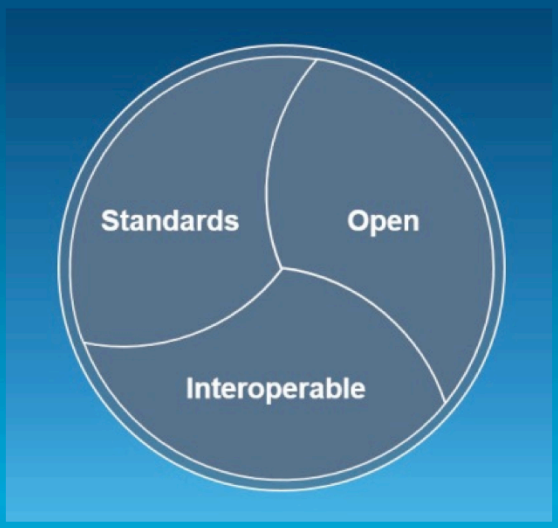
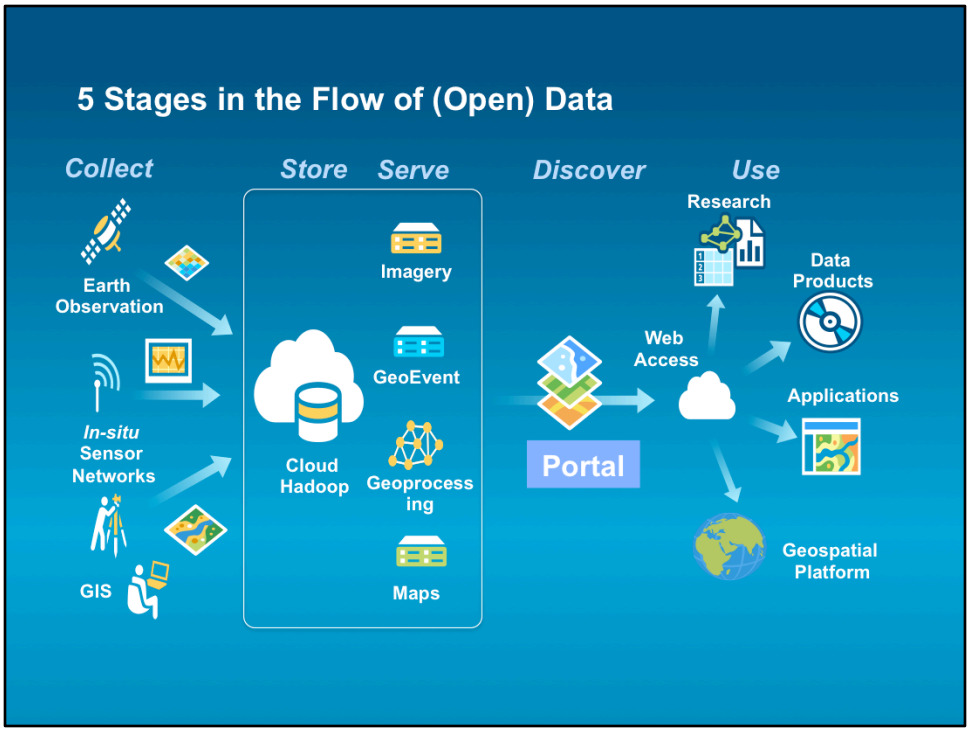
Interoperability
Industry Perspective

