AGILE: GIScience Series, 7, 48, 2026. https://doi.org/10.5194/agile-giss-7-48-2026
Proceedings of the 29th AGILE Conference on Geographic Information Science, 16–19 June 2026.
Eds.: Evelyn Uuemaa, Alexander Kmoch, Holger Virro, and Simon Scheider
This contribution underwent peer review based on a full paper submission.

DGGS-Native Data Cubes: A Design Pattern for Scalable, Distortion-Aware Analytics

Michael Zamudio-Monserratt [1], Evelyn Uuemaa [2], and Alexander Kmoch [2]

[1]INCT/Odisseia, Observatory of Socio-Environmental Dynamics, University of Brasilia, Brazil
[2]Landscape Geoinformatics Lab, Institute of Ecology and Earth Sciences, University of Tartu, Estonia

Correspondence: Michael Zamudio-Monserratt (michael.zamudio03@gmail.com)

Abstract. The rapid growth of geospatial big data has intensified the need for efficient frameworks to store, process, and analyse large-scale, multidimensional datasets. Geospatial data cubes have emerged as a key paradigm for organising spatio-temporal information into analysis-ready structures, enabling scalable analytics across Earth observation and related domains.

This paper presents a synthesis-oriented analysis of recent mathematical and architectural advances in geospatial data cube infrastructures, with a particular focus on multidimensional indexing, sparse computation and compression, and spatial tessellation based on Discrete Global Grid Systems (DGGS). Rather than conducting a systematic review, the study integrates theoretical and system-level contributions to examine how these methods jointly address limitations of projection-dependent raster models, improve storage efficiency, and support consistent multi-resolution analysis.

We argue for DGGS-indexed data cubes, where the spatial dimension is treated as a first-class, hierarchical global grid identifier, serving as a unifying computational substrate that integrates multi-resolution spatial referencing with sparse tensor computation, thereby enabling globally consistent, scalable, and actionable geospatial analytics. This perspective clarifies their role as a foundational component for scalable, reproducible, and globally consistent geospatial analytics, while outlining key challenges and research directions for their operational adoption. By highlighting points of convergence between DGGS-based spatial referencing, cloud-native storage formats, and scalable computational strategies, the paper reframes geospatial data cubes not merely as data storage abstractions, but as integrated computational infrastructures.

Submission Type. theory; review.

BoK Concepts. [IP5-1] Data cubes; [DM3-7] Hierarchical data models; [DA3-6] Cloud and Grid computing

Keywords. Geoinformatics, Discrete Global Grid Systems, DGGS, Geospatial data, Multidimensional indexing

1 Introduction

The increasing volume, variety, and resolution of geospatial data challenge traditional storage and processing approaches. In recent years, data cubes have become a foundational paradigm for managing and analysing multidimensional data, particularly in geospatial contexts, originally developed within business intelligence as Online Analytical Processing (OLAP) abstractions (Baumann, 2021; Wu et al., 2024).

Geospatial data cubes organise spatial and temporal indexing dimensions together with measured variables—such as spectral bands or environmental attributes—within multidimensional array structures, enabling systematic integration, querying, and analysis of large EO and sensor datasets (Lu et al., 2018; Baumann et al., 2019). Recent advances in multidimensional indexing, compression techniques, and matrix-based computation have significantly improved scalability and performance, supporting rapid querying of massive geospatial datasets (Gao et al., 2022; Zhu et al., 2021). Projection-dependent rasters limit scalability and interoperability because their fixed coordinate systems hinder efficient multi-resolution analysis and integration across diverse spatial datasets (Purss et al., 2019; Kmoch et al., 2022b). This article aims to propose directions for a research agenda for DGGS-based geospatial data cubes, positioning DGGS-based spatial indexing as foundational for next-generation infrastructures. Framed as a synthesis of current advances with a call to action, it identifies key research directions to address persistent challenges in global consistency, interoperability, and multi-resolution scalability.

Accordingly, this work adopts a conceptual and integrative approach, focusing on the convergence of geospatial data cube architectures with Discrete Global Grid Systems (DGGS)–based spatial indexing. Rather than conducting a systematic review, the paper follows a synthesis-oriented perspective that integrates theoretical, mathematical, and system-level contributions relevant to contemporary data cube infrastructures. The analysed literature is selected along four core dimensions: DGGS-based spatial indexing, sparse computation and compression, and scalable geospatial analytics architectures.

Accordingly, this paper positions itself as a conceptual synthesis bridging foundational DGGS research with applied geospatial data cube implementations. By organising dispersed methodological advances into a unified framework, it reframes geospatial data cubes not merely as data management abstractions, but as evolving computational infrastructures capable of supporting scalable, reproducible, and multi-resolution geospatial analytics.

