A protocol for linking textual locality descriptions to high resolution spatial polygons data: a method for incorporating spatial uncertainty into biodiversity models

Stephen J. Murphy, PhD

Kelley Erickson and Adam B. Smith

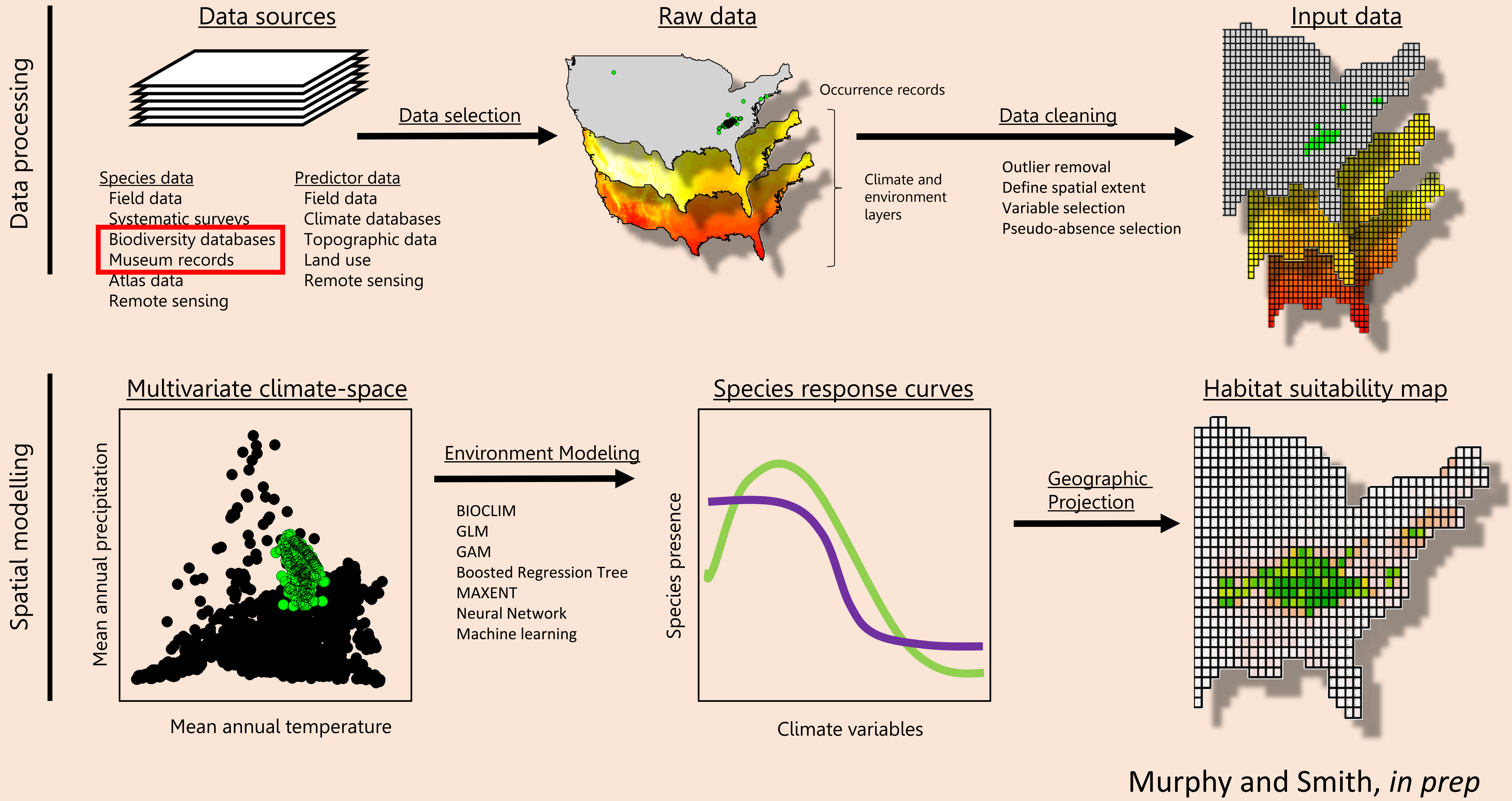
Missouri Botanical Garden

06/11/2019

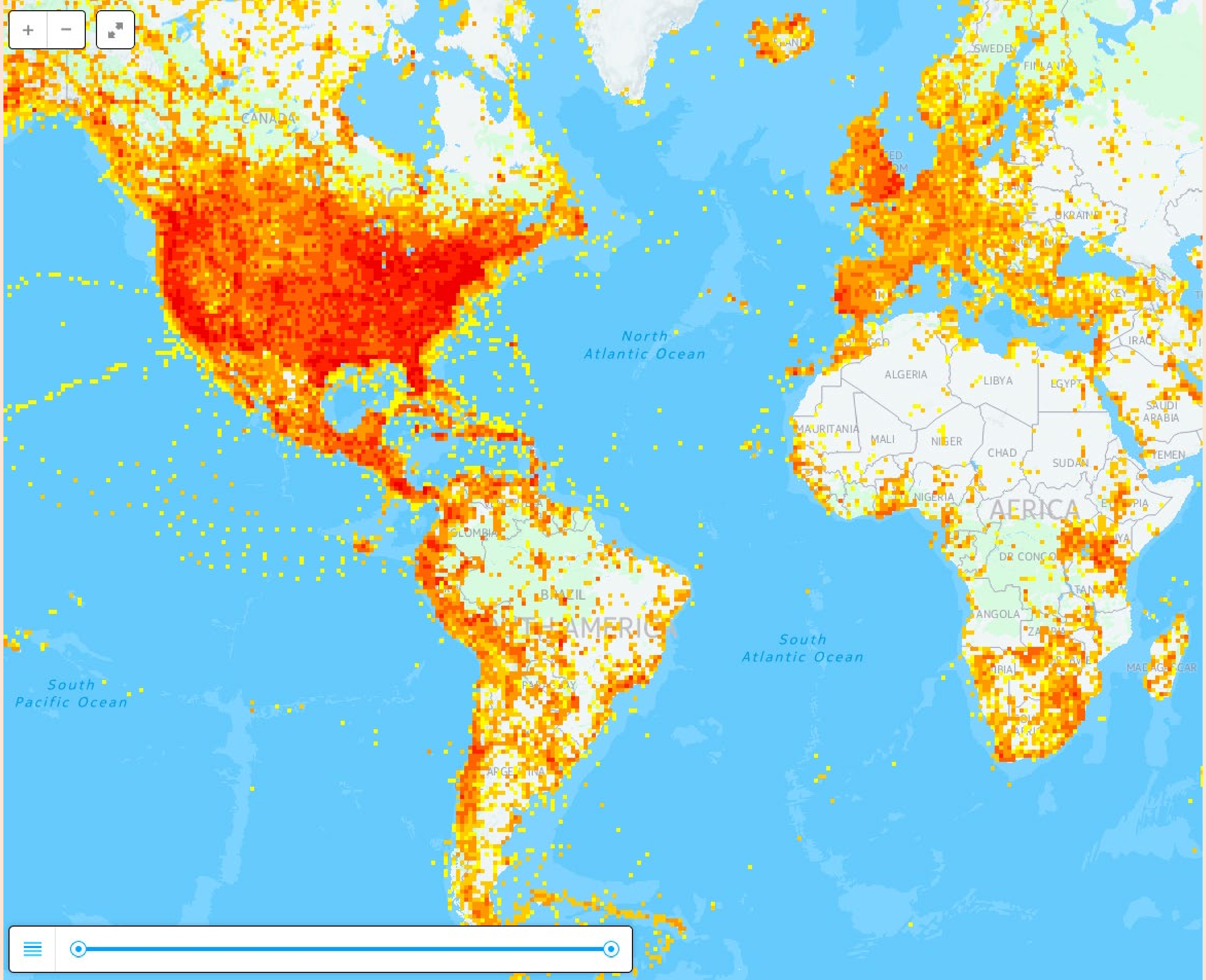
Missouri Botanical Garden

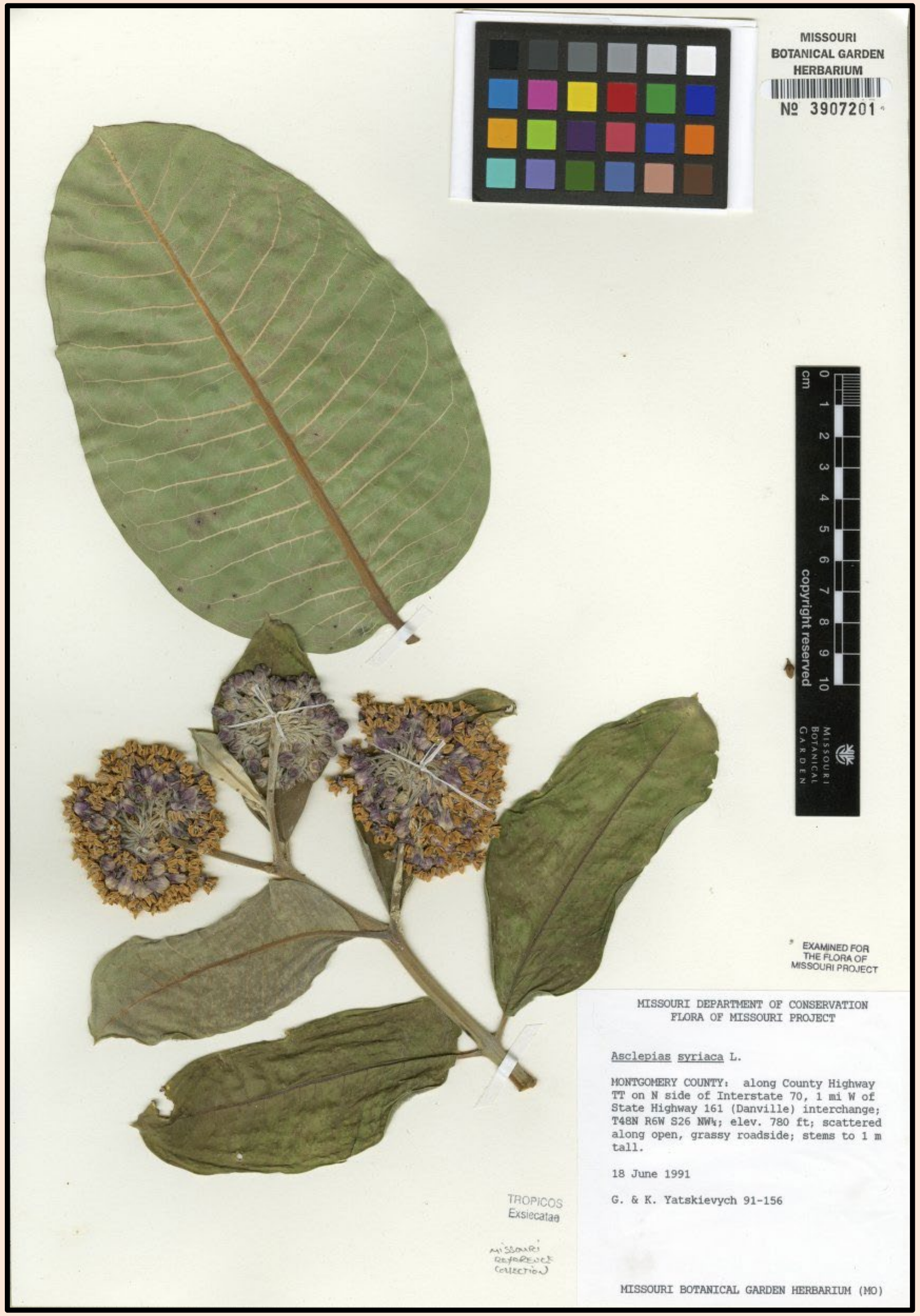
iDigBio
Integrated Digitized Biocollections

Institute of Museum and Library Services



Data processing
Data sources
Species data
Field data
Systematic surveys
Biodiversity databases
Museum records
Atlas data
Remote sensing
Predictor data
Field data
Climate databases
Topographic data
Land use
Remote sensing
Data selection
Raw data
Occurrence records
Climate and environment layers
Data cleaning
Outlier removal
Define spatial extent
Variable selection
Pseudo-absence selection
Input data
Spatial modelling
Multivariate climate-space
