from top section of database home page.

A When adding data, you will work on a draft (unpublished) dataset only accessible to accounts with writer and owner permission. (More info on drafts and published data in Part 4)

For this example, we will work with the Isolate database (Figure 5) to add data manually. In the left menu, we click the (+) button and select "<u>New record</u>", this brings up a window in the tab named *Time and location*. Here the new record can be annotated (Figure 6). Note the tab names! These tabs represent different areas of information covered by the <u>ERGA Sample Manifest</u> and <u>ENA Darwin Tree of Life checklist</u>. In the next example we will fill the minimum sample data required by the ENA DToL checklist.

Isolate										
	1.0 -			Text search		Q	·= 1	<u>u</u> © 🎄	Log	out 💌
	Showing records	1 to 10 out of	10	« ‹	>	»			Page size	25 ~
() ()	Isolate ID	Sampling date	Sampling locatio	n	Sex	Habitat	Species	Specimen	Sequencing	Assembly
0	CA46355/A15 - 000000014	2021-05-20	Søndre Skjælholmen, l Norway	Nesodden, Viken,	Female	•	<u>000000333</u>	<u>000000011</u>	- <u>000000003</u> - <u>000000004</u>	- 000000002 - 000000002
	AL886 - 000000011	2022-04-03	Norway > Norwegian sea		Male	Marine ecosystem	<u>000000358</u>	<u>000000010</u>		
	AL877 - 000000010	2022-08-23	Barents sea		Female	Marine ecosystem	00000358	00000009		

Figure 5 - Screenshot of the top page of the Isolate section in the database.

- 1.0	
Time and location	Isolate features Natural history specimen Collector Taxonomy Related records
ENA Submission -	Record metadata
Time and loca	ation
Sampling date	YYYY-MM-DD or YYYY-MM or YYYY
Sampling location	
③ Sampling coordinates	Latitude N Longitude E 🖛 🐵 🛱
 Collection method 	
 Habitat 	- · · :
③ Depth	Number m 🖛 🕑 🗑
③ Elevation	Number m 🖷 🔅 💼
 Time of day 	
	Cancel Save

Figure 6 - Popup window for adding data to a new record.

By clicking on the tab "ENA Submission" (Figure 6), then choosing "Mandatory Sample", we see the fields required by the ENA DToL checklist (Figure 7). Most fields are free text and information can be added directly by writing or copy pasted. Important information on some of the attributes:

- The <u>Project title</u> is a template sentence (see template naming 2.1.5) and must be written as: "<genus> <species> (<common name>) genome assembly, <ToLID+NUM>". See Figure 8 on how the Project title is written for mRanTar1. For more on the ToLID check the <u>official page</u>.
- The <u>Sampling date</u> is restricted to the ISO 8601 format without time of day.
- The <u>Sampling coordinates</u> are restricted to Decimal degrees (DD).

- The <u>Habitat</u> is regarded as free text if the first column is "-".
 - By selecting "envo" <u>*The Environment Ontology*</u> can be used.
 - To correctly use "envo" the second column requires the numerical identifier of a term to collect and display its environment information.
 - o <u>Ontobee</u> (and similar ontology lookup tools) can be used to search for a relevant term.
 - For example, the mRanTar1 (Figure 8) we have used "03400001" which is the numerical identifier of the term *Area of alpine tundra*. You can find this by searching the relevant sampling terrain in Ontobee.
- Further information on each attribute can be found by hovering over the (i) icon.

Mandatory Sample				
 Project title 				=+ ⊕ ₪
 Collector 				=+ () ()
 Collector affiliation 				=+ ↔ ₪
 Sampling date 	YYYY-MM-DD or YYYY-MM or	• үүүү		=+ ⊕ ₪
Sampling location				=+ ⊕ ₪
 Sampling coordinates 	Latitude	N Longitude	E	=+ ↔
 Habitat 	-	~ :		=+ ⊕ ₪
 Lifestage 				=+ ⊕ ₪
Collection method				=+ ⊕ ₪
(j) Sex				=+ ⊕ ₪
			Cance	el Save

Figure 7 - The minimum sample information required by ENA to publish sequence data under the DToL checklist.