2 Evolution of Geospatial Data Cubes towards DGGS-based Infrastructures

Geospatial data cubes originated as multidimensional abstractions within Online Analytical Processing (OLAP), where dimensions were primarily categorical and non-spatial. As these concepts were adopted in geospatial and Earth Observation (EO) domains, they evolved into array-based models capable of managing large spatio-temporal raster datasets (Baumann, 2021). While EO data cubes enabled systematic access to satellite time series, their reliance on projection-dependent raster grids introduced structural limitations for globally consistent, multi-resolution analysis (Li et al., 2018; Purss et al., 2019).

The increasing scale, resolution, and heterogeneity of geospatial data has therefore motivated a transition towards global geospatial data cube paradigms, where spatial referencing and indexing are treated as core infrastructural components rather than implementation details (Wu et al., 2024; Tian et al., 2022). Within this evolution, Discrete Global Grid Systems (DGGS) have emerged as a key enabler, providing a hierarchical and equal-area discretisation of the Earth's surface that supports consistent analysis across regions and scales.

2.1 Computational and Mathematical Enablers

Scalability in modern geospatial data cubes relies on computational primitives addressing data volume and dimensionality. Sparse representations and compression are key, as spatio-temporal data cubes exhibit intrinsic sparsity in space, time, or variables (Gao et al., 2023). Sparse matrix and tensor models reduce storage while enabling efficient querying and analytics over large multidimensional domains. These models integrate linear algebra operations foundational to many workflows. Mapping data cubes to coordinate-based storage replaces dense traversals with key–value-based data access and windowed lookups, enabling efficient neighbourhood-based stencil and kernel operations.

Rather than mere implementation optimisations, such primitives are embedded within data cube architectures, making computational efficiency a structural system property (Sidiropoulos et al., 2017). This is exemplified by community-driven tools like XDGGS, which extend array-based libraries (e.g., Xarray) to natively support DGGS-indexed data cube computations (Kmoch et al., 2024). The DGGS hierarchy's inherent spatial partitioning aligns naturally with sparse computation and parallel execution, enhancing scalability and efficiency.

2.2 Spatial Indexing and DGGS Integration

Efficient spatial indexing is fundamental to geospatial data cube design. Traditional structures, such as R-trees and quad-trees, support effective spatial querying but are not inherently designed for globally consistent multi-resolution representations. DGGS address this by providing hierarchical indexing schemes that encode spatial resolution and neighbourhood relationships directly (Open Geospatial Consortium, 2017; Hojati et al., 2022).

For example, R-trees are robust and widely used for indexing vector geometries, providing efficient, data-dependent access to spatial objects. Analogously, DGGS offer hierarchical indexing schemes (e.g. H3, IGEO7, HEALPix ZUNIQ) that behave similarly to quad- or oct-trees, while encoding cell identifiers as ordered 64-bit integers. These identifiers support both fine-grained (pixel-level) indexing and more efficient representations as bounding extents defined by the cell geometry (e.g. hexagons or triangles), thereby enabling consistent multi-resolution spatial analysis across global extents.

By mapping observations to discrete, globally indexed cells, DGGS-based data cubes enable seamless integration of spatial and temporal dimensions, preserve spatial locality, and eliminate the need for repeated reprojection or tiling strategies (Li and Stefanakis, 2020; Li et al., 2024). As a result, DGGS facilitate efficient multi-resolution querying, aggregation, and neighbourhood analysis within unified data cube infrastructures.

Recent work demonstrates that DGGS are no longer confined to theoretical formulations but are increasingly supported by a maturing open-source software ecosystem (Kmoch et al., 2022a). Comparative analyses highlight how differences in indexing, traversal, and grid topology directly influence their suitability for data cube integration and large-scale analytics (Purss et al., 2019).

2.3 Tessellation and Distortion-aware Spatial Models

Spatial tessellation determines how the Earth's surface is discretised into analytical units and directly influences aggregation and neighbourhood operations. For large-scale analytics, area preservation is essential, as it ensures statistically comparable spatial units across regions and resolutions. DGGS-based tessellations address this requirement through equal-area, hierarchical cells (Mahdavi-Amiri et al., 2015; Kmoch et al., 2022b).

Among common tessellation strategies, hexagonal and triangular grids provide uniform neighbourhood structures and reduced directional bias, properties that are particularly advantageous for kernel-based operations and spatial interpolation (Uher et al., 2019; Li et al., 2022a). By embedding tessellation choice within a hierarchical spatial reference system, DGGS-based data cubes integrate distortion-aware modelling as a core design principle rather than a post-processing concern.

3 Advances in Data Cube Storage

Storage design directly influences the performance and usability of geospatial data cubes at large scales. Traditional file formats often struggle to scale efficiently with hierarchical indexing and multi-resolution analysis. Recent advances in cloud-optimised storage formats, with spatio-temporal compression and hierarchical strategies, have become central to modern data cube infrastructures.

