python-ags4: A Python library to read, write, and validate AGS4 geodata files

Asitha I. Senanayake [1], **Roger J. Chandler**[2], **Tony Daly**[3], **and Edward Lewis** [4]

1 Norwegian Geotechnical Institute (formerly Fugro USA Marine, Inc.) **2** Bentley Systems **3** Association of Geotechnical and Geoenvironmental Specialists **4** British Geological Survey, Keyworth, UK

Summary

Data gathered from geotechnical, geoenvironmental, and geophysical investigations can be broadly described as "geodata". The AGS4 data format (Association of Geotechnical and Geoenvironmental Specialists, 2011, 2017, 2021b, 2022) is one of the most widely used data transmittal formats for geodata and is used across the world. It is a plain text format consisting of multiple tables of comma-separated values, tied together with a robust data schema and a comprehensive suite of validation rules. The basic structure of AGS4 file is is shown in Figure 1.
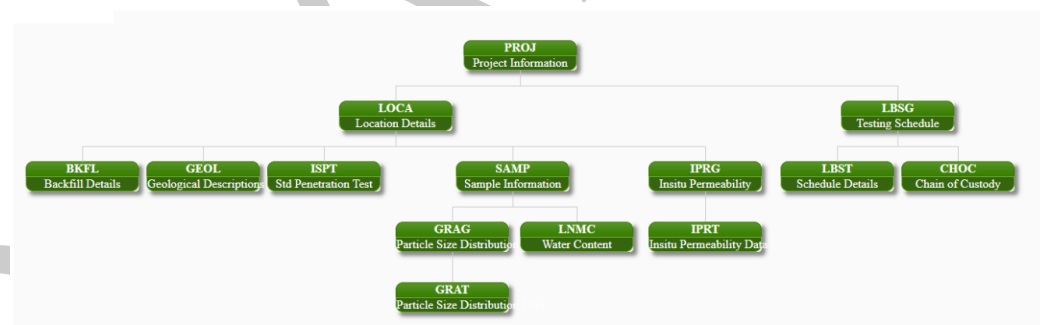
Figure 1: Simplified schematic of AGS4 data structure

Source: Association of Geotechnical and Geoenvironmental Specialists (2022)

python-ags4 is a Python library that provides functionality to read, write, and validate AGS4 geodata files. It provides users with a gateway to access the full power of the Python ecosystem to explore, analyze, and visualize geodata. Pandas DataFrame (The pandas development team, 2020) is the primary data structure used within the library, therefore it can handle relatively large datasets reasonably fast. The data validation module checks the file for compliance with the validation rules and provides a detailed error report. An example error report is shown in Figure 2.

```
Running python_ags4 v0.4.0
Opening file... test_data.ags

  Checking lines...
  Loading tables...
  Checking headings and groups...
  Checking file schema...

File Name:      test_data.ags
File Size:      4 kB
Checker:        python_ags4 v0.4.0
Dictionary:     Standard_dictionary_v4_1_1.ags
Time (UTC):     2022-07-25 00:51:05

AGS Format Rule 1:
  Line 12              Has Non-ASCII character(s).

AGS Format Rule 5:
  Line 46              Contains only spaces.

AGS Format Rule 3:
  Line 63              Does not start with a valid data descriptor.

AGS Format Rule 8:
  Line 85      LLPL    Value 55.1 in LLPL_LL not of data type 2SF. (Expected: 55)
  Line 86      LLPL    Value 155.1 in LLPL_LL not of data type 2SF. (Expected: 160)
  Line 85      LLPL    Value 34.8 in LLPL_PI not of data type 2SF. (Expected: 35)
  Line 86      LLPL    Value 134.8 in LLPL_PI not of data type 2SF. (Expected: 130)
  Line 92      TEST    Value 1.01 in TEST_COL not of data type 1SF. (Expected: 1)

AGS Format Rule 10a:
  Line 92      TEST    Duplicate key field combination: DATA|Location_1
  Line 93      TEST    Duplicate key field combination: DATA|Location_1

AGS Format Rule 10c:
  Line 78      SAMP    Parent entry for line not found in LOCA: BH-AA43
  Line 79      SAMP    Parent entry for line not found in LOCA: BH-AA43
  Line 85      LLPL    Parent entry for line not found in SAMP: Location_1|1.00|1a|Bag|
  Line 86      LLPL    Parent entry for line not found in SAMP: Location_5|2.00|2a|Bag|
  Line -       TEST    Parent group left blank in dictionary.

AGS Format Rule 16:
  Line -       ABBR    ABBR table not found.


File check complete! 16 errors found!
```

