Response to OGC Decision Fusion Standards RFI

Perry R. Peterson, CET B.Eng. P.Eng.

6/10/2010

Kingston | Calgary | Queensland

www.pyixisinnovation.com

Network-Centric Fusion implies fusion ready and therefore must include the process of preparing geospatial information assets so they are self synchronized at any point in their use.

Contents

1 OVERVIEW AND EXECUTIVE SUMMARY

Dear Sir,

At PYXIS we are essentially engineers looking for the answer to a significant problem – Geospatial Integration On-Demand. The problem is a deep and pervasive challenge. The solution therefore requires innovation. In that search we have discovered an elegant solution. However, that solution requires a major paradigm shift. Further, we are often asking the shift from individuals and organizations who have invested careers worth of effort supporting the status quo. This is the innovator's dilemma.

In our years of promoting and evangelizing the establishment of a discrete global grid persistence and patience have garnered preeminent validation. I recall for some years we had to argue that the term Interoperability was not the same as Integration. Seems absurd now, but it was common to review a paper in 1997 that used the words interchangeably. The next resistance we began to hear was that "*There are many grids already. We don't need another one*". For which we began to teach the difference between the traditional analog lattice grid and a discrete tessellation that most digital models are founded on. We had to show that gridding is always a part of the process of data integration, whether implicit or explicit, even with vectors. Some argued that we would have to change all of the legacy approaches to geospatial encoding – for which we demonstrated lossless conversion, statistical sampling, and provided the common practice used with a graphics processor that pixelates all information as a direct analogy to our approach. It's not standards based we'd hear – and so we built processes closely linked to OGC and other standards to demonstrate remarkable, unique, and required, capabilities. Then there is the argument that a discrete global grid is not a standard – we agree that it should be – we require a standard for discrete global grids.

Recently a new argument as surfaced – gridding should be done at the end of the process not at the beginning:

> *"Geospatial and temporal registration of observations is necessary for sensor fusion. There are several approaches on how to achieve such registration including consideration of reformatting in a distributed environment. PYXIS responded to the Fusion RFI promoting the use of a multi-resolution, equal-area, discrete global grid for aiding in the fusion process. Particularly focused on a hexagonal tessellation of the Earth called the optimized Icosahedral Snyder Equal Area aperture 3 Hexagon Grid (ISEA3H). With regard to sensor fusion, PYXIS advocates resampling sensor data to ISEA3H grid to aid in fusion and application of grid specific algorithms for image processing, pixel-to-pixel operations, and fusion. While the ISEA3H tessellation work is impressive and may be useful for an individual fusion algorithm; standardizing data exchange formats to a particular grid is too brittle of an approach for a distributed system. A more robust process is to standardize how a dataset describes its gridding thereby allowing downstream fusion algorithms to re-grid if necessary as late as possible. Such an approach – regrid as little and as late as possible – is supported by the existing SWE* [Sensor Web Enablement] *standards."* (OGC 09-138, 2010)

I strongly agree with this statement within the context of current processes and with a view that the analysis and modeling is the object of the process not the distribution and use of data. The rigors and labor of assembling massive disparate data sources to support analysis and modeling is at the core of what geomaticians do. However, this process idealizes and supports the decision-making to a select few with the wherewithal. We should consider this the top slice of the potential decision-makers pyramid.

GEOINT2, I believe, is something far more powerful. It is about putting the decision making tools into the hand of the decision-makers at the edge of the battle. This is the new world that is distributed, ubiquitous, participatory, and netcentric. It is a pull model. It is rapid. There is no time to assemble data in a fusion engine. There must be a process parallel to the traditional that allows data integration on-demand. And the best way is to present a standard for an optimized discrete global grid. This single effort will fuel a new age of evidence-based decision-making and enable the GEOINT2 cycle.

