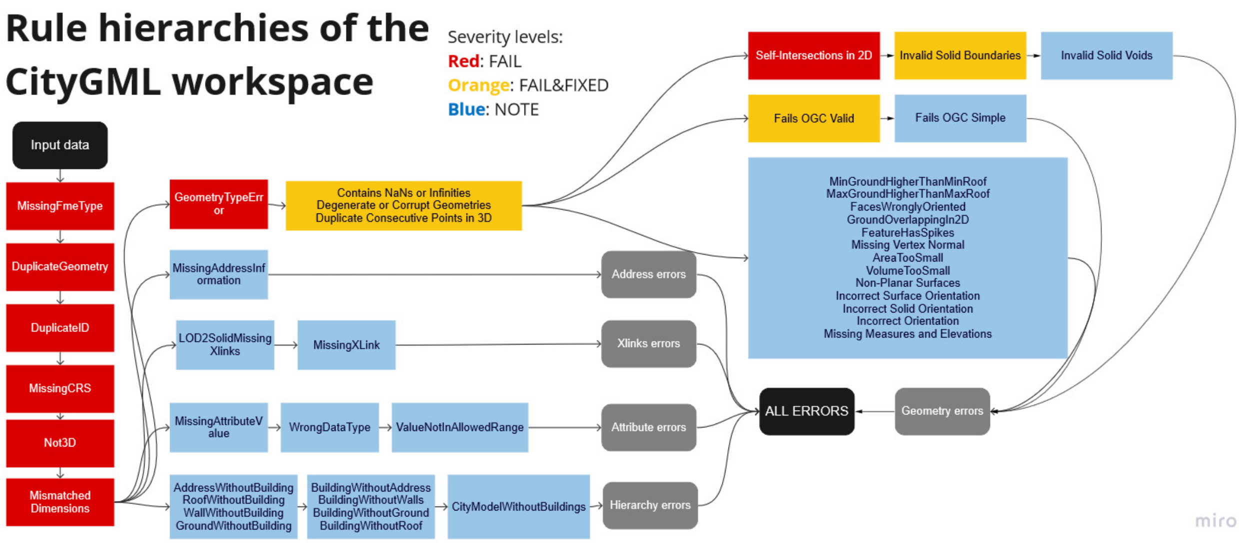

Figure 2. The rule hierarchy and severity levels of the tool.

Users can choose whether they want results in tabular or geometry format. The latter enables visualization of the results.

The tool contains over 40 quality rules, and some of them consist of several sub-criteria. Most quality rules are based on the pre-defined transformers of the FME, like GeometryValidator. Therefore their functionalities could not be edited, since the FME is not an open source software. The rest of the rules were constructed from smaller pieces of transformer, or their algorithms were built completely from the scratch.

The quality rules are classified on the basis of their outcomes. Totally, there are three different options for each rule. If the feature or part of it does not pass the rule, it can either be discarded, repaired or noted. Because of that, the rules have an elaborated hierarchy (see Figure 2), which prevents errors from recurring under the other rules. For example, fixing geometry error in building boundary may produce degenerated faces, which in turn prevents checking for building voids (Safe Software, 2022).

The quality rules and their hierarchy vary between the CityGML and CityJSON formats due to different encoding. Both formats have unique hierarchical structure and different level of complexity (Ledoux et al., 2019). The CityGML contains more mandatory parts related to geometry and attributes, more complex geometries and hierarchies, as well as extended links (xlink:href) to connect building elements (Open Geospatial Consortium, 2021; Ledoux et al., 2019; Hugo, Ledoux and Balázs, Dukai, 2021).

3 Results

The tool was tested by using CityGML 3D building data from Ranua, which is the city in Northern Finland. The area contained 30 640 features, 7660 of them were LoD2 buildings, and rest of them were building parts, like RoofSurfaces and WallSurfaces. The complete data set is available from https://tiedostopalvelu.maanmittauslaitos.fi/tp/kartta?lang=en.

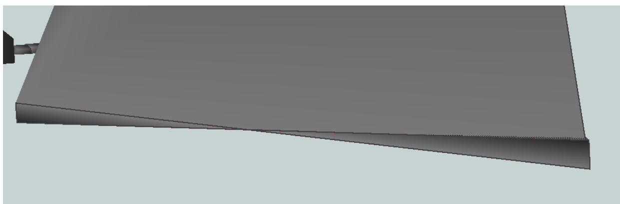

Figure 3. Self-intersection error in the example data set.

By using the default parameters of the tool, totally 15 rules were violated at least once. The most remarkable deficiencies were missing xlinks and addresses, which were reported for every building (total 7660 errors). They are not severe errors alone as they are not required according to the standards (see Open Geospatial Consortium (2021)) but they may cause problems in data integration processes due to requirement of harmonized and interoperable data (Mohammadi et al., 2010). In addition, the tool found totally 424 spikes, 116 self-intersections (see Figure 3), 108 duplicate concecutive points, 90 unclosed solid surfaces, 65 invalid solid orientations and 16 multiple connected components. Four errors had less than ten occurrences, like overlapping grounds or non-planar surfaces. The FME's GeometryValidator transformer (see Safe Software (2022) is able to fix most of the errors automatically.

