



Hochschule Karlsruhe
Technik und Wirtschaft
UNIVERSITY OF APPLIED SCIENCES

BACHELORARBEIT

IM STUDIENGANG KARTOGRAPHIE UND GEOMATIK

Spezifikation von Qualitätsanforderungen für Umweltdaten, Erstellung eines konzeptionellen Vorgehensmodells zur Fehleridentifikation und -verbesserung sowie Erstellung eines Prototyps zur topologischen, geometrischen und inhaltlichen Plausibilisierung von Geodaten des Räumlichen Informations- und Planungssystems (RIPS)

vorgelegt von

Jan Jedersberger

Matrikelnr.: 36410

eingereicht am

07. August 2015

Erstbetreuer: Prof. Dr. Bernhard Bürg

Zweitbetreuer: Prof. Dr. Detlef Günther-Diringer

Drittbetreuer: Dr. Florian Kost

Aufgabenblatt für die Bachelorarbeit

von

Jan Jedersberger

an der

HOCHSCHULE KARLSRUHE – TECHNIK UND WIRTSCHAFT
Fakultät für Informationsmanagement und Medien - Studiengang Kartographie und
Geomatik

in Zusammenarbeit mit der

LUBW – Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg

Thema: Spezifikation von Qualitätsanforderungen für Umweltdaten, Erstellung eines konzeptionellen Vorgehensmodells zur Fehleridentifikation und -verbesserung, sowie Erstellung eines Prototypen zur topologischen, geometrischen und inhaltlichen Plausibilisierung von Geodaten des Räumlichen Informations- und Planungssystems (RIPS)

Die Qualität von raumbezogenen Daten lässt sich nicht allein durch bloße optische Kontrollen feststellen. Kleinste Fehler können weit außerhalb des sichtbaren Bereiches liegen, aber trotzdem große Probleme bei der Weiterverarbeitung, Auswertung und der Datenhaltung verursachen. Dies betrifft auch die Daten des Räumlichen Informations- und Planungssystems (RIPS), welches vom Informationstechnischen Zentrum Umwelt (ITZ) der LUBW verwaltet wird. Im RIPS werden Umweltdaten aus unterschiedlichen Quellen zusammengeführt, die eine Vielzahl verschiedener Fehler aufweisen können. Die Qualitätskontrolle und -sicherung, von Geo- und Umweltdaten, ist daher eine wichtige Aufgabe des ITZ.

Im Rahmen dieser Bachelorarbeit sollen die bisherigen Qualitätssicherungsprozesse evaluiert werden und neue Qualitätssicherungsmaßnahmen vorgeschlagen werden. Es sollen prototypische Prozesse erstellt werden, welche geometrische und topologische Überprüfungen vornehmen, potentielle Fehler in den Daten kennzeichnen und diese gegebenenfalls korrigieren.

Beispielhaft werden in dieser Arbeit Geometriedaten von Wasserschutzgebieten (WSG) verwendet, da diese sehr anspruchsvollen Daten alle Anforderungen an flächenhafte Geometrien abdecken. Die WSG-Daten werden auf diverse geometrische und topologische Fehler geprüft. Während kleinere Geometriefehler auch automatisiert behoben werden können, muss bei größeren topologischen Problemen eine manuelle Behebung erfolgen, da diese möglicherweise topologisch falsch, aber inhaltlich richtig sein können, weshalb automatisiert nur ein Korrekturvorschlag ausgegeben werden kann. Der anhand der Wasserschutzgebietsdaten erstellte Prototyp soll im späteren Verlauf auch für andere Umweltdaten genutzt werden.

Arbeitsschritte:

- Literaturrecherche zur Qualität von Geodaten
- Evaluierung der organisatorischen und technischen Qualitätssicherungsprozesse im ITZ
- Spezifikation von Metriken und Standards für die Geodaten der LUBW
- Einarbeitung in die UIS-Datenstruktur und die Oracle Spatial Datenbank
- Analyse geeigneter Software zur Qualitätskontrolle und -sicherung
- Einarbeitung in die Software FME von Safe Software
- Konzeption und Erstellung prototypischer Prozesse zur Überprüfung von Geodaten auf Geometrie- und Topologiefehler
- Beispielhafte Konzeption eines Prototypen zur automatisierten Korrektur von ausgewählten Fehlertypen

Die LUBW sorgt während des gesamten Bearbeitungszeitraums für folgende Rahmenbedingungen:

- Rechnerarbeitsplatz
- Zugriff auf die kostenpflichtigen Softwares Esri ArcGIS und Safe Software FME
- Zugriff auf Wasserschutzgebietsdaten zu Testzwecken
- Technische und fachliche Betreuung

Alle theoretischen und konzeptionellen Überlegungen, die praktischen Arbeiten, sowie alle Ergebnisse der Arbeit, werden in der schriftlichen Ausarbeitung der Bachelorarbeit, dokumentiert. Des Weiteren ist die Arbeit in einem abschließenden Kolloquium zu präsentieren.

Leiter der Bachelorarbeit: Prof. Dr. Bernhard Bürg
Zweiter Prüfer: Prof. Dr. Detlef Günther-Diringer
Betreuer bei der LUBW: Dr. Florian Kost
Bearbeitungszeit: 3 Monate
Tag der Ausgabe: 07.05.2015
Tag der Abgabe: 07.08.2015
Anschrift des Kandidaten: Jan Jedersberger
Im Roth 8a
76275 Ettlingen
E-Mail: jan.jedersberger@web.de

Prof. Dr. Bernhard Bürg
(Leiter der Bachelorarbeit)

Prof. Dr. Detlef Günther-Diringer
(Zweiter Prüfer)

Dr. Florian Kost
(Betreuer bei der LUBW)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Benutzung der angegebenen Quellen und Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Quellen entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegen.

Karlsruhe, den

Jan Jedersberger

Danksagung

An dieser Stelle möchte ich mich bei Herrn Prof. Dr. Bernhard Bürg und Herrn Prof. Dr. Detlef Günther-Diringer für die Betreuung meiner Bachelorarbeit bedanken.

Bedanken möchte ich mich auch bei Herrn Dipl.-Agrarbiologe Manfred Müller und Herrn Dr. Florian Kost, da sie es mir ermöglicht haben die Arbeit in der Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg zu schreiben. Außerdem möchte ich mich bei Herrn Bernd Weber für die Unterstützung in technischen Fragen, sowie bei Frau Monika Huber und Herrn Egbert Mündörfer für die fachliche Unterstützung bedanken. Auch bei allen weiteren Kollegen des Referats 53.2 möchte ich mich für die Unterstützung und Anregungen bei der Erstellung meiner Arbeit bedanken. Besonderer Dank gilt meiner Familie und meinen Freunden, die mich in allen Lebenslagen unterstützt haben.

Inhaltsverzeichnis

Aufgabenblatt für die Bachelorarbeit.....	I
Erklärung	III
Danksagung.....	IV
Inhaltsverzeichnis.....	V
Abbildungsverzeichnis.....	VII
Tabellenverzeichnis	XI
Abkürzungsverzeichnis	XI
Abstract.....	XII
1 Einleitung.....	1
1.1 Motivation und Zielsetzung	1
1.2 Aufbau der Arbeit.....	3
2 Umfeld	4
2.1 Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg	4
2.2 Umweltinformationssystem Baden-Württemberg	5
2.3 Räumliches Informations- und Planungssystem.....	7
2.4 Daten	8
2.4.1 Datenstruktur des UIS und RIPS	8
2.4.2 Wasserschutzgebiete (WSG)	9
2.4.3 Datenherkunft und -struktur	12
3 Technischer Stand und Konzept	14
3.1 Technische Ausgangslage	14
3.2 Konzept	15
4. Qualität von Geodaten	22
4.1 Kriterien.....	22
4.1.1 Interne Qualität	23
4.1.2 Externe Qualität.....	26
4.1.3 Metadaten.....	27

4.2 Geodatenstandards und Gesetze	28
4.3 Ursachen für Fehler.....	29
4.4.1 Qualitätskriterien und Fehlerursachen am Beispiel Wasserschutzgebiete	30
5 Qualitätssicherungsprozesse in der LUBW	33
5.1 Qualitätssicherungsprozess am Beispiel Wasser-schutzgebiete.....	33
6 Software-Evaluierung für die Umsetzung der Arbeit	35
6.1 ESRI ArcGIS.....	35
6.2 GeoTools.....	36
6.3 GeoKettle	37
6.4 Safe Software FME	38
6.5 Ergebnis der Evaluierung	39
7 Umsetzung.....	41
7.1 Arbeiten mit der FME Workbench.....	41
7.1.1 Reader und Writer	41
7.1.2 Transformer.....	42
7.1.3 Attribute und Listen	43
7.1.4 User Parameter	44
7.1.5 Erstellen und Bearbeiten einer Transformation	45
7.2 Erstellung des Validierungswerkzeuges	48
7.2.1 Übersicht der verwendeten Transformer	48
7.2.2 Die Validierungs- und Korrekturprozesse.....	51
7.2.3 Der Master Workspace.....	76
7.2.4 Steuerung des Werkzeugs.....	79
8 Vorschläge für die Qualitätssicherung der LUBW	83
9 Fazit und Ausblick.....	84
Quellenverzeichnis.....	86
Anhang I – Master Workspace.....	90
Anhang II – Inhalt der beiliegenden DVD	91

Abbildungsverzeichnis

Abbildung 1: Optisch kaum wahrnehmbarer Geometriefehler.....	2
Abbildung 2: Organigramm der LUBW (Quelle: [LUBW_2])	5
Abbildung 3: Gesamtübersicht über das Umweltinformationssystem Baden-Württemberg (Quelle: [UIS_1, S. 16])	6
Abbildung 4: Übersicht über das RIPS (Quelle: [RIPS_2])	7
Abbildung 5: Beispielhafter Auszug einer SDO_GEOMETRY-Spalte	8
Abbildung 6: Karte eines Wasserschutzgebiets mit Zonen.....	11
Abbildung 7: WSG-Zonen und Umringe.....	12
Abbildung 8: Zwei lageidentische Stützpunkte.....	16
Abbildung 9: Nicht geschlossenes Polygon.....	16
Abbildung 10: Richtige Orientierung eines Donutpolygons.....	16
Abbildung 11: Verschiedenen Splitterflächen innerhalb, außerhalb und am Rand eines Polygons.....	17
Abbildung 12: Loch (Fehler) und ausgefülltes Loch (kein Fehler).....	17
Abbildung 13: Fischeschwanz-Polygon	17
Abbildung 14: Selbstüberschneidung	18
Abbildung 15: Spike	18
Abbildung 16: Mehrteilige Geometrie mit mehreren Splitterpolygonen.....	18
Abbildung 17: Verschiedene Anwendungsfälle der Mindestdimensionen	19
Abbildung 18: WSG-Zonen (blau) sind nicht an die ALK gesnappt	19
Abbildung 19: Sich überlappende Polygone	20
Abbildung 20: Lücke zwischen zwei Polygonen	20
Abbildung 21: Sliver Polygone	20
Abbildung 22: Klaffe zwischen zwei Polygonen.....	21
Abbildung 23: Stützpunkte eines Polygons sind nicht an das benachbarte gesnappt	21
Abbildung 24: Verhältnis von objektiven und Subjektiven Qualitätskriterien (Quelle: [ISTL])	22

Abbildung 25: Vergleich von Genauigkeit (Precision) und Richtigkeit (Accuracy) (Quelle: [DEV06, S. 184])	24
Abbildung 26: Übersicht über interne und externe Datenqualität (Quelle: [DEV06, S. 36]).	27
Abbildung 27: Das RIPS als Teil der INSPIRE-Richtlinie (Quelle: [RIPS_1]).....	29
Abbildung 28: Fehlerursachen und Fortpflanzung (Quelle: [DEV06, S. 34]).....	29
Abbildung 29: Zwei sich überlagernde Wasserschutzgebiete	32
Abbildung 30: Beispielhafter Prozess mit dem ModelBuilder (Quelle: [ESRI_2]).....	36
Abbildung 31: Beispielhafte Transformation mit GeoKettle (Quelle: [GEOK_2])	37
Abbildung 32: Beispielhafte Transformation mit der FME Workbench	38
Abbildung 33: Transformer-Fenster mit Parallel Processing Option	39
Abbildung 34: Reader und Writer Feature Types mit Attributen	42
Abbildung 35: Verschiedene Transformer	43
Abbildung 36: Transformer mit Listen und Attributen	43
Abbildung 37: Parameterübersicht.....	44
Abbildung 38: Transformer Parameter als Attribut, User Parameter, und Conditional Value	45
Abbildung 39: Transformer-Auswahl	46
Abbildung 40: Verbinden von Transformern	46
Abbildung 41: Einfügen eines Transformers.....	46
Abbildung 42: Transformer mit fehlenden Parametern	47
Abbildung 43: Aktivierter und deaktivierter Transformer mit Annotations und einem Bookmark.....	48
Abbildung 44: Prozess zur Erstellung des LogFiles.....	52
Abbildung 45: Prozess zum Entfernen von Attributen und der Auswahl des WSG	52
Abbildung 46: Funktionsweise des DonutHoleExtractors, Trennung des äußeren und inneren Rings (Quelle: [SAFE_6])	53
Abbildung 47: Feature, Feature nach der Transformation in Punkte und Ergebnis der Validierung	54
Abbildung 48: Konvertierung in Punkte und Identifizierung von doppelten Punkten	54
Abbildung 49: Filtern von geschlossenen Polygonen (Polylinien) im GeometryFilter.....	55

Abbildung 50: Prüfung auf falsch orientierte Polygone Im GeometryValidator.....	55
Abbildung 51: Abtrennen eines Lochs an der Kante (violett), durch einen Buffer (grün)	56
Abbildung 52: Prüfung auf „Dreipunkt“-Splitter durch extrahieren von Löchern	56
Abbildung 53: Attributtabelle mit Splitterflächen, die über ihren Flächenanteil bestimmt wurden. Bei coord = 4 liegen außerdem Dreipunkt-Splitter vor.	57
Abbildung 54: Prüfung auf sehr kleine Splitterpolygone, durch Berechnung es Anteils an der Gesamtfläche	57
Abbildung 55: Prüfung auf Löcher	58
Abbildung 56: Verschiedene Arten lageidentischer Stützpunkte bei der Identifizierung von Fischeschwanz-Polygonen, rot identifiziert Fischeschwänze	58
Abbildung 57: Datenvorbereitung: Kennzeichnung der Features und Konvertierung in Punkte	59
Abbildung 58: 1. doppelte Punkte gruppiert nach ihrem ursprünglichen Feature, 2. doppelte Punkte gruppiert nach neu dem erzeugten Feature	59
Abbildung 59: Prüfung auf Fischeschwanz-Polygone	60
Abbildung 60: Beispielhafte Konvertierung eines Features mit drei Listen mit je zwei Elementen in einen Datensatz mit zwei Features und drei Spalten.....	60
Abbildung 61: Prüfung auf Selbstüberschneidungen im GeometryValidator	60
Abbildung 62: Prüfung, der Removed-Port gibt	61
Abbildung 63: Definition von Spikes anhand ihres Winkels (Quelle: [SAFE_7])	61
Abbildung 64: Identifizierung zu enger, parallel verlaufender Kanten, die dritte Kante innerhalb des Mindestabstands ist rot markiert.	62
Abbildung 65: Engstelle, Buffer um die einzelnen Kanten und Punkt in der Überschneidung von drei Buffern, der die Engstelle markiert.....	62
Abbildung 66: Prüfung auf die Einhaltung der Mindestabstände durch Verschneiden von Bufferflächen um die einzelnen Stützpunkte	62
Abbildung 67: 1. Multipart-Polygone. 2. getrennte Singleparts, 3. Ergebnis der Validierung	63
Abbildung 68: Prüfung auf Multipart-Polygone, durch Aufteilen und Wiederherstellen der Geometrien	63
Abbildung 69: Verschiedene WSG-Zonen und gekennzeichnete Überlappungsflächen.....	64
Abbildung 70: Prüfung auf Überlappungen zwischen ganzen WSG (oben) und WSG-Zonen (unten)	65

Abbildung 71: Prüfung auf Lücken zwischen Polygonen	65
Abbildung 72: Sliver Polygone, die durch die Anzahl der Überlappungen bestimmt wurden	66
Abbildung 73: Prüfung auf Sliver Polygone.....	66
Abbildung 74: Klaffe (links) erzeugt einen Spike im Umring (rechts)	67
Abbildung 75: Prüfung auf Klaffen.....	67
Abbildung 76: Ergebnis vor und nach dem Ausschneiden der Punkte auf der Außengrenze	68
Abbildung 77: Prüfung der Kantenidentität benachbarter Geometrien	69
Abbildung 78: Import der ALK-Daten in den Prozess und Aufbereitung der Daten für die Analyse.....	69
Abbildung 79: Prüfung der Anpassung an die ALK-Daten, durch Verschneiden der Stützpunkte der Geometrien mit den ALK-Daten.....	70
Abbildung 80: Erstellung des Punktshapes mit den Attributen.....	72
Abbildung 81: Fehler der bei der Korrektur mit Splitterflächen entsteht	72
Abbildung 82: Unbeabsichtigt zusammengefügte Polygone.....	73
Abbildung 83: Korrupte Geometrien, die bei der Korrektur mit Löchern entstehen (rechts) im Vergleich zu richtig korrigierten Geometrien (links).....	73
Abbildung 84: Übersicht Datenvorbereitung.....	73
Abbildung 85: Identifizieren und entfernen doppelter Stützpunkte	74
Abbildung 86: Wiederherstellen der ursprünglichen Polygone.....	75
Abbildung 87: Korrektur von Spikes.....	75
Abbildung 88: Wiederanfügen der Attribute und Wiederherstellung der Multipart-Polygone	76
Abbildung 89: Verkettete Validierungsprozesse, mit Tester und TestFilter für die Parameterabfragen.....	77
Abbildung 90: Korrekturblock mit den Attributanpassungen (Count, InName, OutName) ..	78
Abbildung 91: Parameterübersicht: 1. Input- und Output-Parameter, 2. Testparameter, 3. Werkzeugspezifische Parameter.....	82

Tabellenverzeichnis

Tabelle 1: Fehlerursachen der WSG-Geometrien	31
Tabelle 2: Abschließender Softwarevergleich	40
Tabelle 3: Konvertierung von Oracle Datentypen in FME-interne Datentypen (Quelle: [SAFE_5])	41
Tabelle 4: Übersicht: Input- und Output-Parameter	79
Tabelle 5: Testparameter	81
Tabelle 6: Werkzeugspezifische Parameter	81

Abkürzungsverzeichnis

ALK	Automatisierte Liegenschaftskarte
CSV	Comma Separated Value
DAD	Datenaustauschdienst
ETL	Extract, Transform, Load
FME	Feature Manipulation Engine
GIS	Geoinformationssystem
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Organization for Standardization
ITZ	Informationstechnisches Zentrum Umwelt
LGRB	Landesamt für Geologie, Rohstoffe und Bergbau
LUBW	Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg
OGC	Open Geospatial Consortium
RIPS	Räumliches Informations- und Planungssystem
SDE	Spatial Database Engine
SDO	Spatial Data Option
UDO	Umwelt-Datenbanken und -karten Online
UIS	Umweltinformationssystem
UWB	Untere Wasserbehörde
WSG	Wasserschutzgebiet

Abstract

Die hier vorliegende Bachelorarbeit „Spezifikation von Qualitätsanforderungen für Umweltdaten, Erstellung eines konzeptionellen Vorgehensmodells zur Fehleridentifikation und -verbesserung sowie Erstellung eines Prototypen zur topologischen, geometrischen und inhaltlichen Plausibilisierung von Geodaten des Räumlichen Informations- und Planungssystems (RIPS)“ wurde im Rahmen des Studiengangs Kartographie und Geomatik an der Hochschule Karlsruhe – Technik und Wirtschaft verfasst. Die Arbeit wurde in Zusammenarbeit mit der Landesanstalt für Umwelt, Messungen und Naturschutz Baden Württemberg (LUBW) erstellt.

Ziel der Arbeit ist es, Kriterien für die Qualität von raumbezogenen Daten zu evaluieren und spezifische Kriterien für die Daten der LUBW festzulegen. Die Bedeutung der Geodatenqualität ist nicht zu unterschätzen, da von Geodaten viele weitere Planungs- und Entscheidungsprozesse abhängig sind. Beispielhaft werden für die Bearbeitung dieser Bachelorthesis die Daten der Wasserschutzgebiete verwendet.

Neben der Spezifikation von Qualitätskriterien ist auch die Erstellung eines Prototypen zur Validierung von Geometriedaten Bestandteil dieser Arbeit. Dieser wurde, im Anschluss an eine Evaluierung geeigneter Softwareprodukte, mit der Software FME von Safe Software erstellt.

Schlagwörter: Datenvalidierung – FME – Geodatenqualität – Geometriedaten – Qualitätssicherung – Wasserschutzgebiete

1 Einleitung

In diesem Kapitel werden die Problemstellung dieser Arbeit, sowie deren Zielsetzung erläutert. Außerdem wird das Arbeitsumfeld, in welchem diese Bachelorthesis erstellt wird, behandelt. Abschließend wird der Aufbau der Arbeit beschrieben.

1.1 Motivation und Zielsetzung

Um Daten sinnvoll verwenden zu können, muss ihre Qualität gewährleistet sein. Dies trifft auf raumbezogene Daten, wie auch auf alle anderen Datentypen zu [DEV06, S. 13]. Geodaten unterscheiden sich durch die Kombination von Attributdaten und Geometrien von anderen Daten. Die Geometrien bieten zwar zusätzliche Informationen, sind aber auch Quellen für potentielle Fehler, welche große Probleme bei der Weiterverarbeitung und Analyse der Informationen verursachen können. Selbst geometrische und topologische Fehler, die bereits weit außerhalb des sichtbaren Bereiches liegen, können Ergebnisse stark verfälschen oder die korrekte Bearbeitung und Speicherung verhindern. Mangelnde Qualität von Geodaten ist aber nicht nur ein Problem im Bereich der Daten selbst, Geodaten sind oftmals Grundlagen für Planungs- und Entscheidungsprozesse, die ebenfalls von fehlerhaften Daten beeinträchtigt werden, was zu hohen Folgekosten, rechtlichen Schwierigkeiten und Gefahrensituationen führen kann [LEIT, S. 7]. Daraus folgt, dass der Nutzen von raumbezogenen Daten mit ihrer Qualität steigt und fällt. Dies führt dazu, dass die Bedeutung von Qualitätskontrolle und -sicherung auch im Bereich von Geodaten immer weiter wächst [MAY07, S. 3].

Diese Problematik betrifft auch die Umwelt- und Geodaten des Räumlichen Informations- und Planungssystems (RIPS) der Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW). Da die Daten des RIPS aus vielen unterschiedlichen Quellen stammen und auf verschiedene Weisen erfasst wurden, kann es zu einer Vielzahl von Fehlern kommen, die

von großen, offensichtlichen, bis hin zu kleinsten, nicht sichtbaren Fehlern reichen (siehe Abbildung 1). Die Qualitätssicherung im RIPS ist deswegen eine sehr komplexe und wichtige Aufgabe, die das Informationstechnische Zentrum Umwelt (ITZ) der LUBW stark beansprucht.

Für diese Aufgabe gibt es bereits viele technische und organisatorische Abläufe. Ein Ziel dieser Arbeit ist es jedoch, neue Metriken und Standards für die Geodaten der LUBW zu erarbeiten. Anhand dieser neuen Richtlinien sollen die aktuellen Maßnahmen evaluiert werden und nach Möglichkeiten zur Verbesserung gesucht werden. Im technischen Bereich soll außerdem evaluiert werden, welche Software zur Erfüllung der Qualitätsstandards geeignet ist, da die aktuellen Programme noch nicht vollständig auf dem neuesten Stand sind und es möglicherweise bessere Alternativen gibt. Aber auch die organisatorischen Abläufe sollen evaluiert werden und mögliche Verbesserungen vorgeschlagen werden um neue Programme effektiv einzubinden.



Abbildung 1: Optisch kaum wahrnehmbarer Geometriefehler

Im Rahmen dieser Arbeit sollen, neben der Optimierung der bisherigen Abläufe, basierend auf dem Ergebnis der Software-Evaluierung, neue prototypische Prozesse entwickelt werden. Diese Prozesse sollen Daten automatisiert auf geometrische, topologische und inhaltliche Fehler überprüfen. Diese Fehler sollen in den Daten gekennzeichnet werden, dem Nutzer mit einem Korrekturvorschlag ausgegeben und gegebenenfalls korrigiert werden. Eine allgemeine Fehlerkorrektur ist nicht möglich, da topologische Fehler inhaltlich richtig sein können und im Einzelfall angepasst werden müssen, während geometrische Fehler automatisch korrigiert werden können.

Für diese Arbeit werden beispielhaft die Prozesse der Wasserschutzgebiete (WSG) untersucht, außerdem werden die WSG-Geometrien als Testdaten für

die Entwicklung neuer Prozesse verwendet, da es sich hierbei um besonders anspruchsvolle Daten handelt, die ein breites Spektrum der Anforderungen an Flächengeometrien abdecken. Diese Prototypen und Prozesse sollen generisch auf weitere Datenbestände angewendet werden können.

1.2 Aufbau der Arbeit

Im zweiten Kapitel dieser Bachelorarbeit wird das Umfeld beschrieben, in dem diese Arbeit erstellt wurde, dies schließt auch die Daten, mit denen gearbeitet wird, ein. Im dritten Kapitel soll ein Überblick über den technischen Stand zu Beginn der Arbeit, sowie das Konzept der praktischen Arbeit gegeben werden. Das vierte Kapitel enthält Informationen über Qualitätskriterien und -sicherung im Bereich der Geodaten. Im fünften Kapitel werden Qualitätssicherungsprozesse für Geo- und Umweltdaten innerhalb der LUBW analysiert. Das sechste Kapitel beschreibt die Auswahl der Software für die praktische Umsetzung dieser Arbeit, während im siebten Kapitel diese Umsetzung erläutert wird. Im darauffolgenden Kapitel werden aufbauend auf die bisherigen Ergebnisse Alternativen für die Qualitätssicherung vorgeschlagen. Abschließend folgen im neunten Kapitel das Fazit der Bachelorarbeit sowie ein Ausblick.

2 Umfeld

In diesem Kapitel soll das Arbeitsumfeld, in dem diese Bachelorarbeit erstellt wird, beschrieben werden, außerdem wird auf die im Rahmen dieser Arbeit verwendeten Daten eingegangen.

2.1 Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg

Die Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg ist eine Anstalt des öffentlichen Rechts und untersteht dem Ministerium für Umwelt, Klima und Energiewirtschaft Baden-Württemberg. Die LUBW ging aus der Zusammenlegung der Landesanstalt für Umweltschutz (LfU) und dem Zentrum für Umweltmessungen, Umwelterhebungen und Gerätesicherheit (UMEG) hervor [GBI_1].

Die LUBW dient als Kompetenzzentrum des Landes Baden-Württemberg für die Bereiche Umwelt- und Naturschutz, technischer Arbeitsschutz, Strahlenschutz und Produktsicherheit. Insgesamt sind 550 Mitarbeiter, an fünf Standorten, mit der Lösung diverser Umweltprobleme beschäftigt. Daten werden von der LUBW an Politik und Verwaltung, aber auch an Unternehmen und Privatpersonen weitergegeben [LUBW_1].

Die LUBW besteht aus den sechs Abteilungen:

- Zentrale Dienste,
- Nachhaltigkeit und Naturschutz,
- Technischer Umweltschutz,
- Wasser,
- Informationstechnisches Zentrum Umwelt,
- Messtechnik und Analytik [LUBW_2].

Der genaue Aufbau der einzelnen Abteilungen ist in Abbildung 2 dargestellt.

Im Informationstechnischen Zentrum Umwelt wird die IT-Infrastruktur der LUBW verwaltet. Außerdem werden die Umwelt- und Geodaten des Landes

Baden-Württemberg in Form des Umweltinformationssystems Baden-Württemberg (UIS) und des Räumlichen Informations- und Planungssystems (RIPS) gepflegt.

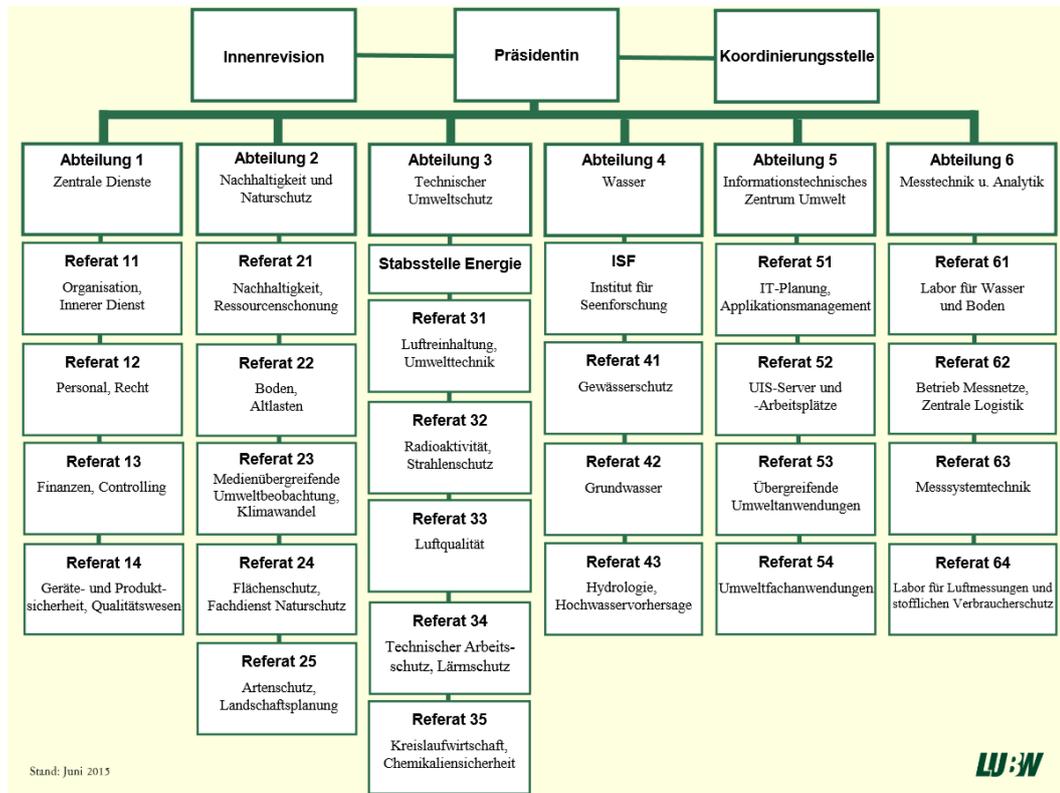


Abbildung 2: Organigramm der LUBW (Quelle: [LUBW_2])

2.2 Umweltinformationssystem Baden-Württemberg

Das Umweltinformationssystem Baden-Württemberg dient der Bereitstellung von aktuellen Umweltinformationen. Hierbei werden eine große Menge verschiedener Daten erhoben, analysiert, bewertet und gespeichert. Die Daten stammen aus unterschiedlichen Quellen, wie manuellen Messungen, automatischen Messnetzen oder sie werden aus bereits bestehenden Datenbanken entnommen. Neben den weit gefächerten Datenquellen, werden auch im Softwarebereich verschiedene Technologien, wie Geoinformationssysteme oder Umweltdatenbanken, verwendet (siehe Abbildung 3).