Mandatory Sample				
 Project title 	Rangifer tarandus platyrhyncus (Svalbard reindeer) genome assembly, mRanT;			
① Collector	Vebjørn Veiberg	=+ ⊕ ₪		
① Collector affiliation	Norwegian Institute for Nature Research	=+ ⊕ ₪		
③ Sampling date	2021-10-19	=+ ⊕ ₪		
③ Sampling location	Svalbard > Nordenskiöld Land > Semmeldalen	=+ ⊕ ₪		
③ Sampling coordinates	77,991317 N 15,371994 E	=+ ⊕ ₪		
① Habitat	envo ~ : 3400001	=+ ⊕ ඕ		
① Lifestage	Adult	=+ ↔		
① Collection method	Animal shot for scientific purpose	=+ ↔		
① Sex	Male	≡₊ ⊕ ₪		
	Ca	Incel		

Figure 8 - Example information filled in for the mRanTar1.

- The information in Figure 8 was simply copy pasted from a prepared text file into the fields, then saved.
- However, the <u>Sampling location</u> was structured into three names for the Country/region, area/municipality, and site name by using the *Add sub-value* button on the right side of the field (For more see section 4.6).
- After saving it may be necessary to updated the table view by refreshing the webpage to see the new record or changes made in the table.

2.2 Modifying a record using the interface directly.

Any record can be modified. To do so, simply click on the record in the table to access its data. In the window popping up (Figure 9), click on the edit button in the upper left corner to start editing. Changes will be kept upon clicking the save button.

φ+	You are watching a draft version of this record.						
	mRanTar1 - 00000002 10						
	Time and location Isolate features Natural history specimen Collector Taxonomy Related records						
	ENA Submission - Record metadata						
	Time and location						
	③ Sampling date	2021-10-19					
	③ Sampling location	Svalbard Nordenskiöld Land Semmeldalen 					
	 Sampling coordinates 	(gmap) 77.991317,15.371994					
	 Collection method 	Animal shot for scientific purpose					
	(i) Habitat	Area of alpine tundra 🛈					
	③ Depth	- m					
	 Elevation 	- m					
	Time of day	15:00					
		Close					

Figure 9 - Popup window with existing values. Note the buttons in the upper left for download, edit, and delete the current record.

2.3 Adding data by uploading a tab-separated file.

The EBP-Nor database also accepts tab-separated tsv files as input for adding new records to the database. This can be helpful if you have many records to add at the same time or have filled a sampling sheet.

• The way the data upload works is by recognizing the tabular headers (first line) of the tsv file. For example, the <u>Sampling date</u> attribute as seen in Figure 9 has a short name *sam:date* (you can see this by hovering over the (i) icon. If a column in the tsv file has the header *sam:date* it will pass the uploaded information into the <u>Sampling date</u> field.

The easiest way to get all the attribute short names is by downloading the data in the table as tsv (Figure 10). Open the tsv file (ebp-nor_isolate_x.x.tsv, see Figure 11) in the program of your choice (e.g., excel or similar) and remove all lines except the header (you might want to keep some data for reference, but remove it before uploading your own data). You can now add you own data into the relevant columns. One line represents one new record added to the database.



Figure 10 - The download functionality as seen from the Isolate database.

Р	Q	R	S	Т	U	1
sam:coll_affi	sam:coll_by	sam:coll_orcid	sam:coordinate	sam:date	sam:depth	Ī
	Morten Helberg			2021-05-20		
UiT Norges Arktiske Universitet	Arve Lynghammer		gmap:66.958,8.181	2022-04-03	540	
UiT Norges Arktiske Universitet	Arve Lynghammer		gmap:73.5337332,-026.4499467	2022-08-23		
Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.556,-21.663	2022-09-02		
Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.556,-21.663	2022-09-02		
Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.098,-19.867	2022-09-03		
Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.098,-19.867	2022-09-03		
Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.098,-19.867	2022-09-04		
Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.556,-21.663	2022-09-02		
Norwegian Institute for Nature Research	Vebjørn Veiberg		gmap:77.991317,15.371994	2021-10-19		

Figure 11 - Small selection of downloaded data from Isolate showing the header line.

