

The PlanetRisk Discrete Global Grid System

Kevin Sahr¹, Mark Dumas², and Neal Choudhuri²

¹Department of Computer Science, Southern Oregon University, Ashland, OR 97520

²PlanetRisk, Inc., 3 Bethesda Center Metro, Suite 508, Bethesda, MD 20814

Introduction

The efficiency of big data processing using modern massively distributed computing architectures is fundamentally limited by the efficiency of the algorithm used to shard the data – to partition that data for distributed storage and processing on multiple computing nodes. Efficiently sharding big data based on geospatial location has proven problematic. The most common geospatial location representations — vectors of floating point numbers representing latitude/longitude or planar map coordinates — do not naturally provide the global integer location index keys required for efficient sharding. The most common geospatial data structures, such as the quadtree and R-tree, have notoriously bad worst-case performance [Kanth et al. 1999; Arge et al. 2004], a serious concern when confronted with truly massive data sets. And because all of these systems are based on the pre-computer paper-based mapping paradigm of rectangular planar regions representing portions of the topologically spherical globe, attempting to use any of these representations on truly global big data sets exposes their intrinsic geometric weaknesses, which include areas of extreme distortion, singularities, and data under/over-sampling.

Efficient real-time analysis of multi-source global geospatial big data requires a common, consistent, multi-resolution geospatial reference frame, which naturally partitions the spherical earth into regular, integer-indexable, shards. In this paper we will systematically evaluate the available design options in order to develop the most advanced such system possible.

System Design

In order to address the issues with traditional approaches to location representation in next generation geospatial systems, researchers have recently focused on location representations based on the multi-resolution recursive partition of spherical polyhedra, called Discrete Global Grid Systems (DGGS) [Goodchild 2010; Goodchild et al. 2012]. Designing a specific DGGS requires making a series of basic design choices [Sahr et al. 2003]. Because DGGS's are the most advanced geospatial discretization approach available to us, we choose to use a DGGS approach to designing the system.

In designing a DGGS, a base polyhedron must first be chosen on which the grid will be constructed. This choice establishes the underlying symmetries of the system. We choose the icosahedron as our base polyhedron, because it has the smallest face size amongst the platonic solids and hence the smallest distortion amongst equivalent face-centered polyhedral projections.

Second, we must choose a method for partitioning the surface of the icosahedron. Researchers in numerous fields (see the survey in [Sahr, 2011]) have consistently concluded that hexagons are the optimal choice for discrete gridding and location representation. The superiority of hexagons

for all forms of discrete location representation stems from their unique geometric properties: amongst the three regular polygons that tile the plane (triangles, squares, and hexagons), hexagons are the most compact, they quantize the plane with the smallest average error [Conway and Sloane 1998], and they provide the greatest angular resolution [Golay 1969]. Unlike square and triangle grids, hexagon grids display uniform adjacency; each cell in a grid of hexagons has six neighbors, all of which share an edge with it, and all of which have centers exactly the same distance away from its center. We therefore choose to use a grid of hexagons. Note that the sphere cannot be fully tiled with hexagons; any hexagonal tiling using an icosahedral basis will always have 12 pentagonal grid cells, one centered at each of the icosahedron's vertices.

The third basic DGGs design choice specifies the orientation of the icosahedron relative to the earth's surface. We choose the standard Terra Cognita orientation, which is symmetrical about the equator and places only one icosahedral vertex/pentagonal cell on land, the minimum number known for a symmetrical orientation (see Figure 1).



Figure 1. A spherical icosahedron with standard Terra Cognita orientation.

Next we choose a method for transforming between the surface of the earth and the surface of the icosahedron upon which the grid of hexagons is constructed. Equal area cells greatly simplify statistical data analysis, and therefore we choose as our transformation the only known equal area icosahedral DGGs projection: the Icosahedral Snyder Equal Area (ISEA) projection [Snyder, 1992]. We further specify that cell center points will be calculated in the ISEA projection space.

