V.V. Dokuchaev
Soil Science Institute

Formalized data and models descriptions for the development of soil digital twins framework

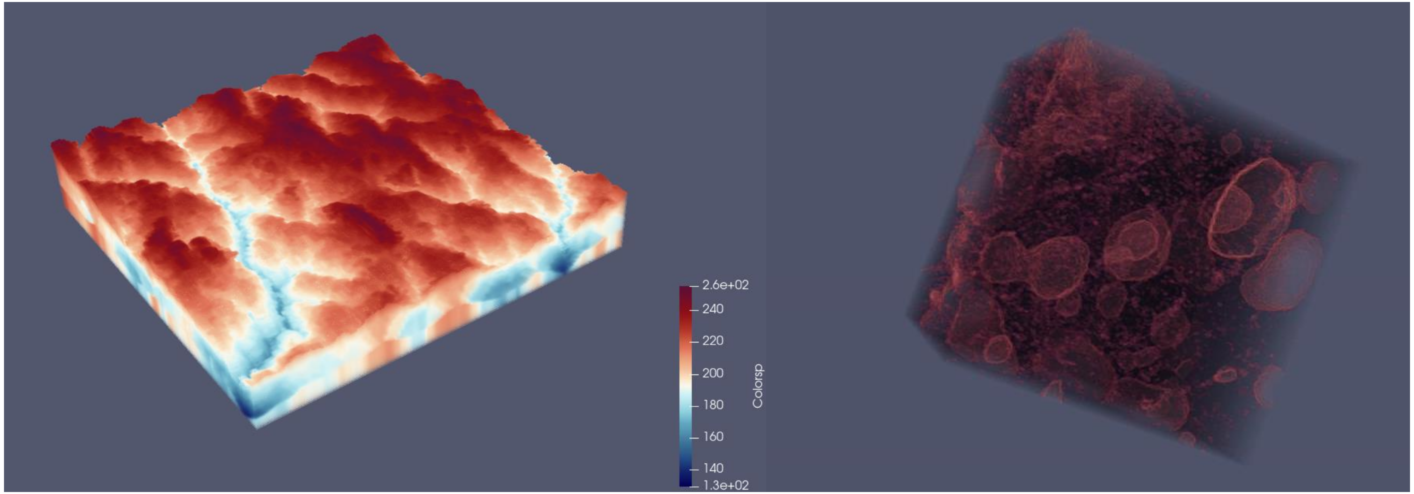
Vasilyeva N.A.*, Vladimirov A.A., Vasiliev T.A., Bolotov A.G., Ivanov A.L.
Dokuchaev Soil Science Institute, Moscow, Russian Federation

Global Symposium on Soil Information and Data
September 25, 2024 | Nanjing, China

- **Digital twin** – **is a virtual representation of an object designed to reflect a physical object accurately.**

  In our case soil digital twin (SDT) is a landscape-scale soil body down to its micro-structures

- Distinct technologies involved in digital twins :
  - Real-time data assimilation from sensors to correct DT real-time state, and
  - NN-acceleration of process-based models for fast simulation of future scenarios
  - 3D visualization analysis



The Soil Digital Twin project

V.V. Dokuchaev
Soil Science Institute

3 year project

Aim – automated data and models dynamic integration for solving various soil problems using digital twins.

We are developing a SDT framework capable of generating different problem-specific SDT.

Development stages:

- 2024: SDT-0 is the information system for automated data and models integration
- 2025: SDT-1 gets real-time data from sensors and represents physical twin current state using data assimilation
- 2026: SDT-2 is a simulator for future scenarios, implemented as a Cellular Automata driven by neural networks (Neural-CA)

Further on, SDT-3 generation should have self-analysis features empowering the system to propose experimental plans that address uncertainties in its outcomes.