A Note that the column "id" must be empty/omitted when uploading new records. This will make sure new records are created with new and unique identifiers.

A Note columns without data can be skipped.

0	Р	Q	R	S	Т	U
id	sam:coll_affi	sam:coll_by	sam:coll_orcid	sam:coordinate	sam:date	sam:depth
	Some museum	Ola Normann	0000-0001	75,-20	2023-01-01	20

Figure 12 - Some fields filled, while the id column is left empty.

Once you have added the information into the respective columns and are happy with the results you need to save or export your data as tab-separated (.tsv). To upload the tsv data do as in Figure 5 by clicking on the (+) button on the table and choose "Upload TSV", then choose "Create/Overwrite" and select your tsv file. Finally click the "Upload" button.

The tsv upload will be format validated, e.g., if text is found in a numeric field it will generate an error.



Figure 13 – Successful upload if no errors in formatting is detected.



Figure 14 - If formatting errors are detected it will show how many records are affected. Here only one.

If you happen to get errors, you can download an annotated _errors.tsv file by clicking on the link in the upload results (under point 1 in Figure 14).

Opening the _errors.tsv you will see the first column is called _error:annotations and will describe which attribute contains an error. It further shows the erroneous value and why it is wrong. In this example the error looks like this: validating: sam:elevation: Value 'none' is not a number. This is obvious as sam:elevation is requiring a numeric value to represent meters above sea level, here it is given the text *none*. This can be corrected by simply removing the value or replacing it with a numeric value. Once corrected the column _error:annotations can be removed and the data uploaded as described.

Note that if you upload 10 records and 2 fails due to errors, the remaining 8 will be uploaded while the failed records will be listed in the _errors.tsv.

2.4 Modifying data by uploading a tab-separated file.

For this to work the record(s) you attempt to modify need to exist in the database, and thus, the simplest approach is to start by downloading the data in the database as described in the top of section 2.1.3. A few columns in the Isolate data is shown in Figure 15where it is marked in green.

A Note that the column "id" must be included with the correct id number when uploading a tsv file to modify existing records. This will make sure the correct records are overwritten.

0							
id	sam:coll_affi	sam:coll_by	sam:coll_orcid	sam:coordinate	sam:date	sam:depth	sam:elevation
ebpnor.isolate:00000014		Morten Helberg			2021-05-20		2
ebpnor.isolate:000000011	UiT Norges Arktiske Universitet	Arve Lynghammer		gmap:66.958,8.181	2022-04-03	540	
ebpnor.isolate:000000010	UIT Norges Arktiske Universitet	Arve Lynghammer		gmap:73.5337332,-026.4499467	2022-08-23		
ebpnor.isolate:00000008	Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.556,-21.663	2022-09-02		
ebpnor.isolate:000000007	Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.556,-21.663	2022-09-02		
ebpnor.isolate:000000006	Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.098,-19.867	2022-09-03		
ebpnor.isolate:000000005	Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.098,-19.867	2022-09-03		
ebpnor.isolate:000000004	Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.098,-19.867	2022-09-04		
ebpnor.isolate:000000003	Globe Institute UiT Norges Arktiske Universitet	Julie Bitz-Thorsen		gmap:75.556,-21.663	2022-09-02		
ebpnor.isolate:00000002	Norwegian Institute for Nature Research	Vebjørn Veiberg		gmap:77.991317,15.371994	2021-10-19		
C	0						

Figure 15 - When updating a set of records the id column with the respective ids must be included.

You might like to perform changes on multiple records, e.g., the name of the person who performed the sampling due to typos or other reasons. Once the changes are made to the sheet, exported/saved as tsv format, it can be uploaded. To upload the tsv data you do as shown in Figure 5 by clicking on the (+) button on the table and choose "Upload TSV", then choose "Update existing" and select your tsv file. Finally click the "Upload" button.

2.5 Template values.

Several fields in the database are restricted to follow a given sentence syntax, e.g., for Species (Table 1). This is to adhere to the unique naming in ENA as well as using DToL standard naming in the umbrella projects. However, ebp-nor_<prj/exp/run> have been added to identify it under EBP-Norway.

