NatureServe

A Network Connecting Science With Conservation

Approaches and Difficulties in Filtering EO Representations Based on Spatial Precision

Northeast Natural Heritage Conference

April 20-21, 2009

Why do we need to be able to identify “Imprecise” data?

- **Some data can be imprecise to the point of not being informative.**
  - **Precise data can be supplemented with lists of species by county/watershed to account for nearby species.**
- **Sometimes it is desireable to “weight” data based on mapping precision for landowner analyses.**
- **Can help prevent reporting species where they obviously don’t belong.**
- **Can affect how you approach fuzzing data for an analysis**

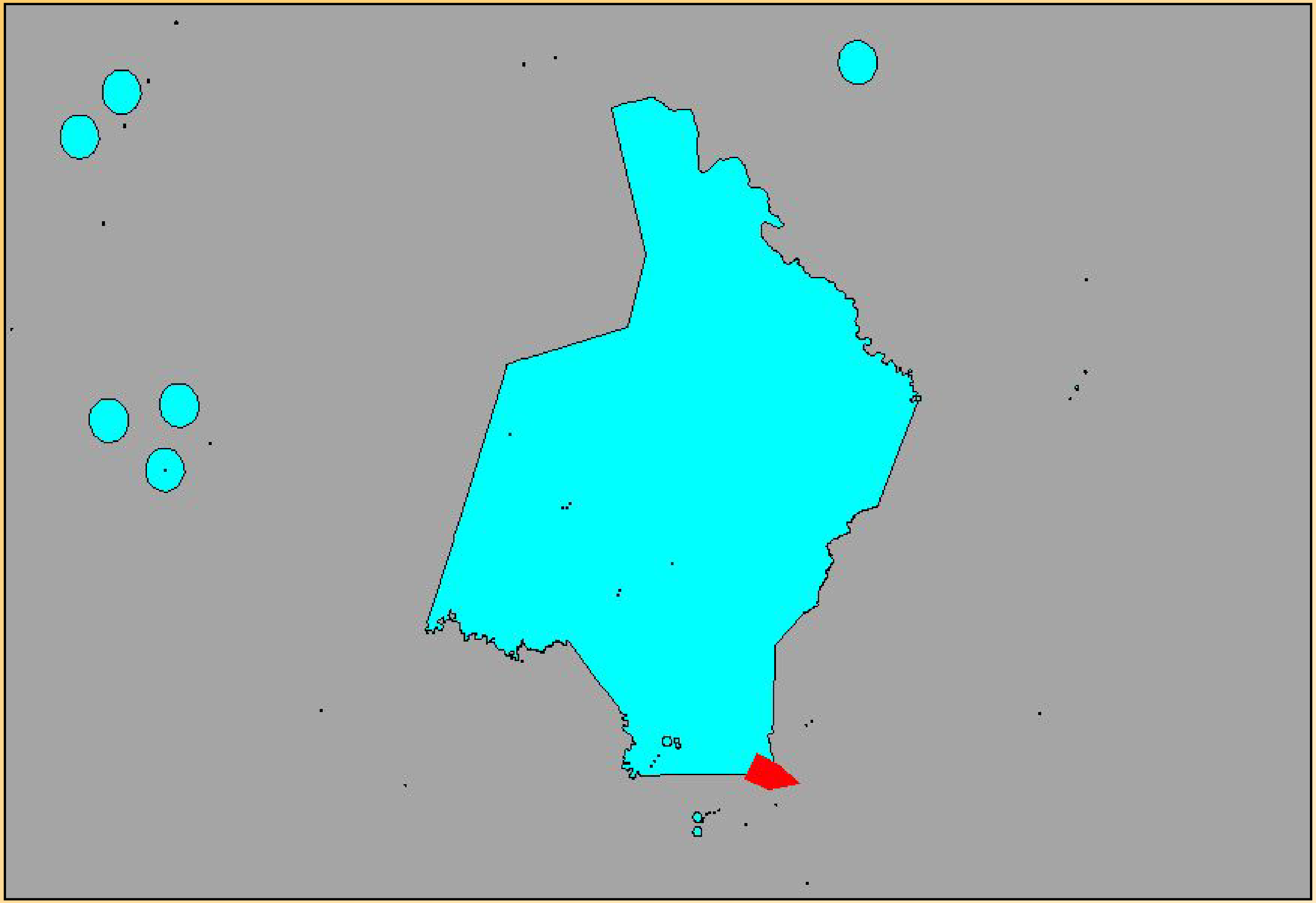
NatureServe

Example 1 – Obvious Misreporting of Species

Species endemic to sand dunes

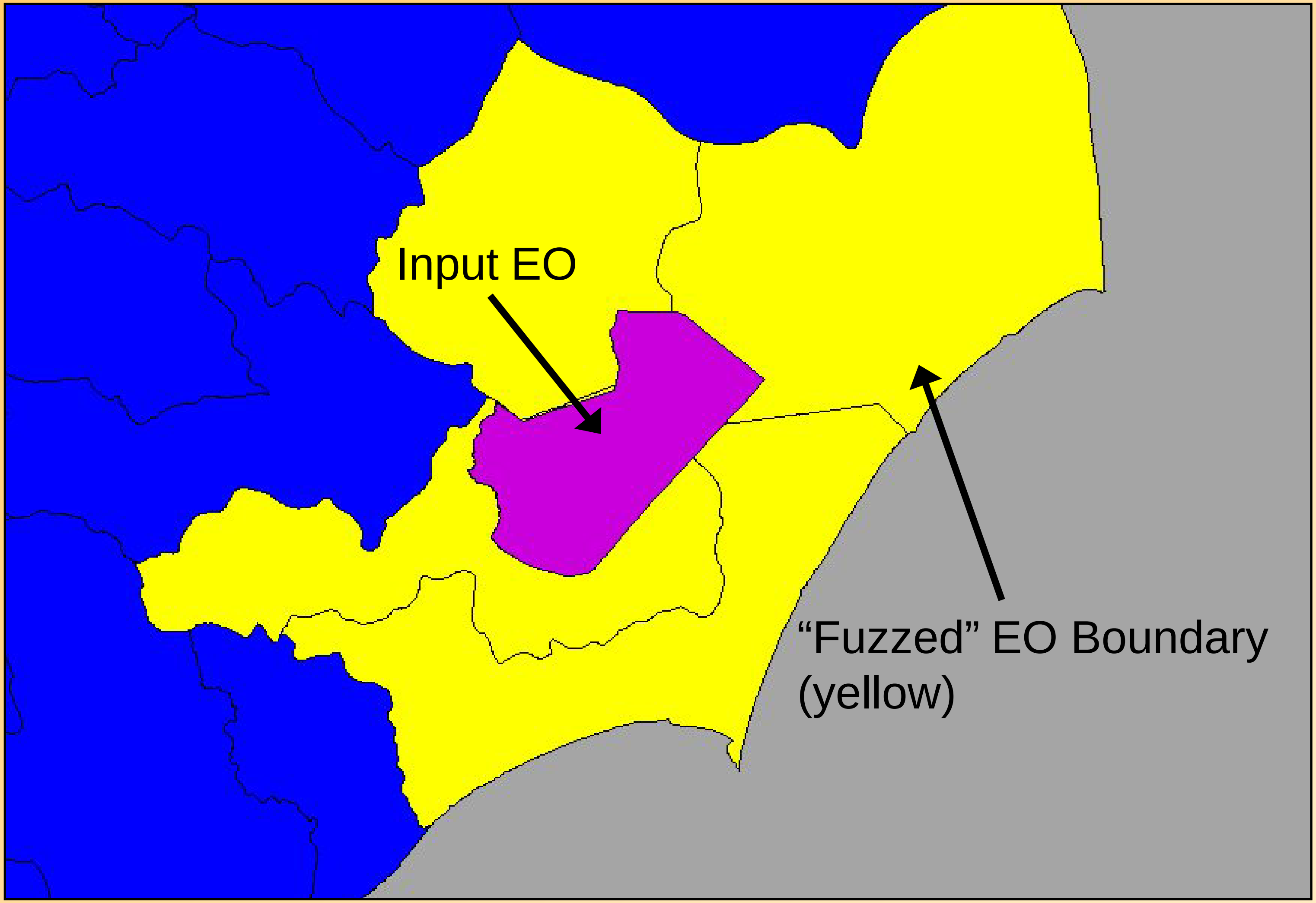
NatureServe

Example 2 – Potential Effects on Landowner Analyses



NatureServe

Example 3 – Affect of Imprecise Data on Fuzzing



“Imprecise” Data – how do you identify it?

Precision Attributes – Then and Now

- **BCD Age:**
  - **Precision (Seconds, Minutes, General, Unmappable, Etc.)**

- **Biotics Age:**
  - **Representation Accuracy (Very Low, Medium, Very High, Etc.)**
  - **Locational Uncertainty Type (Areal Estimated, Areal Delimited, Negligible, Etc.)**
  - **Locational Uncertainty Distance (meters, miles, kilometers, etc.)**