When thinking about SDIs, the most interesting part is readiness for data integration. The good news is that the data did not contain any duplicate identifiers or geometries. Only a few buildings were overlapping and none of them were degenerated or corrupted. Also, semantics were modelled correctly since every buildings contained their mandatory BoundarySurfaces, like Wall- and Roof-Surfaces.

4 Discussion

In addition to the improved data quality itself, the tool helps to assess current data quality level, which is especially important in the context of SDIs. One of the biggest challenges of SDIs is the data integration, which requires known quality. Integration of multi-source data sets is difficult, if data sets are not harmonized, interoperable or in the same level of quality (Mohammadi et al., 2010; Jakobsson and Giversen, 2007).

For that reason, DQA should be an integral part of data management. Data quality should be controlled in all phases of its life cycle (Beare et al., 2010), because objectives of data depend on these phases (Loshin, 2010). For example, data producers compare data to its conceptual model in order to avoid errors, while end users compare data to its suitability for specific uses (ISO 19157). Only one principle is valid for all phases: errors are easier and cheaper to fix as soon as possible (Di Zio et al., 2016).

The proposed tool is not only suitable for DQA of 3D buildings. With a little modifications, it could be used for assuring other data formats or completely other kinds of rules, like business and life cycle rules. In addition, a combination of the DQA and artificial intelligence methods would result in very interesting possibilities. This could further increase automation, and thus decrease human effort as well as error possibility. In the long run, automation enhances overall efficiency of workloads and produce more uniform quality (Beare et al., 2010).

5 Conclusion

This paper proposed an FME-based tool to assure data quality of 3D buildings, demonstrated its results by using the real CityJSON and CityGML data, and finally considered the tool usage in data integration. Further possibilities were discussed.

The tool, its instructions, and all implemented quality rules can be accessed through the project's GitHub repository: https://github.com/opengeospatial/GEOE3.

References

Beare, M., Henriksson, R., Jakobsson, A., Marttinen, J., Onstein, E., Tsoulos, L., Williams, F., Mäkelä, J., De Meulanear, L., Persson, I., and Kavadas, I.: ESDIN Quality Final Report, https://doi.org/10.13140/RG.2.1.3887.5600, 2010.

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., and Shi, W.: Thirty years of research on spatial data quality: achievements, failures, and opportunities, Transactions in GIS, 14, 387–400, 2010.

Di Zio, M., Fursova, N., Gelsema, T., Gießing, S., Guarnera, U., Petrauskienė, J., Quensel-von Kalben, L., Scanu, M., ten Bosch, K., van der Loo, M., et al.: Methodology for data validation 1.0, Essnet Validat Foundation, 2016.

Hugo, Ledoux and Balázs, Dukai: CityJSON Specifications 1.1.0, Delft University of Technology, Mekelweg 5, 2628 CD Delft, Netherlands, 2021.

ISO 19157: ISO/CD 19157-1:2021(E), Geographic information — Data quality – Part 1: General requirements, Standard, International Organization for Standardization, Geneva, CH, 2021.

Jakobsson, A. and Giversen, J.: Guidelines for implementing the ISO 19100 geographic information quality standards in national mapping and cadastral agencies, Eurogeographics Expert Group on Quality, 2007.

Ledoux, H. and Wagner, D.: OGC® CityGML Quality Interoperability Experiment, 2016.

Ledoux, H., Ohori, K. A., Kumar, K., Dukai, B., Labetski, A., and Vitalis, S.: CityJSON: A compact and easy-to-use encoding of the CityGML data model, Open Geospatial Data, Software and Standards, 4, 1–12, 2019.

Loshin, D.: The practitioner's guide to data quality improvement, Elsevier, 2010.

Mohammadi, H., Rajabifard, A., and Williamson, I. P.: Development of an interoperable tool to facilitate spatial data integration in the context of SDI, International Journal of Geographical Information Science, 24, 487–505, 2010.

Open Geospatial Consortium: CityGML Standard, https://www.ogc.org/standards/citygml, 2021.

Safe Software: FME GeometryValidator Documentation, http://docs.safe.com/fme/html/FME_Desktop_Documentation/FME_Transformers/Transformers/geometryvalidator.htm, accessed: 2022-04-05, 2022.

Stoter, J., Ohori, G. A., Dukai, B., Labetski, A., Kavisha, K., Vitalis, S., and Ledoux, H.: State of the Art in 3D City Modelling: Six Challenges Facing 3D Data as a Platform, GIM International: the worldwide magazine for geomatics, 34, 2020.