If we follow the 2.1.1 for adding data manually to the Isolate database, then clicking on the tab "ENA Submission" (Figure 6), then choosing "Mandatory Project", we see the fields required by ENA (Figure 16). This corresponds with the level 4 Project (Study) in the EBP project umbrella.

Mandatory Project		
① Project alias		=+ ⊕ ₪
③ Project description		=+ ⊕ ₪
③ Project title		=+ ⊕ ₪
③ Parent project accessions	- ~ :	=+ ⊕ ₪
 Release date 	YYYY-MM-DD or YYYY-MM or YYYY	=+ ⊕ ₪

Figure 16 - Required information by ENA for the study description.

The values going into "Mandatory Project" and its 3 attributes Project name/alias/description are predetermined template names based on species name and ToLID as shown in Table 2.

A Note that the Sequencing database (Table 4) has a number (NUM) after the ebp-nor_exp and ebpnor_run. This is a unique running number for each experiment-run set (see Figure 4). If a specimen is sequenced once by illumina and once by Pacbio there are two experiment-run sets, one having exp1 and run1, the other exp2 and run2. The order is arbitrary.

Species	Template syntax	Template example
Project name	ToLID	mRanTar
Project title	<genus> <species> (<common< td=""><td>Rangifer tarandus platyrhyncus</td></common<></species></genus>	Rangifer tarandus platyrhyncus
	name>) genome assembly, <tolid></tolid>	(Svalbard reindeer) genome assembly,
		mRanTar
Project alias	ebp-nor_prj_ <tolid></tolid>	ebp-nor_prj_mRanTar
Project description	This project provides the genome	This project provides the genome
	assembly of <genus> <species>,</species></genus>	assembly of Rangifer tarandus
	common name <common name="">,</common>	platyrhyncus, common name Svalbard
	based on a sample provided by	reindeer, based on a sample provided by
	samplers of EBP-Nor.	samplers of EBP-Nor.

Table 1 – Species database.	Project level 3	information fo	or ENA following the	ENA/DToL unique syntax nami	ing.
-----------------------------	-----------------	----------------	----------------------	-----------------------------	------

Table 2- Isolate database. Project level 4 information for ENA following the ENA/DToL unique syntax naming.

Isolates	Template syntax	Template example
Project title	<genus> <species> (<common< td=""><td>Rangifer tarandus platyrhyncus</td></common<></species></genus>	Rangifer tarandus platyrhyncus
	name>) genome assembly,	(Svalbard reindeer) genome assembly,
	<tolid+num></tolid+num>	mRanTar1
Project alias	ebp-nor_prj_ <tolid+num></tolid+num>	ebp-nor_prj_mRanTar1
Project description	This project provides the genome	This project provides the genome
	assembly of <genus> <species>,</species></genus>	assembly of Rangifer tarandus
	common name <common name="">,</common>	platyrhyncus, common name Svalbard
	based on a sample provided by	reindeer, based on a sample provided by
	samplers of EBP-Nor.	samplers of EBP-Nor.

Table 3 - Specimen database. Sample information for ENA following the ENA/DToL unique syntax naming.

Specimen	Template syntax	Template example
Specimen name	<tolid+num>_<tissue></tissue></tolid+num>	mRanTar1_muscle

Table 4 - Sequencing database. Sequencing information for ENA following the ENA/DToL unique syntax naming.

Sequencing	Template syntax	Template example
Experiment alias	ebp-nor_exp <num>_<tolid+num>_<tissue></tissue></tolid+num></num>	ebp-nor_exp3_mrantar1_muscle
Run alias	ebp-nor_run <num>_<tolid+num>_<tissue></tissue></tolid+num></num>	ebp-nor_run3_mrantar1_muscle
Experiment	Same as Experiment alias	Same as Experiment alias
reference		

Part 3 – Linking database records.

3.1 Purpose of linking records.

Connecting records between the databases is possible and required for publishing the dataset (metadata and sequence data) in ENA. In Figure 17 we have examplified the Svalbard reindeer (mRanTar) to show how it is linked from the Species database with records across the other databases. Links can be interacted with to see their respective data.