SDT General Workflow: Initialization in a chosen location

- A task for SDT is chosen from a library of formally described tasks

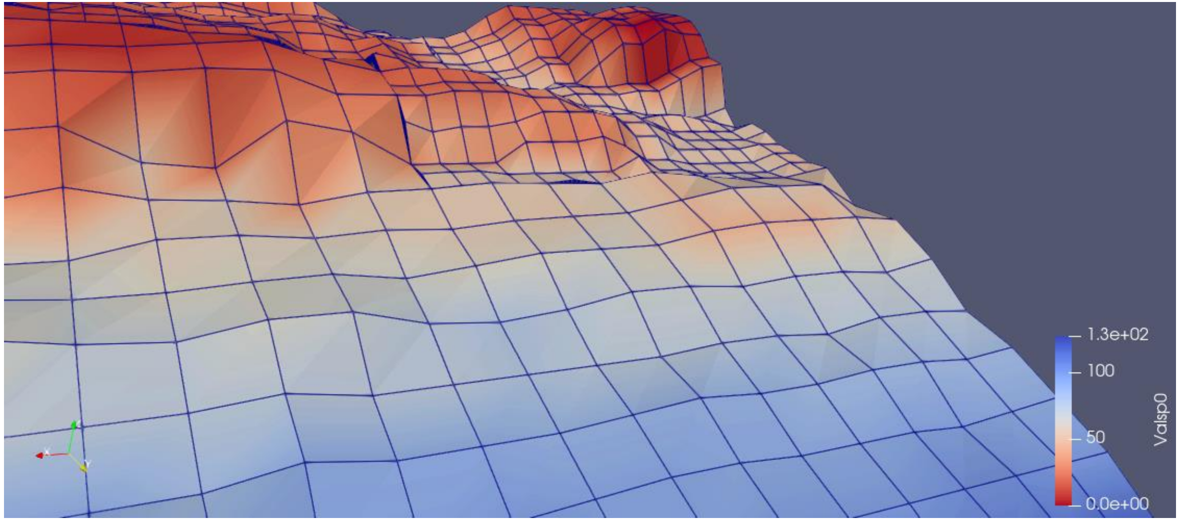
    Formal description of a task for SDT will contain such information as:

        Minimal required accuracy of the results

        Time and space resolutions, etc.

- Simulation mesh is then automatically constructed using all available DEMs and other properties

Example of a layered quadtree construction based on SRTM and available LiDAR DEM (1m)

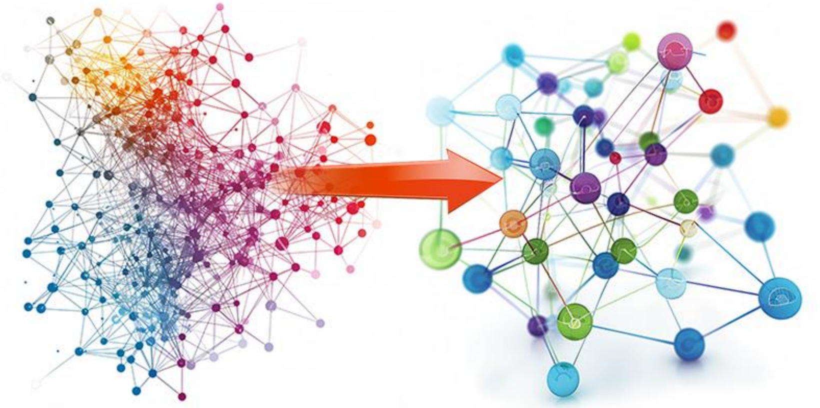


SDT General Workflow: finding solutions for the problem

SDT Libraries can be considered as a graph with relations through input/output variables of its tasks, data and models

- Graph Analysis: SDT evaluates the graph, constructing subgraphs of potential solutions

- Gap Identification: If no solutions are found, SDT recommends necessary missing data or models for specific variables needed to build a solution subgraph.