Representation Accuracy

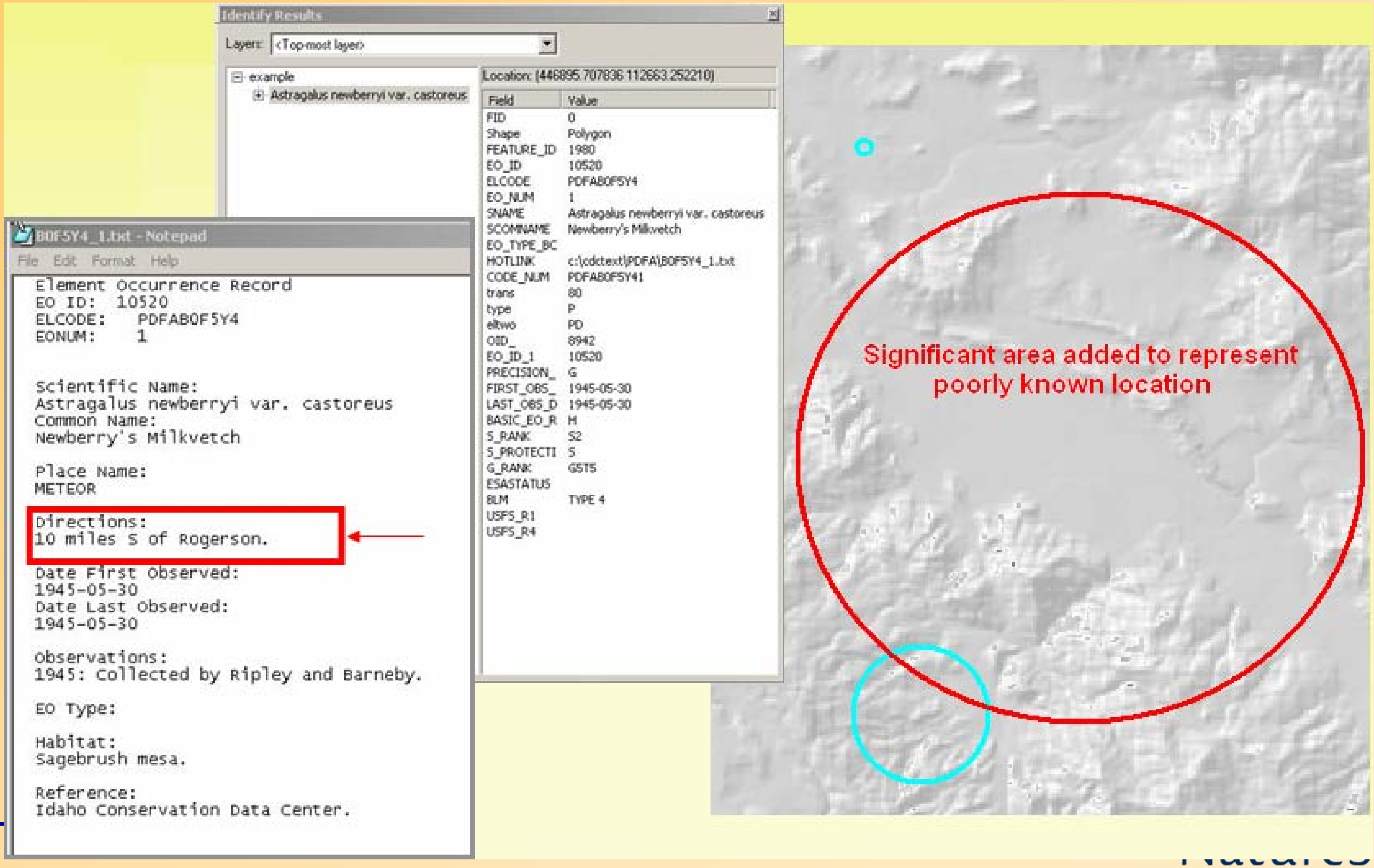
- Value selected from a drop-down menu that indicates the level of accuracy associated with the Element Occurrence Representation (EO Rep). Accuracy varies on the basis of area observed to be occupied by the Element relative to the area contained within the footprint of the EO Rep. Differences in these two values result from incorporation of additional area within the EO boundary to incorporate associated locational uncertainty. Representation Accuracy (RA) provides a common index for the consistent comparison of EO Reps, thus helping to ensure that aggregated data are correctly analyzed and interpreted.

- Domain values for Representation Accuracy Value are
- Very high
- High
- Medium
- Low
- Very Low
- Unknown
- (null) - Not assessed

NatureServe

Low RA

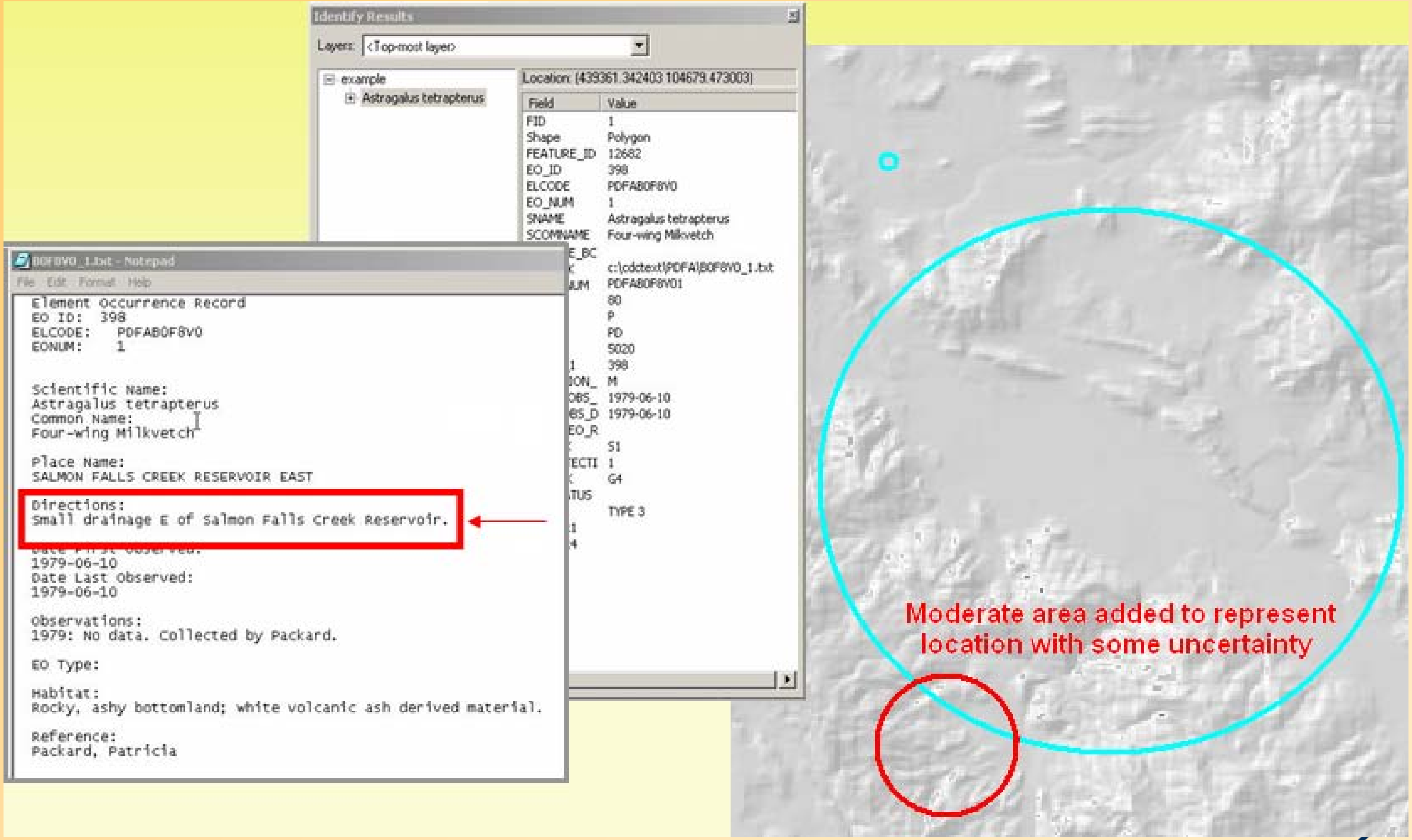
Vague locational data results in significant area added to represent uncertainty and low accuracy.



