Interactive Data Exploration for Geoscience

Christian Beilschmidt ¹, Johannes Drönner ¹, Michael Mattig ¹, Marco Schmidt ², Christian Authmann ¹, Aidin Niamir ², Thomas Hickler ²,³, Bernhard Seeger ¹

¹ University of Marburg, Database Research Group
² Senckenberg Biodiversity and Climate Research Centre (BiK-F)
³ Goethe University Frankfurt, Department of Physical Geography
Motivation
Introduction: Data-Driven Research

• Enabled by data availability
  – satellites, sensors and observations…
• Visualization often triggers new scientific ideas
Explorative Workflows

- Users explore data and various kinds of processing steps
  - Multiple paths are followed \(\rightarrow\) separate workflows
- Requirements:
  - Low latency visualizations
    - Data abstraction and generalization
  - Citations + reproducibility
Challenges

• Visualization
  – Big / heterogeneous data
  – Identify relevant data / quality issues

• Explorative workflows
  – Explorative usage
  – Data lineage / citations

• Time as an integral dimension
  – Data changes over time
  – Detect temporal patterns
Our Approach: VAT - Architecture Overview

**External Tools**

**Users**

**Visualization, Analysis & Transformation System**

**WAVE**

**Front End**

**Mapping**

**Processing Back End**

**Raster Data**

**Vector Data**
WAVE: VAT‘s Graphical User Interface
Visualization: Raster Data

- Parallel requested as tiles
- Aggregation to match map resolution
Visualization: Vector Data

• Polygons and lines: Simplification
• Points: Clustering
Visualization: Identify Relevant Data

- Non-overlapping clustering
  → Allows pattern recognition + data reduction
Visualization: Identify Relevant Data

- Linked table view:
  - Aggregated numeric attributes
  - Representative text attributes
Visualization: Quality Issues
Workflow: Combining Data
Workflow: Combining Data
(Explorative) Workflows: Lineage Tracking
Workflows: Citations

Data Table

<table>
<thead>
<tr>
<th>Citation</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>URI:</td>
<td><a href="http://www.gbif.org/dataset/562785/2.8405.4865.96d4-030a843311b0">http://www.gbif.org/dataset/562785/2.8405.4865.96d4-030a843311b0</a></td>
</tr>
<tr>
<td>Citation:</td>
<td>Western Australian Museum: Western Australian Museum provider for OZCAM</td>
</tr>
<tr>
<td>License:</td>
<td><a href="http://www.gbif.org/dataset/2c132d9e-dc6b-11de-822a-b3009c80f862">http://www.gbif.org/dataset/2c132d9e-dc6b-11de-822a-b3009c80f862</a></td>
</tr>
</tbody>
</table>

| License: | public domain |
| URI:     | http://www2.jpl.nasa.gov/srtm/ |

| License: | http://www.iucnredlist.org/ |
| URI:     | 16 |
Support for Temporal Operations

- All datasets are time-series
  - Individual temporal validity
- Operations create new time-series
Time as an Integral Dimension

January

<table>
<thead>
<tr>
<th>Data Table</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>basisofrecord</td>
</tr>
<tr>
<td></td>
<td>gbifid</td>
</tr>
<tr>
<td>29.1 ± 0.1</td>
<td>PRESERVED_SPECIMEN</td>
</tr>
<tr>
<td>28</td>
<td>PRESERVED_SPECIMEN</td>
</tr>
</tbody>
</table>
Time as an Integral Dimension

April

<table>
<thead>
<tr>
<th>Data Table</th>
<th>Citation</th>
<th>Temperature</th>
<th>basisofrecord</th>
<th>gbifid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>23.25 ± 0.0499992</td>
<td>PRESERVED_SPECIMEN</td>
<td>1.10123e+09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22.4</td>
<td>PRESERVED_SPECIMEN</td>
<td>1.06662e+09</td>
</tr>
</tbody>
</table>
Time as an Integral Dimension

July

<table>
<thead>
<tr>
<th>Data Table</th>
<th>Citation</th>
<th>Temperature</th>
<th>basisofrecord</th>
<th>gbifid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>14.9 ± 0.0999999</td>
<td>PRESERVED_SPECIMEN</td>
<td>1.10123e+09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.2</td>
<td>PRESERVED_SPECIMEN</td>
<td>1.06662e+09</td>
</tr>
</tbody>
</table>
User Evaluation

- Use-cases with 15 domain experts in biodiversity
  ➔ Overall positive feedback, minor changes
Conclusion / Future Work

• Our approach to interactive data exploration for geoscience
  – Low latency visualization
  – Explorative workflows
  – Time as an integral dimension

• Future Work
  – Adaptive applications / Application builder
  – Connectivity to external tools e.g. R
  – Support for complex time patterns / aggregation
    e.g. Average; Jan-Apr; 1995-2000;