Figure 2: Example error report

Statement of Need

This library fulfills the following needs of the engineering and scientific community that uses AGS4 geodata files:

- Provide a transparent and easily accessible tool to validate AGS4 geodata files.
- Provide access to the Python ecosystem to users of AGS4 geodata.
- Provide a cross-platform tool to work with AGS4 geodata files.

Having a free and open-source tool that can validate AGS4 files will help producers and receivers of data check that the geodata file is of the required standard. It also lowers the barrier to entry for those who are not familiar with this data format and fosters data sharing and collaboration. python-ags4 has already been adopted by the British Geological Survey to work with its large repository of publicly available AGS4 geodata files.

Python has a rich ecosystem of packages that can be utilized to analyze and visualize data. The ability to easily import geodata from AGS4 files will greatly enhance the ability of engineers and scientists to interpret data and improve knowledge of ground conditions. A Jupyter Notebook with examples of how to import data to Pandas and GeoPandas (Jordahl et al., 2020) for statistical and spatial analyses and visualize using Matplotlib (Hunter, 2007) is provided in the git repo of this library. We provide an example for users in the git repo.

The command-line interface included with the library is a convenient and easy to use tool to work with AGS4 geodata files. It is cross-platform and has been tested in Linux, Windows, and Mac environments.

The current state of practice for validation of AGS4 files is to use one of two proprietary software, namely KeyAGS Checker which is freeware and gINT AGS Checker which is part of

the widely used `gINT` geotechnical data management suite. They are only available for the Windows operating system and can provide slightly different results in some cases. Therefore, validators based on `python-ags4` are expected to become the standard in the future and provide a single source of truth for the validity of AGS4 files (Association of Geotechnical and Geoenvironmental Specialists, 2021a; Chandler, 2021).

The open-source geotechnical software package groundhog also provides functionality to import AGS4 data to Pandas DataFrames, therefore it can be used as an alternative to `python-ags4` in this regard. However, the main focus of that package is general purpose geotechnical engineering calculations so it does not provide functionality to validate and export AGS4 files.

Ongoing Projects

`python-ags4` has been implemented in two free and open-source software projects, making the library more accessible to non technical users.

- Windows application offering the functionality of the library
- API and GUI developed by the British Geological Survey, offering the functionality of the library on the web.

It has also been used in commercial projects. (e.g. https://digitalgeotechnical.com/apps-home/)

Acknowledgements

The authors acknowledge the work done by the AGS Data Format Working Group to develop the AGS4 data format and the support it has extended to develop this library.

References

Association of Geotechnical and Geoenvironmental Specialists. (2011). *Electronic Transfer of Geotechnical and Geoenvironmental Data*.

Association of Geotechnical and Geoenvironmental Specialists. (2017). *Electronic Transfer of Geotechnical and Geoenvironmental Data*.

Association of Geotechnical and Geoenvironmental Specialists. (2021a). *AGS validator (beta)*. https://www.ags.org.uk/data-format/ags-validator

Association of Geotechnical and Geoenvironmental Specialists. (2021b). *Electronic Transfer of Geotechnical and Geoenvironmental Data*.

Association of Geotechnical and Geoenvironmental Specialists. (2022). *Electronic Transfer of Geotechnical and Geoenvironmental Data*.

Chandler, R. (2021). *AGS checkers, diggs and the future world of data transfer*. https://communities.bentley.com/products/geotechnical1/b/geoblog/posts/ags-checkers-diggs-and-the-future-world-of-data-transfer

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Jordahl, K., Bossche, J. V. den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020). *Geopandas/geopandas: v0.8.1* (Version v0.8.1) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.3946761

87 The pandas development team. (2020). *Pandas-dev/pandas: pandas* (latest) [Computer
88   software]. Zenodo. https://doi.org/10.5281/zenodo.3509134