NatureServe

Medium RA

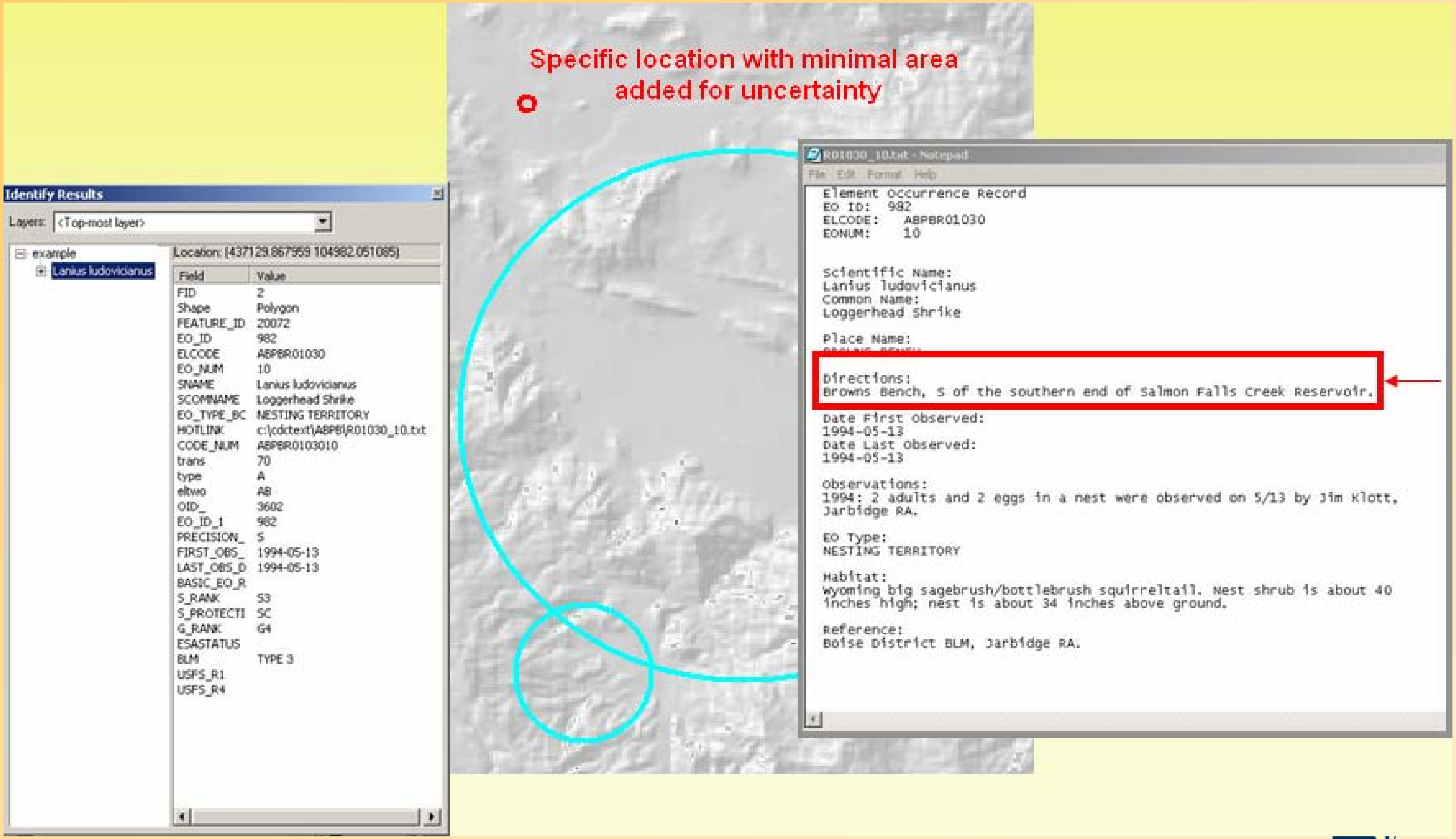
As the quality of locational data increases, accuracy increases.



