The GeoPlatform:

How service-oriented architecture impacts your organization

Torrin Hultgren

EPA National Geospatial Support Team

July 15, 2014

The vision of SOA

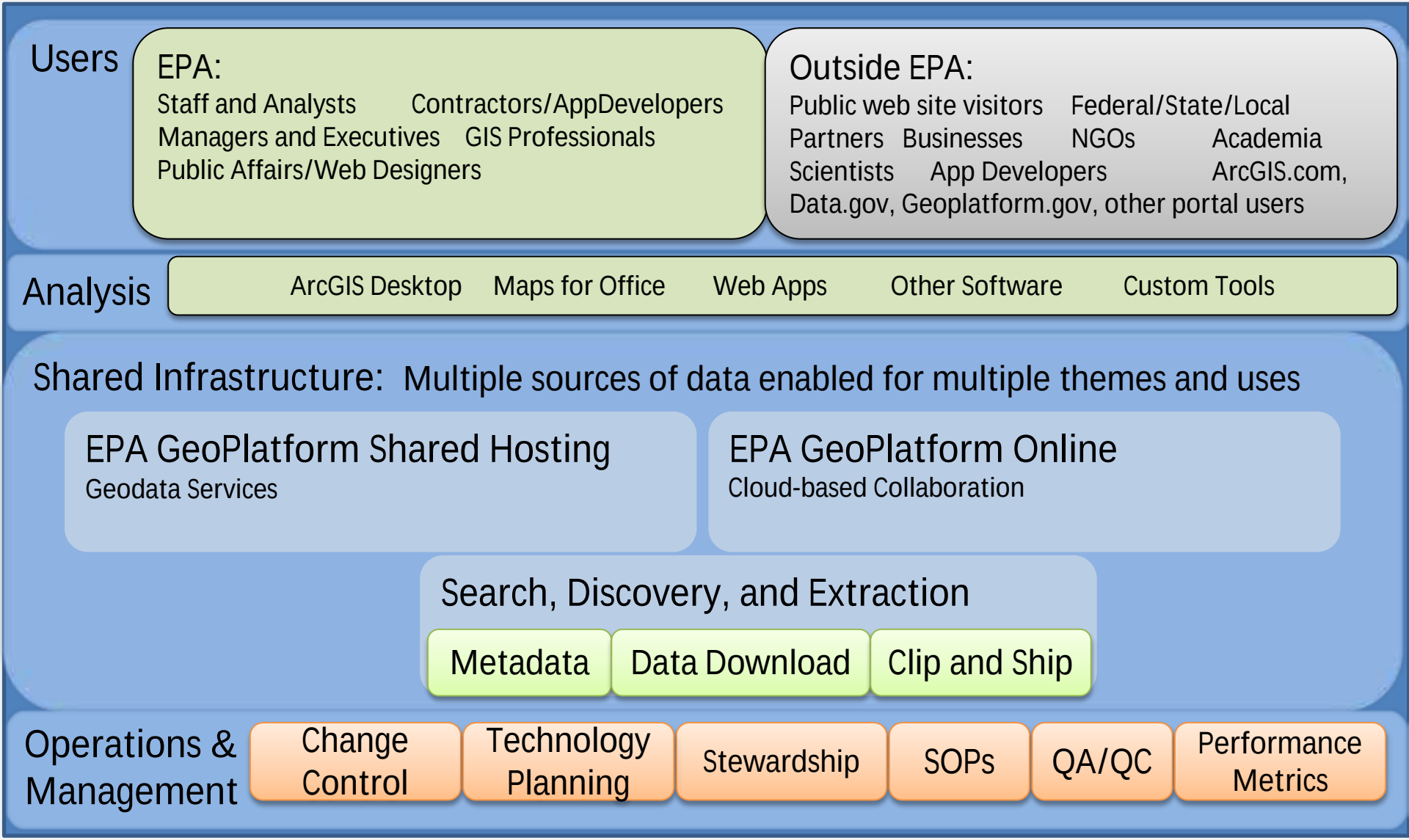
- An ecosystem in which everyone is responsible for managing and publishing their granular, reusable services
- Effort focused on everyone's area of expertise, not obtaining and processing someone else's data
- Decreases redundancy, improves efficiency and lowers costs
- Faster development cycles - ability to adapt to changing requirements
- Lowers barrier to entry – more people authoring more maps
- Transparency and data openness fosters higher quality

Sounds great, but isn't everyone doing this already?