The final design choice we must make is to choose a recursive method for constructing multiple resolutions of hexagon grids, ideally one in which the grids at different resolutions have a regular hierarchical relationship that can be exploited during analysis. Such grids are often constructed by specifying a sequence of apertures, or ratios between cell areas at a given grid resolution and the next coarser resolution, with research primarily focusing on the three “central place” apertures 3, 4, and 7 [Sahr 2003] (see Figure 2).

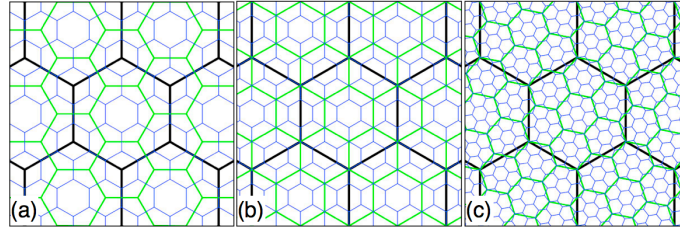


Figure 2. Three resolutions of three multi-resolution hexagon grid apertures: (a) aperture 3, (b) aperture 4, and (c) aperture 7.

In the case of an icosahedral/hexagonal DGGs the sequence begins with 12 fixed-size pentagons (forming a spherical dodecahedron), usually designated as resolution 0. A chosen sequence of apertures then determines the cell area and intercell-spacing at each coarser grid resolution. It would be desirable to have one grid resolution with cells as close as possible to 1 km^2 in area, since that is both a generally familiar size as well as a reasonable base granularity for analysis. We evaluated all possible combinations of apertures 3, 4, and 7, and found that a sequence of eleven resolutions containing three aperture 3's, two aperture 4's, and six aperture 7's creates a grid of hexagons with an area of 1.002464 km^2 , the closest it is possible to achieve to 1 km^2 in area with this approach.

Any sequence consisting of the required mix of apertures will achieve our 1 km^2 cell area, but we want to arrange their ordering for optimal value. We chose to do an initial aperture 4 in order to center cells on the north and south pole, which lie on the mid-points of icosahedron edges in the Terra Cognita orientation (see Figure 3).

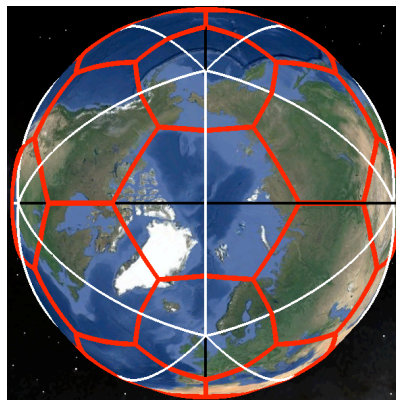


Figure 3. Cell boundaries for the first grid resolution overlaid on base icosahedron. Note the cell centered on the North Pole.

We choose to order the remaining apertures from smallest to largest, minimizing the number of cells in the coarser resolution of the DGGs to enable efficient coarse filtering operations. We also specify that additional grid resolutions, beyond our 1 km^2 resolution 11, should have an aperture of 7. Ordering the apertures as we do also has the effect of making the system geometrically a pure aperture 7 system at the finest resolutions; it should be noted that, of the three central place apertures, aperture 7 is geometrically the most efficient choice for sharding,

since it forms unambiguous, balanced hierarchies that are optimally compact and “hexagon like” across resolutions.

We have now completed the design choices needed to specify our DGGS. Table 1 gives some characteristics of the first 12 resolutions of the system, and Figure 4 illustrates the cell boundaries of one grid resolution.