Mean annual precipitation
Mean annual temperature
Environment Modeling
BIOCLIM
GLM
GAM
Boosted Regression Tree
MAXENT
Neural Network
Machine learning
Species response curves
Species presence
Climate variables
Geographic Projection
Habitat suitability map
Murphy and Smith, in prep


CANADA
North
Atlantic Ocean
South
Atlantic Ocean
South
Pacific Ocean
SWEDEN
ALGERIA
LIBYA
EGYPT
MAURITANIA
MALI
NIGER
CHAD
SUDAN
SAUDI
ARABIA
YEMEN
AFRICA
ANGOLA
BRAZIL



MONTGOMERY COUNTY: along highway TT on N side of Interstate 70, 1 mi W of State Highway 161 (Danville) interchange; T48N R6W S26 NW $^{1}/_{2}$; elev. 780 ft; scattered along open, grassy roadside; stems to 1 m tall.

INT. J. GEOGRAPHICAL INFORMATION SCIENCE
VOL. 18, NO. 8, DECEMBER 2004, 745–767

Taylor & Francis
Taylor & Francis Group

Research Article

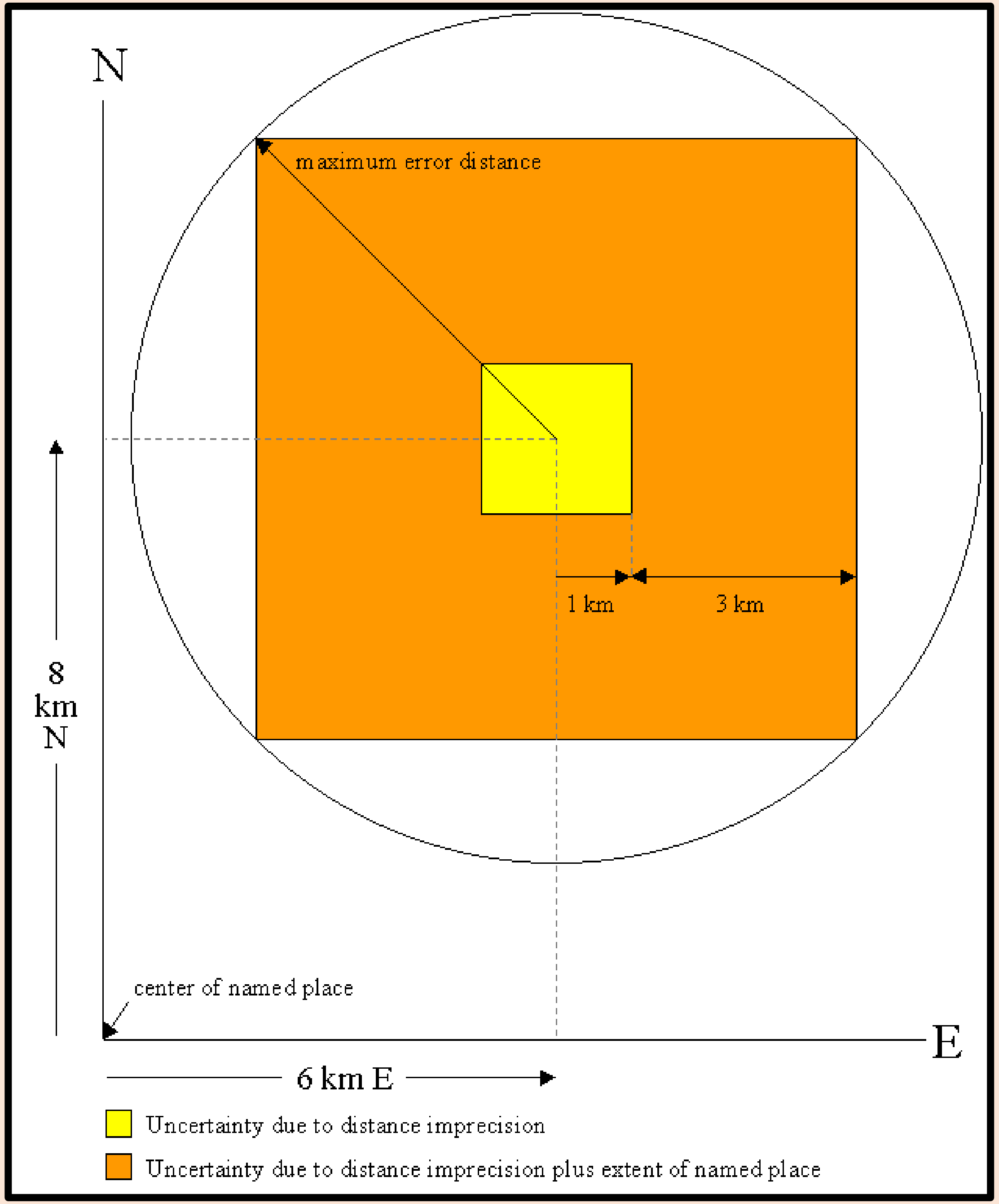
The point-radius method for georeferencing locality descriptions and calculating associated uncertainty

