



## Global Reference Grids for Big Earth Data

Robert G. Gibb, Matthew B.J. Purss, Zoheir Sabeur, Peter Strobl & Tengteng Qu

To cite this article: Robert G. Gibb, Matthew B.J. Purss, Zoheir Sabeur, Peter Strobl & Tengteng Qu (2022): Global Reference Grids for Big Earth Data, Big Earth Data, DOI: [10.1080/20964471.2022.2113037](https://doi.org/10.1080/20964471.2022.2113037)

To link to this article: <https://doi.org/10.1080/20964471.2022.2113037>



© 2022 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the International Research Center of Big Data for Sustainable Development Goals, and CASEarth Strategic Priority Research Programme.



Published online: 01 Sep 2022.



Submit your article to this journal [↗](#)



Article views: 100



View related articles [↗](#)



View Crossmark data [↗](#)

## Global Reference Grids for Big Earth Data

Robert G. Gibb<sup>a</sup>, Matthew B.J. Purss<sup>b</sup>, Zoheir Sabeur<sup>c</sup>, Peter Strobl<sup>d</sup> and Tengteng Qu<sup>e</sup>

<sup>a</sup>Manaaki Whenua – Landcare Research, Palmerston North, New Zealand; <sup>b</sup>Pangaea Innovations Pty. Ltd, Canberra, Australia; <sup>c</sup>Department of Computing and Informatics, Talbot Campus, University of Bournemouth, Bournemouth, UK; <sup>d</sup>European Commission, Joint Research Centre (JRC), Ispra, Italy; <sup>e</sup>College of Engineering, Peking University, Beijing, China

The emerging field of Discrete Global Grid Systems (DGGS) provides a way to organise, store and analyse spatio-temporal data at multiple resolutions and scales (from near global scales down to microns). DGGS partition the entire planet into a discrete hierarchy of global tessellations of progressively finer resolution zones (or cells). Data integration, decomposition and aggregation are optimised by assigning a unique spatio-temporal identifier to each zone. These identifiers are encodings of both the zone's location and its resolution. As a result, complex multi-dimensional, multi-resolution spatio-temporal operations are simplified into sets of 1D array and filter operations. DGGS are therefore particularly suited for efficient multi-source data processing, storage, discovery, transmission, visualisation, computation, analysis, and modelling.

DGGS are supported by both the Open Geospatial Consortium (OGC) and the International Organization for Standardization (ISO) TC211 standards (OGC Abstract Specification – Topic 21<sup>1</sup>, ISO 19170-1<sup>2</sup>). These published specifications support 2D equal-area DGGS of the Earth's surface. Current work led through both OGC and ISO/TC-211 is drafting standards to specify 3D (3D & equi-volume)<sup>3</sup>, 4D (spatio-temporal)<sup>4</sup> and axis-aligned<sup>5</sup> DGGS, as well as OGC API DGGS<sup>6,7</sup> interface encodings for DGGS infrastructures. The continued effort to develop international standards for DGGS will support the implementation of standardised interoperable Global Reference Grid Infrastructures that can support efficient and scalable integration of Big Earth Data across multiple organisations around the world.

Introducing DGGS, we mentioned both global grids and spatio-temporal identifiers. So, which is more important for a DGGS – the choice of global grid geometry or the structure of the spatio-temporal identifier? When people first encounter DGGS, they often initially focus on the “grid” structure – after all, grid is writ large in the DGGS name, and this has a superficial familiarity with grids used in traditional raster systems. However, as people's understanding of DGGS matures, they come to understand that the simplicity of the spatio-temporal identifiers is what gives DGGS their power and that the discrete hierarchy of global grids merely provides structure for the spatio-temporal identifiers. This is now reflected by the choice of words describing DGGS Reference Systems in the recently published OGC & ISO Specifications – “Spatial Referencing using Identifiers with structured geometry”. So, the emphasis on global grids in both the acronym Discrete Global Grid Systems and the title of this special issue “Global Reference Grids for Big Earth Data”

**CONTACT** Robert Gibb  [gibbr@landcareresearch.co.nz](mailto:gibbr@landcareresearch.co.nz)  Manaaki Whenua – Landcare Research, Palmerston North, New Zealand

© 2022 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the International Research Center of Big Data for Sustainable Development Goals, and CASearch Strategic Priority Research Programme.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

may be misplaced or possibly even misleading. How do the papers in this Big Earth Data Special Issue help us appreciate the difference?

