

SPATIAL BIG DATA ORGANIZATION, ACCESS AND VISUALIZATION WITH ESSG

Lixin WU^{a, b, c}, Jieqing Yu^a, Yizhou Yang^d, Yongji Jia^a

^aSchool of Environment Science and Spatial Informatics, **China University of Mining and Technology, Xuzhou, China** (awulixin@263.net)

^bAcademy of Disaster Reduction and Emergency Management, Beijing Normal University, Beijing, China

^cIoT (Mine Perception) Center, China University of Mining and Technology, Xuzhou, China

^dInstitute of Geo-informatics and Digital Mine Research, Northeastern University, Shenyang, China

Commission: IV-WG5

KEY WORDS: Spatial big data, Global Spatial Grid, ESSG, Spatial Infrastructure, Visualization, Mapping, Internet/Web, Cloud

ABSTRACT:

There are hundreds of spatial reference frame (SRF) being applied, and the great difference among SRFs has blocked the share of global data on planet Earth. A conceptual spheroid of radius 12,800km and a spheroid degenerated octree grid method are applied to produce an earth system spatial grid (ESSG), which is of natural characteristics to be applied as a new common SRF. A triple CTA is designed as ESSG-based data structure to organize the big data of planet Earth, and a 2D table of a unique label and limitless records for time slices and attribute values is present to record the data of each grid. The big data on planet Earth can hence be gridded and interrelated without discipline gaps and SRF obstacles. An integral data organization mode is designed, and three potential routes are presented for users to access shareable global data in cloud environment. Furthermore, with global crust, atmosphere, DEM, and satellite image being examples, the integrated visualization of global large objects is demonstrated.

1. INTRODUCTION

The big data on our planet Earth is the common property of human beings to recognize the mother planet Earth^[1]. There are 1.2 zetabytes of electronic data produced each year, and 80% about of current data are spatial data, which comprise of the main part of big data. Spatial data are location-based, which is the fundamental data and knowledge source for people to recognize the real world. With the rapid development of earth observations, the global spatial data increases sharply in resolution, variety and volume. For example, the stored data in NASA had reached 5 petabytes and increase 4.5 terabytes per day. The planet Earth is a comprehensive, highly structured ellipsoid in that its core, mantle, lithosphere, coversphere and atmosphere are orderly organized from its interior to surface, and to exterior. Hence, the data sets produced on planet Earth should be recorded, organized and represented referring to the geometric shape and spherical structure of the planet Earth for better understanding, easy access and more applications.

A lot of works on the share of scientific data had been conducted since the end of last century, which has facilitated effectively the collecting, storing, managing, exchanging and sharing of geoscience data^[2]. Among them the most important works include: 1) the World Data Center (WDC) set up in 1957 and the Committee on Data for Science and Technology (CODATA) set up in 1966, which are under the umbrella of International Council of Science Union (ICSU), 2) the International Oceanographic Data and Information Exchange (IODE) and the data exchange system of World Meteorological Organization (WMO), which are under the umbrella of United Nations Educational, Scientific and Cultural Organization (UNESCO), and 3) the DataCORE (Data Collection of Open Resources for Everyone) plan initiated in the GEO Work Plan 2012-2015 (<http://www.earthobservations.org/>), which is a GEO Common Infrastructure (GCI) for Global Earth Observation

System of System (GEOSS).

Due to the difference in geographical region, discipline domain, spatial scale and culture history, there are hundreds of spatial reference frames (SRFs), including space datum and coordination system, had been and are being applied for the spatial data of planet Earth. The SRFs could be classified as map projection system, Euclidean spheroid system, surface grid and spheroid grid, and each has several branches. For example, the spheroid system can be further classified by origin as geocenter system (including ECEF2000, WGS84, North American NAD83^[3], and European ETRS89^[4]) and ellipse centered coordination system (Beijing 54 and Beijing 80). The spheroid grid can be further classified by generating mechanism as center-toward radial extension system from spherical grid (including spheroid latitude-longitude-radius system^[5-7], spheroid tri-prism system^[8, 9], and cube-sphere grid^[10-12]) and surface-toward extension system from 3D Euclidean grids (including Yin-Yang grid^[13], blocking hexagon grid^[14, 15] and voxel filling grid^[16]). The abundant SRFs applied for global data together with the continuously changing geoid and north polar, have brought with the great difficult in spatial interrelation among data sets, especially among the data sets of different discipline domain, such as geophysics, topography, atmosphere and oceanology, which obstacle seriously the integral analysis to global data and to the coupling study on Earth system process.