We would ask you to consider carefully the immense value that would be unlocked to the world if we can generalize these GEOINT2 capabilities and drive them lower on the pyramid. On-Demand Geospatial Integration IS the grand challenge and the bottleneck that is holding it up. There is no conflict by supporting the adage "*regrid as little and as late as possible"* and ALSO *"process your data into a standard Discrete Global Grid"* as a product for those who want it. It can be/should be a parallel process. This approach is supported heavily by DOD netcentric directives and doctrine.

In the 1990's there were a series of cancer studies sponsored by the Clinton administration. I recall it as a "war on cancer". Significant research was completed, included geospatial work. One study looked at the correlation between cervical cancer and industrial pollution sources. We know now with some certainty that the risk of cervical cancer can be linked to the human papillomavirus. However, at that time we collected massive amounts of geodata to support the industrial pollution hypothesis – $9 million of the $10 million of this particular research budget went to data collection and preparation. Data was fused, models were run, conclusions were drawn, and papers were published. Some years later I made a request for the data to support further investigation but was told, besides the tabular results that supported the conclusion, none of the raw data could be found. Why?? I was told there was no "medium".

This is our point. This is the paradigm shift. There is no medium. Imagine a medium where Earth observations and interpretations can be discovered and shared as seamlessly as any information on the Web. The grid is the medium.

Why pre-process to a discrete global grid as a process in GEOINT2:

1. Global – An optimized discrete global grid is a global reference model, and thus it can address the entire planet equally – including the poles. As it is built upon a multi resolution grid it can be used to represent arbitrarily sized locations, from continents to birdbaths. It can make meaningful use of data stored at different resolutions, without the need to conflate between scales.
2. Data Agnostic – the discrete global grid enables rapid integration of multiple datasets, regardless of scale, origin, spatial resolution, format, datum, projection, and temporality. It harmonizes vectors and raster imagery. It provides applications consistent, data independent, geospatial reference model for integration. The chief result will be the provision of Integrated Geospatial Intelligence On-Demand.
3. Fast – Gridding or reprojecting data takes a considerable amount of computer resources. When dealing in GBytes and TBytes of data, processing data can be measured in months. Efficiencies are gained and duplication is avoided when data is processed as one product at the source. Using data that is prepared on a discrete global grid is VERY VERY FAST – online fast.

4. Accessible – Geospatial data comes in numerous forms, scales, sizes, and formats. Due to this complexity data is generally locked up in silos inaccessible. A discrete global grid allows ALL forms of geospatial data to be resolved to a simple cell containing attributions of that location. Access is ubiquitous and easy across multi spatial scales and massive data sizes.
5. Searchable – Analogue coordinates cannot be searched unless pre-federated into tiles/trees with spatial indices. Spatial indexing is generally done by an integrator at the application end, not the geospatial-intelligence end, of the spectrum. Source data is currently, at best, catalogued. However, in a discrete global grid each cell is indexed and thus searchable very early in the process – as soon as it is georeferenced – Level 2 – a global grid product can be produced. The mere act of gridding to a standard aligned discrete global grid federates the data with all other disparate data that has also been gridded, providing distributed search capabilities. Further, this indexing can be harmonized with semantic search creating true location based searching synchronizing it with the digital indexing of other MultiInt assets including IP network addressing promising powerful cyber-spatial capabilities.
6. Inquiry - The process of gridding promotes location from an attribute of a data record to a high level object in the record – it becomes an object not merely an attribute (consider the conventional database design where location is the attribute of an aircraft as opposed a discrete global grid design where an aircraft is passing through a cell location.) This is a significant turning of the paradigm – location can be the thing. Further, Filters, Queries, Topology, Spatial Analysis, Image Processing on multisource of data are generally simplified to set theory, database operations and spreadsheet calculations. Simulations and other models fit naturally on the grid.
7. Maintenance – In the global grid architecture, there is only the data creator and the decision maker. The discrete global grid drives the maintenance process back to the source creator, where single sources of data can be created, managed, distributed, secured, and archived directly by those most familiar with its properties, as opposed to a GIS integrator who is typically managing scores for data sources and secondary and tertiary products – a unsustainable process in the future world of GEOINT2.
8. Temporal – Simply add a discrete timestamp to a cell and you have the basis for critical event processing, animation, simulation, and tracking
9. Distributable – By the nature of the discrete grid, ALL data is tiled and ready for transmission in practical portions necessary in a networked environment. Data is easily compressed and encrypted. It conforms well to the distribution within a peer-to-peer network, distributed processing of cloud networks, and service chaining architecture. There tiles are spatially aligned with any other tiles produced permitting multi source integration upon access by the end-user.
10. Reuse – Products of multisource data analysis and modeling can be republished as new data content on the grid, promoting a medium for sharing results and reuse of assets.
11. Netcentric – Many of the characteristics above denote netcentric capabilities. The prime characteristic is the ability to enable a pull model, where as the end-user has the power. The discrete global grid provides the basis for such an environment and business model.

We present here a number of options for showcasing a parallel decision fusion environment where a valuable fusion ready data product is created by processed source data to a discrete global grid.

Respectfully submitted by

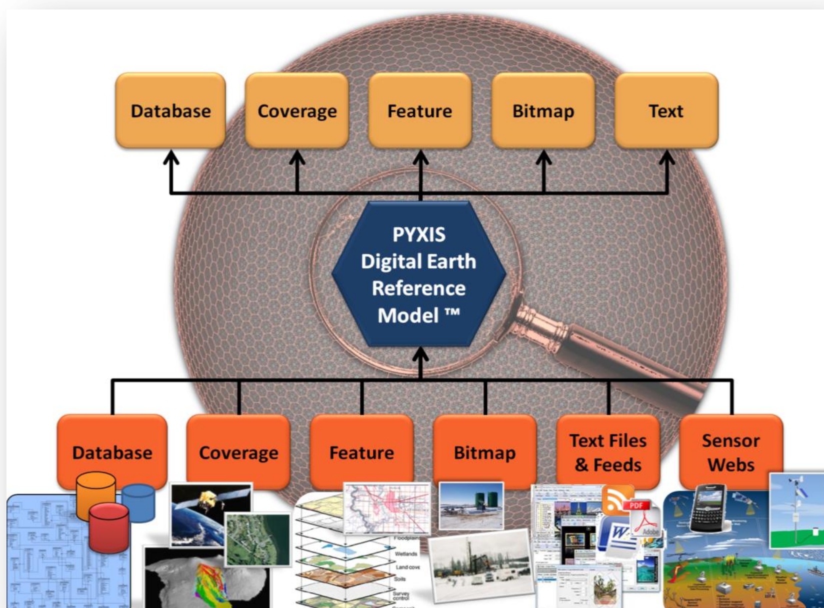
Perry R. Peterson

2 ELABORATION

2.1 DEFINITION AND MODELS OF DECISION FUSION

We cannot augment wisdom but we can augment intelligence by providing appropriate information to a decision-maker in a timely manner. Earth location promises to be an organizational structure for knowledge discovery that can close the gap on evidence-based decision-making. Organizing massive disparate silos of data by Earth location and making the information accessible for complex and consumer use is the hallmark vision of the emerging GeoWeb, Al Gore's The Digital Earth, Cisco's Planetary Skin, and NGA's requirements for On-Demand Geospatial-Intelligence or GEOINT2.

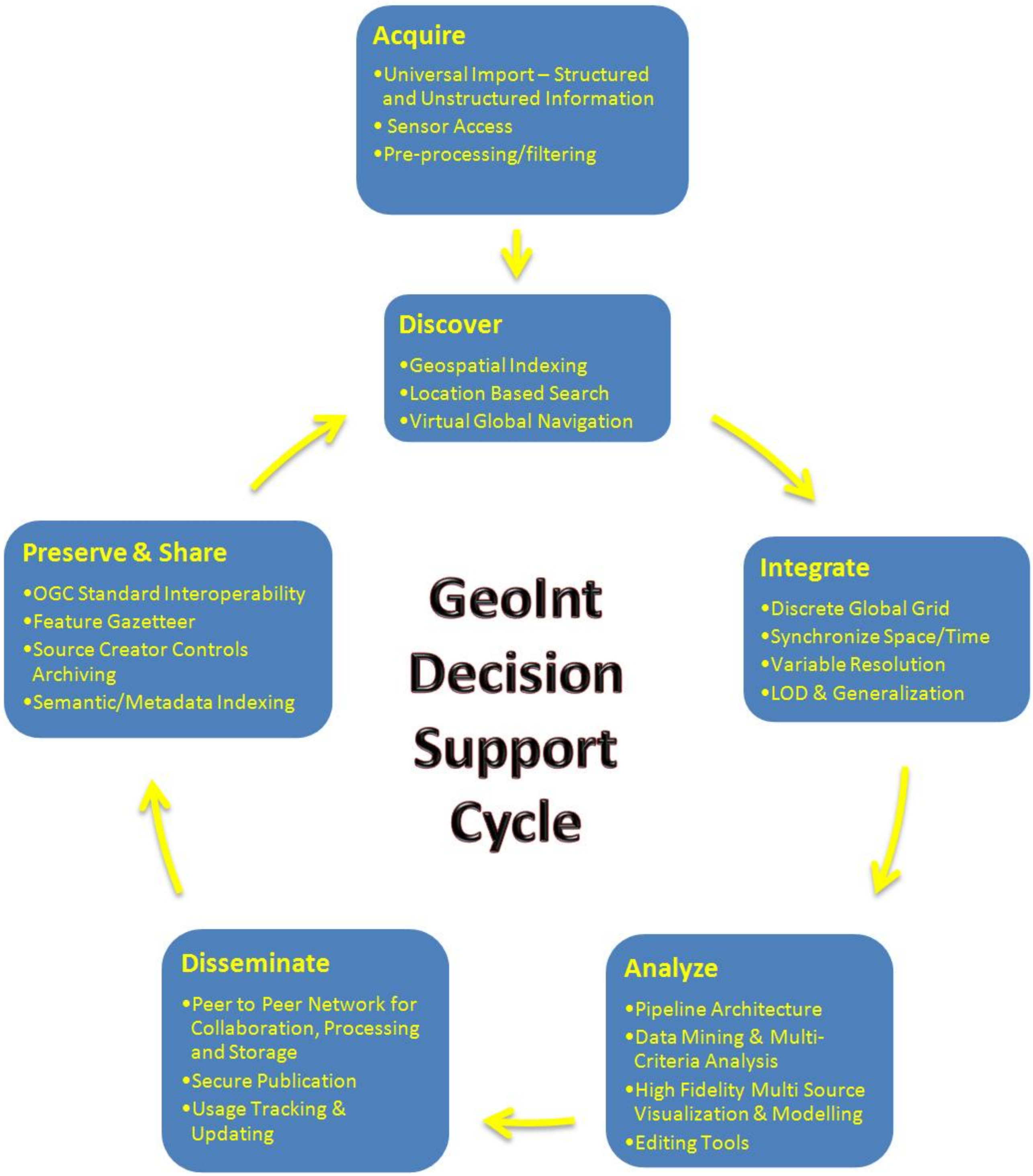
At PYXIS we hold that Data Integration, of which Fusion can be considered analogous, is the bottleneck to the GeoInt Cycle and a central challenge. It is not that data fusion/integration is a challenge in a traditional sense. There is a whole industry dedicated to integrating two of more sources of geospatial information to create single products. It is that the current process cannot *"handle and analyze the future volume, velocity and variety of information at our disposal."* (Clapper, 2006) Further, *"when working in near or actual real time, there is no time to wait for multiple forms of intelligence (MULTI-INT) to become integrated through conversion to GIS layers, nor will the intelligence demands of tomorrow fit into this simple model. Integration is indeed one of the central challenges."* (Priorities for GeoInt, NGA 2006)