Diese Vielzahl an Daten und Systemen lässt sich in die Kategorien Übergreifende Komponenten, Fachkomponenten und Basiskomponenten unterteilen:

- Übergreifende Komponenten sind Systeme, die zur Zusammenführung und fachübergreifenden Nutzung, von Daten verschiedener Umwelt- und Zuständigkeitsbereiche, dienen. Beispiele hierfür sind der Daten- und Kartendienst UDO (Umwelt-Datenbanken und -karten Online) oder das Räumliche Informations- und Planungssystem.
- Fachkomponenten sind verschiedene Systeme welche bei der Erfassung und Bearbeitung von Umweltfachdaten eingesetzt werden, wie das Fachinformationssystem Naturschutz oder die Grundwasserdatenbank.
- Basiskomponenten sind Komponenten, welche nicht ausschließlich zur Analyse und Bearbeitung von Umweltinformationen geschaffen wurden, aber dennoch für das UIS verwendet werden. Hierzu gehören verschiedene statistische Daten oder Geodaten, wie die Automatisierte Liegenschaftskarte (ALK) [UIS_1].

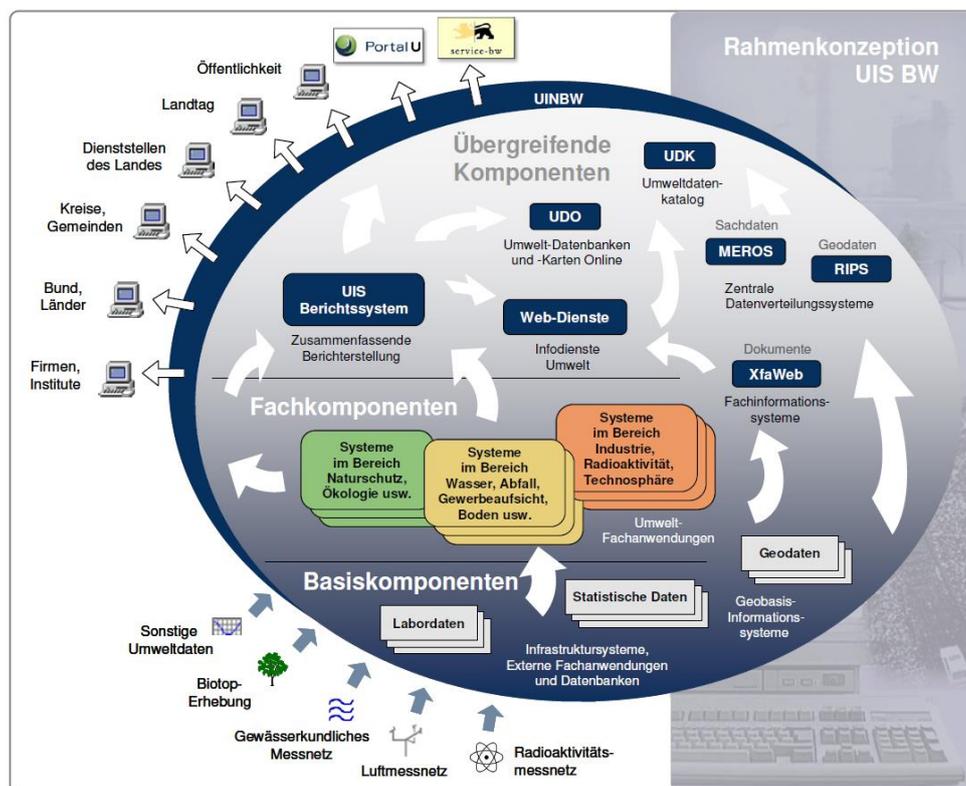


Abbildung 3: Gesamtübersicht über das Umweltinformationssystem Baden-Württemberg (Quelle: [UIS_1, S. 16])

2.3 Räumliches Informations- und Planungssystem

Ein bedeutender Bestandteil des UIS ist das Räumliche Informations- und Planungssystem, hierbei handelt es sich um eine ressortübergreifende Geodateninfrastruktur. Hauptaufgabe des RIPS ist die Unterstützung von Umweltfachverfahren durch die Bereitstellung von Geobasisdaten und Geofachdaten. Die Daten des RIPS sind zu einem Teil öffentlich zugänglich, viele Daten sind jedoch nur von kommunalen und staatlichen UIS-Arbeitsplätzen aus verfügbar (siehe Abbildung 4).

Die zentrale Komponente des RIPS ist die themenübergreifende UIS-Datenbank, in der alle umweltbezogenen Fach- und Geodaten des Landes verwaltet werden [RIPS_1]. Die Daten werden also an verschiedenen Stellen von verschiedenen Erfassern aufgenommen und in die Datenbank eingespeist. Beispiele für die Daten im RIPS sind die Daten des UDO oder der Potenzialatlas Erneuerbare Energien der LUBW.

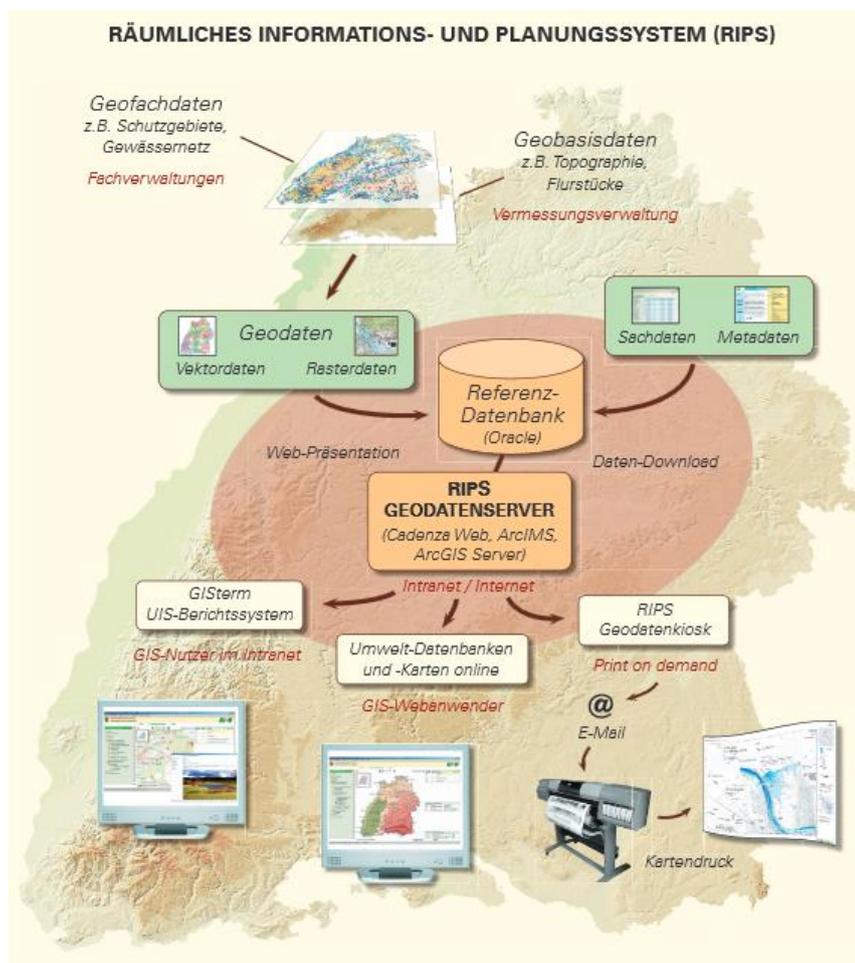


Abbildung 4: Übersicht über das RIPS (Quelle: [RIPS_2])

im Fall von mehreren Tabellen eines Themas, bei historischen Ständen, wiederholt sich der Name zwar, aber die Eindeutigkeit wird durch den Zusatz einer Jahreszahl, bei den historischen Ständen, gewahrt. Der aktuelle Stand der Wasserschutzgebietszonen erhält so den Namen UIS_0100007500200036.

Die im RIPS verwalteten Geodaten können in zwei Gruppen eingeteilt werden, die Geobasisdaten und die Geofachdaten. Die Basisdaten werden von der Vermessungsverwaltung bereitgestellt. Sie umfassen die Automatisierte Liegenschaftskarte (ALK), digitale Gelände-, Höhen- und Landschaftsmodelle, sowie digitale topographische Karten und Orthophotos. Die Fachdaten enthalten neben ihrer geographischen Information, zusätzliche Daten zu verschiedenen umweltrelevanten Themenbereichen. Hierzu gehören beispielsweise die Wasserschutzgebiete. Die Geometrien der Fachdaten werden in Bezug auf die Basisdaten erstellt. Für die Fachdatenverwaltung werden spezielle Fachanwendungen eingesetzt, welche dabei helfen, die Daten, bereits beim Erfassen und Digitalisieren, an viele Vorgaben anzupassen. Ein Beispiel für eine solche Fachanwendung ist WawiG, die speziell für wasserwirtschaftliche Gebiete, darunter auch Wasserschutzgebiete, konzipiert wurde.

2.4.2 Wasserschutzgebiete (WSG)

Wasserschutzgebiete sind Flächen, in denen besondere gesetzliche Regelungen zum Schutz der Wasserversorgung gelten. Laut dem Wasserhaushaltsgesetz der Bundesrepublik Deutschland können WSG unter folgenden Voraussetzungen festgesetzt werden:

„Soweit es das Wohl der Allgemeinheit erfordert,

- 1. Gewässer im Interesse der derzeit bestehenden oder künftigen öffentlichen Wasserversorgung vor nachteiligen Einwirkungen zu schützen,*
- 2. das Grundwasser anzureichern oder*

3. das schädliche Abfließen von Niederschlagswasser sowie das Abschwemmen und den Eintrag von Bodenbestandteilen, Dünge- oder Pflanzenschutzmitteln in Gewässer zu vermeiden, kann die Landesregierung durch Rechtsverordnung Wasserschutzgebiete festsetzen.“ [WHG_1]

Die Festsetzung ist Aufgabe der Landesregierung und kann von dieser an andere Landesbehörden weitergegeben werden.

Geographisch umfassen die WSG den Einzugsbereich einer oder mehrerer Trinkwasserentnahmestellen und werden selbst in drei Zonen unterteilt (siehe Abbildung 6). In diesen gelten verschieden strenge Regelungen für die Nutzung des Areals. Die Zone I umfasst die direkte Umgebung der Entnahmestelle, in einem Radius von mindestens 10m. Hier sind jegliche anderweitige Nutzungen des Gebietes, sowie das Betreten der Zone für Unbefugte verboten. Die Zone II umfasst das Gebiet, für welches das Wasser 50 Tage benötigt, um es bis zum Brunnen zu durchqueren. In der Zone II sind Nutzungsarten, welche die Oberfläche des Bodens verletzen, wie Bebauung oder landwirtschaftliche Nutzung, verboten. Die Zone II soll verhindern, dass bakteriellen Verunreinigungen im Wasser bis zur Entnahmestelle vordringen. Die Zone III erfasst das gesamte Einzugsgebiet der Wasserentnahmestelle. Die hier geltenden gesetzlichen Regelungen beziehen sich vor allem auf den Umgang mit wassergefährdenden Stoffen, wie die Verwendung von Schädlingsbekämpfungsmitteln in der Landwirtschaft. Die Zonen II und II können in einigen Fällen noch jeweils in die Zonen A und B unterteilt werden. In diesen Zonen A und B gibt es nochmals Abstufungen bei den gesetzlichen Regelungen [WIKI_1].

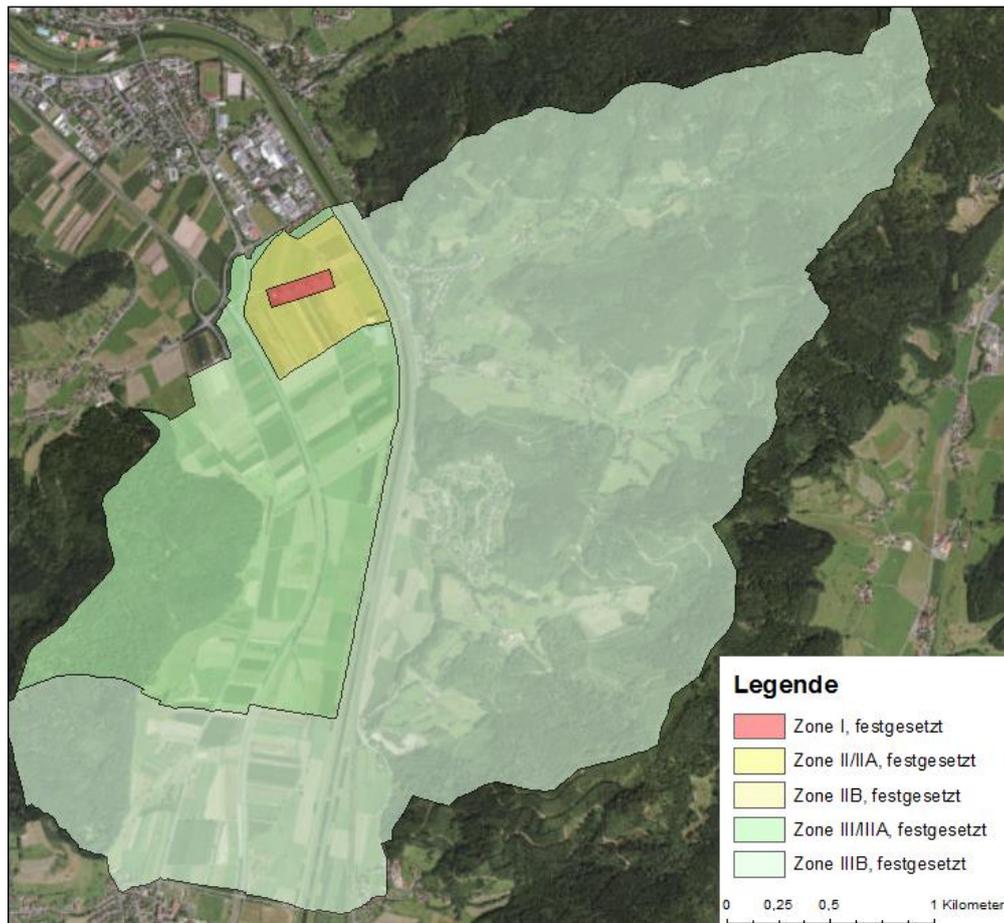


Abbildung 6: Karte eines Wasserschutzgebiets mit Zonen

Die genauen Regelungen für die Nutzung der einzelnen Zonen werden bei der rechtskräftigen Festsetzung des Wasserschutzgebietes in einer Rechtsverordnung festgelegt [VER_1].

Wasserschutzgebiete durchlaufen bei ihrer Festsetzung mehrere Phasen, beginnend mit der fachtechnischen Abgrenzung. Die fachtechnische Abgrenzung wird nach der Beantragung des Wasserschutzgebietes und der Erstellung eines hydrogeologischen Gutachtens durch das Landesamt für Geologie, Rohstoffe und Bergbau (LGRB) durchgeführt. [VER_2]

Anschließend befindet sich das Wasserschutzgebiet im Verfahren. Während dieses Zeitraums werden betroffene Gemeinden über die Pläne ein WSG einzurichten, informiert, und die Pläne werden für betroffene Bürger zugänglich gemacht. Während dieser Zeit besteht die Möglichkeit, Einsprüche gegen das WSG zu erheben. Abschließend wird das WSG mittels einer Rechtsverordnung festgesetzt.

2.4.3 Datenherkunft und -struktur

Die WSG-Daten stammen von den unteren Wasserbehörden (UWB) der Landratsämter und Bürgermeisterämter der Stadtkreise. Die Daten werden von den UWB oder von ihnen beauftragten Ingenieurbüros aufgenommen und in lokale Datenbanken der Behörden eingespeist. Über den Datenaustauschdienst (DAD) gelangen die WSG-Daten in die Referenzdatenbank der LUBW.

Die Geometriedaten jedes Wasserschutzgebiets bestehen aus je einem Polygon für jede Schutzzone, wobei mehrteilige Zonen als Multipart-Polygone gespeichert werden. Für das gesamte WSG existiert zusätzlich noch ein Umring-Polygon, welches alle Zonen umfasst (siehe Abbildung 7).



Abbildung 7: WSG-Zonen und Umringe

Im Datenbestand der LUBW befinden sich Geometrien für geplante, fachtechnisch abgegrenzte, im Verfahren befindliche und rechtskräftig festgesetzte Wasserschutzgebiete, zusätzlich existieren noch Daten für aufgehobene Gebiete. Im Fall einer Gebietsänderung wird in den Daten ein neues Schutzgebiet angelegt und der alte Stand wird als aufgehobenes Gebiet in der Datenbank belassen. WSG-Geometrien werden in zwei Maßstabsbereichen geführt, der Maßstabsbereich M1 entspricht einem Erfassungs-Maßstab von 1:10.000 oder größer, als Datengrundlage für die Geometrien wird in diesem Bereich die Automatisierte Liegenschaftskarte verwendet. Diese Genauigkeit ist notwendig, da die Geometrien einen Rechtsbezug haben. Der Bereich M2 entspricht einem Erfassungs-Maßstab

von kleiner 1:10.000 bis einschließlich 1:50.000, als Grundlage für die M2-Geometrien dient das Digitale Basis-Landschaftsmodell (Basis-DLM). Diese Geometrien werden für topographische Übersichten verwendet. Diese Bachelorarbeit bezieht sich vorrangig auf die M1-Geometrien.

2.4.3.1 Anpassung an die ALK

Die ALK bildet die Erfassungsgrundlage der Wasserschutzgebiete, jedoch ist sie nicht flächendeckend vorhanden. Durch die ALK ist es möglich beim Digitalisieren einen exakten Raumbezug herzustellen, da alle Objekte der ALK eindeutig definierbar sind [WAABIS, S 1.7]. Für die Digitalisierung der Wasserschutzgebiete sind vor allem ALK-Punkte, ALK-Linien und ALK-Flurstücke von Bedeutung. Die ALK-Punkte enthalten Objekte wie trigonometrische Punkte oder Grenzpunkte. ALK-Linien enthalten Objekte wie Grenzen von Verwaltungseinheiten, Gebäudelinien, topographische Linien und verschiedene Leitungen. Die ALK-Flurstücke enthalten Flurstücksgrenzen. Durch diese Daten ist es möglich die Abgrenzung eines WSG exakt zu bestimmen.

3 Technischer Stand und Konzept

In diesem Kapitel soll die technische Ausgangslage zu Beginn der Bachelorarbeit beschrieben werden. Außerdem wird das Konzept für die Umsetzung des praktischen Teils der Arbeit vorgestellt.

3.1 Technische Ausgangslage

Im ITZ wird zum gegenwärtigen Zeitpunkt bereits ein mit der Software FME erstellter Prozess (vgl. Kapitel 8 „Umsetzung“) zur Überprüfung der Wasserschutzgebietsdaten eingesetzt. FME ist eine Software zur Verarbeitung und Analyse von raumbezogenen Daten. Aus verschiedenen Gründen soll dieser Prozess durch ein neues Werkzeug zur Datenvalidierung ersetzt werden. Der alte Prozess wurde für eine ältere Version von FME konzipiert und kann nur noch bis zur Version 2013 zuverlässig eingesetzt werden, während die neusten Versionen ihn nicht mehr vollständig unterstützen. Da beispielsweise die Funktionen der verwendeten Transformer, den Geoverarbeitungswerkzeugen von FME, abgeändert wurden oder diese nicht mehr in den damaligen Formen existieren. Des Weiteren ist die Wartung und Aktualisierung des Werkzeuges, aufgrund seines sehr komplexen und unübersichtlichen Aufbaus, nur schwer möglich. Der Gesamtprozess wurde um Custom Transformer, spezielle durch den Nutzer erstellte Transformer, aufgebaut, welche nicht vom ITZ, sondern von einer anderen Stelle, entwickelt wurden. Da die Custom Transformer ebenfalls komplex und teilweise ineinander verschachtelt aufgebaut sind, ist eine Wartung, mit vertretbarem Aufwand, nur durch Entwickler möglich. Der Gesamtprozess wurde ebenfalls in Einzelschritten weiter verarbeitet, weshalb er auch für erfahrene FME-Nutzer sehr undurchsichtig ist. Langfristig kann mit diesem Werkzeug keine zuverlässige Datenvalidierung garantiert werden.

3.2 Konzept

Der im Rahmen dieser Bachelorarbeit erstellte Prototyp soll die Geometrien von umweltbezogenen Geodaten auf verschiedene geometrische und topologische Fehler überprüfen. Für den Nutzer soll ein Shapefile ausgegeben werden, in dem die Fehler gekennzeichnet sind. An der Position des jeweiligen Fehlers soll sich im Fehlershape ein Punkt befinden, in dessen Attributen der Fehlertyp angegeben ist. Mit Hilfe dieses Fehlershapes soll es dem Anwender ermöglicht werden, seine Datensätze so anzupassen, dass sie den Qualitätsanforderungen entsprechen.

Hierfür soll für jeden Fehlertypen mit der Software FME von Safe Software eine so genannte Workbench (siehe Kapitel 8.1) angelegt werden, welche alle zur Analyse der Daten notwendigen Prozesse ausführt.

Der Anwender soll die Möglichkeit erhalten, auszuwählen welche Überprüfungen vorgenommen werden. Es soll möglich sein, nur einzelne ausgewählte Fehlertypen zu überprüfen oder im Voraus zusammengestellte Überprüfungen für bestimmte Datentypen, wie Wasserschutzgebiete oder Umringe, durchzuführen. Des Weiteren soll für ausgewählte Fehlertypen auch die Option bestehen, sie mit Hilfe des Prototyps automatisiert korrigieren zu lassen. Dies wird jedoch nur im Falle einfacher Geometriefehler möglich sein, bei denen ausgeschlossen werden kann, dass der Fehler in den Daten den realen Gegebenheiten entspricht und deswegen gewollt ist.

In Absprache mit den für Wasserschutzgebiete verantwortlichen Mitarbeitern des ITZ wurde ein Katalog mit verschiedenen geometrischen und topologischen Fehlern zusammengestellt. Diese Fehler sollen, mit Hilfe des hier erstellten Prototyps, erkannt werden.

Im Einzelnen werden die Eingangsdaten auf folgende geometrische Fehler und Probleme untersucht:

1. Doppelte Stützpunkte: Hierbei handelt es sich um zwei Stützpunkte, welche über die gleichen Koordinaten verfügen. Stützpunkte, deren Abstand zueinander weniger als ein durch den Nutzer festgelegter Mindestabstand beträgt, werden ebenfalls als doppelte Stützpunkte behandelt.

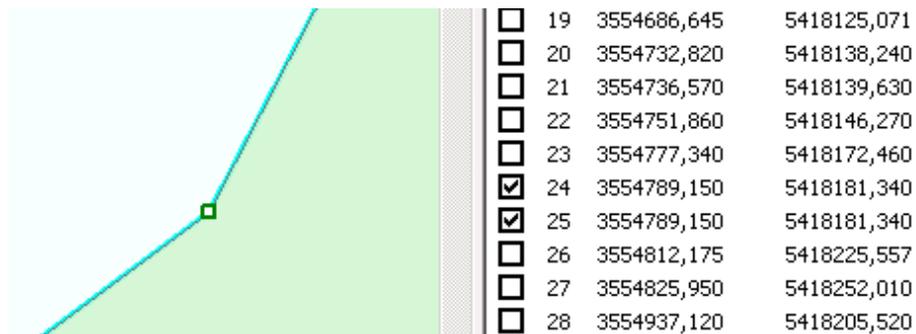


Abbildung 8: Zwei lageidentische Stützpunkte

2. Nicht geschlossene Polygone: Nicht geschlossene Polygone sind keine Flächengeometrien, sondern Liniengeometrien (Polylinien).

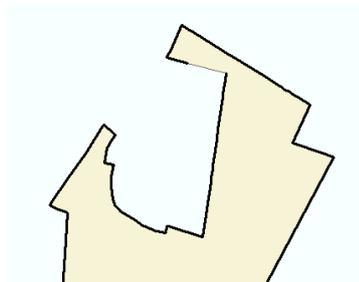


Abbildung 9: Nicht geschlossenes Polygon

3. Orientierung der inneren und äußeren Ringe: Im Falle von Donutpolygons müssen der innere und äußere Ring des Polygons über verschiedene Orientierungen verfügen. Dies ist an der Reihenfolge der einzelnen Stützpunkte erkennbar.

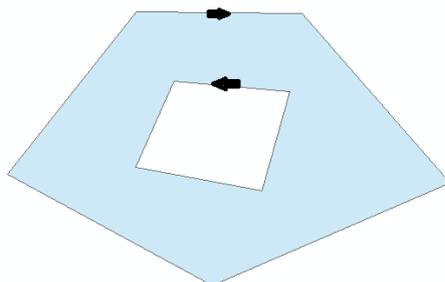


Abbildung 10: Richtige Orientierung eines Donutpolygons

4. Splitterflächen: Hierbei handelt es sich um kleinste Flächen, welche mit der eigentlichen Geometrie ein Multipart-Polygon bilden. Außerdem umfassen die Splitterflächen kleinste Löcher. Flächen und Löcher, welche nur über drei Stützpunkte verfügen, sind ebenfalls mit hoher Wahrscheinlichkeit Splitterflächen.

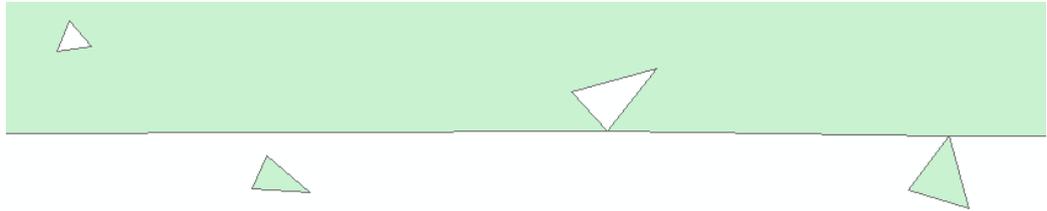


Abbildung 11: Verschiedenen Splitterflächen innerhalb, außerhalb und am Rand eines Polygons

5. Löcher: Löcher in den Geometrien können gewollt sein, jedoch kann es sich, gerade bei kleineren Löchern, um Fehler handeln. Löcher, welche durch weitere Polygone (beispielsweise eine andere WSG-Zone) ausgefüllt werden, sind in der Regel beabsichtigt.

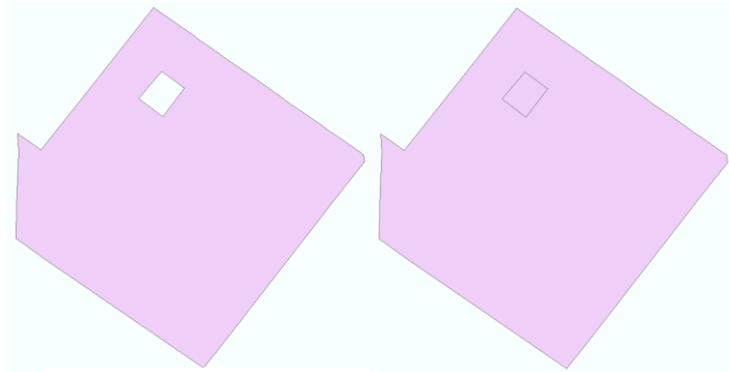


Abbildung 12: Loch (Fehler) und ausgefülltes Loch (kein Fehler)

6. Fischeschwanz-Polygone: Als Fischeschwanz bezeichnet man Multipart-Polygone, in denen sich zwei Teilflächen in nur einem Punkt berühren. Zwei verschiedene Flächen, welche sich in nur einem Punkt berühren, sind keine Fischeschwänze.

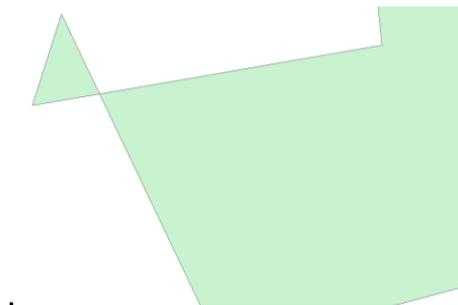


Abbildung 13: Fischeschwanz-Polygon

7. Selbstüberschneidungen: Hierbei handelt es sich um Flächen, bei denen Kanten sich mit anderen Kanten derselben Fläche kreuzen.

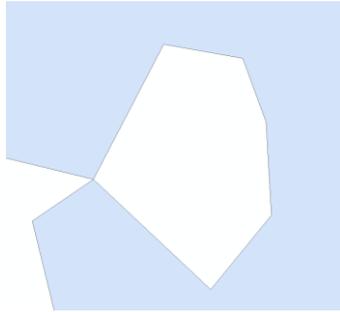


Abbildung 14: Selbstüberschneidung

8. Spikes: Spikes sind lange, dünne Spitzen an einem Polygon. In der Regel ist von drei aufeinander folgenden Stützpunkten der mittlere stark versetzt. Dies kann sowohl nach innen als auch nach außen auftreten.

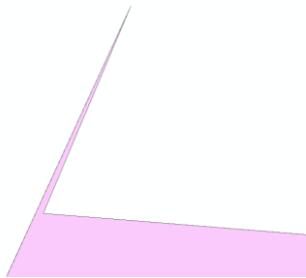


Abbildung 15: Spike

9. Mehrteilige Geometrien: Multipart-Polygone können beabsichtigt sein, da beispielsweise eine WSG-Zone aus mehreren Flächen bestehen kann. Ist dies nicht der Fall, können Multipart auf zusätzliche Flächen, wie Splitter, hinweisen.

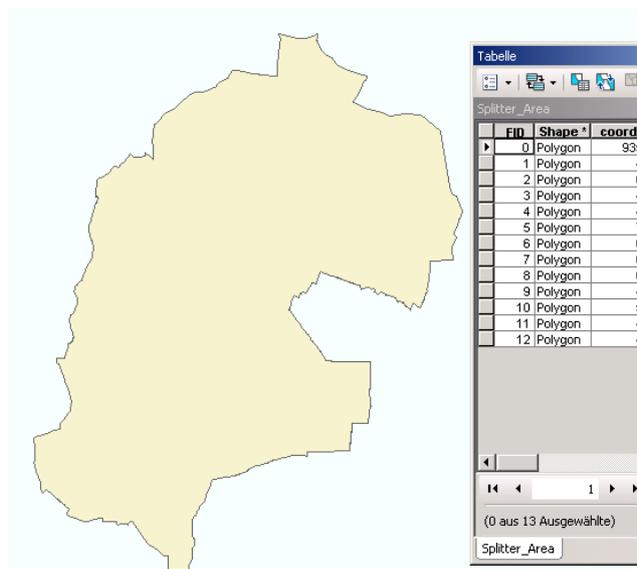


Abbildung 16: Mehrteilige Geometrie mit mehreren Splitterpolygonen

10. Mindestdimensionen: Teilgeometrien müssen einen bestimmten, durch den Nutzer festgelegten, Abstand zueinander aufweisen. Dies schließt auch die Abstände zwischen den einzelnen Ringen von Donutpolygonen mit ein. Außerdem muss die Länge der Kanten von einzelnen Flächen ebenfalls diesem Abstand entsprechen.