However, current works on global data share focus on metadata and data standards, less consideration is paid to data compatibility and to space interrelation dealing with the difference in SRF and resolution. The WDC emphasizes on open consortium and metadata management to solve inconsistency on data archiving process, and to provide catalogue service on internet. The users of WDC need to understand the data release mode and to enter the data center

to download what he wanted. There are conspicuous differences in metadata, data standard and spatial reference among different data centers, and there lacks of data inter-manipulation interface for space interrelation and share service. The art of WMO is similar to that of WDC, in that WMO provides mainly data index and service for member counties with metadata. CODATA guides the collecting a, archiving and storing of global scientific data with data standards, so as to facilitate global data share.

Besides, the universal software for shared data download is much lacked. Although GoogleEarth, VirtualEarth, GeoGlobe and Tianditu have provided browse and inquiry service for global images, DEM and raster-based thematic maps, they do not provide data download and spatial analysis services. WorldWind provides data to users with pictures even if it has ability in automatic updating of satellite images and monitoring on global climate. ArcGIS Online is limited in surface data and cannot meet the demands for more geoscience researches, even if it has provides online mapping and service based on cloud technology. SkylineGlobe has built massive 3D terrain data for user editing and analysis, while it is not interrelated with other geoscience data sets.

To solve the problem on global big data share resulted from the difference among SRFs, and to develop a new software technology for global big data share access, this paper introduce a SDOG-based earth system spatial grid (ESSG), which is capable to be applied as a new global SRF for global big data interrelation. A prototype is developed based on ESSG, and the grid linkage, integrated organization, web-based access and global visualization on spatial big data are reached with a few examples.

2. INTRODUCTION TO EARTH SYSTEM SPATIAL GRID (ESSG)

2.1 Definition of conceptual spheroid

In consideration of the approximate ellipsoid shape and the actual size of planet Earth, but not limited to it, a conceptual spheroid is defined to serve as the base of the common global SRF of planet Earth. For the convenience of bi-sectional subdivision computation and in consideration of atmosphere and near-space environment, the radius of the conceptual spheroid is supposed to be 12,800km ($2 \times 6,400\text{km}$), which is double of the approximate equator radius, 6,378.137km, of the planet Earth. As in figure 1, to make the origin, reference and spherical representation of the conceptual spheroid matching with the most adopted geographical Longitude-Latitude-Geodetic Height (BLH), World Geodetic System (WGS), and Earth-Centered-Earth-Fixed (ECEF) rectangular Cartesian coordination system, we make the earth center of mass (including ocean and atmosphere, and regard "The Earth's center of mass does not move"^[17]) being the center of the conceptual spheroid, and define the spatial reference (equator plane, polar, and latitude/longitude system) according to the usual equator plane (X/Y plane), the Z-axis as defined by Conventional Terrestrial Pole (BIH1984.0), the X-axis as pointing from the origin to the intersection of Greenwich 0°-meridian line and the equator, the Y-axis as composing a right-handed 3-D XYZ system with X- and Z-axis, and the usual longitude-latitude graticule system as that defined/applied by BLH, WGS, ECEF and International Terrestrial Reference Frame (ITRF, <http://itrf.ensg.ign.fr/>).