Dawn Wright
Esri Chief Scientist

esri

RCN OceanObs Panel, December 14, 2014, San Francisco

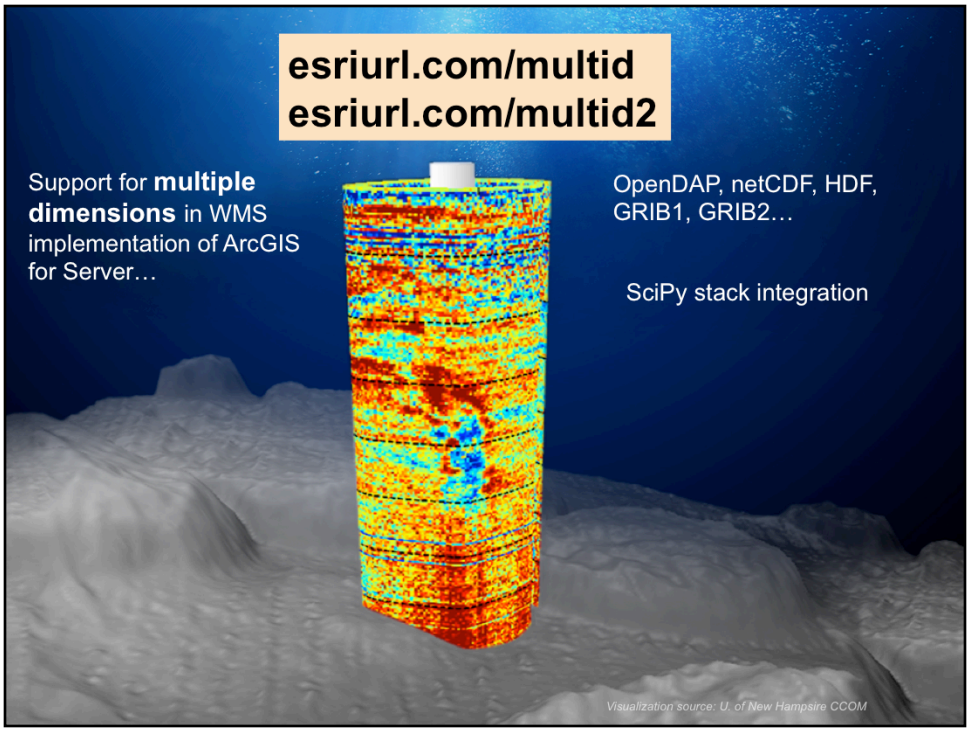

Standards
Open
Interoperable


5 Stages in the Flow of (Open) Data
Collect
Store
Serve
Discover
Use
Earth Observation
In-situ Sensor Networks
GIS
Cloud Hadoop
Imagery
GeoEvent
Geoprocessing
Maps
Portal
Web Access
Research
Data Products
Applications
Geospatial Platform

Interoperability Through Standards & Specifications

COLLECT

- Support for netCDF, HDF, GRIB, OpenDAP
- OGC sensor observation services and WaterML
- Metadata and SDI: ISO TC211, Profiles of ISO TC211, FGDC, and harvest metadata from catalogs (OGC CSW, OpenSearch), INSPIRE



URLs are to **videos**
GRIB = Gridded Information in Binary Form. GRIB is a WMO format for gridded data used by meteorological centers for storing and exchanging meteorological charts and other patterns of wind, sea state, temperature, etc. In other words, GRIB files are computer generated forecast files.

Interoperability Through Standards & Specifications

SERVE

Interoperability Through Standards & Specifications

OGC Services → Geoservices REST (e.g., ArcGIS Server)
SOAP, XML →JSON (everything is a URL)

- WMS (Web Map Service) → Map Service
- WFS (Web Feature Service) → Feature Service
- WCS (Web Coverage Service → Image Service
- WMTS (Web Map Tile Service) → Cached Map Service
- WPS (Web Processing Service) → GeoProcessing Service