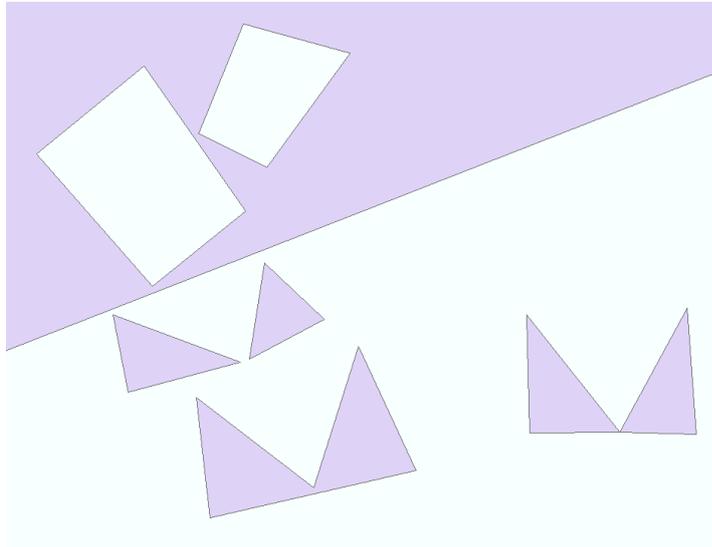


Abbildung 17: Verschiedene Anwendungsfälle der Mindestdimensionen

11. Anpassung an die Grunddaten: Die einzelnen Stützpunkte der Polygone müssen an entsprechende Geometrien in den Grunddaten gesnappt sein. In diesem Fall müssen die Stützpunkte entweder an Punkte, Linien oder Flurstücke der Automatisierten Liegenschaftskarte gesnappt sein (siehe 2.4.3.1).

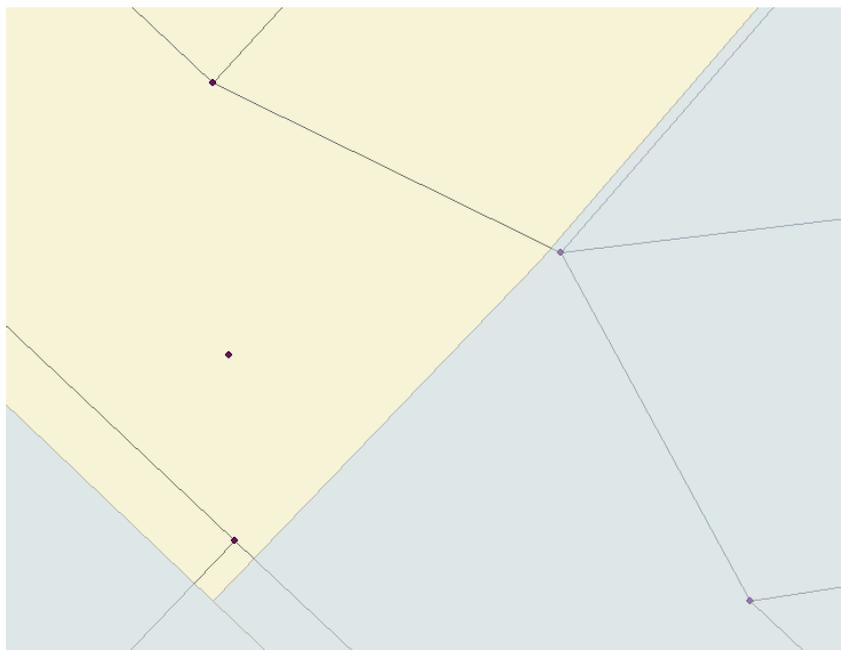


Abbildung 18: WSG-Zonen (blau) sind nicht an die ALK gesnappt

Des Weiteren sollen folgende topologische Fehler gefunden werden:

1. Überlappungen: Zwei Polygone dürfen sich nicht überlagern.



Abbildung 19: Sich überlappende Polygone

2. Lücken: Die Kanten zweier Polygone müssen aneinander anliegen, es darf keine Lücken geben, unabhängig davon ob die Lücken einen Teil der gemeinsamen Kante trennen oder ob die Lücken die beiden Kanten vollständig voneinander trennen.

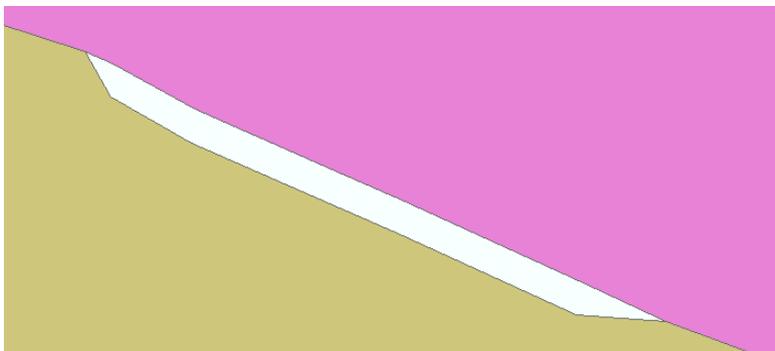


Abbildung 20: Lücke zwischen zwei Polygonen

3. Sliver Polygone: Hierbei handelt es sich um kleinste, sehr schmale Lücken und Überlagerungen an der Grenze zweier Polygone.

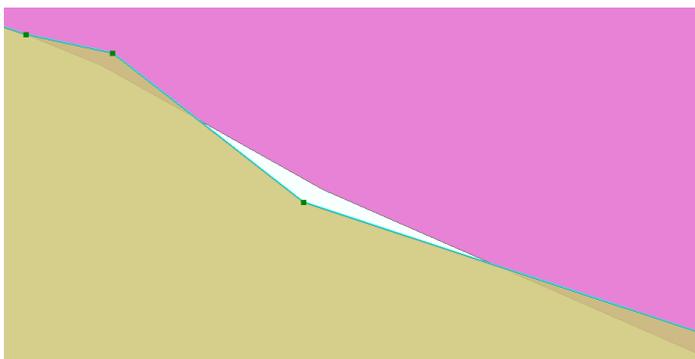


Abbildung 21: Sliver Polygone

4. Klaffen: Klaffen sind eine spezielle Art von Lücken, sie treten an den äußersten Punkten, aneinander angrenzender Kanten zweier Polygone auf.

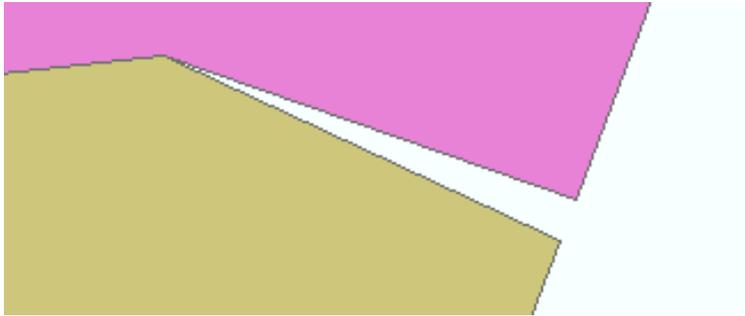


Abbildung 22: Klaffe zwischen zwei Polygonen

5. Anpassung an benachbarte Geometrien: Bei aneinander angrenzenden Polygonen müssen die Stützpunkte eines Polygons an die des anderen gesnappt sein, andernfalls können minimale Lücken und Überlagerungen entstehen.

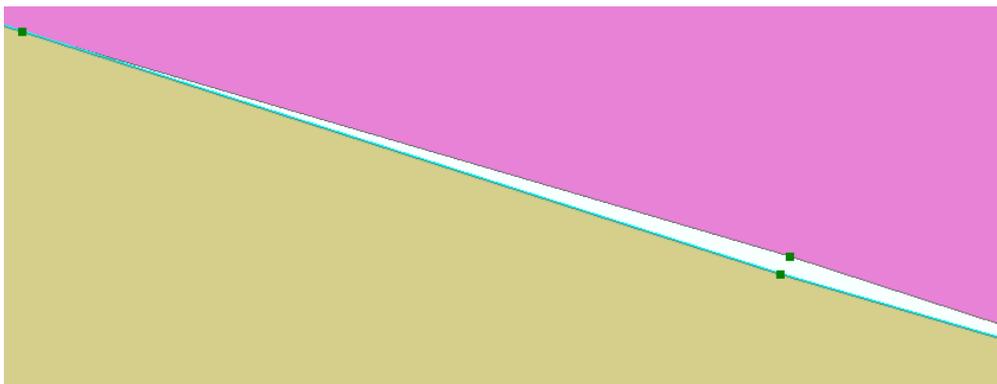


Abbildung 23: Stützpunkte eines Polygons sind nicht an das benachbarte gesnappt

4. Qualität von Geodaten

In diesem Kapitel wird die Qualität von Geodaten behandelt. Es sollen Qualitätskriterien für raumbezogene Daten ermittelt werden. Ferner sollen Geodatenstandards analysiert werden und Ursachen für Qualitätsmängel betrachtet werden.

4.1 Kriterien

Die Qualität von Geodaten lässt sich von zwei Seiten aus betrachten, zum einen gibt es die interne Qualität, die an der Richtigkeit des Datensatzes gemessen werden kann, zum anderen gibt es die externe Qualität, die an den Ansprüchen und Bedürfnissen des Nutzers gemessen werden kann [DEV06, S. 36]. Während sich die interne Qualität anhand von objektiven Kriterien messen lässt, hängt die externe Qualität sowohl von subjektiven Eindrücken als auch vom aktuellen Verwendungszweck ab. Um die bestmögliche Datenqualität zu erhalten muss gewährleistet sein, dass sowohl die interne als auch die externe Qualität so hoch wie möglich sind [ISTL]. Die Abbildung 24 stellt das Verhältnis von subjektiver und objektiver Bewertung (Assessment) dar. Während der Quadrant I den schlechtesten möglichen Fall darstellt, stellt Quadrant IV den besten möglichen Fall dar. Hier sind sowohl subjektive als auch objektive Kriterien „gut“. In den Quadranten II und III ist zumindest jeweils eine der Bewertungen gut, was die bewerteten Daten zumindest teilweise brauchbar macht.

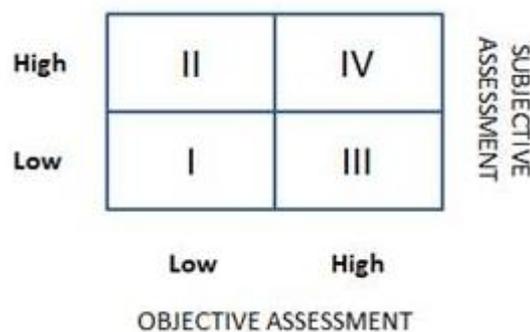


Abbildung 24: Verhältnis von objektiven und Subjektiven Qualitätskriterien (Quelle: [ISTL])

4.1.1 Interne Qualität

Eine hohe interne Qualität lässt sich anhand von verschiedenen Parametern feststellen. Ein solcher Parameter ist die Aktualität der Daten. Aufgrund der sich ständig verändernden realen Vorlagen von Geodaten, ist es wichtig, dass die Daten auf dem neuesten Stand sind, da Analysen mit veralteten Datensätzen zu weniger verlässlichen oder nicht brauchbaren Ergebnissen führen. In der Kartographie gilt so beispielsweise, dass eine Karte zum Zeitpunkt ihrer Fertigstellung bereits veraltet ist, da sich ihre Datengrundlage schon wieder verändert hat. Jedoch ist dies auch vom jeweiligen Thema abhängig, da sich nicht alle Daten gleich schnell und stark verändern.

Ein weiteres wichtiges Kriterium ist die Richtigkeit, wobei die Richtigkeit im Fall von raumbezogenen Daten von drei Faktoren abhängt. Der bedeutendste Faktor ist die räumliche Richtigkeit. Sie ist abhängig von der Abweichung der digitalen Koordinaten von ihren realen Vorbildern, also der absoluten Lage der Geometrieobjekte im Koordinatensystem und der relativen Lage der Objekte zueinander. Die zeitliche Richtigkeit bezieht sich auf die Tatsache, dass Geodaten schnell an Aktualität verlieren können. Allerdings können auch nicht aktuelle Daten sehr nützlich sein, wenn sie sich einem bestimmten Zeitpunkt zuordnen lassen, um zum Beispiel Entwicklungen darzustellen. Die Aktualität ist insofern ein Teilaspekt der zeitlichen Richtigkeit, dass Daten nur mit den entsprechenden Angaben auch sicher als aktuell gelten können [GEOG, S. 179 ff]. Ohne die entsprechenden Informationen kann die Aktualität nicht garantiert werden. Der dritte Faktor ist die thematische Richtigkeit, sie bezieht sich einerseits auf die Richtigkeit der Attributdaten, andererseits bezieht sie sich auch darauf, ob die richtigen, für das Thema notwendigen, Geometrien im Datensatz enthalten sind.

Das nächste Kriterium, die Genauigkeit, ist eng mit der Richtigkeit verknüpft. Die Genauigkeit beinhaltet ebenfalls räumliche, zeitliche und thematische Aspekte und drückt die Bearbeitungstiefe der Daten aus. Sie wird durch die technischen Möglichkeiten bei der Datenaufnahme begrenzt. Auch die steigende Menge der Daten bei höherer Genauigkeit ist ein Aspekt der

sowohl bei der Speicherung als auch bei der Verarbeitung beachtet werden muss. Wird bei abnehmender Genauigkeit der Daten eine bestimmte Toleranz überschritten, ist auch die Richtigkeit der Daten nicht mehr gewährleistet. Die räumliche Genauigkeit kann durch eine höhere Dichte an aufgenommenen Punkten erhöht werden, wodurch die Daten auch in größeren Maßstabsbereichen dargestellt werden können und ein höherer Detaillierungsgrad sichtbar wird. Die zeitliche Genauigkeit kann durch häufigere Datenaufnahmen und Aktualisierungen gesteigert werden. Dadurch können Veränderungen besser aufgezeigt werden und damit die Aktualität gesteigert werden. Die thematische Genauigkeit bezieht sich auf die Attributwerte, bei qualitativen Attributen beschreibt sie die Detailliertheit der Klassifizierung, wie die Anzahl der Klassen, nach denen die Daten eingeteilt wurden, bei quantitativen Attributen die Anzahl der Nachkommastellen oder die Stärke der Rundung von Werten.

Auch eine hohe Genauigkeit garantiert jedoch keine Richtigkeit, während weniger genaue Daten richtiger sein können, als genauere Daten (siehe Abbildung 25).

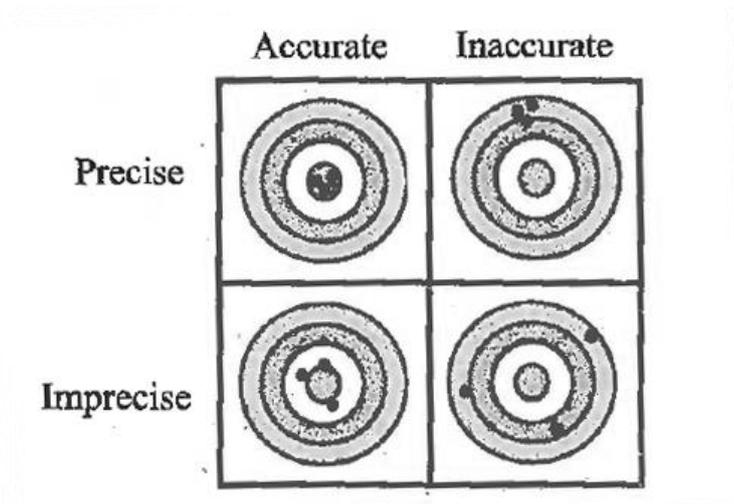


Abbildung 25: Vergleich von Genauigkeit (Precision) und Richtigkeit (Accuracy) (Quelle: [DEV06, S. 184])

Ein weiteres Kriterium ist die Vollständigkeit, sie bezieht sich in erster Linie auf die Attribute des Datensatzes. Hier kann zwischen Pflichtattributen, optionalen Attributen und nicht durchgehend anwendbaren Attributen unterschieden werden [LOSH, S. 9]. Pflichtattribute dienen oftmals als

Primärschlüssel zur Unterscheidung der einzelnen Datensätze oder sie sind thematisch vorgegeben. Optionale Attribute können zwar in jedem Datensatz angegeben werden, sind aber nicht zwingend vorgegeben, wobei auch bei den optionalen Angaben gilt, dass vollständigere Daten die Qualität der gesamten Daten erhöhen und sie für weitere Verwendungszwecke brauchbar machen können. Die nicht durchgehend anwendbaren Attribute beinhalten Werte, die nicht für alle Zeilen erfasst werden können und deswegen in einigen Zeilen nicht angegeben werden können. Neben den Sachdaten müssen auch die Geometrien alle Gebiete, die für das dargestellte Thema notwendig sind, abdecken.

Das Kriterium der Konsistenz beschreibt den gedanklichen Zusammenhang der Daten und deren Vereinbarkeit miteinander, dies umfasst sowohl räumliche, zeitliche und inhaltliche Aspekte der Daten selbst als auch die Anpassung der Daten an den gesamten Aufbau der Datenstruktur. Dies ist vor allem bei Daten aus verschiedenen Quellen von großer Bedeutung [ISTL], da es hier zu Abweichungen bei der Datenaufnahme und Datenverarbeitung kommen kann. Die hat zur Folge, dass Geometrien nicht übereinstimmen und zum Beispiel Lücken aufweisen, Daten verschieden aktuell sind oder Sachdaten mit unterschiedlicher Genauigkeit aufgenommen und wiedergegeben werden oder sich komplett unterscheiden. Hinzu kommt bei Daten aus verschiedenen Quellen ein unterschiedlicher Aufbau, der an eine einheitliche Datenstruktur angepasst werden muss. Konsistenz existiert sowohl innerhalb eines Datensatzes als auch zwischen verschiedenen Datensätzen, außerdem müssen Daten für verschiedene Zeitpunkte konsistent sein [LOSH, S. 9]. Um konsistente Geodaten zu erhalten ist es wichtig, Bedingungen festzulegen, darunter fallen auch topologische Regeln [DEV06, S. 191], mit denen die Anordnung von Geometrien geregelt wird. Die Konsistenz vereinfacht die Bearbeitung von Geodaten, indem sie sie vereinheitlicht und so ihre Übersichtlichkeit erhöht.

4.1.2 Externe Qualität

Die externe Qualität von Geodaten bezieht sich auf deren Gebrauchstauglichkeit. Diese hängt von vielen Faktoren ab, was zur Folge hat, dass Geodaten für einen Nutzer sehr gut zu gebrauchen sind, während die gleichen Daten für einen anderen Nutzer nutzlos sind [DEV06, S. 39]. Dies kann verschiedene Gründe, wie die zu Verfügung stehenden Programme oder die Erfahrung im Umgang mit Geodaten, haben.

Die externe Qualität ist eng mit der internen Qualität verknüpft, da eine steigende interne Qualität die Geodaten für mehr Aufgaben nutzbar macht, aber ihrer Gebrauchstauglichkeit auch schaden kann, wenn Datensätze zu groß werden um, abhängig von den Hard- und Softwarevoraussetzungen der Nutzer, verarbeitet werden zu können. Jedoch ist zu beachten, dass Daten nicht nur für einen einzigen Zweck brauchbar sind, da von mehrfach verwendbaren Daten eine zukünftige Zeit- und Kostenersparnis ausgeht.

An erster Stelle steht der Verwendungszweck der Daten. Von ihm hängt ab, in welchem Umfang interne Qualitätskriterien wie Aktualität, Genauigkeit und Vollständigkeit mindestens erfüllt sein müssen. So ist die benötigte Aktualität oft davon abhängig, wie schnell und stark sich die realen Vorbilder der Daten ändern, hinzu kommt noch der zu bearbeitende Zeitpunkt. Die benötigte Genauigkeit hängt davon ab, welcher Maßstab für die Darstellung und Auswertung der Daten sinnvoll ist. Die benötigte Vollständigkeit hängt davon ab, welche Teile des Datensatzes verarbeitet werden sollen und ob sie zwingend notwendig oder Zusatzinformationen sind. Wichtig ist außerdem, dass alle Daten miteinander vereinbar sind, da auch einzelne Abweichungen zu verfälschten Ergebnissen führen. Außerdem sind neue Datensätze schneller zu gebrauchen, wenn sie mit den bereits vorhandenen Daten konsistent sind und der Bearbeiter keine zusätzlichen Arbeitsschritte durchführen muss.

Für den Nutzer können auch weitere subjektive Kriterien wie die Aufbereitung und Darstellung der Informationen wichtig sein, dies hängt auch mit seiner Erfahrung im Umgang mit Geodaten zusammen.

Abbildung 26 fasst den Zusammenhang von internen und externen Qualitätskriterien zusammen. Während für die Bewertung der Internen Qualität einmalig feste Vorschriften festgelegt werden können, gibt es bei der externen Qualität so viele Bewertungskriterien wie es Nutzer gibt.

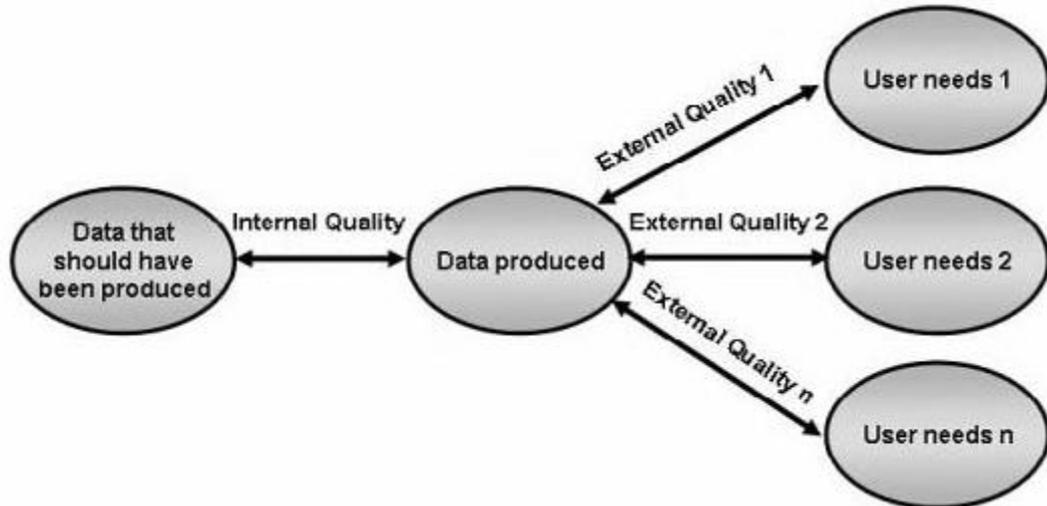


Abbildung 26: Übersicht über interne und externe Datenqualität (Quelle: [DEV06, S. 36])

4.1.3 Metadaten

Die Verlässlichkeit der Daten spielt ebenfalls eine große Rolle. Neben ihrer Quelle werden auch Informationen über die Datenaufnahme und ihren Aufbau benötigt. Um diese Informationen zu vermitteln werden Metadaten verwendet. Bei Metadaten handelt es sich um zusätzliche Daten, welche die eigentlichen Daten beschreiben. Diese Zusatzinformationen geben Auskunft über die Daten in Bezug auf Herkunft, Aufbau, Schema, Projektion, Ausdehnung und Nutzung. Außerdem werden in ihnen auch Qualitätsinformationen wie Stand, Genauigkeit und Konsistenz dokumentiert. Metadaten nehmen eine besondere Rolle unter den Qualitätskriterien ein, da sie Daten zur internen Qualität enthalten und mit ihren Informationen Geodaten leichter analysiert und bearbeitet werden können, womit sie auch zur Gebrauchstauglichkeit beitragen. Aufgrund dieser besonderen Rolle wurden für Metadaten von raumbezogenen Daten auch gesetzliche Standards, wie die INSPIRE-Richtlinie (siehe Kapitel 4.2) eingeführt. Sie

sollen gewährleisten, dass die Metadaten selbst eine hohe Qualität aufweisen.

4.2 Geodatenstandards und Gesetze

Für Geodaten existieren verschiedene Standards und gesetzliche Vorgaben, diese sollen zu einer einheitlich hohen Datenqualität beitragen. Durch die Einhaltung solcher Standards wird die Kompatibilität von Geodaten verschiedener Quellen erhöht. Für Geodaten existieren Standards der International Organisation for Standardization (ISO), die mit den Nummern ISO 191xx gekennzeichnet sind [WIKI_2]. Speziell für die Geodatenqualität existiert der Standard ISO 19113, der Qualitätskriterien wie Vollständigkeit, Konsistenz sowie räumliche, zeitliche und thematische Genauigkeit festlegt [WIKI_3]. Weitere ISO Geodatenstandards beinhalten unter anderem Richtlinien für Metadaten, Web Map und Web Feature Services oder die Geographic Markup Language (GML).

Um innerhalb der Europäischen Union einheitliche Geodatenstandards zu erreichen, wurde im Jahr 2007 die „Richtlinie zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE)“ verabschiedet. Durch INSPIRE werden Behörden innerhalb der EU dazu verpflichtet, ihre vorhandenen Geodaten entsprechend den INSPIRE-Richtlinien aufzubereiten und verfügbar zu machen, dies betrifft alle Behörden auf nationaler, regionaler und lokaler Ebene. [WIKI_4]. Durch die vereinheitlichten Standards und Vorgaben für Geodaten und Geodatendienste, sowie deren Metadaten, soll ermöglicht werden, dass raumbezogene Daten europaweit verfügbar sind und verwendet werden können, ohne dass der jeweilige Nutzer die Daten erst für sein Land anpassen muss. Außerdem kann der Nutzer davon ausgehen, dass die Daten über eine entsprechende Qualität verfügen.

In Deutschland wird INSPIRE mittels des Geodatenzugangsgesetzes umgesetzt, in dem der Aufbau der Geodateninfrastruktur Deutschland (GDI-DE) geregelt ist. Die Geodaten des RIPS sind als Teil der Geodateninfrastruktur Baden-Württemberg (GDI-BW) innerhalb der GDI-DE

ebenfalls an die Vorgaben der INSPIRE-Richtlinie angepasst (siehe Abbildung 27).

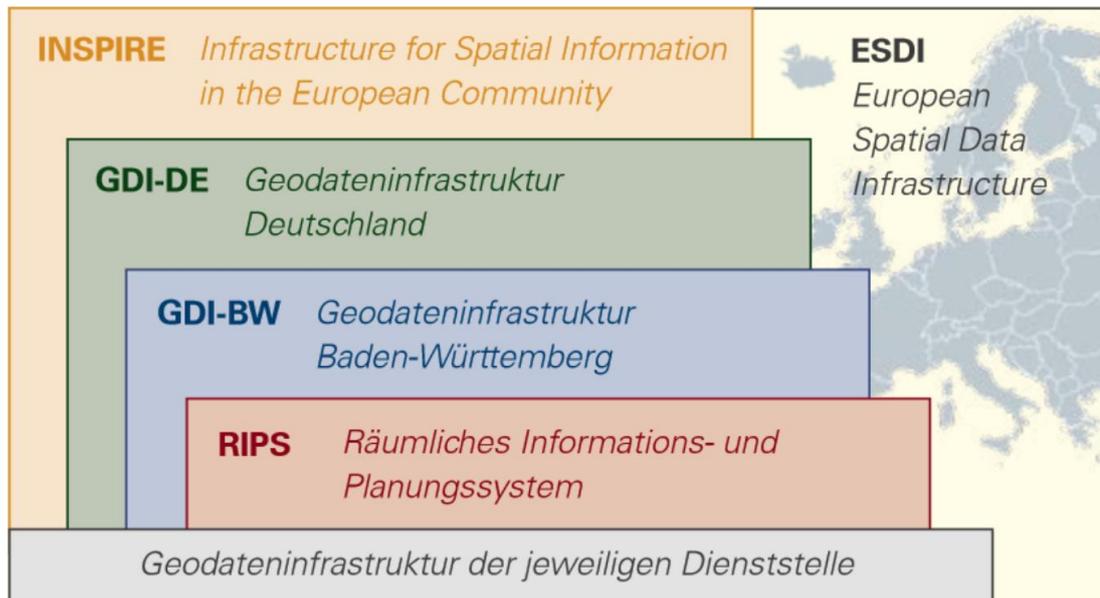


Abbildung 27: Das RIPS als Teil der INSPIRE-Richtlinie (Quelle: [RIPS_1])

4.3 Ursachen für Fehler

Fehler in Geodaten führen unweigerlich zu einer schlechteren internen und externen Datenqualität. Ursachen hierfür finden sich in allen Bereichen der Datenaufnahme, -verarbeitung und -haltung. In vielen Fällen sind nachlässig vorgehende Bearbeiter oder fehlerhafte Technik für Fehler verantwortlich, aber auch die mangelnde Qualität der Quelldaten und in ihnen enthaltene Fehler pflanzen sich in neue Datensätze fort (siehe Abbildung 28).

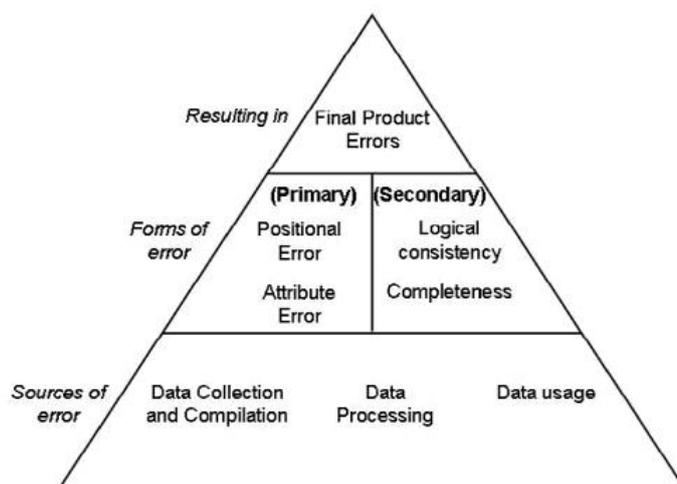


Abbildung 28: Fehlerursachen und Fortpflanzung (Quelle: [DEV06, S. 34])

Gerade im technischen Bereich können Fehler auf viele Arten entstehen. Dies betrifft nicht nur defekte Aufnahmegeräte oder Software-Bugs, auch richtig funktionierende Programme können Fehler verursachen, was zum Teil auch an Bedienungsfehlern liegt. Auch die undurchsichtige Datenstruktur von Datenbanken und Verarbeitungssoftware ist ein Faktor, der Probleme verursacht. Zusätzlich haben verschiedene Softwareprodukte verschiedene interne Standards, wodurch sie verschiedene Datenstrukturen als richtig ansehen. Die hat zur Folge, dass bei Prozessen, in denen verschiedene Softwares angewendet werden, Daten entsprechend den Vorgaben einer Software verarbeitet werden. Das nächste Programm erkennt diese Daten jedoch als fehlerhafte Daten und verarbeitet sie nicht oder nicht vollständig weiter, da beispielsweise fehlerhafte Features von der Weiterverarbeitung ausgeschlossen werden. Neben der Datenstruktur unterscheiden sich auch Toleranzen und die Genauigkeit in verschiedenen Programmen und Datenformaten. Dies kann beispielsweise zur Folge haben, dass zwei Datensätze, die durch verschiedene Softwares verarbeitet wurden, bei einer Verschneidung der beiden kleinste Splitterpolygone erzeugen. Das Shapefile-Format ist beispielsweise toleranter gegenüber fehlerhaften Daten als das File-Geodatabase-Format, da es keine XY-Toleranzen oder topologische Regeln speichern kann.