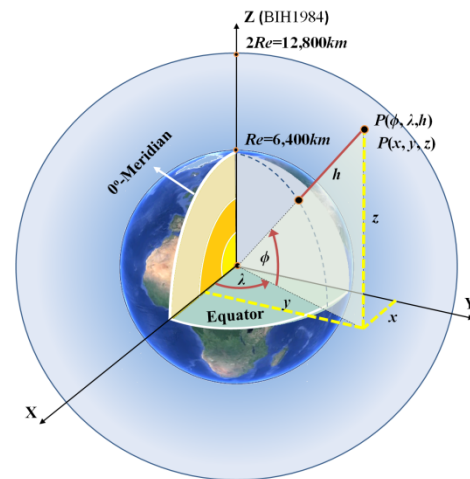


Fig. 1 The conceptual spheroid and definitions of its spatial references

2.2 Production of ESSG

Firstly, the conceptual spheroid is subdivided into eight octants with its equator plane, 0°- and 90°-meridian plane. Afterward, applying a specially designed hierarchical grid subdivision mechanism, named Spheroid Degenerated Octree Grid (SDOG)^[18], each octant is hierarchically divided into a SDOG tree of smaller and smaller sub-grids with the tree's depth increase gradually. The combination of the eight SDOG trees produces a hierarchical and spherically-structured Earth System Spatial Grid (ESSG)^[19], which is a hierarchical and multiple granularity subdivision of planet Earth including its interior, surface and exterior. The radial thickness of grid at subdivision level 10, 20, and 30 is 25km, 25m and 2.5cm about, respectively. The ESSG has excellent characteristics in spherical structure, continuous coverage, hierarchical structure, approximate size, definite frame, unique code, multiple granularities, and geographical consistency, which makes it being able to act a common global SRF.

2.3 Linking ESSG to other SRFs

A unique octal code, in linear 1-D but not 3-D as XYZ and BLH, can be generated for each grid^[20]. Since the spatial reference of SDOG-based ESSG is equivalent to that of ECEF, there exist a simple and direct conversion between ESSG code and ECEF coordination^[20]. With ECEF being bridge, coordination conversion algorithms have been developed for linking ESSG to other SRFs including geocenter coordination, map projection, spherical Discrete Global Grid (DGG) and spheroid grid. The encoding and decoding algorithm of ESSG^[20] makes ESSG being possible to link all data sets using current SRFs as in figure 2.

3. BIG DATA ORGANIZATION WITH TRIPLE CTA

3.1 ESSG-based triple data structure

The grid element in ESSG is a 3-D voxels in manifold space, which is not only a 3-D spatial element of the tessellated space of planet Earth, but also a 3-D container for all data related with the grid of corresponding location and granularity. The grid element has meanings both in space and attribute. The basic geometrical parameters, including grid

position, shape and granularity, are the chief space meanings embodied in grid code, while the other parameter values, such as mass, material, contains, temperature, and stress, are the attributes. If the grid is a crust element, the grid attributes are consisted of rock type, rock density, rock porosity, rock temperature, elastic modular, Poisson's ratio, seismic velocity and so on. If the grid is a terrain element, the grid attributes are consisted of vegetation type, vegetation index, surface

temperature, soil moisture, image texture, spectrum value, biological diversity and so on. If the grid is an ocean element, the grid attributes are consisted of water salinity, temperature, ionic concentration, biological diversity and so on. If the grid is an atmosphere element, the grid's attribute are consisted of temperature, pressure, aerosol quantity, electronic density, ionic concentration and so on.