Cloud-native formats like Zarr (Khan et al., 2024) and Apache Parquet suit large, multidimensional geospatial datasets. Zarr enables chunked, compressed, and distributed storage of array-based data, allowing access to subsets of large data cubes. Parquet's columnar storage efficiently supports heterogeneous attributes such as spatial indices, temporal identifiers, and measured variables. In DGGS-based data cubes, where grid cell identifiers replace explicit geometries, Parquet remains relevant without requiring geometry extensions.

Spatio-temporal compression plays a key role in reducing data volume. Time-series data exhibit gradual temporal variation, making them suited to differencing and temporal aggregation strategies (Faghmous and Kumar, 2014; Ma, 2023; Fernandes et al., 2024). Spatial compression techniques based on multi-resolution transforms, like wavelets, reduce storage requirements by representing spatial patterns across hierarchical scales (Yang et al., 2024). These approaches align with DGGS-based indexing, where hierarchical cell structures support distortion-aware, multi-scale access.

Hierarchical storage mechanisms enable efficient navigation across resolutions, supporting global overviews and local analysis (University of Calgary and Stefanakis, 2023). By combining cloud-native formats with DGGS-inspired indexing and compression strategies, modern data cubes achieve scalable, distortion-aware storage suited to global geospatial analytics in climate monitoring, environmental research, and urban analysis (Munteanu, 2024).

4 Computational Query Optimisation for Large-Scale Data Cubes

This section focuses on how computational methods can enable efficient querying and real-time analytics in large-scale geospatial data cubes. As datasets reach petabyte scale, query optimisation is critical for responsive access to spatio-temporal information across multiple resolutions.

Advanced linear algebra and signal processing techniques are central here. For example, Fourier transforms efficiently decompose periodic or cyclic time-series data such as climate or oceanographic observations, into frequency components, enabling compact representations and accelerated temporal analysis (Vu et al., 2022). These methods extend beyond traditional raster grids to spherical DGGS frameworks like HEALPix (Górski et al., 2005) and more recently to hexagonal DGGS representations (Li et al., 2022b), illustrating their spectral applicability across grid types. These spectral analyses are directly supported by neighbourhood operations of DGGS-based indexes.

DGGS-based hierarchical spatial identifiers enable efficient localisation of queries across adjacent cells (Yao et al., 2019), and significantly improve spatial lookup performance including spatial joins and interpolation tasks. Probabilistic models embedded within the cube structure—using Gaussian processes, Bayesian inference, and Monte Carlo sampling—support queries returning confidence intervals or distributions, prioritising computation in variable regions while avoiding unnecessary processing in stable areas(Cuzzocrea and Gunopulos, 2010; Wang et al., 2017; Raveendran and Sofronov, 2021).

While data cube discussions have largely focused on raster and chunked-array representations, vector data, such as boundaries, networks, and point observations, remain essential in many geospatial workflows. Modern columnar formats such as GeoParquet offer efficient storage and querying for vector data (Khan et al., 2024), and can be combined with DGGS-based hierarchical indexing to support hybrid workflows where needed.

Finally, advances in DGGS geometry and mapping algorithms enhance interoperability between grid systems. Bidirectional mappings between icosahedral and rhombic triacontahedral DGGS enable reuse of algorithms and datasets across tessellations, supporting flexible, scalable data cube architectures(Huang et al., 2024a, b). Together, these developments show how query optimisation in DGGS-indexed cubes arises from aligning mathematical methods, hierarchical spatial indexing, and system-level design.

5 Case studies and method comparison

Recent applications in geosciences illustrate how the theoretical and architectural principles discussed in previous sections are being operationalised within contemporary geospatial data cube systems. Rather than providing an exhaustive survey, this section highlights selected examples that demonstrate how DGGS-based indexing, hierarchical representations, and scalable computation enable practical advances across climate modelling, satellite data processing, and environmental analysis.

5.1 Climate modelling

Climate modelling represents a critical application domain for geospatial data cube infrastructures, as the increasing dimensionality of atmospheric and environmental variables exceeds the capabilities of traditional projection-dependent grids. Early climate simulation frameworks already explored geodesic discretisations of the Earth to minimise distortion and support global consistency. Building on these foundations, DGGS generalise geodesic principles into hierarchically indexed, equal-area spatial reference systems that integrate naturally with data cube architectures.

When embedded within data cubes, DGGS enable consistent multiscale representations of climate variables and support distributed computation across global domains. This facilitates the integration, querying, and analysis of heterogeneous climate datasets while preserving geometric and topological consistency across Earth system simulations. Recent work has demonstrated how adaptive data cube frameworks combining Earth observation data, machine learning, and context-aware processing can support environmental monitoring and climate-related applications at scale (Temenos et al., 2024).

