Note on data repository structure supporting the DST optimization pipeline

April 2026

Deliverable 1.1a

Project name	Embedding the new nitrogen regulation in farmer and advisor decision support
Project acronym	PlantPerform
Project ID	NNF26SA0115432
Project duration	April 2026 – March 2029
Project Coordinator	Innovation Centre for Organic Farming (ICOEL)
Project website	https://icoel.dk/om-os/projekter/novo-nordisk-fonden/2026/plantperform/
Work package	WP1
Work package leader	Tommy Dalgaard
Deliverable number	D1.1a
Deliverable title	Note on data repository structure supporting the DST optimization pipeline
Authors	Mette Vestergaard Odgaard, Morten Birk, Anton Rasmussen, Morten Winther Vestenaa, Anders Gjern, Kenneth Hansen, Morten Graversgaard, Tommy Dalgaard et al.
Version	1
Dissemination level	PU - Public

Table of Contents

1. Background - Context and rationale 4
2. Definition of data repository 4
3. Overall data flow as from March to June 2026 4
4. Raw data and data from existing models 5
   4.1 Raw data 6
   4.2. Data from models 7
5. Data processing – pipeline 7
6. Data repository and maintenance 8
7. Next steps 8
8. References 8

1. Background - Context and rationale

The following document illustrates the data repository structure of data included in development of the Decision Support Tool (DST) as from March to June 2026, and is therefore the first of two deliverables. Hence, new data repositories inputs can be added and followed by new data-processing which will result in an updated PlantPerform data-repository. An update follows in June 2027.

The project "PlantPerform" sets out to further develop an existing DST developed in the EU project Trans4Num (Trans4Num). The landscape-scale tool shows the most optimal placement of a given nature-based solution (in this case more grass in the rotation) in a geographical space given existing geographical variations and constraints where the latter is set by the user. In PlantPerform the DST will be made relevant at farm scale while also including information from the landscape scale. The DST tool will be showcased with multiple agricultural consultants though workshops and courses developed in PlantPerform. After the project the tool will be used in agricultural planning.

2. Definition of data repository

A data repository is in the literature defined as being a collection of data including detailed metadata. These repositories span from being relatively small project-based to institutional (such as ERDA for Aarhus University) and finally more global repositories such as Dataverse, FigShare, Dryas, Zenodo etc.. The repository can function as storage of data or as a pipeline to ease access. Hence, any collection of data is in principle a data repository, and in the life cycle of data creation various repositories can be linked beginning from a collection of raw data to the final ready data. This can be referred to as bronze, silver, and gold data were bronze is identified as being rather raw including duplicates and other errors, silver is defined as more polished data, and finally, gold gives data ready for use in specific work/research.

The large repositories tend to be less detailed and do not streamline or maintain data. Hence, these can end up conflicting with the FAIR data principles by not being Findable, Accessible, Interoperable, and Reusable (EU commission) and end up with an overall low usability (Wilkinson et al., 2016).

In this project, a data repository is built up holding the scripted code to build the tool and metadata for all input data (as appendix H from the Final Application PlantPerform). The repository will be hosted by ICOEL and will function as a smaller project-based repository containing meta data belonging to the final polished data ready to feed directly into the DST. The life cycle of the data belonging to PlantPerform will be elaborated in this document.

3. Overall data flow as from March to June 2026

Data used to calculate data input to the DST includes raw data and data outputs from existing models (Figure 1). The raw data is hosted by various owners in separate repositories (Figure 1a) before being processed (Figure 1b) and finally entering the Data input repository (Figure 1c). In the Data input repository data is structured in: 1) transition tables, 2) predefined crop rotation based on real data, and 3) a geographic table. Finally, the data in the Data input repository can be loaded to the DST backend and used in the DST (Figure 1d) and finally displayed in the PlantPerform Data repository (Figure 1e). The processing and final data structure of the data in the PlantPerform data repository will be elaborated on in section 6 and 7 regarding data processing and repository.