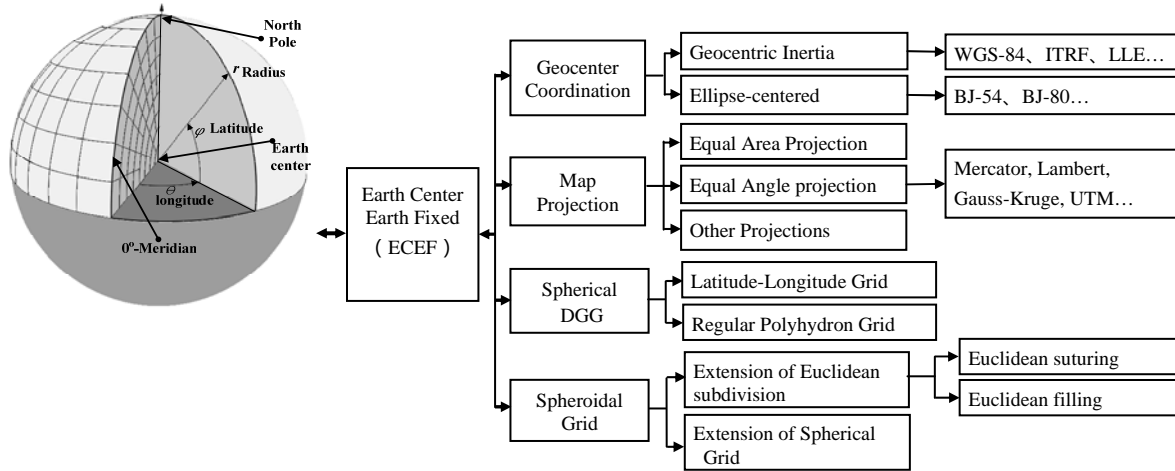


Fig. 2 Link of ESSG with other spatial reference frames

The attributes of a grid in ESSG can be obtained from manifold spatial interpolation on earth observation data sets. A time-stamp can be assigned to the attribute values, so as to get a historical record of the dynamic change the attribute. A triple (CTA) is so presented as a new data structure to organize global big data^[21,22], which take the octal code (C) of a grid as the spatial identifier, the time record (T) as a time-stamp, and the attribute values (A) related with the grid as parameter fillings. Any data sets collected with a time-stamp on anywhere of the planet Earth, can be interpolated as a set of C-T-A values and input to the ESSG at corresponding granularity. The C is the unique label of the triple, while T and A are recorded in an open 2-D table. For the self-description properties of CTA, unlimited data could be recorded for a grid and all the attributes of the grid are interrelated automatically. This provides excellent properties for grid data integration, fusion, interrelate analysis and visualized representation, which is the nature characteristics of SDOG-based ESSG to realize global big data interrelation and share service.

3.2 Big data organization

Referring to the resolutions of observation data sets and the application demands on grid size, the observation data sets can be resampled or re-gridded into ESSG at a certain granularity level, so as to get the value of spatial C and attribute A for the elements of triple CTA. The assembling of massive C and A for global attributes at various granularity is really a spatial big data. Nevertheless, the values of A are changeable, and the time axe will be limitless which lead the C-T-A to be indescribable big data. Fortunately, the C-T-A data of a grid is independent of other grids, which make C-T-A being inherent independent for distributed and cloud storage. The basic concept of ESSG-based big data organization is that each grid is a 2-D table with C being its unique label, and the distributed big data base is a massive set of 2-D tables (Figure 3). Since C is related with grid size, i.e., related with grid's subdivision level, the 2-D tables are spatially interrelated with grid size.

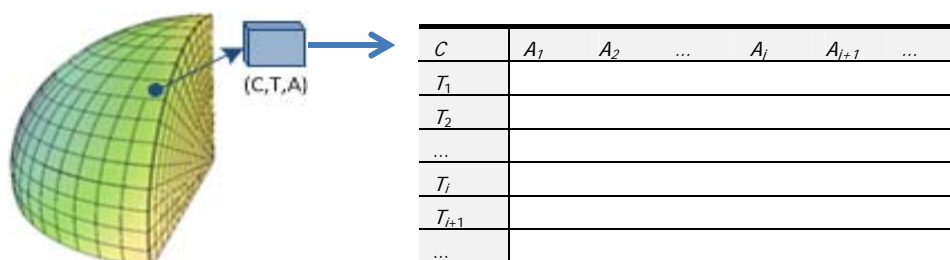


Figure 3 CTA-based big data organization

With the fundamental geometrical concepts, point, line, polygon, and body, in Geographical Information System being applied, the 3-D point, curve, surface and body in the manifold space of planet Earth can be geometrical represented by a singular grid or by a group of adjacent grids

with node, edge or face conjunctions. Hence the spatial features and non-spatial attributes of the object can be represented by the component grids of the object. We apply Oracle to organize, manage and maintain orderly the ESSG-

based *C-T-A* data, which will support quick access and easy inquiry both in *C/S* and cloud environment.

3.3 Big data access

An ESSG-based access mode is designed for global big data share (Figure 4). The basic components of ESSG-based global big data share access system are comprised of ESSG server, CTA data base and three engines being G-Link Engine, Conversion Engine and Visualization Engine, respectively. The G-Link Engine serves for grid linking of ESSG to outside sharable data sets either in data base or in data file format. Conversion Engine provide coordination

conversion between grid *C* and coordination of other SRFs, which makes outside sharable data sets can be tracked, linked and input to a Virtual Pool (V-Pool) in data structure of CTA for temporary or long-term keeping, so as to speed up the access to global big data. The Visualization Engine serves user a glance of the data sets he or she downloaded from CTA data base or from outside sharable data sets directly. It will provide also a powerful visualization service to support analysis on big volume of data, such as global image, global DEM, regional DEM, and HDF-EOS atmosphere data (Figure 4).