4.4.1 Qualitätskriterien und Fehlerursachen am Beispiel Wasserschutzgebiete

Die Geometrien der Wasserschutzgebiete durchlaufen keinen einheitlichen Entstehungsprozess, sie werden von den zuständigen unteren Wasserbehörden erfasst und digitalisiert, wobei hierfür in manchen Fällen auch Ingenieurbüros beauftragt werden. Durch die große Anzahl verschiedener Bearbeiter ergibt sich eine weite Spanne unterschiedlicher Datenqualitäten. Zusätzlich werden die Daten mit verschiedenen GIS-Softwaresysteme wie ESRI ArcGIS oder Disy GISterm digitalisiert und in die jeweiligen Datenbanken eingespeist. In der nachfolgenden Tabelle sind

verschiedene Geometrie- und Topologiefehler aufgeführt, die gegen die Qualitätsstandards der LUBW verstoßen, sowie mögliche Ursachen dieser Fehler:

Fehlertyp	Ursache(n)
Doppelt oder zu nah gesetzte Stützpunkte	Beim Digitalisieren wurde unsauber gearbeitet. Können auch beim Vereinigen mehrerer Zonen zu einem Umring entstehen.
Polygone sind nicht geschlossen	Die Geometrie wird als Polylinie anstatt als Polygon angelegt.
Orientierung von inneren und äußeren Ringen ist identisch	Ein möglicher Software-Fehler, da Programme wie ArcGIS, beim Digitalisieren die Orientierung automatisch anpassen.
Der Datensatz enthält Splitterflächen	Datensätze mit unterschiedlicher Genauigkeit oder Qualität werden miteinander verschnitten
Geometrien haben Löcher	Löcher können gewollt sein, aber auch durch Fehler bei Verschneidungen entstehen.
Multipart-Polygone berühren sich selbst in einem Punkt	Fehler beim Digitalisieren, z.B. durch falsch eingestelltes Snapping.
Polygone berühren sich selbst oder überschneiden sich mit sich selbst	Fehler beim Digitalisieren, z.B. durch falsch eingestelltes Snapping oder Digitalisieren der Stützpunkte in der falschen Reihenfolge. Es ist softwareabhängig, ob es als Selbstüberschneidung oder Berührung gespeichert wird.
Polygone haben Spikes	Folgen von Verschneidungen mit nicht genau zueinander passenden Flächen.
Minstdimensionen werden nicht eingehalten	Fehler beim Digitalisieren.
Multipart Geometrien	Multiparts können beabsichtigt sein, aber sie können auch durch Splitterflächen entstehen.
Flächen (WSG-Zonen) liegen nicht exakt aneinander an	Entstehen durch falsche Snapping-Einstellungen oder unsauberes Digitalisieren
Lücken und Überlappungen zwischen Geometrien bzw. Sliver-Polygone, sowie Klaffen zwischen zwei Polygonen	Folge von nicht exakt aneinander liegenden Flächen Bei der Erstellung des Umrings entstehen durch diese Lücken Löcher.
Geometrien sind nicht an die Grunddaten angepasst	Entstehen durch unsauberes Digitalisieren oder falsch eingestelltes Snapping.

Tabelle 1: Fehlerursachen der WSG-Geometrien

Anhand der Tabelle 1 wird deutlich, dass viele Fehler durch falsche Programmeinstellungen und unsorgfältiges Arbeiten verursacht werden. So können Fehler bei Verschneidungen beispielsweise durch versehentliches Verschieben eines Polygons entstehen oder durch den Einsatz falscher oder unvollständiger Grunddaten, wie einer nicht mehr aktuellen ALK.

Eine weitere Schwierigkeit ergibt sich an den Grenzen von Verwaltungseinheiten. Da für jeden Landkreis eine andere untere Wasserbehörde zuständig ist, kommt es an den Grenzen zu Konflikten in den Daten, da beim Digitalisierungsprozess oft nur die Daten des eigenen Landkreises beachtet werden (siehe Abbildung 29).

Die geometrische und topologische Richtigkeit der Wasserschutzgebiete ist als messbare Größe zwar ein internes Qualitätskriterium, aber sie betrifft auch die externe Qualität, da korrupte Geometrien von einigen Softwares nicht verarbeitet werden können und somit nicht brauchbar sind. Auch die Auswertung der Daten ist davon betroffen, topologische Fehler beeinträchtigen beispielsweise die Analyse der Flächen der WSG. Die Einhaltung vieler interner Qualitätskriterien hängt von den unteren Wasserbehörden ab, da diese die Verantwortung für die Richtigkeit und Aktualität tragen und Veränderungen oder neue Wasserschutzgebiete aufnehmen und weitergeben müssen. Auch die Vollständigkeit hängt von der Zuarbeit der UWB ab.



Abbildung 29: Zwei sich überlagernde Wasserschutzgebiete

5 Qualitätssicherungsprozesse in der LUBW

Dieses Kapitel soll die aktuellen organisatorischen und technischen Qualitätssicherungsprozesse im ITZ beschreiben und analysieren. Hierbei werden in erster Linie die Wasserschutzgebiete als Beispiel verwendet.

In allen Fachbereichen sind die Hauptprobleme der Qualitätssicherung die verschiedenen Datenquellen. Um den Dateneingang zu vereinheitlichen, werden die Daten über Fachanwendungen in die Datenbanken eingespeist. Die Programme verfügen bereits über Funktionen zur Validierung der Daten und können so Fehler bereits vor der Verarbeitung im ITZ abfangen, Fachanwendungen wie das speziell für wasserwirtschaftliche Gebiete entwickelte WawiG verwenden hierfür die integrierte GIS-Software GISterm, welche einen Teil der Fehler abfangen kann. Weitere Fehler werden beim Einlesen in die Datenbanken abgefangen. Des Weiteren werden von Seiten des ITZ auch Schulungen für die Erfasser angeboten, in denen der Umgang mit den Fachanwendungen und die Anforderungen an die Daten erläutert werden. Es besteht außerdem eine enge Zusammenarbeit zwischen den einzelnen Fachbereichen und den Entwicklern im ITZ, so werden auch bereichsübergreifend Qualitätskriterien und Softwareanforderungen festgelegt.

5.1 Qualitätssicherungsprozess am Beispiel Wasserschutzgebiete

Der aktuelle Qualitätssicherungsprozess der Wasserschutzgebiete ist ein komplexer Kreislauf, der sowohl das ITZ der LUBW als auch die unteren Wasserbehörden mit einschließt. Dadurch ist eine große Anzahl Personen, verteilt auf 45 Behörden, involviert. Durch die Vielzahl an beteiligten Stellen erhöht sich auch die Anzahl der für die Qualitätssicherung verwendeten Programme und Datenbanken, da nicht alle UWB über die gleiche Softwareinfrastruktur verfügen.

Der Kreislauf beginnt bei den unteren Wasserbehörden, wo die WSG-Geometrien, mit Softwares wie ArcGIS, WawiG, GISterm oder dem GEO Shapeimport in die jeweiligen Datenbanken importiert werden. Mittels des Datenaustauschdienstes (DAD) werden monatlich Datensätze, die neu in die UWB-Datenbanken aufgenommen wurden, in die UIS-Referenzdatenbank übertragen. Mit FME erfolgt die automatische Übertragung der Daten in das GEO-Schema der UIS SDE-Datenbank, hierbei erfolgt auch eine Geometrie-Validierung, bei der korrupte Geometrien ausgeschlossen werden. Für die Daten der SDE-Datenbank wird, ebenfalls mit FME, ein Topologiecheck durchgeführt. Die so gefundenen geometrischen und topologischen Fehler werden im nächsten Schritt manuell von Mitarbeitern des ITZ korrigiert. Hierbei werden die ArcUIS-Tools eingesetzt, eine vom ITZ erstellte ArcGIS-Erweiterung, die dem Bearbeiter topologische Fehler wie fehlende Kantenidentität oder Planarität anzeigt [ARCUIS, S. 22 f.]. Nach der Korrektur werden die Daten testweise in die Datenbanken eingelesen und überprüft, bis diese vollständig valide sind. Die korrigierten Geometrien werden als E-Mail an die UWB versandt, wo sie erneut in die lokalen Datenbanken importiert werden.

Dieser Prozess ist innerhalb des ITZ aktuell sehr arbeits- und zeitaufwändig, wobei die Verantwortung für die Qualität der Wasserschutzgebietsdaten bei den UWB liegt. Die Ursachen vieler Qualitätsmängel liegen oft bei den Wasserbehörden. Deswegen werden von Mitarbeitern der LUBW auch regelmäßig Schulungen angeboten, um die Mitarbeiter der UWB bei der Qualitätssicherung zu unterstützen. Auch bei den UWB werden bereits Datenvalidierungen vorgenommen, bei den Datenbankimporten werden die Geometrien durch ArcGIS, GISterm und WawiG automatisch überprüft. Diese Tests sind oft zu tolerant, wodurch fehlerhafte Geometrien in die Datenbanken gelangen. Aus diesem Grund wird mit FME eine zusätzliche Validierung durchgeführt, da mit FME eine genauere Datenvalidierung möglich ist.

6 Software-Evaluierung für die Umsetzung der Arbeit

In diesem Kapitel werden verschiedene Softwares zur Geodatenverarbeitung vorgestellt, deren Eignung für die Umsetzung dieser Bachelorarbeit untersucht wurde. Das Hauptaugenmerk wurde auf die technischen Voraussetzungen der einzelnen Programme gelegt. Hierbei wurden sowohl kostenpflichtige als auch Open Source Softwares untersucht. Neben klassischer Geoinformationssoftware wurden eine Java-Library und zwei Spatial ETL Produkte untersucht. Spatial ETL (Extract, Transform, Load) bezeichnet Programme, die raumbezogene Daten aus Datenbanken oder Datensätzen extrahieren, sie dann so weiterverarbeiten, wie der Nutzer sie benötigt und sie in einem passenden Format abspeichern [SAFE_2, S. 1]. Die wichtigsten Bewertungskriterien sind der angebotene Funktionsumfang, einschließlich Funktionen zur Automatisierung sowie die Bedienbarkeit, vor allem für neue Benutzer der Software.

6.1 ESRI ArcGIS

ArcGIS ist eine kostenpflichtige GIS-Software, die aus mehreren Programmen wie dem Darstellungsprogramm ArcMap oder dem Datenverwaltungsprogramm ArcCatalog besteht. ArcGIS wird von der Firma ESRI herausgegeben und ist aktuell in der Version 10.3 erhältlich. ArcGIS enthält Tools zur Verarbeitung und Analyse von Geodaten. Welche Tools für den Nutzer verfügbar sind, hängt von der Lizenzierungsstufe und den installierten Erweiterungen ab. Mit Hilfe des integrierten ModelBuilder (siehe Abbildung 30) lassen sich die einzelnen Tools miteinander kombinieren, wodurch eigene Geoverarbeitungsprozesse erstellt werden, die wiederholt verwendet werden können [ESRI_2]. Zusätzlich können in ArcGIS Python-Skripte zur Datenverarbeitung eingesetzt werden, sowohl als Ersatz für Tools als auch im ModelBuilder. Im Bereich der Vektordaten beschränkt sich ArcGIS auf GIS-Formate wie Shapefile sowie auf Geodatenbankformate.

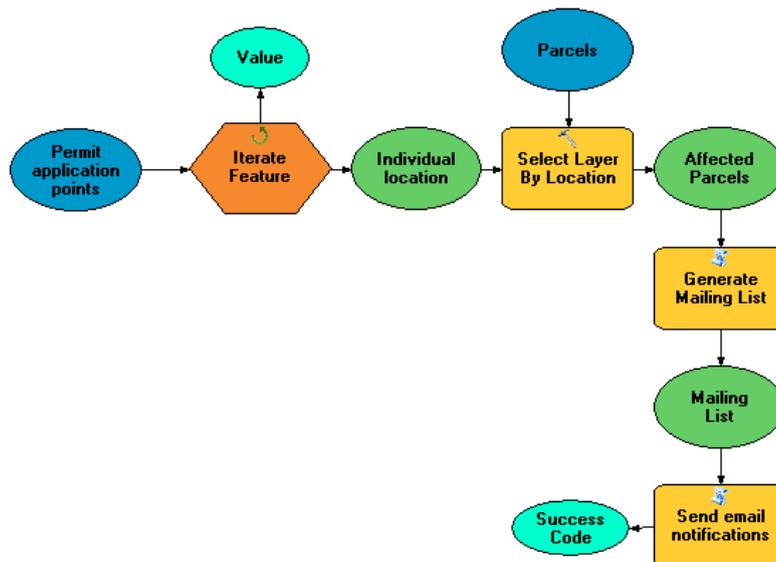


Abbildung 30: Beispielhafter Prozess mit dem ModelBuilder (Quelle: [ESRI_2])

Dank der großen Auswahl an Tools ist der ModelBuilder gut für die Umsetzung der Arbeit geeignet. Aber viele Tools, darunter auch grundlegende Verschneidungsoperationen, benötigen eine hohe, kostspielige Lizenzierungsstufe.

Ein mit dem ModelBuilder erstelltes Model lässt sich leicht bedienen und kann wie jedes andere Tool von ArcGIS eingesetzt werden, dadurch ist es sehr flexibel einsetzbar und kann alle Inputdaten verarbeiten, die das passende Datenformat besitzen.

6.2 GeoTools

GeoTools ist eine Open Source Java-Library, welche eine Sammlung verschiedener Klassen der Programmiersprache Java enthält, die zur Verarbeitung, Darstellung und Manipulation von Geodaten verwendet werden können [GEOT_1]. Aktuell ist die Version 14.x verfügbar. GeoTools bietet Klassen, mit denen es möglich ist, Raster- und Vektordaten, Daten in Datenbanken, sowie Daten in XML basierten Sprachen zu verarbeiten. Zu den unterstützten Formaten gehören unter anderem Oracle Spatial und Shapefile. Außerdem halten die Werkzeuge von Geotools die vom OGC vorgegebenen Standards ein. Neben den bereits vorhandenen Klassen und Methoden besteht die Möglichkeit, mit verschiedenen Extensions die

Funktionen von GeoTools zu erweitern [GEOT_2]. Dadurch ist der Nutzer in der Lage, mit Java eigene Software zur Geodatenverarbeitung zu schreiben. Die GeoTools-Library verfügt bereits über Klassen zur Validierung von Topologien, jedoch ist die Auswahl sehr eingeschränkt und viele Überprüfungen, die der fertige Prototyp ausführen soll, wären damit nicht durchführbar.

6.3 GeoKettle

GeoKettle ist eine quelloffene Spatial ETL Software, welche auf Kettle, einer ETL Software für nicht raumbezogene Daten, basiert, aktuell ist die Version 2.5 verfügbar [GEOK_1]. Auf einer graphischen Oberfläche kann der Nutzer Transformationen und Jobs erstellen (siehe Abbildung 31). Transformationen verarbeiten Daten, sie sind aus einzelnen Arbeitsschritten (Steps) zusammengesetzt, welche durch Verbindungen (Hops) geordnet werden. Jobs verwalten wiederum die Transformationen. Die Programme zum Ausführen der Jobs und Transformationen werden über die Kommandozeile aufgerufen. Als In- und Output können diverse Geodaten-, Tabellen- und Datenbankformate gewählt werden [GEOK_2].

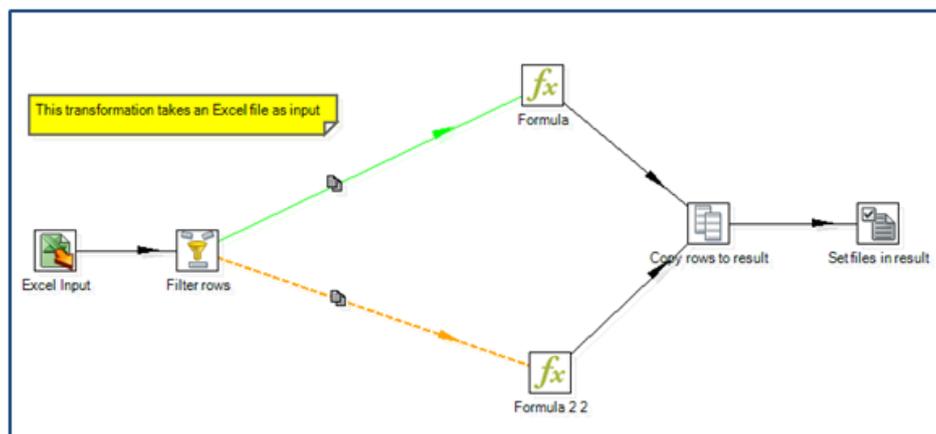


Abbildung 31: Beispielhafte Transformation mit GeoKettle (Quelle: [GEOK_2])

GeoKettle bietet größtenteils Funktionen zur Analyse von Sachdaten. Steps zur Verarbeitung von Geometriedaten machen nur einen sehr geringen Teil aus, vor allem im Bereich der Validierung und Korrektur von Geodaten ist die Step-Auswahl nur sehr gering.

6.4 Safe Software FME

FME (Feature Manipulation Engine) ist eine kostenpflichtige Spatial ETL Software, die von der Firma Safe Software entwickelt wird. Derzeit ist die Version 2015 auf dem Markt. Mit FME ist es möglich, Geometrie- und Sachdaten sowie Daten in Datenbanken und Webservices zu verarbeiten und zu analysieren. Mit der aktuellen Version können 371 Formate gelesen beziehungsweise geschrieben werden, darunter alle gängigen GIS-, CAD-, Raster- und Datenbankformate [SAFE_1]. Diese Flexibilität wird dadurch ermöglicht, dass FME alle Eingangsdaten für die Bearbeitung in ein internes Datenformat konvertiert und erst beim Schreiben der Ausgangsdaten wieder in ein anderes Format umwandelt. FME ist in verschiedenen Versionen für Server und Desktop verfügbar, außerdem gibt es spezielle Versionen für die Arbeit mit Datenbanken oder anderen Programmen wie ESRI ArcGIS. Über eine graphische Benutzeroberfläche, der FME Workbench, ist es möglich, Datentransformationen zu erstellen und auszuführen (siehe Abbildung 32).

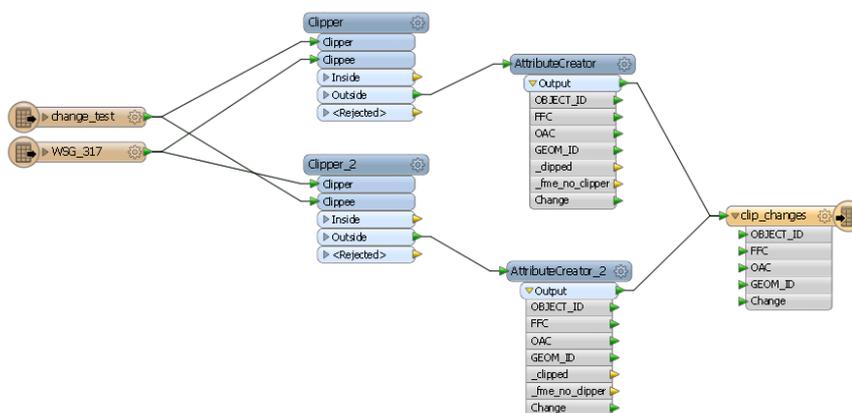


Abbildung 32: Beispielhafte Transformation mit der FME Workbench

Hierfür stehen verschiedene Transformer zur Verfügung, mit denen Geometrie- und Attributdaten bearbeitet werden können. Diese Transformer lassen sich auf der Workbench beliebig anordnen und so zu benutzerdefinierten Transformationsprozessen verbinden. Zusätzlich zu den angebotenen Transformern besteht die Möglichkeit, Python-Skripte zu integrieren. Eine weitere wichtige Komponente ist der FME Data Inspector, mit dem erstellte Daten und Zwischenergebnisse betrachtet werden können.

Für diese Arbeit wurden die FME Desktop Editionen für ESRI und Oracle getestet. Beide Versionen eignen sich gut für die Umsetzung der Arbeit, wobei die Oracle-Edition den Vorteil mit sich bringt, in die vom ITZ verwendeten Oracle Spatial Datenbanken schreiben zu können, während die ESRI-Edition diese nur auslesen kann. Durch die große Auswahl an Transformern und die sich dadurch ergebenden Möglichkeiten zur Datenverarbeitung eignet sich FME gut für die Umsetzung dieser Arbeit. Des Weiteren läuft die Datenverarbeitung mit FME sehr schnell ab. Außerdem verfügt FME über eine einfache und übersichtliche Benutzerführung, wodurch die Software auch für neue Nutzer intuitiv erlernbar ist.

Ein weiterer Vorteil ist die Ausnutzung der Rechnerleistung durch FME. Die Software versucht, falls große Datenmengen verarbeitet werden müssen, selbstständig die Verwendung des Arbeitsspeichers zu optimieren. Zusätzlich bieten einige Transformer die Möglichkeit, den Bearbeitungsvorgang in mehrere Prozesse aufzuteilen [SAFE_3] (siehe Abbildung 33).

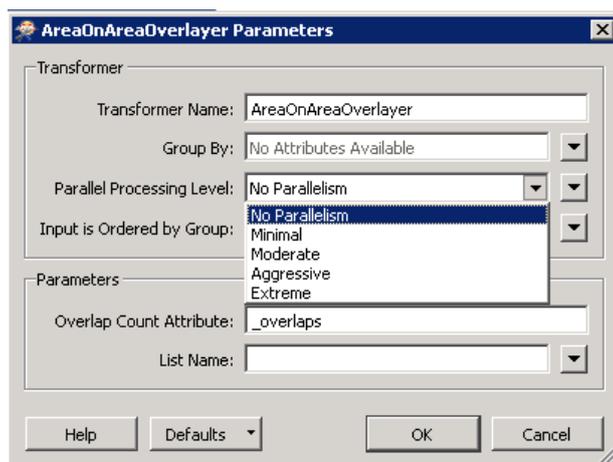


Abbildung 33: Transformer-Fenster mit Parallel Processing Option

6.5 Ergebnis der Evaluierung

Durch ihre große Auswahl an Geoverarbeitungswerkzeugen bieten FME und ArcGIS gute Voraussetzungen, um diese Arbeit umzusetzen, zusätzlich sind die Funktionen in beiden Softwares durch Python-Skripte noch erweiterbar. GeoTools hat ebenfalls eine sehr große Auswahl an Funktionen, aber hier fehlen Klassen, die für die Umsetzung wichtig wären. Bei GeoTools wäre es aber möglich, weitere über die Library hinausgehende Klassen zu programmieren. Die Werkzeugauswahl in Geokettle ist zu klein und nur

bedingt geeignet um den Prototypen damit zu erstellen. ArcGIS und FME sind neben der größeren Werkzeugauswahl auch aufgrund der Möglichkeit mit einer graphischen Oberfläche zu arbeiten gegenüber GeoTools im Vorteil. Gegenüber dem ModelBuilder von ArcGIS ist die FME Workbench, im Vorteil da die Workbench benutzerfreundlicher und intuitiver zu bedienen ist. Außerdem laufen Geoverarbeitungsprozesse mit FME schneller und stabiler ab, was aus Erfahrungswerten abgeleitet wurde. Für ArcGIS spricht jedoch, dass es für neue Benutzer des Prototyps leicht zu erlernen ist, da ArcGIS vielen Nutzern bekannter ist als FME. Da FME auch für neue Benutzer leicht erlernbar ist, kann es diesen Nachteil gegenüber ArcGIS ausgleichen.

	ArcGIS	GeoKettle	GeoTools	FME
Funktionsumfang	Groß, kann mit Python erweitert werden. Viele Tools sind gut für die Umsetzung geeignet.	Nicht ausreichend	Nicht ausreichend, kann mit Java erweitert werden	Groß, kann mit Python erweitert werden. Viele Transformer sind gut für die Umsetzung geeignet.
Bedienbarkeit/Benutzerfreundlichkeit	GIS-Kenntnisse benötigt	Ähnlich wie FME, jedoch komplizierter aufgebaut	Hoher Programmieraufwand	Intuitiv erlernbar
Flexibilität/Automatisierung	Nur für Shapefile und Geodatenbanken geeignet. Automatisierung möglich. Dynamische Abläufe leicht erstellbar. Batch Processing gut möglich.	Automatisierung und Batch Processing möglich	Automatisierung ist möglich, Flexibilität mit erhöhtem Programmieraufwand	Beliebige Formate für In- und Output, Automatisierung möglich. Dynamische Abläufe leicht erstellbar. Batch Processing gut möglich.
In der LUBW eingesetzt?	Ja	Nein	Nein	Ja

Tabelle 2: Abschließender Softwarevergleich

Aufgrund dieser Argumente wurde FME ausgewählt, um diese Arbeit umzusetzen, da es eine große Auswahl an Werkzeugen für die Geodatenverarbeitung bietet. Außerdem ist FME absturzsicherer und es nutzt die technischen Voraussetzungen des Rechners besser aus.

7 Umsetzung

In diesem Kapitel soll die Umsetzung des praktischen Teils dieser Bachelorarbeit beschrieben werden. Da die Umsetzung mit dem Programm FME Workbench, als Teil des FME Softwarepakets, erfolgt, wird zusätzlich der Umgang mit dieser beschrieben.

7.1 Arbeiten mit der FME Workbench

Auf der graphischen Benutzeroberfläche der FME Workbench können durch Platzieren und Verbinden so genannter Reader, Writer und Transformer Prozesse zur Datentransformation und -analyse erstellt werden.

7.1.1 Reader und Writer

Reader sind die Funktionen der Workbench, die Datensätze eines bestimmten Formates lesen können. Auf der Workbench werden die einzelnen Feature Types des Readers dargestellt, welche anschließend weiterverarbeitet werden können. Der Reader selbst ist nur im Navigationsfenster der Benutzeroberfläche dargestellt. Im Reader werden die Daten in das interne FME-Format konvertiert, das für die Transformationen benötigt wird. In Tabelle 3 sind beispielhaft Datentypen von Oracle und die entsprechenden FME-Datentypen aufgelistet

Oracle Data Type	FME Data Type
char(width)	fme_char(width)
float	fme_real32
number(width,decimal)	fme_decimal(width,decimal)
double	fme_real64
integer	fme_int32
logical	fme_boolean
smallint	fme_int16

Tabelle 3: Konvertierung von Oracle Datentypen in FME-interne Datentypen (Quelle: [SAFE_5])

Writer sind die Gegenstücke der Reader, zum Schreiben von Daten. Die Datentypen von Readern und Writern müssen nicht identisch sein und können frei kombiniert werden. Die Writer konvertieren das interne FME-Format beim Schreiben der Outputdaten in das entsprechende Datenformat. Die auf der Workbench dargestellten Reader/Writer Feature Types entsprechen den Input- beziehungsweise Outputdatensätzen des Prozesses (siehe Abbildung 34). Die Feature Types bestehen wiederum aus Features, welche den einzelnen Daten, wie Geometrien oder Zeilen einer Tabelle, entsprechen.

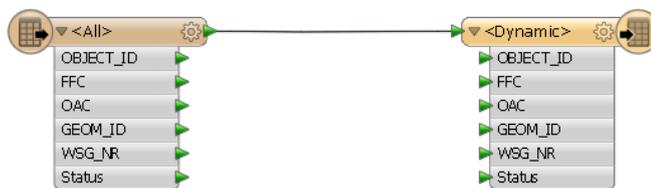


Abbildung 34: Reader und Writer Feature Types mit Attributen

7.1.2 Transformer

Transformer sind die einzelnen Werkzeuge, aus denen die Transformationen, zwischen Reader und Writer bestehen. Transformer führen verschiedene Operationen zur Verarbeitung von Geometrie-, Raster-, Sach- und Attributdaten aus, des Weiteren existieren noch Transformer für Aufgaben wie das Aufrufen von URLs. Sie können die Daten beliebig vieler Reader und anderer Transformer aufnehmen und sie an beliebig viele Transformer und Writer weitergeben. Abhängig von ihren Funktionen verfügen Transformer über verschiedene In- und Outputports (siehe Abbildung 35). Bei jedem Transformer kann der Nutzer über eine Auswahl verschiedener Parameter Einfluss auf dessen Funktionsweise nehmen. Für die Parameter stehen in vielen Fällen Default-Werte zur Verfügung, bei einigen Transformern müssen Parameter durch den Nutzer festgelegt werden, damit er verwendet werden kann.

Es besteht ebenfalls die Möglichkeit, mehrere Transformer mit ihren Parameter-Einstellungen und Verbindungen zu einem Custom Transformer

zu verbinden, wodurch Teile des gesamten Transformationsprozesses wiederverwendet werden können.

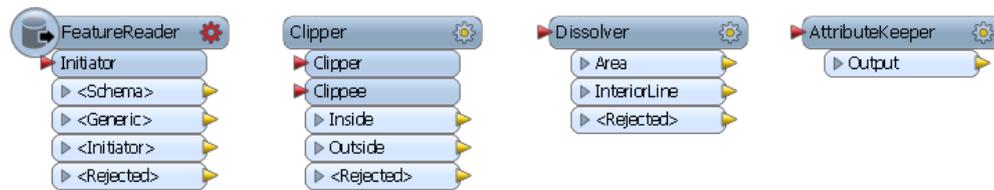


Abbildung 35: Verschiedene Transformer

7.1.3 Attribute und Listen

Die Attribute eines Reader oder Writer Feature Types sowie der Daten an den Output-Ports der Transformer lassen sich durch Ausklappen eines Dropdown-Menüs anzeigen. Attribute werden von den vorherigen Schritten mit übernommen und ihre Werte werden bei gleichbleibendem Attributnamen automatisch weitergegeben, bei einer Namensänderung oder dem Übertragen von Werten auf ein anderes Attribut können durch den Nutzer Verbindungen hergestellt werden.

Listen sind eine Sonderform von Attributen, die mit einem Array vergleichbar ist (siehe Abbildung 36). Sie enthalten mehrere Attributwerte und werden beispielsweise angelegt, wenn eine Geometrie, aus mehreren Geometrien gebildet wird, welche über gleiche Attribute verfügen. Des Weiteren können Listen durch verschiedene Transformer erzeugt werden. Die Listen der neuen Geometriedaten enthalten die Werte aller Ausgangsdaten für dieses Attribut [SAFE_4]. Da Listen ein spezieller Datentyp sind, der nur mit den internen FME-Datenformaten kompatibel ist, können sie nicht direkt in den Outputdaten ausgegeben werden, sondern müssen zuerst in einfache Attribute transformiert werden.

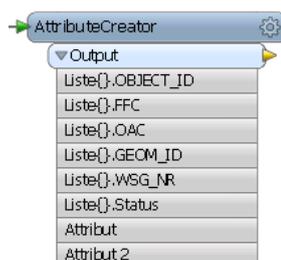


Abbildung 36: Transformer mit Listen und Attributen

7.1.4 User Parameter

User Parameter werden zur Steuerung von Workspaces eingesetzt. So können die Inputs von Readern, Writern und Transformern mit Parametern verknüpft werden. Diese können mit verschiedenen Werten, wie Pfaden, Texten, Zahlen, Passwortabfragen oder auch frei definierbaren Abfragen belegt werden, welche bei der Ausführung der Transformation als Werte für den jeweiligen Input verwendet werden können. Durch den Einsatz von User Parametern können Werte, wie Toleranzen oder die zu verarbeitenden Daten, durch den Nutzer flexibel angepasst werden, die Möglichkeit einen Parameter beliebig oft in einer Transformation zu verwenden, senkt außerdem den Wartungsaufwand, da nicht jeder Wert einzeln aktualisiert werden muss. Über die Auswahlmöglichkeit „Create User Parameter“ in den Transformer-Fenstern lassen sich alle Input-Felder mit Parametern verknüpfen, des Weiteren können User Parameter im Navigationsfenster angelegt und bearbeitet werden. Im Navigationsfenster sind User Parameter durch ein violettes Zahnrad-Icon gekennzeichnet, während Transformer Parameter ein grünes Icon besitzen und Workspace Parameter ein gelbes. Transformer und Workspace Parameter lassen sich jedoch in User Parameter konvertieren.