Interoperability Through Standards & Specifications

SERVE

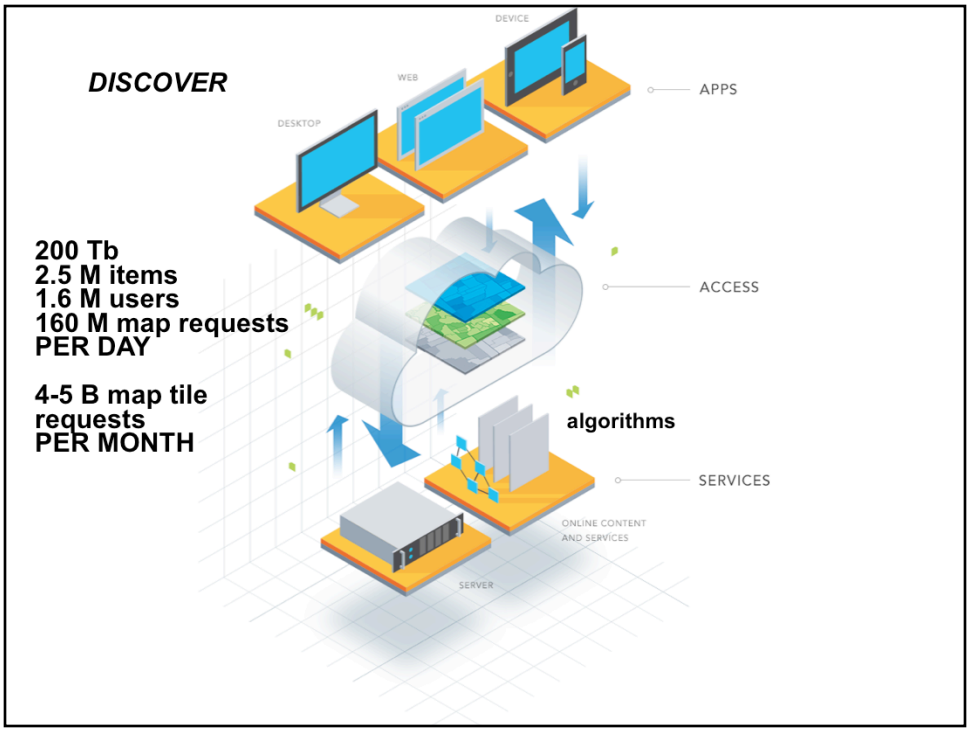
- Support for KML too
- **Python** for scripting, development of custom analytic methods, workflow automation
- Data streaming from different sources using open-source Koop

Geoservice REST specification easy to interrogate and understand
Others are building on top of this – OpenLayers and GDAL creating drivers and libraries for reading and understanding GeoServices
52N built a geometry service using the REST spec
Arc2Earth built out a feature service building on the REST spec
Esri R&D Centers providing ability to work with open source projects such as Leaflet web API – fearure services and map services and working with AGOL
Koop is a middleware server to convert other data types into feature services for data streaming

Interoperability Through Standards & Specifications

STORE

- Apache Hadoop HDFS for storing unstructured data
- Use open-source GDAL library for reading/writing rasters
- Use open-source Computational Geometry Algorithms Library (CGAL) for analysis of 3D data
- Read/Write/Transform hundreds of file formats
- Support various COTS and open-source relational DBMS for storing standard geometry data types



- One example of a discovery platform is ArcGIS Online, www.arcgis.com/home
- **Focused on cloud computing, cloud GIS**
- Apps Are the Window into the Platform
- People who are not GIS professionals can do self-service mapping on any device using focused apps and then make decisions based on the most current, accurate information.
- Esri ArcGIS Online community: 160 million requests per day, 200 Terabytes of data, 2.5 million items, and 1.6M users - See more at: http://blogs.esri.com/esri/esri-insider/2013/09/30/2512/#sthash.NiZx78nX.dpuf

Esri and GEOSS have an MOU
Esri working on a prototype that harvests/connects - CEOS, IOOS, CUASHI, OneGeology, DataONE, EarthCube, AGOL (ibringing in a considerable GIS community into GEOSS)
ArcGIS Desktop with a CSW client can also search the DAB - find a netCDF file and open in ArcGIS desktop
Open-source Geoportal Server (in use in 8 NOAA data centers) used to enhance ArcGIS Online SaaS with specific capabilities related to metadata, federated search, and metadata harvesting.