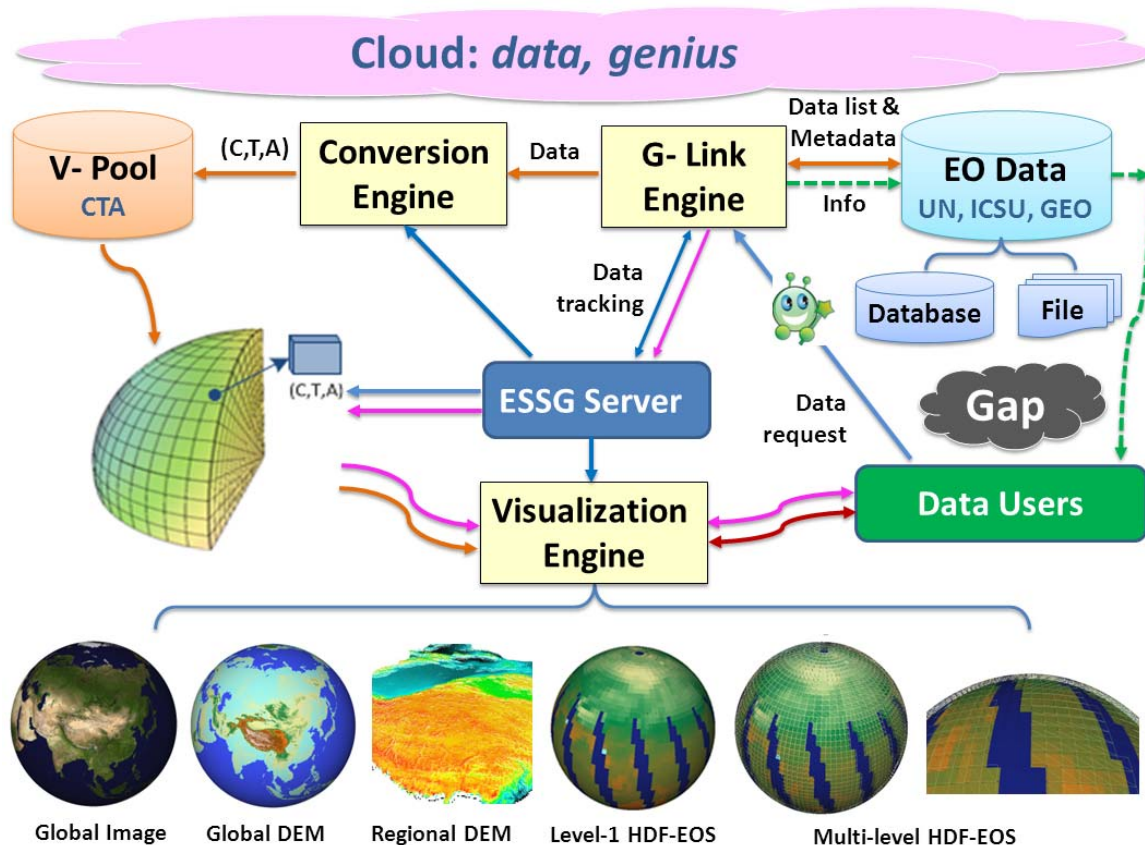


Figure 4. ESSG-based CTA organization and cloud-based access to big data on the planet Earth

When a user inquiry global data by via of G-Link engine from Web and cloud environment, the inquiry will be converted automatically to be standard requisition information (spatial range, time section, and attribute names), and then the G-Link engine embodied with many genius, as plug-ins being capable of data finding, will find the desired data from outside sharable data set in clouds, or find the desired data in CTA data base through ESSG server. Later, the user will have three routes to download data, 1) to download from outside sharable data base matched in data list and metadata (along the dashed green lines); 2) to download existed data from CTA data base after pre-visualization (along the pink lines), and 3) to download new data from CTA data base coming from V-Pool after coordination conversion (along the brown lines). Hence, the user can access and get the desired global data sets easily with genius even he or she know nothing about the data format, spatial reference and record mode of the sharable data. Thus the black gap between users and Earth observation data can be filled completely.