NatureServe

High RA

Very specific locational data allow precise mapping with high accuracy.



NatureServe

Pros and Cons of Using Representation Accuracy for MJD Projects

- **Pros:**
  - **Single field (like BCD Precision)**
  - **Standard methodology (in theory)**
  - **Standard domain values**

- **Cons:**
  - **Not filled out for a majority of records**
  - **Not 100% sure if it’s been filled out consistently**
  - **Aggregate value of Source Feature accuracy (OK when just one Source Feature but can be a problem when there are many SFs.**
  - **Not sure if it’s totally reliable (i.e. are some programs not using it, or tracking “precision” with other fields?)**

NatureServe

Spatial Precision – Filtering Methods

1. **Using RA value only**
2. **Using RA value if it is populated, and Locational Uncertainty when RA is null.**
3. **Area review (where RA is null)**
4. **Weighting based on spatial precision**
5. **Determining presence based on overlap percentage**

- **Drawbacks: 2 through 5 can be bulky and somewhat subjective.**

NatureServe

Filtering Methods - Using RA Value Only

- "You either got it or you don't."
- Because RA is filled out for a minority of records, leaves a large number that are labeled "unknown".

Attributes of Fote_eo_polys_041708.shp

Survey_dat	Lu_summary	Est_rep_ac	Rep_ac_com	Conf_x_cd	Conf_x_des
1982-05-14	Areal - Delimited / (1)	Very High		Y	Confident full extent of EO is k
2005-09-16	Negligible / (2)	Very High		?	Uncertain whether full extent
2005-04-10	Areal - Estimated 100/M (2)	Very High		?	Uncertain whether full extent
2002-09-16	Areal - Estimated 50/M (2)	Very High		?	Uncertain whether full extent
1968-04-28	Areal - Estimated 200/ (1)	Very High		N	Confident full extent of EO is I
2004-04-07	Negligible / (1), Areal - Estimat	Very High		N	Confident full extent of EO is I
1995-06-24	Areal - Estimated 400/ (1)	Very High		N	Confident full extent of EO is I
1980-09-26	Areal - Delimited / (1)	Very Low	G precision record updated to i	?	Uncertain whether full extent
1995-06-09	Areal - Delimited / (1)	Very Low			

NatureServe

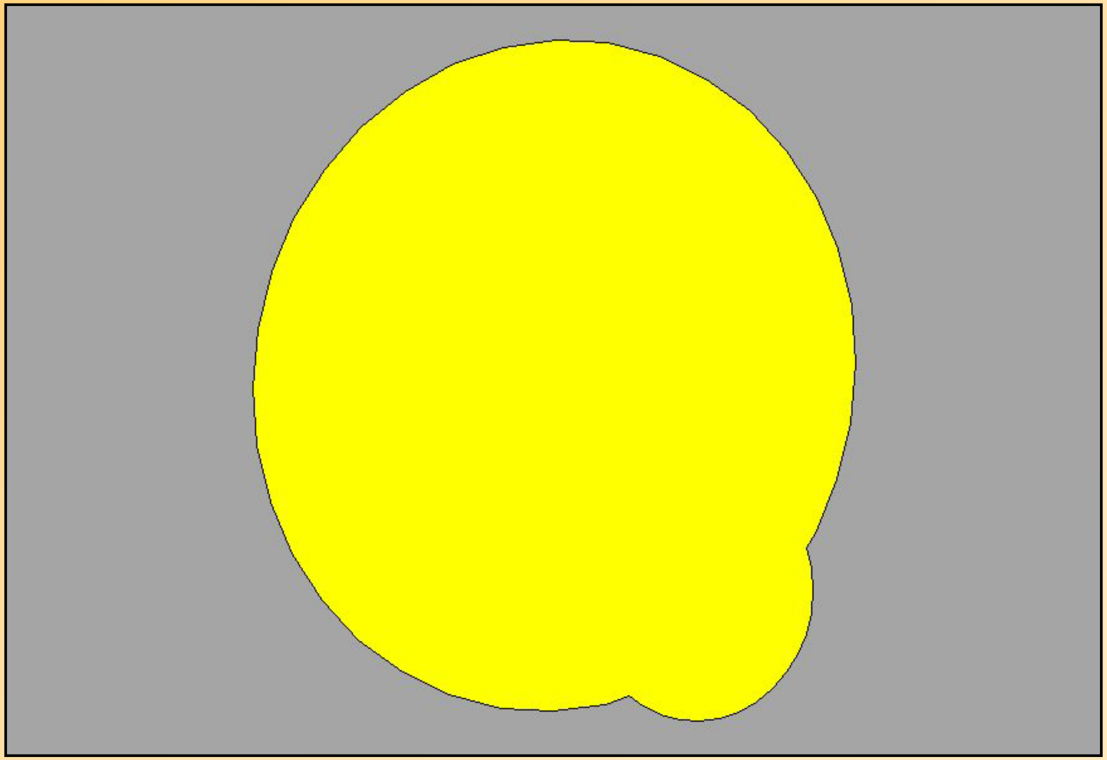
Filtering Methods - Using RA and Locational Uncertainty

