Rümmler, Arne
Chair of Geoinformatics, TU Dresden

How to evaluate and reduce the complexity of geospatial provenance graphs?

01.09.2020

PROVENANCE

- Provenance describes the origination history of data

PROVENANCE

- Provenance describes the origination history of data
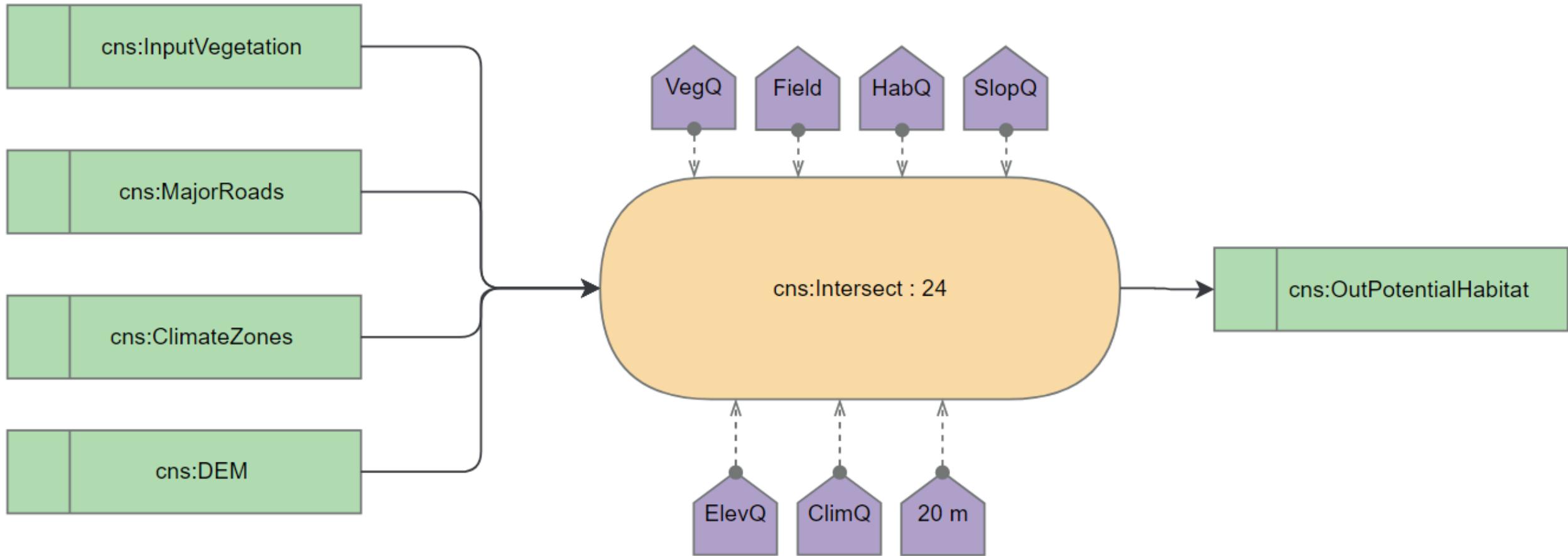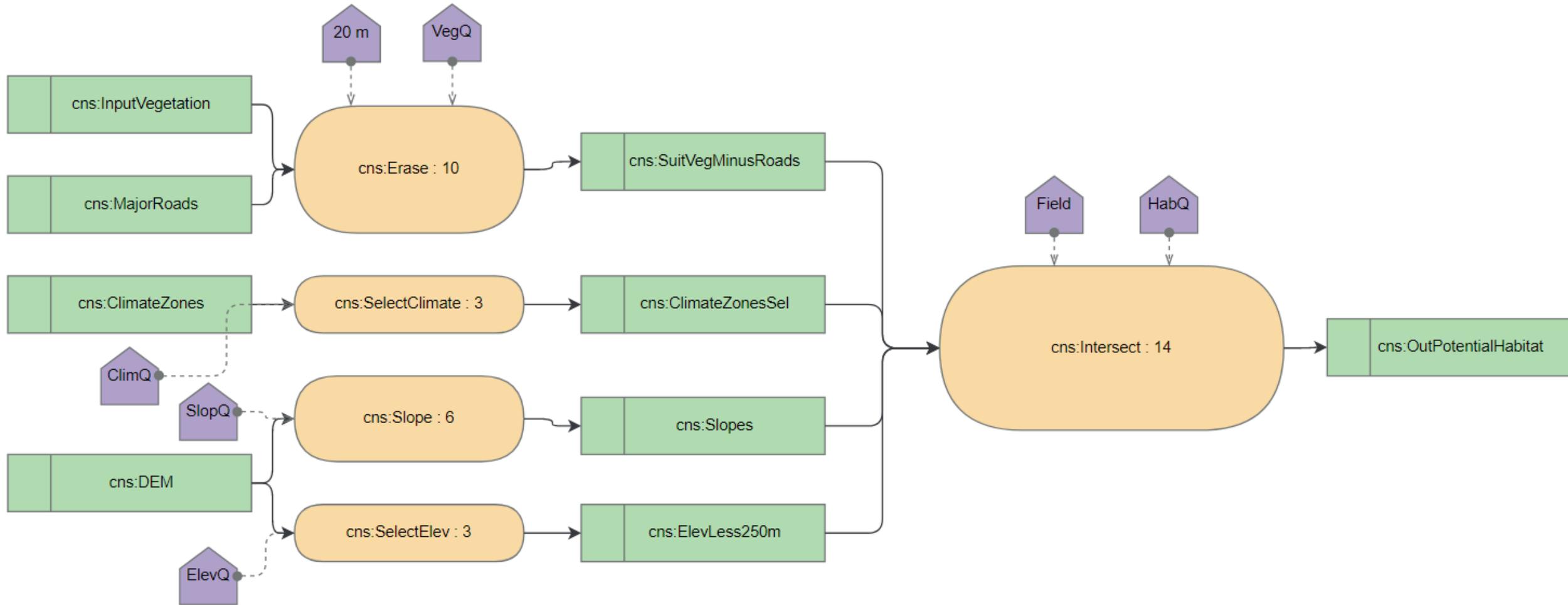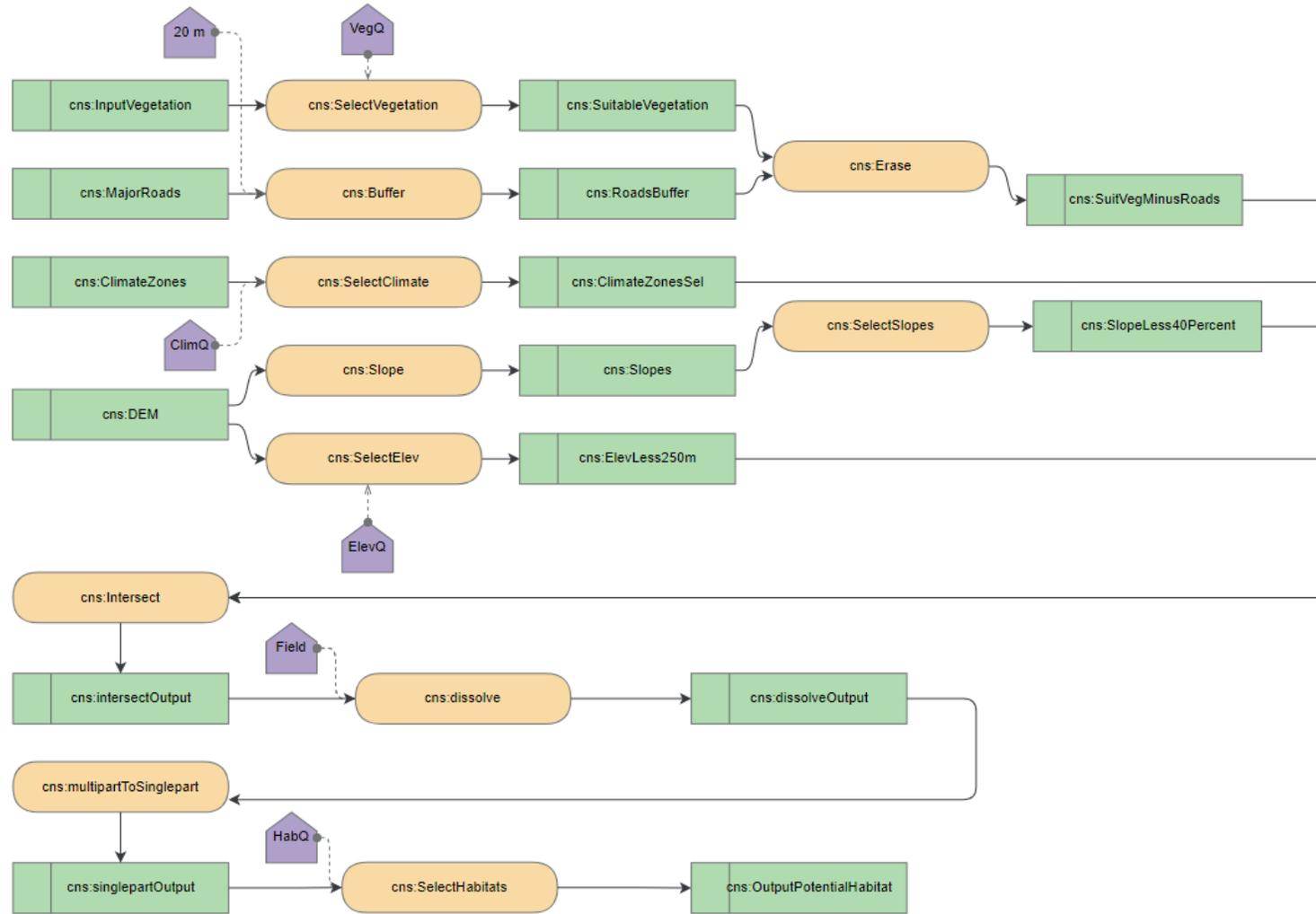- Workflow graphs are a widely known form of provenance