User Parameter existieren in den Varianten „published“, „private“ und „FME Server“ (siehe Abbildung 37), letztere beinhalten spezielle Informationen für den Einsatz des Workspaces in FME Server. Published Parameter haben den Vorteil, dass sie auch außerhalb ihres Workspaces bearbeitet werden können. Dadurch können Parameter in mehreren Workspaces gleichzeitig verwendet werden und aufeinander abgestimmte Prozesse ausgeführt werden. Private Parameter können dagegen nur in ihrem jeweiligen Workspace bearbeitet und aufgerufen werden.

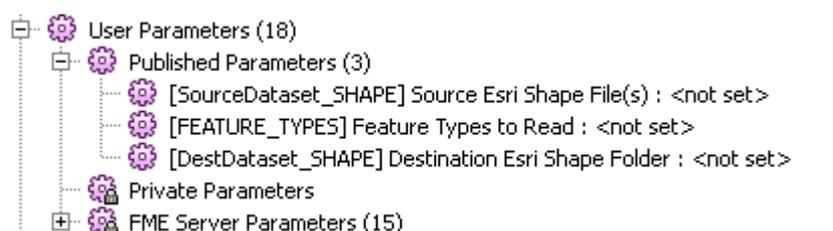


Abbildung 37: Parameterübersicht

Als Transformer Parameter können anstelle von User Parametern auch Attributwerte oder bedingte Werte (Conditional Values) eingesetzt werden. Diese haben den Vorteil, dass ihre Werte im Gegensatz zu Parametern im Verlauf der Transformation dynamisch angepasst werden können.

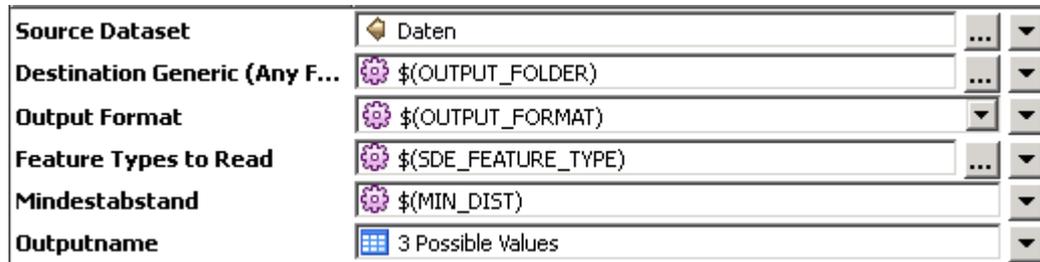


Abbildung 38: Transformer Parameter als Attribut, User Parameter, und Conditional Value

7.1.5 Erstellen und Bearbeiten einer Transformation

FME bietet zur Erstellung von Transformationen die Möglichkeit mit einer leeren Oberfläche (*Blank Workspace*) oder einer Vorlage (*Create workspace from Template*) zu arbeiten, eine weitere Möglichkeit besteht darin, sich einen Workspace mit einem beliebigen Reader und Writer generieren zu lassen (*Generate Workspace*). Um den Reader zu erstellen, müssen ein Datenformat sowie Eingangsdatensätze gewählt werden, der Writer benötigt ebenfalls ein Datenformat sowie einen Speicherort für den Outputdatensatz. Formatabhängig kann dies ein Ordner oder eine Datenbank sein. Zusätzlich besteht noch die Möglichkeit, zwischen einem dynamischen oder statischen Schema zu wählen, ein dynamisches Schema ermöglicht es FME darauf zu reagieren, wenn der Aufbau der Inputdaten verändert wird oder diese vollständig ausgetauscht werden. Weitere Reader und Writer können über die *Add Reader/Writer* Buttons jederzeit zum Workspace hinzugefügt werden.

Transformer lassen sich über das Transformer Gallery Fenster per Drag and Drop zur Oberfläche hinzufügen, alternativ können sie durch Eintippen ihres Namens ausgewählt werden, während man im Main-Fenster arbeitet (siehe Abbildung 39).

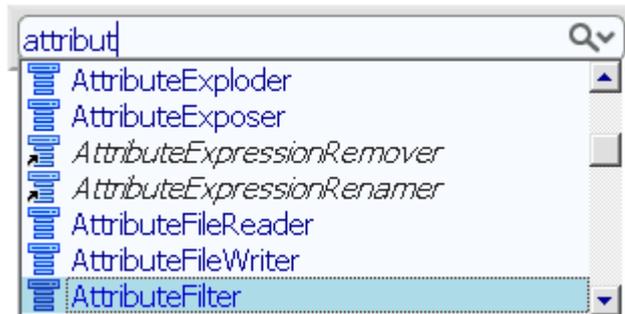


Abbildung 39: Transformer-Auswahl

Jedes Objekt verfügt über Input-Ports auf seiner linken und Output-Ports auf seiner rechten Seite. Diese lassen sich durch Ziehen eines Output-Ports zu einem Input-Port eines anderen Objektes verbinden (siehe Abbildungen 40).

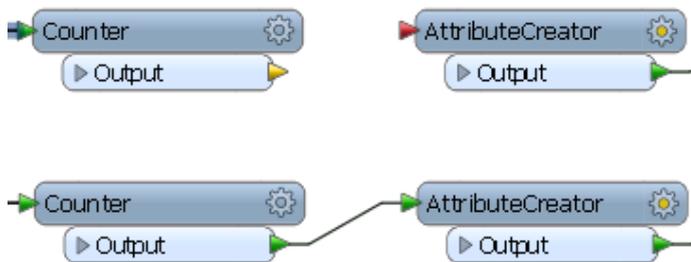


Abbildung 40: Verbinden von Transformern

Durch diese Verbindungen wird der Datenfluss der Transformation gesteuert. Verbindungen lassen sich durch Ziehen auch an anderen Ports platzieren und die Anzahl der ein- beziehungsweise ausgehenden Verbindungen eines Ports ist nicht begrenzt. Es besteht außerdem die Möglichkeit, Transformer in Verbindungen einzufügen, indem der beim Verschieben angezeigte pinkle Pfeil auf eine Verbindung geschoben wird, bis diese ebenfalls pink unterlegt ist (siehe Abbildung 41).



Abbildung 41: Einfügen eines Transformers

Bei Transformern, die mit einem Writer Feature Type verbunden sind, können ebenfalls die Attribute des Transformers mit denen des Feature

Types verbunden werden. Bei gleichem Namen wird diese Verbindung von FME automatisch vorgenommen, was an der grünen Färbung des Ports erkennbar ist, während Attribute, denen keine Inputdaten zugewiesen werden, durch einen roten Port gekennzeichnet werden, die Spalten bleiben beim Schreiben der Daten leer. Über das Zahnrad-Symbol lassen sich die Eigenschaften der Feature Types sowie die Parameter der jeweiligen Transformer anzeigen und bearbeiten. Durch die Farbe des Zahnrades wird angezeigt ob der Nutzer Einstellungen vornehmen muss (rot), die Default-Einstellungen bestätigen soll (Gelbes Zentrum) oder ob die Parameter bereits angepasst wurden (keine Färbung). In den Eigenschaften der Writer Feature Types, können weitere Attribute angelegt werden, welche die Outputdatensätze enthalten sollen, im Reader Feature Type lässt sich einstellen ob alle Feature Types des Readers getrennt oder gemeinsam auf der Oberfläche dargestellt werden sollen. Die Transformer-Parameter sind vom jeweiligen Transformer abhängig, Parameter, die nicht durch Default-Werte belegt werden können, wie zu verwendende Attribute, sind rot unterlegt und müssen durch den Nutzer angegeben werden (siehe Abbildung 42).

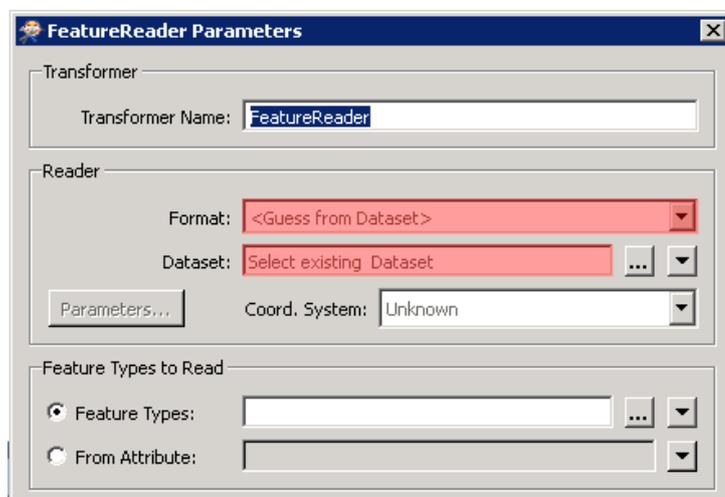


Abbildung 42: Transformer mit fehlenden Parametern

Um die Arbeit mit Transformationen zu erleichtern, bietet die Workbench mehrere Optionen, die man an markierten Objekten und Verbindungen mittels Rechtsklick durchführen kann. Durch das Erstellen von Bookmarks können Objekte gruppiert werden und durch Annotations können Kommentare angefügt werden (siehe Abbildung 43). Indem Objekte mit den Funktionen *disable*, beziehungsweise *enable* aktiviert oder deaktiviert

werden, ist es möglich Teilstücke von Transformationen zu testen. Für die Analyse von Zwischenergebnissen lässt sich über *Connect Inspector* die Software FME Data Inspector aufrufen, wodurch die Zwischenergebnisse in diesem dargestellt werden können und festgestellt werden kann, ob der Bearbeitungsprozess korrekt abläuft.

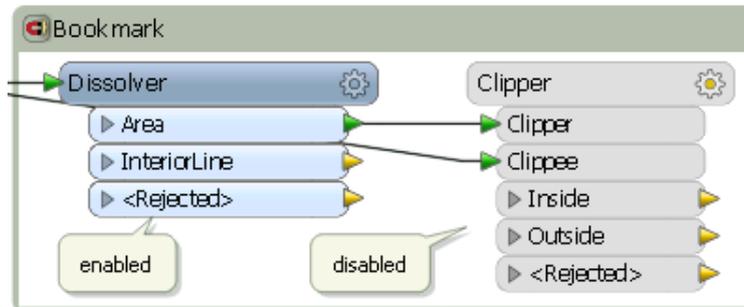


Abbildung 43: Aktivierter und deaktivierter Transformer mit Annotations und einem Bookmark

7.2 Erstellung des Validierungswerkzeuges

In diesem Kapitel werden der Erstellungsprozess und die Funktionsweise des Validierungswerkzeuges für Geodaten beschrieben. Für sämtliche der unter Kapitel 3.2 „Konzept“ beschriebenen Fehlertypen, werden mit FME Workspaces angelegt, welche die Eingangsdaten auf jeweils einen Fehler untersuchen. Zusätzlich werden beispielhafte Korrekturprozesse für doppelte Stützpunkte und Spikes erstellt und dem Validierungswerkzeug hinzugefügt. Die einzelnen Validierungsprozesse können über einen Master Workspace aufgerufen werden, in dem die einzelnen Prozesse ausgewählt werden können.

7.2.1 Übersicht der verwendeten Transformer

In diesem Teilkapitel werden, zur besseren Übersicht und dem besseren Verständnis der darauffolgenden Beschreibungen, die Funktionen der verwendeten FME Transformer kurz beschrieben.

- *Aggregator*: Aggregiert alle Geometrien zu einer. Die Geometrien können nach Attributen gruppiert aggregiert werden, wodurch nur Geometrien mit dem gleichen Attributwert zusammengefügt werden.
- *AreaBuilder*: Erstellt, wenn möglich, aus allen Geometrien einzelne Flächen, z.B. Multipart-Polygon -> Singlepart-Polygon
- *AreaCalculator*: Erstellt ein Attribut mit dem Flächeninhalt des jeweiligen Features.
- *AreaOnAreaOverlayer*: Verschneidet alle Polygone miteinander und erstellt neue Polygone, die ein Attribut besitzen das die Anzahl der Überlagerungen der Inputdaten angibt.
- *AttributeCreator*: Erstellt ein Attribut, für das auch ein Default Wert angegeben werden kann.
- *AttributeExposer*: Macht ausgewählte Attribute für die Transformation verwendbar, die in den Inputdaten möglicherweise fehlen.
- *AttributeKeeper*: Entfernt alle Attribute des Datensatzes, nur ausgewählte Attribute bleiben erhalten.
- *AttributeRemover*: Entfernt ausgewählte Attribute.
- *Bufferer*: Erstellt eine Bufferfläche um die Eingangsdaten entsprechend eines angegebenen Wertes.
- *CenterPointReplacer*: Ersetzt Polygone durch ihren Mittelpunkt.
- *Chopper*: Zerteilt Geometrien an ihren Stützpunkten in Punkte oder Linien.
- *Clipper*: Schneidet die Geometrien die über den Clipper-Port eingegeben werden aus denen aus, die über den Clippee-Port eingegeben werden.
- *CoordinateCounter*: Erstellt ein Attribut, das die Anzahl der Stützpunkte des Objekts enthält.
- *CoordinateRemover*: Entfernt Stützpunkte am Anfang oder Ende einer Geometrie.
- *Counter*: Erstellt ein Attribut, in dem die einzelnen Features nummeriert werden.
- *Creator*: Erstellt Features innerhalb einer Transformation.
- *Deaggregator*: Trennt Features, z.B. Multipart-Polygone zu Singlepart-Polygonen.

- *Dissolver*: Löst alle sich berührenden Einzelgeometrien zu einer Geometrie auf. Die Geometrien können nach Attributen gruppiert aufgelöst werden, wodurch nur Geometrien mit dem gleichen Attributwert zusammengefügt werden.
- *DonutHoleExtractor*: Trennt den inneren und äußeren Ring eines Donut-Polygons und gibt diese durch getrennte Ports aus.
- *FeatureReader*: Liest externe Datensätze wie ein Reader aus, und führt zusätzliche räumliche Abfragen anhand des Transformer-Inputs aus. Dadurch muss nicht der vollständige externe Datensatz gelesen werden.
- *GeometryFilter*: Filtert Features nach ihrem Geometrietyp.
- *GeometryValidator*: Führt verschiedene Geometrieverifizierungen durch und kann die Daten teilweise reparieren.
- *LineCloser*: Verbindet den Anfangs- und Endpunkt von Polylinien miteinander.
- *LineOnLineOverlayer*:
- *ListConcatenator*: Wandelt eine Liste in ein Attribut um.
- *ListExploder*: Erstellt identische Features für jedes Listenelement, die über jeweils ein Listenelement als Attribut verfügen. Sollten mehrere Listen vorhanden sein, wird jedes Element an der gleichen Stelle in der Liste ein Attribut des neuen Features.
- *NeighborFinder*: Findet benachbarte Geometrien innerhalb einer bestimmten Toleranz. Die Geometrien können nach einem Attribut gruppiert werden, sodass nur Geometrien mit dem gleichen Attributwert gefunden werden.
- *PointOnLineOverlayer*: Verschneidet Punkte und Linien miteinander, ein Attribut gibt die Anzahl der Überlagerungen an.
- *PointOnPointOverlayer*: Verschneidet Punkte miteinander, ein Attribut gibt die Anzahl der Überlagerungen an.
- *SliverRemover*: Entfernt Sliver Polygone
- *Sorter*: Sortiert die Features nach einem beliebigen Attribut.
- *SpatialFilter*: Filtert Daten nach räumlichen Abfragen.
- *SpikeRemover*: Entfernt Spikes und gibt ein Fehlershape aus.
- *StatisticsCalculator*: Wertet Attributwerte aus, z.B. nach den höchsten Wert.

- *Terminator*: Bricht eine Transformation ab, wenn die durch den Nutzer festgelegten Voraussetzungen erfüllt sind.
- *Tester*: Filtert die Daten durch Abfragen, die durch den Nutzer festgelegt werden können. Es existieren Outputports für „bestanden“ und „nicht bestanden“.
- *TestFilter*: Filtert die Daten nach mehreren Abfragen, für die einzelne Outputports existieren.
- *TopologyBuilder*: Erstellt Topologien Anhand der InputDaten, ignoriert Multipart-Polygone.
- *VertexCreator*: Erstellt Punkte anhand von Attributwerten.
- *WorkspaceRunner*: Startet einen externen FME-Workspace. Die Published Parameter des Workspace werden zu Transformer Parametern des WorkspaceRunner.

[SAFE_8, S. 12 ff]

7.2.2 Die Validierungs- und Korrekturprozesse

Alle Validierungsprozesse verfügen über einen gemeinsamen Grundaufbau. Jeder Workspace enthält zwei Reader, den ArcSDE Reader, um auf die UIS-Datenbank zugreifen zu können. Sowie einen Generic Reader, der jedes dateibasierte Datenformat lesen kann, um Shapefiles verarbeiten zu können. Außerdem enthält jeder Prozess einen Generic Writer, um das Fehlershape als Shapefile oder File Geodatabase Feature Class, schreiben zu können. Damit die Eingangsdaten der Prozesse flexibel ausgetauscht werden können, werden alle Workspaces mit einem dynamischen Schema angelegt. Das Input Feature des Generic Readers, die zu lesenden Feature Classes des ArcSDE Readers, das entsprechende SDE-Dataset, der Speicherort und das Datenformat des Outputs werden als Published User Parameter gespeichert, um sie vom Master Workspace aus steuern zu können.

Jeder Workspace enthält zusätzlich einen CSV-Writer, der ein LogFile im CSV-Format (Comma Seperated Value) ausgibt. Diese Datei enthält den Namen des eingesetzten Validierungs- oder Korrekturtools sowie die Anzahl der gelesenen und geschriebenen Objekte. Damit kann der Nutzer überwachen, ob die Datensätze in den einzelnen Prozessen vollständig

verarbeitet wurden. Der Pfad und Name der CSV-Datei werden ebenfalls als Published User Parameter angegeben. Hierfür werden die Input- und Outputfeatures gezählt und der jeweils höchste Wert ermittelt. Diese Daten werden zu einer Zeile zusammengefasst und mit weiteren Attributen, wie den durchgeführten Validierungsprozess und den Zeitpunkt der Validierung, vor oder nach der Korrektur, ergänzt. In Abbildung 44 ist diese Vorgehensweise mit den verwendeten Transformatoren dargestellt.

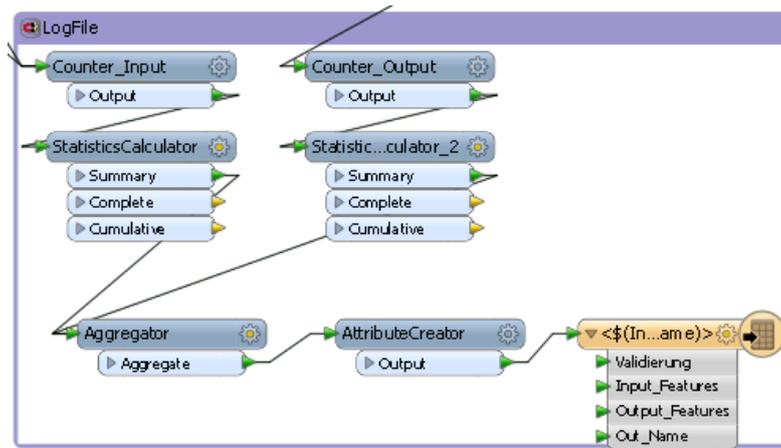


Abbildung 44: Prozess zur Erstellung des LogFiles

Die Validierungswerkzeuge enthalten nach den Readern eine Gruppe von Transformatern, die alle Attribute entfernen, welche nicht für die Weiterverarbeitung benötigt werden. Dies vermindert den Rechenaufwand bei großen Datenmengen. In einigen Prozessen werden bestimmte Attribute wie die Wasserschutzgebietsnummer benötigt, diese bleiben in den entsprechenden Prozessen erhalten. Darauf folgt eine Abfrage des Parameters mit der Angabe ob der gesamte Datensatz überprüft werden soll. Ist dies der Fall, wird mit der jeweiligen Validierung fortgefahren. Anderenfalls folgt die Abfrage des Parameters, in dem die gewünschte WSG-Nummer angegeben ist. Die Daten mit der entsprechenden Nummer werden anschließend gefiltert und an den Validierungsprozess weitergegeben (siehe Abbildung 45).

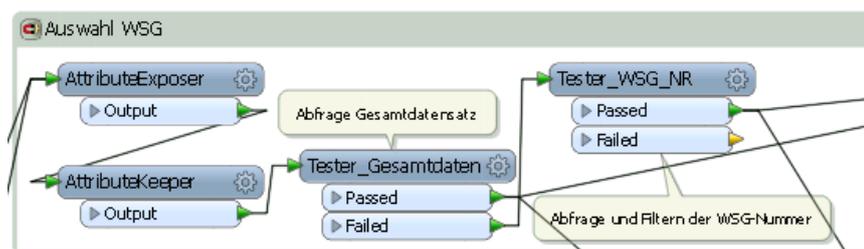


Abbildung 45: Prozess zum Entfernen von Attributen und der Auswahl des WSG

Die als Fehlershape ausgegebenen Geometrien werden mit einem Attribut versehen, welches Informationen zum jeweiligen Fehler beinhaltet.

Die Korrekturprozesse verfügen über einen Generic-Reader und einen Shapefile-Writer, der die korrigierten Daten ausgibt sowie einen Generic-Writer, der ein Shapefile ausgibt, das die veränderten Features kennzeichnet. Außerdem wird auch hier ein LogFile geschrieben. Zusätzlich verfügen die Korrekturwerkzeuge über einen Terminator Transformer, der den aktuellen Prozess abbricht, sollte ein leerer Output geschrieben werden. Dadurch gilt der Korrekturprozess als fehlgeschlagen und die Attribute im Master Workspace werden nicht aktualisiert. Dies ermöglicht ein korrektes Fortfahren mit der nächsten Korrektur, die sonst einen leeren Input hätte.

Die Funktionsweisen der Validierungsprozesse sind in den Kapiteln 7.2.2.1 bis 7.2.2.16 beschrieben, darauf folgen die ausgewählten Korrekturprozesse.

7.2.2.1 Prüfung auf doppelt oder zu nah gesetzte Stützpunkte

FME bietet bereits eine Funktion um lageidentische Stützpunkte zu ermitteln. Da jedoch auch Punkte innerhalb eines Abstandes von 5 mm als problematisch angesehen werden, ist der Einsatz dieser Funktion hier nicht möglich. Der gewünschte Mindestabstand zweier Punkte wird als Published User Parameter angegeben, wodurch er beim Einsatz des Werkzeuges durch den Nutzer frei ausgewählt werden kann.

Um die doppelten Stützpunkte zu identifizieren, werden die ursprünglichen Geometrien in Punkte umgewandelt und nach dem ursprünglichen Feature der Punkte sortiert, da Punkte aus verschiedenen Features die gleichen Koordinaten haben dürfen.

Zu Beginn werden die inneren und äußeren Ringe des Polygons getrennt um diese getrennt untersuchen zu können (siehe Abbildung 46).

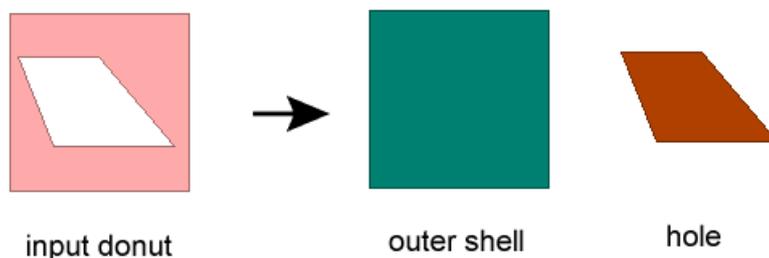


Abbildung 46: Funktionsweise des DonutHoleExtractors, Trennung des äußeren und inneren Rings (Quelle: [SAFE_6])

Danach werden die einzelnen Objekte nummeriert. Anschließend wird der jeweils letzte Stützpunkt der Geometrien entfernt, da der Start- und Endpunkt eines Polygons von FME als zwei identische Punkte angesehen werden. Diese identischen Punkte würden sonst im Fehlershape erscheinen, obwohl sie keine Probleme verursachen.

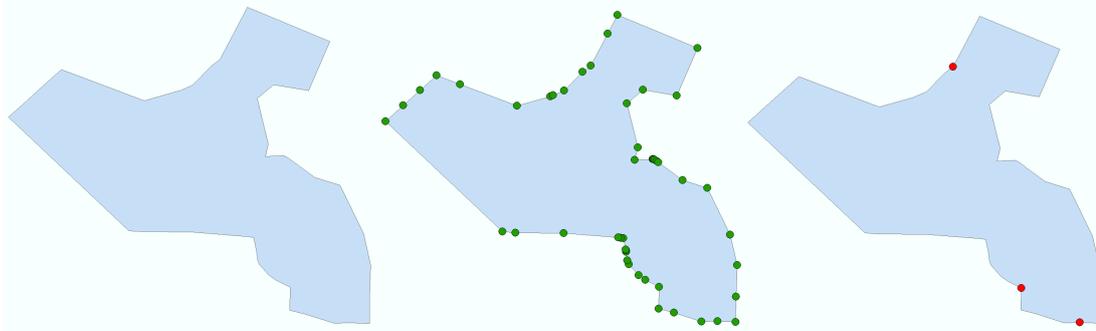


Abbildung 47: Feature, Feature nach der Transformation in Punkte und Ergebnis der Validierung

Im nächsten Schritt werden die Stützpunkte der Polygone in Punktfeatures konvertiert (siehe Abbildung 47). Anschließend werden Punkte die sich innerhalb des festgelegten Abstandes zueinander befinden ausgewählt. Hierbei werden die Punkte über die zuvor erstellte Nummerierung gruppiert. So wird sichergestellt, dass nur Punkte ausgewählt werden, die aus demselben Ursprungsfeature stammen. In Abbildung 48 wird die Vorgehensweise mit den einzelnen Transformern dargestellt.

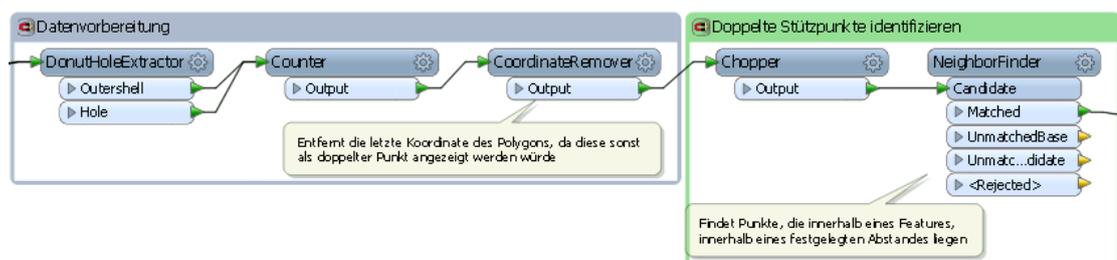


Abbildung 48: Konvertierung in Punkte und Identifizierung von doppelten Punkten

7.2.2.2 Prüfung auf geschlossene Polygone

Die Überprüfung auf geschlossene Polygone geht nach dem Ansatz vor, dass nicht geschlossene Polygone, keine Polygone im eigentlichen Sinne sind. Polygone sind vollständig, von miteinander verbundenen Punkten, eingegrenzte Flächen sind [ESRI_3]. Ein nicht geschlossenes Polygon ist nicht vollständig eingegrenzt und somit keine Fläche, sondern eine Polylinie.

Deshalb können sie über ihren Geometrietyp aus dem Datensatz herausgefiltert werden.

Die Validierung erfolgt durch das Filtern der Features nach ihrem Geometrietyp. Hier werden Polylinien aus dem Gesamtdatensatz herausgefiltert und als Fehlershape ausgegeben (siehe Abbildung 47).

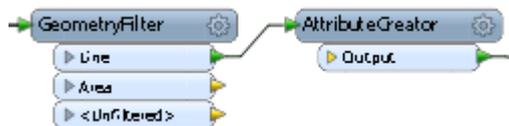


Abbildung 49: Filtern von geschlossenen Polygonen (Polylinien) im GeometryFilter

7.2.2.3 Prüfung der Orientierung von Polygonen

FME verfügt mit dem GeometryValidator Transformer bereits über ein Validierungswerkzeug für einige Geometriefehler, darunter auch die Orientierung von Polygonen.

Die Ergebnisse dieses Transformers werden als Liste an die Daten angefügt. Anschließend wird für jedes Listenelement ein Geometrieobjekt, mit dem jeweiligen Fehler als Attribut, erstellt. Schließlich werden die Features gefiltert, deren Fehlerattribut eine falsche Orientierung angibt (siehe Abbildung 50).



Abbildung 50: Prüfung auf falsch orientierte Polygone Im GeometryValidator

7.2.2.4 Prüfung auf Splitterflächen

Um Splitterflächen aus den Daten zu filtern, werden zwei Ansätze verfolgt. Zum einen wird nach dreieckigen Flächen und Löchern, so genannten Dreipunkt-Splittern, gesucht, welche aus Erfahrung oft Splitterflächen sind. Zum anderen wird nach kleinsten Flächen und Löchern gesucht, deren

Fläche nur einen minimalen Anteil der Gesamtfläche ausmacht. Da beide Ansätze verschiedene Arten von Splitterflächen erkennen, werden in diesem Validierungsprozess beide Methoden durchgeführt.

Dreipunkt-Splitter können sowohl als Löcher als auch als Polygone auftreten. Da Splitterpolygone oft Teile von Multipart-Polygonen sind, werden zuerst alle Multiparts in Singleparts aufgeteilt. Um Löcher aufzuspüren, welche die Kante des Polygons berühren, wird ein Buffer um die Fläche gelegt, der die Löcher vollständig von der äußeren Kante abtrennt (siehe Abbildung 51). Dies ist notwendig, da Löcher, die die Kante berühren, von FME nicht als solche behandelt werden.

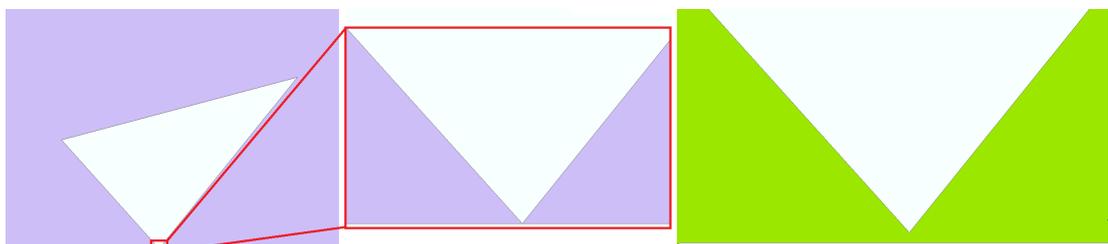


Abbildung 51: Abtrennen eines Lochs an der Kante (violett), durch einen Buffer (grün)

Anschließend werden sämtliche Löcher in Polygone umgewandelt. Im nächsten Schritt wird die Anzahl der Koordinaten aller Polygone ermittelt, diese entspricht der Anzahl der Ecken plus eins, da der Start und Endpunkt eines Polygons zwar identisch sind, aber doppelt gezählt werden. Durch einen Tester werden alle in diesen Schritten erstellten Flächen ausgewertet. Alle Flächen mit vier oder weniger Koordinaten, beziehungsweise drei Ecken, werden im Fehlershape ausgegeben. Diese Vorgehensweise ist in Abbildung 52 mit den verwendeten Transformern dargestellt.

