



# DGGSTools: An Open Source Python Package for the Manipulation of Vector and Raster Datasets in the rHEALPix Discrete Global Grid System

Sergio Martin-Segura <sup>1</sup>, Rubén Béjar <sup>1</sup>, and F. Javier Zarazaga-Soria <sup>1</sup>

<sup>1</sup>Aragon Institute for Engineering Research, Universidad Zaragoza, Zaragoza, Spain

Correspondence: Sergio Martin-Segura ([segura@unizar.es](mailto:segura@unizar.es))

**Abstract.** DGGSTools is a Python package and command-line application to process geospatial datasets within the spatial reference provided by the rHEALPix Discrete Global Grid System (DGGS). Its main objective is to provide the tools necessary to transform geospatial datasets to equivalent ones following the rHEALPix DGGS, and to allow its users to work with both raster and vector versions of those datasets as preferred. Once these equivalent datasets have been produced with DGGSTools, the users can obtain the benefits of working with a DGGS, intended to facilitate the efficient integration and retrieval of heterogeneous global data, while using their preferred GIS tools and workflows for the analysis of those datasets.

This short paper describes the main characteristics of the first public version of DGGSTools. Although a small part of its code had already been used in some geospatial research experiments, this is the first version which is publicly available. It includes a good number of automatic tests, it is properly packaged as a Python wheel to facilitate its installation, and offers a command-line application so the package can be used without writing a single line of code.

**Keywords.** discrete global grid system, rheapix, raster, vector

## 1 Introduction

Discrete Global Grid Systems (DGGS) are spatial reference frameworks that associate information with multi-resolution grids of uniquely identified cells; they are proposed as mechanisms to facilitate the efficient integration of heterogeneous spatial data (Purss, 2017). They integrate map projections, geocoding systems, i.e., mechanisms to construct unique identifiers for each cells, and the previously mentioned multi-resolution grids.

DGGSs provide a hierarchical tessellation of the Earth's surface, where each tile at a certain level represents a smaller area than its parent tile at a coarser level (Sahr et al., 2003). There are multiple strategies for tiling the Earth. These have led to various systems proposed by both academia and industry, such as Uber H3, DGGrid, Google S2, OpenEaggr, and rHEALPix (Kmoch et al., 2022).

The rearranged Hierarchical Equal Area isoLatitude Pixelization (rHEALPix) is a cubic geodesic DGGS, based on quadrangular cells that are squares once they are projected Gibb (2016). Quadrangular cells have some practical advantages over hexagons or triangles. They simplify the re-projection, rewarping, and resampling of raster datasets, which are normally based in square pixels too, and, unlike other shapes like hexagons, all the area covered by a quadrangular cell will also be covered by its parent cell in the immediately higher level of the multi-resolution grid. They also have some disadvantages; for example, hexagonal cells are a better fit for dynamical systems where inter-cell distances are important, because a tessellation based on hexagons guarantees that the centroids of all the neighbouring cells are at the same distance (Bowater and Stefanakis, 2018).

Datasets in the rHEALPix system can be stored and manipulated in existing EO Data Cubes as shown in (Béjar et al., 2023), as long as this is done with care. They can also be stored as vector data, as for example in the OGC Testbed-16 (Gibb et al., 2021), where they stored each tile as a record identified by the DGGS cell identifier.

There are various libraries that support the rHEALPix projection, such as PROJ (PROJ contributors, 2022), and other libraries that can manipulate datasets in that projection such as GDAL (GDAL/OGR contributors, 2024) or, only for raster datasets, rasterio (Gillies et al., 2013–). Besides this, the rHEALPix DGGS library (Gibb et al., 2013–) supports the rHEALPix DGGS, although it does not provide any capability to work with spatial datasets. In summary, each of these libraries focuses on a specific area and

none of them meets all the requirements of someone that wants to work with rHEALPix datasets.

This paper proposes a Python package, that can be used both as a library and a command-line application, DGGSTools, which allows to manipulate datasets in rHEALPix while keeping them compatible with existing GIS programs, libraries and formats. Section 2 describes this package, discusses its necessity and shows its main intended uses. Section 3 provides the information required to use it. Finally, Section 4 summarizes the main results and details the next steps for the project.

## 2 DGGSTools

DGGSTools (IAAA Laboratory, 2024) is a command line application and Python package which is focused on processing vector and raster data in the rHEALPix Discrete Global Grid System, although supporting other DGGs is considered as future work.