Species name	Unique ID	Common name (nor.)	Common name (engl.)	Family	Isolate	Specimen	Sequencing	Assembly
Rangifer tarandus platyrhyncus	<u>000000231</u>	Svalbardrein	Svalbard reindeer	<u>Cervidae</u>	00000002	<u>000000002</u>	- 000000001 - 000000002 - 000000005 - 000000006	 00000001 00000001 00000001 00000001 00000001

Figure 17 - The Svalbard reindeer linked with its isolate, specimen, sequencing and assembly records. Note that there is only one assembly (that 4 are listed is due to a small bug).

The Unique ID (short name "id") is used to link the records between databases. This linkage can be illustrated as shown in Figure 18.



Figure 18 - Linking of unique id numbers between the databases.

Tips and perquisites for making a link between two records in different databases:

- \Rightarrow Need to have writer permission and be logged into both databases the two records are saved.
- ⇒ Connect adjacent records as shown in Figure 18, e.g., from Species you connect a Specimen. From Isolate you can connect with a Specimen. If a connection exists, e.g., Sequencing with assembly, it will automatically be included if a Specimen becomes connected with the Sequencing. It is not possible to connect Species directly with Sequencing, Assembly or Isolates.
- \Rightarrow Know the id numbers of the records you like to link.

3.2 Example linking an Isolate and Specimen record.

For this example, we have created two dummy records in the Isolate and Specimen databases (Figure 19 and 20) that we will connect. In the Isolate database the id of the dummy record is 000000015 and in the Specimen it is 000000018.

Isolate										
	1.0 -			Text search		٩	= L	<u>h</u> © 🕸	Log	out 🝷
	Showing records	1 to 11 out of	11	« «	>	>>>		I	Page size	25 ~
⊕ •	Isolate ID	Sampling date	Sampling location	n	Sex	Habitat	Species	Specimen	Sequencing	Assembly
0	Test_isolate - 000000015	1999	Norway		-	forest				

Figure 19 - Dummy record in Isolates which is not connected to any other record.

Specim	ien						
	1.0 -		Text search	٩		L 🕸	Log out 🔹
	Showing records 1 to 17 out of 1	7	« < >	»		Page s	ize 25 ~
Φ¥							
⊕	Specimen name and ID	Organism part	or tissue	Species	Isolate	Sequencing	Assembly
	Test_specimen - 000000018	skin					

Figure 20 - Dummy record in Specimen which is not connected to any other record.

- Simply click on the record in one of the databases (Figure 19). We use Isolate in this example.
 - It is arbitrary which database is used to connect the adjacent record, here Isolate or Specimen can be selected since we are creating a bidirectional connection between these two databases.
- Click on Edit mode and go to the tab Related records (Figure 21).
- Here (Figure 22) you will see a link button [%] to Specimen.
 - \circ $% \ensuremath{\mathsf{Note}}$ Note that relative to the Isolate database you cannot connect directly to any other database.

- Clicking on the link button brings up an input field (figure 23).
 - Here we have copy-pasted the id of the dummy specimen.
- We then click OK and Save the record.
- Now the two records are successfully linked as shown from the table view in Isolates (Figure 24 and 25).
 - \circ $\;$ The web browser might need to be reloaded to see the linkage.
- Note that links can easily be broken by Editing the linkage in Related records.

@+	bu are watching a draft version of this record.	A
	Test_isolate - 000000015	
	Time and location Isolate features Natural history specimen Collector Taxonom	y Related records
	ENA Submission - Record metadata	
	Time and location	
	③ Sampling date	1999
	③ Sampling location	Norway
	③ Sampling coordinates	-
	③ Collection method	-
	③ Habitat	forest
	() Depth	- m
	() Elevation	135 m
	() Time of day	-
		Close

Figure 21 - The test isolate default view.

φ+	You are watching a c	raft version of this record.
	Test_isolate	- 00000015 10
	Time and location	Isolate features Natural history specimen Collector Taxonomy Related records
	ENA Submission -	Record metadata
	Related recor	ds
	(i) Species	- · · · · · · ·
	 Specimen 	
	(i) Sequencing	- • · · · · · ·
	 Assembly 	- v : -
		Cancel Save

Figure 22 - Test isolate on Related records tab.