PROVENANCE

- Provenance describes the origination history of data

- Workflow graphs are a widely known form of provenance

- Provenance data can be standardized (e.g. PROV-O)

PROVENANCE

- Provenance describes the origination history of data
- Workflow graphs are a widely known form of provenance
- Provenance data can be standardized (e.g. PROV-O)
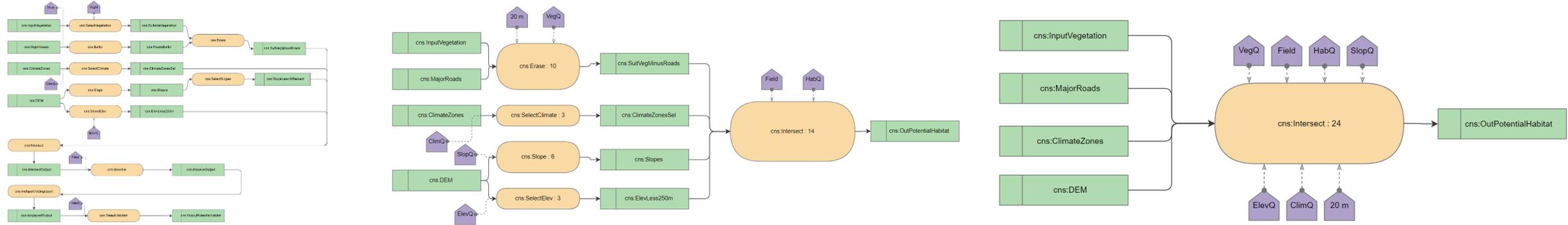- Standardized provenance graphs can be linked

PROVENANCE

- Provenance describes the origination history of data

- Workflow graphs are a widely known form of provenance

- Provenance data can be standardized (e.g. PROV-O)

- Standardized provenance graphs can be linked

- Provenance graphs can become indefinitely large

PROVENANCE

- Provenance describes the origination history of data
- Workflow graphs are a widely known form of provenance
- Provenance data can be standardized (e.g. PROV-O)
- Standardized provenance graphs can be linked
- Provenance graphs can become indefinitely large

How to make use of such provenance data?

Leverage standardized provenance data to enhance the understandability of the origination history of geospatial data.

How to get from the left to the right?

-> **Process-significance based generalization**

Process complexity analysis by graph pattern

- 2 Inputs, 3 Parameter, 1 Output -> Complexity **C = 2 + 3 + 1 = 6**
- Alternative complexity calculation possible, e.g. sqrt(params)
- Further patterns can be included

Prepare	Analyze	Model
▪ Selection ▪ Transformation ▪ Reclassification ▪ Geometric Computation ▪ Visualization	▪ Neighborhood Analysis ▪ Map Overlay ▪ Network ▪ Statistics	▪ Model (back-box-modules, neural networks, …)
1	2	3

- Subcategories derived from Albrecht (Albrecht, J. Universal analytical GIS operations: a task-oriented systematization of data structure-independent GIS functionality 1998.)
- Value serves a multiplier