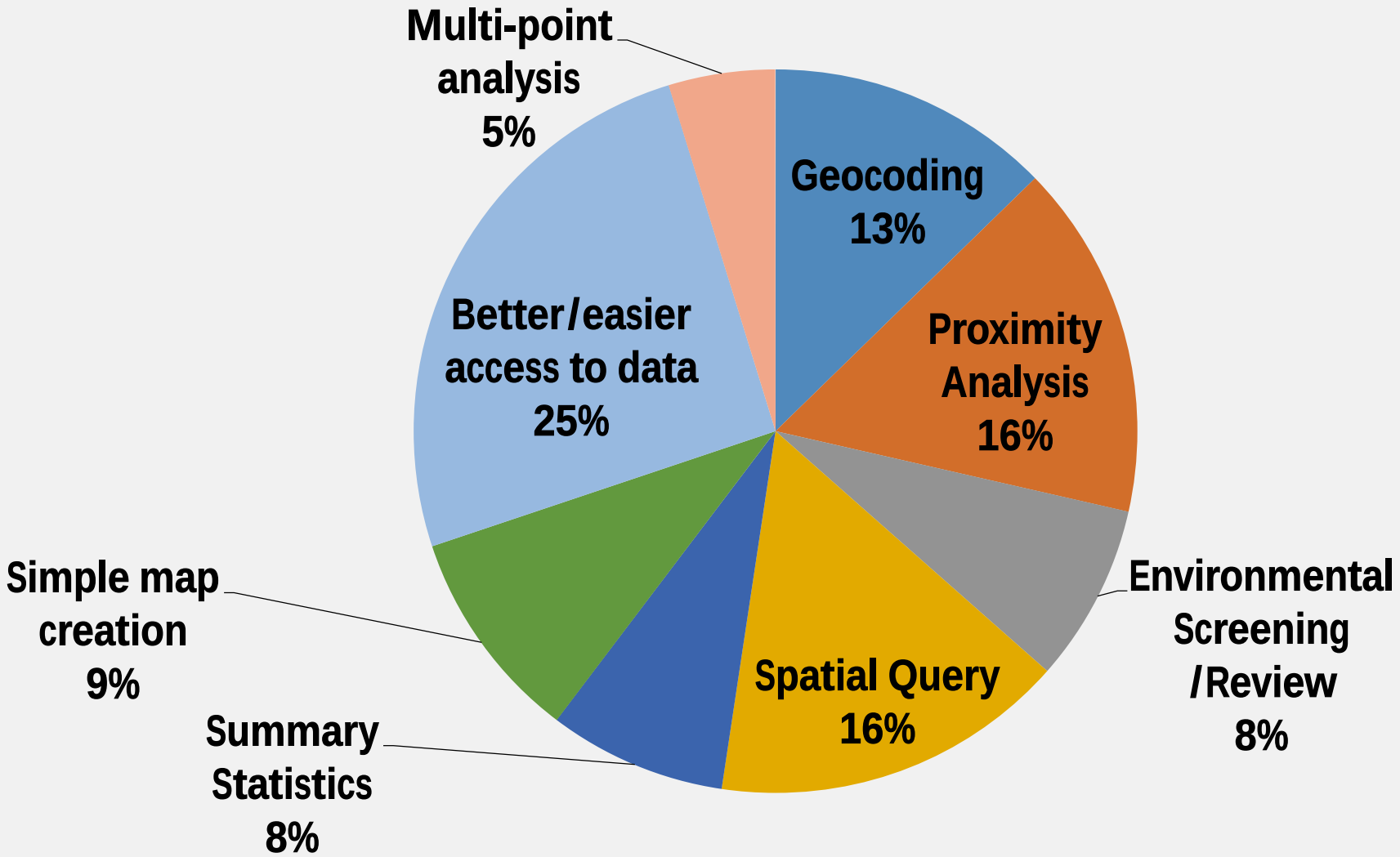
- Service Oriented Architecture as a technology isn't new
- Adoption has been spotty, or isolated, or stovepiped
- What are the barriers to publishing services?
  - No requirements, less control, less credit/more criticism, costs
- What are the barriers to consuming services?
  - Discovery, fit, documentation, reliability, change control
- SOA isn't something you can buy, it requires organizational change
  - Roles, responsibilities, expectations, workflows, trust
- We have a success story to share, but thorny challenges remain

