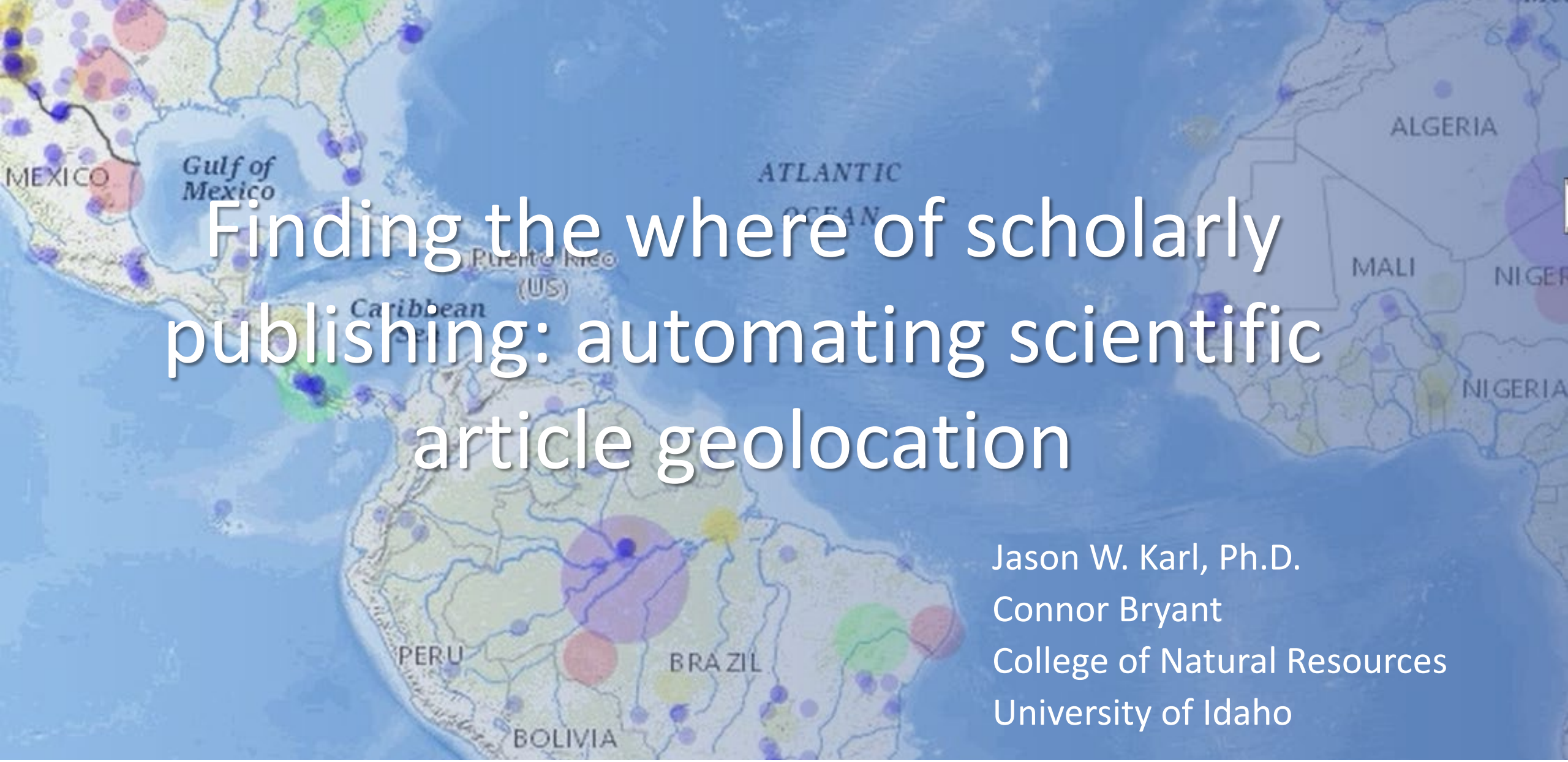


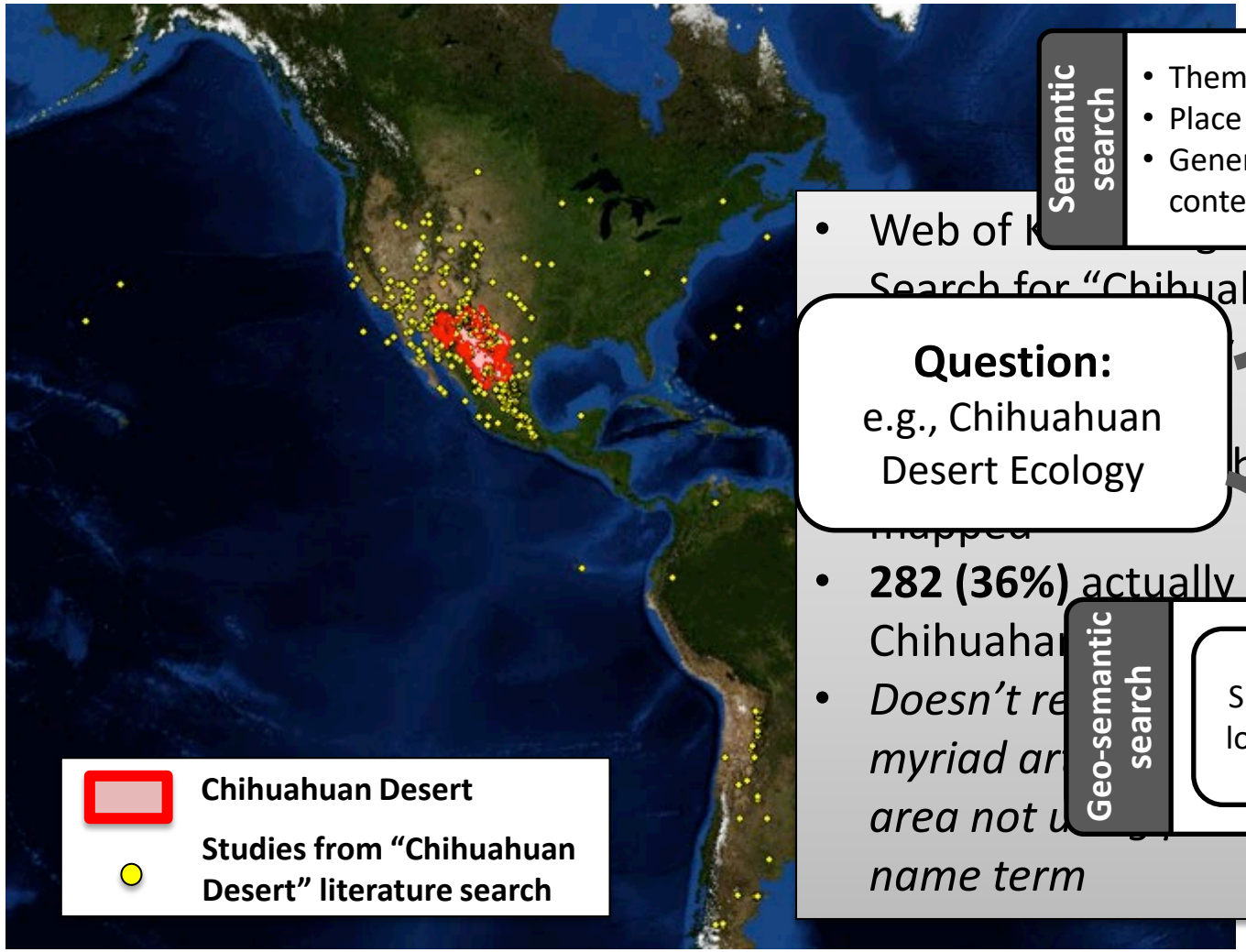
Finding the where of scholarly publishing: automating scientific article geolocation

Jason W. Karl, Ph.D.
Connor Bryant
College of Natural Resources
University of Idaho

JOURNALMAP
Research. Reimagined.

University of Idaho

What happens if you want to search for research from a specific place?



Question:
e.g., Chihuahuan Desert Ecology

- Web of K… Search for "Chihuahu…n
- …be mapped
- **282 (36%)** actually in Chihuahan…
- *Doesn't re… myriad ar… area not u… name term*

