

Контроль качества векторных геопространственных данных

Екатерина Подольская, к. т. н., ГИС-аналитик

Содержание

- Типы геопространственных данных
- Процедуры контроля качества данных
- Проверка геометрии и атрибутов при контроле качества
- Пример организации контроля качества в геоинформационной базе данных

Часть 1. Контроль качества данных в ArcGIS (ESRI)

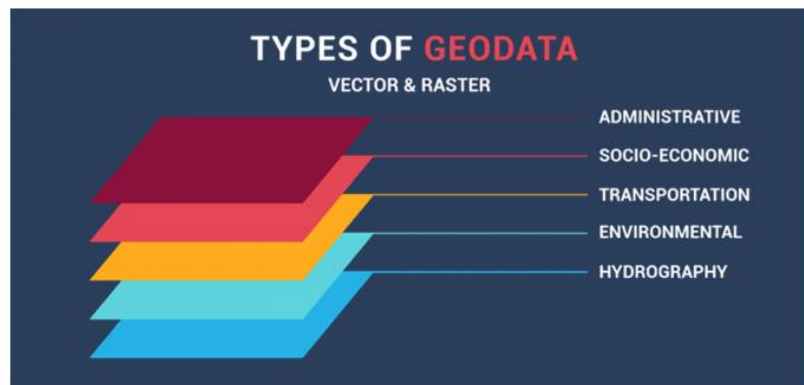
- Определение
- Основы
- Реализация
- Способы конфигурирования (настройки) проведения контроля качества
- Инструменты контроля качества данных в ArcGIS ArcMap
- Топология
- ArcGIS Data Reviewer

Часть 2. Контроль качества данных в QGIS (Open source)

- Плагины контроля качества в QGIS
- Инструменты: Delete duplicate geometries, check validity, delete holes, fix geometry, remove null geometries, remove duplicate vertices

Ссылки

Типы геопространственных данных



<https://gisgeography.com/what-is-geodata-geospatial-data/>

Процедуры контроля качества данных в ArcGIS и QGIS

- Проверка необходимого набора файлов

Векторные форматы: SHP, GDB

Форматы проектов: MXD, QGZ

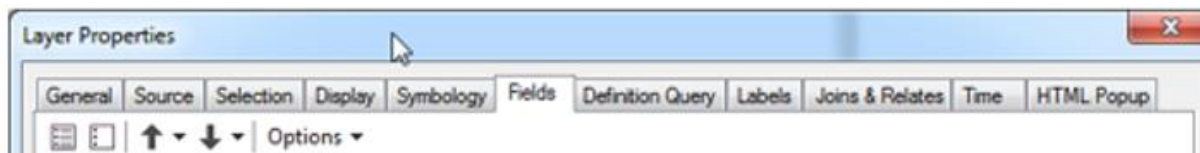
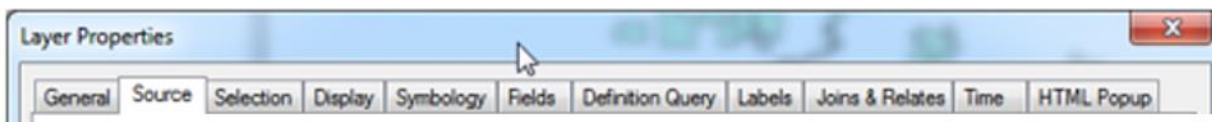
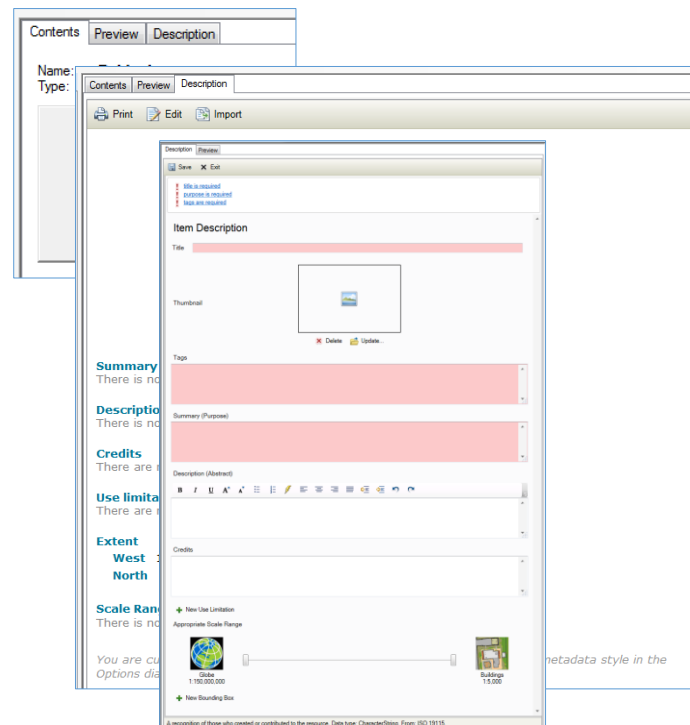
- Изучение Свойств и Метаданных

- Изучение Картографической проекции/
Координатной системы данных

- Изучение геометрии данных
(точка, линия, полигон)

- Изучение атрибутов данных

(названия полей, числовой и текстовый тип, ненулевые поля)



<http://zansae-geonode.org/documents/513/download>

Проверка геометрии и атрибутов при контроле качества

Существуют две категории для проверки:

Геометрия

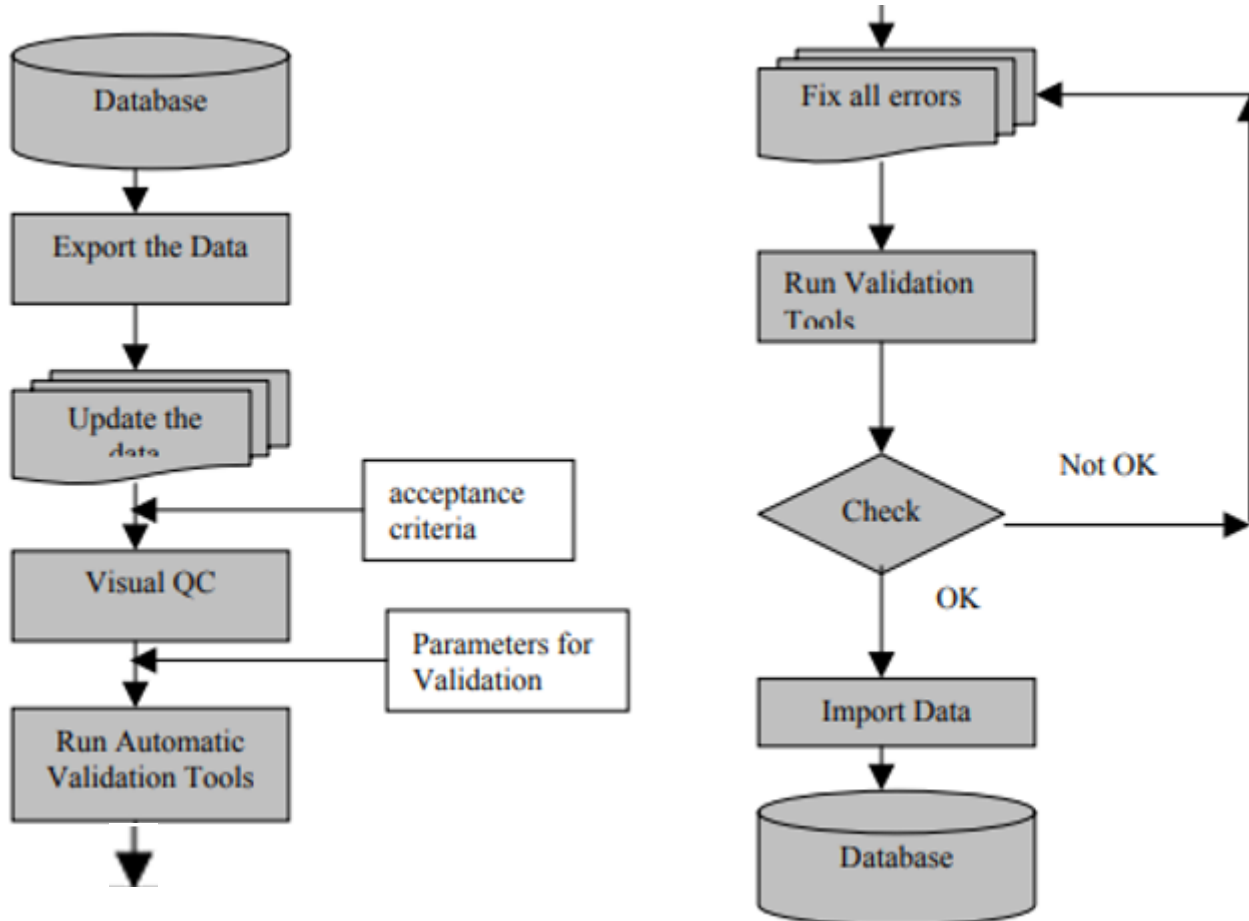
Большая часть коммерческих и Open source GIS-приложений имеют наборы инструментов для проверки ошибок геометрии (например, ArcGIS [Checking and Repairing Geometries](#), QGIS [Check Geometry Validity](#), или GRASS [v.clean](#))

Табличные данные (или Атрибуты)

Для базы геоданных процесс ввода атрибутивных данных должен начинаться с создания набора доменов, что позволит избежать ошибок при последующем вводе. Например, в ArcGIS для таблиц и наборов данных [Attribute Domain](#), в QGIS создается [attribute list](#)

<https://gis.stackexchange.com/questions/57243/automated-quality-control-and-assurance-geodata?rq=1>

Пример организации контроля качества в геоинформационной базе данных



https://www.isprs.org/proceedings/xxxvi/2-c43/postersession/hedge_parameshwa_et_al.pdf

Часть 1: Контроль качества данных с ArcGIS (ESRI)

Определение оценки качества

Quality control

ArcMap 10.3 | [Other versions](#) ▾

- Domains
- Data rendering
- Label expressions

Data quality is regulated by a quality assurance (QA) program through quality control (QC) processes. QA is an all-encompassing management approach that combines technical aspects of quality, qualitative methods, and human resources in a system designed to meet the customer's expectations. QC is an individual task or set of tasks performed at a given level of the production process aimed at ensuring the integrity of a product, output, or action.

QC does not have a prescribed time in the production process other than as a last step before data delivery. However, this does not mean that's the first time QC is performed. QC can be performed in process and at any time during production. Some QC is inherent in the database and occurs on the fly, such as attribution with domains, while other processes must be applied such as minimum feature detection and visual data validation. Early detection and feedback are essential in successfully producing high-quality, cost-effective data.

<http://desktop.arcgis.com/en/arcmap/10.3/guide-books/extensions/maritime-charting/quality-control.htm>



QC Quality Control— Processes or tools to identify errors that are already in the data.