5.2 Satellite data processing

The rapid growth of high-resolution satellite imagery from missions such as Landsat, Sentinel, and commercial sensors has intensified demands on data cube architectures. Several studies have demonstrated how spatio-temporal data cubes reduce heterogeneity in satellite time series by representing observations within consistent multidimensional structures (Xi et al., 2019; Xu et al., 2020).

DGGS-based approaches have gained particular attention in this context, as they address limitations of conventional tiling schemes related to overlap, redundancy, and global consistency. For example, DGGS-aligned grid systems improve statistical consistency across resolutions and reduce unnecessary data duplication compared to projection-based raster tiling (Thompson et al., 2022; Bauer-Marschallinger and Falkner, 2023). Comparative evaluations of DGGS implementations—including H3, rHEALPix, and ISEA variants—have further clarified their respective trade-offs for satellite data integration and analysis (Salgues et al., 2023).

5.3 Environmental spatial data analysis

Environmental monitoring applications further illustrate the analytical benefits of DGGS-integrated data cubes. By combining satellite observations with in situ measurements, data cubes support scalable analysis of air quality, land use, irrigation, and ecosystem dynamics. Studies have shown that distortion-aware DGGS frameworks can significantly improve predictive accuracy in global-scale environmental models by avoiding area biases inherent in traditional projections (Wagner et al., 2024).

DGGS-based analytical frameworks have also demonstrated competitive performance relative to raster models while offering improved interoperability and reproducibility across heterogeneous datasets (Law and Ardo, 2025). These examples highlight how grid choice, indexing strategy, and system architecture jointly influence analytical outcomes, reinforcing the role of DGGS as a foundational spatial component in next-generation geospatial data cube systems.

6 Discussion and conclusions

Building on the theoretical and system-level advances reviewed, this section synthesises the broader implications of contemporary geospatial data cube research and highlights the strategic challenges shaping future developments. By connecting mathematical foundations with operational architectures, the discussion emphasises the transition from traditional storage-oriented models towards integrated, scalable computational infrastructures.

6.1 Synthesis of theoretical advances

Contemporary geospatial data cubes have evolved beyond their original role as OLAP-style storage abstractions into mathematically grounded computational frameworks (Baumann et al., 2016). The integration of DGGS-based spatial referencing with multiscale hierarchical indexing would provide geometric consistency and mitigate long-standing issues related to cartographic distortion.

The key to this change is using sparse linear algebra, tensor representations, and space-filling curves. These make optimization and parallelism a basic part of the system, not just extra details. In this setting, DGGS act as a common spatial reference. They help make algorithms efficient and keep geographic data organized. This allows for large-scale and sustainable analysis of global data (Robertson et al., 2020).

Recent work further advances data cube architectures by treating DGGS grid cells as spatial tokenizers for AI models. Instead of viewing AI models as external consumers of static outputs, these approaches position data cubes as active computational substrates that support distributed inference and on-demand model execution (Kmoch et al., 2025), thereby reducing data movement and enhancing efficiency in large-scale Earth observation applications.

Instead of developing a full theory of geocomputational infrastructures here, we aim to synthesise convergent design patterns, i.e. DGGS-based indexing, sparse tensor methods, and cloud-native storage, and point towards the need for a stronger theoretical foundation on the intersection. Rather than framing these systems as purely implementation challenges, this perspective highlights the importance of a unifying framework capable of representing a broad range of geocomputations, including both field-based and object-based (raster and vector) representations. In this context, the identified design patterns are not only architectural solutions but conceptual building blocks that inform future developments.

6.2 Practical implications and outlook

From a systems perspective, combining DGGS with cloud-native storage (e.g. Zarr, Parquet) enables infrastructures that can operate at petabyte scale while still supporting interactive analytics. Hierarchical indexing and sparse/tensor methods reduce costly reprojections and allow seamless transitions between global and local scales, improving latency for decision-making in domains such as climate monitoring and urban planning (Gao et al., 2022; Yue et al., 2024). Yet these advances demonstrate rather potential than fully mature solutions.

Furthermore, rapid growth in Earth observation data has outpaced semantic interpretability, leaving persistent challenges in preprocessing and analysis. Semantic Earth Observation Data Cubes (SEOD) offer one promising path by coupling numerical observations with machine-interpretable semantics (Augustin et al., 2019). Lastly, projection-dependent raster workflows still impose structural inefficiencies, such redundant storage and avoidable computational overhead due to (temporal reprojection needs) (Sousa, 2018).