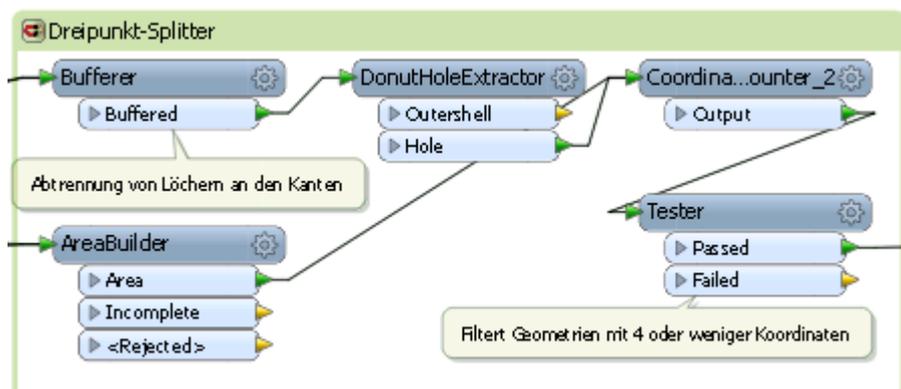


Abbildung 52: Prüfung auf „Dreipunkt“-Splitter durch extrahieren von Löchern

In der zweiten Methode wird die Fläche der ursprünglichen Polygone berechnet. Darauf folgend werden wie im ersten Ansatz die Multipart Polygone getrennt und Löcher extrahiert. Für die nun erstellten Flächen wird ebenfalls die Fläche berechnet und der Anteil der Polygone an der Gesamtfläche berechnet. Darauf werden alle Flächen, deren Anteil kleiner als 0,0001 % ist, für das Fehlershape aufbereitet. Dieser Wert wurde nach der Analyse von Beispieldaten festgelegt (vgl. Abbildung 53)

FID	Shape *	coord	area	area gesam	anteil
0	Polygon	935	10773244,840138	10773247,001988	99,99998
1	Polygon	4	0,29135	10773247,001988	0,000003
2	Polygon	6	0,12095	10773247,001988	0,000001
3	Polygon	4	0,013	10773247,001988	0
4	Polygon	4	0,0667	10773247,001988	0,000001
5	Polygon	7	0,66575	10773247,001988	0,000006
6	Polygon	6	0,1538	10773247,001988	0,000001
7	Polygon	6	0,1393	10773247,001988	0,000001
8	Polygon	6	0,1098	10773247,001988	0,000001
9	Polygon	4	0,0449	10773247,001988	0
10	Polygon	5	0,41075	10773247,001988	0,000004
11	Polygon	4	0,0063	10773247,001988	0
12	Polygon	4	0,13925	10773247,001988	0,000001

Abbildung 53: Attributtabelle mit Splitterflächen, die über ihren Flächenanteil bestimmt wurden. Bei coord = 4 liegen außerdem Dreipunkt-Splitter vor.

Löcher bilden keinen Teil der Gesamtfläche und werden bei ihrer Berechnung nicht berücksichtigt. Aber Löcher, die im Vergleich nur über so eine geringe Fläche verfügen, können als Splitter angesehen werden. Abbildung 54 stellt die Vorgehensweise des zweiten Ansatzes mit den verwendeten Transformern dar.

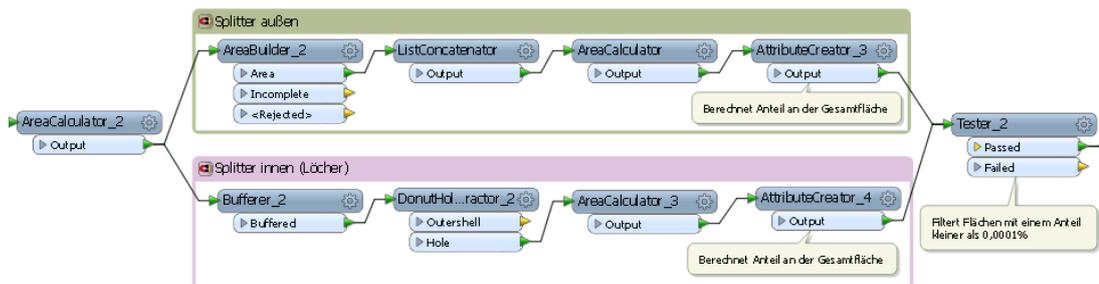


Abbildung 54: Prüfung auf sehr kleine Splitterpolygone, durch Berechnung es Anteils an der Gesamtfläche

7.2.2.5 Prüfung auf Löcher

Löcher können mit einem DonutHoleExtractor identifiziert werden. Zusätzlich wird die Fläche des Lochs berechnet. Bei Löchern mit einer Fläche größer als 1,5 m² kann davon ausgegangen werden, dass das Loch beabsichtigt ist,

da es beispielsweise eine Aussparung für eine andere WSG-Zone ist. Der Wert 1,5 m² wurde gewählt, da dieser der aktuell kleinsten WSG-Zone entspricht. Abbildung 55 stellt die Vorgehensweise dieses Prozesses dar.

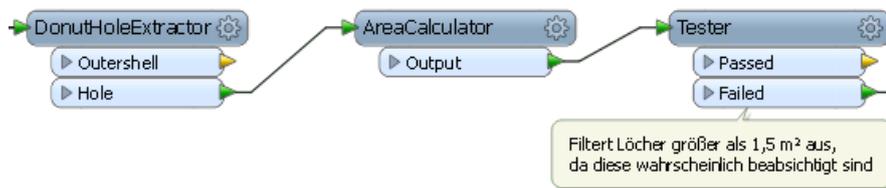


Abbildung 55: Prüfung auf Löcher

7.2.2.6 Prüfung auf Fischeschwanz-Geometrien

Um Fischeschwanz-Geometrien zu identifizieren, müssen Multipart-Polygone darauf untersucht werden, ob sich ihre einzelnen Flächen in nur einem Punkt berühren. Geometrien verschiedener Multipart-Polygone, die sich in nur einem Punkt berühren, werden nicht als fehlerhaft angesehen. Um dies zu erreichen werden die Eingangsdaten auf identische Stützpunkte untersucht. Wobei Stützpunkte, die nur in einem einzigen Teil eines Multipart-Polygons liegen (Abbildung 56: grüner Punkt) sowie Stützpunkte, die in verschiedenen Polygonen liegen (schwarzer Punkt), vom Ergebnis der Untersuchung ausgenommen werden. Das Ergebnis enthält somit nur doppelte Stützpunkte, welche auf verschiedenen Teilen eines Multipart-Polygons liegen (roter Punkt).

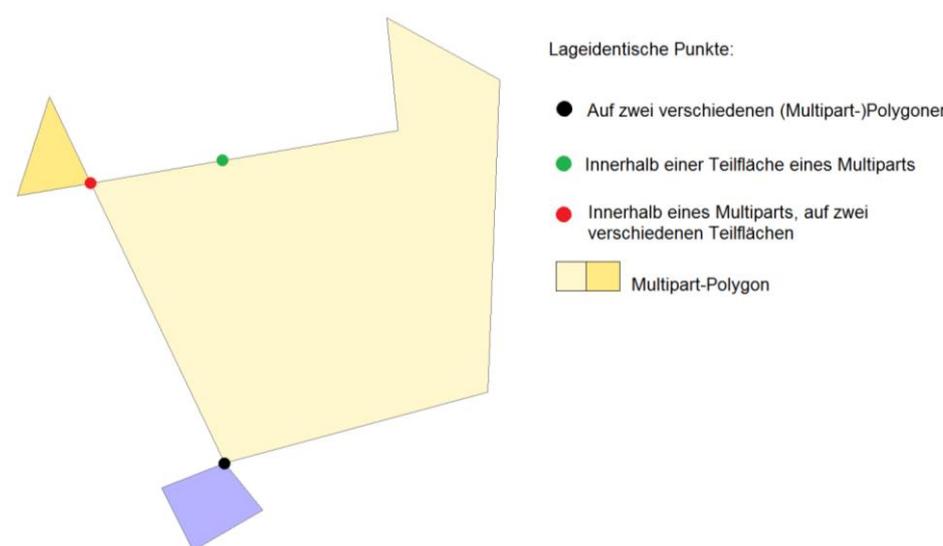


Abbildung 56: Verschiedene Arten lageidentischer Stützpunkte bei der Identifizierung von Fischeschwanz-Polygonen, rot identifiziert Fischeschwänze

Für diesen Prozess werden im ersten Schritt die ursprünglichen Geometrien mit einer eindeutigen ID versehen, die einzelnen Teile von Multipart-Polygonen erhalten so eine gemeinsame ID. Anschließend werden die inneren und äußeren Ringe getrennt, um diese getrennt voneinander untersuchen zu können. Außerdem werden die Multipart-Polygone in einzelne Geometrien unterteilt. Die so neu erzeugten Einzelfeatures erhalten ebenfalls eine weitere eindeutige Nummerierung. Danach werden die Stützpunkte der Features in Punktgeometrien transformiert, wobei die Punktgeometrien beide IDs behalten. Dieser Prozess ist in Abbildung 57 mit den entsprechenden Transformatoren dargestellt.

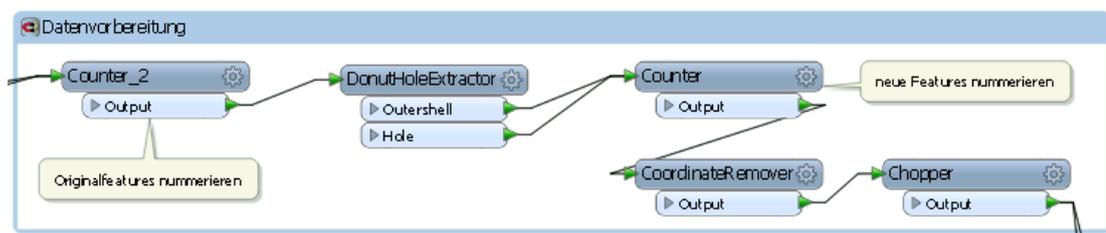


Abbildung 57: Datenvorbereitung: Kennzeichnung der Features und Konvertierung in Punkte

Aus diesen Punkten werden alle lageidentischen Punkte herausgefiltert. Wobei sie anhand der zuvor erzeugten IDs einmal nach ihrem ursprünglichen (Multipart-)Polygon und einmal nach ihrem neu erzeugten (Singlepart-)Feature gruppiert werden (siehe Abbildung 58). Die Punkte, die nach ihrem neuen Polygon gruppiert wurden, werden anschließend aus dem nach den ursprünglichen Features gruppierten Datensatz ausgeschnitten (siehe Abbildung 59).

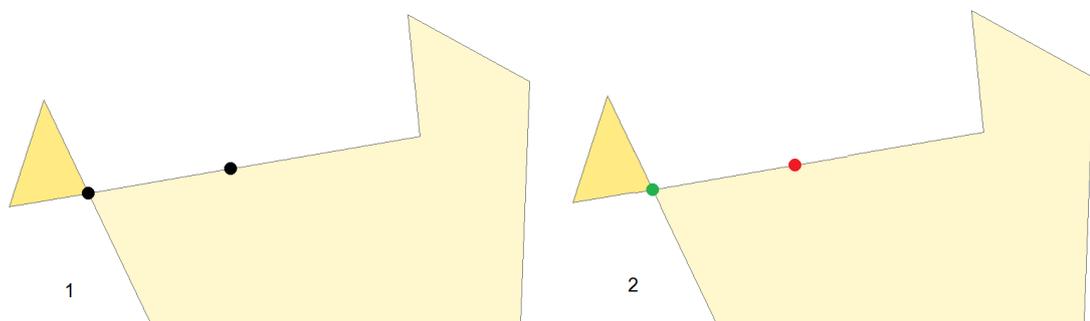


Abbildung 58: 1. doppelte Punkte gruppiert nach ihrem ursprünglichen Feature, 2. doppelte Punkte gruppiert nach neu dem erzeugten Feature

Somit bleiben nur die lageidentischen Punkte übrig, die innerhalb eines Multipart-Polygons auf zwei verschiedenen Einzelpolygonen lagen.

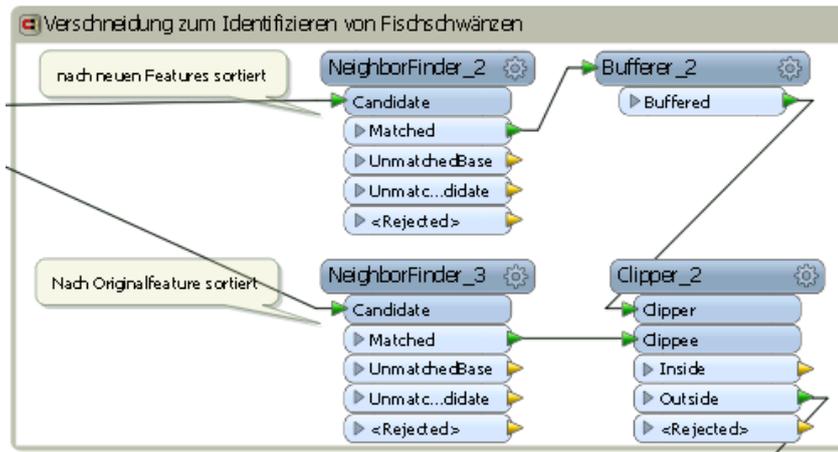


Abbildung 59: Prüfung auf Fischeschwanz-Polygone

7.2.2.7 Prüfung auf Selbstüberschneidungen

Für die Überprüfung auf Selbstüberschneidungen bietet FME bereits die Möglichkeit, den GeometryValidator Transformer einzusetzen. Die Ergebnisse des GeometryValidator werden als Liste an die Daten angefügt. Außerdem werden Listen mit den Koordinaten der Fehler erzeugt. Diese Listen werden in Attribute konvertiert, wobei für jedes Listenelement ein Feature angelegt wird und die Listen zu Spalten des hierbei erzeugten Datensatzes werden (siehe Abbildung 60).

	location sample x	location sample y	issue found
	3495300.86	5455020.89	Self-Intersections in 2D
▶	3493850.2212	5456422.2329	Self-Intersections in 2D

Abbildung 60: Beispielhafte Konvertierung eines Features mit drei Listen mit je zwei Elementen in einen Datensatz mit zwei Features und drei Spalten.

Mit den angegebenen Koordinaten werden dann Punkte erzeugt. Diese Punkte, welche die Position von Selbstüberschneidungen darstellen werden ausgefiltert und im Fehlershape ausgegeben. Abbildung 61 stellt diesen Prozess mit den verwendeten Transformern dar.

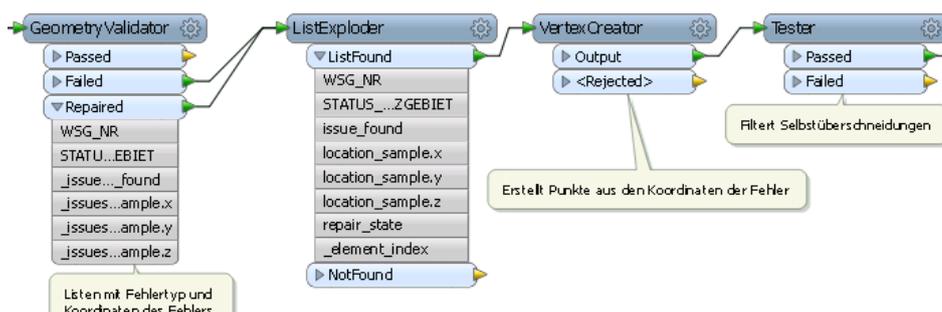


Abbildung 61: Prüfung auf Selbstüberschneidungen im GeometryValidator

7.2.2.8 Prüfung auf Spikes

FME bietet mit dem SpikeRemover bereits einen Transformer, der Spikes korrigieren kann (siehe Abbildung 62). In diesem Transformer ist es möglich, Spikes anhand ihrer Länge oder ihres Winkels zu identifizieren.

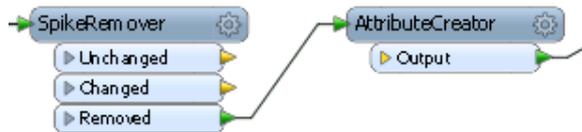


Abbildung 62: Prüfung, der Removed-Port gibt

Für die Daten der LUBW ist eine Erkennung über den Winkel die bessere Alternative, da Spikes in sehr unterschiedlichen Längen vorkommen. Der Winkel unterhalb dessen ein Spike vorliegt (siehe Abbildung 63). als Published User Parameter angegeben, wodurch er durch den Nutzer frei einstellbar ist.

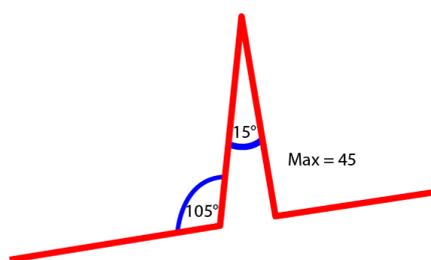


Abbildung 63: Definition von Spikes anhand ihres Winkels (Quelle: [SAFE_7])

7.2.2.9 Prüfung auf Einhaltung der Mindestabstände

Dieser Prozess soll überprüfen, ob die Mindestabstände von einzelnen Teilen eines Objektes zueinander eingehalten werden. Dies beinhaltet die Abstände einzelner Teile von Multipartpolygonen, den Abstand innerer Ringe zum äußeren Ring und zueinander, sowie Engstellen in Polygonen. Diese Probleme können dadurch identifiziert werden, dass sich mindestens drei Kanten des Polygons innerhalb des als Mindestabstand bestimmten Radius befinden. Dies schließt auch parallel verlaufende, zu eng liegende Kanten ein, da nur Polygone geprüft werden, müssen die Objekte über weitere Kanten verfügen, die die dritte Kante für die Untersuchung bilden (siehe Abbildung 64). Der Mindestabstand kann über einen Published User Parameter gewählt werden.

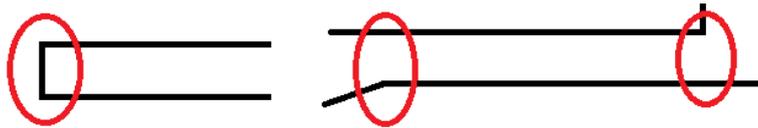


Abbildung 64: Identifizierung zu enger, parallel verlaufender Kanten, die dritte Kante innerhalb des Mindestabstands ist rot markiert.

Um eine korrekte Validierung zu gewährleisten, können in diesem Prozess nur einzelne Geometrieobjekte untersucht werden. Deshalb werden die einzelnen Features mit einer eindeutigen ID versehen. Ohne diese ID können sich überlagernde Objekte gleichzeitig untersucht werden, welche zu fehlerhaften Ergebnissen führen. Außerdem kann der Rechenaufwand so groß werden, dass der Prozess wegen zu geringem Arbeitsspeicher abbricht. Im nächsten Schritt werden die einzelnen Kanten der Eingangsdaten in Linienobjekte umgewandelt. Im nächsten Schritt wird ein Buffer mit der Breite des halben Mindestabstandes um die Linien gelegt.

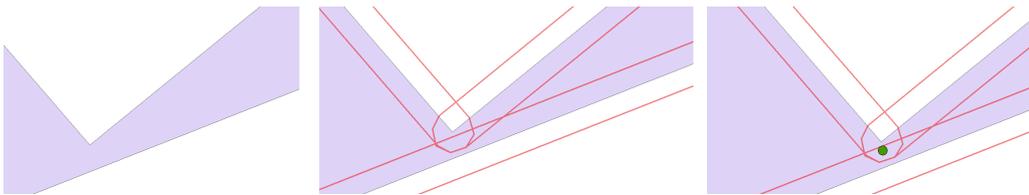


Abbildung 65: Engstelle, Buffer um die einzelnen Kanten und Punkt in der Überschneidung von drei Buffern, der die Engstelle markiert

Die Buffer werden, gruppiert nach der anfangs generierten ID, verschnitten. Hierbei wird ein Attribut erstellt, das angibt, wie viele Flächen bei dieser Verschneidung übereinandergelegt wurden. Anschließend werden alle Flächen gefiltert bei denen dieser Wert drei oder mehr beträgt. Nebeneinanderliegende Flächen werden nun Zusammengefasst und durch ihren Mittelpunkt ersetzt (siehe Abbildung 66). Die Punkte werden im Fehlershape ausgegeben und befinden sich jeweils in der Mitte einer Engstelle. Die vollständige Vorgehensweise ist in Abbildung 66 dargestellt.

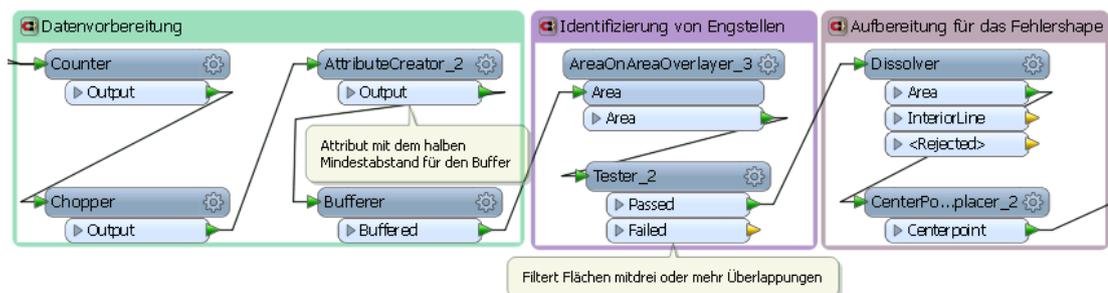


Abbildung 66: Prüfung auf die Einhaltung der Mindestabstände durch Verschnitten von Bufferflächen um die einzelnen Stützpunkte

7.2.2.10 Prüfung auf Multipart-Polygone

Zur Identifizierung mehrteiliger Geometrien müssen die ursprünglichen Polygone erst in einzelne Flächen aufgeteilt werden, damit beim Wiederherstellen der Multipart Polygone die zusammengeführten Flächen gezählt werden können.

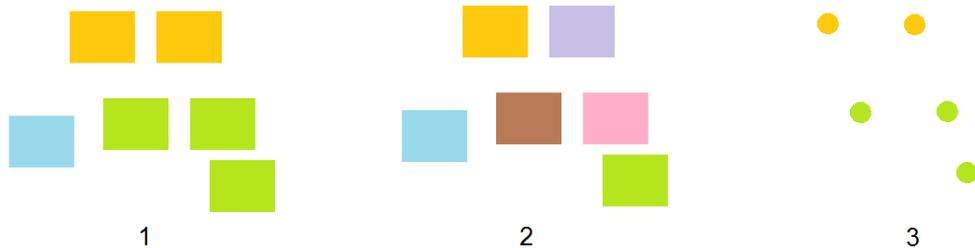


Abbildung 67: 1. Multipart-Polygone, 2. getrennte Singleparts, 3. Ergebnis der Validierung

Im ersten Schritt werden die ursprünglichen (Multipart-)Geometrien einer ID versehen und danach in einzelne Flächen umgewandelt. Diese werden anhand ihrer ID wieder zusammengefügt. Dabei wird ein Attribut erstellt, welches die Anzahl der zu einem Polygon zusammengeführten Flächen enthält. Alle Features, bei denen dieser Wert zwei oder mehr beträgt, werden gefiltert und mit einer neuen ID versehen. Diese ID kennzeichnet zusammengehörige Teilflächen für das Fehlershape. Die neuen Features werden wieder aufgeteilt und in Punkte konvertiert. So erhält man für jede Teilgeometrie einen Punkt, der ein Attribut enthält, in dem abgelesen werden kann, welche Teile zusammen gehören (siehe Abbildung 67). Diese Daten werden im Fehlershape ausgegeben. Abbildung 68 fasst diesen Prozess mit den verwendeten Transformern zusammen.

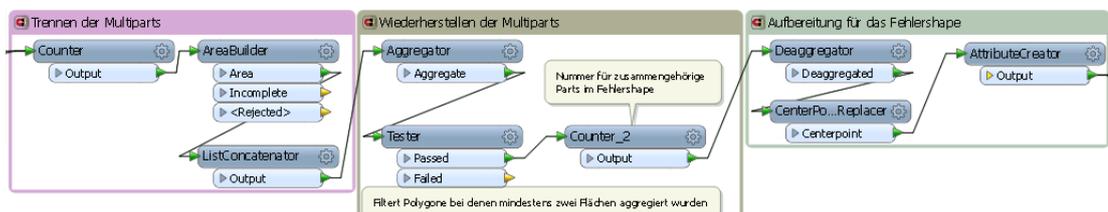


Abbildung 68: Prüfung auf Multipart-Polygone, durch Aufteilen und Wiederherstellen der Geometrien

7.2.2.11 Prüfung auf Überlappungen

Überlappungen stellen in erster Linie zwischen verschiedenen Wasserschutzgebietszonen ein Problem dar. Überlappungen zwischen Geometrien verschiedener Festsetzungsstände sind keine Fehler und werden von diesem Prozess daher nicht überprüft.

Um nur Überlappungen zwischen Geometrien eines Wasserschutzgebiets zu untersuchen, wird ein Attribut benötigt, mit dem die Wasserschutzgebiete unterschieden werden können, wie die WSG-Nummer. Deshalb wird dieses Attribut durch den Nutzer als Published User Parameter angegeben. Sollte kein geeignetes Attribut existieren, wird das „*ALLE*“ Attribut verwendet, welches im Verlauf des Prozesses für alle Features angelegt wird. Dieses Attribut dient nur dazu, eine Validierung durchführen zu können, wenn keine Attribute verfügbar sind.

Überlappungen innerhalb eines WSG können durch eine Verschneidung der einzelnen Polygone identifiziert werden. Da die Eingangsdaten nach dem zuvor gewählten Attribut gruppiert werden, werden nur die Überlappungen zwischen den gruppierten Attributen gefunden. Das Ergebnis dieser Verschneidung enthält auch die Überlappungsflächen (siehe Abbildung 69). Diese werden mit einem Attribut, das die Anzahl der Überlagerungen der Inputdaten enthält herausgefiltert, wenn der Werte des Attributes größer als eins ist.

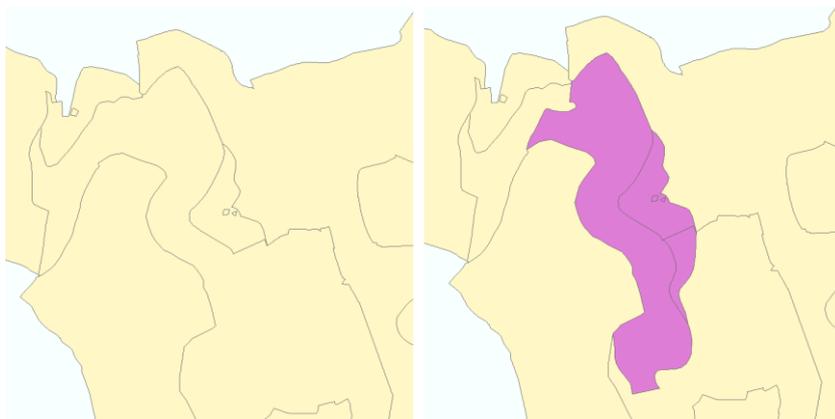


Abbildung 69: Verschiedene WSG-Zonen und gekennzeichnete Überlappungsflächen

In einem zweiten Durchgang dieses Prozesses (siehe Abbildung 70) werden die Daten auf Überlappungen zwischen verschiedenen Wasserschutzgebieten untersucht. Hierzu werden die Geometrien vor der

Verschneidung nach dem Gruppierungsattribut aufgelöst. Sodass die Verschneidung nun die kompletten Wasserschutzgebiete als Input erhält. Hierdurch werden nur Überlagerungen zwischen ganzen WSG identifiziert. Die Polygone der sich überlagernden Flächen werden als Fehlershape ausgegeben.

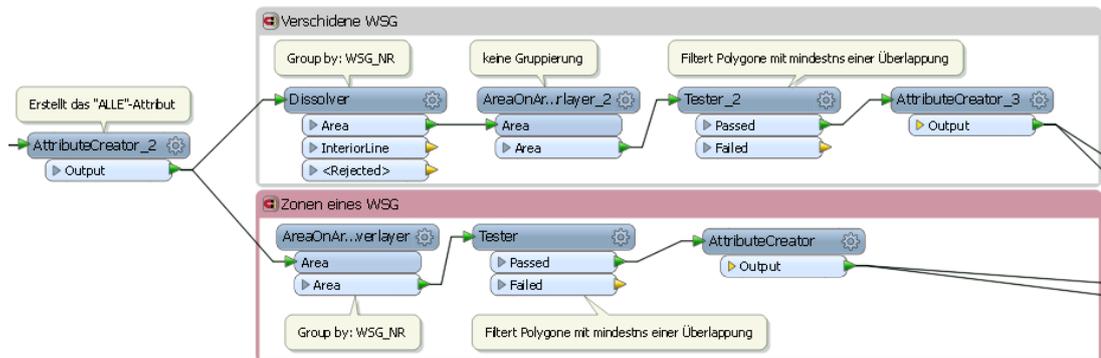


Abbildung 70: Prüfung auf Überlagerungen zwischen ganzen WSG (oben) und WSG-Zonen (unten)

8.2.2.12 Prüfung auf Lücken zwischen Geometrien

Lücken zwischen einzelnen Zonen eines Wasserschutzgebiets sind Fehler, während Lücken zwischen verschiedenen Wasserschutzgebieten beabsichtigt sein können, weshalb diese hier nicht geprüft werden. Um nur Lücken zwischen Geometrien eines WSGs zu untersuchen, werden, wie in 7.2.2.12 beschrieben, die entsprechenden Attribute verwendet.

Zur Identifizierung der Lücken werden die einzelnen Geometrien, nach dem gewählten Attribut gruppiert, aufgelöst (siehe Abbildung 71). In den so entstandenen Polygonen bilden die Lücken aus den Originalfeatures Löcher. Diese Löcher lassen sich extrahieren und in Polygone umwandeln. Diese Polygone werden anschließend im Fehlershape ausgegeben.



Abbildung 71: Prüfung auf Lücken zwischen Polygonen

7.2.2.13 Prüfung auf Sliver Polygone

FME bietet mit dem SliverRemover bereits einen Transformer zur Reparatur von Sliver Polygonen. Dieser kann auch zum bloßen Erkennen von Sliver Polygonen verwendet werden, indem sein Ergebnis mit den Ausgangsdaten abgeglichen wird. Der Abgleich erfolgt hierbei durch eine Verschneidung der reparierten mit den ursprünglichen Daten. Die Output Features verfügen hier über ein Attribut, das die Anzahl der Überlagerungen bei der Verschneidung angibt. Bei Sliverflächen gibt es entweder eine (Lücke) oder drei (Überlappung) Überlagerungen, alle anderen Flächen haben zwei Überlagerungen (siehe Abbildung 72).