3. BIG DATA VISUALIZATION

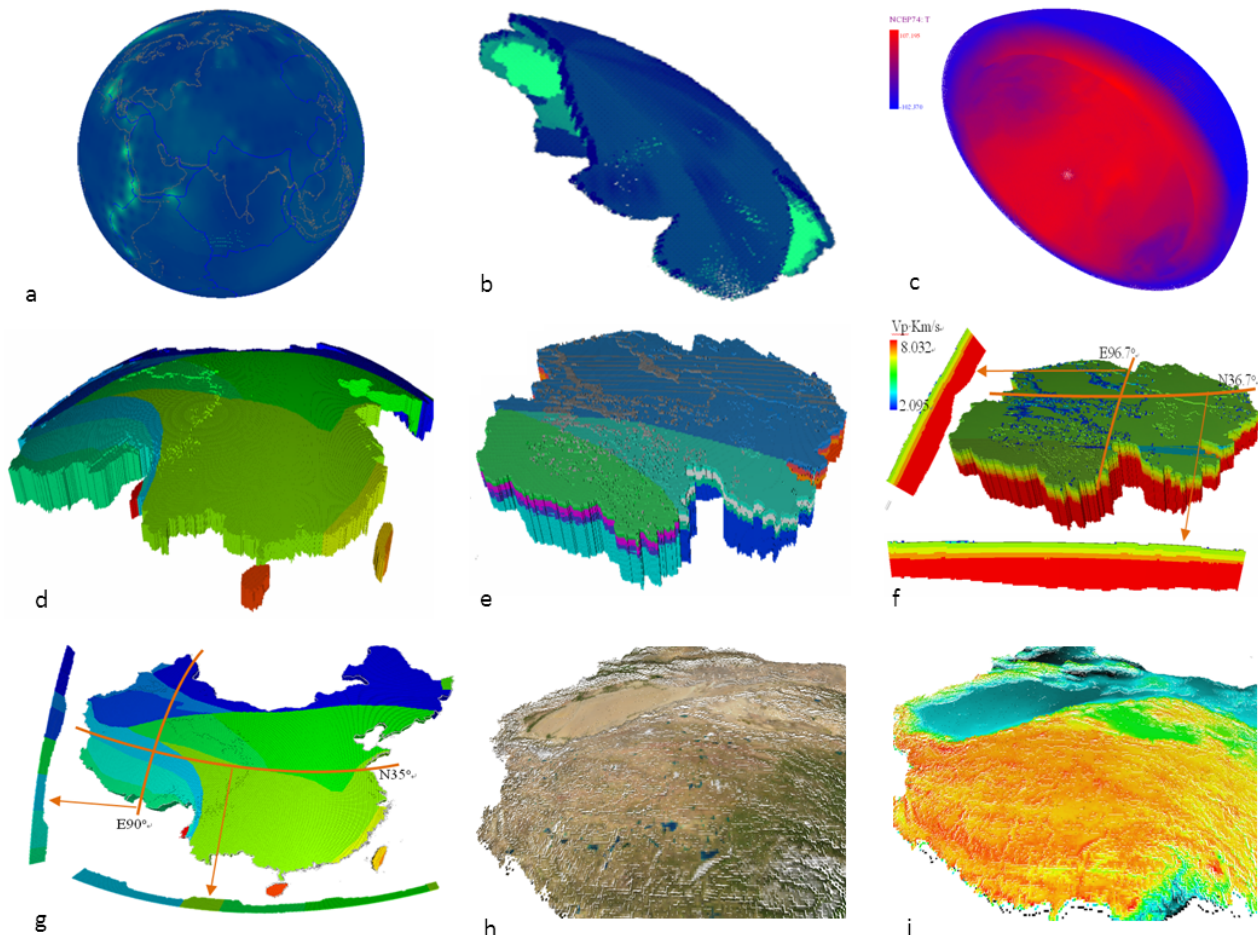
A prototype system named **GASE** (Global Data Spatial Interrelate) is developed with C++ based open source graphic library (Qt) and graphic rendering library (Coin3D), and with CUDA programming and GPU acceleration technologies. The functions of multiple hierarchical organization, dynamic dispatching, texture rendering and parallel visualization are realized for global big data.

With global atmosphere temperature data observed by the channel 4-14 of microwave sensors (AMSU-A1) embodied at NASA-17 satellite being example, the temperature data of spherical resolution $3.3^{\circ} \times 3.3^{\circ}$, recorded in strip-shaped HDF-EOS for 11 atmosphere elevations (2, 5, 12, 25, 50, 80, 180, 270, 400, 700, 1000 hpa), are gridded into grid temperature of grids at subdivision level 11 (spherical size being 12.5 Km

about). The temperature, at single elevation or all elevations, can be visualized in ESSG (Figure 4).

We have also linked global plate data, global terrain data, global crust thickness data, local seismic velocity data, local geo-layer data from WDC, United States Geological Survey (USGS) and China, global atmosphere data from National Oceanic and Atmospheric Administration (NOAA), Blue

Marble image and 30m DEM from NASA. All the data are gridded into ESSG as A_i of corresponding grid size, and the integral visualization of the data sets is shown in figure 5. Here, the crust grids have attributes in geology age, strata temperature^[23] and seismic p -velocity, while the atmosphere grids have attributes in temperature^[24,25]. It shows that the big data of planet Earth can be orderly interrelated, combined, integrated and organized with ESSG.



a-global lithosphere temperature, b-north American lithosphere temperature, c-semi-sphere atmosphere temperature
 d-China crust, e-Qinhai crust in west China, f-seismic velocity of Qinhai crust
 g-sections of China crust, h-Satellite image of west China, i-terrain DEM of west China

Figure 5. ESSG-based integral visualization of global big data on planet Earth

4. REMARKS AND FUTURE WORK

The SDOG-based ESSG and ESSG-based triple data structure CTA has overcome the bottleneck of global big data share resulted from spheres limitation of planet Earth and coordination difference among SRFs. An new common global SRF is built on ESSG, and a new global big data share mode is presented on the common SRF. Not only the interior, surface and exterior data of planet Earth can be organized integrally with the common SRF, but also any institutions and any people can contribute to global data share service. No matter where you are, you will be able to upload your scientific data on our planet Earth for share if you have a computer or a iPad, and if you can connect to the internet. In the cloud, a virtual data pool could be built for frequently accessed data sets so as to strengthen the data serving capability and to improve the access speed. In the cloud,

particular tracking tools could be developed to serve for automatic tracking and spatial interrelating of user interested data; in the cloud, a data pushing service could be developed to track the state of user interested data and push automatically the new information or data sets to the user as soon as the data sets is updated, so as so to support the user's rapid response.

Since the ESSG-based CTA data structure has unique virtues in distributed storage and parallel computing, the distributed organization, cloud management, easy exchange, free access, and integrated analysis (spatial statistics, spatial reasoning, data mining, etc.) on the big data of planet Earth will get facilitated by the rapid developing cloud technology and parallel computation. To keep the most recent value of sharable data, and to keep the recorded data in CTA data base

in accordance with the state of outside sharable data sets^[26], an automatic tracking and updating mechanism is designed and a remote updating technology is developed^[27]. The interrelated interface between ESSG-based CTA and other global data sets will be designed and the auto-tracking genius for data updating will be developed gradually.

Acknowledgement

This study is funded by National Natural Science Foundation of China (Grant No. 40930104).

Reference

1. UNITED NATIONS GLOBAL PULSE. Big Data for Development: Challenges & Opportunities[R]. New York, 2012.
<http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>. [2013-8-1]
2. Sun J.L., Lin H. Earth System Research and Scientific Data[M]. Beijing: Science Press, 2009: 538..
3. Wikimedia. North American Datum [EB/OL]. July 17, 2013.
http://en.wikipedia.org/wiki/NAD83#North_American_Datum_83_and_WGS84. [2013-8-1].
4. Wikimedia. European Terrestrial Reference System 1989 [EB/OL], July 30, 2013. URL: http://en.wikipedia.org/wiki/European_Terrestrial_Reference_System_1989. [2013-8-1].
5. Zhu J.S, Cao J.M, Li X.G. The reconstruction of preliminary three-dimensional earth's model and its implications in China and adjacent regions. *Adv Earth Sci.*, 1997,40(5): 628-648.
6. Bassin, C., G. Laske, G. Masters. The Current Limits of Resolution for Surface Wave Tomography in North America[J]. *EOS Trans AGU*, 2000, 81: F897.
7. Zhao, D. Global tomographic images of mantle plumes and subducting slabs: insight into deep Earth dynamics[J]. *Physics of the Earth and Planetary Interiors*, 2004, 146(1-2): 3-34.
8. Baumgardner, J. R. Three-dimensional treatment of convective flow in the earth's mantle[J]. *Journal of Statistical Physics*, 1985, 39(5): 501-511.
9. Ballard, S., J. R. Hipp, C. J. Young. Efficient and Accurate Calculation of Ray Theory Seismic Travel Time through Variable Resolution 3D Earth Models[J]. *Seismological Research Letters*, 2009, 80(6): 989-999.
10. Harder, H., U. Hansen. A finite-volume solution method for thermal convection and dynamo problems in spherical shells[J]. *Geophysical Journal International*, 2005, 161(2): 522-532.
11. Stemmer, K., H. Harder, U. Hansen. A new method to simulate convection with strongly temperature- and pressure-dependent viscosity in a spherical shell: Applications to the Earth's mantle[J]. *Physics of the Earth and Planetary Interiors*, 2006, 157(3-4): 223-249.
12. Tsuboi, S., Komatitsch, D., Ji, C., et al. Computations of Global Seismic Wave Propagation in Three Dimensional Earth Model[M]. In: Labarta, J., Joe, K., Sato, T. (Eds.), *High-Performance Computing*. Springer Berlin / Heidelberg, 2008, 434-443.
13. Kageyama A, Sato T. The "Yin-Yang Grid": An overset grid in spherical geometry[J]. *Geochemistry Geophysics Geosystems*, 2004, 5(9): 1-15.
14. Komatitsch, D., J. Ritsema, J. Tromp. The Spectral-Element Method, Beowulf Computing, and Global Seismology[J]. *Science*, 2002, 298(5599): 1737-1742.
15. Stadler, G., M. Gurnis, C. Burstedde. The Dynamics of Plate Tectonics and Mantle Flow: From Local to Global Scales[J]. *Science*, 2010, 329(5995): 1033-1038.
16. Braun, J., C. Thieulot, P. Fullsack. DOUAR: A new three-dimensional creeping flow numerical model for the solution of geological problems[(2008)]. *Physics of the Earth and Planetary Interiors*, 2008, 171(1-4): 76-91.
17. Earl F. Burkholder. The 3-D Global Spatial Data Model. CRC Press, p.301
18. Wu L.X, Yu J. Q. Global 3D2Grid Based on Sphere Degenerated Octree and Its Distortion Features[J]. *Geography and Geo-information Science*, 2009, 24(1):1-4
19. Wu L.X, Yu J. Q. Earth System Spatial Grid and Its Application Mode[J]. *Geography and Geo-information Science*, 2012, 28(01): 7-13.
20. Yu J. Q, Wu L.X. On Coding and Decoding for Sphere Degenerated 2-Octree Grid[J]. *Geography and Geo-information Science*, 2009, 24(1):5-9
21. Yu J.Q. SDOG-based Earth System Spatial Grid and Its Application and Three Dimensional Modeling[D]. Beijing Normal University, 2012.6: 103
22. Yu J.Q., Wu L.X., Li, Z.F., et al. A SDOG-based intrinsic method for three-dimensional modelling of large-scale spatial objects[J]. *Annals of GIS*, 2012, 18(4): 267-278
23. Yu J.Q., Wu L.X., Zi G.J., et al. SDOG-based Multi-Scale 3D Modeling and Visualization on Global Lithosphere[J]. *Science in China (D)*, 2012, 55(6): 1012-1020
24. Li Z.F., Wu L.X., Yu J.Q., et al. Parallel Visualization of Earth System Grid Data Based on Visit and its Performance Analysis[J]. *Journal of Geomatics Science and Technology*, 2012, 29(2):144-148
25. Bao H.G. Rapid visualization of global image and DEM based on SDOG-ESSG[D]. China University of Mining & Technology (Beijing), 2013.
26. Li X.J. ESSG-based sharing modes and methods for the earth system multi-scale data[D]. Beijing Normal University, 2012
27. Wang H. Geoscience sharing data auto-updating for the ESSG-based data sharing server[D]. China University of Mining & Technology (Beijing), 2013