EPA GeoPlatform Concept

EPA GeoPlatform Architecture



EPA GeoPlatform Business Case



Publishing Barriers – No requirements

- Fundamental premise of SOA is that not every use case can be foreseen - services will be discovered and recombined freely
- How do you publish with no clear requirements?
- We authored a checklist of best practices to guide authorship of generic, reusable services
- Publish the same data in multiple open formats
- Design services to be modular and flexible
  - Include only semantically similar layers
  - Separate labels, enable dynamic rendering

Publishing Barriers – Access Control

- Data owners often feel possessive about their data
- Addressing at the organizational and federal level via mandates:
  - Data required to be published openly by default unless a valid justification is given for restricting access
- Continuous outreach documenting success stories built on open services ensures data owners continue to receive credit
- Reassure data owners that transparency may air dirty laundry but will get it cleaner faster – avoid organizational blame
- Data misers do not win friends in an organization
- Services and data portals do support access controls

Publishing Barriers – Message Control

- Data owners often believe that their data is unique and complex and therefore they need to guide users carefully through it
- Result is often overweight, kitchen-sink style maps that imitate desktop functionality but still lack a message and overwhelm users
- Solution: New application development model:
  - Tightly focused maps/apps – simple UI, specific message
  - Agile development – flexible, lightweight, adaptable
  - Works best with flexible, modular, reusable services!
- High quality data with thorough metadata will ensure that services will be used appropriately

Publishing Barriers – Bureaucratic and Financial

- Historically the bureaucratic burden to publishing a map service or application was significant
- GIS Applications can be complex, involve many IT groups, and many reviews for content and technology alignment
- EPA has internal hosting cost recovery mechanism, GIS complexity also drives up those costs
- Result: only the very largest, best funded applications survived, which means no ecosystem of services
- Solution: communal “GeoPlatform” hosting environment with no direct cost and streamlined review process, limits on application size and complexity to control costs