Abbildung 72: Sliver Polygone, die durch die Anzahl der Überlagerungen bestimmt wurden

Die Flächen mit zwei Überlagerungen können aus dem Ergebnis entfernt werden, während die übrigen Polygone als Fehlershape ausgegeben werden. Abbildung 73 stellt den Ablauf dieses Prozesses dar.

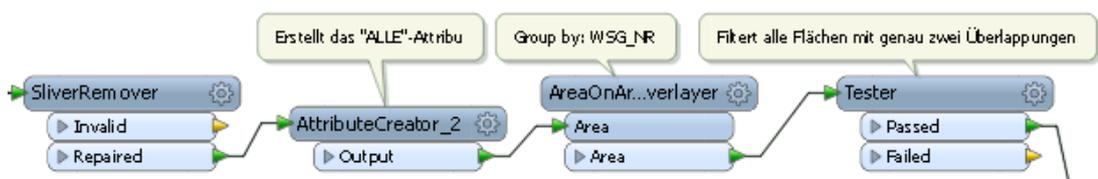


Abbildung 73: Prüfung auf Sliver Polygone

7.2.2.14 Prüfung auf Klaffen

Da Klaffen zwischen zwei Wasserschutzgebietszonen bei der Bildung des Umrings zu Spikes werden (siehe Abbildung 74), können sie auch ähnlich wie diese identifiziert werden. Hierzu werden die Spikes eines WSG-Umrings mit denen einzelner Zonen abgeglichen.

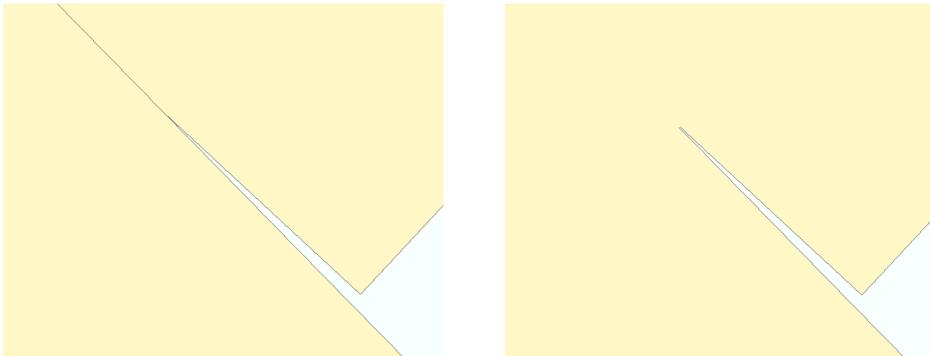


Abbildung 74: Klaffe (links) erzeugt einen Spike im Umring (rechts)

Zuerst werden die Einzelfeatures nach einem durch den Nutzer gewählten Attribut (siehe 7.2.2.11) aufgelöst und so der Umring gebildet. Dieser Umring wird, wie in 7.2.2.8, mit einem SpikeRemover untersucht. Gleichzeitig werden die Löcher aus dem Umring extrahiert und die Spikes in den Originaldaten identifiziert. Anschließend werden aus den Spikes der Umringe die Spikes der Originalfeatures und die Löcher der Umringe ausgeschnitten, wodurch nur die Klaffen in den Ergebnisdaten enthalten bleiben. Die Löcher werden ausgeschnitten, da sehr schmale Löcher ebenfalls als Klaffen identifiziert werden. Abbildung 75 stellt den Ablauf dieses Prozesses mit den verwendeten Transformatoren dar.

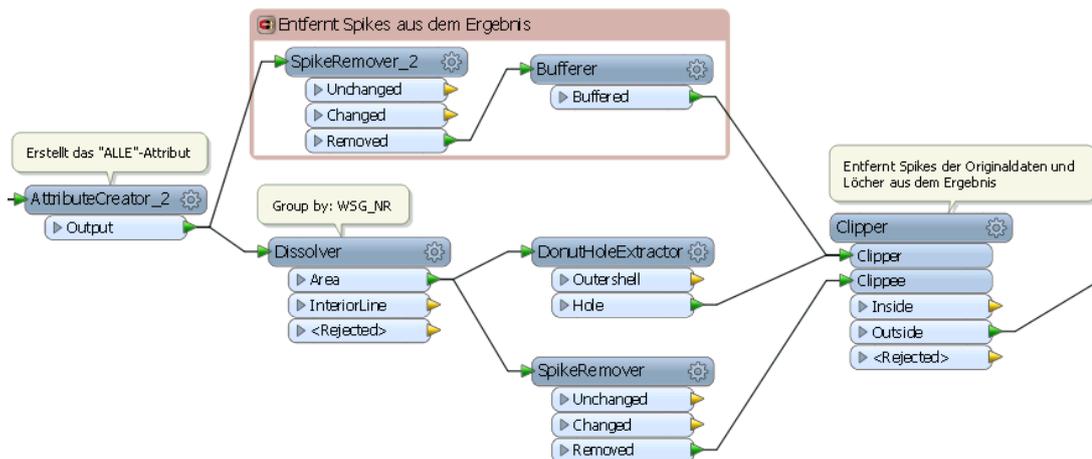


Abbildung 75: Prüfung auf Klaffen

7.2.2.15 Prüfung der Kantenidentität benachbarter Geometrien

Um die Kantenidentität von benachbarten Geometrien zu überprüfen muss bei jedem Stützpunkt eines Polygon überprüft werden, ob ein Stützpunkt eines anliegenden Polygons über identische Koordinaten verfügt. Dieser Prüfprozess soll aneinandergrenzende Zonen eines Wasserschutzgebiets überprüfen. Außerdem sollen die Außengrenzen der WSG ausgenommen werden, da es an diesen Stellen nicht zwangsläufig benachbarte Flächen gibt. Dieser Prozess benötigt ein Attribut zur eindeutigen Zuordnung einer Geometrie zu einem Wasserschutzgebiet (siehe 7.2.2.11).

Die Stützpunkte der Eingangsdaten werden in Punktfeatures konvertiert. Anschließend werden die erstellten Punkte nach dem Zuordnungsattribut gruppiert, miteinander verschnitten. Dabei wird ein Attribut erzeugt, das beinhaltet, wie viele Punkte sich dabei überlagert haben. Sollte der Wert des Attributs 0 betragen, bedeutet dies, dass der Punkt mit keinem weiteren Punkt übereinstimmt und somit nicht mit einem Punkt einer benachbarten Geometrie identisch ist.

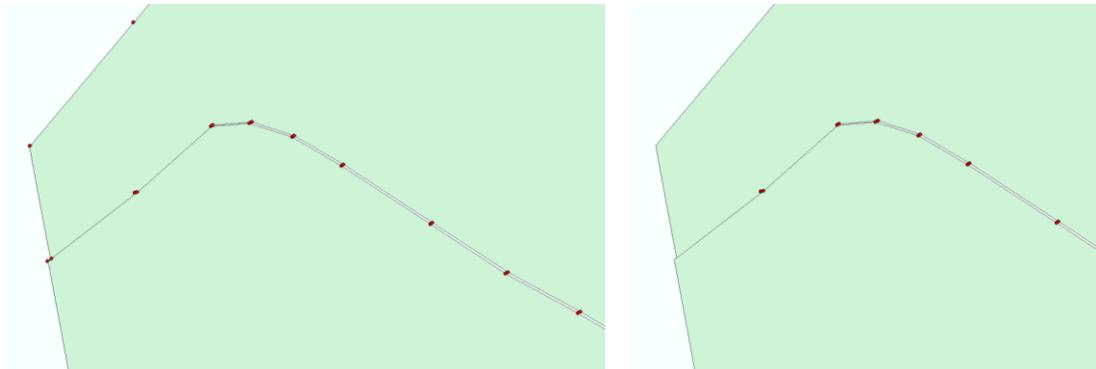


Abbildung 76: Ergebnis vor und nach dem Ausschneiden der Punkte auf der Außengrenze

Gleichzeitig werden die Inputdaten nach dem Zuordnungsattribut aufgelöst, wodurch die Außengrenze des Wasserschutzgebiets entsteht. Im nächsten Schritt werden die Punkte, die auf der Außengrenze liegen, ausgeschnitten (siehe Abbildung 76). Hierdurch bleiben nur die Punkte übrig, die über keinen identischen Punkt verfügen und an einer Grenze zweier Zonen liegen. Abbildung 77 stellt diesen Prozess mit den verwendeten Transformern dar.

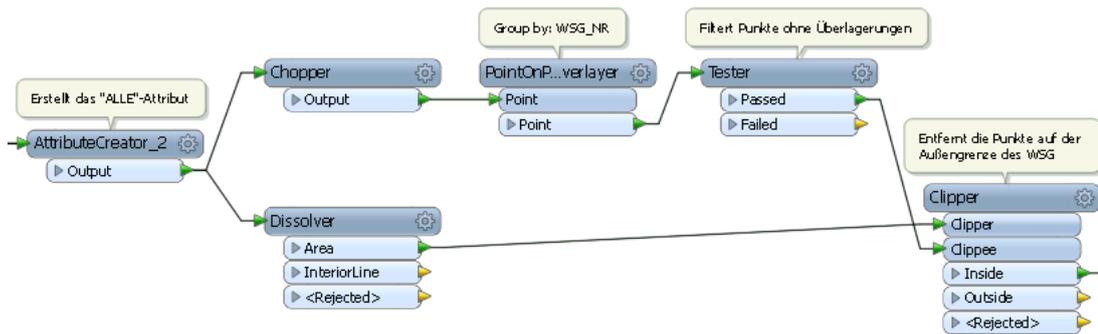


Abbildung 77: Prüfung der Kantenidentität benachbarter Geometrien

7.2.2.16 Prüfung der Anpassung an die ALK-Daten

Dieser Prozess soll überprüfen, ob alle Stützpunkte der WSG-Geometrien an die ALK-Daten gesnappt sind. Dies schließt ALK-Punkte, Linien und Flurstücke ein. Da für jede Prüfung die gleichen ALK Daten benötigt werden, werden diese nicht über einen Reader in den Prozess integriert. Stattdessen wird ein FeatureReader Transformer eingesetzt, welcher die aktuellen ALK-Daten während der Transformation aus der SDE-Datenbank ausliest. Sämtliche Attribute der ALK-Daten werden entfernt. Dies dient dazu, den Rechenaufwand zu verringern, da die Attribute der ALK-Daten für den Validierungsprozess nicht benötigt werden (siehe Abbildung 78)

Im ersten Prüfschritt werden die Stützpunkte der Input-Geometrien in Punktfeatures umgewandelt und mit einem Attribut zur Kennzeichnung versehen.

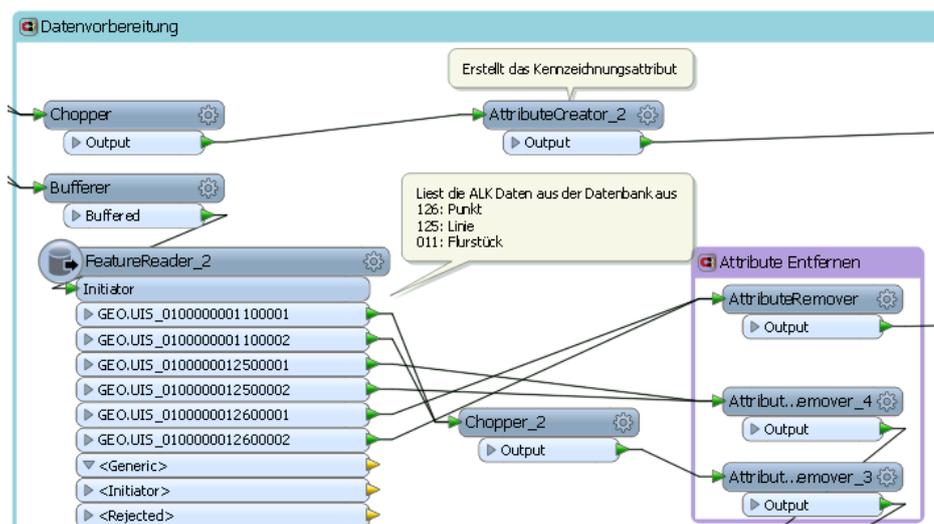


Abbildung 78: Import der ALK-Daten in den Prozess und Aufbereitung der Daten für die Analyse

Im ersten Prüfschritt werden die Stützpunkte der WSG-Geometrien mit den ALK-Punkten abgeglichen. Hierzu werden beide Punktdatensätze verschnitten und mit einem Attribut versehen, welches die Anzahl der Überlagerungen enthält. Jedoch werden hierbei die beiden Punktdatensätze zu einem Datensatz vermischt. Schließlich werden die Punkte gefiltert. Erst werden die ALK-Punkte, die über das Kennzeichnungsattribut verfügen wieder aus dem Datensatz entfernt. Dann werden Punkte die Überlagerungen mit anderen Punkten aufweisen, aus dem Datensatz entfernt. Diese sind an die ALK-Punkte gesnappt sind und müssen deshalb nicht weiter untersucht werden.

Im zweiten Prüfschritt wird das Snapping an die ALK-Linien sowie die Kanten der ALK-Flurstücke getestet. Die Kanten der Flurstücke werden hierfür in Linien umgewandelt. Dies verhindert, dass Punkte als gesnappt gewertet werden, die nur innerhalb der Flurstücksfläche liegen. Außerdem ermöglicht es die Anpassung an ALK-Linien und ALK-Flurstücke in einem Schritt zu testen. Die Punkte, die im vorherigen Schritt übrig geblieben sind, werden nun mit den ALK-Linien und Flurstückskanten verschnitten. Hier wird ebenfalls ein Überlappungsattribut erstellt, alle Punkte deren Überlappungsattribut nun über den Wert 0 hat, werden als Fehlershape ausgegeben, da diese an keine ALK-Geometrien angepasst sind. Abbildung 79 stellt die Verschneidungs- und Testprozesse mit den verwendeten Transformern dar.

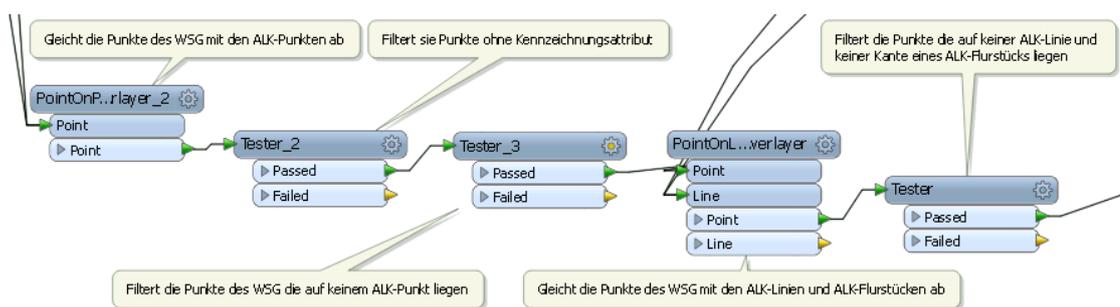


Abbildung 79: Prüfung der Anpassung an die ALK-Daten, durch Verschneiden der Stützpunkte der Geometrien mit den ALK-Daten

7.2.2.17 Korrekturvorbereitung

Die Korrekturvorbereitung dient dazu, die zu bearbeitenden Daten in einem Shapefile zu speichern, damit die Korrekturwerkzeuge auf einen Datensatz zugreifen können, der verändert werden kann. Dies sichert die Originaldaten vor Veränderungen und umgeht mögliche fehlende Schreibrechte. Außerdem werden die Sachdatenattribute der Ausgangsdaten, die bei den Korrekturen nicht benötigt werden, zwischengespeichert und entfernt, um den Rechenaufwand der Korrekturwerkzeuge zu verringern.

Dieser Prozess verfügt über einen SDE-Reader und einen Generic-Reader, sowie einen Shapefile-Writer. Außerdem wird auch in diesem Prozess ein LogFile geschrieben.

Zuerst werden alle Sachattribute der Wasserschutzgebiete in die Transformation integriert (siehe Abbildung 80), während alle anderen Attribute entfernt werden.

Diese Transformation verfügt ebenfalls über die Testprozesse zur Auswahl des zu bearbeitenden Wasserschutzgebiets. Nach der Auswahl werden die Daten an den Shapefile-Writer weitergegeben, der sie für die Korrekturprozesse zwischenspeichert, der Writer Feature Type verfügt über keine Attribute.

Die Daten mit den Attributen werden in Singlepart-Polygone konvertiert. Die durch den AreaBuilder als Listen gespeicherten Attribute werden anschließend wieder in normale Attribute konvertiert. Abschließend werden die Polygone in Punkte konvertiert und von einem weiteren Shapefile-Writer für das Wiederherstellen der Attribute in den korrigierten Daten zwischengespeichert. Diese Punkte können in einem späteren Schritt mit den korrigierten Geometrien verschnitten werden. Hierdurch werden die Attribute wieder an die korrigierten Daten angefügt.

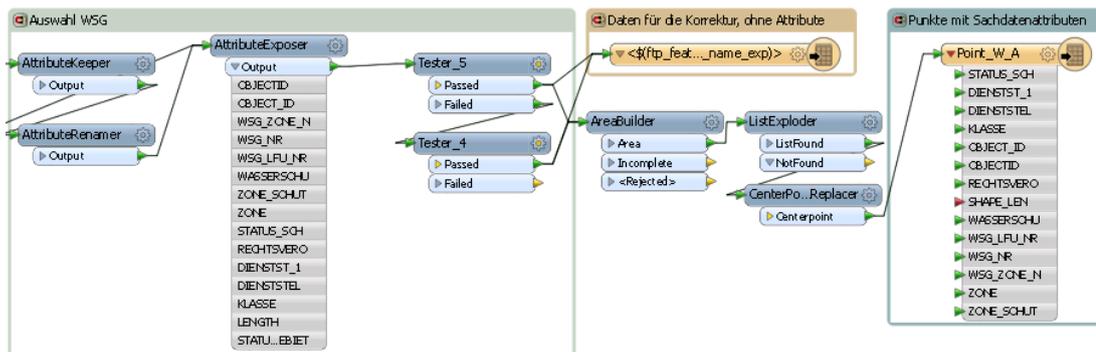


Abbildung 80: Erstellung des Punktshapes mit den Attributen

7.2.2.18 Korrektur doppelter Stützpunkte

Die Korrektur doppelter Stützpunkte erfolgt durch das Aufteilen der Geometrien in ihre einzelnen Stützpunkte, das Entfernen der fehlerhaften Punkte sowie das Wiederherstellen der Geometrien. Im Fall von Stützpunkten, die nicht identisch sind, aber innerhalb eines zu kleinen Abstands liegen, werden beide Punkte durch einen Punkt ersetzt, der genau in der Mitte der ursprünglichen Punkte liegt. Splitterpolygone am Rand einer Geometrie werden bei der Korrektur ausgenommen, da eine automatische Korrektur hier weitere Fehler hervorrufen würde und ohnehin manuell nachbearbeitet werden muss (siehe Abbildung 81).



Abbildung 81: Fehler der bei der Korrektur mit Splitterflächen entsteht

Vor der eigentlichen Datenverarbeitung erfolgt die Vorbereitung. Zuerst werden die Flächen in für interne FME-Standards topologisch korrekte Flächen umgewandelt. Dies verhindert kleinere Fehler wie Flächen, die unbeabsichtigt zusammengefügt werden (siehe Abbildung 82). Der gesamte Vorbereitungsprozess ist in Abbildung 84 dargestellt.

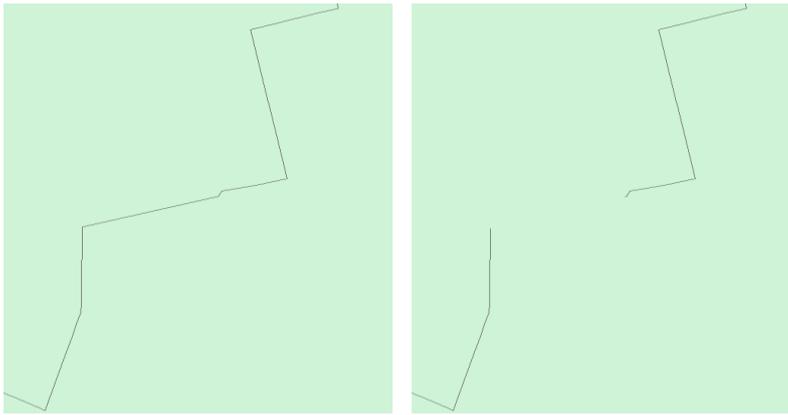


Abbildung 82: Unbeabsichtigt zusammengefügte Polygone

Die Geometrien werden außerdem mit einer eindeutigen ID versehen und Löcher werden entfernt. Dase die bei der Weiterverarbeitung korrupte Geometrien erzeugen würden (siehe Abbildung 83).



Abbildung 83: Korrupte Geometrien, die bei der Korrektur mit Löchern entstehen (rechts) im Vergleich zu richtig korrigierten Geometrien (links)

Die Splitterflächen, die Probleme verursachen können, werden mit einem vom Rest des Polygons abgetrennt und extrahiert. Mit einem Buffer werden wie Splitterflächen vergrößert, damit mit ihnen alle in ihnen liegenden Punkte aus dem Datensatz ausgeschnitten werden können (siehe Abbildung 85, oben).

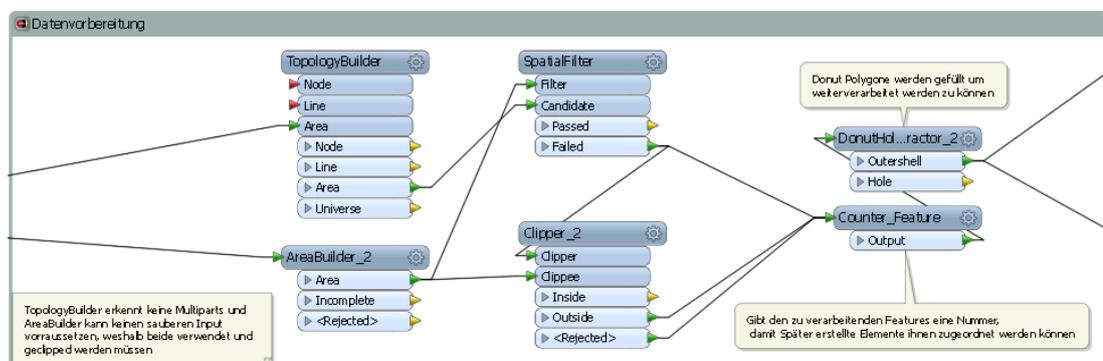


Abbildung 84: Übersicht Datenvorbereitung

Im ersten Teil der Korrektur werden die doppelten Stützpunkte identifiziert. Hierzu wird der jeweils letzte Stützpunkt der Polygone entfernt, da dieser mit dem ersten Stützpunkt identisch ist und das Ergebnis verfälschen würde (siehe 7.2.2.1). Anschließend werden die Stützpunkte in Punktfeatures transformiert und mit einem weiteren Counter werden die Punkte nummeriert um in einem späteren Schritt ihre ursprüngliche Reihenfolge wiederherstellen zu können. Im folgenden Schritt werden nun alle Punkte, die innerhalb eines, durch einen Published User Parameter angegebenen, Mindestabstands liegen identifiziert. Alle Punkte für die kein benachbarter Punkt gefunden wurde, werden an den Wiederherstellungsteil (siehe Abbildung 86) weitergegeben. Doppelte und zu nahe beieinanderliegende Punkte, werden weiterverarbeitet. Anschließend werden die Punkte ausgeschnitten, die innerhalb einer Splitterfläche liegen und ebenfalls an den Wiederherstellungsteil weitergegeben. Alle doppelten Punkte erhalten nun einen Buffer, mit der breite des Mindestabstands. Die aneinander angrenzenden Buffer werden zu einer Fläche zusammengefügt. Diese Fläche wird durch ihren Mittelpunkt ersetzt, der nun ebenfalls an den Wiederherstellungsteil weitergegeben wird (siehe Abbildung 85). Alle so erzeugten Punkte werden zusätzlich als Shapefile ausgegeben um es dem Nutzer zu ermöglichen, die einzelnen Korrekturen nachzuvollziehen.

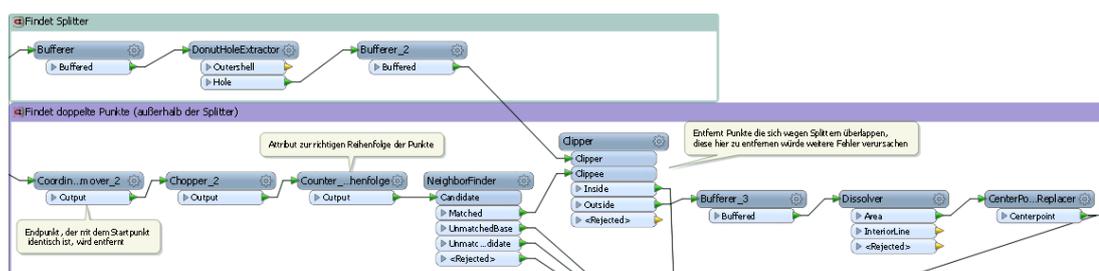


Abbildung 85: Identifizieren und entfernen doppelter Stützpunkte

Im letzten Teil der Korrektur erfolgt die Wiederherstellung der Polygone. Alle in den vorherigen Schritten ausgefilterten oder erzeugten Punkte werden nun nach ihrer Reihenfolge sortiert. Anschließend werden alle Punkte, entsprechend ihrer Reihenfolge, in einer Polylinie umgewandelt. Diese Polylinien werden geschlossen und in Polygone konvertiert. Abschließend werden Löcher für niedrigere WSG-Zonen in die Polygone eingefügt. Da die ursprünglich vorhandenen Löcher während der Datenvorbereitung

geschlossen wurden. Abbildung 86 stellt die Wiederherstellung der Polygone mit den verwendeten Transformatoren dar.

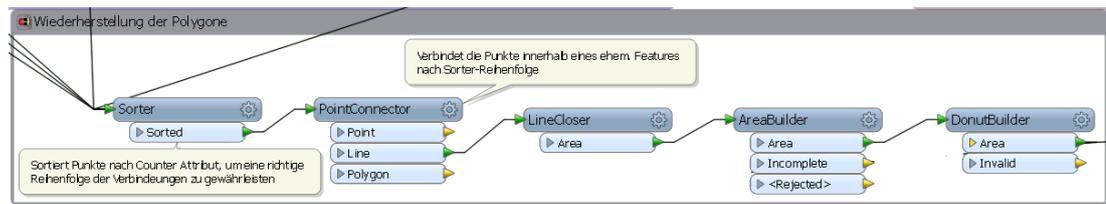


Abbildung 86: Wiederherstellen der ursprünglichen Polygone

7.2.2.19 Korrektur von Spikes

Das Werkzeug zur Entfernung von Spikes setzt den von FME angebotenen SpikeRemover Transformer ein. Der Winkel der Spikes definiert, kann durch einen Published User Parameter angegeben werden.

Die entfernten Spikes werden zusätzlich als Shapefile ausgegeben, damit der Nutzer die Korrekturen nachvollziehen kann (siehe Abbildung 87).

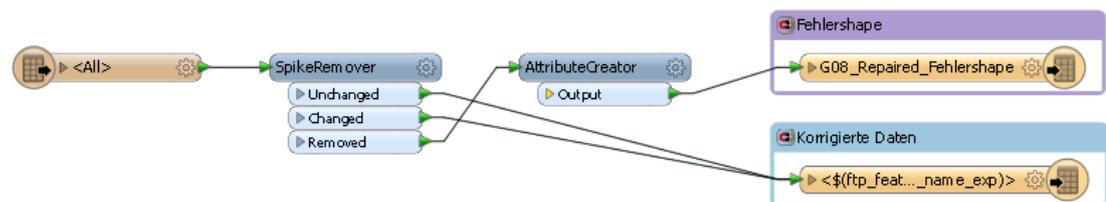


Abbildung 87: Korrektur von Spikes

7.2.2.20 Korrigierte Features mit Attributen versehen

Dieser Prozess verfügt über zwei Generic Reader, einen für die korrigierten Geometrien und einen für die in der Korrekturvorbereitung erstellten Punkte mit den Sachdatenattributen. Außerdem gibt es einen Shapefile Writer, sowie die Transformer zum Erstellen des LogFile.

Die Attribute der Punkte werden zuerst in den Workspace integriert. Anschließend werden die Punkte mit den Sachdatenattributen mit den korrigierten Polygonen verschnitten, wodurch diese mit Attributen versehen werden. Abschließend werden Polygone die über die exakt gleichen Attributwerte verfügen miteinander verschmolzen, wodurch die ursprünglichen Multipart-Polygone wiederhergestellt werden.

Diese räumliche Verschneidung wird durchgeführt, da eine Zuordnung über Attribute wie IDs, nicht möglich ist. Innerhalb der Korrekturprozesse werden alle Attribute entfernt und die Daten so umgewandelt, dass nicht entfernte Attribute verloren gehen oder verändert werden würden.

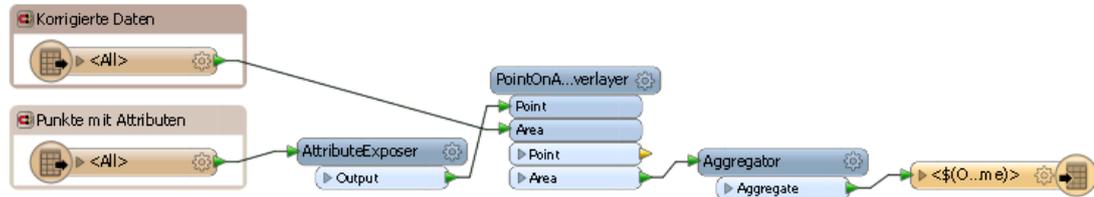


Abbildung 88: Wiederanfügen der Attribute und Wiederherstellung der Multipart-Polygone

7.2.3 Der Master Workspace

Der Master Workspace enthält für jede Validierung und für jede Korrektur eine Gruppe von Transformatoren, welche die Ausführung des jeweiligen Prozesses steuert. Die Validierungsprozesse werden in einer Reihe nacheinander durchgeführt, ebenso wie die Korrekturen. Eine vollständige Übersicht über den Master Workspace befindet sich in Anhang I.

Da der Master Workspace die Daten nicht direkt verarbeitet, sondern Ihre Verarbeitung steuert, wird auf einen Reader verzichtet. Anstelle des Readers wird ein Creator Transformer eingesetzt, der ein Feature erstellt, welches nur dazu dient die Transformer des Master Workspace anzustoßen. Der Writer ist ein Shapefile-Writer, welcher nur das Feature ausgibt, mit dem der Prozess gestartet wurde.

Im ersten Schritt des Master Workspace wird abgefragt ob eine Korrektur, eine Validierung oder beides durchgeführt werden soll. Hierfür wird ein durch den Nutzer veränderbarer Published User Parameter ausgelesen. Von hier aus wird entweder zu den Validierungs- oder zu den Korrekturwerkzeugen weitergeleitet, welche über WorkspaceRunner Transformer aufgerufen werden.