- Examples:
 - GP tools (Select by Attribute, Select by Location), GP Model, Data Reviewer Batch Jobs

GP - Геообработка


http://proceedings.esri.com/library/userconf/proc18/tech-workshops/tw_1846-39.pdf

- Полнота и подробность
- Правильность и достоверность (варианты перевода validity)
- Логическая последовательность
- Физическая последовательность
- Относительная целостность
- Позиционная точность

**Комплексный характер ГИС-данных
(Rules and Standards for Spatial Data Quality
in GIS environments, 2003)**

- Necessity of Analysis. The client will have certain ideas for the development of data. However, in most cases, he may not be fully aware of what is the best that can be done. It is the job of the GIS analyst to assess carefully the requirements of the user and suggest what is best suited for his needs. Such an analysis will help in evolving the needed specifications.
- Cost. With any type of information system, carefully planning prior to the acquisition or the creation of data generally increases the degree of success and efficacy. On the contrary, lack in planning increases the possibility of a mismatch between the system's capabilities and the user needs, becoming just a waste of money. The cost of data development is quite high and the customer should be advised and shown the method through which he can achieve his aim within his estimate.
- Accuracy. Another major issue for the creation of data relates to accuracy, completeness and timeliness. Digitising information to be included in a GIS is not simple and straightforward. As information is included or excluded to fit with the application, the accuracy and completeness of that information can be compromised. In addition the conversion of existing records, without verifying the accuracy of the information with the data subject, can mean that the quality of the newly digitised information is poor as it is out-of-date or incomplete.
- Scaleability Issues. It is important to develop data in order to maintain future expansion always possible. If this is not considered during the creation of data then different kinds of spatial data pertaining to the same area cannot be attached with each other because of the difference in the projection systems and the attribute information.
- Lack of quality and availability of base data. Data quality is essential when a GIS is used to make decisions that, potentially, could adversely impact the data subject. Without an accurate information, any potential operational efficiency or benefit, may be compromised. Many efforts in the creation of data get stuck either because the base data is unavailable or because it is very poor in quality.
- Permissions from authorities. This policy on data development needs to be reviewed.
- Standards/Formats. A standard format needs to be developed for spatial data because to convert data from different formats causes the loss of data and their quality gets reduced.
- Symbology. Standard sets of symbols should be developed in order to be used with different kind of applications which use GIS.

https://www.isprs.org/proceedings/xxxvi/2-c43/postersession/hedge_parameshwa_et_al.pdf



- **Daily or when editing**
 - Step in a workflow
 - Current extent
- **Periodically (weekly or monthly)**
 - Scheduled, automated
 - Full database
- **New or Updated Dataset**
 - When received
 - Full dataset



ultimate Geospatial Data Validation CHECKLIST

safe.com/qa

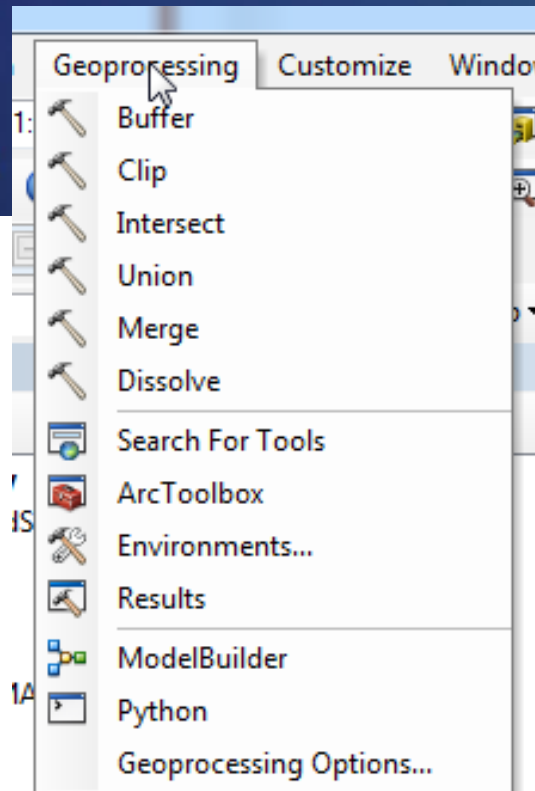
- Check the schema**
 - ☑ Feature type (i.e. layer, level, table ...) names
 - ☑ Attribute names and types
 - ☑ Coordinate system
 - ☑ Allowed geometries
- Check the data values**
 - ☑ Correct data type for the field
 - ☑ Within the valid range/domain
 - ☑ Check for duplicates (e.g. of a unique key)
 - ☑ Are null values allowed? If so, are the null types consistent (NaN, infinity, empty strings, etc.)?
- Validate the geometry**
 - ☑ Self-intersections
 - ☑ Degenerate or corrupt geometries
 - ☑ Null geometries
 - ☑ Vertices with missing normals
 - ☑ Texture coordinates, for geometry with a texture
 - ☑ Invalid solid boundaries
 - ☑ Invalid solid voids
 - ☑ Non-planar surfaces
 - ☑ Duplicate consecutive points, in 2D or 3D
- Compliance to standards**
 - ☑ OGC
 - ☑ INSPIRE
 - ☑ Other international standards or trade standards
 - ☑ Your company's standards
- Format-specific QA/QC**
 - ☑ CAD data: ensure the robust extraction of layers, geometry, text, line types, blocks, extended entity data
 - ☑ XML / JSON: validate the syntax or schema
 - ☑ Tabular data: ensure values pass logical tests; check integration with spatial details
 - ☑ Databases: check the data and geometry before attempting to load it into a central repository
 - ☑ Point clouds: check for correct components and values
- Workflow-based validation**
 - ☑ Detect differences in an updated version of the same data
 - ☑ Validate submitted data and immediately give feedback to stop bad data from being processed
 - ☑ Other requirements for your workflow
- Repairing and reporting bad data**
 - ☑ Map the schema to fit the destination data model
 - ☑ Geometry manipulation
 - ☑ Enforce compliance with your standards
 - ☑ Flag the bad data and return it for human analysis
 - ☑ Measure and describe the quality in a standardized way
 - ☑ Send the report