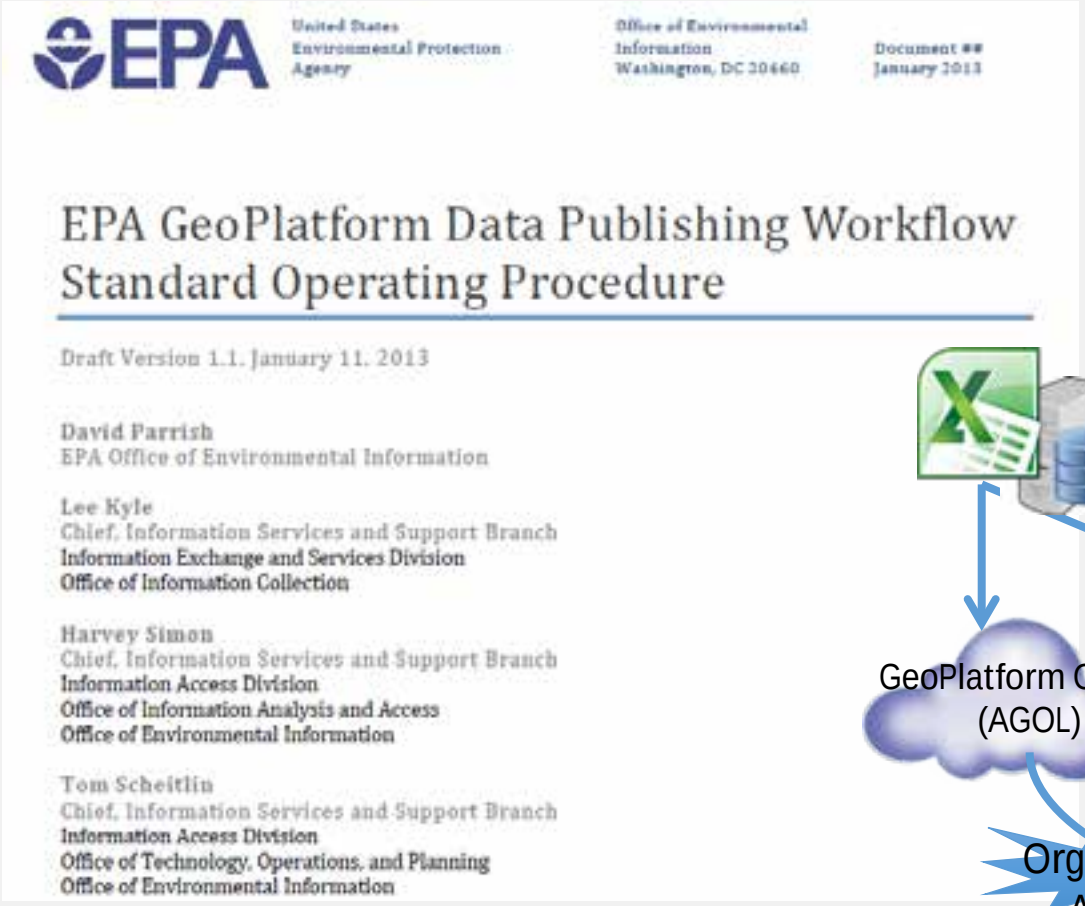
Bureaucratic and Financial Gaps

- Federal authority to operate (ATO) in ArcGIS Online cloud
  - EPA still operating in provisional mode
  - FISMA low certification recently approved for USDA
- Feature service hosting still outstanding, as well as FISMA moderate
- Federal security requirements mean cloud solutions rarely cheaper, usually much more expensive than on-premises solutions

Publishing Barriers – Technical Knowledge

- Publishing geospatial data requires significant expertise: subject matter, cartography, metadata, as well as databases, application servers, web programming and cloud services
- ESRI is really helping by providing simple, easy to use tools and templates:
  - Esri Maps for Office
  - ArcGIS.com viewer, FlexBuilder and new JavaScript app builder
  - Story Map templates and other application templates
- We can't eliminate all the complexity, some business cases and data are simply rich and complex
- We have developed SOP documents, best practices checklists and are conducting ongoing training and outreach
- We also rely on traditional GIS staff (analysts, managers, and developers) to educate new users and review products

Process for publishing maps and data



EPA

United States Environmental Protection Agency

Office of Environmental Information
Washington, DC 20460

Document ##
January 2013

EPA GeoPlatform Data Publishing Workflow Standard Operating Procedure

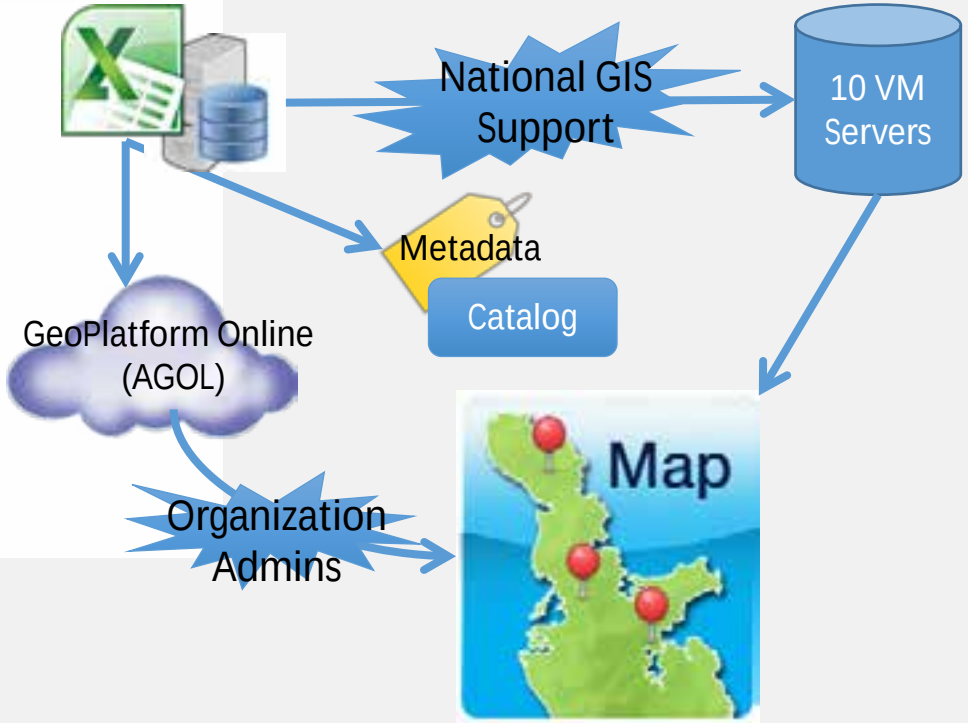
Draft Version 1.1, January 11, 2013

David Parrish
EPA Office of Environmental Information

Lee Kyle
Chief, Information Services and Support Branch
Information Exchange and Services Division
Office of Information Collection