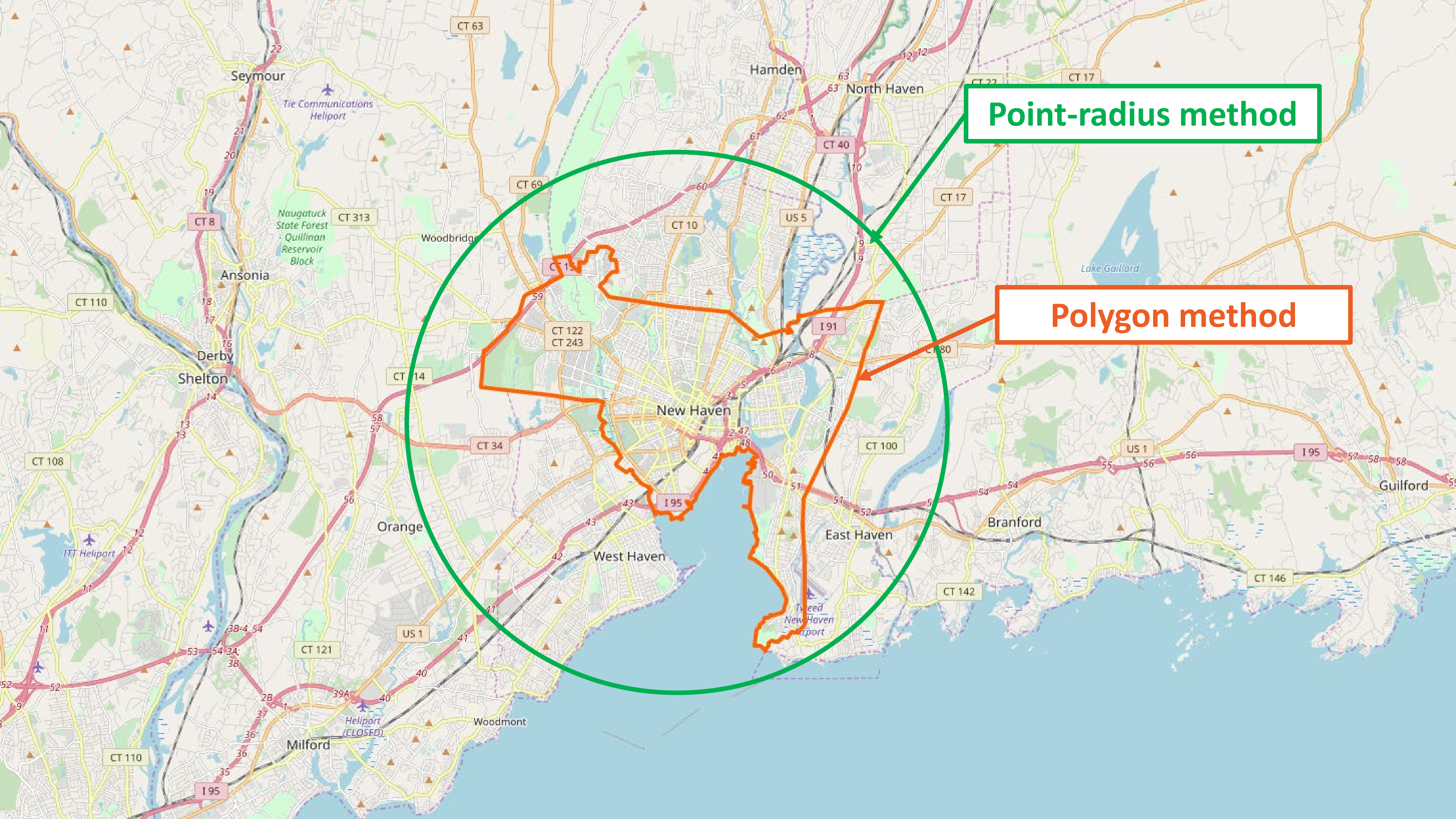
JOHN WIECZOREK*
Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA;
e-mail: tuco@socrates.berkeley.edu

QINGHUA GUO
Department of Environmental Sciences, Policy & Management, 151 Hilgard Hall #3110, University of California, Berkeley, CA 94720, USA

and ROBERT J. HIJMANS
Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA

Natural history museums store millions of specimens of geological, biological, and cultural entities. Data related to these objects are in increasing demand for investigations of biodiversity and its relationship to the environment and anthropogenic disturbance. A major barrier to the use of these data in GIS is that collecting localities have typically been recorded as textual descriptions, without geographic coordinates. We describe a method for georeferencing locality descriptions that accounts for the idiosyncrasies, sources of uncertainty, and practical maintenance requirements encountered when working with natural history collections. Each locality is described as a circle, with a point to mark the position most closely described by the locality description, and a radius to describe the maximum distance from that point within which the locality is expected to occur. The calculation of the radius takes into account aspects of the precision and specificity of the locality description, as well as the map scale, datum, precision and accuracy of the sources used to determine coordinates. This method minimizes the subjectivity involved in the georeferencing process. The resulting georeferences are consistent, reproducible, and allow for the use of uncertainty in analyses that use these data.




Point-radius method
Polygon method
New Haven
West Haven
East Haven
Hamden
North Haven
Branford
Guilford
Orange
Woodbridge
Milford
Woodmont
Seymour
Ansonia
Derby
Shelton
Lake Gaillard
Tweed New Haven Airport
Naugatuck State Forest Quillinan Reservoir Block
Tie Communications Heliport
ITT Heliport
Heliport (CLOSED)

Methods in Ecology and Evolution
British Ecological Society
Methods in Ecology and Evolution 2015, 6, 424–438 doi: 10.1111/2041-210X.12242

SPECIAL FEATURE
NEW OPPORTUNITIES AT THE INTERFACE BETWEEN ECOLOGY AND STATISTICS

Bias correction in species distribution models: pooling survey and collection data for multiple species

William Fithian[1]*, Jane Elith[2], Trevor Hastie[1] and David A. Keith[3]

[1]Stanford University, Department of Statistics, 390 Serra Mall, Stanford, CA, USA 94305, USA; [2]School of Botany, University of Melbourne, Parkville, VIC 3010, Australia; and [3]Centre for Ecosystem Science, University of New South Wales, Sydney 2052, NSW, Australia

Su
1. I unk syst
2. V com max ence cies
3. V pres data amo
4. I for s estir

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B
Phil. Trans. R. Soc. B (2012) 367
doi:10.1098/rstb.

Review

Incorporating uncertainty in predictive species distribution modelling

Colin M. Beale[1,*] and Jack J. Lennon[2]

[1]*Department of Biology, University of York, Wentworth Way, York YO10 5DD, UK*
[2]*The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, UK*

Motivated by the need to solve ecological problems (climate change, habitat fragmentation and bi logical invasions), there has been increasing interest in species distribution models (SDM Predictions from these models inform conservation policy, invasive species management and d ease-control measures. However, predictions are subject to uncertainty, the degree and source which is often unrecognized. Here, we review the SDM literature in the context of uncertain focusing on three main classes of SDM: niche-based models, demographic models and proce based models. We identify sources of uncertainty for each class and discuss how uncertainty c be minimized or included in the modelling process to give realistic measures of confiden around predictions. Because this has typically not been performed, we conclude that uncertai in SDMs has often been underestimated and a false precision assigned to predictions of geograp cal distribution. We identify areas where development of new statistical tools will improve predictions from distribution models, notably the development of hierarchical models that link different types of distribution model and their attendant uncertainties across spatial scales. Finally, we discuss the need to develop more defensible methods for assessing predictive performance, quantifying model goodness-of-fit and for assessing the significance of model covariates.

Keywords: distribution model; bioclimate envelope; spatial analysis; niche model; process-based model; climate impacts

Methods in Ecology and Evolution
BRITISH ECOLOGICAL SOCIETY
Methods in Ecology and Evolution 2017, **8**, 1566–1573 doi: 10.1111/2041-210X.12793

Bias correction of bounded location errors in presence-only data

Trevor J. Hefley*[,1], Brian M. Brost[2] and Mevin B. Hooten[3]

[1]*Department of Statistics, Kansas State University, Manhattan, KS, USA;* [2]*Marine Mammal Laboratory, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle, WA, USA; and* [3]*U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Department of Statistics, Colorado State University, Fort Collins, CO, USA*

Summary
1. Location teristics at th to unreliable
2. We expla be used to c within arbitr
3. We illustr *cana*) using with a simula
4. In our an five-fold cha estimates tha

Key-word Poisson po

Research
Ecography 39: 305–316, 201
doi: 10.1111/ecog.0120
Subject Editor: Miguel Araújo. Editor-in-Chief: Miguel Araújo. Accepted 8 April 201

Using measurement error models to account for georeferencing error in species distribution models

Jorge Velásquez-Tibatá, Catherine H. Graham and Stephan B. Munch

J. Velásquez-Tibatá (jivelasquezt@gmail.com) and C. H. Graham, Dept of Ecology and Evolution, Stony Brook Univ., Stony Brook, NY 11794 USA. Present address of JV-T: Inst. de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia. – S. B. Munch, Southwes Fisheries Science Center, National Oceanic and Atmospheric Administration, 110 Shaffer Road, Santa Cruz, CA 95060, USA.