res	aperture	#cells	hex area (km ²)	inter-node spacing (km)			
				min	max	mean	std dev
0	N/A	12	N/A	7,053.6500	7,053.6500	7,053.6500	0.0000
1	4	42	12,751,646.9098	3,526.8300	4,003.0200	3,764.9200	238.5950
2	3	122	4,250,548.9699	2,050.2700	2,445.8000	2,210.5000	134.6020
3	3	362	1,416,849.6566	1,143.1000	1,360.3800	1,280.7700	76.5518
4	3	1,082	472,283.2189	680.8020	835.8620	742.0270	44.6220
5	4	4,322	118,070.8047	340.2800	420.1550	371.2960	22.3994
6	7	30,242	16,867.2578	121.5960	157.3700	140.3000	8.5704
7	7	211,682	2,409.6083	48.6057	60.7706	53.0575	3.2118
8	7	1,481,762	344.2298	17.2579	22.7613	20.0503	1.2484
9	7	10,372,322	49.1757	6.9437	8.7313	7.5797	0.4589
10	7	72,606,242	7.0251	2.4609	3.2634	2.8645	0.1789
11	7	508,243,682	1.0036	0.9919	1.2494	1.0828	0.0656
12	7	3,557,705,762	0.1434	0.3514	0.4667	0.4092	0.0256

Table 1. The first 12 resolutions of the PlanetRisk DGGS.

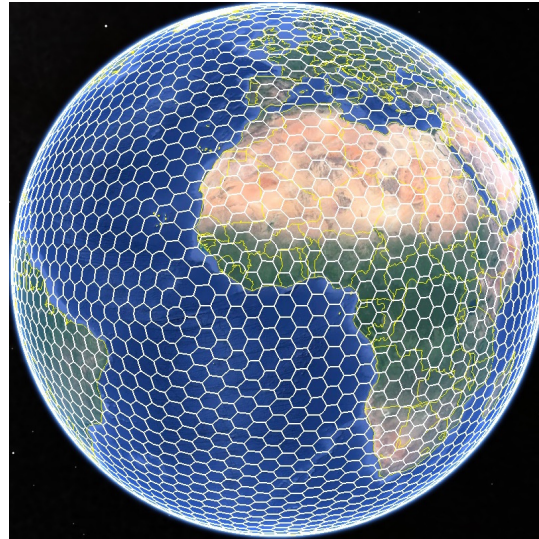


Figure 4. Grid cell boundaries for resolution 5 of the PlanetRisk DGGS.

Conclusions

In this paper we have developed a discrete location reference system for global geospatial big data analysis, optimized for efficient location-based sharding. We evaluated the most advanced design options available to us, making optimal choices at each step, to develop what we believe to be the most advanced such system currently known: the PlanetRisk Discrete Global Grid System (DGGS).

References

- L. Arge, M. de Berg, H. J. Haverkort, and K. Yi. The priority R- tree: A practically efficient and worst-case optimal R-tree. In *Proceedings of ACM Management of Data (SIGMOD)*, pages 347–358, 2004.
- Conway, J. H., and N. J. A. Sloane. 1998. *Sphere packings, lattices, and groups*. New York, New York: Springer-Verlag. 679p.
- Golay, J.E. 1969. Hexagonal parallel pattern transformations. *IEEE Transactions on Computers* C-18(8): 733-9.
- Goodchild, M.F., 2010. Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science*, 2010(1), pp. 3-20.
- Goodchild, M.F., Guo, H., Annoni, A., Bian, L., de Bie, K., Campbell, F., Craglia, M., Ehlers, M., van Genderen, J., Jackson, D., Lewis, A.J., Pesaresi, M., Remetey-Fülöpp, G., Simpson, R., Skidmore, A., Wang, C., Woodgate, P., 2012. Next-generation digital earth. *Proceedings of the National Academy of Sciences of the United States of America*, 109(28), pp. 11088–11094.
- K. V. R. Kanth and A. K. Singh. Optimal dynamic range searching in non-replicating index structures. In *Proceedings of International Conference on Database Theory (ICDT)*, pages 257–276, 1999.
- Sahr K, White D, Kimerling AJ. 2003. Geodesic discrete global grid systems. *Cartography and Geographic Information Science* 30(2):121-134.
- Sahr, K. 2011. Hexagonal discrete global grid systems for geospatial computing. *Archives of Photogrammetry, Cartography and Remote Sensing*, 22:363-376.
- Snyder, J. P. 1992. An equal-area map projection for polyhedral globes. *Cartographica* 29(1): 10-21.