National Geodatabase Replication at the Bureau of Land Management

Jeff Safran – Northrop Grumman
George Heine – Bureau of Land Management

Esri Federal User’s Conference
Building National Spatial Data Infrastructures
Washington, D.C.
January 20, 2011
BLM Public Lands

The Bureau of Land Management's mission is to sustain the health, diversity, and productivity of the public lands for the use and enjoyment of present and future generations.
Why do we need replication?

- The BLM Spatial data tends to be state-centric, stored in geodatabases and shapefiles at the state and local level.
- In order to create seamless, national datasets, the BLM has selected Esri’s geodatabase replication technology to compile data at a national level.
- ArcGIS is the Bureau standard GIS package.
- Capability to create a national dataset with COTS tools and no custom developed applications.
- National datasets can be continually updated and kept current via automated synchronization.
Data Standards

• In order to compile seamless national datasets of geospatial data, each state needs to store geospatial data in the same format, and schema.

• National Data Standards are developed for each national dataset

• Input from all interested users of the data from all BLM states are encouraged to participate in the Data Standards creation process
National Datasets

• Currently the following National Datasets have been compiled using replication to “harvest” data from the state data centers:
  – Grazing Allotments and Pastures
  – Taylor Grazing Act district boundaries
  – Areas of Critical Environmental Concern (ACEC)
  – Land Use Planning Areas (LUPA)

• There are numerous other data standards (20+) in various stages of the Data Standards and compilation process
Replication Requirements

- All Data must be hosted in a state “Publication” ArcSDE enabled database
- RDBMS platform can be any Esri supported platform, allowing the states to maintain autonomy over their infrastructure
- Data in the Parent database must be in the standard schema, and registered as versioned
- The BLM does not use named versions in the national replica infrastructure
- The sde.DEFAULT version is the parent and child version on all databases
- All replicas on all databases are “owned” by the same Active Directory user, a “Service Account”
One way replication

• Each State Data Center is a parent and the National Operations Center is the Child in all replicas
• The NOC only pulls data from the states, and does not edit or send data back to the state publication database
• One Way replicas are simpler and are not likely to lead to corruption of the parent data as we discovered was possible with two way replicas
Mixed hardware/database environment

- RDBMS Platforms
  - Oracle
  - MS SQL Server
- Server Operating Systems
  - Windows 2003
  - Windows 2008 R2
- Citrix vs Desktop
  - XenApp 4.5
  - XenApp 5.0
  - XenApp6.0
  - Desktop Software Only

- ArcGIS Versions
  - 9.2
  - 9.3
  - 9.3.1
  - 9.3.1 SP 1
  - 9.3.1 SP 2
  - 10.0 SP 1
BLM Replication Architecture

7 State Publication Databases
7 NOC State QC Databases (SQL Server)

Typical State Publication

ILMOC(ST) (At NOC)

National Compilation Database (SQL Server)

Typical SDE States

NOC Assisted States

NOC Assisted States (AZ, CO, NV, WY, ES): NOC Will perform QC and pull data directly from state publication database hosted at the NOC. No separate QC Database is required

5 State Publication Databases

5- Replicas (1 per NOC Assisted state) ILMOCSRVRA User account

NOC Asst. States (ILMNVPUB)

ArcGIS Server Geoporta

One-way replica

One-way replica

One-way replica

One-way replica
Lessons learned

- Two-Way Replicas tend to be far more complex and more prone to unintentional data corruption
- Using named versions rather than Default adds additional unnecessary complexity
- Using an Active Directory “Service” account greatly simplifies replica setup and maintenance
- Active Directory Authentication greatly simplifies user access to databases
Data Quality

- As data was compiled from the BLM States into a national dataset, it quickly became apparent that the data was of varying quality.
- Quality Control plans have been developed for each dataset to ensure all data from all states meets a minimum criteria for attribute and spatial quality before being published.
- Geoprocessing models have been developed to perform automated QC upon data Synchronization and deliver reports of errors.
Web Services

• Internally to the BLM, ArcGIS Service Web Services have been created
• These services allow the states to see the “live” picture of the national datasets. This allows the states to see the data from their adjacent states, as well as for national level mapping to be performed
Replication back to states

- Some states have requested a “working” copy of the national datasets in order to perform operations such as snapping to their neighbors and Geoprocessing.

- The NOC has set up one way replicas back to these states of the national datasets, thus allowing the data to travel “full circle” back to the originator.
Future directions

• Geocommunicator.gov
  – At some point in the near future, the data that has been compiled from the states via national geodatabase replication will become the data presented to the public via Geocommunicator.
  – This will be as simple as another 1-way replica being created to the back-end ArcSDE database for Geocommunicator

• ArcGIS 10
  – The BLM plans to upgrade to ArcGIS 10 in mid-2011
Questions?

• Contact Information:
  – Jeff Safran
    • National GIS Software Support
    • Northrop Grumman
    • (303)236-2315
  – George Heine
    • USDOI BLM
    • (303)236-0099