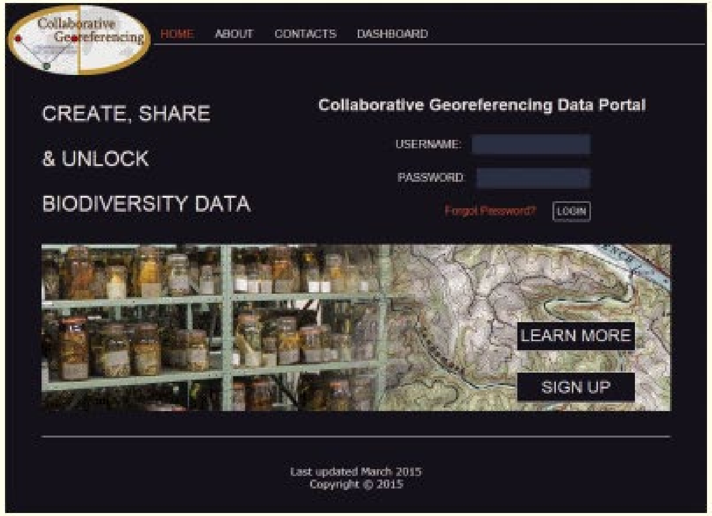
Georeferencing error is prevalent in datasets used to model species distributions, inducing uncertainty in covariate values associated with species occurrences that result in biased probability of occurrence estimates. Traditionally, this error has been dealt with at the data-level by using only records with an acceptable level of error (filtering) or by summarizing covariates at sampling units by using measures of central tendency (averaging). Here we compare those previous approaches to a novel implementation of a Bayesian logistic regression with measurement error (ME), a seldom used method in species distribution modeling. We show that the ME model outperforms data-level approaches on 1) specialist species and 2) when either sample sizes are small, the georeferencing error is large or when all georeferenced occurrences have a fixed level of error. Thus, for certain types of species and datasets the ME model is an effective method to reduce biases in probability of occurrence estimates and account for the uncertainty generated by georeferencing error. Our approach may be expanded for its use with presence-only data as well as to include other sources of uncertainty in species distribution models.

Country	State Province	County	Municipality	Locality	Latitude	Longitude	uncertainty in m
US	Ohio			Old fields, Rockland, Ohio.			
US	Kansas			4 mi. W. Lawrence.			
US	Rhode Island			Prudence Island, Narragansett Bay.			
US	Minnesota			Fort Snelling.			
US	Kentucky			Pendleton Co. Roadsides. Near Falmouth.			
US	Ohio			County: Greene. Yellow Springs.			
US	Wisconsin			Vicinity of Delavan.			
US	Minnesota			Douglas Co. Camp Carlos for Boys.			
US	North Carolina			The Southern United States. Near Chapel Hill.			
US	West Virginia			On old grade about Â½ mile from iron bridge across Blackwater River, Davis. Canaan Valley, Tucker County.			
US	Illinois			Sandy shores of Lake Michigan, Lak View. Chicago.			
US	Rhode Island			Prudenee Island, Narragansett Bay.			
US	Kansas			Mitchell Co. 7 mi. n. Tipton.			
US	Maine			Town of Cumberland. Sandy embankment.			
US	Connecticut			The Vicinity of Green's Farms, Conn.			
US	Nebraska			Weeping Water.			
US	Michigan			Low, sandy ground near railroad. Stevensville.			
US	Iowa			Southeastern Iowa. S.E. 1/4 Sec. 5, T-75N, R-2W, Louisa Co.			
CA	Ontario			Roadside, east of Morristown, Puslinch Tp., Wellington County, Conc. X, lot 35, Ont.			
US	Vermont			Southern Vermont. Manchester.			
US	Illinois			Rock Island Arsenal.			
US	Tennessee			Near Rockwood.			
US	Tennessee			Roane County. White Oak Creek and Lake. Along road leading to the head of W.O. Lake.			
US	Minnesota			Itasca Park. The Headwaters of the Mississippi River. Opening in N. Pine. County Clearwater. Sec. SW-SE-35 T. 143 R. 36.			
US	Ohio			Hamilton Co.			
US	New York			Northern New York. Canton.			
US	Nebraska			Beatrice.			
US	New York			Watkins Glen.			
US	New York			Van Cortlandt, N. Y.			
US	Nebraska			Northeastern Nebraska. Aten. Cedar Co.			
US	Missouri			Prairie along r.r., about 2 mi. west of Amity, De Kalb Co.			
US	Connecticut			New Haven.			
US	Illinois			Peoria County. Woodroad. Peoria Heights.			
US	Wisconsin			Crandon, Wis.			
US	Kansas			Woodson Co. T24S, R14E, SW corner Sec.21.			
US	Iowa			Fayette Co.			
US	Kentucky			Edmonson Co. Mammoth Cave National Park. MCNP-Joppa Ridge-Near Joppa Church.			
US	West Virginia			Bridge over Blackwater River, Davis. Canaan Valley, Tucker County.			
US	Kansas			Douglas County. 1 mile north of Lawrence. Edge of field.			
US	Indiana			Michigan City.			
US	Virginia			South Holston Valley. S.W. Virginia. S. Fork Holston River, at St. Clair's Bottom, Smyth Co.			
US	Wisconsin			Western Wisconsin. La Crosse County. Dry, sandy prairie. Midway Scientific Area. (T. 17 N.; R. 7 W; Sect. 30). Onalaska Twp.			
US	Virginia			Stony Man Mountain and vicinity in the Blue Ridge, near Luray, Va.			
US	Wisconsin			Camp Douglas.			

Home | Web Application | Collaborative Georeferencing | Developer Resources | Education & Outreach | Support and Contacts

GEOLocate

A Platform for Georeferencing Natural History Collections Data



For Users:

- Overview
- GEOLocate Web Clients
- Collaborative Georeferencing
- Education & Outreach

Brief overview (video) of the GEOLocate Project.

For Developers:

- Web Services
- Embeddable Web Client




Web Applications

Georeference collections data using your web browser. Quick and easy georeferencing.

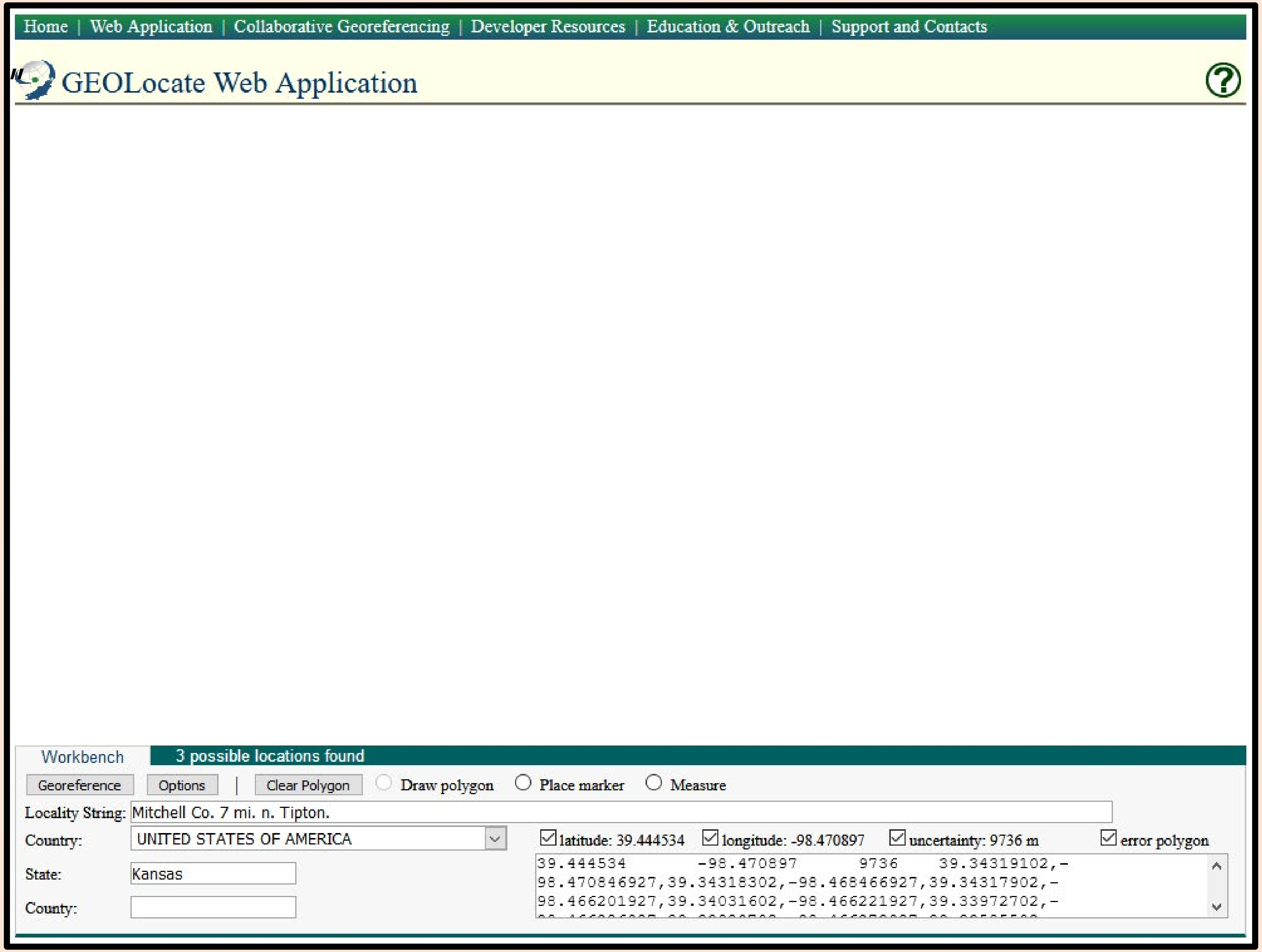
Web Services

Integrate georeferencing into your own databases and applications using GEOLocate webservices.