For automation of this process are required:

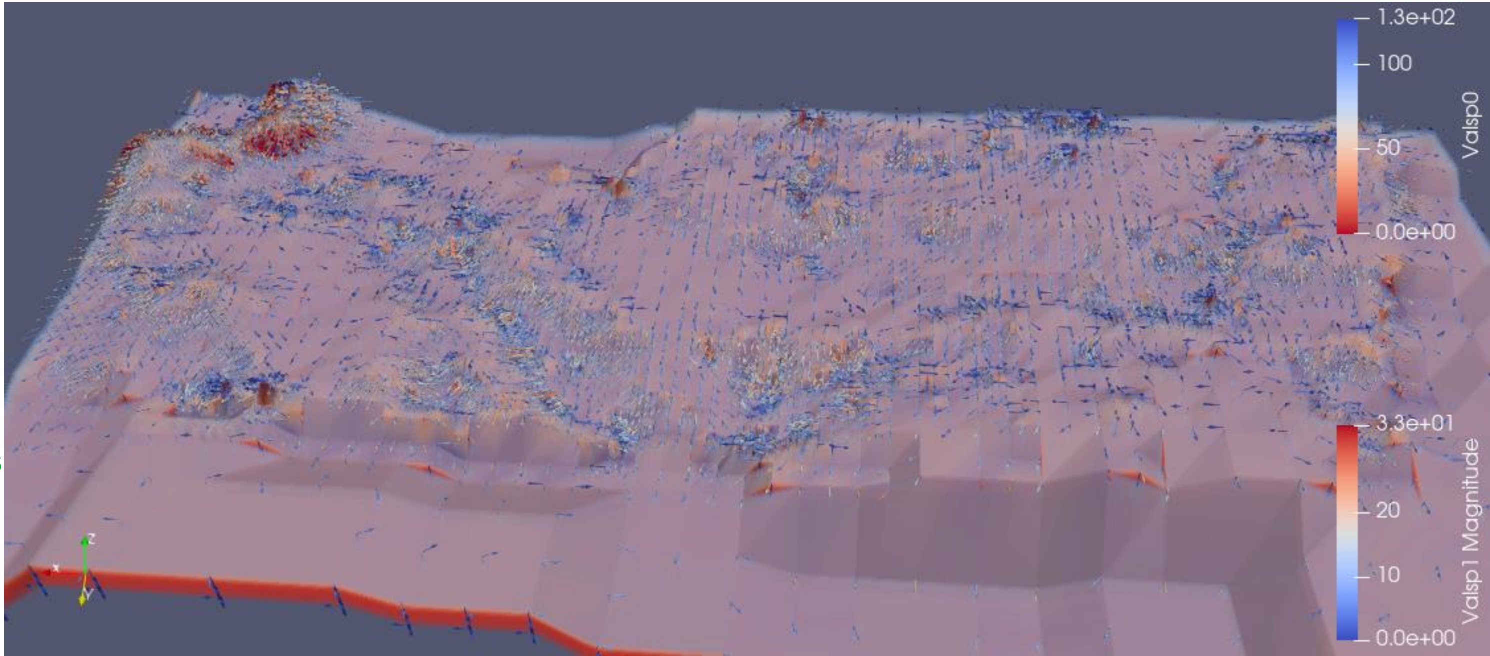
- **formal descriptions of data and models** (including input/output variables, uncertainties, applicability ranges, etc.), and
- **a common thesaurus for variable names**.

