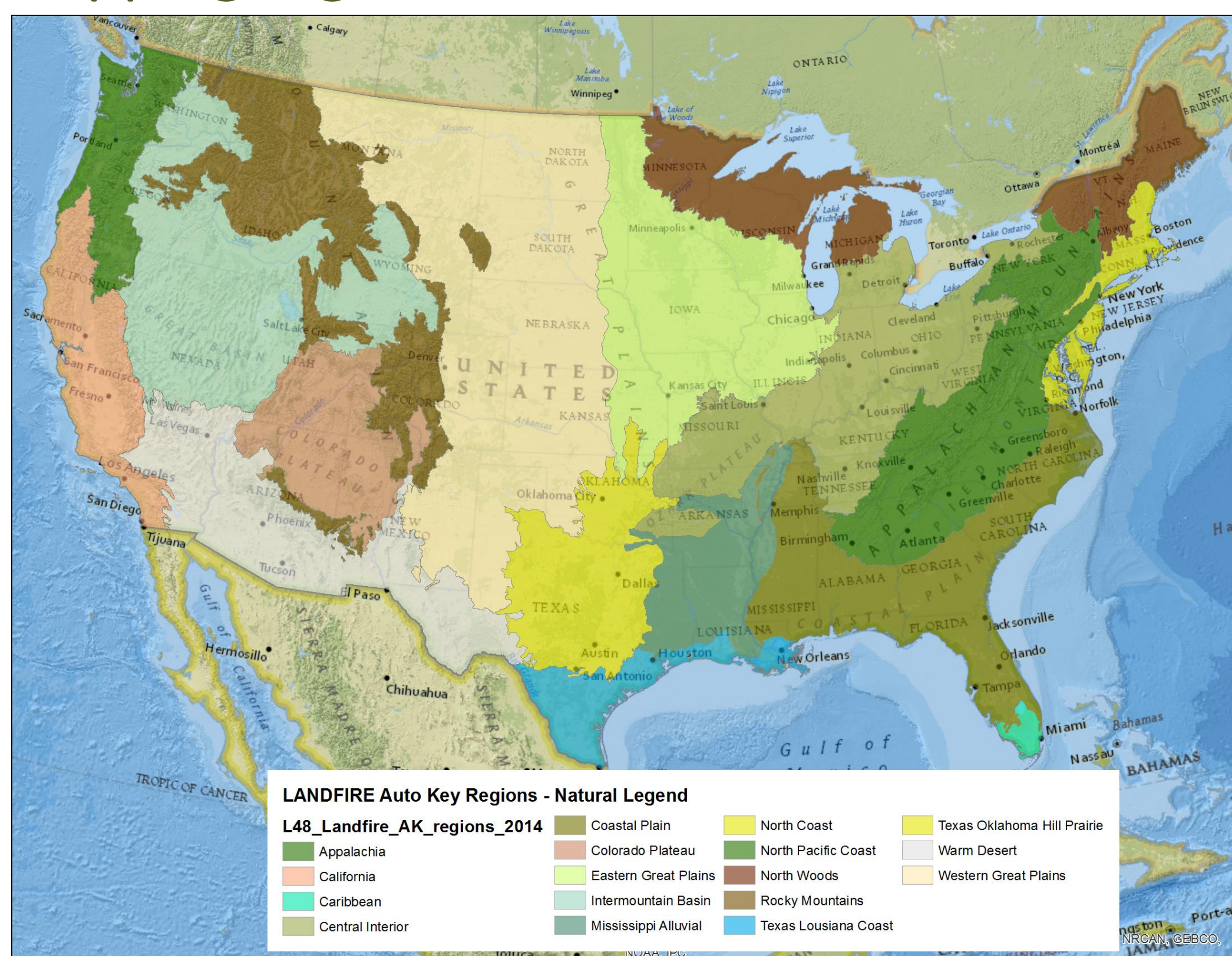


## Automated Vegetation Sample Labeling Tools and their Application to National Mapping.

Patrick Comer & Marion Reid, NatureServe and Donald Long, USDA Forest Service.

### Background

Rapid progress is being made in national-scaled vegetation mapping with standard classifications, mapping methods and tools, and vegetation sample data. Inductive modeling methods rely on vegetation sample data, and currently over 500,000 samples, each labeled to the target map legend class, are used for mapping nationally. In order to integrate sample data from across multiple sources, and consistently label each to a standard classification concept, automated labeling tools are required. This poster describes an automated labeling process and tools that have been developed and refined for use in mapping regions of the conterminous USA.