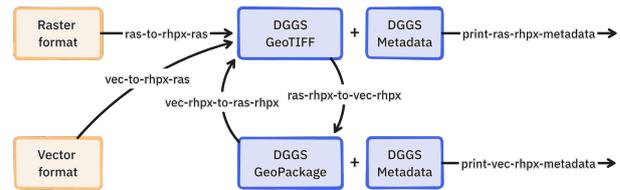
The rHEALPix projection is implemented in the PROJ software library (PROJ contributors, 2022). As any other DGGs, the rHEALPix system includes this projection, but also a multi-resolution hierarchy of cells and a method for providing a unique identifier for each of those cells. The rHEALPixDGGs library (Gibb et al., 2013–) provides an implementation of this DGGs including, for instance, the mapping between cells and pairs of coordinates in the rHEALPix projection.

As described in (Béjar et al., 2023), geospatial data following the rHEALPix DGGs can be stored and processed in both vector and raster formats depending on our objectives, or depending on the geospatial processing toolchain that we intend to use. However there is not any software, as far as we know, that can easily transform geospatial data in other coordinate reference systems or map projections to the rHEALPix projection, or vice versa, while, at the same time, fulfilling the constraints and requirements imposed by the rHEALPix grid. And much less if we want to work equally with either vector or raster data, and to be able to transform between those two models without losing information. DGGSTools exists to provide exactly that functionality.

### 2.1 Main functionalities

The main functionalities provided by the first public version of this tool are explained in the following points. Figure 1 illustrates how these tools are related:

- F1. Support for the GeoTIFF, for raster, and the GeoPackage, for vector, formats. Other formats are supported for the input datasets, but the produced datasets are in one of those two formats.
- F2. Transforming a raster dataset, in a common GIS format and coordinate reference system, to a GeoTIFF



**Figure 1.** A diagram displaying the transformations the tool can perform

in the rHEALPix DGGs. This includes warping to the rHEALPix projection and resampling the input raster dataset so the result is in a given rHEALPix DGGs resolution, i.e., each pixel corresponds exactly with an rHEALPix cell. It also adds the necessary metadata to the GeoTIFF file so the parameters of the rHEALPix grid, and not only the rHEALPix projection which is of course included as any other CRS in the GeoTIFF, can be retrieved. The resulting dataset is a normal GeoTIFF, and it can be processed by any GIS application, as long as it can use the rHEALPix projection which is included, for example, in the widely used PROJ library (PROJ contributors, 2022)

- `dggstools ras-to-rhpx-ras`.

- F3. Transforming a GeoTIFF dataset produced by DGGSTools, in the rHEALPix DGGs, to a vector dataset in the GeoPackage format, with a point for each cell in the original (its centroid). This vector dataset is equivalent to the raster version, up to the point that you can obtain back the raster dataset from this vector dataset, with only minor differences related to implementation details of the GeoTIFF format. The vector dataset is a normal vector dataset, that can be processed by any GIS application, as long as it can use the rHEALPix projection and has support for the GeoPackage format. This vector dataset includes the identifier for each cell, i.e., for each point, and a column with one attribute per each band in the GeoTIFF.

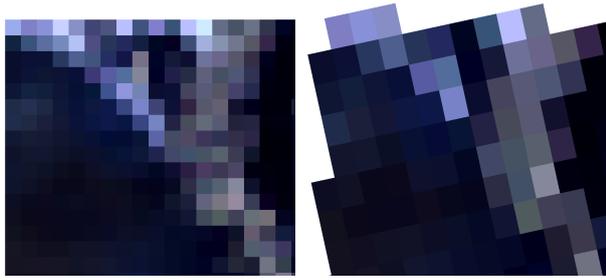
- `dggstools ras-rhpx-to-vec-rhpx`.

- F4. Transforming a vector dataset in rHEALPix produced by DGGSTools with the operation described in F3, to a raster dataset in GeoTIFF which is similar to the one that was used as the input to that operation.

- `dggstools vec-rhpx-to-ras-rhpx`

- F5. Taking a vector dataset with polygons in a common GIS format and coordinate reference system, projected or not, and creating a GeoTIFF dataset with a rasterization of those polygons that follows the constraints of the rHEALPix projection and grid.

- `dggstools vec-to-rhpx-ras`



**Figure 2.** Left: Landsat image; right: the same image in rHEALPix with  $N_{\text{side}}$  3 and resolution 11

F6. Taking a GeoPackage or GeoTIFF produced by DGGSTools, and printing some metadata so the user can find out which parameters of the rHEALPix are used in that file. This includes the parameters of the rHEALPIX grid, such as the  $N_{\text{side}}$  and the resolution index, which do not have a specific place in the GeoTIFF or GeoPackage format specifications.