Alexander Kmoch et al. from the Department of Geography, Institute of Ecology and Earth Sciences at the University of Tartu compare the broad functionality and usability of five of the more prominent open-source DGGs libraries, and explore the global variation in geometric distortion and zone area for the DGGs(s) supported by each library. Results are presented as maps, histograms and tables. While a small portion of the variation is due to coding choices, the dominant source of variation is an inherent characteristic of the geometry of each DGG and therefore the results provide a very useful introduction and guide to the compromises inherent in the geometry selection for a DGG. The paper also makes the distinction between those libraries that focus on providing grids that satisfy particular criteria, and those libraries that also generate identifiers that support efficient processing by providing spatial functions that are driven by the DGG zone's identifier.

The next two papers provide examples of machine learning that use DGGs to manage many highly heterogeneous data sources. Bing Han et al. extend the 2D surface DGGs examined by Kmoch et al. to 3D DGGs using the GeoSOT-3D DGGs from the College of Engineering, Peking University. They model site selection of emergency airports in the Yangtze river delta. This is an area with high population density that is prone to flooding. So, emergency airports need to be dynamically chosen in response to rapidly evolving situations, and sub-optimal choices have potentially severe consequences. Andrew Rawson et al. from the University of Southampton use their `dggridpy` library to deploy Big Data technology across North America to assess the likelihood of maritime risk associated with 200,000 shipping incidents during the last few decades in the context of harsh environmental conditions and vessel movements using Automatic Identification System (AIS) data. While both these papers focus primarily on their application, they also describe how using DGGs helps with their analyses. Bing Han et al. describes how the use of GeoSOT-3D's 3D spatial identifier allows them to use a traditional database structure with a single column containing the 3D identifier for their analysis. Andrew Rawson et al. initially focus on DGGs as grids and describe how the advantages to their analysis of both equal-area and uniform nearest-neighbor properties supported their choice of a DGG using Icosahedral Snyder Equal Area projection with grids of hexagonal zones. For the analyses, once again table-based data structures are used indexed by zone identifiers to feed data through the machine learning pipeline.

The remaining papers address three different forms of challenges associated with DGGs – extreme scaling, pathways to adoption and future research. Xinyu Tang et al. from College of Land Science and Technology, China Agricultural University, Beijing, and colleagues address the wider challenge of Big Data remote sensing analytics at peta- and exa-scale. While they do not consider DGGs specifically, the issues they discuss and the solutions they examine do refer to techniques such as space filling curves that are widely used for spatial identifier logic in DGGs. This paper underscores that moving to peta- and exa-scale requires special consideration, whatever technology is used – whether traditional raster or DGGs. Even vector systems at giga-scale struggle, let alone at peta- to exa-scale, so it would be very interesting to see research explicitly exploring the issues Xinyu Tang et al. discuss in the context of DGGs identifiers and diverse vector and raster source data.

Jeffery Thompson et al. from Minnesota Supercomputing Institute, University of Minnesota, and from the National Snow and Ice Data Center, University of Colorado tackle

the challenge of adoption by adding DGGs capability to an existing package – Equal-Area Scalable Earth Grid (EASE-Grid 2.0) – that is already widely used by the Earth Observation (EO) community. Their proposal underlines how key DGGs characteristics such as a pre-defined hierarchy of nested grids and structured cell indices substantially benefit global scale EO data storage and interoperability. At the same time, it demonstrates the dilemma in which other assets like seamless global coverage or a consistent subdivision pattern across the grid hierarchy are being sacrificed to data legacy and user comfort in pursuit of easier acceptance. This raises the question whether this initial version of EASE-DGGs is a sufficient endpoint in DGGs adoption or whether it is a valuable first step towards the use of DGGs by the Earth Science community. The challenge will be to ensure that this is just a first step towards a wider adoption of DGGs, so that in time the community can realise the full DGGs vision described by Goodchild (Goodchild, 2019) and address the shortcomings of the early vector and raster decisions whose legacy continues to hold back the global community (Goodchild, 2018).