Interoperability Through Standards & Specifications

USE

- Open-source SDKs and APIs for developers on various hardware platforms and operating systems
- Open-source template apps for info products and custom apps (esri.**github**.io)
- Content mgmt systems: open-source (Drupal, WordPress, Joomla) and commercial (Microsoft SharePoint, IBM WebSphere, Oracle WebCenter, Kentico CMS)
- ArcGIS Online = *geospatial* content management system

See the ocnres URL for info on Esri's ocean GIS initiative
See the scicomm URL to see Esri's involvement in ESIP, EarthCube and GEO.
We are tracking RDA and will be a partner in ODIP2 if its funding continues.
For a complete view of Esri's support for OGC specifications beyond KML, please visit http://www.opengeospatial.org/resource/products/compliant#ESRI.
For a complete view of Esri's interoperability program, see http://resources.arcgis.com/en/help/main/10.2/index.html#/What_is_GIS_interoperability/003700000001000000/.

Extra Slides

GIS Tools for Hadoop

Big Data Spatial Analytics for the Hadoop Framework

esri.github.com/gis-tools-for-hadoop

View project on GitHub

Looking at data without location, most of the time seems like looking at just part of a story. Including location and geography in analysis reveals patterns and associations that otherwise are missed. As Big Data emerges as a new frontier for analysis, including location in Big Data is becoming significantly important.

Data that includes location, and that is enhanced with geographic information in a structured form, is often referred to as Spatial Data. Doing Analysis on Spatial data requires an understanding of geometry and operations that can be preformed on it. Enabling Hadoop to include spatial data and spatial analysis is the goal of this Esri Open Source effort.

GIS Tools for Hadoop is an open source toolkit intended for Big Spatial Data Analytics. The toolkit provides different libraries:

- Esri Geometry API for Java: A generic geometry library, can be used to extend Hadoop core with vector geometry types and operations, and enables developers to build MapReduce applications for spatial data.
- Spatial Framework for Hadoop: Extends Hive and is based on the Esri Geometry API, to enable Hive Query Language users to leverage a set of analytical functions and geometry types. In addition to some utilities for JSON used in ArcGIS.
- Geoprocessing Tools for Hadoop: Contains a set of ready to use ArcGIS Geoprocessing tools, based on the Esri Geometry API and Spatial Framework for Hadoop. Developers can download the source code of the tools and customize it; they can also create new tools and contribute it to the open source project. Through these tools ArcGIS users can move their spatial data and execute a pre-defined workflow inside Hadoop.

is maintained by Esri.

This page was generated by GitHub Pages using the Architect theme by Jason Long.

Hadoop is an open source sw framework that allows you to run clustering and targeting analysis on huge datasets. Finance, online retail, user behavior from search get all the buzz, but Esri is working on *spatial* data, including ocean data (e.g, AIS ship feeds).

Industry is a very important player in the open data space, and practicing interoperability therein. As Maryland Governor Martin O'Malley says, **"cooperation is the new competition."**

PARTNERSHIPS: Ocean exploration has been in the realm of academics and federal agencies, but it will be critical to partner with industry in the DATA SCIENCE space. We are working on solving problems that you need to pay attention to. Be willing to partner with the new, emerging ocean data industry, as well as the instrument providers.
In this new world of an **ocean DATA industry**:
According to PlanetOS, the data acquisition market is currently $80 BILLION including ships, buoys, satellites, AUVS, ocean communication
The data management market will be $5 billion, including software and associated costs

Several ocean IT companies including the industry partners here, are members of the World Ocean Council, a unique international BUSINESS alliance for corporate ocean responsibility, collaborating on issues of stewardship of the seas, including multisectoral voluntary ocean obs/ships of opportunity for science, CMSP, and of course improved data infrastructure, sharing for science. As partners, we are all working on different aspects of data science. MX is even building a marine operating system to streamline big data flows, as well as machine learning tools to improve data quality (e.g., automatically detecting sensor miscalibration, or automatically removing artifacts in ocean satellite data due to cloud cover).

RDA fostering public-private partnerships focusing on data use, data quality
From the NRC Critical Infrastructure report:
It would be beneficial for federal agencies to periodically examine and adopt data management practices that come from beyond the ocean sciences, as well as approaches to grow access to and use of community-wide facilities. Proven efforts from beyond the ocean sciences can be very informative and helpful. Community-specific organizations that focus on data use and data quality will also be valuable to the ocean sciences (e.g., NSF EarthCube, AGU Earth and Space Informatics, Research Data Alliance)