### Sample Data

• LANDFIRE Reference Database includes sample plots with field-collected & map-derived attributes:

- Coordinates
- Species composition
- Species abundance
- Stand structure
- Date sampled
- Ecoregion
- Elevation
- Landform & bioclimate
- Expert applied labels for validation

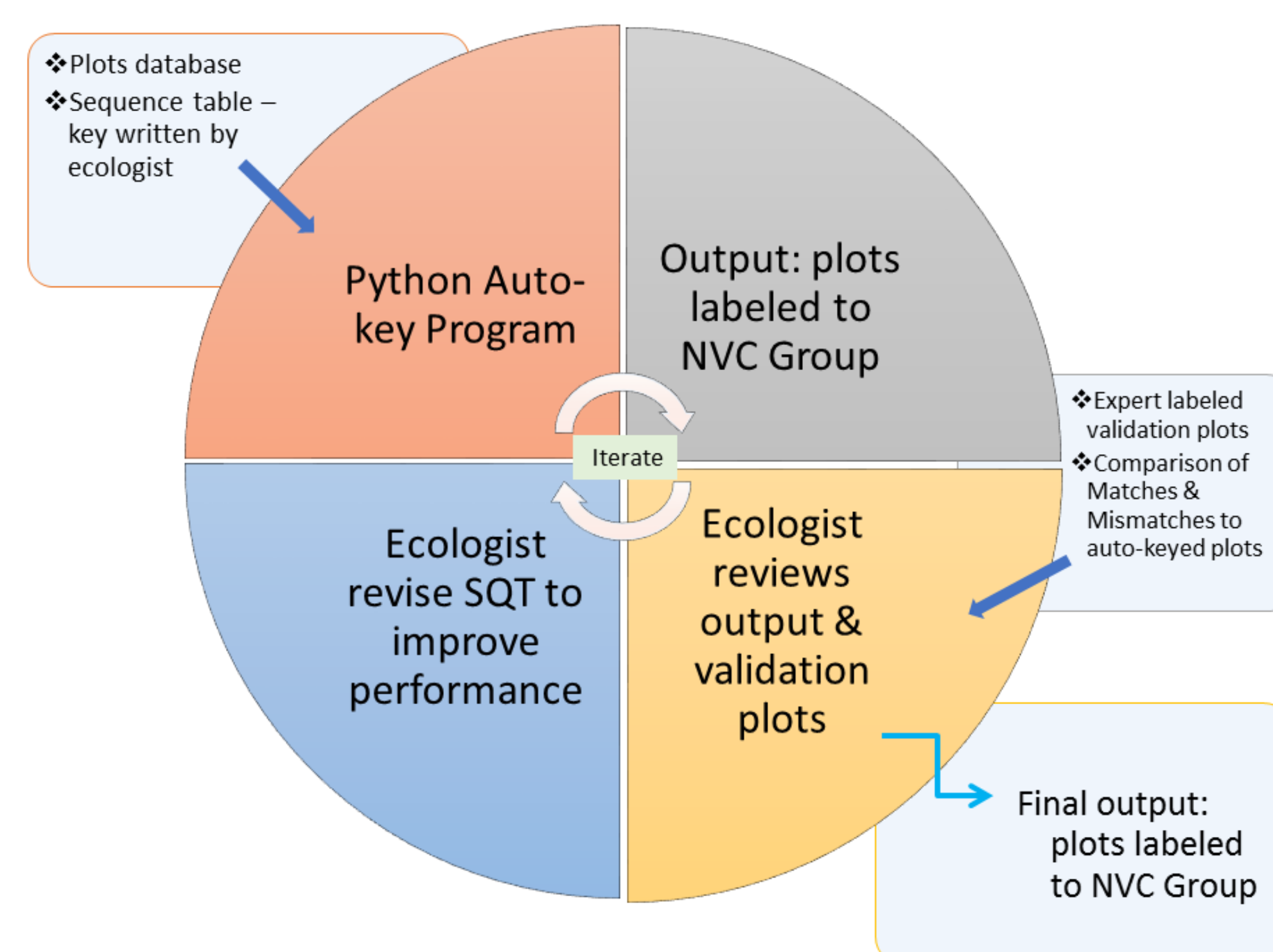
• Plots and attributes are provided for keying in a MS Access DB