<http://cdn.safe.com/wp-content/uploads/2014/11/Data-QA-Checklist.png>

Способы конфигурирования (настройки) проведения контроля качества

- **Automated Data Validation**

- Geoprocessing Tools & Models (Toolbox)
- Python Scripts
- ArcGIS Data Reviewer



- **Manual Data Validation**

- ArcMap Document
- ArcGIS Data Reviewer
- Checklist

ArcMap/ArcCatalog

http://proceedings.esri.com/library/userconf/proc18/tech-workshops/tw_1846-39.pdf

Инструменты контроля качества данных в ArcGIS ArcMap

- Топология и ее правила

Топологические правила определяют возможные пространственные отношения между объектами

Geodatabase topology rules and topology error fixes

ArcMap 10.3 | Other versions ▾

▲ Available with Standard or Advanced license.

- Polygon rules
- Line rules
- Point rules

🔗 License: You can create simple, temporary topological relationships between features in ArcGIS for Desktop Basic. Creating or editing geodatabase topology requires an ArcGIS for Desktop Standard or ArcGIS for Desktop Advanced license.

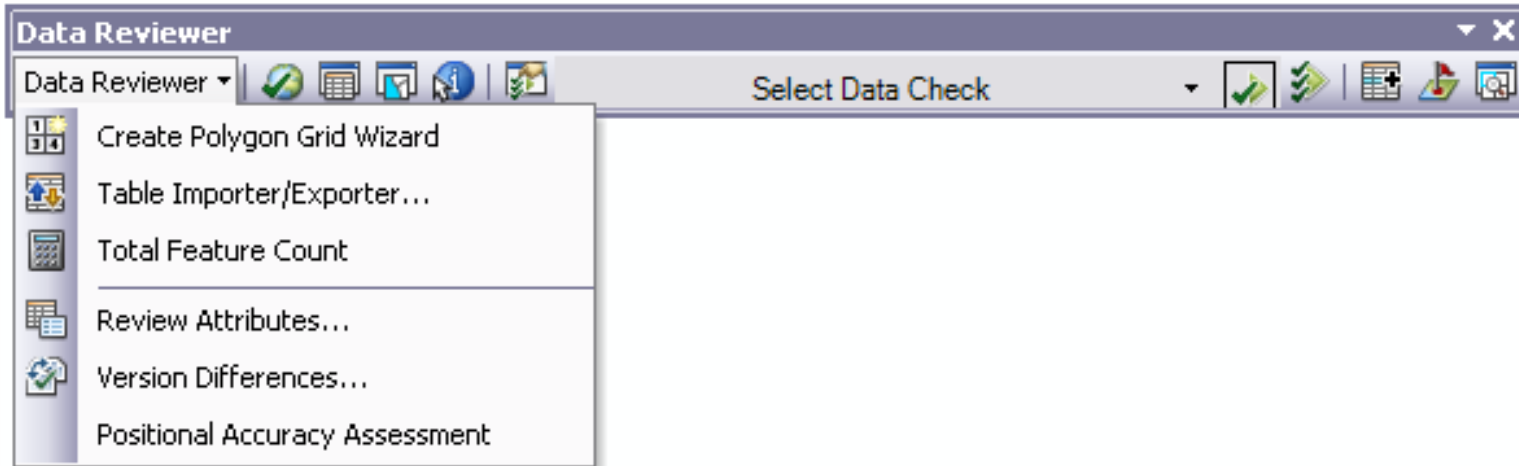


<http://desktop.arcgis.com/en/arcmap/10.3/manage-data/editing-topology/geodatabase-topology-rules-and-topology-error-fixes.htm#GUID-50241B19-F084-474C-A861-79AC69441D58>

- ArcGIS Data Reviewer

<https://gisdigest.wordpress.com/2015/05/02/arcgis-data-reviewer/>

Arcgis Data Reviewer



Data Reviewer toolbox

The Data Reviewer geoprocessing toolbox contains tools that allow you to create and delete a Reviewer session, execute a batch job, and compare the schemas of two geodatabases. You can also write the output from another geoprocessing tool to the Reviewer table when it is a feature class, table, or layer.

