1

Log in

Σ MAS

📄 Summary: NEW UniMus:Natur

This application is valid and may be submitted. When you have submitted the application it can no longer be changed.

Submit this application

📄 1 Project Description and Funding

Project title	UniMus:Natur
Project short name	unimus-natur
Project description (summary)	The UniMus-Natur project is migrating the Natural History Museum's legacy dataset (MUSIT) to the Specify 7 platform. This transition is vital for securing millions of natural history records and maintaining their availability for international biodiversity research. We request NIRD resources to host a fully containerized staging environment orchestrated via Kubernetes. This validated environment consists of a live Specify 7 web instance, worker nodes, and MariaDB databases running as Docker microservices. Our methodology involves an ETL pipeline where data is

1

Log in

Σ MAS
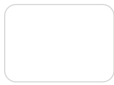
critical; it enables us to verify the migration results through the actual user interface and API in a controlled Kubernetes environment. The requested resources will drive this containerized stack and support the computationally intensive, iterative processes required to transform, clean, and validate the complete dataset before production release.

Primary field of science.	Biosciences
Secondary field(s) of science.	Geosciences
Research Council projects	N/A
Other funding sources	Funded by BOTT Universities (Bergen, Oslo, Trondheim, Tromsø)
Does any part of your project contain commercial funding?	False
Commercial funding ratio (%)	0
I confirm that I have acknowledged / I will acknowledge the use of the national e-infrastructure in our publications, and that my list of publications in	True

Log in

Σ MAS

publication list explanation Projects that use Sigma2 resources are invoiced according to the [contribution model](#).	N/A
I confirm that my project is obliged to comply with the contribution model. This may incur payment for the use of resources.	yes
Cost center (kostnadssted)	28001500
Internal activity number / string	106353100
I confirm that I have read and understood our Acceptable Use Policy (AUP) & MAS Privacy Notice	True

📄 2 Usage and output

Relevant usage experience	Our team has extensive specialized experience with NIRD resources through the GBIF Norway project (NS8095K), where we manage scalable services using Kubernetes and Minio object

1

Log in

Σ MAS
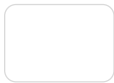
Measurable output

Specify 7 staging environment.
The primary output is the secured continuity and enhanced accessibility of over 15 million natural history records, which serve as the empirical foundation for numerous Master's and PhD theses at the BOTT universities annually. The museum's data generated more than 1200 peer-reviewed publications in 2025 alone (tracked by GBIF), a high-volume output we expect to maintain or exceed by modernizing to the Specify 7 platform. This project ensures continued global exploitation and dissemination through direct integration with the Norwegian Species Map Service (Artskart) and GBIF, safeguarding the data's compatibility with international standards for future research, conservation, and policy-making.

📄 3 Storage Resources

	2025.2	2026.1	2026.2	2027.1	2027.2
DL	2	2	2	2	2
DP					
Storage request justification	We request 2 TiB of project storage on NIRD to support the migration of the MUSIT database and potential future attachment of media assets.				
Do you need backup?	yes				
Report underusage	New application				

📄 4 Project characteristics

1

Log in

Σ MAS

Estimated data volume and technical requirements
Access level requirements
Access policies

📄5 Simplified Data Management Plan

DMP file	None
DMP metadata	
Link to DMP	https://github.com/UniMus-Natur/migration/raw/refs/heads/main/docs/documents/data-management-plan.docx
Valid start	2026-01-20
End for project	2036-01-01
Data management after project completion	Data will be published and archived
Most used file formats	['csv', 'json']
Published datasets related to project	All university museum collections https://www.unimus.no/nedlasting/datasett/naturhistorie/
Size of input dataset	2-10GB
Will you use output data as input	no
I/O pattern.	
Operations per second (IO/s)	100-1000

📄6 Access and Data policy

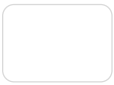|	
Data management plan	plan-exists
No plan needed explanation	

This application is valid and may be submitted. When you have

1

Log in

Σ MAS

Σ Sigma2