3 🛡 0 🕂	New 🖉 Edit Page 📵	Edit with Eler	JavaScript from "ebp-nor.sfb.uit.no"
• •	You are watching a	draft versio	Enter the identifier of the record you want to link to.
	Test_isolate	∍-00(OK Cancel
	Time and location	Isolate fea	tures Natural history specimen Collector l'axonomy Related records
	ENA Submission -	Record m	etadata
	Related recor	rds	
	 Species 	-	✓ : -
	③ Specimen	ę	
	 Sequencing 	-	✓ : -
	③ Assembly	-	✓ : -
			Cancel Save

Figure 23 - The pop-up field for inputting the Specimen id number.

Isolate										
	1.0 -			Text search		٩		<u>u</u> 🕲 👳	Log	out 🝷
■ +	Showing records	1 to 11 out of	11	« «	>	»		l	Page size	25 ~
() ()	Isolate ID	Sampling date	Sampling location	n	Sex	Habitat	Species	Specimen	Sequencing	Assembly

Figure 24 - Test isolate linked with Specimen record 00000018.

Specim	nen					
	1.0 -	Text se	earch	۹ 🔳	<u>u</u> 🕸	Log out 🔹
■ ►	Showing records 1 to 17 out of 1	7	« < > »		Page	size 25 🗸
⊕ ▶	Specimen name and ID	Organism part or tissue	Species	s Isolate	Sequencing	Assembly
	Test_specimen - 000000018	skin		00000015		

Figure 25 - Test Specimen linked with Isolate record 00000015.

Part 4

Other information about the database functionalities.

4.1 User roles in the EBP-Nor database.

There are four membership roles independent for each of the 5 databases, where the higher tier always includes the privileges of the lower:

- Visitor / Pending request
 - Can only browse published data and download single records.
- Reader
 - Can download all published data.
- Writer
 - Access to editing draft entries (add, remove, modify including TSV file operations).
- Owner
 - Can edit database configuration and membership.

4.2 Applying for a role.

Navigate to the database you like to work with and use the log in button (Fig. 2) to first register/log in to Life Science Login. Afterwards you can apply for a writer role in the database if you intend to work and edit entries. Follow the same progress for the other databases if these are relevant for you to edit. You might contact one of the database owners once you have applied for a role since there are no current automatic notifications sent.

	Log in	•	
е	25	Lo	ogin
		R	egister

4.3 Functionalities as writer.

Navigation. The Contextual biodata framework (CBF) is the system which provides all functionalities in the databases and the links connecting them as a relational database. Its foundation is the MongoDB database. Fig. 3 below shows the basic functions for navigating the databases. Clicking on a listed record brings up all the attributes related to it.

Note 1. If you are a writer/owner, you will see the draft version of the database if logged in. Given that version 1.0 is published for all visitors to see, you as a writer will see version 1.1 in italic (Fig. 3B) and can easily switch to view v1.0.

Note 2. If changes are made to any attribute of any record in draft version *1.1* the CBF will publish this automatically at 2 AM in the night. Afterwards, visitors will see version 1.1 and you as a writer/owner will see version *1.2* deprecate a version if needed, to make it unavailable for visitors.

4.4 Quick search and adding filters.

In the search box, shown in Fig. 3, you may type in a search word and hit enter to perform a quick search. Clicking on the <u>magnifier button-icon</u> brings up the filtering functions. Simple to use but can look daunting due to the script showing up.

A filter consists of statements. A statement holds the attribute to search within, the search function to use, and what to search for. See screenshot below for a simple example of filtering the Scientific name attribute only.

- 1. Upper: The attribute to apply the filter. Example: Scientific name
- 2. Middle: The type of filter (equals, regex, before and after (for date fields), ontologies/compact IRI, CV (Controlled Vocabulary) and so forth.
- 3. Lower: Input and/or secondary options for the filter type. E.g., case sensitive, what ontology, what CV.

Filters	
AND	Add statement
Now filter	Scientific name 🗸
	equals ~
	Alces alces
	Add
	? Clear all Apply

An uncomplicated guide to make a filter showing the 2 genera Saxifraga and Labrus:

Click on the magnifier button in the species database.