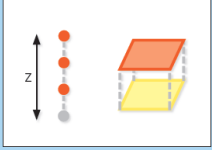
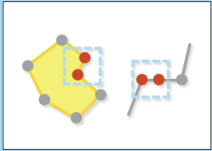
- Data Reviewer Tools.tbx
 - Create Reviewer Session
 - Delete Reviewer Session
 - Execute Reviewer Batch Job
 - Schema Compare
 - Write to Reviewer Table

<https://desktop.arcgis.com/en/arcmap/latest/extensions/data-reviewer/a-quick-tour-of-data-reviewer.htm>
<https://desktop.arcgis.com/en/arcmap/latest/extensions/data-reviewer/essential-data-reviewer-vocabulary.htm>

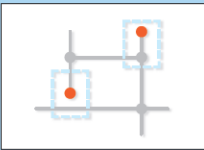
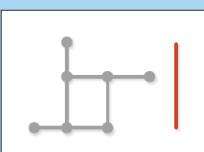

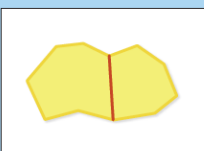
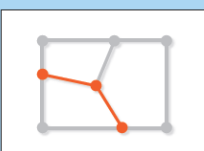
Инструменты проверки в Arcgis Data Reviewer

- Проверка базы данных (validation)
- Проверка таблиц
- Оценка пространственного параметра
- Проверка по умолчанию
- Топология
- Дополнительные проверки
- Проверки полигонов и полилиний
- Оценка значений Z
- Проверка объекта в объекте
- Дублирование геометрии

Примеры: Дубликаты

Duplicate Geometry	Finds features of the same geometry type that are collocated	
Duplicate Vertex	Searches for vertices in selected polyline or polygon feature classes that are within a specified tolerance of each other	

Примеры: Проверки топологии

Find Dangles	Within a database topology, finds polyline features that have nodes that are within a tolerance but not connected to other features in the database topology Comparison to Topology: Line must not have dangles	
Orphan	Finds single polyline features that are not connected in the database topology	
Unnecessary Nodes	Finds features that share a node and have identical attributes in editable fields Comparison to Topology: Line must not have pseudo nodes	
Unnecessary Polygon Boundaries	Finds adjacent polygon features that share a boundary and have identical attributes in editable fields	
Topology Rules	Returns the geometry of features that violate the topology rules that have been defined for a feature dataset in the geodatabase	

<https://www.esri.com/~media/Files/Pdfs/library/fliers/pdfs/arcgis-data-reviewer-checks.pdf>

Проектирование Reviewer workspace в геобазе данных

Introduction to storing the Reviewer workspace in a geodatabase in PostgreSQL

ArcMap 10.7 | [Other versions](#) ▾

⚠ Available with Data Reviewer license.

■ [PostgreSQL installation](#)

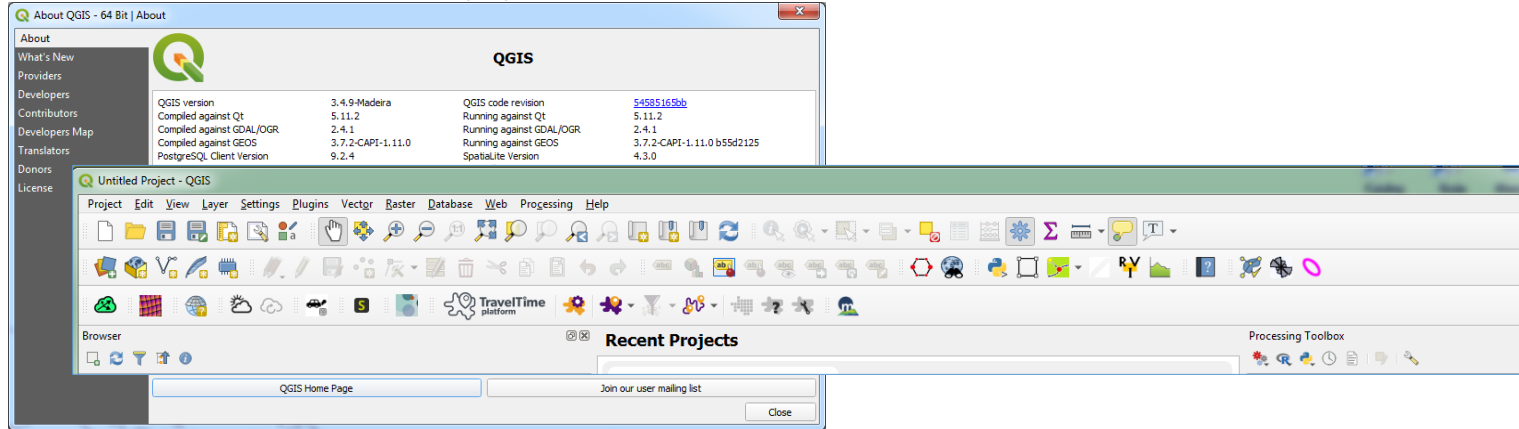
ArcGIS Data Reviewer for Desktop is a data quality-control management application. It requires that you have a workspace defined when using a Reviewer session to store the results found during your data review process. It creates a new feature dataset and tables in the geodatabase that you have identified as your Reviewer workspace. The feature dataset includes point, line, and polygon feature classes, which contain error geometries for the features that have been written to the Reviewer table. When stored in a enterprise geodatabase, the Reviewer workspace supports versioning.

This guide book is intended for database administrators as recommendations for establishing the workspace in a geodatabase in PostgreSQL. The enterprise geodatabase uses ArcGIS technology as the gateway between geographic information system (GIS) clients and PostgreSQL.