Further, the future of GEOINT can not presume an architecture dependant on a central decision fusion engine. Information and the decision-makers are distributed, though the environment netcentric. The netcentric reality provides opportunities to rely on the scalability of the network, variable node sizes and capabilities, distribution of services, and shared processing and storage. We believe a peer-to-peer architecture rather than a central integrator will prevail.

We have identified gaps in the standards and addressed these gaps in our submission to RFI 1 (PYXIS 2009). We have reviewed 09-138 Fusion Standards Study Engineering Report-1. In this second submission we would like to clarify our position and present decision fusion processes, use cases, technology and products, we currently deploy within our GEOINT2 environment.

To this end, the PYXIS innovation has developed and deployed technologies, testbeds, and prototype applications that support this vision. Our process conforms to the GEOINT2 decision-making cycle and every step has consideration to a standards based approach. We promote information posting, persistence, and reuse.



Acquire

Data Acquisition – The DISCRETE GLOBAL GRID allows import of any geo-referenced information of any datum, projection, or format. It is data agnostic. The close packed properties of the grid provide statistically optimal conditions for extracting values from any conventional geodata source. The discrete grid acts like a graphics processor rendering diverse information to pixels. Data to the discrete grid can be pre-processed, facilitating transmission and discovery, or gridded on-the-fly at level of detail.

Discover

Data Discovery - The PYXIS hierarchical indexing on the DISCRETE GLOBAL GRID provides properties for fast analytical processes necessary for geospatial search, spatial relationships, and topological operations. This means a user can navigate and search across multiple databases by geographic location filtering by intersecting regions of interest – a screen view, polygon, place name, etc.

Integrate

Rapid Scale Independent Data Integration – The DISCRETE GLOBAL GRID is defined as a global reference model thus it can address uniquely and uniformly the entire planet from continents to birdbaths. The multi resolution grid is aligned and tightly packed. It acts like a spreadsheet of cells over the planet surface with special mathematical properties to rapidly aggregate and decompose information described in each cell to courser or infinitely finer resolutions. As the DISCRETE GLOBAL GRID is built upon an optimized multi resolution grid it can be used to represent arbitrarily sized locations, from continents to birdbaths and therefore preserve meaningful use of data stored at different resolutions without the need to re-project between scales. High fidelity global data integration is a product of encoded data to the discrete grid. Data encoded to the discrete grid is therefore synchronized with all other data sources that use the gird, effectively integrated and fused ready for decision support applications.

Analyze

Multi-Source Data Analysis – The grid, its tiling structure, and indexing, provide ideal characteristics for timely and efficient data distribution, processing, reuse and preservation appropriate for data mining, knowledge discovery, high fidelity multisource visualization, spatial-analysis, image processing, distributed simulation and modeling. PYXIS Information Processing Engine (PIPE) is a pipe and filter architecture adapted to multi-source geodata processing and used to perform common and complex operations on the discrete grid. PIPE is extensible allowing customized process and application development. It is essential enabling service chaining.

Disseminate

Data Distribution & Export- PyxNet is a P2P network to facilitate distributed processing and storage of PYXIS data tiles. A license service provides secure publication over PyxNet. The PYXIS index is also compatible with conventional coordinates and suitable for use in a database or other geospatial applications. Export functions provide interoperable access to the grid as standard coverages.

Preserve

Control to the Content Owner – The PYXIS Feature Gazetteer acts to resolve semantic indexing and location (place) indexing provided by the DERM. This service provides a mechanism whereby the content creator is providing information directly to the decision maker. By eliminating the reliance on a middle manual integration process, data preservation is in the hands of those with the most interest in and knowledge of the data source and out of the hands of a large integrator who must deal with increasing volumes, velocity and variability of data.

2.2 INFORMATION FOR DECISION FUSION

Section blank in this response.

2.3 SERVICES FOR DECISION FUSION

We have created 4 technologies in support of decision fusion.

1. A **Digital Earth Reference Model (DERM)** which includes a library of resources in extensable C++ encoding the mathematics and processed to generate and use information on the ISEA3H discrete global grid.
2. A pipe and filter (pipeline) architecture called the **PYXIS Information Processing Engine** or PIPE that enables a service chaining on multi sources of data.
3. **PYXnet Encrypted Peer to Peer Network** tuned to the tiling and distribution requirements of a discrete global grid.
4. A **Feature Gazetteer** that encodes and indexes massive disparate feature data sources to the DERM allowing for fast location based searching across un-federated geodata.

We have also created 4 appkcations that permit content providers the capability of distributing their geospatially encoded assets directly to end-use decision-makers.

1. **PYXIS WorldView** is a GeoWeb Browser that can search, navigate, visualize and analyze multi-source data that is readily available on the Web. This tool provides a prototype of processes and technologies that are enabled when geospatial information assets are self synchronized at any point in their use. It is supported by the DERM, PIPE and PYXNet. WorldView is highly extensible with Plug-in and Scripting Capabilities exposed.
2. **PYXIS GeoWeb StreamServer** – provides interoperable image, coverage and feature serving tuned to the DERM.
3. **PYXIS License Publication Service** – provides encrypted content control by the content provider – republishing, security, transaction, messaging, recall, update, use statistics, flow control
4. **PYXIS Software Development Kit/Environment** – C++ with Java, C#, .net wrapper, allows developers to design, build and deploy custom applications and processed on the platform

We have several pilot projects underway that OGC and supporting partners may be interested in observing or participating in.

Alberta Integrated Resource Management Pilot Porject

Goal – to enable GeoWeb type decision support network with the highly regulatory and environmentally sensitive Alberta Oil and Gas sector.

Lead – the PYXIS innovation, Adamlabs

Partners – Alberta Government, Syncrude, Tectera Centre of Excellence, University of Calgary, Natural Resources Canada, Oracle, CAE, and several key Oil and Gas producers and geodata content providers.

Planetary Skin www.planetaryskin.org

Goal - Planetary Skin Institute aims to address the challenge posed by human growth and rising resource demand with the opportunity presented by the data volumes and information processing capabilities have grown, driven by new sensor networks and a whole host of emerging information and communications technologies to help decision-makers manage scarce resources and risks more effectively in a changing world.

Lead – Cisco Business Solutions Group, NASA

Web Based GeoInt

Goal – To provide a web based medium for searching across massive features based on location and semantic content.

Lead – Harris Corporation, NGA

Public Safety Decision Support

Goal – to create a common operating picture by assimilating disparate data geospatial and temporally encoded data sources in support of critical emergency decision making and response

Lead – the PYXIS innovation, CAE inc.

We would be please to meet with your committee to review these capabilities further and in more detail.

3 ORGANIZATION DESCRIPTION

the PYXIS innovation is a privately held Canadian corporation that has focused on the development of technologies and products to support military GEOINT 2.0 and commercial GeoWeb 2.0 applications. Their vision is to close the gap within the evidence-based decision-making cycle by enabling decision makers to pull geospatial data assets directly from the data producer without an intermediate manual data processing layer - key to which are the requirements of *Geospatial Integration On-Demand*.

Their work has included use of OGC and NATO STANAG standards for interoperability. Their approach is supported by Canadian Department of National Defence (LF ISTAR, J2 GeoInt, MCE, DRDC) and Natural Resources Canada GeoConnections program which has been Canada’s main contribution to promoting ISO TC211, OGC and other open standards. the PYXIS innovation is a member of the Canadian General Standards Board (Geomatics).