- `dggstools print-vec-rhpx-metadata`
- `dggstools print-ras-rhpx-metadata`

## 2.2 Architecture

The main modules of this project are the command line interface entrypoint that exposes the use cases and parses the arguments in `__main__`, the multiple data transformation processes in the `rhpx` module, and the `utils` modules that provide support for them.

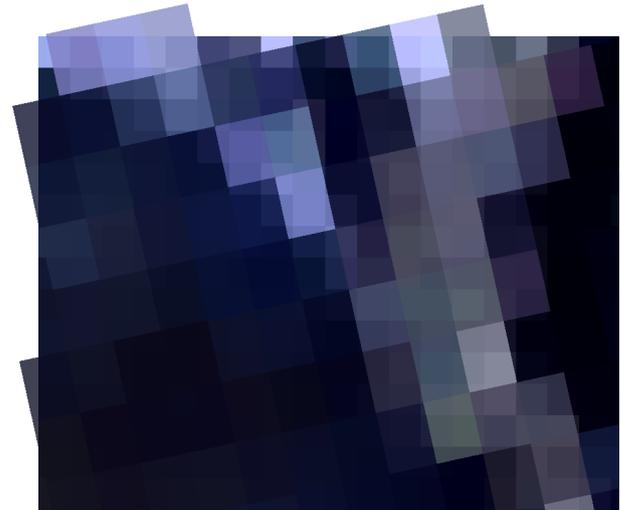
This project uses a number of libraries to provide compatibility with a wide range of spatial formats, both vector and raster, in many of its use cases, and to reuse existing implementations of the rHEALPix projection and the rHEALPix system. The most significant ones are `tt-rhealpixdgs` and `PROJ`, mentioned before, `geopandas` (den Bossche et al., 2024), used to manipulate vector datasets and `rasterio`, also mentioned before, used to process raster datasets.

An automatic pipeline is used to build and test the library. This pipeline is implemented using GitHub Actions. Its purpose is to fetch the latest version of the source code from the repository, retrieve all its dependencies, execute the tests and build the Python wheel package.

## 2.3 Use examples

Although the main functionality has been described in Section 2.1, some examples can provide a better insight into the use of the library and its possibilities. All the examples in this section have been produced by the automatic tests included in the DGGSTools repository, and use the example datasets included there.

Figure 2 shows, on the left side, part of a 30-meter resolution, Albers Equal Area WGS84 Landsat image. The

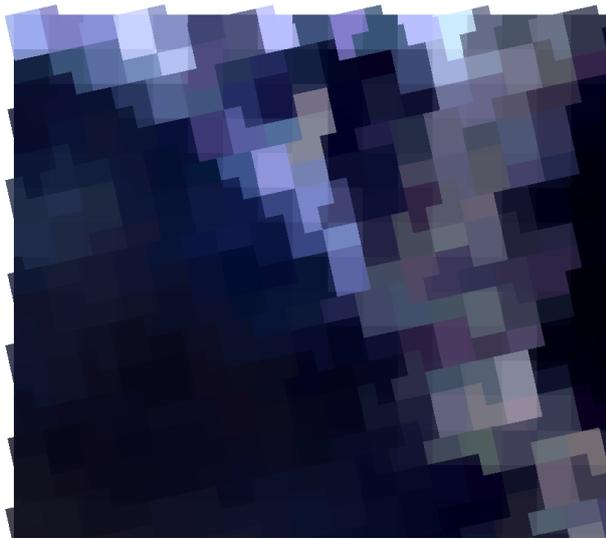


**Figure 3.** Bottom layer: Landsat image; top layer: the same image in rHEALPix with  $N_{\text{side}}$  3 and resolution 11, with an 80% of transparency.

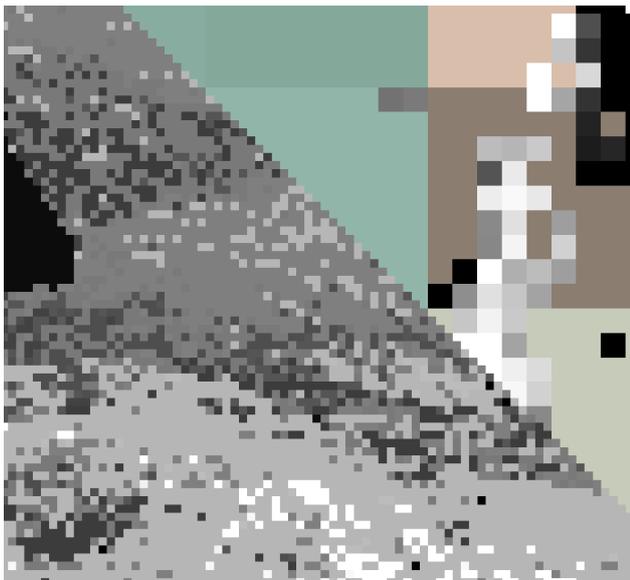
right side shows the same image after being reprojected to rHEALPix, and constrained to the rHEALPix grid with  $N_{\text{side}}$  3 and resolution index 11. The right side is shown as it looks like after projecting to the Albers Equal Area WGS84 reference system, i.e., when loading both images on a desktop GIS application, they overlap as shown in Figure 3. This is also the reason why the rHEALPix cells are not squares in that Figure: they have been projected to Albers Equal Area to show them along with the original dataset. All the images are GeoTIFF files.

If we want a higher resolution in the rHEALPix version, we just need to choose one when calling the `dggstools ras-to-rhpx-ras` command. For example, Figure 4 shows the same two layers as Figure 3 but with the top layer, the rHEALPix version, created with resolution index 12.

In Figure 5 we can see three layers which have been reprojected and constrained to the rHEALPix grid. The bottom layer, in color, is a hypsometric map of the whole Earth originally in geodetic coordinates on the WGS 1984 ellipsoid, and with a resolution of 0.33 decimal degrees per pixel. The intermediate layer, black and white with medium-sized cells in the Figure, is a snow cover map of Europe originally with a resolution of 0.005 decimal degrees per pixel, and also in geodetic coordinates. The top layer, smaller black and white cells in the Figure, is a land cover map originally with a resolution of 485 m per pixel and a sinusoidal (Sanson-Flamsteed) map projection. The map shows an area between Spain and France, and it uses the rHEALPix projection; this is why cells are shown as squares there. These three layers have been reprojected and constrained to the rHEALPix grid with  $N_{\text{side}}$  3 and three different resolutions: 4, 6 and 7, from bottom to top. The Figure shows that they are perfectly aligned to the rHEALPix grid.

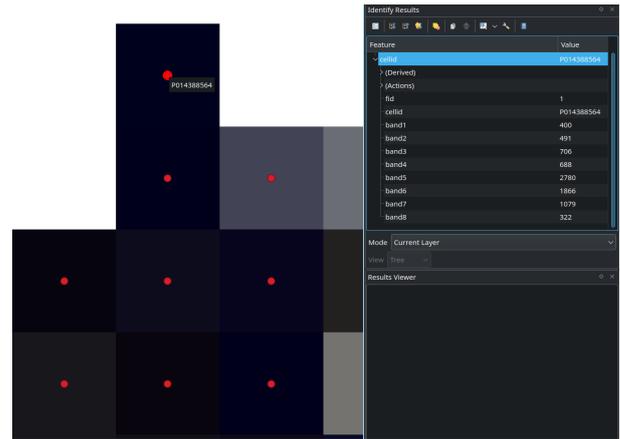


**Figure 4.** Bottom layer: Landsat image; top layer: the same image in rHEALPix with  $N_{side}$  3 and resolution 12, with an 80% of transparency.



**Figure 5.** Three layers reprojected to an rHEALPix grid with DGGSTools to show that the cells are perfectly aligned

Figure 6 shows the previously used Landsat image reprojected to rHEALPix,  $N_{side}$  3, resolution index 9. On top of it, there is a vector layer with points which has been produced with the `dggstools ras-rhpx-to-vec-rhpx` command. Each point corresponds to the centroid of a cell, and the attributes stored along with it, also shown in the Figure, include both the cell identifier of that cell and all the values of the 8 bands in the Landsat image. The point layer is a GeoPackage file. The coordinates of those points could be deduced from the cell identifiers, as they are both unique and well-known identifiers for each cell. However, a GeoPackage layer needs coordinates to be painted by GIS desktop



**Figure 6.** Landsat image in rHEALPix, and the GeoPackage equivalent dataset on top.



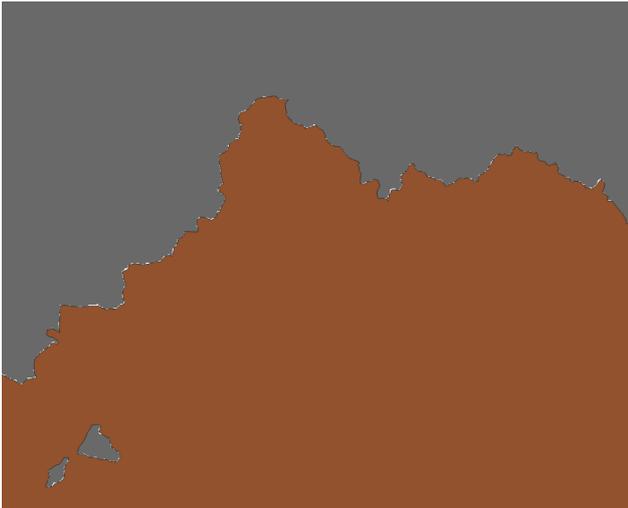
**Figure 7.** North-West of Aragon, Spain, as an rHEALPix raster.

applications, so they are included too. The GeoPackage can be used to generate the original GeoTIFF again using the `dggstools vec-rhpx-to-ras-rhpx` command. A figure showing the result would not very helpful, as that result is undistinguishable from the original.

Figure 7 shows the result of applying the `dggstools vec-to-rhpx-ras` command to a vector dataset with the Aragon region, originally in UTM 30N with the ETRS89 datum. This result is a raster image, a GeoTIFF, in rHEALPix with  $N_{side}$  3, and resolution 10. The original vector file is shown as another layer on top in Figure 8.

### 3 Methods

This section provides the information that is necessary to access to the DGGSTools package and its source code.



**Figure 8.** North-West of Aragon, Spain, as an rHEALPix raster with the original vector dataset on top.

### 3.1 Data and Software Availability Section

Research data and code supporting this publication are available on GitHub, <https://github.com/IAAA-Lab/dggstools>, and can be accessed using the following DOI: <https://doi.org/10.5281/zenodo.10659260>.

The computational workflow supporting this publication is executed via automatic testing published under the same license and repository as the code. The repository is self-documented with all the necessary details such as dependencies and instructions to reproduce the test experiments. The expected run time is currently around 3 minutes.

## 4 Conclusion and Future Work

This paper has presented the first public version of DGGSTools. A small part of its code had been created in some previous experiments (Béjar et al., 2019, 2023), but that code solved narrow and specific problems, and required a Python development environment, or a Jupyter notebook, to be used. Besides that, it was not thoroughly tested and it was not intended to be reused. This paper is a substantial advance from that initial ideas: besides being the first publicly available version, a lot of new functionality has been developed, many automatic tests have been created to improve the overall quality, and the whole project has been given a command line interface, and has been packaged to make it easier for others install and use it as a command-line tool, and not only as a software library.

The main objective of DGGSTools is to facilitate the use of DGGSS in general, although for now is focused on the rHEALPix DGGSS in particular, in geospatial research projects. DGGSTools provides a way to transform cell-based geospatial data to and from three different models without losing any information: a raster model, in the Geo-

TIFF format, a vector model with a pair of coordinates for each cell, in the GeoPackage format, and a less spatial model without explicit coordinates but with unique, well-known identifiers for each cell. We think that this tool may facilitate the collaboration of different geospatial researchers. Being able to easily transform to and from the three different models allows to work with very different toolchains in different stages of the same workflow. This would mean that even different researchers working in different laboratories, but collaborating on the same project, can use their preferred tools and models, knowing that the transformation of datasets required between those stages is automatic and avoids unnecessary problems such as those, usually small but certainly quantifiable, caused by the incongruent alignment of raster datasets. Standardizing on a DGGSS for a given research project would allow to obtain some of the benefits they provide, briefly described in Section 1, and using a package such as DGGSTools makes that standardization as simple as possible by providing a tested solution to some of the problems that need to be solved to do that.

DGGSTools is a young project, and it has many rough edges. By publishing it in its current state, we hope that the geospatial research community can both benefit from it, even with those rough edges it has also being used in several experiments, and also help us to shape its future development, to find out bugs and to improve its quality. As future work, we plan to improve the documentation and the tests, so the library can be used with confidence. Publishing it in a public Python repository is also in the near horizon. The package has some functionality which does not have, yet, a command-line tool. For example the code that was used in the experiments in (Béjar et al., 2019). We will also work to make that available via the command-line interface. Beyond that, we also want to explore other DGGSS; although rHEALPix has a number of advantages given its quadrangular cells, and it might be difficult, or impossible, to replicate every functionality with other DGGSS, extending DGGSTools so at least some of its possibilities can be used with other DGGSS might prove itself useful.

*Acknowledgements.* This paper is part of the R&D project PID2020-113353RB-I00 supported by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033/) and the project T59\_23R supported by the Aragon Regional Government.

## References

- Bowater, D. and Stefanakis, E.: The rHEALPix Discrete Global Grid System: considerations for Canada, *Geomatica*, 72, 27–37, <https://doi.org/10.1139/geomat-2018-0008>, 2018.
- Béjar, R., Latre, M. Á., López-Pellicer, F. J., Nogueras-Iso, J., and Zarazaga-Soria, F. J.: On the problem of providing unique identifiers for areas with any shape on

- Discrete Global Grid Systems, in: Accepted Short Papers and Posters from the 22nd AGILE Conference on Geo-information Science, Limassol, Cyprus, 17-20 June, [https://agile-online.org/images/conference\\_2019/documents/short\\_papers/58\\_Upload\\_your\\_PDF\\_file.pdf](https://agile-online.org/images/conference_2019/documents/short_papers/58_Upload_your_PDF_file.pdf), [retrieved: October, 2020], 2019.
- Béjar, R., Lacasta, J., Lopez-Pellicer, F. J., and Nogueras-Iso, J.: Discrete Global Grid Systems with quadrangular cells as reference frameworks for the current generation of Earth observation data cubes., *Environmental Modelling & Software*, 162, 105 656, 2023.
- den Bossche, J. V., Jordahl, K., Fleischmann, M., McBride, J., Richards, M., Wasserman, J., Badaracco, A. G., Snow, A. D., Ward, B., Gerard, J., Tratner, J., Perry, M., Farmer, C., Hjelle, G. A., Taves, M., ter Hoeven, E., Cochran, M., rraymondgh, Bartos, M., Bell, R., Culbertson, L., Caria, G., Eubank, N., Roggemans, P., sangarshanan, Flavin, J., Rey, S., Gardiner, J., and Kaushik: *geopandas/geopandas: v0.14.3*, <https://doi.org/10.5281/zenodo.10601680>, 2024.
- GDAL/OGR contributors: GDAL/OGR Geospatial Data Abstraction software Library, Open Source Geospatial Foundation, <https://doi.org/10.5281/zenodo.5884351>, 2024.
- Gibb, R., Car, N. J., and Raichev, A.: rHEALPixDGGS: a Python package that implements the rHEALPix Discrete Global Grid System (DGGS), <https://github.com/manaakiwhenua/rhealpixdggs-py>, 2013–.
- Gibb, R., Cochran, B., and Purss, M.: OGC Testbed-16: DGGS and DGGS API Engineering Report, OGC Public Engineering Report OGC 20-039r2, Open Geospatial Consortium, <http://www.opengis.net/doc/PER/t16-D017>, accessed: 2023-04-XX, 2021.
- Gibb, R. G.: The rHEALPix Discrete Global Grid System, in: *IOP Conference Series: Earth and Environmental Science*, vol. 34 of *IOP Conference Series: Earth and Environmental Science*, p. 012012, <https://doi.org/10.1088/1755-1315/34/1/012012>, 2016.
- Gillies, S. et al.: Rasterio: geospatial raster I/O for Python programmers, <https://github.com/rasterio/rasterio>, 2013–.
- IAAA Laboratory: DGGSTools: Vector and Raster Data Manipulation for Discrete Global Grid Systems, IAAA Laboratoy (<https://www.iaaa.es>), <https://doi.org/10.5281/zenodo.10659260>, 2024.
- Kmoch, A., Matsibora, O., Vasilyev, I., and Uemaa, E.: Applied Open-Source Discrete Global Grid Systems, *AGILE: GIScience Series*, 3, 1–6, <https://doi.org/10.5194/agile-giss-3-41-2022>, 2022.
- PROJ contributors: PROJ coordinate transformation software library, Open Source Geospatial Foundation, <https://doi.org/10.5281/zenodo.5884394>, 2022.
- Purss, M., ed.: The OpenGIS Abstract Specification - Topic 21: Discrete Global Grid Systems Abstract Specification), OGC 15-104r5, Open Geospatial Consortium, 2017.
- Sahr, K., White, D., and Kimerling, A. J.: Geodesic Discrete Global Grid Systems, *Cartography and Geographic Information Science*, 30, 121–134, <https://doi.org/10.1559/152304003100011090>, 2003.