Majid Hojati et al. from the Department of Geography and Environmental Studies, Wilfrid Laurier University, Waterloo and University of Saskatchewan pick up on recent developments in OGC's DGGs standards, and explore the implications compared with traditional raster and vector-based GIS. The authors take readers on deep dives into particular issues discussing both the superficial pros and cons and also exposing deeper practicalities that will need to be addressed through future research, richer implementations and evolving best practices.

So what do these papers tell us about the relative importance of grids and identifiers? Alexander Knoch et al. demonstrate the trade-offs between different grids that we need to be aware of when we select a DGGs for our application. Jeffery Thompson et al. show that a DGGs-compliant spatial index can be developed for the Earth Observation community's EASE-Grid v2 system of grids, thereby enabling DGGs capabilities for that community, and Xinyu Tang et al. and Andrew Rawson et al. show the power of leveraging DGGs spatial identifiers for complex analysis of mixed raster and vector data sources. So, the choice of grids is an important consideration in DGGs selection, but it is the indexes that drive analyses.

The challenge of leveraging and benefiting from the wisdom and insights we can gain from Big Earth Data is significant and increasingly being recognised by those who are beginning to drown in a sea of data that is rising at an exponential rate. Our ability to increase our compute capacity to keep up with this growth in data is now limited by the laws of physics, and we have over-reached Moore's Law, as evidenced most obviously by the switch from single processor to parallel processing in the early 2000s (Hilbert, 2016; Karl, 2018; Rydning, Reinsel, & Gantz, 2018).

The papers in this issue clearly demonstrate that there is no single DGGs which is ideal for every use case nor are the advantages of using DGGs technologies realisable at no cost; however, it is also apparent that established techniques for global geospatial data storage and analysis have reached their limits and are now acting as an impediment rather than a facilitator for realising the combined power of a "Digital Earth". To solve this problem, we need to rethink the way we are storing, managing, sharing and working with our data, and DGGs offers a very promising opportunity.

The perceived lack of established and available DGGs solutions, which has been a barrier to implementation, is rapidly becoming an issue of the past. The papers in this

Special Issue provide a solid argument for the application and adoption of DGGs to the Big Earth Data paradigm and some of the challenges that we as a community must resolve to truly realise the concept of the “Digital Earth”.

We would like to thank all contributing authors, including those whose papers were not selected in this Special Issue. We would also like to express our hearty gratitude to those anonymous reviewers. Last but not least, special thanks go to the Executive Editor-in-Chief Dr Changlin Wang and the Assistant Editor Dr Linlin Guan for their great assistance.

## Notes

1. Gibb, R.: OGC 20-040r3, Topic 21 – Discrete Global Grid Systems – Part 1 Core Reference System and Operations and Equal Area Earth Reference System. Open Geospatial Consortium (2021). <http://www.opengis.net/doc/AS/dggs/2.0>.
2. ISO: ISO 19170-1:2021, Geographic information – Discrete Global Grid Systems Specifications – Part 1: Core Reference System and Operations, and Equal Area Earth Reference System. International Organization for Standardization, Geneva (2021). <https://www.iso.org/standard/32588.html>.
3. Gibb, R. (Ed): OGC 20-047r0, Topic 21 - Discrete Global Grid Systems – Part 2 Three Dimensional and Equi-volume DGGs Reference Systems. Open Geospatial Consortium (Draft Abstract Specification – in Prep).
4. Gibb, R (Ed): OGC 20-048r0, Topic 21 - Discrete Global Grid Systems – Part 3 Spatio-temporal DGGs Reference Systems. Open Geospatial Consortium (Draft Abstract Specification – in Prep).
5. Gibb, R. (Ed): OGC 20-049r3, Topic 21 - Discrete Global Grid Systems – Part 4 Axis Aligned DGGs Reference Systems. Open Geospatial Consortium (Draft Abstract Specification – in Prep).
6. Purss, M. (Ed): OGC 21-038r0, OGC API – Discrete Global Grid Systems – Part 1: Core. Open Geospatial Consortium (Draft Specification – in Prep). <https://github.com/opengeospatial/ogcapi-discrete-global-grid-systems>.
7. Purss, M. (Ed): OGC API – Discrete Global Grid Systems – Part 1: Core. Open Geospatial Consortium (Draft schema – in Prep). <https://developer.ogc.org/api/dggs/index.html>.

## Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Notes on contributors



**Robert Gibb** is passionate about the environment and dedicated to addressing the climate, ecosystem and societal crises the world faces, and he developed the rHEALPix DGGs as a technology to address these issues. Robert recently retired after a 35-year career as a GI researcher and innovator with Manaaki Whenua Landcare Research, New Zealand, where he led the Informatics Team. He is now a Research Associate in Informatics, focusing on OGC & ISO standards. Robert was co-founder and co-chair of the OGC DGGs SWG, is an OGC co-chair of the OGC–ISO TC 211 Joint Advisory Group and editor of the recently published DGGs Specifications ISO/19170-1:2021 & OGC AS Topic 21 – Pt 1 v2.



**Matthew B.J. Purss** is a geophysicist, data scientist, thought leader and entrepreneur with over 24 years' experience in the exploration, research and government sectors. He is a world leader in the standardisation, development and implementation of DGGs technologies through his roles as Founding Co-Chair of the OGC DGGs Standards and Domain Working Groups, Co-Chair of the OGC-ISO/TC211 Joint Advisory Group and Co-Chair of the OGC Points of Interest Standards Working Group. He is the Co-Founder and Chief Executive Officer of Pangaea Innovations Pty. Ltd., a spatial data technology startup company developing commercial applications of 3D & 4D DGGs technologies.



**Zoheir Sabeur** investigates natural phenomena, behaviour and processes using experimental sensor observations for science and knowledge discovery. He uses OGC DGGs standards on spatial and temporal Earth data for enabling the scalability of his multi-modal data fusion and deep learning distributed methods. These aim at capturing advanced situational awareness in the context of understanding potential critical events and forecasting. Zoheir has been working in this domain for more than 25 years and validating his research and development on environmental, health and security driven intelligent agents. He is currently Professor of Data Science and Artificial Intelligence at Bournemouth University, Department of Computing and Informatics, United Kingdom.



**Peter Strobl**, during more than 30 years in a broad range of Earth Observation topics, Peter recognised spatial representation and architecture as a key to interoperability of diverse geospatial data. He works as a Senior Scientist in the European Commission's Joint Research Centre advising the Copernicus Earth Observation programme on questions of references and standards and on the development of future sensors and data technologies. He is a member or chair of various working groups and expert panels within OGC, CEOS, ESA and NASA/USGS.



**Tengeng Qu** is currently a Research Assistant Professor of the Department of Aeronautics and Astronautics in the College of Engineering, Peking University of China. Her research interests focus on geospatial big data analysis and global subdivision grids. Since 2019, she has served as a Standard Expert of Geographic Information Science in OGC DGGs SWG and ISO/TC211. In particular, she is now in charge of the drafting of "OGC 20-049r3, Topic 21 – Discrete Global Grid Systems – Part 4 Axis Aligned DGGs Reference Systems".

## References

- Goodchild, M. (2018). Reimagining the history of GIS. *Annals of GIS*, 24(1), 1–8. doi:<https://doi.org/10.1080/19475683.2018.1424737>
- Goodchild, M. (2019). Preface. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 54(1), 1–3. doi:<https://doi.org/10.3138/cart.54.1.preface>
- Hilbert, M. (2016). *Enabling digital development. The data revolution*. World Development Report, 244–247.
- Karl, R. (2018). <https://github.com/karlrupp/microprocessor-trend-data>
- Rydning, D. R. J. G. J., Reinsel, J., & Gantz, J. (2018). *The digitization of the world from edge to core*. Framingham: International Data Corporation.