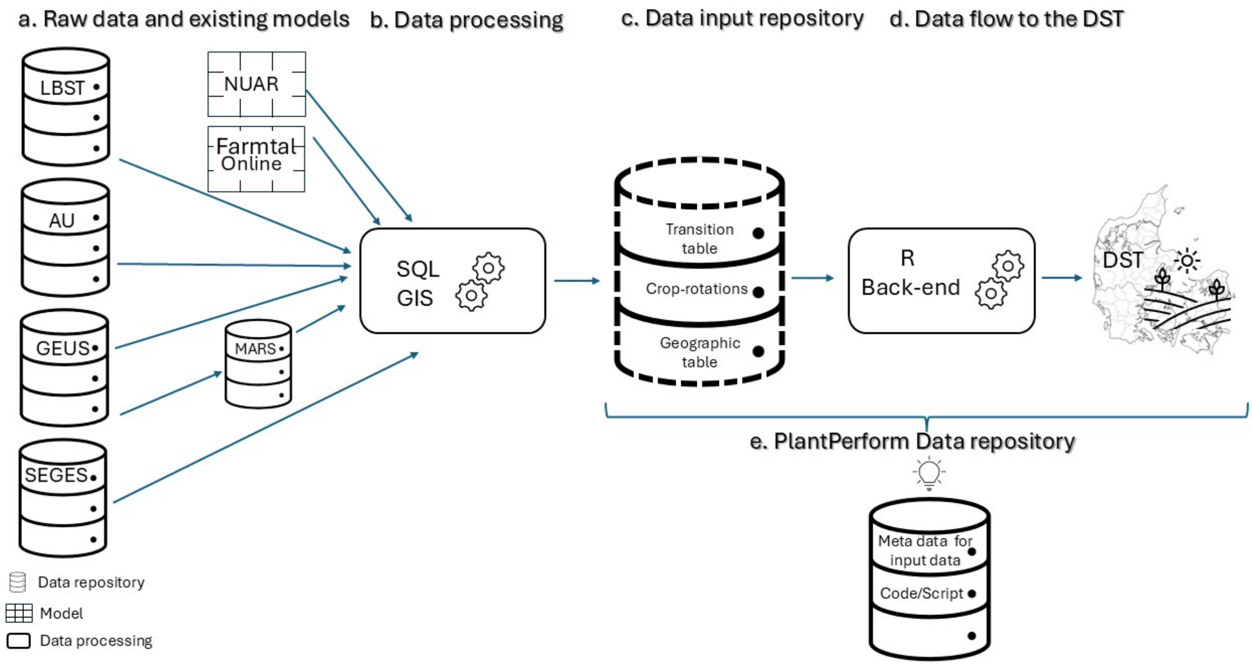

Figure 1: Illustration of data flow and processing from repositories to the DST. a) shows relevant data repositories and existing models, b) shows data processing, c) shows the Data input repository including transition tables, crop rotations, and geographic tables, d) Illustrates the coding of the DST and the final usage in the DST, and e) illustrates the Data repository in PlantPerform. The dashed line in *c) Data input repository* indicated that this is a gathering of various data not completely in a defined repository yet. This will be further elaborated in D1.1b.

4. Raw data and data from existing models

The data used to create the input data to the DST are divided into <u>raw data</u> and <u>data outputs form models</u> (Figure 1a and Table 1). These data have various usages in the final DST (Table 1).

Table 1: Data used to create the DST including raw data and data outputs from models. The columns indicate data repository or model and how the data is used in relation to the DST. Underline text highlights the three data structures in the data input repository (see also Figure 1).

Data Aggregated to field scale	Data repository or model	Usage of data in DST					
		Geo. unit	Transition table - unique geographic identifier	Transition table - Direct effect of measure	Crop rotation	Geographic table - Spatial effect of measure	
							Model constraint
Raw data							
Fields and crop code	LBST	x	x		x		x
Clay and sand	AU		x	x		x	
N retention	GEUS					x	x
HNV	AU						x
Transition plans	MARS						x
Biogas and biogasplants	AU						x
Data outputs from models							
N Leaching	NUAR			x			
Yield	FarmTal Online			x			x
Field economy	ICOEL			x			x

4.1 Raw data

Field data including crop details origin from the Danish farming agency (LBST 2025, LBST data repository (Figure 1a and Table 1)). The field data (years 2011-2025) has a national extent with field boundaries as the spatial unit. It includes information on yearly field size, crop type, and owner ship indicated by CVR number. Field data was - in combination with clay and sand - used as geographical unit, to build the unique geographic identifier in the transition table, and as model constraint.

Clay and sand originate from the Danish soil maps (Adhikari et al., 2013, AU data repository, (Figure 1a and Table 1)). The soil maps have a 10×10m resolution and Danish extent, and each show percentage of sand, clay, humus, and silt, and have been combined into 11 soil classes (Adhikari et al., 2013) where classes 1-4 refer to sand and 5-10 refer to clay (class 11 is humus). The soil map were aggregated to field scale giving the majority soil-class reclassified into sand and clay. Clay and sand were - in combination with field data - used to build the unique geographic identifier in the transition table, in calculation the effect, and indirectly as spatial effect of measure.

The Nitrogen (N) retention map in 100×100m resolution and Danish extent (Højberg et al., 2025, GEUS data repository (Figure 1a and Table 1)) defines the amount of N being denitrified from the agricultural rootzone to the coast ($NO_3^- \rightarrow N_2 + N_2O$) in percentage. In combination with N leaching from an agricultural field to the rootzone and the overall coastal N-reduction goal, the N retention/N load can be interpreted as N vulnerability. The retention map was aggregated to field scale giving the mean N retention of a field. N retention was used to calculate the spatial effect of the measure (N load = N leaching *(100-N retention)/100) and as model constraint (Table 1).

HNV (High Nature Value) is used as model constraint (Ejrnæs et al., 2012, AU data repository, (Figure 1a and Table 1)). The map is in 10x10m resolution with values from 0-14 where 14 indicate high nature value based on landscapes, habitats, Farm management, species information. The idea is that there should not be agriculture on fields with an HNV >5.