### Sequence Table

Row #	Mapping Code Group Name	Total absolute cover criteria, by lifeform				Species 1 Relative cover 1	Species 2 Relative cover 2	USFS Ecoregion Elevation (m)		
		All	Tree	Shrub	Pete Herb					
210	6094 North Pacific Mountain Hemlock - Silver Fir Forest & Tree Island	>10%	>10%	>25%	>10%	Abies amabilis, Abies procera, Callitropsis nootkatensis, Tsuga mertensiana	>30%	Abies magnifica	M242A, M242B, M242D	>1450
215	6096 Sierra-Cascade Red Fir - Mountain Hemlock Forest	>10%	>10%	>10%		Abies concolor, Abies magnifica, Abies procera, Picea breweriana, Pinus monticola, Tsuga mertensiana	>50%	Abies concolor, Abies magnifica		>1600

- Sequence table is written for each of the regions in the map to the left
- Number of types ranges from 20 to over 100; number of criteria rows in the largest tables approaches 250
- The plots are tested sequentially against each row of criteria and must meet ALL of the criteria to be labeled to the type in the row; i.e. if the plot doesn't meet ALL row 210 criteria it is tested against row 215; and so on.
- The logic for criteria is an AND between columns (e.g. >10% total cover of vascular plants AND >10% cover of trees AND >25% cover shrubs) and an OR within columns, where there are lists (e.g. of several species or ecoregions)
- For more information: [https://www.landfire.gov/documents/NatureServe\\_Auto-Keys\\_Report\\_December\\_2015.zip](https://www.landfire.gov/documents/NatureServe_Auto-Keys_Report_December_2015.zip)

### Auto-Keying Process



- ~17,900 expert labeled validation plots are available for the conterminous USA
- Queries result in comparison of expert to auto-key labeled plots; plot labels are compared as by-type matches or mismatches and statistics (e.g., %matches) generated for review by-type
- Key development is iterative, and subsequent new samples can be input and labeled as they come available

### Results

Auto-Key Region (AKR, see map)	# of Plots	# of expert labeled plots	% of expert to total plots	Overall validation agreement	# of natural Groups
Western Great Plains	27,337	1,686	6%	69%	78
Inter-mountain Basins	50,034	1,992	4%	81%	77
Rocky Mountains	80,478	3,390	4%	70%	77
North Pacific Coast	44,929	1,722	4%	73%	68
Warm Desert	17,338	705	4%	86%	63
Colorado Plateau	23,735	527	2%	88%	61
California	37,862	1,965	5%	76%	52
Appalachia	22,999	1,171	5%	56%	48
North Woods	50,195	1,058	2%	68%	39
Coastal Plain	20,221	937	5%	40%	35
TX-OK Hill Prairie	4,043	552	14%	90%	33
Eastern Great Plains	5,691	513	9%	80%	32
Texas-Louisiana Coast	7,087	259	4%	87%	32
Central Interior	26,021	756	3%	79%	30
North [east] Coast	3,517	259	7%	78%	28
Mississippi Alluvial	5,822	442	8%	78%	21

Auto-Key labeling performance varies by region with a low of 40% along the Atlantic and Gulf Coastal Plain to a high of 90% in the TX-OK Hill Prairie. Here no "fuzzy" assessment of accuracy was used, and so some error is likely explained by confusion among similar types.

### Discussion

In order to map at regional and national scales, aiming for 30m pixel resolution spatially, and US NVC Group (or similar) thematic resolution, it is now clear that a current data set approaching 1M samples is needed for the conterminous USA. The LANDFIRE reference database is the first attempt to compile comprehensive georeferenced vegetation sample data for the United States.

Auto-Keys were designed for application within a given region defined using Forest Service ECOMAP ecoregions. This facilitated efficient treatment of all types most distinct from neighboring regions. Regional expert review of a subsample of vegetation plots enabled iterative evaluation and of improvement of each autokey and reporting of key performance. Continued challenges to automated sample plot labeling can be tied to the quality and completeness of sample plot data and remaining uncertainties in current knowledge to distinguish among relatively similar vegetation types.

Auto-keys are an innovative method for rapidly and efficiently labeling thousands of vegetation samples. They should facilitate efficient utilization of field sample data for producing high-resolution vegetation-based map products for a wide variety of conservation applications.

Technical Report citation: Reid, M., P. Comer, B. Lundberg, J. Smith, J. Drake, D. Faber-Langendoen, M. Hartness, G. Rittel, S. Menard, C. Nordman, M. Pyne, K. Schulz, L. Sneddon, and J. Teague. 2015. Developing auto-keys for LANDFIRE Vegetation Mapping: 2014-2015 CONUS Project Report. Report prepared for Inter-Agency LANDFIRE Program by NatureServe, Arlington VA. 69p + Appendices and Data Tables.

Key partners for this effort:

