GIScience Workshop Notes:

Dawn Intro

Bridging CI and GIScience – A Holistic Approach
Shaowen Wang

TerraGrid demo

Data simulation service (using SOA)
Analysis centric view of GIS rather than navigation-centric

Evaluate spatial analysis quality

Service used by several hundred users – several overlapping clusters; saved the data (20,000 data points); then to straightforward spatial interpolation (computing intensive because of nearest neighbor search of all of these points); Open Science grid resource hookup from his lab; k-neighbor values of 5,15; submitted job using a community account provided by CI communities to aggregate users together to XX privileges for them to access data and resources; computation is lended on a resource at UIUC but also to Open Science Grid; result is an interpolation overlain on top of the map of the US (ESRI spatial Analyst would have not have been able to do this in a day but the calculation was performed live in a matter of minutes); then he zoomed in to show the interpolation resolution

Will get to flow chart of the above process to demystify what occurred

What is Cyberinfrastructure?
Applications, data, e-community, more!

CI Evolution
- number of supercomputer centers
- NSF networking initiatives
- Then PACI with substantial funding of NSF
- Then Terascale facility
- Now it's "cyberinfrastructure"
- All moving toward knowledge communities combined with data; digital resources combined with high end networking and computing resources

Integration/Holism

Motivation – What's Beyond/Next
- our digital globe and digital earth vision, and as assisted by vendors such as Google

Background Slide

Review Slide

- domain specific CI activities such as GEON and NEON; driven by those communities rather than us (GIScience) which is cross-cutting

CI Complexity

- turning vision into action is not easy

Managing CI Complexity

- problem solving combined with user practice

User is in the middle to emphasis human-centered interaction, around this is middleware

Middleware of GISolve

Computational Intensity = Wattage?!!
Knowledge about computational intensity; from supercomputing the driver has been to push envelope of quantity and capability; geography has very intensive spatial analysis problems; we need to know how intense the computation will need to be and how geography problems make that so;
Every problem does NOT have to be so computationally intense though to be part of a CI; data problems, knowledge community problems are key

NSF Vision certainly led us to the opportunities to work on problems
Data intensity
virtual organizations

Shaowen's talk later about unique contributions that GIScience community stands to make to CI community

how spatial characteristics contribute to characteristics of spatial analyses computations

Domain Decomposition and Task Scheduling

- Divide and conquer in order to solve a large multiscale problem; and then move them to the right resources including people

GISolve Workflow – back to the demo

- all web services that could be shared by others to call out and access
- mini nature of Teragrid with all kinds of services; these are moved to multiple portal servers to load balance the requests
- security access also

TeraGrid
– for regular users they don't know where a job goes, where computing actually happens out on the network, everything is virtualized across a spectrum (cloud computing); the user just sees the front end

- how are algorithms certified? In order to contribute an algorithm, you have to be registered with a community account; and you must have a certificate for security control; then you can play with your code in a secure, controlled environment to test it; when ready for release there is a manager to take you through several steps to make there is no malicious action in the code
- Paper is in preparation about this whole process; clouds with user having a centric view

Open Science Grid
- another example of national infrastructure
- funded by DOE and NSF
- $30 million budget for 30 (?) years

GISolve Services

SOA approach
- for those interesting in understanding how things are glued together, this is how things work up through tiers

Spatio-Temporal Data Handling and Visualizing

Bayesian Geostats Modeling
- notoriously computational intensive
- multiple supercomputers busy linked through high performance networks
- supercomputing problem, when rendered in GISolve it becomes a collaborative problem-solving process; Monte Carlo simulation in order to for users to see if Markov chains converge and if there is good validation results; users have to look at this together and tune the parameters; same analysis environment, they compare notes, use text messaging to refine; spatial stats people were really excited about this; now that the computing is feasible it becomes a collaboration challenge

Analyses Supported by the Gateway
- references to papers recently published

production env that he demoed is half year behind prototyping and research env they are working on

Summary of demo – Integrated CI-based Workbench

Education and Outreach
- TeraGrid is also comparing NanoHops, which is an exemplar of CI community project to support nanotech computation and community building effort; 60,000 users worldwide compared to GISolve which has hundreds of users
- How to grow user base and get community engaged in building their own CI environment

Conclusions
GISsolve demos a holistic approach to bridging CI and GIScience
GISolve is effective for teaching

CIGI – CI and Geo(graphic) Virtual Organization
- CyberInfrastructure and Geospatial Information Lab
  - great to see the integration of the 4 moving across borders
  - GIScience community IS getting across all 4 borders in diagram

Disciplines Involved in the CIGI VO
  - they all see the value from the VO, one of over 24 in Open Science Grid
  - CIGI is the only GIScience VO in the Open Science Grid
  - Serving at-large VO community chair as well

Global Climate Change Impact
  - Nature article in revision, project supported by CDC

Digital Watershed Cyberenvironments
WATERS project led by James D. Myers of NCSA

Ongoing R&D

Half million-supercomputer hours used?

GeoCyberinfrastructure
Phil Yang

- personal view of CI, will show several examples from his research group, www.cisc.gmu.edu
- "GeoCyberinfrastracture" is a term that was discussed prior to formation of new specialty group in AAG

Introduction
- animation of cloud top, predictable water concentration;
- animation shows that datasets are still missing
- without background on data do not understand what it is
- data are collected from different sensors
- where will potential flooding occur
- many global and env issues can benefit from datasets like this
- resolution of data is global, very coarse rez; need more high rez datasets
- more data should be collected for dealing with challenges
- need to integrate the existing heterogeneous and large amounts of datasets

- need new models and algorithms to mine and maintain the intrinsic relationships between phenomena and data collected; for both GIScience and domain scientists
- need different science domains to collaborate with each other and share knowledge so that CI works more effectively FOR them
- need to facilitate data, info, and knowledge sharing for research, development, education, and communication → this is what we mean by cyberinfrastructure

Example: Southwestern US Dust Storm Forecasting
- movement of dust storms
- dust forecasting performance

Traffic Simulation
- loading on servers is static, but when we drive in metro areas, the traffic flow is very different; taking a different route can saves time
- need to do real time loading to help us select the best load
- need to integrate different data sets; instead of using static load, they use a different algorithm to do the loading in real time and cut the roads into different sections

Concurrent User Request
- each user submits request, loaded on fly, dispatched to different servers (CI)
- which request to send to which server?
- Near real-time loading can get to thousands of concurrent users; will overload facilities

George Mason Computing Environment Diagram
- cluster with 224 CPUs
- they connect to SURAGrid environment

LamdaRail
- transmit datasets at 1 gig bps connections on the network

Dust Storm Forecasting
trying to improve rez of dust forecasting
model simulation; different parcel size
performance evaluation (how many CPUs to leverage to support the simulation?
Compared to prior, can leverage more CPUs with higher rez; 3k rez model running by using different number of CPU cores
- New Mexico; coverage is around 1.5 states – they can then extrapolate to how many more CPUs will be needed for national coverage
- Algorithm for this??

How much computing power for 1 km global water cycle simulation?

Another example of near real-time loading
Route change with congestion on roads in Clifton, VA

- don't want to waste CPU cycles
- takes too many computing resources (port connections)
- they developed new middleware to help with near real-time loading; allowed 8 simulation jobs to be submitted using only 4 CPU cycles
- help us to get good throughput within grid infrastructure

Sharing & Integrating Data Globally
- another animation, but with familiar background map; snow cover change on a global scale
- interoperability; using web map services (WMS) to put up data

Earth Information Exchange
Semantic Search
- access to several catalogs including Geospatial One-Stop, NOAA, NASA catalogs
- can pool together results of searching all of these into one interface
- use ontologies to navigate our search
- through semantic search you can link to catalogs served from different locations; don't have to know which catalog is being accessed but result of search comes back through one interface to the user

World Wind WMS View
3D to look at different places
- goes to server to fetch data which is gradually updated
- using catalog services for the web (CSW)
- can access other data sets such as Terra Modis sensor (fire outbreaks?)

What's Missing Now?
- CI can really help us to enable scientific research and facilitate development of apps that we need, HOWEVER….
- GEOSS and Digital Earth envision to leverage all possible resources
- GEOSS is the Global Earth Obs System of Systems – still at early stage, still talking about portals and protocols, lot of work needs to be done
- Digital Earth – do we really have a Digital Earth as envisioned by Al Gore? No, but we have a journal ☺
- Microsoft Virtual Earth and Google Earth
- We are still far from initial vision
- Data integration and collection needed (e.g., earthquake in China – could data be collected and integrated quickly enough for emergency response, decision support)
- We still need technology beyond the current; consider data collection but mgmt and support, sharing, processing, analysis to support different applications
- Forthcoming book on *Advanced GeoInformation Science* on these topics, based on trainings given at GMU

Need more collaboration between computer scientists, geography, Earth science
High Performance Computing still widely needed

Middleware is needed (NSF middleware initiative)
Gaps in workforce needs and education of students; not many courses on geocyberinfrastrucutre
More applications needed to showcase how CI could help us; e.g., emergency response (Wiley Thompson thesis at Oregon State)

"To Name a Few"
- funding agencies – NSF, USGS, NASA, EPA, DOE
- industry – IBM-Google, Microsoft
- GIScience
- AAG CISG – cyberinfrastructure specialty group
- UCGIS
- GIS&T Body of Knowledge
- More

Rene Sieber – evolution of CI from supercomputing is understood but what about social sciences and humanities; they need data but are not computationally intensive
Social science has a different set of challenges – semantic searches, ontological issues are key; SDI stuff is key
- need interdisciplinary research; need the input of domain disciplines; not just GIS solution but how fluid dynamics would change in the dust storm
- social science needs to be included in the "domain science" realm

Barbara Poore
- social science including in GEON

how social interaction happens between scientific communities
2007 Understanding Infrastructure Dynamics Tensions and Designs workshop report, U. of Michigan
NSF Virtual Organization initiative

Andre Skupin
- trying to define a new discipline when it might be more appropriate to define a new subdiscipline within GIScience
- geographic is more holistic label than geospatial
- don't forget about collaborations intellectual paradigms that already exist

---

Knowledge as CI: Toward Geographic Knowledge Systems
Rob Raskin

-transition from data to knowledge
- data is really at the level of numbers where knowledge is more about facts; information is in the middle and gives us the intermediate step between data and knowledge
semantics = meaning of concepts

ontology can be used to make some efforts to STORE knowledge
What is Spatial Knowledge?

Ontology: Knowledge as CI

- context comes from SHARED understanding and agreements
- content is a dictionary that is readable by both humans and machines in the form of a namespace (URL, URI pointer)
- the authority for all of this is anyone, but the best will win out – perception is everything; people go to Wikipedia rather than online Encyclopedia Britannia

Enterprise-wide Ontology
What is the meaning of terms across different domains? Different groups are accumulating knowledge but talking to each other or sharing between each other in terms of concepts
smart discovery is key – Google does not understand the meaning of terms but neither does ArcGIS!
Knowledge commons = institutional memory for spatial concepts

Why Point to an Ontology?
Ontologies are static things that you point to which then enables machine-to-machine communication
Associate data with its context – you can therefore say anything that you want to say about a dataset; accuracy, assumptions of measurements, what do you mean when you say "climatological mean" for example; so there it goes BEYOND mere metadata because you can actually define what you mean by the terms that you are using

Use Case: Global Warming Query

Ontology Representations
Ontology first coined by philosophers to indicate "all that is known by people"
We now deal with "all that is known (by computers)"

Ontology Rep: Triple
Ontology Rep: Visual
Plate Tectonic Ontology
Ontology of an Organization (ESIP)

XML – humans and computers can read this fairly well so it is a reasonable compromise

RDF and OWL are XML languages
OWL lets you define much richer concepts of semantics, cardinality restrictions, etc.
Both RDF and OWL give us the building blocks
W3C is the leader in developing these languages
Open world assumption = if you haven't stated a fact, it could still be true, unless you say something is false (more consistent with the web and the notion of incremental

knowledge); but in the database world, whatever is in the db tell you the whole truth; if is not present in the tuple it is false
Ontologies let us add facts incrementally which is a great advantage

SWEET
For Earth Science but also has spatial concepts
Numerics
Space
Time
Data and information
Data and information services

Why Use an Upper Level Ontology?

- allows for common concepts to be across physical and social sciences for examples; common definitions

SWEET 1.0 Ontologies
2000 concepts in about a dozen ontologies

SWEET Spatial Ontologies

SWEET 2.0 Modular Design
No longer a 1:1 relationship between facets and files
Files now organized by subject to help domain specialists
Instead of 12 large ontologies there are 100 small ontologies
Including Social Sciences
Of course many concepts don't fit in just one area; ok, you just repeat them in several different places
Ontologies need to capture ambiguity of the English language (English in THIS case)
What is the dependency of a term/topic? If it needs to go in more than one place, it can

Community Agreements: How to Optimally Use OWL

- need to have workshops for different domain specialists to come together to populate ontologies
- geological sciences have done this – plate tectonics, climate
- Marine Metadata Interoperability has done this for ocean sciences
- domain specialists (world leading authorities) with ontologist
- spatial decision support systems, U of Redlands - can build, mix and match, and end-to-end system for decision support
- pdf here = probability density function
- getting agreement, and get people to consistently work together to create a knowledge infrastructure to go with the hardware infrastructure
- NSF has had much fewer solicitations for knowledge; still focused on hardware; the more proposals we submit, the more they will realize the importance of the knowledge work

Planetary Ontologies project

Ontology Development Tools: CMAP
http:/cmap.ihmc.us/coe (export and imports OWL and supports 80% of the features)
visually is so much more powerful than XML

Resources
Sweet.jpl.nasa.gov
PlanetOnt.org

We need to see knowledge as an integral part of CI
We have to work together; requires more meetings, workshops; add a workshop of this nature into your NSF proposal

Karen Kemp – social science could be in the middle and other stuff could go on outside as opposed to SWEET 2.0 visualization

- the point is that anything can be inserted into the center; if your center of universe is different, you can switch the representation
- representation concepts end up coming in first

Renee Sieber working on Chinese historical biographic and geographic data project ontologies for GUI, query broker, all by way of a CI

Roland Viger - if a person were coming into this community from his own context how would he reorganize the existing structure to meet his needs? How to get people who don't normally talk to have a common language and move forward?
Let's put 2 ontologies side by side and compare and discuss; these are just conceptual maps. It is within the boxes that things really happen. Opening something like this on a community basis is really important

What would be a grand challenge in ontology? Go back to the "Use Case: Global Warming Query" slide and switch the process? Also, Tim's work on folksonomy and link that with ontology? And something that scales up and down

Scott – NSF document that he can pass to the group

UCGIS Knowledge Web
Tim Nyerges

Overview of UCGIS as an organization

Knowledge Web at way of:

- Increasing the "value-added" for UCGIS as an organization to re-energize the membership and increase participation

- creating a way for the UCGIS Research and Education committees to collaborate on a common initiative

Can we carry the BoK to version 2 and make it the core of the Knowledge Web for those 2 thrusts of the organizations?

GIScience has always been a CI of sorts; has always been a knowledge infrastructure

Using the term "dynamics" along with CI is very attractive; dynamics of many things is a huge research challenge

Cyber-enabled Discovery and Innovation (CDI) cuts across NSF's Cyberinfrastructure Vision for 21st Century Discovery
The GIScience Knowledge Web includes all 4

1) high performance computing **("petabytes for peta flops")**
2) data, data analysis, and visualization
3) virtual organizations
4) learning and workforce development

How does Digital Earth and SDI interplay with CI?

SDI does not always get you to all of the data that you need, hence the NSF DataNet initiative to assist with this

This workshop is running under an open world assumption! There are many more perspectives out there that can be taken into an account

GISKW – 3 Components (keep it simple in a triple)

1) Body of Knowledge – Research and Education
2) Workshops and Agenda – Research
3) Model Curriculum – Education

GISKW BoK R&E Component
AAG has offered up their Manuscript Central to assist (it is a CI of sorts for journals)
5 managing editors for the next version of BoK so UCGIS is actively looking for funding for this

the scientific glue that disciplines need to be effective at larger scales

GISKW Research Component

"Networking" has not gone away
"Supercomputing" has not gone away
We don't know what the next "word" beyond "cyberinfrastructure" will be – NSF has not gotten that far yet

GISKW Education Component
What are the pathways though various educational processes that can be brought together?

GISKW Education Component
Medium Term

- what about on your own campus; use a GIScience approach to collaborate on your own campus among multiple units; e.g., 20+ units on campus doing GIS or GIScience; there is an opportunity in GIScience and CI to move forward with an interdisciplinary perspective

GISKW Information Architecture
GIS BoK already has an organizing structure within it that can be leveraged
The center of the world for mathematicians is math; for us it is space and time
Social sciences are much harder because everything changes so rapidly

Why do we still not have the ultimate collaborative infrastructure in place? Do we even know the criteria for such a platform?

UCGIS Seeking Partners for 3-Component Effort

NSF is definitely a partner because of the CI potential for research and education

Tim – This often comes down to competing with computer scientists for this kind of money.
It's easy to be a social sci in CI and do science for society; the computer scientists are not adept at that so cannot compete

USGS
NGA
ESRI – could not have done the BoK without their $100K funding to support the workshops

Jeremy – can GIS be rolled into "geoinformatics" – politically how does GIScience stack up to geoinformatics – could be the reverse also
Tim: Geoinformatics a more practical approach to the Sciences than GIScience (nature of tools and techniques). Geoinformatics does not have to explore the nature of knowledge on a head of a pin. Ultimately they both overlap on each other and would depend on your world view.

Renee – NSF, AAG, UCGIS refer to the U.S. but other countries have bigger pots of money such as Canada.
Nick and Tim will align forces of UCGIS and GEOID. Developments in China and in Europe with AGILE are in play as well.

Jeff – How do we see this going beyond 77 members of UCGIS for the broader, international GIScience community?
Access on a computer lends itself to different levels of access.
AAG and UCGIS hold copyright to BoK. So it must begin with the managing editors, the relationship with the AAG, etc. But the BoK will be vetted worldwide by some level of community. Primary access

Scott – NSF does allow and support international collaboration through subcontracts, etc. And they like to see proposals funded internationally. What kinds of things would countries help support.
David Lightfoot at NSF – BYOB – "bring your own budget" – NSF will fund Americans working in collaboration with internationals

Phil – scope of GIScience, geoinformatics, Earth science community does not always think of itself as part of the GIScience community. How to deal with that dynamic? It's about how to collaborate rather than putting one community above another in importance. We are still learning how to work with other (cognitive psychology literature, etc.). Large scale collaborations (folksonomy over ontology is key here, thousands of people making common meaning for the effective of democracy)

The National Map
Lynn Usery

National Map is critical component of NSDI

Structure and practices relationships to Geospatial Data Program of USGS
NSDI frames the NGP

FGDC not really in the production business anymore

National Map is the integrated topo content portion of the NGP

There was no Google Earth or Virtual Earth when the vision for the National Map was launched. They are working on viewer capability in today's "Googlized world"

Right now there are too many access points to The National Map and they are trying to get down to one

Going from National Map 1.0 → 2.0 → 3.0

National Map 1.0:
Close to having complete 10-m coverage of US (90% there)
With NGA there is imagery over 130 urban areas at 1-foot resolution. City of Atlanta by itself is 1 Tb
Google Earth is using USGS geographic names, and has finally branded it as such

Hydrography coverage
Elevation data
Partnerships with states, liaisons with 42 of the states
For 8 data layers, 4 of them are vector; in coop with ESRI there is a best practices data model for transportation

National Map 2.0:
Changing from distributed to centralized databases
USGS has had a problem in that they get whatever partners have but now everything will go to centralized and will be distributed only through USGS guidance
GOS (Geospatial One-Stop) will be primary viewer
Developing a production capability for graphics (problems with hazards and homeland security)
Guided by a 2-year tactical plan
Many planned products and services
Can take data at highest resolution and generalize to lower resolutions

National Map 3.0:
National coverage 8 data layers
Seamless elevation, imagery, land cover, hydrography
Maps will be distributed as geo pdfs
Ontology-drive
Vision for this comes from an NRC report that was done for the USGS
Quality aware data model
Spatio-temporal data model
Going from layer-based to feature-based

USGS Center of Excellence for Geospatial Information Science (CEGIS)
Lynn Usery

Cegis.usgs.gov
All projects, proposals on this web site (after 4-6 months of clearance)

Research is guided by previous plan about geography research at USGS, 2 of them about GIScience

CEGIS is a virtual center within the USGS and there are affiliated scientists as well (Keith Clarke, Barbara Buttenfield, Cindy Brewer); post-docs as well (2 in attendance of the CI workshop)
They rely on CI to make things work

Maintain world-class expertise, leadership
In recent years, USGS has moved away from cartography and GIScience basic research and wants to return to that

NRC Study Overview (and how USGS has implemented those recommendations)

Among current projects that CEGIS is working on:
Ontology is at the core
Generalization
Data integration (fitting together data to make topo map)
Multi-rez raster data
More

Geographic Feature Ontology
USGS has been working on feature lists for years (essentially ontologies)

Ontology Project Status
Specialist Meeting with UCGIS in Feb 2009

Barbara Poore is PI on User Centered Design for Web Services
Cindy Brewer is PI on Electronic Topographic Map Design
Lynn Usery is PI on Data Integration (can you integrate based merely on the metadata)
Larry Stanislavski of SAIC is PI on Generalization (change rez/scale from 1:24K to 1:100K)
Multi-Rez Raster Data (pull map down and get it in desired projection, involves parallel computing)

Ontology feeds into their long-term projects

Questions
Skeptics of VGI, Open Street Map, etc.
USGS was one of the first orgs to use volunteers to collect information (National Map Corps back in 2000). Now USGS is out of this business because they have such a backlog of data from volunteers that they can't get to the processing. Started out with just names collection, then to structures info (but USGS represents a building with center of building, not just entrance). USGS would like to get back to the volunteerism.

Few people know about the National Map Corps

Are web services part of 2.0?
National Map is built on web services; catalog of web services
Layers currently available, but not all as web services

Bill Manley - Kudos to National Map, amazing resource
Example of months of effort and tens of thousands saved – 0.3 m orthoimagery in area of interest downloaded in a day
$2M spent on orthoimagery, $25M worth is available through partnerships
LIDAR for the nation

Critique of claims of neogeography:
Citizen scientist vs. domain scientist

Related to geospatial vs. geographic; reinventing something that already exists so that you forget the past
For example, Neogeography's definition of geography is mapmaking by professional cartographers (inadequate)
So we have to be more pro-active in defining what GIscientists do, geographers do, rather than letting other people decide and define for us what we are doing
Bunge's Theoretical Geography is slowly being lost and re-invented because geographers are not speaking out on this
For the sake of intellectual progress

Barbara Poore – user-centered work is more than just interface; includes the foundational question of who are the users and who are the potential users? Citizens? Neogeographers/ This realm of people has been excluded even from internal surveys. Theses are the people who will open up the future to us. Disciplinary boundary policing may hold GIScience community back. If you had a project that would enlist people with projects like the National Map could be incredibly productive. Can community police itself rather than central rules (GPS right to center of building)

Renee – citizens are technological partners; they advance the data, sharing, inferential geography issue; they should be looked upon as partners and should be invited, on the ground

Wikipedia as equivalent of open street map; free and Ciitizendium where there is some editing and authority, and quality control

Urban health, food, accidents, things that lead to bad health; they are presenting a lot of data that was not first thought of; rich creation of community-based GIS; scholars could not think of these data because we don't live in those conditions; citizens can add to our knowledge; it is contextual knowledge, part of their daily lives
Yes, there is difficulty in handling the processing

The wisdom of crowds – how to aggregate data; under what conditions
Financial page of the New Yorker, how businesses can leverage the wisdom of crowds

In health GIS, conventional GIS may neglect issues that are actual on-the-ground, issues that lead to poor health; a lot to be learned from citizen groups

Could not CI contribute to solving the problem of processing all of these data? Maybe, maybe not.

Andre - keep having to defend what USGS is doing; firefighters chose to use these kinds of data because they trusted them. They used USGS data rather than Open Street Map

Orthoimage

Special issue of *Cartographica*?

Neogeography issues in CI
Crossing the domains from physical to social science for CI
Andre, Rina, Renee, Barbara, neogeographers

Bridge between info and knowledge
We supply sufficient evidence to each other and we develop trust
Go after evidence; what is the character of that, that then leads to trust of the source
Evidence has to be "authoritative" and rhetorical as opposed to assertive – use of wikis and people vote on whether they think evidence is reputable – we don't talk enough about the evidence issues to unpack and how it links to

There is trust and taking things ON trust without having the evidence

What are the similarities between world of words and the world of geographic space.
How is trust established on wikipedia on written word vs. maps and GPS recordings.

The National Map
Nmcatalog.usgs.gov/catalogwms/based

We may need smarter computer rather than high performance computing

---

SOA

Sharing of data, methods, services – those technologies
semantic web services

FGDC best practices in geospatial SOA

How Web 2.0 and 3.0 and its impact

Carl Reed of OGC – OGC standards and SOA research

Bin Li – automated approach for service discovery and how RSS is used to discovery geospatial services

Chris Mattman, Jim, 3 – how to deploy SOA for various applications

Problem still remains that service discovery is still a problem not jus tin GIScience but mainstream in general.
If we want to integrate both data and services into CI, a system architecture design is important

specific services
how to find them
how to make services reliable

GIScience 2008
Joint Discussion Session Between SOA & Cyberinfrastructure Workshops
2:00-4:00 p.m., Tuesday, September 23, 2008

Sustainability – deployment, development
(after we build it, what then?)
refers to ontologies as well
libraries to absorb cost? Sysadmin?

Permanent entity/community
ESIP Federation, need more groups like this (AAG?)
GUIs get left out, updating over time
Tools to make use of ontologies, to build, visualize – more tools to support more ontologies

SUSTAINABILITY
Long-term objective of a center
NSF program managers?
Sustainability remains a challenge, how to fund? Cost a huge issue
Many universities share a sysadmin with an entire school/college

1) Archival side (after project is done)
2) Evolution side (more complex)
myUSGS.gov is a start

1) Operational model for system, with transition plans
2) Amazon web services, other vendor services

Examples at Indiana University?
Katy Börner model of getting CI money from NSF and used that to hire professional programmers (implement CI ideas only to create sustainability; get framework, architecture down)
Open Source, Development Paradigm

NSF Vision for Cyberinfrastructure – but vigorous planning process to plan next CI calls for CISE/ENG, cross-cutting
Science Community Participation Process
TeraGrid – next phase planning, GISolve involved
Next Generation CI

NSF is aware of GIScience but could be even more aware

Grand challenge of competing with Computer Science; need to have a CS/IS/Engr person on proposal for CISE; but CDI (Cyber-Enabled Discovery & Innovation) and other programs are broader

DataNet proposals require a sustainability plan. These are huge proposals for up to $20 million but how about "sustainability" in much smaller proposals?

Proposal on sustainability as main focus might fly with CDI for example; how is this all getting out into society; a good "business model" for example for web services, an SOA for a community
Also, give scientists access to public algorithms (could be part of UCGIS Knowledge Web also?); this is sustainability also
SOA is all about this
Ontologies going to a community (e.g., SourceForge). Is there a SourceForge for ontologies (OOR, Open Ontologies Registry)

Advanced data and visualization services – 2-3 projects at $3M over 5 year cycle
4 types of services - $27M per year for 5 years

Sustainability issue – Two important questions at outset, before going after money: Does our research demand CI capabilities and if so, what are they?
Are we going to build CI, and if so, how do we go about it?

Existing CI can help smaller projects
UCGIS Knowledge Web an example of thinking this through

PPGIS – don't underestimate what citizen scientist can do; can they share code, methods, services too?

What do we need and when we have this, how is our research improved?
Community information systems in extremely poor neighborhoods (little technology but value the potential; can see the power of spatial)
Where to get the funding for this?
Community groups devastated by budget cutbacks

Prior interaction with Dan Atkins – integration as a key component, knowledge communities
Work with TeraGrid, plug into common CI projects
Mosaic created with $20K; sometimes most successful outcomes don't have to be funded at huge level
Foundations?
Google provides CI, partnership with IBM for data-centric computing
A disciplinary community to support smaller projects within the community

CDC on a notifiable disease system? WhoIsSick.org HealthMap.org; value back to the community at a fraction of the larger cost; CDC

How can CI interact more with GeoWeb; GeoCommons.org – users upload data and GeoCommons renders as map, table, chart, tag cloud – can we go beyond map, make rendering distributed, etc. within a Web 2.0 model

NSF is about funding big science for the nation, to meet national needs
- e.g., nationwide network of communities in crisis?

What kinds of big questions need to be answered and then go to appropriate funding agency accordingly

Indiana University Social Efforts and Vulnerability Indicators project?
Funded by foundations; Des Moines United Way to integrate their community information system into their existing structure
PARTNERSHIPS!

Not just providing a service, but there is original, compelling research that comes out of the projects
NSF Geography and Regional Science

January 20, 2009??? ☺

Opportunities for Action:
1) PNAS special issue – Dawn and Shaowen

2) Advanced Geoinformation Science book - Phil
Linkages to Earth Science community

3) Cartographica article submissions or perhaps special issue – Jeremy

4) DataNet solicitation which would involve multiple institutions – Nancy; preliminary proposal deadline is Nov. 10th;

5) Springer book series on Cyberinfrastructure - Xuan

6) AAG Annual Meeting in Las Vegas – propose a session and CISG and GISS groups would co-sponsor; Oct. 16th deadline

7) CIGI Virtual Organization, part of Open Science Grid, sponsored by NSF and DoE – Shaowen
DoE certificate allowing access to large repositories around the country

8) USGS Specialist Meeting on Ontologies with UCGIS, February, 2009, Washington, DC – position papers for travel funding – coordinated through UCGIS

USGS personnel cannot be PIs on NSF projects, but they can be co-PIs, senior personnel; USGS has big datasets, juicy problems

10) www.GISolve.org - request an account – Shaowen; access to services

11) CISG of AAG – newest group of AAG; vision document/white paper to inform NSF; value of CI for geographic science
Education Committee to educate geographers on value of CI
Outreach Committee – article for AAG newsletter
Awards Committee – student awards at Annual Meeting

THEY COULD USE MORE VOLUNTEERS!

12) CDC Knowledge Base – team.cdc.gov, jtobias@cdc.gov - he can submit to sitescape team; access to CDC enterprise GIS knowledge base; past PPTs, organizational memory; links to other places on the web; some similar content to GIS&T BoK; vendor-specific info; code sets, snippets (e.g., SAS) for generating maps on the fly; 475 users around the world

13) UCGIS Winter Assembly, first week of February, 2009; Workshop on next steps with UCGIS Knowledge Web (Friday morning); Washington, DC; focus on transportation as well
www.ucgis.org