- **Use RA for records that have it, and set criteria based on LU for those that don’t.**
  - **example: select records where ((RA = “Very Low”) or RA is null and (LU = Areal Estimated and LU Distance > 8,000 m))**
- **Problems**
  - **Works well for records based on one Source Feature, or where all SFs are of the same LU Type, but not when EO is based on multiple SFs of multiple LU Types**
  - **Sometimes LU filtering logic doesn’t hold up**
    - **example: filter out EOs that are Areal Estimated and greater than 8,000 m LU distance, but then find records with these same values that have an RA of “High”.**
  - **Time Consuming**
    - **Lots of manual review; iterative process**
    - **Some programs use different or multiple LU Distance Units**

NatureServe

Filtering Methods - Using RA and Locational Uncertainty

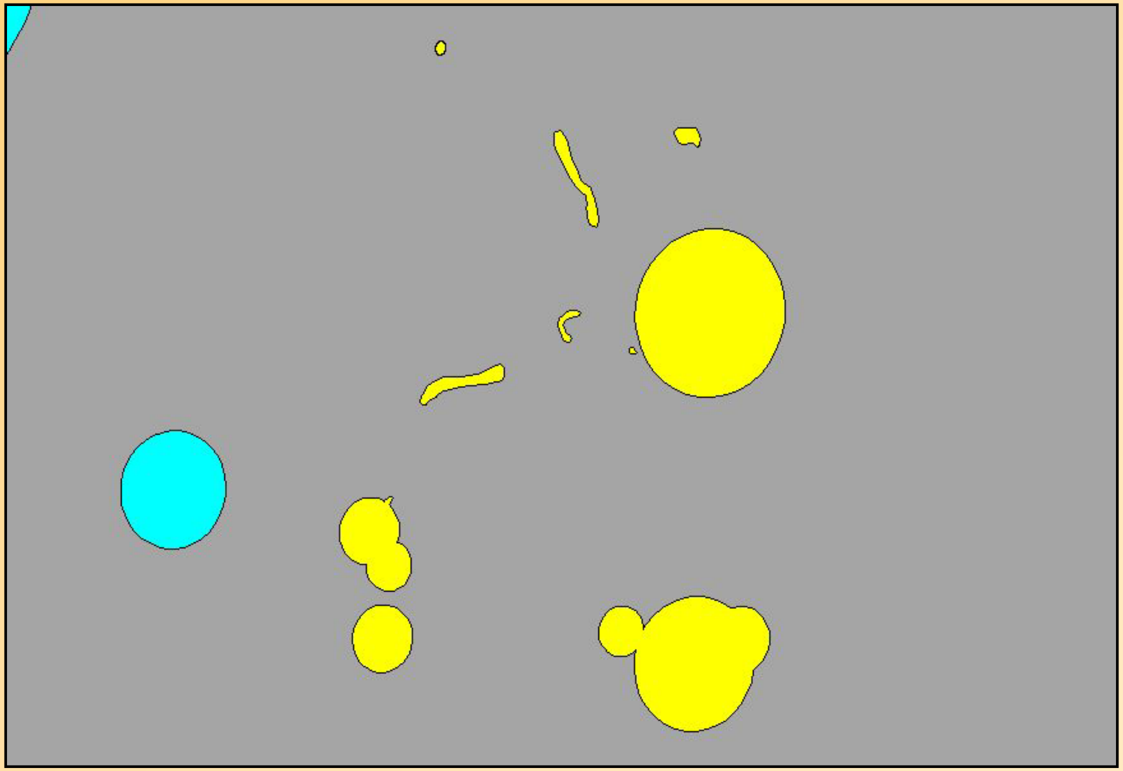
- **Problem Example 1:** Locational Uncertainty Info = Areal - Estimated 4000/ (1), Areal - Estimated 1500/ (1), Areal - Estimated 800/ (1)



NatureServe

Filtering Methods - Using RA and Locational Uncertainty

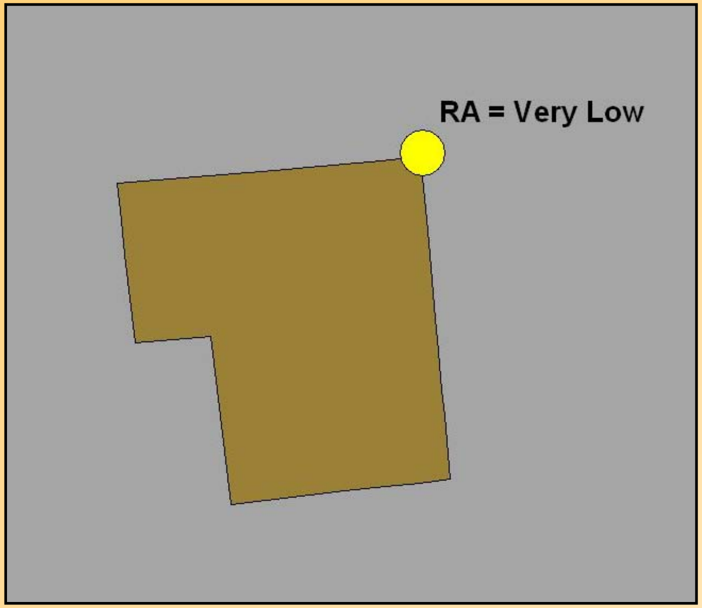
- **Problem Example 2:** Negligible / (7), Areal - Estimated 3000/M (2), Areal - Estimated 4000/M (3), Areal - Estimated 10000/M (1), Areal - Estimated 8000/M (1)