SDT General Workflow: when all resources are provided SDT runs

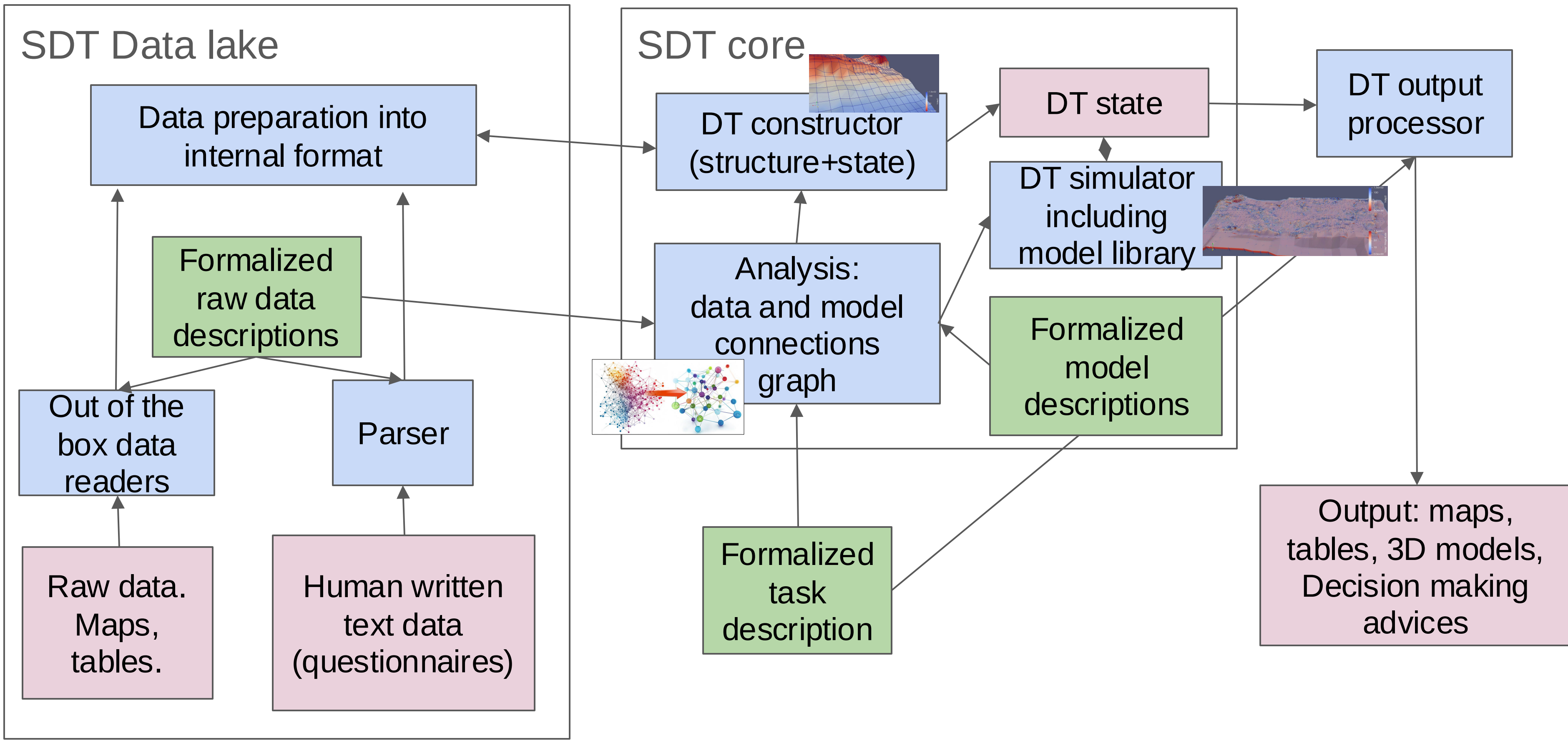
Example of water redistribution and infiltration simulation by the CA,
where state variables in SDT (Neural-CA) are supposed to be updated by neural networks (physically-informed or physically aware NNs trained on the results of various process-based models).

Training of NNs on simulation results of different models also **requires automation** of:
- data preparation,
- scenario generation,
- simulation runs.

Such automation **becomes possible with formalized descriptions** of data and models.