<https://desktop.arcgis.com/en/arcmap/latest/extensions/data-reviewer-guide/admin-dr-sql-server/introduction-to-storing-the-data-reviewer-workspace-in-an-enterprise-geodatabase-in-sql-server.htm>

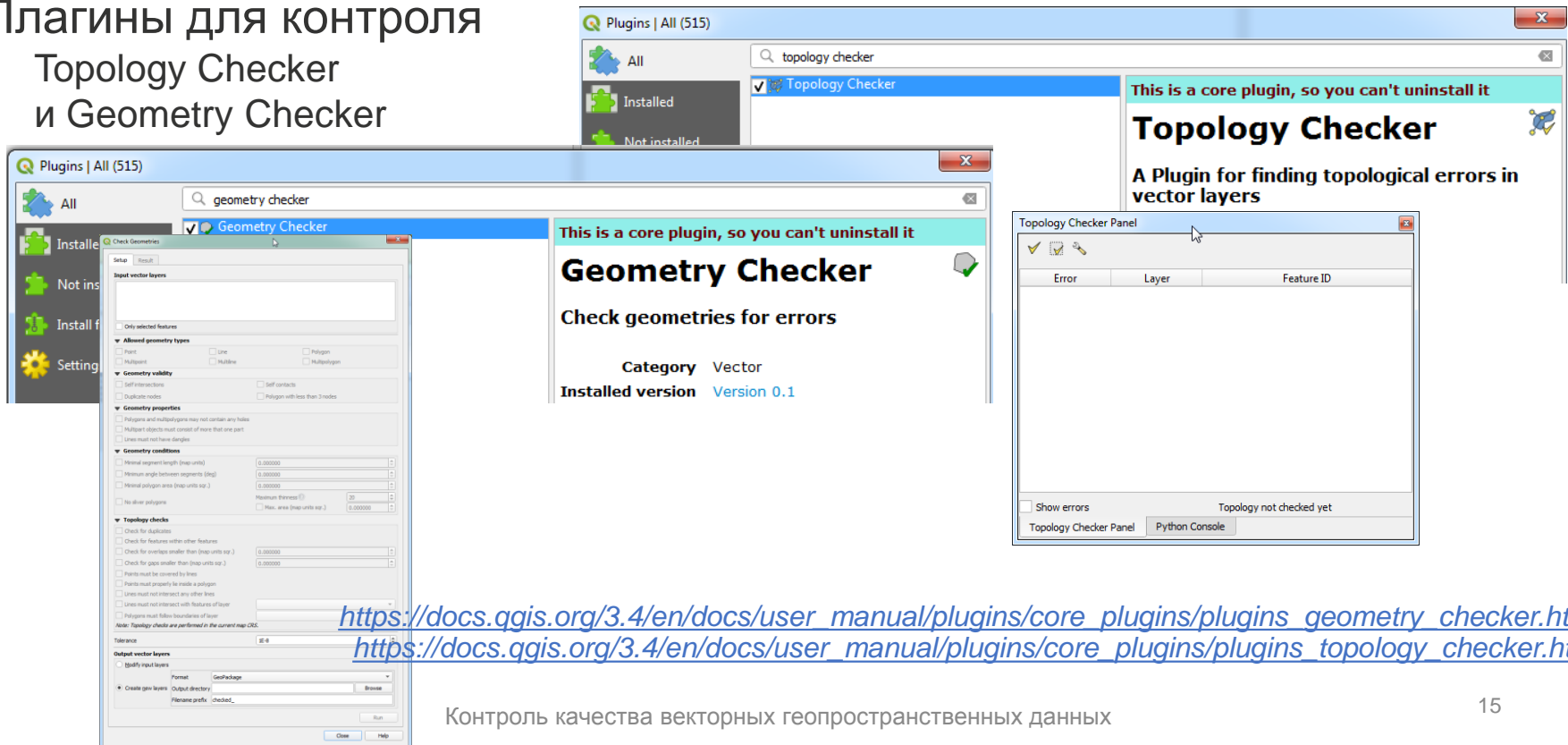
Часть 2: Контроль качества данных в QGIS (Open source)

