



The HUMBOLDT Workflow Design and Construction Service

Daniel Fitzner (FHG-IGD)

Outline

Workflows in HUMBOLDT

- The structure of HUMBOLDT Workflows (graphs? trees? workflow patterns?)

The HUMBOLDT Workflow Service (WDCS)

- Purpose and Scope
- Stack of functionality
- Current Status → Demo
- Questions to the scenarios

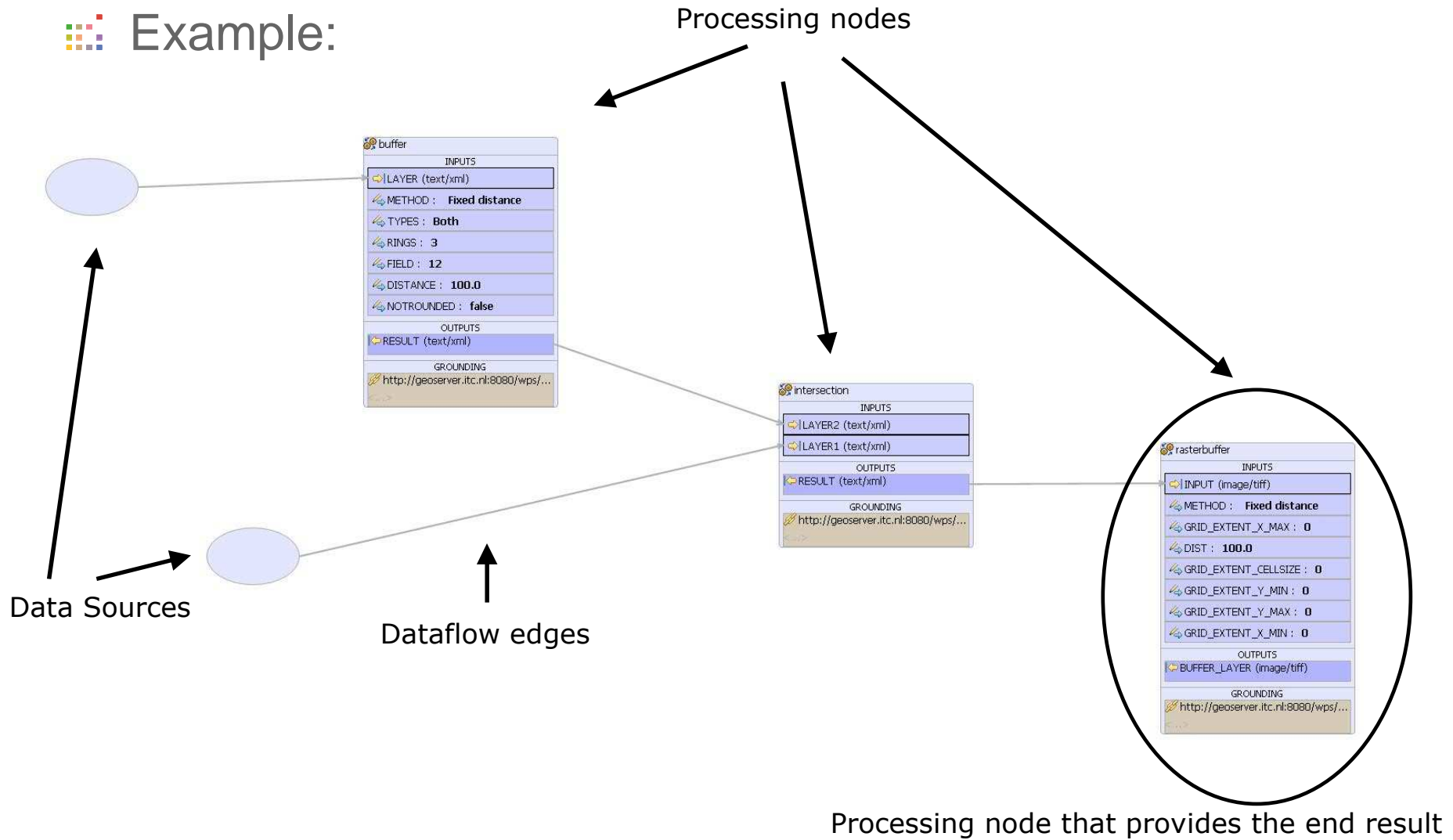
■ The structure of HUMBOLDT workflows

□ In HUMBOLDT, workflows are not arbitrary graph structures, but:

acyclic and **directed** graphs with a single node with no outgoing edge

- the nodes with no incoming edges are **data sources**
- the nodes with incoming and outgoing edges are **processing components**
- The **directed edges** represent the **dataflow** between nodes
- Finally, the node with no outgoing edge is the processing node that will provide the **end result**

Example:



Service-based Harmonisation in HUMBOLDT

- ▣ Harmonisation in HUMBOLDT = Applying a chain of geoprocessing components (a workflow graph) to source datasets
- ▣ **Fully manual** → GIS based, a human passes around the results → Out of the scope of HUMBOLDT
- ▣ **Manual Workflow Definition, Automated Execution:** A human designs the workflow graph manually → Only the execution is automated
- ▣ **Automated Harmonization:** Building the workflow graph automatically
 - **Inputs** (to the WDCS):
 - A **set of constraints** on a geospatial dataset (i.e. the HUMBOLDT Context)
 - **Process Metadata** for a set of processes (e.g. WPS DescribeProcess Responses + some additional semantic metadata)
 - **Data Source Metadata** for a set of data sources
 - **Output** (of the WDCS):
 - A **workflow graph** whose execution (in the MS) results in a dataset satisfying the set of constraints

Basic Functionality of the WDCCS

- Allow users to visually compose the workflow graph out of geoprocessing components (out of WPS / java processes / WSDL?) and data sources →

Manual Workflow Definition, Automated Execution

- Allow users to register processes to the system
- Exports such workflows in different workflow dialects via a WSDL / SOAP interface (priority: the dialect used by the HUMBOLDT Mediator Service)
- Help the user in the composition process, by:
 - type checking inputs and outputs when connecting
 - type checking user entered input values
 - Some small graphical features that make the composition process easier

Long-term functionality

- ▣ Allow users to just specify **thematic constraints** (or data source discovery queries) as input nodes (input nodes that do not contain a binding to concrete data sources, e.g. WFS) and dynamically discover them → **HUMBOLDT Basic Workflows**
- ▣ Allow users registering processing components to formalize **cross-input** and **cross input-output** parameter constraints → ~Formalized GIS Environments
- ▣ Perform **automated harmonisation**, i.e. building the workflow graph automatically for „harmonization“ processing components

■ Status – What is already implemented

- Registering WPS offline (provide WPS URLs in the source code)
- **Manual** and **Visual** Workflow Definition and XML-Export to WDCS
Workflow Model
- Simple type checking based on WPS Process Descriptions
- Small graphical features that support the user in the composition
process
- Some issues of the long term functionality

Next Steps

- Clarify the appearance of workflows in the scenarios
- Implement:
 1. Registration (+ annotation) GUI for processes (WPS, java)
 2. Connection to data source catalogue (i.e. the HUMBOLDT IGS) + discovery GUI for data sources
 3. Export to workflow execution model (Mediator Service Model) via a WSDL / SOAP interface
 4. Implement long-term functionality (connect all things already implemented + implement additional functionality)

Questions to be answered by the scenarios

What are the processing components in your workflows?

- Only services provided by HUMBOLDT or some specific processing?
- If specific: WPS, WSDL / SOAP, java code, something else?

How does the workflow graph look like?

- If we have an answer to the first question (+ some time) or only WPS, you can already use the HUMBOLDT Workflow GUI to specify your workflow graph

What are the data sources?

- Only WFS or some other service type / format?

Outlook

- ▣ In the upcoming weeks, we will provide a first installable version for playaround (the source code can already be downloaded from the community website)
- ▣ ...
- ▣ Thanks, any **questions** or **suggestions**?