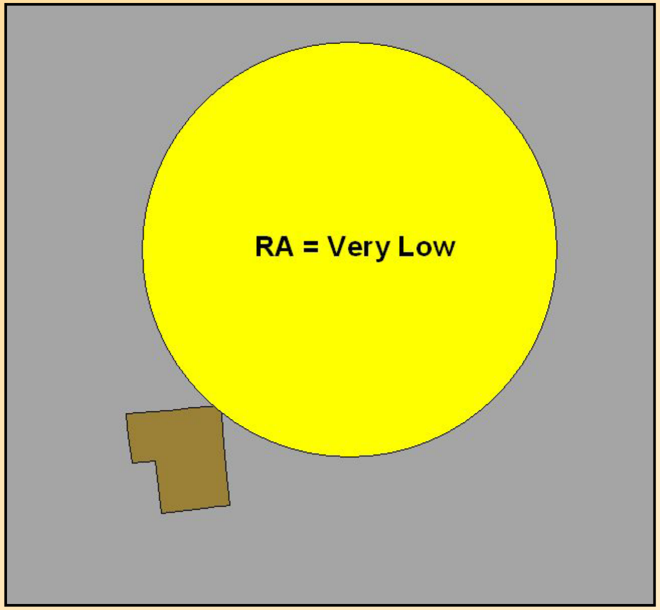
NatureServe

Filtering Methods - Using RA and Locational Uncertainty

- **Another problem is that there is some threshold at which the size of an EO relative to an overlay feature can be more important than spatial precision of the EO itself. This can be difficult to determine and apply.**

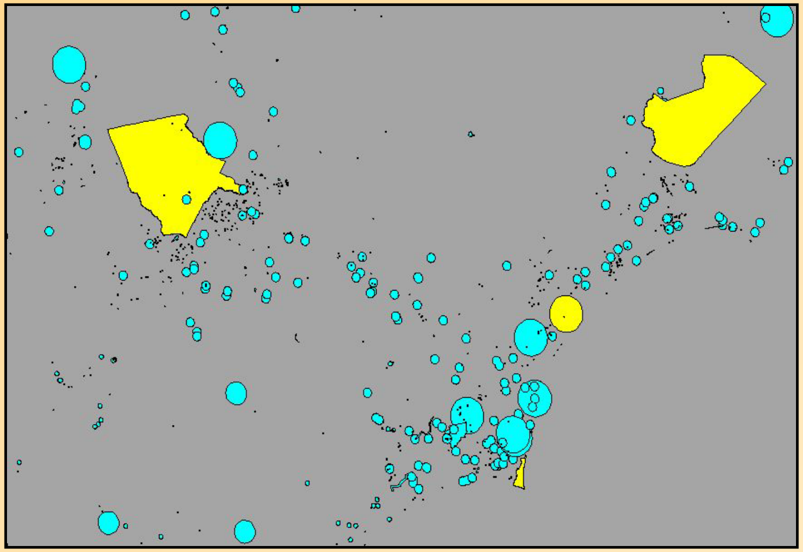


VS.



Filtering Methods – Area Review

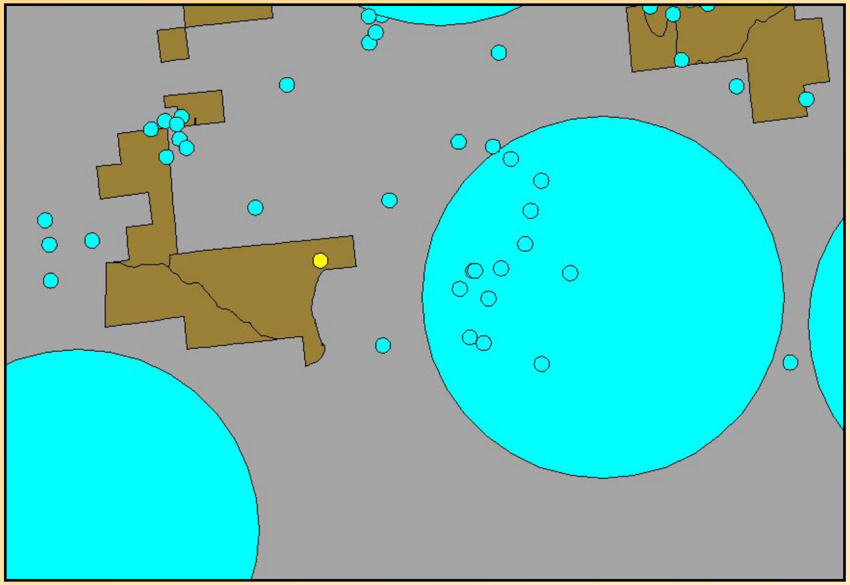
- **Find all shapes larger than "X" square miles and review their attributes to try and assign a precision, and/or send to NHPs for review.**
- **Problems:**
  - **Size "cutoff" can seem arbitrary**
  - **Size alone is not confirmation of an imprecise record**
  - **Time Consuming – manual process; need to build in funding and turn-around time for program review.**