Under add statement chose the following dropdowns from the top: Scientific name, matches (regex), Case insensitive, type in "Saxifraga*," click Add. Now only change the last input by removing "Saxifraga*" and type in "Labrus*" instead, click Add again. Your filter should look like the screenshot below. Click Apply to filter the table accordingly.

Insensitive search does not consider caps. And in regex the asterisk (*) means that anything proceeding does not count and will show up after applying the filter.

Filters					
	OR	{"\$or": [{"attrs.tax:sci_name.obj.val": {\$regex: /Saxifraga*/i]}, {"attrs.tax:sci_name.obj.iri": {\$regex: /Saxifraga*/i]}]}		Add statement Scientific name ~	
AND	OR {"\$or": [{"attrs.tax:sci_name.obj.val": {\$regex: /Labrus*/i}}, {"attrs.tax:sci_name.obj.iri": {\$regex: /Labrus*/i}}]}			matches (regex) ~ Case insensitive ~ Labrus*	
New filter	r			Add	
				? Clear all Apply	

Result:

1.1 🔻					1 filter(s	s) applied	٩	ت <u>الل</u> ۵	\$	L	.og out 👻
Showing rec (total datase	ords 1 to 6 o t count: 145	ut of 6)				« « » »				Page size	25 ~
Scientific name	Synonym name	Family	Sampling status	Isolate name(s)	Isolate	Specimen name(s)	Specimen	Sequencing name(s)	Sequencing	Assembly name	Assembly
Labrus bergylta (000000362)	-	Labridae	-	-		-		-		-	
Saxifraga cotyledon (000000384)	-	Saxifragaceae	-	-		-		-		-	
Saxifraga osloensis (000000291)	-	Saxifragaceae	-	-		-		-		-	
Saxifraga adscendens (000000432)	-	Saxifragaceae	-	-		-		-		-	
Saxifraga tridactylites (000000418)	-	Saxifragaceae	-	-		-		-		-	
Saxifraga hyperborea (000000424)	-	Saxifragaceae	-	-		-		-		-	
Showing rec (total datase	ords 1 to 6 o t count: 145	out of 6				« « » »				Page size	25 ~

4.5 Field types.

In edit mode the simplest field types are free text and are recognized by not having any dropdown choices (controlled vocabulary), nor ontology associated. However, free text fields can be mandatory as shown below, preventing the record from being saved if not filled.

▲ Isolate name	0	=+ (
	This field is compulsory.		

Controlled vocabularies, shown below, are pre-defined choices and are recognized by a small down-arrow on the field's right-hand side.

(i) Specimen id mix risk	✓ -	
	No Not applicable Yes	ncel Save

Date fields follow the ISO8601 format and accept year, month, date in truncated variants.

 Sampling date 	YYYY-MM-DD or YYYY-MM or YYYY	[=+	Ð	Ū	
			_			'

The *environment material* and *habitat* attributes are examples where ontology use is supported. Ontologies are recognized by 2 fields, a drop-down menu, and an input field. By selecting the "-" no ontology will be used, and the input field works as a free text field. By selecting the "envo" you can choose to use the Environment Ontology (ENVO) and input identifiers representing a class term. E.g., to link the envo class "temperate forest" you must input the corresponding identifier "01001805" and save the record (use <u>Ontobee</u> or other envo ontology search tools to find classes and their identifiers). This will make CBF collect the ontology class information and make a link to ebi.ac.uk/ols for the given term and show a tooltip containing the term definition.

 Risk of mixed samples 	No		✓ =+ ⊕ 🗑
③ Environment material	✓ envo	:	=+ ⊕ 🗊

	 Environment material 	envo	× :	01001805		=+	Ð	Ī
--	--	------	-----	----------	--	----	---	---

Result:

 Environment material 	Temperate forest ()
 Risk of mixed samples 	envo:01001805 No
 Environment material 	Description: A forest ecosystem which is subject
③ Environment sample name	to temperate climate 123_01 conditions.