Collaborative Georeferencing

Build communities, share data, relate records across collections and improve verification efficiency.

"Mitchell Co. 7 mi. n. Tipton"



"Prudence Island, Narragansett Bay."


INT. J. GEOGRAPHICAL INFORMATION SCIENCE
VOL. 18, NO. 8, DECEMBER 2004, 745–767

Taylor & Francis
Taylor & Francis Group

Research Article

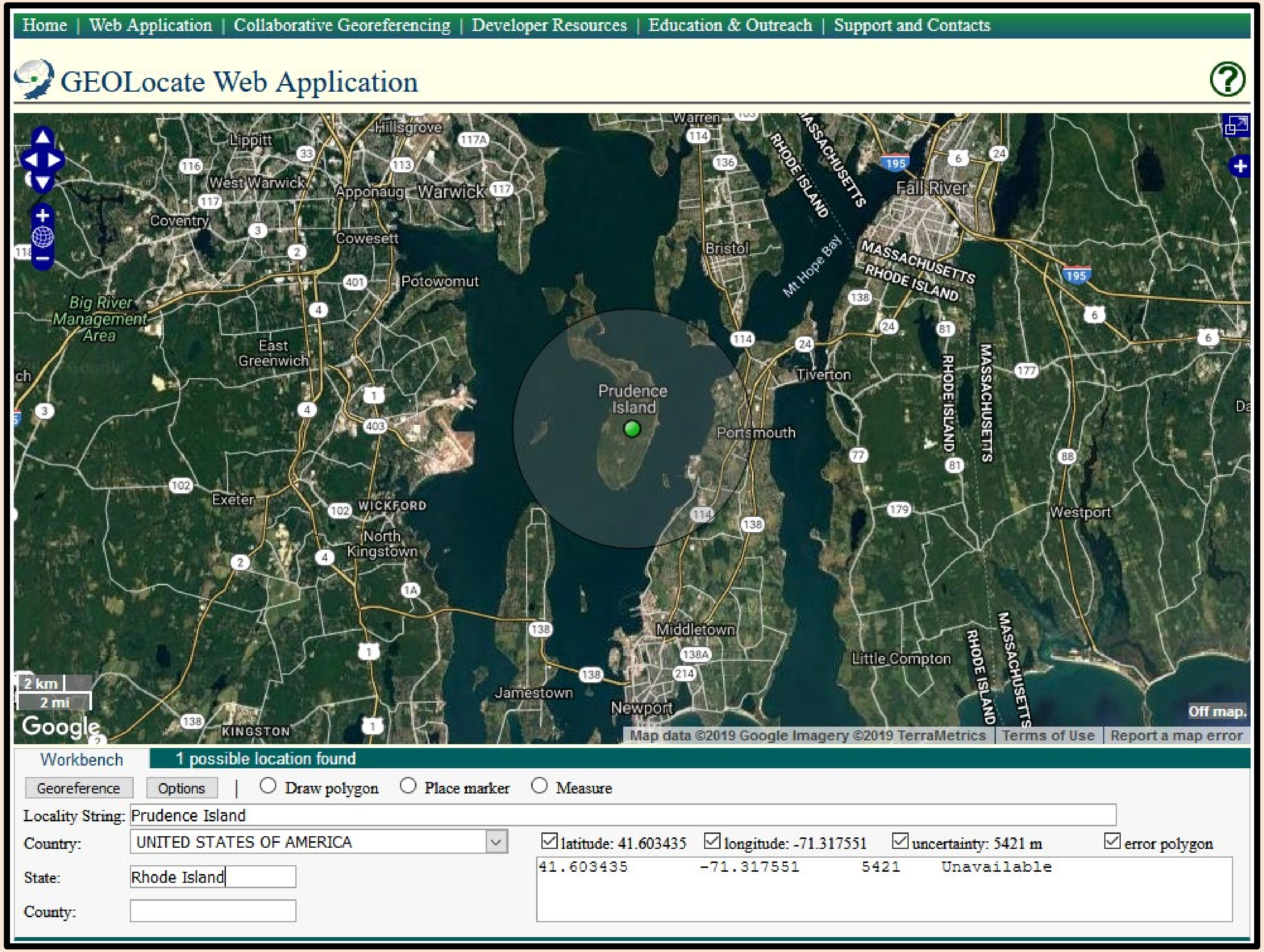
The point-radius method for georeferencing locality descriptions and calculating associated uncertainty

JOHN WIECZOREK*
Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA;
e-mail: tuco@socrates.berkeley.edu

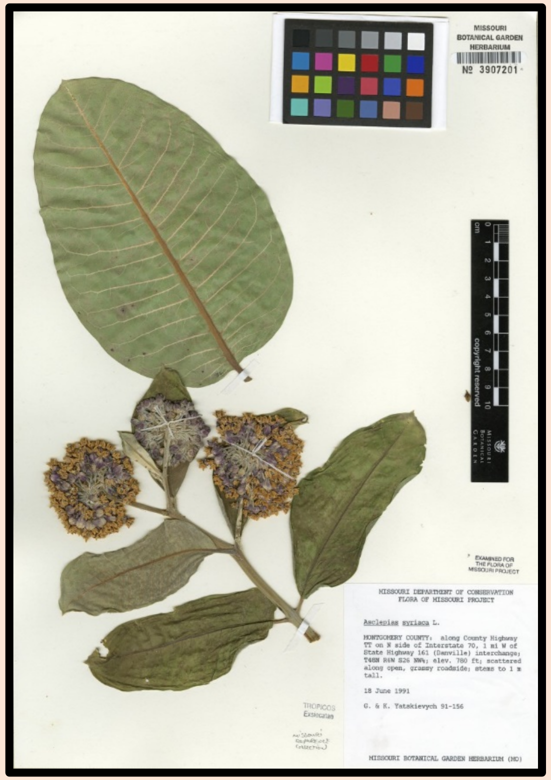
QINGHUA GUO
Department of Environmental Sciences, Policy & Management, 151 Hilgard Hall #3110, University of California, Berkeley, CA 94720, USA

and ROBERT J. HIJMANS
Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA

Natural history museums store millions of specimens of geological, biological, and cultural entities. Data related to these objects are in increasing demand for investigations of biodiversity and its relationship to the environment and anthropogenic disturbance. A major barrier to the use of these data in GIS is that collecting localities have typically been recorded as textual descriptions, without geographic coordinates. We describe a method for georeferencing locality descriptions that accounts for the idiosyncrasies, sources of uncertainty, and practical maintenance requirements encountered when working with natural history collections. Each locality is described as a circle, with a point to mark the position most closely described by the locality description, and a radius to describe the maximum distance from that point within which the locality is expected to occur. The calculation of the radius takes into account aspects of the precision and specificity of the locality description, as well as the map scale, datum, precision and accuracy of the sources used to determine coordinates. This method minimizes the subjectivity involved in the georeferencing process. The resulting georeferences are consistent, reproducible, and allow for the use of uncertainty in analyses that use these data.



How do we achieve accuracy in uncertainty?



Identify the feature of minimal area and generate a spatial polygon defining its boundary

→ "Prudence Island, Narangaset Bay" →

basic protocol

- Parsing of textual information into discrete geolocational features
- Identify minimum area feature
- Geo-validation of parsed components
- Assign spatial polygon to minimum area features

basic protocol

- Parsing of textual information into discrete geolocational features
- Identify minimum area feature
- Geo-validation of parsed components
- Assign spatial polygon to minimum area features

New R package in development

- 'GeoParse' function
  - Admin level 0 name (Country)
  - Admin level 1 name (State or province)
  - Admin level 2 name (County, Municipality)
  - Admin level 3 name (Town or city)
  - Street name
  - Address number
  - Zip or other postal code
  - Place names and other features
  - Elevation data

Basic Regular Expressions in R

Cheat Sheet

Character Classes

[[:digit:]] or \\d	Digits; `[0-9]`
\\D	Non-digits; `[^0-9]`
[[:lower:]]	Lower-case letters; `[a-z]`
[[:upper:]]	Upper-case letters; `[A-Z]`
[[:alpha:]]	Alphabetic characters; `[A-z]`
[[:alnum:]]	Alphanumeric characters `[A-z0-9]`
\\w	Word characters; `[A-z0-9_]`
\\W	Non-word characters
[[:xdigit:]] or \\x	Hexadec. digits; `[0-9A-Fa-f]`
[[:blank:]]	Space and tab
[[:space:]] or \\s	Space, tab, vertical tab, newline, form feed, carriage return
\\S	Not space; `[^[:space:]]`
[[:punct:]]	Punctuation characters; `!"#$%&'()*+,-./:;<=>?@[]^_`{|}~`
[[:graph:]]	Graphical char.; `[[:alnum:][:punct:]]`
[[:print:]]	Printable characters; `[[:alnum:][:punct:]\\s]`
[[:cntrl:]] or \\c	Control characters; `\n`, `\r` etc.

