Building a National SGCN Dataset

Moving form 56 Disparate Plans to One Integrated Product
Presentation Overview

- NBII Species Mashup
- LIVE DEMO
- Species of Greatest Conservation Need (SGCN) as the first production – level implementation within the Species Mashup framework
- LIVE DEMO
What is a “Mashup?”

“Web page or application that uses or combines data or functionality from two or many more external sources to create a new service” …Wikipedia

What is the NBII Species Mashup?

A technical framework – database, content management tool and application interface – that enables rapid and efficient integration of distributed data sources related to species via web services or other Web 2.0 methods
NBII Species Mashup – Main Components

- **Species** – Scientific Names (CN’s optional)

- **Data Sources** – Scientific Names are connected to Data Sources (ITIS, Natureserve or any relevant regional or local data source) by cross-walking names to unique IDs

- **Tags** – Scientific Names can be associated with Tags. These Tags or Topics are used to create customized views of the Mashup.
Within this framework, a user, (most likely an NBII CM) needs to be able to edit the three main components and then connect them together.
NBII Species Mashup – Pilot Project

Pilot implementation

- Hawaii Federally Listed T & E Species

Process

- Determine Hawaii T & E species list
- Upload species names and associated tags to database
- Identify and establish data sources
- Cross-walk and upload species name to data source connections
- Create data views, web services and APIs for local data sources to facilitate data harvesting and display
NBII Species Mashup – The Dashboard

- **Species Section** – Upload scientific names and tags
  - Required format
  - error checking
  - validation

- **Tags Section** – Set up customized mashup views
  - Browse Panel – custom left-column in application interface (taxonomic and spatial filtering)
  - Data Sources – set data sources that display in that particular mashup; set order of display
NBII Species Mashup – The Dashboard

- **Data Source Section** – Establish data sources
  - Connect mashup application to web services and code that will enable the data to be harvested and displayed
  - NBII programmers have established many relevant national level data sources (taking advantage of existing web services and creating new ones) for consumption in mashup
    - ITIS, NatureServe, GBIF, NBII, Google

- **Data Source UIDs** – Upload cross-walk of UIDs to names
NBII Species Mashup – Summary

What does the NBII Species Mashup mean to me?

- Easily integrate and meaningfully present species information
- Allows me to focus on the most important part of my job
SGCN Mashup – Team Structure

Mashup Working Group

- 8 – 10 people from several NBII Nodes (led by Jean Freeney)
- Made up of program managers, content / data managers and programmers
- Focus on collaboration and collective decision making

Content Sub-group

- 3 – 4 people focused on content / data management
SGCN Mashup – Main Objectives

- Create SGCN mashups for public consumption at multiple spatial scales
  - State, region, nation
  - Implementation on NBII regional node websites

- Create a value-added base product that enables comparison and analysis across states
  - Regional or national SGCN distribution
  - Gap analysis
SGCN Mashup – Early Considerations

What factors affected implementation?

- Technical architecture of mashup system
  - Database structure; connecting names to data sources
  - Unique ID and linking field is Scientific Name text string

- Presentation as a public product
  - Establish general nomenclature standards to promote consistency (CN capitalization, ssp. vs. subsp. etc..)
SGCN Mashup – Early Considerations

What factors affected implementation?

- SGCN species name sources
  - Do we interact with 56 government entities in the beginning or the end?
  - Use readily available WAP sources (PDFs) on the web (but case by case basis)

- Species name extraction from WAP documents
  - Can we do this in an automated fashion using a tool like UBIO?
  - Not really! Must manually extract
SGCN Mashup – Implementation

Desired Product

- National list of SGCN species that consists of unique (no text string or taxonomic duplicates) scientific names

General Methodology

- Established business process, content management and specific data element standards and guidelines

  - Excel template with embedded verification and QAQC tools
  - Field level data standards
  - Standardized taxonomic, spatial, and topic tags
  - Standardized QAQC process
SGCN Mashup – Implementation

Specific Methodology

- Content Group divided up states; obtained copy of WAPs
- Determine SGCN species in each plan
- Extract scientific and common names into Excel template
- Copy data into new template and archive original extraction
- Assign standardized taxonomic, geographic and topic tags
SGCN Mashup – Implementation

Specific Methodology Cont’d

- QAQC process
  - Cursory review of the entire list
  - Import the state list into a master list
  - Detailed review of names in common across states
  - Resolve errors, inconsistencies and nomenclature issues in scientific names across states
  - Document and verify any changes made to the list content
  - Follow-up with states on any unresolved issues
Methodology Notes

- 100% taxonomic correctness not the driving factor
- Data aggregation and data resolution was the focus
- Intelligence in the process was driven by comparison of all SGCN lists
- Pretty sure mistakes were made (23000 records)
- Currently in verification phase
SGCN Mashup – Summary Stats

Total Unique SGCN Species in Nation:

- App. 12740

Total Species Records from all 56 Plans:

- App. 23410

Increase in the Number of Grey Hairs on my Head:

- Too many to count!
SGCN Mashup – Lessons Learned

See Document
NBII Species Mashup – Live Demo

LIVE DEMO
Mahalo!

Sky Harrison
USGS / NBII / PBIN
sharrison@usgs.gov
808-984-3722