Контроль качества данных в QGIS

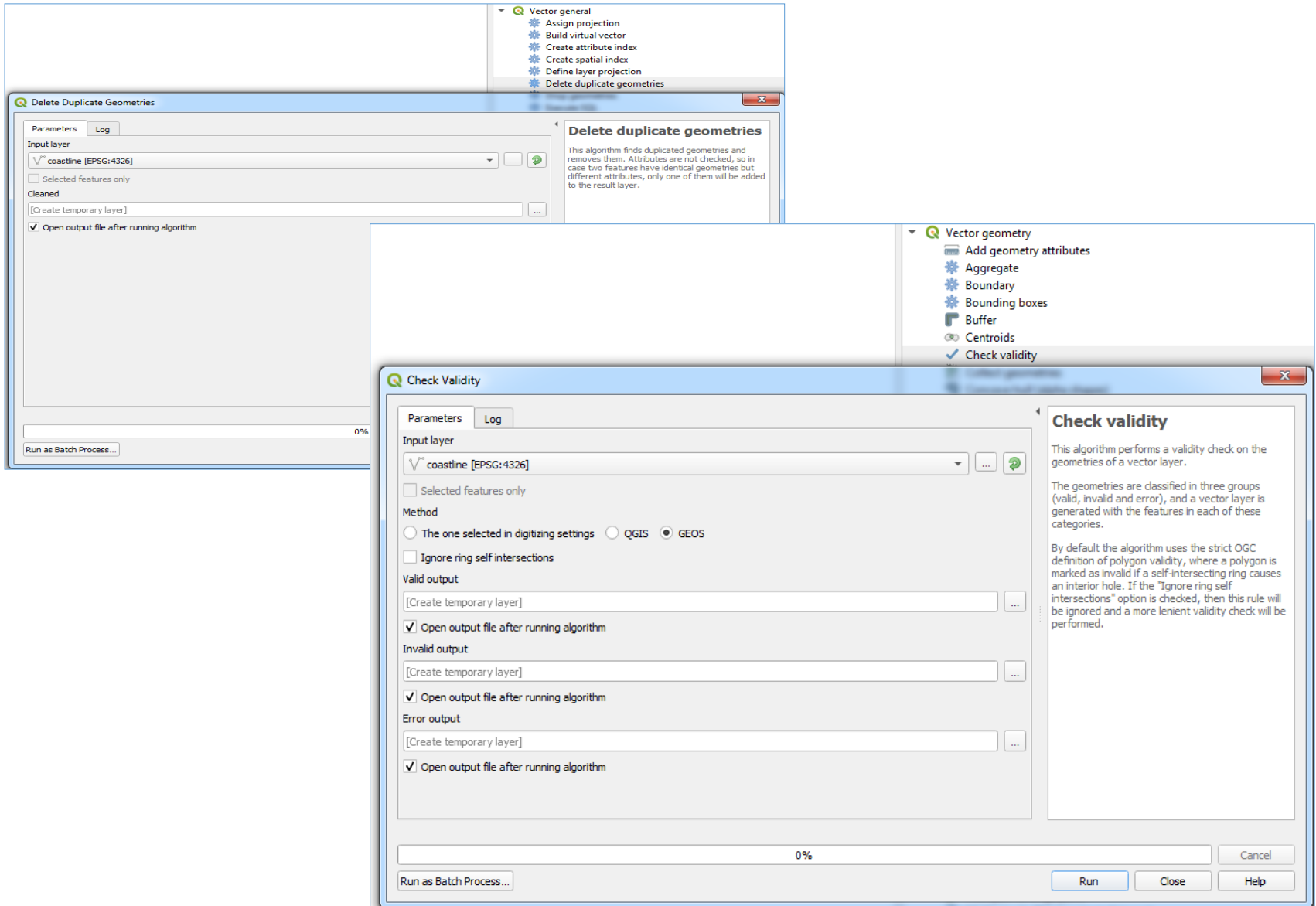


Плагины для контроля

- Topology Checker
- и Geometry Checker



Инструменты QGIS - 1/3



Инструменты QGIS - 2/3

The image displays the QGIS interface, focusing on the Vector Geometry tools menu and two specific tool dialog boxes.

Vector Geometry Tools Menu:

- Vector geometry
 - Add geometry attributes
 - Aggregate
 - Boundary
 - Bounding boxes
 - Buffer
 - Centroids
 - Check validity
 - Collect geometries
 - Concave hull (alpha shapes)
 - Concave hull (k-nearest neighbor)
 - Convert geometry type
 - Convex hull
 - Create layer from extent
 - Create wedge buffers
 - Delaunay triangulation
 - Delete holes

Delete Holes Dialog:

- Parameters** | Log
- Input layer:** [Empty dropdown]
- Selected features only
- Remove holes with area less than:** 0.000000
- Cleaned:** [Create temporary layer]
- Open output file after running algorithm
- Delete holes**

This algorithm takes a polygon layer and removes holes in polygons. It creates a new vector layer in which polygons with holes have been replaced by polygons with only their external ring. Attributes are not modified.

An optional minimum area parameter allows removing only holes which are smaller than a specified area threshold. Leaving this parameter as 0.0 results in all holes being removed.

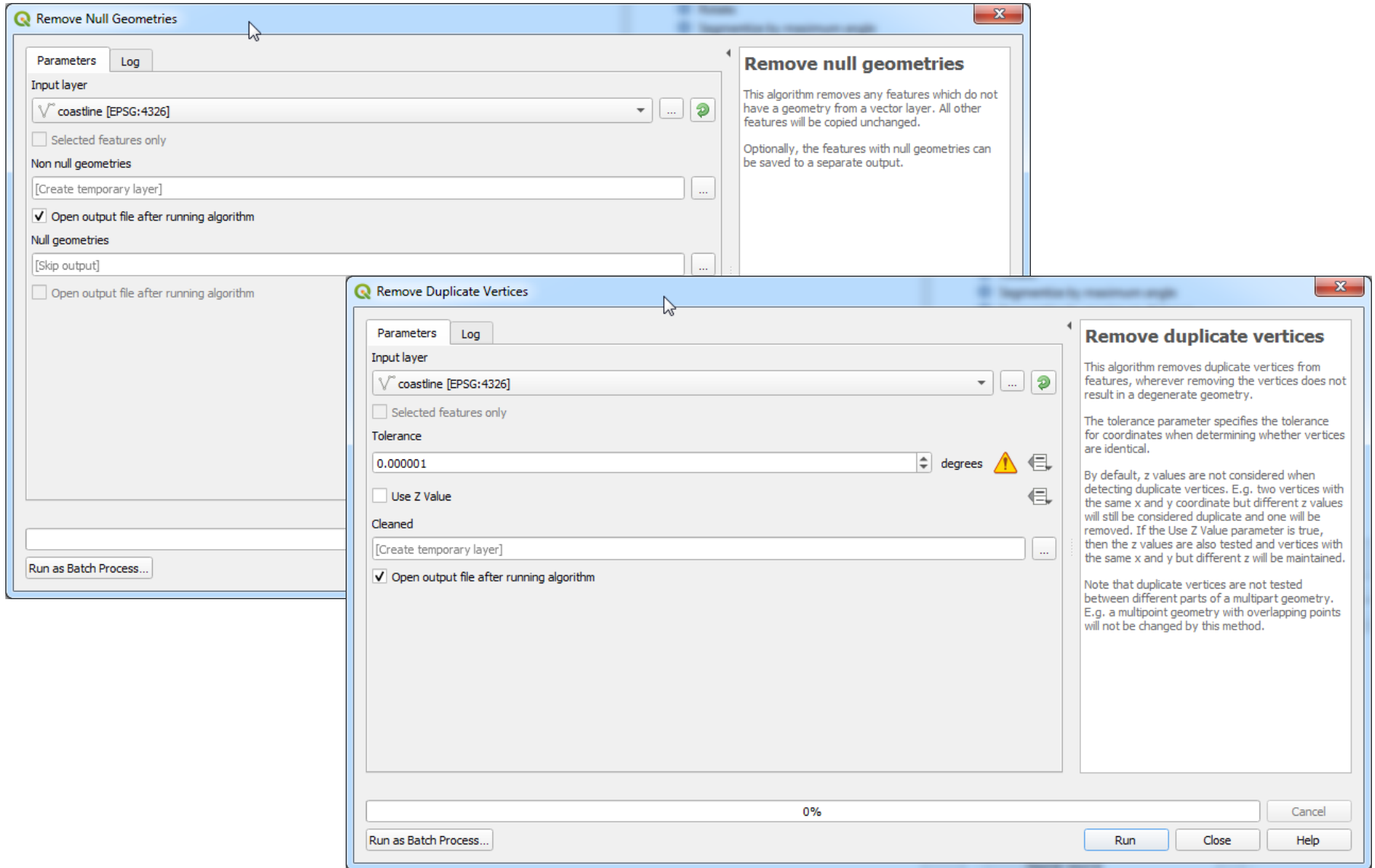
Fix Geometries Dialog:

- Parameters** | Log
- Input layer:** coastline [EPSG:4326]
- Selected features only
- Fixed geometries:** [Create temporary layer]
- Open output file after running algorithm
- Fix geometries**

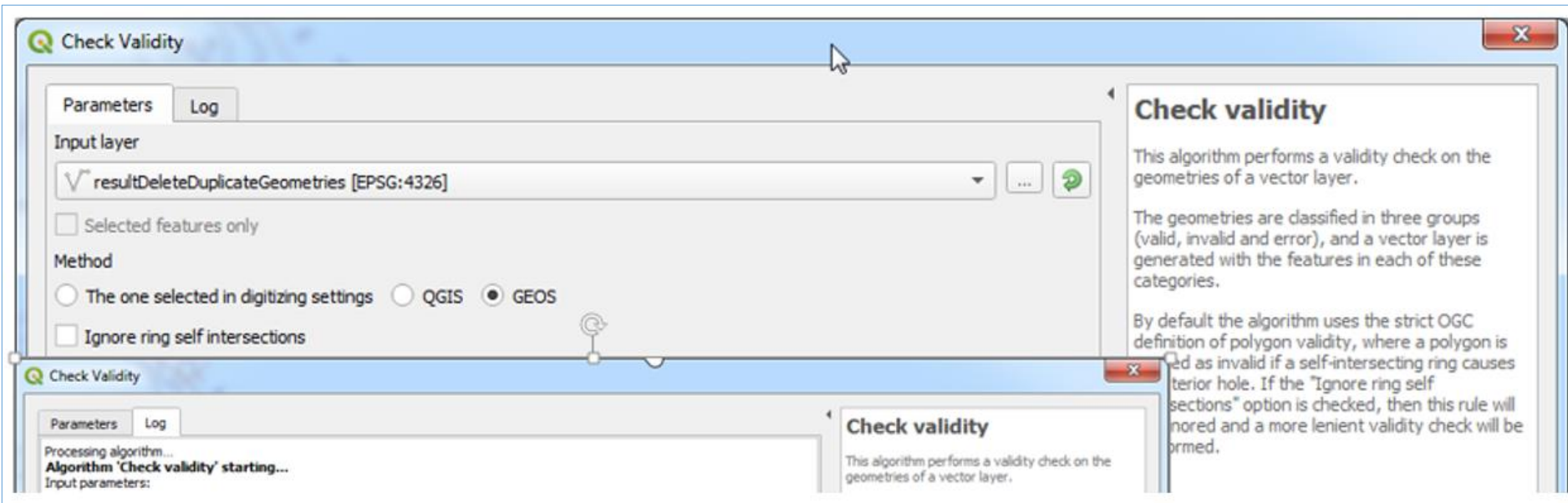
This algorithm attempts to create a valid representation of a given invalid geometry without losing any of the input vertices. Already-valid geometries are returned without further intervention. Always outputs multi-geometry layer.

NOTE: M values will be dropped from the output.

Инструменты QGIS - 3/3



Пример: Проверка на геометрические ошибки



Name	Type	Size
CheckValidity_ErrorOutput.gpkg	GPKG File	96 KB
CheckValidity_InvalidOutput.gpkg	GPKG File	96 KB
CheckValidity_ValidOutput.gpkg	GPKG File	112 KB

Ссылки

<https://www.gislounge.com/gis-data-a-look-at-accuracy-precision-and-types-of-errors/>

<https://esriaustraliatechblog.wordpress.com/2014/09/22/data-quality-checks-and-arcgis-10-x/>

<https://gisdigest.wordpress.com/2015/05/23/topology-checker-vs-geometry-checker/>

https://en.wikipedia.org/wiki/Free_and_open-source_software

<https://desktop.arcgis.com/en/arcmap/latest/extensions/data-reviewer/data-reviewer-custom-steps-and-tokens.htm>