Special Metacharacters

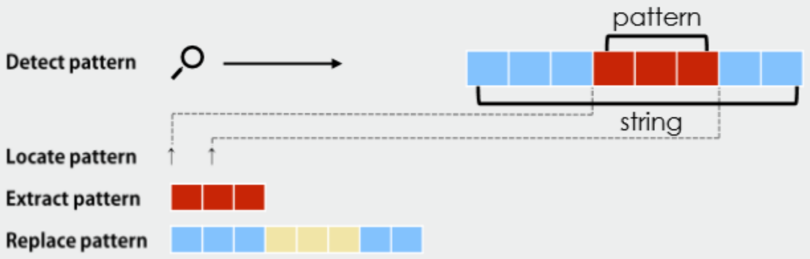
\n	New line
\r	Carriage return
\t	Tab
\v	Vertical tab
\f	Form feed

Lookaraounds and Conditionals*

(?=)	Lookahead (requires PERL = TRUE), e.g. `(?=yx)`: position followed by 'xy'
(?!)	Negative lookahead (PERL = TRUE); position NOT followed by pattern
(?<=)	Lookbehind (PERL = TRUE), e.g. `(?<=yx)`: position following 'xy'
(?<!)	Negative lookbehind (PERL = TRUE); position NOT following pattern
?(if)then	If-then-condition (PERL = TRUE); use lookaheads, optional char. etc in if-clause
?(if)then|else	If-then-else-condition (PERL = TRUE)

*see, e.g. http://www.regular-expressions.info/lookaround.html
http://www.regular-expressions.info/conditional.html

Functions for Pattern Matching




```
> string <- c("Hiphopopotamus", "Rhymenoceros", "time for bottomless lyrics")
> pattern <- "t.m"
```

Detect Patterns

grep(pattern, string)
`[1] 1 3`

grep(pattern, string, value = TRUE)
`[1] "Hiphopopotamus"`
`[2] "time for bottomless lyrics"`

grepl(pattern, string)
`[1]  TRUE FALSE  TRUE`

stringr::str_detect(string, pattern)
`[1]  TRUE FALSE  TRUE`

Locate Patterns

regexpr(pattern, string)
find starting position and length of first match

gregexpr(pattern, string)
find starting position and length of all matches

stringr::**str_locate(string, pattern)**
find starting and end position of first match

stringr::**str_locate_all(string, pattern)**
find starting and end position of all matches

Split a String using a Pattern

strsplit(string, pattern) or stringr::**str_split(string, pattern)**

Extract Patterns

regmatches(string, regexpr(pattern, string))
extract first match `[1] "tam" "tim"`

regmatches(string, gregexpr(pattern, string))
extracts all matches, outputs a list
`[[1]] "tam" [[2]] character(0) [[3]] "tim" "tom"`

stringr::**str_extract(string, pattern)**
extract first match `[1] "tam" NA  "tim"`

stringr::**str_extract_all(string, pattern)**
extract all matches, outputs a list

stringr::**str_extract_all(string, pattern, simplify = TRUE)**
extract all matches, outputs a matrix

stringr::**str_match(string, pattern)**
extract first match + individual character groups

stringr::**str_match_all(string, pattern)**
extract all matches + individual character groups

Replace Patterns

sub(pattern, replacement, string)
replace first match

gsub(pattern, replacement, string)
replace all matches

stringr::**str_replace(string, pattern, replacement)**
replace first match

stringr::**str_replace_all(string, pattern, replacement)**
replace all matches

Character Classes and Groups

.	Any character except \n
|	Or, e.g. `(a|b)`
[...]	List permitted characters, e.g. `[abc]`
[a-z]	Specify character ranges
[^...]	List excluded characters
(...)	Grouping, enables back referencing using \\N where N is an integer

Anchors

^	Start of the string
$	End of the string
\\b	Empty string at either edge of a word
\\B	NOT the edge of a word
\\<	Beginning of a word
\\>	End of a word

Quantifiers

*	Matches at least 0 times
+	Matches at least 1 time
?	Matches at most 1 time; optional string
{n}	Matches exactly n times
{n,}	Matches at least n times
{,n}	Matches at most n times
{n,m}	Matches between n and m times

General Modes

By default R uses *POSIX extended regular expressions*. You can switch to *PCRE regular expressions* using `PERL = TRUE` for base or by wrapping patterns with `perl()` for stringr.

All functions can be used with literal searches using `fixed = TRUE` for base or by wrapping patterns with `fixed()` for stringr.

All base functions can be made case insensitive by specifying `ignore.cases = TRUE`.

Escaping Characters

Metacharacters (`. * +` etc.) can be used as literal characters by escaping them. Characters can be escaped using `\\` or by enclosing them in `\\Q...\\E`.

Case Conversions

Regular expressions can be made case insensitive using `(?i)`. In backreferences, the strings can be converted to lower or upper case using `\\L` or `\\U` (e.g. `\\L\\1`). This requires `PERL = TRUE`.

Greedy Matching

By default the asterisk `*` is greedy, i.e. it always matches the longest possible string. It can be used in lazy mode by adding `?`, i.e. `*?`.

Greedy mode can be turned off using `(?U)`. This switches the syntax, so that `(?U)a*` is lazy and `(?U)a*?` is greedy.

Note

Regular expressions can conveniently be created using rex::**rex()**.

NorthAmerica_Level_1 - Excel

F111

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
61	US	United States	United States	Illinois	Illinois	Illinois	Illinois	IL			Illi					
62	US	United States	United States	Indiana	Indiana	Indiana	Indiana	IN			Indi	Ill				
63	US	United States	United States	Iowa	Iowa	Iowa	Iowa	IA								
64	US	United States	United States	Jarvis Island		Jarvis Island										
65	US	United States	United States	Johnston Atoll		Johnston Atoll										
66	US	United States	United States	Kansas	Kansas	Kansas	Kansas	KS			Kans	Kan				
67	US	United States	United States	Kentucky	Kentucky	Kentucky	Kentucky	KY			Ky					
68	US	United States	United States	Kingman Reef		Kingman Reef										
69	US	United States	United States	Louisiana	Louisiana	Louisiana	Louisiana	LA								
70	US	United States	United States	Maine	Maine	Maine	Maine	ME								
71	US	United States	United States	Maryland	Maryland	Maryland	Maryland	MD			Md					
72	US	United States	United States	Massachusetts	Massachusetts	Massachusetts	Massachusetts	MA			Mass					
73	US	United States	United States	Michigan	Michigan	Michigan	Michigan	MI			Mich					
74	US	United States	United States	Midway Islands		Midway Islands										
75	US	United States	United States	Minnesota	Minnesota	Minnesota	Minnesota	MN			Minn					
76	US	United States	United States	Mississippi	Mississippi	Mississippi	Mississippi	MS			Miss					
77	US	United States	United States	Missouri	Missouri	Missouri	Missouri	MO			Mo					
78	US	United States	United States	Montana	Montana	Montana	Montana	MT			Mont					
79	US	United States	United States	Navassa Island		Navassa Island										
80	US	United States	United States	Nebraska	Nebraska	Nebraska	Nebraska	NE			Nebr	Neb				
81	US	United States	United States	Nevada	Nevada	Nevada	Nevada	NV			Nev					
82	US	United States	United States	New Hampshire	New Hampshire	New Hampshire	New Hampshire	NH			NH	N.H.	N. H.			
83	US	United States	United States	New Jersey	New Jersey	New Jersey	New Jersey	NJ			N.J.	N. J.				
84	US	United States	United States	New Mexico	New Mexico	New Mexico	New Mexico	NM			N. Mex	N.Mex	New Mex			
85	US	United States	United States	New York	New York	New York	New York	NY			N.Y.	N. Y.				
86	US	United States	United States	North Carolina	North Carolina	North Carolina	North Carolina	NC			N.C.	N. C.				
87	US	United States	United States	North Dakota	North Dakota	North Dakota	North Dakota	ND			N. Dak	N Dak	North Dak	N Dakota		
88	US	United States	United States	Ohio	Ohio	Ohio	Ohio	OH								
89	US	United States	United States	Oklahoma	Oklahoma	Oklahoma	Oklahoma	OK			Okla					
90	US	United States	United States	Oregon	Oregon	Oregon	Oregon	OR			Oreg	Ore				
91	US	United States	United States	Palmyra Atoll		Palmyra Atoll										
92	US	United States	United States	Pennsylvania	Pennsylvania	Pennsylvania	Pennsylvania	PA			Penn	Pa.				
93	US	United States	United States	Rhode Island	Rhode Island	Rhode Island	Rhode Island	RI			R.I.	R. I.				
94	US	United States	United States	South Carolina	South Carolina	South Carolina	South Carolina	SC			S.C.	S. C.				
95	US	United States	United States	South Dakota	South Dakota	South Dakota	South Dakota	SD			S.Dak	S. Dak	South Dak	S Dakota	S. Dakota	
96	US	United States	United States	Tennessee	Tennessee	Tennessee	Tennessee	TN			Tenn					
97	US	United States	United States	Texas	Texas	Texas	Texas	TX			Tex					
98	US	United States	United States	Utah	Utah	Utah	Utah	UT								
99	US	United States	United States	Vermont	Vermont	Vermont	Vermont	VT			Vt.					
100	US	United States	United States	Virginia	Virginia	Virginia	Virginia	VA			Virg	Va.				
101	US	United States	United States	Wake Island		Wake Island										
102	US	United States	United States	Washington	Washington	Washington	Washington	WA			Wash					
103	US	United States	United States	West Virginia	West Virginia	West Virginia	West Virginia	WV			W.Va	W Va	W. Va	West Va	West Virg	W. Virg
104	US	United States	United States	Wisconsin	Wisconsin	Wisconsin	Wisconsin	WI			Wisc	Wis				
105	US	United States	United States	Wyoming	Wyoming	Wyoming	Wyoming	WY			Wyom	Wyo				