Ontologies can also be used when referring to publications and personnel. For publications Digital Object Identifiers (DOI) and PubMed identification numbers (PMID) can be used, and ORCIDs (Open Researcher and Contributor Identifier) can be used for registered people. These can be selected similarly as described above for the envo ontology. By inputting values as the listed examples below will provide links for attributes like *original publication* and *Collector ORCID*. Links will lead to the DOI target, PubMed page or ORCID page.

- DOI example
 - o 10.1002/ece3.5991
- PubMed example
 - o **32184981**
- ORCID example
 - o 0000-0002-1825-0097

Coordinate fields also has 2 inputs, the Latitude and Longitude. The CBF accepts the easy-to-use Decimal Degrees (DD) format which looks like this example from the Trondheim fjord: 63.511722, 10.485166. The format only consists of numbers. To be valid the Latitude (first number) must be within –90 and 90, and Longitude (second number) must be within –180 and 180. Saved coordinates for isolates will be shown on the map and will get a google map link. Note that the more decimal numbers are used the more accurate the coordinates are. E.g., 63.511722, 10.485166 is more accurate than 63.5, 10.5.

Γ						
	(i) Sampling coordinates	63.511722	Ν	10.485166	Е	=+ ↔ 🛅
L						

4.6 Values, Added values, Added sub-values, Delete button.

When editing the add value button is used for making a list. This can be convenient for situations where more than one information item needs to be stored, like listing 2 names. This function can be used for all attributes. On exporting the data as a tsv these values will be separated by a pipe symbol e.g., Ola Normann | Kari Normann.

(i) Collector	Ola Normann	Add value 🗩 💼
	Kari Normann	=+ ⊕ 🗊

Result:

(i) Collector	- Kari Normann
	- Ola Normann

The Add sub-value function is useful for making lists of hierarchical values, like a sampling location where 3 levels (Region, Municipality, Site) can be listed. On exporting the data as tsv these values will be separated by the less/more than symbol: Troms>Tromsø>Kvaløysletta.

 Sampling location 	Troms		Ac	Add sub-value 🚘 🕣 🛅					
	>	Tromsø	=+						
	>	Kvaløysletta	≡+	Î					

The added values and added sub-values can also be combined as shown in the example below. An added value (-) can get added sub-values (>).

 Collector 	- Kari Normann
	> Oslo
	- Ola Normann
	> Bergen

Delete button.

The delete button is used to clear the input field for any data, but if used on an added value or sub-value it will remove it. Note, clearing the main value will clear all added values and sub-values.

4.7 Linking external media.

By using the § character most values in a field can be linked with an external URL link. In the example below the collection method (a Swab sample) has been linked with a document on the sampling guidelines for molds. The input field can be given a value e.g., "Swab sample" followed by the paragraph character (§) followed by an URL. Any media supported by an URL link can be made.

- [value]§[URL]
- Swab sample§https://aemtek.com/wp-content/uploads/2019/10/AEMTEK-Mold-Sampling-Guide.2019.pdf

Collection method

Swab sample

4.8 Error messages.

If editing or adding data in a database while not logged in, you will get a 400-error. E.g., if you create a new isolate record from the Species database (See option 1 for linking records above) but you are not logged into the Isolate database.

JavaScript from "ebp-nor.sfb.uit.no"		
400: A client err Missing parame	ror occurred eter: ver	
ic a		ОК

Errors can occur when using the quick search function (2022-06-07). This is currently a reported bug in the CBF which is to be corrected.



Save often if you are a writer! If you have been inactive on the site for a while you might get the error below. A sort of halfway logged-out state. You must reload the page in your browser to correct it. Work is ongoing to improve CBF on this issue by moving it to NIRD for greater stability.

Internal server error

Bad User Info response, error= {"error_description":"Invalid access token: eyJraWQiOiJyc2ExliwidHlwljoiSldUliwiYWxnljoiUlMyl RfG_ru5Dr5mR4f0FWFoeXDjjD4XZzRpXcMjam3Jcm-QlUmmY5klhpUjAl4buP1PQlYCKPpBaG-F-FcKhmjpK2A4NJe2Sx5Y7Gi4s8DAdid1tQrbd2UqJRF eMTfQ60dvunCorVHmOllu251GobR0RsHakw","error

×