Der Master Workspace enthält für jede Überprüfung, die durchgeführt werden kann, einen User Parameter, der die Auswahlmöglichkeiten „Ja“ und „Nein“ enthält. Vor jedem Validierungsprozess wird der jeweilige Abfrage-

User-Parameter ausgewertet. Falls im Parameter „Ja“ angegeben ist, wird die Transformation zum entsprechenden Überprüfungs-WorkspaceRunner weitergeleitet, bei „Nein“ wird mit der nächsten Abfrage fortgefahren. Vor der Abfrage der einzelnen Parameter wird ein weiterer Parameter abgefragt, der die Information enthält ob das gesamte Validierungspaket für Wasserschutzgebiete durchgeführt werden soll. Hier wird ebenfalls ein Parameter mit „Ja“ oder „Nein“ abgefragt. Bei „Ja“ wird die Einzelabfrage übersprungen und die Validierung durchgeführt, bei „Nein“ wird mit der Einzelabfrage fortgefahren. Durch die Angabe von „Ja“ bei der WSG-Validierung werden so alle Detailsinstellungen ignoriert (siehe 7.2.3.2). Nach der durchgeführten Validierung wird der Master Workspace mit der nächsten Abfrage fortgesetzt. Die einzelnen Korrektur- und Validierungsprozesse warten grundsätzlich bis der vorherige Prozess abgeschlossen wurde um eine zu große Anzahl gleichzeitig laufender Prozesse zu vermeiden. Abbildung 89 zeigt den Ablauf der Validierungsabfragen.

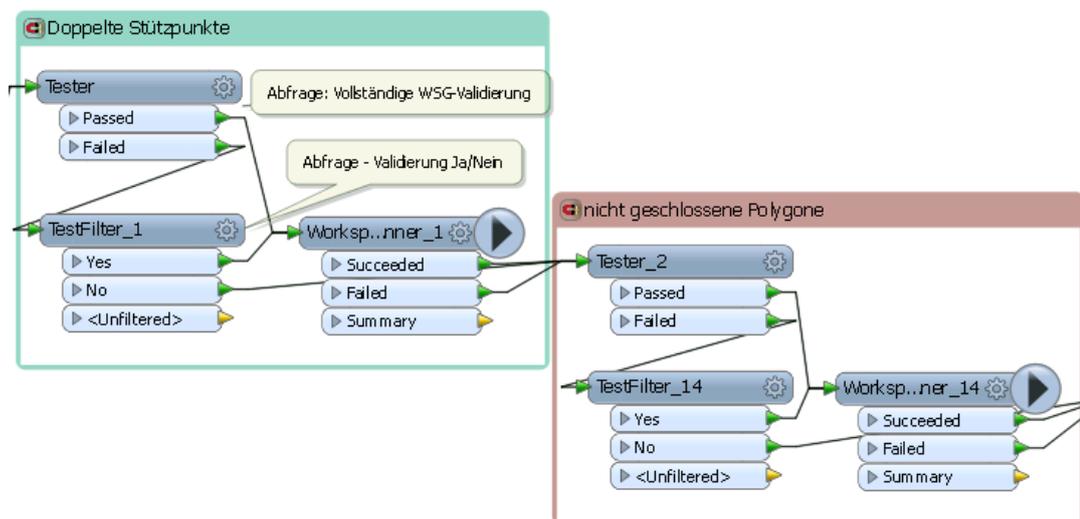


Abbildung 89: Verkettete Validierungsprozesse, mit Tester und TestFilter für die Parameterabfragen

Die Validierungsprozesse lesen immer dieselben Inputdaten, während im Korrekturbereich das Ergebnis eines Prozesses die Ausgangsdaten des nächsten Prozesses bildet. Das Ergebnis der letzten Korrektur bildet gleichzeitig den Input der Validierung der korrigierten Daten, falls diese durchgeführt werden soll und im entsprechenden Parameter „Ja“ ausgewählt wurde.

Für die Durchführung der Korrekturen werden die Datensätze als Shapefile exportiert. Die Auswahl der Inputdaten erfolgt dynamisch über ein Attribut, welches dynamisch nach jeder erfolgreich durchgeführten Korrektur angepasst wird. Dieses Attribut gibt den Pfad des Shapefiles an den WorkspaceRunner weiter. Da Attribute dynamisch angepasst werden können, werden sie hier anstelle von Parametern eingesetzt. Diese dynamische Anpassung ermöglicht es, beliebige Korrekturen durchzuführen oder auszulassen und auf fehlgeschlagene Transformationen zu reagieren. Des Weiteren können so alle Zwischenergebnisse gespeichert werden, anstatt sie zu überschreiben, wodurch der gesamte Korrekturprozess für den Nutzer nachvollziehbar bleibt.

Zusätzlich zu den Abfragen und dem WorkspaceRunner werden in jedem Korrekturblock die Attribute mit den Namen und Pfad der In- und Outputdaten dynamisch angepasst. Zwischen der Abfrage und dem WorkspaceRunner wird zuerst ein Zähler („count“) hochgezählt, welcher anschließend in den Namen des Outputdatensatzes („OutName“) eingefügt wird. Im Fall eines erfolgreichen Durchlaufs des WorkspaceRunners wird das Attribut erneut angepasst, das Ergebnis der Korrektur wird in den Pfad der Eingangsdaten („InName“) eingefügt. Dieser Ablauf ist in Abbildung 90 dargestellt.

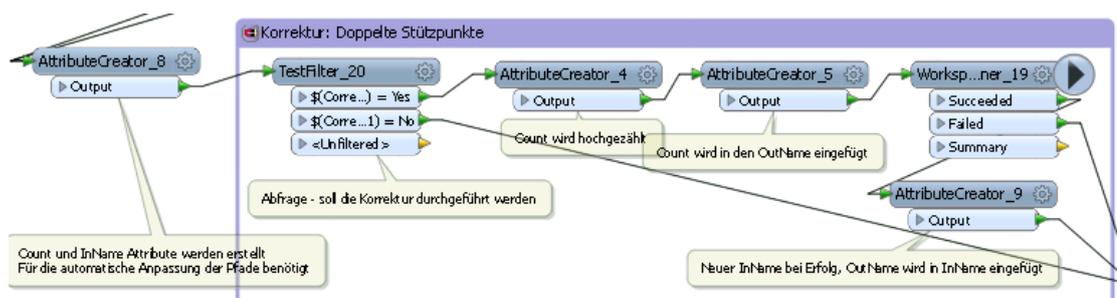


Abbildung 90: Korrekturblock mit den Attributanpassungen (Count, InName, OutName)

7.2.4 Steuerung des Werkzeugs

Im Master Workspace lässt sich das Werkzeug über verschiedene User Parameter steuern. Diese Parameter steuern Input- und Outputdaten sowie die durchzuführenden Validierungen und Korrekturen. Außerdem lassen sich Werte wie Toleranzen hier einstellen. Abbildung 91 zeigt die Eingabemaske mit der beim Starten des Master Workspace die Parameter gewählt werden können.

7.2.3.1 Input- und Output-Parameter

Diese Parameter steuern welche Daten verarbeitet werden sollen. Außerdem werden der Speicherort und das Format der Outputdaten verwaltet.

Parameter	Verwendung	Anmerkung
Input_Shape	Inputdaten (Shapefile)	Als Input kann entweder ein Shapefile oder ein Datensatz aus der SDE ausgewählt werden, wenn ein SDE-Datensatz ausgewählt wird, muss bei ArcSDE_Dataset „sde“ eingetragen werden.
SDE_Feature_Type	Inputdaten (SDE)	
ArcSDE_Dataset	SDE-Dataset	
WSG_NR	WSG-Nummer des zu überprüfenden WSG	Wird bei All_Features „No“ ausgewählt, kann bei WSG_NR eine beliebige WSG-Nummer eingegeben werden. Ist „Yes“ ausgewählt, wird der vollständige Datensatz validiert.
All_Features	Vollständigen Input-Datensatz validieren Yes/No	
Output_Format	Datenformat der erzeugten Fehlershapes	z.B. Shapefile oder File Geodatabase
Output_Folder	Ordner in dem die erzeugten Fehlershapes und die LogFiles gespeichert werden	
LogFile-Name	Name der LogFiles	

Tabelle 4: Übersicht: Input- und Output-Parameter

7.2.3.2 Testparameter

Diese Parameter steuern welche Validierungen und Korrekturen durchgeführt werden sollen.

Parameter	Verwendung	Anmerkung
Korrektur_oder_Validierung	Soll eine Korrektur, Validierung oder beides durchgeführt werden	
Test_nach_Korrektur	Soll nach der Korrektur eine zusätzliche Validierung durchgeführt werden.	
WSG_Validierung	Sollen Alle Validierungen für WSG durchgeführt werden?	Yes: Auswahlen der Test-Parameter werden ignoriert und mit Yes überschrieben No: Auswahlen der Testparameter werden ausgeführt
Test_G01	Doppelt oder zu nah gesetzte Stützpunkte	Benötigt den Mindestabstand-Parameter
Test_G02	Nicht geschlossene Polygone	
Test_G03	Falsch orientierte Polygone	
Test_G04	Splitterpolygone	
Test_G05	Löcher innerhalb eines Polygons	
Test_G06	Fischschwanz-Polygone	
Test_G07	Selbstüberschneidungen	
Test_G08	Spikes	Benötigt den Spike-Winkel-Parameter
Test_G09	Mindestabstände	Benötigt den Mindestabstand-Parameter
Test_G10	ALK-Anpassung	Benötigt den ALK_Toleranz-Parameter
Test_G11	Multipart-Polygone	
Test_T01	Überlappungen	Benötigt den Group_by_Attribut-Parameter
Test_T02	Lücken zwischen Polygonen	Benötigt den Group_by_Attribut-Parameter
Test_T03	Sliver Polygone	Benötigt den Group_by_Attribut-Parameter
Test_T04	Klaffen	Benötigt den Group_by_Attribut-Parameter
Test_T05	Snapping an benachbarte Geometrien	Benötigt den Group_by_Attribut-Parameter

Correct_G01	Korrektur von doppelten Stützpunkten	Benötigt den Mindestabstand-Parameter
Correct_G08	Korrektur von Spikes	Benötigt den Spike_Winkel-Parameter

Tabelle 5: Testparameter

7.2.3.3 Werkzeugspezifische Parameter

Diese Parameter umfassen alle Parameter, die für die Durchführung einzelner Validierungs- und Korrekturwerkzeuge benötigt werden. In erster Linie sind dies Toleranzparameter sowie Angaben zu Attributen, die für die jeweilige Transformation benötigt werden.

Parameter	Verwendung	Anmerkung
Mindestabstand	Mindestabstand zwischen Stützpunkten sowie zwischen den einzelnen Ringen und Teilflächen eines Polygons.	Wird für die Validierung und Korrektur von doppelten Stützpunkten sowie die Prüfung der Mindestabstände benötigt.
Spike_Winkel	Winkel unterhalb dessen ein Spike vorliegt.	Wird für die Validierung und Korrektur von Spikes benötigt.
ALK-Toleranz	Erlaubter Abstand eines Stützpunktes zu einem ALK Objekt.	Wird für die Prüfung der ALK-Anpassung benötigt.
Group_by_Attribut	Attribut nachdem die einzelnen Features, bei der Verarbeitung, gruppiert werden sollen.	Wird für die Prüfung auf Überlappungen, Lücken, Sliver Polygone, Klaffen und die Anpassung an benachbarte Geometrien benötigt. Wenn kein passendes Attribut (z.B. WSG_NR) vorhanden ist, kann auch „ALLE“ ausgewählt werden.

Tabelle 6: Werkzeugspezifische Parameter

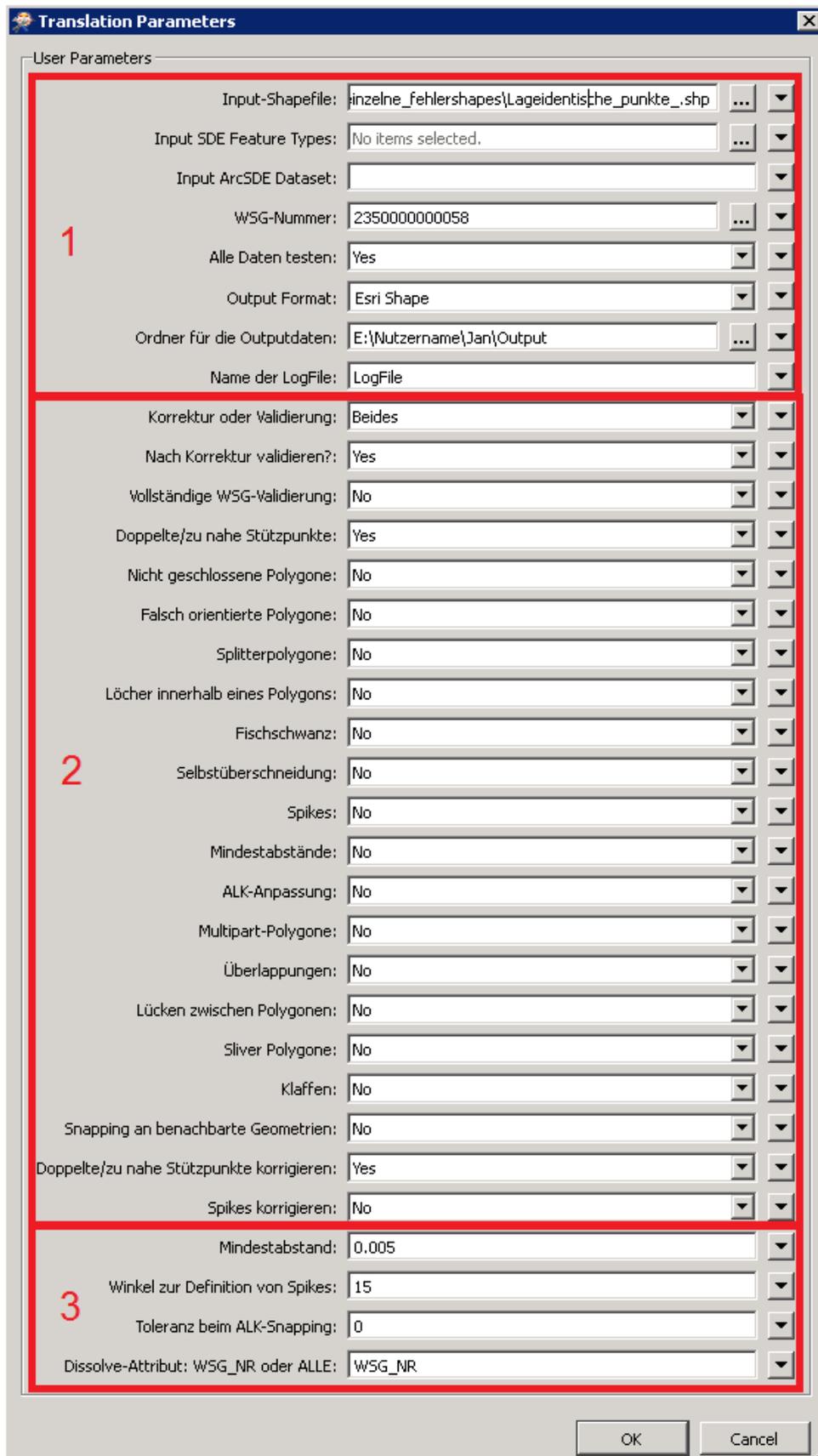


Abbildung 91: Parameterübersicht: 1. Input- und Output-Parameter, 2. Testparameter, 3. Werkzeugspezifische Parameter

8 Vorschläge für die Qualitätssicherung der LUBW

Aktuell liegt die Verantwortung für die Qualitätssicherung der Geodaten beim ITZ der LUBW, während die Verantwortung für die Erfassung und Digitalisierung und damit für die Qualität der Daten bei anderen Behörden liegt. Hierdurch entsteht ein zusätzlicher Arbeitsaufwand für die Mitarbeiter des ITZ. Daher sollte ein gewisser Teil der Qualitätssicherungsaufgaben ebenfalls an die verantwortlichen Stellen abgegeben werden. Diese können von der Durchführung der Validierung bis hin zu Korrekturen reichen. Eine Möglichkeit wäre, das im Rahmen dieser Arbeit erstellte Werkzeug den unteren Wasserbehörden zur Verfügung zu stellen. Behörden, die über FME verfügen, können das Werkzeug direkt einsetzen. Dies kann beispielsweise mit FME Server durchgeführt werden, wodurch eine Validierung via Internet möglich wäre. Dies ist außerdem unabhängig von der Softwareinfrastruktur der einzelnen Behörden. Der Einsatz einer Onlinevalidierung bringt jedoch auch Probleme mit sich. Zum einen müssen datenschutzrechtliche Aspekte beachtet werden, da es sich bei Daten wie den Wasserschutzgebieten um sensible Daten handelt. Zum anderen muss auch der vergleichsweise hohe Speicherplatzbedarf auf den Servern, auf denen die Validierungen durchgeführt würden, beachtet werden. Außerdem muss die Auslastung der jeweiligen verantwortlichen Stellen beachtet werden, da auch hier Qualitätssicherungsmaßnahmen einen zusätzlichen Arbeitsaufwand bedeuten würden.

Ein weiterer Vorschlag ist die Vereinheitlichung der Softwareinfrastruktur, da verschiedene Systeme verschiedene Ansprüche an Daten haben. Außerdem werden Datenkonvertierungen notwendig, bei denen Daten beschädigt werden können. So kann beispielsweise eine Konvertierung in das Shapefile-Format vermieden werden, da es eine hohe Fehlertoleranz aufweist. Durch einheitliche Systeme mit einheitlichen Vorgaben an die Daten lassen sich festgelegte Standards einfacher durchsetzen. Dies ist jedoch mit hohen Kosten verbunden.

9 Fazit und Ausblick

Das primäre Ziel dieser Arbeit war es die Qualitätssicherung des ITZ mit einem Werkzeug zur Geodatensvalidierung zu unterstützen. Dieses Werkzeug soll in die aktuellen Qualitätssicherungsprozesse integriert werden und diese so voranbringen.

Die Analyse der Qualitätskriterien für Geodaten bildet die Grundlage für die Festlegung von spezifischen Kriterien für die Daten der LUBW. Diese Kriterien wurden schließlich, in Abstimmung mit den verantwortlichen Mitarbeitern, festgelegt.

Mit Hilfe von FME wurde ein Prototyp dieses Werkzeugs erstellt, der aktuell 16 verschiedene geometrische und topologische Validierungen durchführen kann. Außerdem wurden zwei Prozesse zur Datenkorrektur in den Prototypen integriert. Da die aktuellen Prozesse beispielhaft mit den Wasserschutzgebietsdaten entwickelt wurden, sind die Prozesse auf diese hin angepasst. Dies ist daran erkennbar, dass einige Validierungsprozesse noch von speziellen Attributen wie der WSG-Nummer abhängig sind. Einige geometrische Validierungen sind bereits universell einsetzbar, während die topologischen Validierungen noch an die Datenstruktur der Wasserschutzgebiete angepasst sind.

Denkbar ist eine Weiterentwicklung des Werkzeugs für weitere Datenbestände wie die Mähwiesen. Neben der Erweiterung des Funktionsumfangs des Prototyps ist auch eine Optimierung der bisherigen Funktionen geplant. Dies umfasst beispielsweise die Verbesserung der Performance einzelner Werkzeuge sowie die weitere Flexibilisierung des Daten-Inputs.

Der Einsatz des Werkzeugs in den aktuellen Qualitätssicherungsprozessen innerhalb des ITZ ist bereits möglich. Somit kann das Ergebnis dieser Arbeit die Arbeit mit dem bisherigen veralteten FME-Prozess unterstützen, wodurch eine zuverlässige Datensvalidierung auch weiterhin möglich sein wird.

Während die Datensvalidierung durch automatisierte Prozesse leicht optimiert werden kann, ist eine automatische Datenkorrektur nur bedingt umsetzbar. Einerseits können vermeintliche Fehler in den Daten beabsichtigt sein und

Ausnahmen in den vordefinierten Regeln bilden. Auf der anderen Seite kann die Korrektur eines Fehlers neue Probleme verursachen. Das Entfernen eines falsch gesetzten Stützpunktes kann beispielsweise die Anpassung an die ALK oder benachbarte Geometrien beeinträchtigen, was zu Lücken oder Überlappungen führen kann. Deshalb sollte eine automatische Korrektur grundsätzlich nochmals überprüft werden, bevor die Daten weiterverwendet werden.

Eine weitere Option für die zukünftige Entwicklung des Werkzeugs ist der Einsatz mit FME Server. Da jedoch die Entwicklung mit FME Desktop im Vordergrund stand, wurden bisher keine Tests mit FME Server durchgeführt. Damit wäre auch eine Online-Validierung möglich und neben dem ITZ würden auch andere Behörden wie die UWB vom Ergebnis dieser Arbeit profitieren.

Quellenverzeichnis

[ARCUIS]

ArcUIS-Tools Benutzerhandbuch, Ministerium für Umwelt, Klima und Energiewirtschaft Baden-Württemberg, 2012

[DEV06]

Fundamentals of Spatial Data Quality, Rodolphe Devillers, Robert Jeansoulin ISTE, London, 2006

[ESRI_1]

<http://help.arcgis.com/de/arcgisdesktop/10.0/help/index.html#/00050000001000000>

Abgerufen: 07.07.2015 08:38

[ESRI_2]

<http://help.arcgis.com/de/arcgisdesktop/10.0/help/index.html#/002w00000001000000>

Abgerufen: 10.06.2015 13:02

[ESRI_3]

<http://support.esri.com/en/knowledgebase/GISDictionary/term/polygon>

Abgerufen: 24.07.2015 09:40

[GBI_1]

http://www.lubw.baden-wuerttemberg.de/servlet/is/61060/gbl_15_2005_landesanstalt_umwelt_messungen_und_naturschutz_b-w.pdf?command=downloadContent&filename=gbl_15_2005_landesanstalt_umwelt_messungen_und_naturschutz_b-w.pdf

Abgerufen: 04.05.2015 9:54

[GEOG]

Geographical Information Systems – Principles, Techniques, Applications and Management
P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind
Wiley, New York, 2005

[GEOK_1]

<http://www.spatialytics.org/projects/geokettle/>

Abgerufen: 09.06.2015 14:25

[GEOK_2]

http://docs.spatialytics.com/doku.php?id=en:spatialytics_etl:004_transformations#definition

Abgerufen: 09.06.2015 14:27

[GEOT_1]

<http://geotools.org/about.html>

Abgerufen: 09.06.2015, 09:26

[GEOT_2]

<http://docs.geotools.org/latest/userguide/geotools.html>

Abgerufen: 09.06.2015 10:49

[ISTL]

<http://www.istl.org/12-winter/article1.html>

Abgerufen: 18.05.2015 08:20

[LUBW_1]

<http://www.lubw.baden-wuerttemberg.de/servlet/is/31656/>

Abgerufen: 04.05.2015 8:15

[LUBW_2]

<http://www.lubw.baden-wuerttemberg.de/servlet/is/61060/>

Abgerufen: 04.05.2015 8:33

[RIPS_1]

http://www.fachdokumente.lubw.baden-wuerttemberg.de/servlet/is/94173/RIPS_Faltblatt_2009_Web.pdf?command=downloadContent&filename=RIPS_Faltblatt_2009_Web.pdf

Abgerufen: 06.05.2015 14:12

[RIPS_2]

http://www.fachdokumente.lubw.baden-wuerttemberg.de/servlet/is/93175/10_rips.pdf?command=downloadContent&filename=10_rips.pdf

Abgerufen: 17.07.2015 08:51

[UIS_1]

<http://www.fachdokumente.lubw.baden-wuerttemberg.de/servlet/is/90750/rkuis2006.pdf?command=downloadContent&filename=rkuis2006.pdf>

Abgerufen: 06.05.2015 15:37

[MAY07, S.3].

Data Quality Assessment, Arkady Maydanchik
Technics Productions, LLC, Bradley Beach, NJ, 2007

[LEIT]

Leitfaden zur Datenqualität in Planungsbüros und Behörden
Runder Tisch Geoinformationssysteme, München 2005

[LOSH]

D. Loshin Monitoring Data Quality Performance Using Data Quality Metrics

[ORAC_1]

http://docs.oracle.com/cd/B19306_01/appdev.102/b14255/sdo_objrelschem.htm#CBBDAFFA

Abgerufen: 07.07.2015 09:06

[SAFE_1]
<http://www.safe.com/fme/format-search/#!>
Abgerufen: 10.06.2015 07:45

[SAFE_2]
<http://cdn.safe.com/resources/whitepapers/Spatial-ETL-Whitepaper.pdf>
Abgerufen: 10.06.2015 07:57

[SAFE_3]
http://docs.safe.com/fme/html/FME_Desktop_Documentation/FME_Desktop_Help.htm#./Subsystems/FME_Transformers/Content/parallel_processing/parallel_processing.htm%3FTocPath%3DFME%2520Transformers|_____9
Abgerufen: 10.06.2015 14:35

[SAFE_4]
http://docs.safe.com/fme/html/FME_Workbench/FME_Workbench.htm#Workbench/lists_working_with.htm
Abgerufen: 17.06.2015 10:06

[SAFE_5]
http://docs.safe.com/fme/html/FME_Workbench/FME_Workbench.htm#Workbench/feature_types_attributes.htm
Abgerufen: 22.07.2015 14:10

[SAFE_6]
http://docs.safe.com/fme/html/FME_Transformers/FME_Transformers.htm#Transformers/donutholeextractor.htm
Abgerufen: 31.07.2015 16:30

[SAFE_7]
http://docs.safe.com/fme/html/FME_Transformers/FME_Transformers.htm#Transformers/spikeremover.htm?Highlight=spikeremover
Abgerufen: 31.07.2015 17:12

[SAFE_8]
http://cdn.safe.com/resources/fme/FME_Transformer-Reference-Guide.pdf
Abgerufen: 03.08.2015 17:04

[VER_1]
Verordnung des Landratsamtes Rastatt zum Schutz des Grundwassers im Einzugsgebiet der Wassergewinnungsanlage "Rheinwald" der Stadtwerke Karlsruhe GmbH – Verordnungstext

[VER_2]
Verordnung des Landratsamtes Rastatt zum Schutz des Grundwassers im Einzugsgebiet der Wassergewinnungsanlage "Rheinwald" der Stadtwerke Karlsruhe GmbH
Anlage 2 – Hydrogeologisches Gutachten zur Abgrenzung eines Wasserschutzgebietes für das Wasserwerk Rheinwald

[WAABIS]
WAABIS-Regelwerk zur Geodatenführung - Anleitung 3: Digitalisierung von Wasserschutz-, Quellenschutz- und Überschwemmungsgebieten
Ministerium für Umwelt und Verkehr Baden-Württemberg (Hrsg.), 2003

[WHG_1]

http://www.gesetze-im-internet.de/whg_2009/BJNR258510009.html#BJNR258510009BJNG000800000

§ 51

Abgerufen: 08.05.2015 11:02

[WIKI_1]

<http://de.wikipedia.org/wiki/Wasserschutzgebiet>

Abgerufen: 08.05.2015 10:48

[WIKI_2]

http://en.wikipedia.org/wiki/List_of_International_Organization_for_Standardization_standards

Abgerufen: 18.05.2015 10:15

[WIKI_3]

http://de.wikipedia.org/wiki/ISO_19113

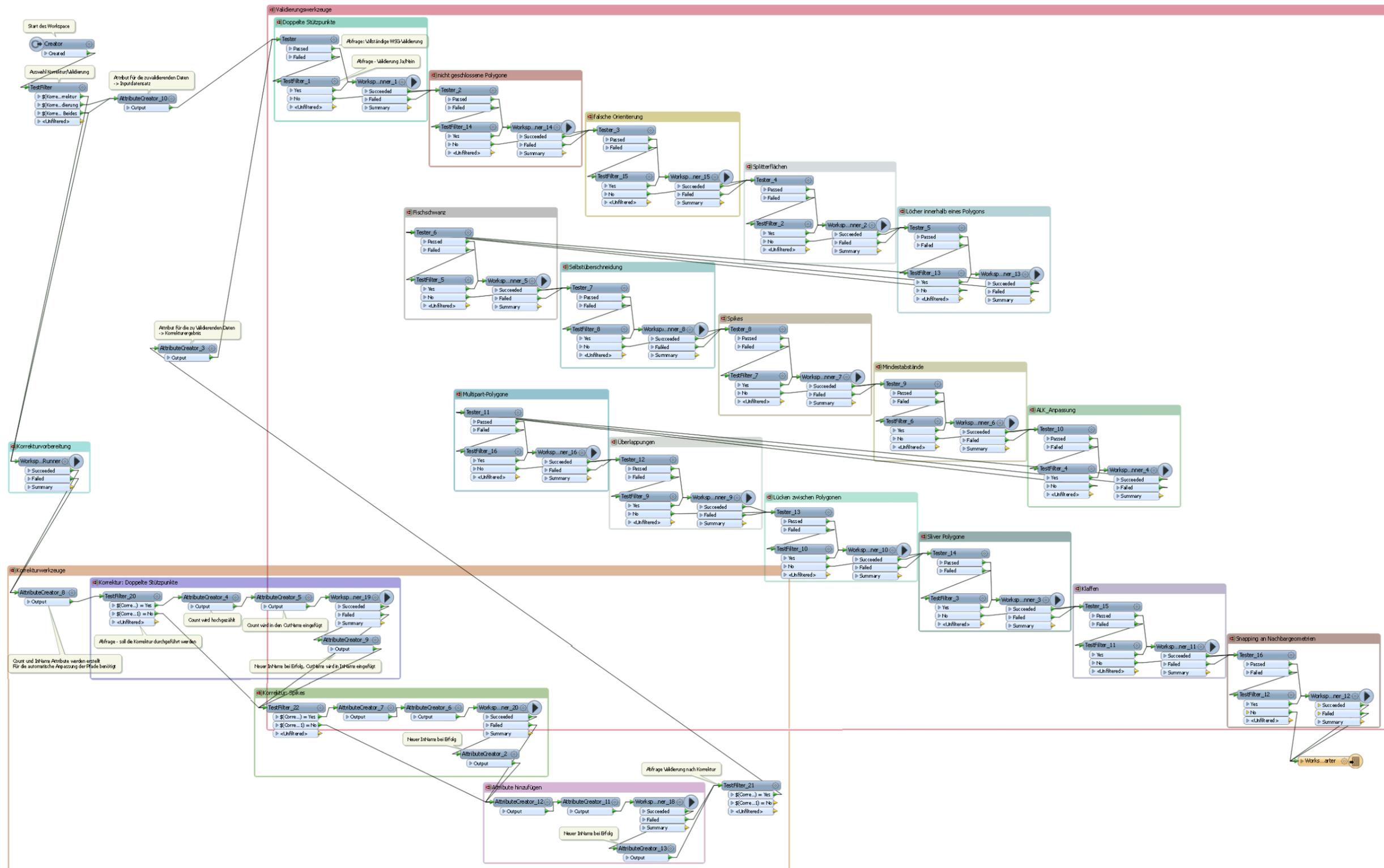
Abgerufen: 18.05.2015 10:23

[WIKI_4]

http://de.wikipedia.org/wiki/Infrastructure_for_Spatial_Information_in_the_European_Community

Abgerufen: 18.05.2015 13:03

Anhang I – Master Workspace



Anhang II – Inhalt der beiliegenden DVD

Die beiliegende DVD enthält eine digitale Version dieser Bachelorarbeit im PDF-Format. Außerdem befinden sich die einzelnen Workspaces des Korrekturwerkzeugs auf der DVD.