NorthAmerica_Level_1

Ready

parsed text - Excel

	E	F	G	H	I	J	K	L	M
1	**Locality string**	**Country**	**State**	**County**	**Township**	**Town**	**Offset**	**Named Place**	**Other text**
2	Old fields, Rockland, Ohio.		Ohio			Rockland			
3	4 mi. W. Lawrence.						4 mi. W.		
4	Prudence Island, Narragansett Bay.							Prudence Island; Narragansett Bay	
5	Fort Snelling.								Fort Snelling
6	Pendleton Co. Roadsides. Near Falmouth.			Pendleton		Falmouth	Near		
7	County: Greene. Yellow Springs.			Greene		Yellow Springs			
8	Vicinity of Delavan.					Delavan	Vicinity of		
9	Douglas Co. Camp Carlos for Boys.			Douglas					Camp Carlos for Boys
10	The Southern United States. Near Chapel Hill.					Chapel Hill	Near		
11	On old grade about Â½ mile from iron bridge across Blackwater River, Davis. Canaan Valley, Tucker County.			Tucker		Davis	about; 1/2 mile from	Blackwater River	An old grade; iron bridge across
12	Sandy shores of Lake Michigan, Lak View. Chicago.					Chicago		Lake Michigan	Sandy shores of; Lak View
13	Prudenee Island, Narragansett Bay.							Prudenee Island; Narragansett Bay	
14	Mitchell Co. 7 mi. n. Tipton.			Mitchell		Tipton	7 mi. n.		
15	Town of Cumberland. Sandy embankment.					Cumberland			Sandy embankment
16	The Vicinity of Green's Farms, Conn.		Connecticut				Vicinity of		The; Green's Farms
17	Weeping Water.					Weeping Water			
18	Low, sandy ground near railroad. Stevensville.					Stevensville	near		Low, sandy ground; railroad
19	Southeastern Iowa. S.E. 1/4 Sec. 5, T-75N, R-2W, Louisa Co.		Iowa	Louisa			Southeastern; S.E.		1/4 Sec. 5, T-75N, R-2W
20	Roadside, east of Morristown, Puslinch Tp., Wellington County, Conc. X, lot 35, Ont.		Ontario	Wellington	Puslinch	Morristown	east of		Roadside; Conc. X, lot 35
21	Southern Vermont. Manchester.		Vermont			Manchester	Southern		
22	Rock Island Arsenal.					Rock Island Arsenal			
23	Near Rockwood.					Rockwood	Near		
24	Roane County. White Oak Creek and Lake. Along road leading to the head of W.O. Lake.			Roane			Along		White Oak Creek and Lake; road leading to the head
25	Itasca Park. The Headwaters of the Mississippi River. Opening in N. Pine. County Clearwater. Sec. SW-SE-35 T. 143 R. 36.			Pine			Headwaters; N.	Itaska State Park; Mississippi River	The; of the; Opening in; Clearwater. Sec. SW-SE-35 T
26	Hamilton Co.			Hamilton					
27	Northern New York. Canton.		New York			Canton	Northern		
28	Beatrice.					Beatrice			
29	Watkins Glen.					Watkins Glen			
30	Van Cortlandt, N. Y.		New York			Van Cortlandt			
31	Northeastern Nebraska. Aten. Cedar Co.		Nebraska	Cedar		Aten	Northeastern		
32	Prairie along r.r., about 2 mi. west of Amity, De Kalb Co.			De Kalb		Amity	along; about; 2 mi. west		Prairie; r.r.
33	New Haven.					New Haven			
34	Peoria County. Woodroad. Peoria Heights.			Peoria		Peoria Heights			Woodroad
35	Crandon, Wis.		Wisconsin			Crandon			
36	Woodson Co. T24S, R14E, SW corner Sec.21.			Woodson			SW		T24S, R14E,; corner Sec.21.
37	Fayette Co.			Fayette					
38	Edmonson Co. Mammoth Cave National Park. MCNP-Joppa Ridge-Near Joppa Church.			Edmonson			near		MCNP-Joppa Ridge-; Joppa Church.
39	Bridge over Blackwater River, Davis. Canaan Valley, Tucker County.			Tucker		Davis	about; 1/2 mile from	Blackwater River	An old grade; iron bridge across
40	Douglas County. 1 mile north of Lawrence. Edge of field.			Douglas		Lawrence	1 mile north of		Edge of field
41	Michigan City.					Michigan City			
42	South Holston Valley. S.W. Virginia. S. Fork Holston River, at St. Clair's Bottom, Smyth Co.		Virginia	Smyth			S.W.; S.	South Holston Valley; Holston River	Fork Holston River, at St. Clair's Bottom
43	Western Wisconsin. La Crosse County. Dry, sandy prairie. Midway Scientific Area. (T. 17 N.; R. 7 W; Sect. 30). Onalaska Twp.		Wisconsin	La Crosse	Onalsaska		N		Dry, sandy prairie. Midway Scientific Area. (T. 17; R.
44	Stony Man Mountain and vicinity in the Blue Ridge, near Luray, Va.		Virginia			Luray	vicinity; near		Stony Man Mountain and; in the Blue Ridge,
45	Camp Douglas.					Camp Douglas			

Sheet1

Ready

basic protocol

- Parsing of textual information into discrete geolocational features
- Identify minimum area feature
- Geo-validation of parsed components
- Assign spatial polygon to minimum area features

basic protocol

- Parsing of textual information into discrete geolocational features
- Identify minimum area feature
- Geo-validation of parsed components
- Assign spatial polygon to minimum area features

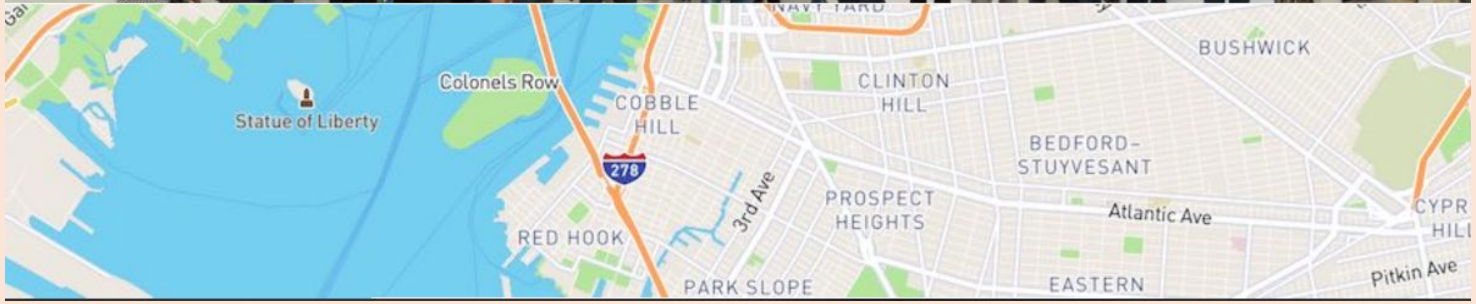
basic protocol

- Parsing of textual information into discrete geolocational features
- Identify minimum area feature
- Geo-validation of parsed components
- Assign spatial polygon to minimum area features

OpenStreetMap US | Blog | Community | Join OSM US | Jobs | Start Mapping

OpenStreetMap US

We help grow **OpenStreetMap**, the free and editable map of the world, in the United States.



