**Fig. A.2** Visualization of the cleaning procedure steps.



For the automated cleaning of the dataset, we used the Python package gdal and the QGIS API. The cleaning procedure contained the following steps:

1. Identify and delete features with duplicate geometry. To do so, we collected features with the same centroid and area and randomly chose one to remain in the dataset. The others were removed (gdal).
2. Check for invalid geometries and make them valid (gdal).
3. Remove features with no geometry (gdal).
4. Remove features with line geometries (gdal).
5. Remove loose lines that are not part of any polygon (QGIS API).
6. Identify overlapping fields, extract the overlapping sections, and assign the extracted sections randomly to one of the original overlapping fields. This was carried out in several steps:
	1. Identify overlaps of neighbouring fields, collect IDs of the fields that contributed to the overlap and save the overlaps and collected IDs to a new dataset (henceforth called the intersection dataset). Thus, each feature in the intersection dataset has a geometry representing the overlap and an attribute containing the IDs of the original overlapping fields (gdal).
	2. Because multiple fields can overlap, disassemble all overlaps into the contributing pieces by slicing it up into its component parts to create a new dataset containing the overlap geometries, which no longer overlap (QGIS API).
	3. Remove the overlaps from the original overlapping fields by calculating the difference between the original overlapping fields and the intersection datasets (QGIS API).
	4. Add all features of the intersection dataset to the original overlapping fields dataset (QGIS API). This dataset now contains the original overlapping fields with overlaps removed, as well as polygons that represent the former overlaps (overlap polygons). Each of these overlapping polygons was associated with a list of IDs, representing the connection to the original overlapping fields.
	5. Randomly assign one of the IDs to each overlap polygon (gdal).
	6. Based on this ID, each overlap polygon is merged with one of its original overlapping fields (QGIS API).