Transition plans are defined as areas selected to be takes out of production by the green tripartite and the Land Use Transition Consultant (ustagningskonsulenter) (MARS 2026, MARS data repository, (Figure 1a and Table 1)). The transition plans were downloaded from MARS April 2026 and used as model constraint where a field were assigned 1 if it is part of a transition plan and otherwise 0. These planes change ongoing and needs to be re-downloaded for new modeling.

Biogas and biorefinery indicate calculated distance to biogas or biorefinery and is used as model constraint (Table 1). Two point data sets and a detailed line road data-set was used to measure distances from field center to nearest Biogas and biorefinery. Road data included information on one ways direction, bridges and motorway exists. Ack GIS network analyst was used (ESRI 2024).

4.2. Data from models

Modelling data coming from already existing models were mainly used to calculate the effect of a given change in management practice (Figure 1ab). These were, for now, N leaching, economy, and yield.

N Leaching was based on NUAR (Ny Uledningsbaseret ArealRegulering for kvælstof) (Eriksen et al., 2024). Which is a model build in excel drawing on various constants and norm values and relevant climatic grids.

Yield is currently being developed but will be drawing on data from Farmtal Online (https://farmtalonline.dlbr.dk/Navigation/NavigationTree.aspx).

Field economy is currently being developed but will be drawing on data from Farmtal Online (https://farmtalonline.dlbr.dk/Navigation/NavigationTree.aspx). In the current version Economy is drawn from NAUR outputs and ICOEL expert knowledge.

5. Data processing – pipeline

The various data from the existing data repositories and outputs form existing models (Table 1, Figure 1a) is processed using SQL and GIS (Figure 1b) and result in transition tables, defined crop-rotations, and geographic table (Table 1, Figure 1c) which is finally processed using R and thereby used in the DST (Figure 1d), where the transition table illustrates the effect of transition from one crop to another (effects of measure), the crop rotation defines the order of crops, and the geographic table illustrates the spatial effect of the measure due to spatial variability in the geophysical variables. Finally, data is issued in the PlantPerform Data repository (Figure 1e).

6. Data repository and maintenance

All metadata presenting the processed data (transition tables, crop rotations, and the geographic table) (Figure 1c) and the code/Script of the DST (Figure 1d) will be issued in the PlantPerform Data repository (Figure 1e). This repository will be centralized and controlled by the project to prevent irrelevance and risk of conflicting with the FAIR data principles (Wikinson et al., 2016). The Repository will be hosted by ICOEL and is displayed on the project homepage with open access. The DST tool can draw directly from the Data input repository which is currently hosted by various data owners and not centralized at any specific host (Figure 1c). This avoids large data repositories which tend to be less detailed and do not streamline or maintain data (Wilkinson et al., 2016). Hence, only metadata and the code is stored in the PlantPerform data repository (Figure 1e).

7. Next steps

In the next project phase from August 2026 to June 2027 work will be made regarding process automation of data flow from models to the tool. Hence, the Data input repository will be refined (Figure 1c). Moreover, based on feedback from the farmer-advisor meetings, new features may be added to the tool. This will probably not change the overall data repository structure presented in this document in short being: A Data input repository feed the DST and final metadata and scripts are stored in the PlantPerform data repository hosted by ICOEL.

8. References

Adhikari, K., Bou Kheir, R., Greve, M.B., Bøcher, P.K., Malone, B.P., Minasny, B., McBratney, A.B. and Greve, M.H., 2013. *High-Resolution 3-D Mapping of Soil Texture in Denmark. Soil Science Society of America Journal*, 77(3), pp.860–876. https://doi.org/10.2136/sssaj2012.0275

Ejrnæs R, Skov F, Bladt, J, Fredshavn J, Nygaard B. 2012. Udvikling af en high nature value (HNV) indikator. Videnskabelig rapport fra DCE – Nationalt Center for Miljø og Energi. Raport number 40.

Eriksen, J. (redaktør), Ørum, J.E., Hvid, S.K., Olsen, J.V., Hansen, E.M., Larsen, S.U., Giannini-Kurina, F., Thomsen, I.K., 2024. Ny Udledningsbaseret Arealregulering for kvælstof (NUAR) – analyse af metode, kvotetildelingsmodeller og omkostninger. 116 sider. Rådgivningsnotat fra DCA – Nationalt Center for Fødevarer og Jordbrug, Aarhus Universitet Leveret: 24.05.2024.
ESRI. 2024. ArcGis Pro. Ver. 3.4.3

Højberg, Anker Lajer; Thodsen, Hans; Børgesen, Christen Duus, 2025, "National nitrogen retention maps for Denmark", https://doi.org/10.22008/FK2/O4XLAQ, GEUS Dataverse, V1

LBST 2025 https://landbrugsgeodata.fvm.dk/

MARS - Multifunktionel Areal Registreringssystem - https://mars.sgav.dk/
SGAV Styrelsen for Grøn Arealomlægning og Vandmiljø.

PlantPerform 2026 Homepage

Trans4num project website - https://trans4num.eu/da/

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., *et al.*, 2016. *The FAIR Guiding Principles for scientific data management and stewardship*. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18