NatureServe

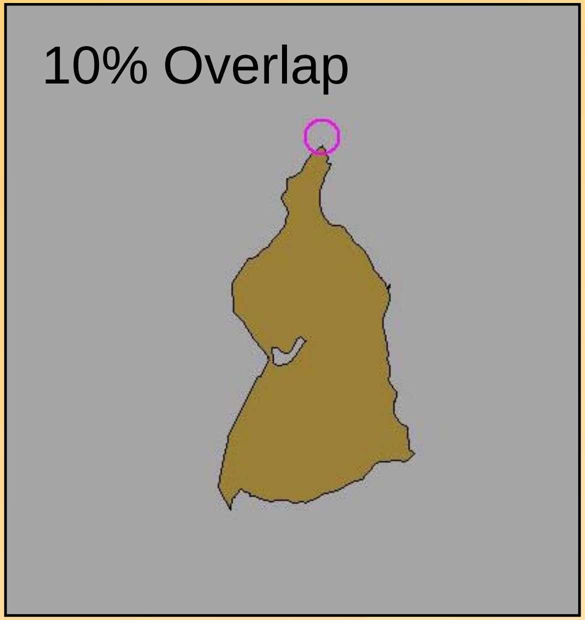
Filtering Methods – Weighting

- **Assign a spatial precision category based on a combination of RA, LU, Area Review, and geographic overlap.**
- **High, Medium, Low, Unknown, and "Subsumed"**
- **Subsumed EOs are completely enclosed by an analysis input feature, which trumps the RA and LU values.**

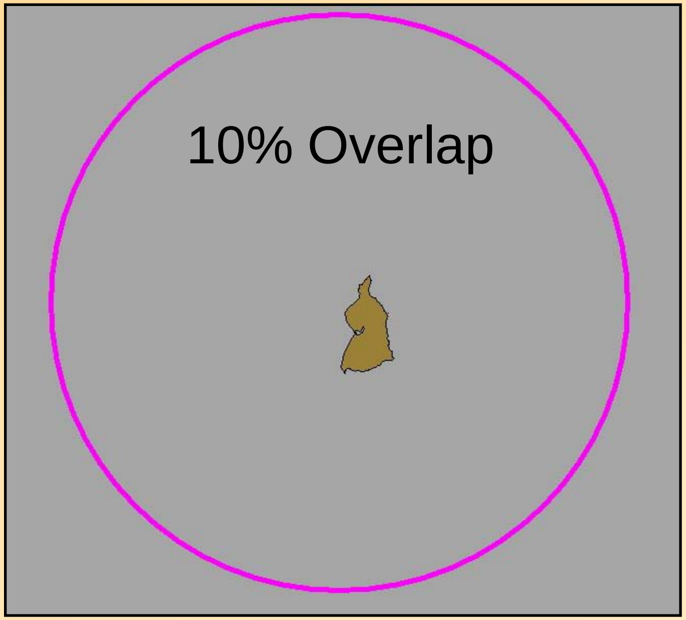


Filtering Methods – Overlap

- **Accepting or rejecting the presence of a species in a project area based on percent overlap.**
- **Problems**
  - What is the “correct” overlap cutoff percentage?
  - The total area of the EO vs the total area of the overlap polygon can be a factor that needs accounted for.



VS.



Possible Solutions for Identifying Imprecise Data

- **Automate the population of Representation Accuracy.**
  - **Write a script that automates the RA logic and assigns an RA value to an EO based on Source Feature attributes where possible, and flags EOs that require a manual review.**
- **Create a new "Imprecise" flag field**
  - **Programs know their data best, and since programs may be flagging things different ways, this would be a "Y/N" field that tells NatureServe if a program deems a record imprecise based on their judgement.**
- **There isn't one – what we're doing now is the best way.**
- **None of the above – Other Ideas?**

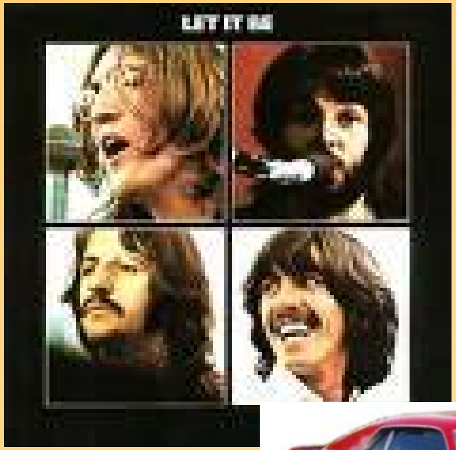
NatureServe

Open Discussion

- Is identifying imprecise data an issue for any programs in the room?
- If so, how are programs handling this locally?
- Are there any ideas for other ways to handle this issue that have not been mentioned?
- General comments/questions?

Bonus Slide!

- **What are the arguments for, and against, continuing to use 1970 as a cutoff year for filtering out EO data based on Last Observed Date?**



NatureServe