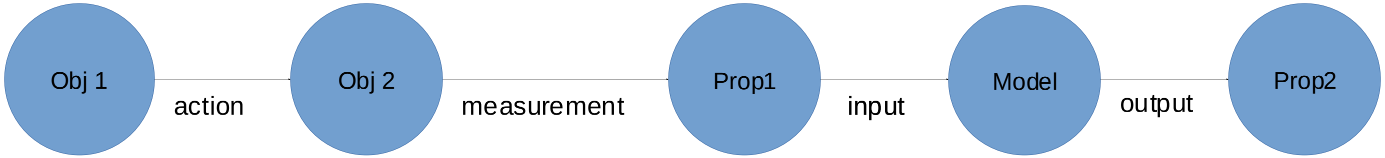
SDT data flows

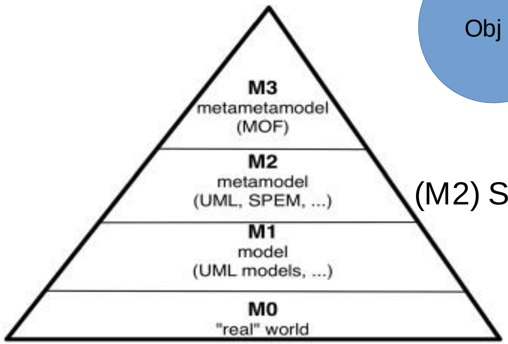


Model-Driven Architecture (MDA) for formal descriptions

Knowledge graph, where object types and properties are represented as nodes, and actions, such as measurements or sampling are represented as edges.

(M3) Ontological model





Model hierarchy in MDA

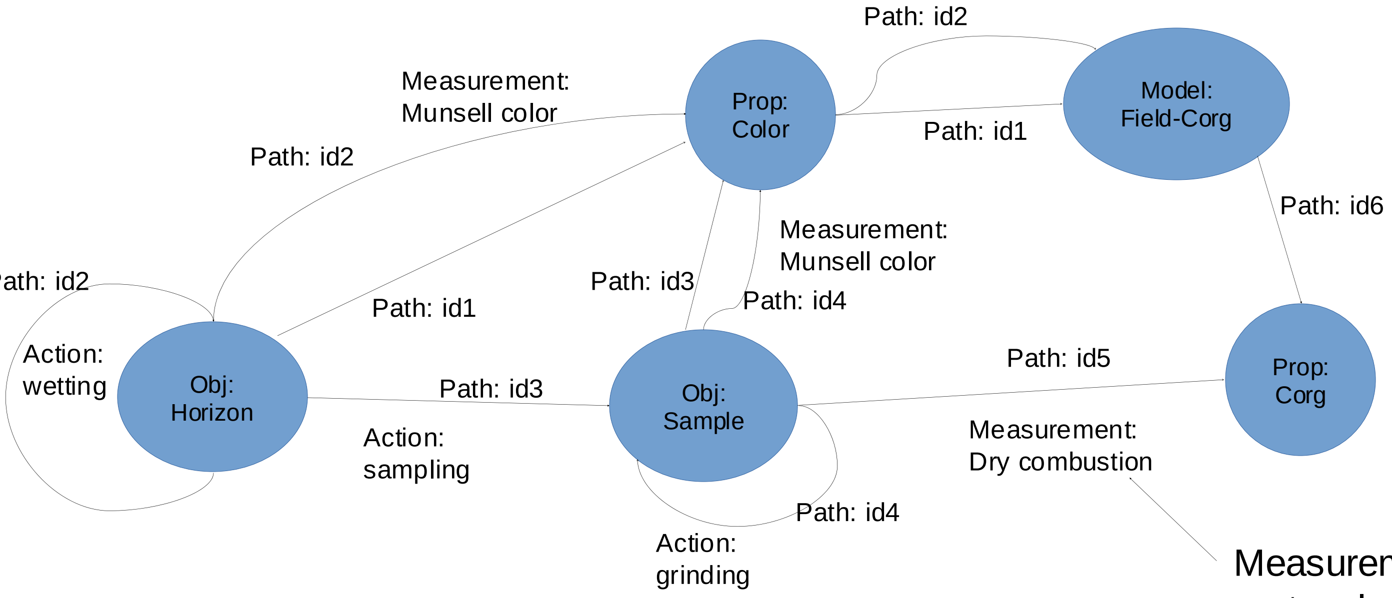
(M2) Structure of entities, common thesaurus

Each data element then is associated with a graph path connecting object with property (for example, carbon content of a specific soil fraction obtained by specific method from a soil horizon)

All object nodes and unique paths to properties of the graph get their ID and **automatically comprise a growing common thesaurus** for:

- object types
- actions (including all soil preparation protocol), and
- soil properties (including all its measurement protocol)

(M2) Illustrative knowledge graph



Measurement protocols can be inserted in any level detalization

- Thank you for your attention!

We acknowledge the Ministry of Science and Education of Russian Federation Project grant agreement No.075-15-2024-545 of 24.04.2024