Common spatial data problems in biodiversity data sets

Dr. Donovan Kirkwood

don@ecological.co.za  donovankirkwood
Unnecessary spatial complexity – NOT AN ERROR, but a problem for serving data
Example of typical source data misalignment relative to Demarcation Board Local Municipality boundary in a Biodiversity Assessment.
Examples of spurious sliver polygons in source data
Shifts + duplication
(not using snapping tools or coordinate system issues)
Overlapping polygons
Identical OR similar
- Often with important attribute data in each
(caused by merge instead of union, intersect or update operations)
Overlapping polygons
Identical OR similar
- Often with important attribute data in each

(caused by not using autocomplete polygon and/or merge instead of union, intersect or update operations)
Other kinds of dirty data
Data brought in from Google Earth
– twisted loops
- Automatic inclusion of Z & M data
DIGITISING
Inappropriate or inconsistent spatial accuracy and scale
Crude generalisation methods
Structures and process-reporting problems in data sets with external custodians
<table>
<thead>
<tr>
<th>ISSUE</th>
<th>PROBLEM</th>
<th>Tools to AVOID</th>
<th>Tools to FIX</th>
</tr>
</thead>
</table>
| Topology – overlaps, gaps (esp Cads) | • Area measurement  
    • Querying or processing attributes | • | • |
| Slivers (resulting processing, often coincident but not identical features) | • Inaccuracy, larger data sets but no extra info, visual perception of unreliability | • | • Mostly manual + |
| Shifts – slivers + gaps + overlaps | • Good workflow with coordinate systems.  
    • Visual checking  
    • Prevent unintended moves while editing by set sticky move tolerance | • | • Mostly manual |
| | • Projection | • | • |
| | • attributes too many obscure | • | • |
| | • archive source data properly | • | • |
| | | • | • |
WORKFLOWS?

• Set up data accuracy and production requirements BEFORE producing/digitising
• Training and standards for common issues
• Use topology tools and Identify/Repair Geometry tools on every data set before using/processing
• Resolve individual source layers BEFORE processing.