Search Where is this? Go

Way: Peabody Museum of Natural History (217341969)

switch name with alt_name. alt_name was the official name

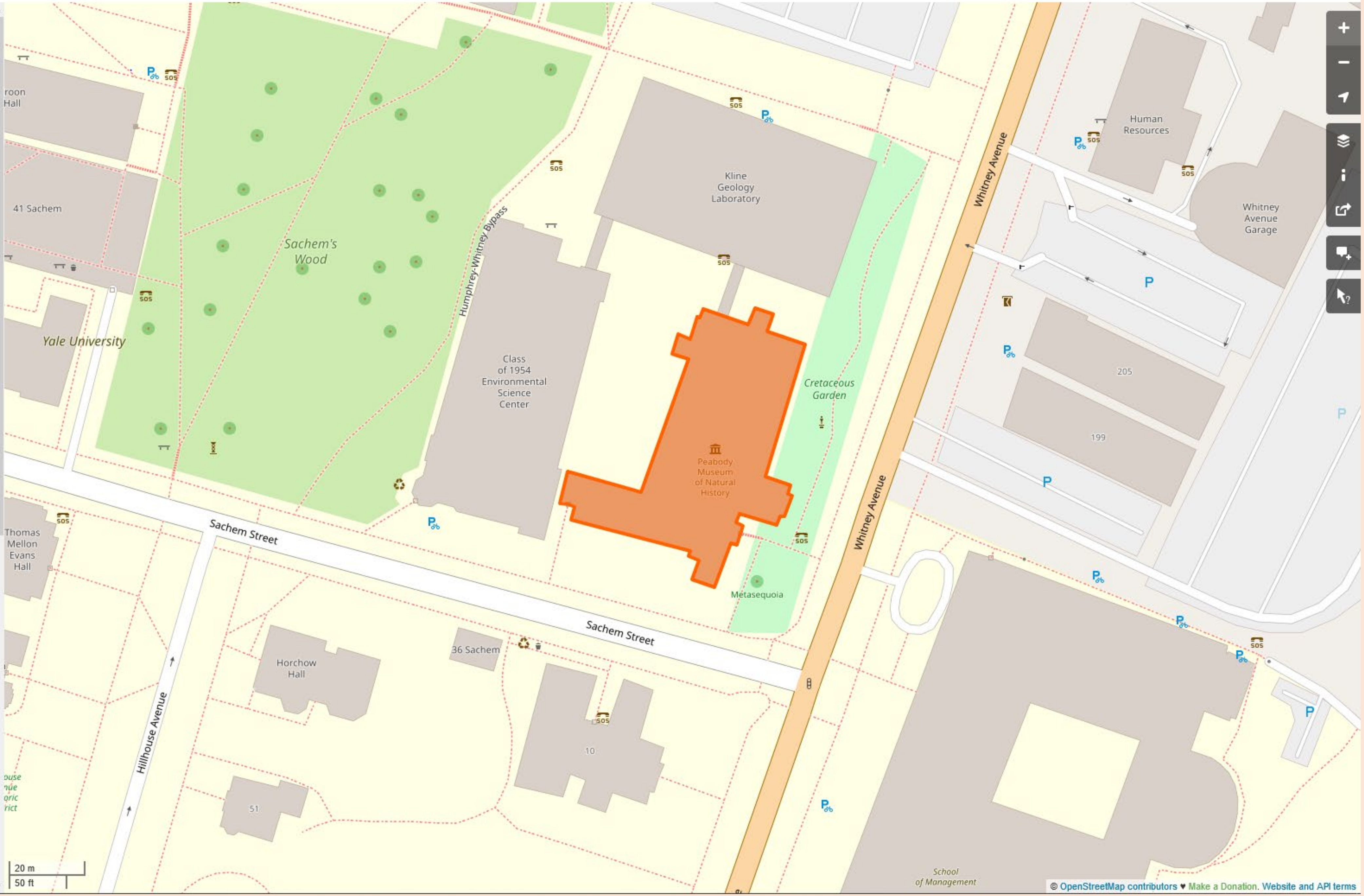
Edited about 2 years ago by snodnipper
Version #12 · Changeset #48060588

Tags

addr:city	New Haven
addr:housenumber	170
addr:postcode	06511
addr:state	CT
addr:street	Whitney Avenue
alt_name	Yale Peabody Museum
building	university
name	Peabody Museum of Natural History
operator	Yale University
tourism	museum
website	http://peabody.yale.edu/
wikidata	Q122945
wikipedia	en:Peabody Museum of Natural History

Nodes

- 2266144279
- 2370071336
- 2371516019 (part of way .... 228473369)
- 3688863560
- 3688863557
- 3688863549
- 3688863545
- 2266144283
- 2266144285
- 2266144288



Search Where is this? Go

Way: Cretaceous Garden (553644538)

Small walking park associated with the Yale Peabody Museu

Edited over 1 year ago by pterman
Version #1 · Changeset #55507987

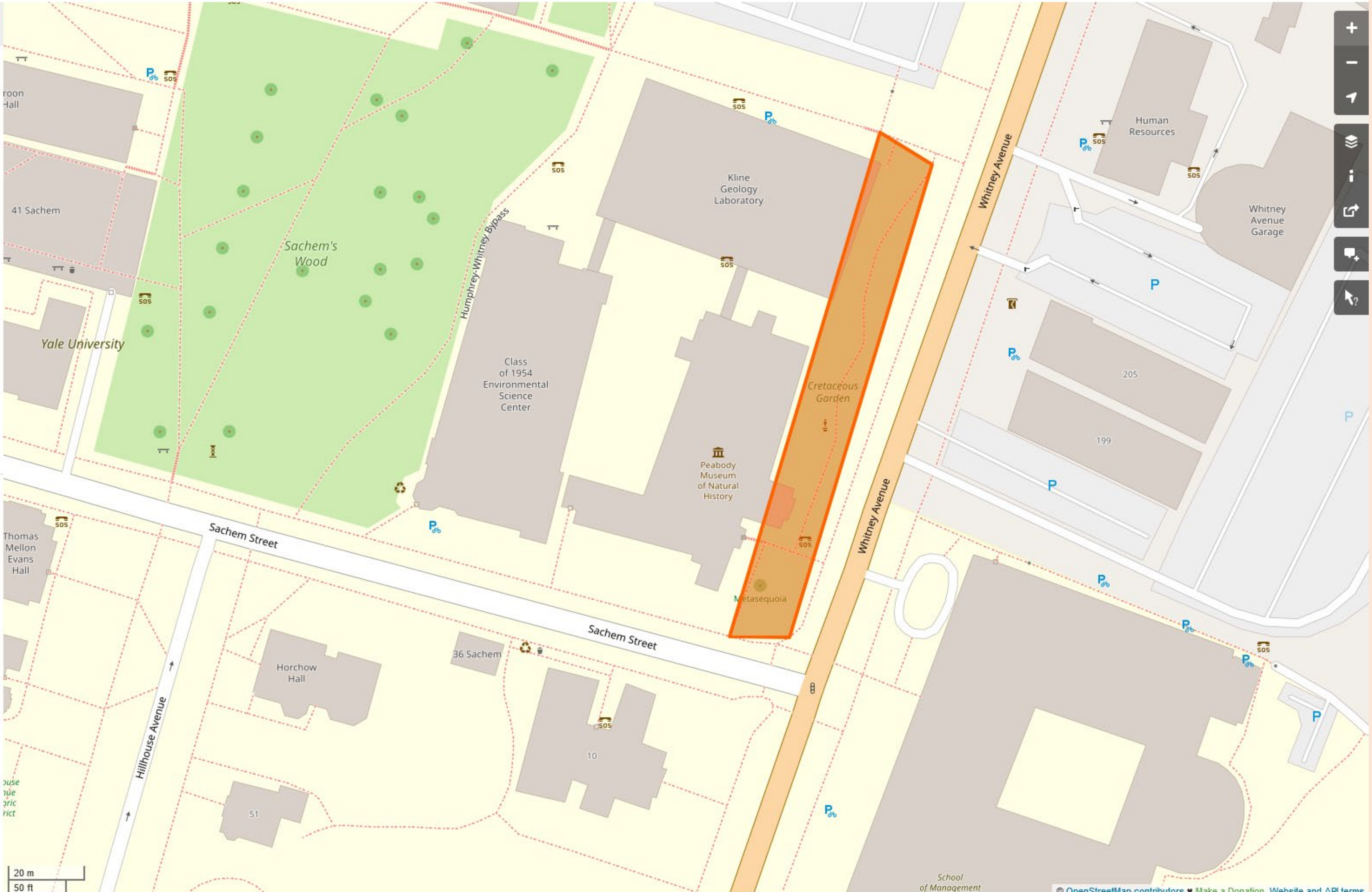
Tags

leisure	park
name	Cretaceous Garden

Nodes

2948023637 (part of ways 291344513 and 291344514)
3470426143 (part of way 339835104)
2193446378 (part of way 216539466)
2370071316 (part of ways 216539466 and 228325198)
2948023637 (part of ways 291344513 and 291344514)

Download XML · View History



Alternate spellings and phrasings

"Ft. Snelling"

- Fort Snelling
- Fort Snelling State Park (Dakota County)
- Fort Snelling State Park (Hennepin County)
- Fort Snelling Unorganized Territory
- Unorganized Territory of Fort Snelling
- Fort Snelling National Cemetery
- Fort Snelling Golf Course

Future Directions