We advocate shifting focus from tessellation schemes to end-to-end operational workflows that prioritise scalable DGGS-indexed addressing and chunking (or sharding) strategies, and reproducible, analysis-ready data pipelines (Kmoch et al., 2024). In parallel, the emergence of “AI Cubes” underscores the need to embed model retrieval and distributed inference inside the cube, enabling GeoAI workflows that are spatially explicit and natively multiscale (Yue et al., 2024). Progress also depends on rigorous, community-maintained benchmarks that compare DGGS-indexed cubes to projection-dependent baselines across tasks (aggregation fidelity, neighbourhood operations, time-to-answer, storage footprint).

In sum, the field is shifting from static storage containers to DGGS-indexed computational infrastructures. Sustained impact will come less from isolated algorithmic advances and more from coherent, end-to-end system design that aligns mathematical compute performance (sparse tensor methods), spatial referencing (hierarchical, distortion-reduced DGGS), and cloud-optimised architecture. Such alignment is essential for robust, interoperable, and truly actionable geospatial analytics at global scale.

Data and Software Availability

This paper does not contain associated data or software beyond the extended literature collection.

Declaration of Generative AI in writing

The authors declare that generative AI tools were utilised during the preparation of this manuscript to support language editing, improve grammar and clarity, and restructure text for coherence and readability. Furthermore, these tools assisted in the improved articulation of conceptual frameworks, particularly regarding English language expression. The use of generative AI did not involve the generation of original scientific content, primary research data, methodological designs, or substantive conclusions. All scientific interpretations, the selection of literature, conceptual framing, and final editorial decisions remain the sole responsibility of the authors.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

All authors contributed equally to the conceptualisation, research, technical review, and final editing of the manuscript.

Acknowledgements

This work was funded by the Estonian Research Agency (grant number PRG1764, PSG841), Estonian Ministry of Education and Research (Centre of Excellence for Sustainable Land Use (TK232)), and by the European Commission (ERC, WaterSmartLand, 101125476, Interreg-BSR, HyTruck, #C031). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Augustin, H., Sudmanns, M., Tiede, D., Lang, S., and Baraldi, A.: Semantic earth observation data cubes, Data, 4, https://doi.org/10.3390/data4030102, iSBN: 4366280447574, 2019.

Bauer-Marschallinger, B. and Falkner, K.: Wasting petabytes: A survey of the Sentinel-2 UTM tiling grid and its spatial overhead, ISPRS Journal of Photogrammetry and Remote Sensing, 202, 682–690, https://doi.org/10.1016/j.isprsjprs.2023.07.015, 2023.

Baumann, P.: A general conceptual framework for multi-dimensional spatio-temporal data sets, Environmental Modelling & Software, 143, 105 096, https://doi.org/10.1016/j.envsoft.2021.105096, 2021.

Baumann, P., Mazzetti, P., Ungar, J., Barbera, R., Barboni, D., Beccati, A., Bigagli, L., Boldrini, E., Bruno, R., Calanducci, A., Campalani, P., Clements, O., Dumitru, A., Grant, M., Herzig, P., Kakaletris, G., Laxton, J., Koltsida, P., Lipskoch, K., Mahdiraji, A. R., Mantovani, S., Merticariu, V., Messina, A., Misev, D., Natali, S., Nativi, S., Oosthoek, J., Pappalardo, M., Passmore, J., Rossi, A. P., Rundo, F., Sen, M., Sorbera, V., Sullivan, D., Torrisi, M., Trovato, L., Veratelli, M. G., and Wagner, S.: Big Data Analytics for Earth Sciences: the EarthServer approach, International Journal of Digital Earth, 9, 3–29, https://doi.org/10.1080/17538947.2014.1003106, 2016.

Baumann, P., Misev, D., Merticariu, V., and Huu, B. P.: Datacubes: Towards Space/Time Analysis-Ready Data, in: Service-Oriented Mapping, edited by Döllner, J., Jobst, M., and Schmitz, P., pp. 269–299, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-72434-8_14, series Title: Lecture Notes in Geoinformation and Cartography, 2019.

Cuzzocrea, A. and Gunopulos, D.: Efficiently Computing and Querying Multidimensional OLAP Data Cubes over Probabilistic Relational Data, in: Advances in Databases and Information Systems, edited by Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Catania, B., Ivanović, M., and Thalheim, B., vol. 6295, pp. 132–148, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-15576-5_12, series Title: Lecture Notes in Computer Science, 2010.

Faghmous, J. H. and Kumar, V.: Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities, in: Data Mining and Knowledge Discovery for Big Data, edited by Chu, W. W., vol. 1, pp. 83–116, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-40837-3_3, series Title: Studies in Big Data, 2014.

Fernandes, V., Carvalho, G., Pereira, V., and Bernardino, J.: Analyzing Data Reduction Techniques: An Experimental Perspective, Applied Sciences, 14, 3436, https://doi.org/10.3390/app14083436, 2024.

Gao, F., Yue, P., Cao, Z., Zhao, S., Shangguan, B., Jiang, L., Hu, L., Fang, Z., and Liang, Z.: A multi-source spatio-temporal data cube for large-scale geospatial analysis, International Journal of Geographical Information Science, 36, 1853–1884, https://doi.org/10.1080/13658816.2022.2087222, 2022.

Gao, J., Ji, W., Chang, F., Han, S., Wei, B., Liu, Z., and Wang, Y.: A Systematic Survey of General Sparse Matrix-matrix Multiplication, ACM Comput. Surv., 55, 1–36, https://doi.org/10.1145/3571157, 2023.

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M.: HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere, The Astrophysical Journal, 622, 759, https://doi.org/10.1086/427976, 2005.

Hojati, M., Robertson, C., Roberts, S., and Chaudhuri, C.: GIScience research challenges for realizing discrete global grid systems as a Digital Earth, Big Earth Data, 6, 358–379, https://doi.org/10.1080/20964471.2021.2012912, 2022.

Huang, X., Dai, J., Ben, J., Zhou, J., and Ding, J.: Bidirectional mapping between rhombic triacontahedron and icosahedral hexagonal discrete global grid systems, International Journal of Digital Earth, 17, 2324 952, https://doi.org/10.1080/17538947.2024.2324952, 2024a.

Huang, X., Ding, J., Ben, J., Zhou, J., Liang, Q., and Dai, J.: Advancing digital earth modeling: Hexagonal multi-structural elements in icosahedral DGGS for enhanced geospatial data processing, Environmental Modelling & Software, 172, 105 922, https://doi.org/10.1016/j.envsoft.2023.105922, 2024b.

Khan, J., E, S., and Ramesh, R.: Performance Evaluation of Geospatial Images based on Zarr and Tiff, https://doi.org/10.48550/ARXIV.2411.11291, version Number: 1, 2024.

Kmoch, A., Matsibora, O., Vasilyev, I., and Uuemaa, E.: Applied open-source Discrete Global Grid Systems, AGILE GIScience Ser., 3, 1–6, https://doi.org/10.5194/agile-giss-3-41-2022, 2022a.

Kmoch, A., Vasilyev, I., Virro, H., and Uuemaa, E.: Area and shape distortions in open-source discrete global grid systems, Big Earth Data, 6, 256–275, https://doi.org/10.1080/20964471.2022.2094926, 2022b.

Kmoch, A., Bovy, B., Magin, J., Abernathey, R., Coca-Castro, A., Strobl, P., Fouilloux, A., Loos, D., Uuemaa, E., Chan, W. T., Delouis, J.-M., and Odaka, T.: XDGGS: A community-developed Xarray package to support planetary DGGS data cube computations, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4-W12-2024, 75–80, https://doi.org/10.5194/isprs-archives-XLVIII-4-W12-2024-75-2024, conference Name: ISPRS ICWG IV/III/II
FOSS4G (Free and Open Source Software for Geospatial) Europe 2024 – Academic Track - 1–7 July 2024, Tartu, Estonia Publisher: Copernicus GmbH, 2024.

Kmoch, A., Sahr, K., Chan, W. T., and Uuemaa, E.: IGEO7: A new hierarchically indexed hexagonal equal-area discrete global grid system, AGILE GIScience Ser., 6, 1–6, https://doi.org/10.5194/agile-giss-6-32-2025, 2025.

Law, R. M. and Ardo, J.: Using a discrete global grid system for a scalable, interoperable, and reproducible system of land-use mapping, Big Earth Data, 9, 29–46, https://doi.org/10.1080/20964471.2024.2429847, 2025.

Li, D., Shen, X., and Wang, L.: Connected Geomatics in the big data era, International Journal of Digital Earth, 11, 139–153, https://doi.org/10.1080/17538947.2017.1311953, 2018.

Li, M. and Stefanakis, E.: Geospatial Operations of Discrete Global Grid Systems—a Comparison with Traditional GIS, J geovis spat anal, 4, 26, https://doi.org/10.1007/s41651-020-00066-3, 2020.

Li, M., McGrath, H., and Stefanakis, E.: Multi-resolution topographic analysis in hexagonal Discrete Global Grid Systems, International Journal of Applied Earth Observation and Geoinformation, 113, 102 985, https://doi.org/10.1016/j.jag.2022.102985, 2022a.

Li, M., Tousignant, C., Chaudhuri, C., and Chabbouh, A.: Utilizing serverless framework for dynamic visualization and operations in geospatial applications, International Journal of Digital Earth, 17, 2392 835, https://doi.org/10.1080/17538947.2024.2392835, 2024.

Li, X., Li, J., Tong, X., Li, S., and Fan, H.: The employment of quasi-hexagonal grids in spherical harmonic analysis and synthesis for the earth's gravity field, Journal of Geodesy, 96, 89, https://doi.org/10.1007/s00190-022-01653-6, 2022b.

Lu, M., Appel, M., and Pebesma, E.: Multidimensional Arrays for Analysing Geoscientific Data, IJGI, 7, 313, https://doi.org/10.3390/ijgi7080313, 2018.

Ma, X.: High-resolution image compression algorithms in remote sensing imaging, Displays, 79, 102 462, https://doi.org/10.1016/j.displa.2023.102462, 2023.

Mahdavi-Amiri, A., Alderson, T., and Samavati, F.: A Survey of Digital Earth, Computers & Graphics, 53, 95–117, https://doi.org/10.1016/j.cag.2015.08.005, 2015.

Munteanu, A.: Data Cubes and Cloud-Native Environments for Earth Observation: An Overview, SCPE, 25, https://doi.org/10.12694/scpe.v25i6.4999, 2024.

Open Geospatial Consortium: Topic 21: Discrete Global Grid Systems Abstract Specification, https://portal.opengeospatial.org/files/15-104r5.pdf, last accessed Feb 18, 2024, 2017.

Purss, M. B., Peterson, P. R., Strobl, P., Dow, C., Sabeur, Z. A., Gibb, R. G., and Ben, J.: Datacubes: A Discrete Global Grid Systems Perspective, Cartographica, 54, 63–71, https://doi.org/10.3138/cart.54.1.2018-0017, 2019.

Raveendran, N. and Sofronov, G.: A Markov Chain Monte Carlo Algorithm for Spatial Segmentation, Information, 12, 58, https://doi.org/10.3390/info12020058, 2021.

Robertson, C., Chaudhuri, C., Hojati, M., and Roberts, S. A.: An integrated environmental analytics system (IDEAS) based on a DGGS, ISPRS Journal of Photogrammetry and Remote Sensing, 162, 214–228, https://doi.org/10.1016/j.isprsjprs.2020.02.009, 2020.

Salgues, G., Cadau, E. G., Pessiot, L., Gaudissart, V., Enache, S., Gascon, F., Boccia, V., and Strobl, P.: A Candidate DGGS (Discrete Global Grid System) for Sentinel-2: First Outcomes, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, pp. 4978–4981, IEEE, Pasadena, CA, USA, https://doi.org/10.1109/IGARSS52108.2023.10281749, 2023.

Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., and Faloutsos, C.: Tensor Decomposition for Signal Processing and Machine Learning, IEEE Trans. Signal Process., 65, 3551–3582, https://doi.org/10.1109/TSP.2017.2690524, 2017.

Sousa, L. M. d.: Remaining gaps in open source software for Big Spatial Data, PeerJ Preprints, pp. 1–8, https://doi.org/10.7287/peerj.preprints.27215v2, 2018.

Temenos, A., Temenos, N., Tzortzis, I. N., Rallis, I., Doulamis, A., and Doulamis, N.: C2A-DC: A context-aware adaptive data cube framework for environmental monitoring and climate change crisis management, Remote Sensing Applications: Society and Environment, 34, 101 171, https://doi.org/10.1016/j.rsase.2024.101171, 2024.

Thompson, J. A., Brodzik, M. J., Silverstein, K. A. T., Hurley, M. A., and Carlson, N. L.: EASE-DGGS: a hybrid discrete global grid system for Earth sciences, Big Earth Data, 6, 340–357, https://doi.org/10.1080/20964471.2021.2017539, 2022.

Tian, R., Zhai, H., Zhang, W., Wang, F., and Guan, Y.: A Survey of Spatio-Temporal Big Data Indexing Methods in Distributed Environment, IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, 15, 4132–4155, https://doi.org/10.1109/JSTARS.2022.3175657, 2022.

Uher, V., Gajdoš, P., Snášel, V., Lai, Y.-C., and Radecký, M.: Hierarchical Hexagonal Clustering and Indexing, Symmetry, 11, 731, https://doi.org/10.3390/sym11060731, 2019.

University of Calgary and Stefanakis, E.: Hierarchical Data Models, GIS&T BoK, 2023, https://doi.org/10.22224/gistbok/2023.4.3, 2023.

Vu, V. T., Pettersson, M. I., Batra, A., and Kaiser, T.: Fourier Transform of SAR Data Cube and 3-D Range Migration Algorithm, IEEE Trans. Aerosp. Electron. Syst., 58, 2584–2591, https://doi.org/10.1109/TAES.2021.3134724, 2022.

Wagner, S., Stenzel, F., Krueger, T., and De Wiljes, J.: Drivers of global irrigation expansion: the role of discrete global grid choice, Hydrol. Earth Syst. Sci., 28, 5049–5068, https://doi.org/10.5194/hess-28-5049-2024, 2024.

Wang, Z., Ferreira, N., Wei, Y., Bhaskar, A. S., and Scheidegger, C.: Gaussian Cubes: Real-Time Modeling for Visual Exploration of Large Multidimensional Datasets, IEEE Trans. Visual. Comput. Graphics, 23, 681–690, https://doi.org/10.1109/TVCG.2016.2598694, 2017.

Wu, J., Gan, W., Chao, H.-C., and Yu, P. S.: Geospatial Big Data: Survey and Challenges, IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, 17, 17 007–17 020, https://doi.org/10.1109/JSTARS.2024.3438376, 2024.

Xi, W., Du, S., Wang, Y.-C., and Zhang, X.: A spatiotemporal cube model for analyzing satellite image time series: Application to land-cover mapping and change detection, Remote Sensing of Environment, 231, 111 212, https://doi.org/10.1016/j.rse.2019.111212, 2019.

Xu, D., Ma, Y., Yan, J., Liu, P., and Chen, L.: Spatial-feature data cube for spatiotemporal remote sensing data processing and analysis, Computing, 102, 1447–1461, https://doi.org/10.1007/s00607-018-0681-y, 2020.

Yang, Y., Yuan, G., and Li, J.: SFFNet: A Wavelet-Based Spatial and Frequency Domain Fusion Network for Remote Sensing Segmentation, IEEE Trans. Geosci. Remote Sensing, 62, 1–17, https://doi.org/10.1109/TGRS.2024.3427370, 2024.

Yao, X., Li, G., Xia, J., Ben, J., Cao, Q., Zhao, L., Ma, Y., Zhang, L., and Zhu, D.: Enabling the Big Earth Observation Data via Cloud Computing and DGGS: Opportunities and Challenges, Remote Sensing, 12, 62, https://doi.org/10.3390/rs12010062, 2019.

Yue, P., Wang, K., Xu, H., Gong, J., and Xiang, L.: From Geospatial Data Cube to AI Cube: the Open Geospatial Engine (OGE) Approach, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., X-4-2024, 441–446, https://doi.org/10.5194/isprs-annals-X-4-2024-441-2024, 2024.

Zhu, L., Su, X., and Tai, X.: A High-Dimensional Indexing Model for Multi-Source Remote Sensing Big Data, Remote Sensing, 13, 1314, https://doi.org/10.3390/rs13071314, 2021.

Appendix A: Key tools for DGGS-based data cube architectures.

This appendix provides an overview of key software tools and frameworks that support or are compatible with DGGS-based data cube architectures.

Table 1. Key tools for DGGS-based data cube architectures.

Tool / Framework	Timeline / Version	Language	Type	Key Feature	Illustrative Applications
H3	2018 / v4.1 (2024)	C / Py / JS	Indexing Library	Hexagonal hierarchical grid (Aperture 7)	Spatial indexing; maritime risk; land-use.
DGGRID	1999 / v8.0 (2023)	C / Python	Grid Generation	Equal-area DGGS generation (ISEA, IGEOT)	Custom DGGS construction and infrastructure.
rHEALPix	2018 / OGC Std.	C++ / Py	Native Grid System	Interrupted equal-area HEALPix projection	Earth observation; global climate modelling.
Open Data Cube (ODC)	2017 / v1.8 (2024)	Python	Data Management	ARD management and STAC indexing	Satellite time series; IDEAS framework.
Zarr	2017 / v3 (2024)	Py / C / JS	Storage Format	Cloud-optimized, chunked n-D arrays	Scalable storage for n-D satellite data.
GeoParquet	2021 / v1.1 (2024)	Spec.	Storage Format	OGC columnar geospatial vector format	Vector data integration; land-use mapping.
GDAL	1998 / v3.10 (2024)	C++ / Py	Geospatial Library	Multi-format abstraction and conversion	Data ingestion and format translation.
xarray	2014 / v2024.11	Python	Library	Labeled n-D arrays with Dask integration	Array-based analysis; general cube analytics.
PyTorch	2016 / v2.6 (2025)	Py / C++	ML Framework	Tensor ops; TorchCompile optimization	Deep Learning on SITS; optimized kernels.
TensorFlow	2015 / v2.18 (2024)	Py / C++	ML Framework	Tensor ops on GPU/TPU; Keras 3	EO data classification; compute graphs.
Apache Spark	2014 / v3.5 (2024)	Scala / Py	Distributed Computing	Massive parallel processing (H3 via HexTile)	Scalable data cube processing; maritime risk.
Google Earth Engine	2010 / 2015 (Public)	JS / Py	Cloud Platform	Multi-petabyte EO data catalog and analysis	Environmental monitoring; global change.
C2A-DC	2024 / Initial Rel.	Python	Framework	Context-aware adaptive and multiscale cubes	Climate change monitoring and modelling.