Harvey Simon
Chief, Information Services and Support Branch
Information Access Division
Office of Information Analysis and Access
Office of Environmental Information

Tom Scheitlin
Chief, Information Services and Support Branch
Information Access Division
Office of Technology, Operations, and Planning
Office of Environmental Information



Service Publishing Checklist Sampler

- Performance really matters – test, tune, optimize, test
  - Use local data in Web Mercator projection with indices
- Services are multiscale – set appropriate visibility, generalization
- Services don't do relational data well – flatten, reorganize
- Include all/only appropriate attributes and aliases
- Enable all practical endpoints and formats
- Publish source data in parallel
- Produce thorough and complete metadata
- Register service, crosslink metadata

Usage Barriers – Resource Discovery

- Finding resources can be a big challenge:
  - Many services and downloads aren't indexed by Google
  - Many services aren't registered in ArcGIS Online or other portals
  - Often best resources are buried in clutter of ArcGIS Online or other portals
- Carrot and stick approach with metadata:
  - EPA Metadata Editor and Metadata Styleguides make authorship easy
  - Registration of complete metadata a requirement for hosting
- System of federated metadata harvesting from producers through Environmental Dataset Gateway up to Data.gov
- Active partner in federal *project open data* initiative to make datasets more accessible and machine-readable

Usage Barriers - Gaps in Resource Discovery

- We are still struggling with tagging within the ArcGIS Online Environment:
  - Need to simultaneously promote and standardize keyword usage
  - Need to better match tagging with used search terms
  - Need to find better mechanisms for weighting search results to emphasize preferred datasets
- Creation of service metadata is still detached from dataset metadata
  - Need better, more seamless synchronization
- Metadata standards and guidance are still misaligned
- Goal is federated stewardship, still a clear need for a catalog librarian

Usage Barriers – Resource Fit

- All too often, even after discovering a service, users will say “that’s great, but…”
  - “I don’t like the cartographic choices”, “I wish it had certain additional attributes”, etc.
- Publish the same data in multiple open formats – service, download, and web-accessible file (csv, kml, GeoRSS, GeoJSON)
- Design services to be modular and flexible
  - Include only semantically similar layers
  - Separate labels, enable dynamic rendering
- Provide feedback to service owner – engage in constructive dialog

Usage Barriers - Documentation

- A service can look perfect, but still be unusable without accompanying metadata, data dictionary, other documentation
- Even if it is perfectly usable, many users won't trust it without documentation
- EPA Metadata Editor makes authorship easy – success story for FGDC CSDGM metadata, in beta for ISO and Data.gov Common Core
- EPA Metadata Styleguides provide templates, boilerplate language, and map between MXD, REST, ArcGIS Online, and XML elements.
- Service layer metadata often doesn't belong in metadata catalogs, but should be referenced in service layer descriptions
- Cross-linking between REST pages, ArcGIS Online registration and full FGDC metadata is crucial

Usage Barriers – Reliability and Change Control

- Common concerns about services include excessive downtime, poor performance, or unanticipated changes
- Internal solutions
  - Migrated hosting infrastructure to load-balanced Linux cluster – high performance and stability, can perform maintenance without interruptions
  - Service performance often about map authorship – developed best practices checklist based on Esri recommendations and internal findings
  - Developed some change control communication protocols, but also requested enhancements to ArcGIS Server and ArcGIS Online to better facilitate communications between service owners and service consumers
- External solutions
  - Active engagement with external organizations, sharing of best practices and feedback, development of communities of interest

Summary - Organizational Change

- Democratization of map authorship using modular services and lightweight datasets
  - Everyone should know first aid, but we still need doctors
  - Everyone should be able to make a map, but we still need GIS Specialists
- GIS community needs to educate this new user base and review their work
- Data owners need to focus on data quality, multiple publication channels, and documentation
- Application developers need to move away from kitchen sinks, towards lightweight applications that tell a clear and coherent story
- A successful SOA ecosystem requires full participation, engagement, and communication from all parties
- Build organizational governance and success metrics around these goals

Torrin Hultgren
Director, Geospatial Solutions

Innovate!, Inc.
Hultgren.Torrin@epa.gov