Semantic search
- Theme
- Place name
- General context

Geo-semantic search
- Specific location
- Similar areas

Results

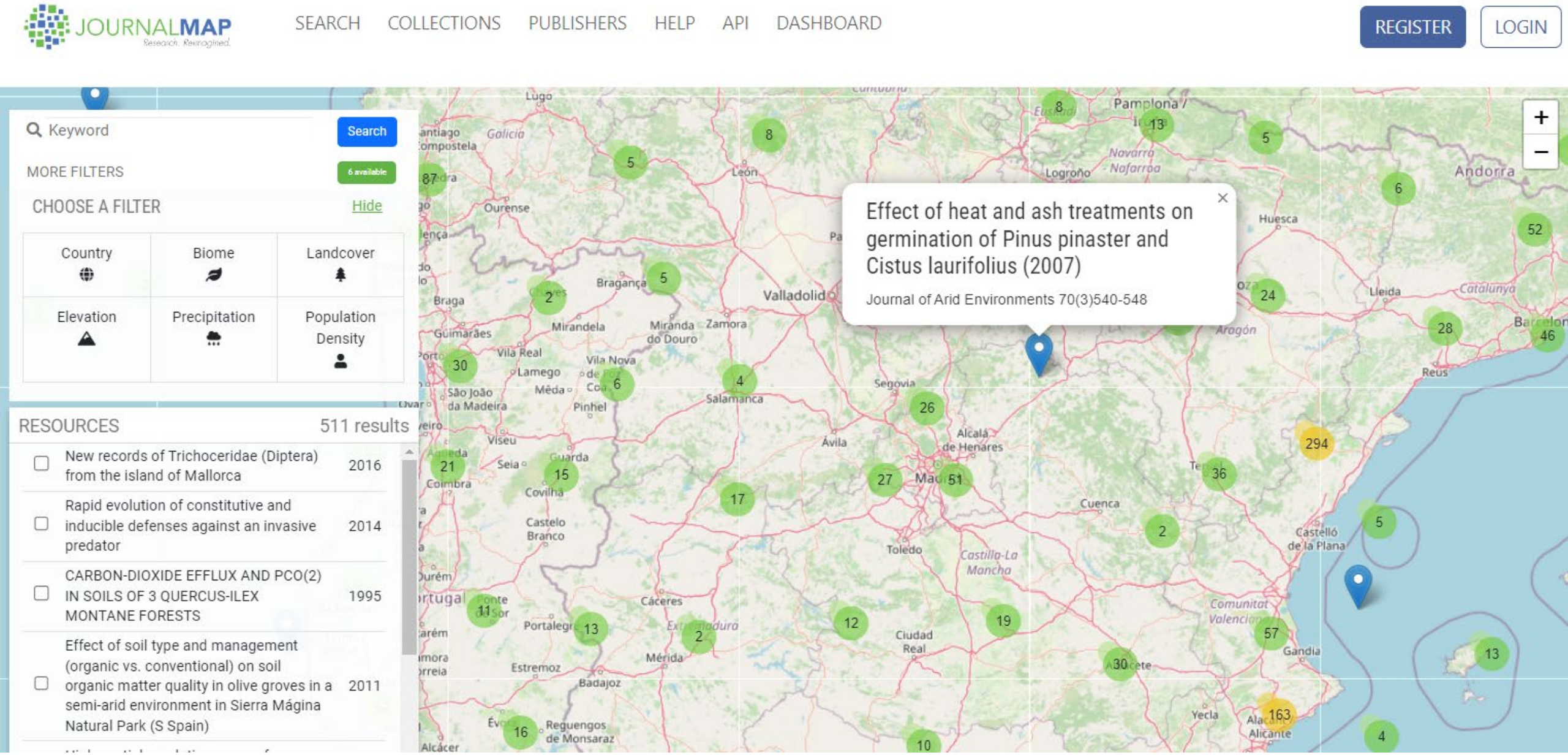
Relevant to theme

Relevant to location

Relevant by similar context

Source: Karl et al. (2013)

https://journalmap.org - Geographic-based literature searching



Challenge: Extracting location information from published studies

JOURNALMAP
Research. Reimagined.

>90% of ecology/agriculture studies report a study location, but use countless ways to do it*

<50% of ecology/ studies report geographic coordinates for their study area

tats within the Craters of the Moon National Preserve in southern Idaho, USA (Fig. 1). The Craters of the Moon area was created during at least eight volcanic eruptive periods that blanketed portions of the Snake River Plain with basaltic lava flows between 15,000 and 2,000 years ago (Owen 2008). Within the area of the lava flows, some areas were not covered by the lava and formed kipukas—islands of vegetation surrounded by lava. Two of these kipukas, Little Park (3,150 ha) and Laidlaw Park (35,040 ha) were sampled for this study.

2. Study area

For this study we considered the Bureau of Land Management's (BLM) 97,308-ha Wildhorse grazing allotment in southern Idaho (Fig. 1, 43.028°N, 113.864°W). The majority of the study area is in public ownership with the Bureau of Land Management (BLM) being the largest single land steward—managing approximately 93,317 ha (95.8%) of the study area. Approximately 1305 ha (1.3%) of the study area is in private ownership, and 2843 ha (2.9%) managed by the state of Idaho. The dominant land use in the study area is cattle and sheep grazing.

STUDY AREA

This study was conducted in the West Central Sage-grouse planning area in west-central Idaho, which included parts of Washington, Adams, Gem, and Payette counties (centered at 44°26′N, 116°38′W). The 374,700 ha planning area was established to conserve a small and isolated population of sage-grouse that was considered the most likely to be extirpated within the state (Idaho Sage-grouse Advisory Committee 2006). Exact population numbers are unknown but it was estimated that the population was significantly

Table 1. Locations of study sites and acquisition dates of the color infrared digital aerial photographs used in this study.

Study area	Site	Geographic coordinates	Average elevation, m	Average annual precipitation,[1] cm	Photo acquisition dates
Idaho	1	Lat 43°12′35.9″N, long 116°44′15.6″W	1 227	28	27 August 2008
	2	Lat 43°6′32.1″N, long 116°46′33.8″W	1 629	69	27 August 2008
	3	Lat 43°3′58.1″N, long 116°45′23.2″W	2 082	89	27 August 2008
Nevada	1	Lat 36°21′45.4″N, long 114°25′1.9″W	366	15	17 March 2009
	2	Lat 36°22′12.5″N, long 114°26′50.9″W	470	16	17 March 2009
	3	Lat 35°17′50.7″N, long 115°33′14.0″W	1 601	24	17 March 2009
New Mexico	1	Lat 32°34′37.0″N, long 106°0′42.0″W	1 230	26	23 August 2008
	2	Lat 32°29′38.8″N, long 105°40′59.7″W	1 641	48	23 August 2008
	3	Lat 32°22′33.8″N, long 105°39′25.6″W	1 440	37	23 August 2008

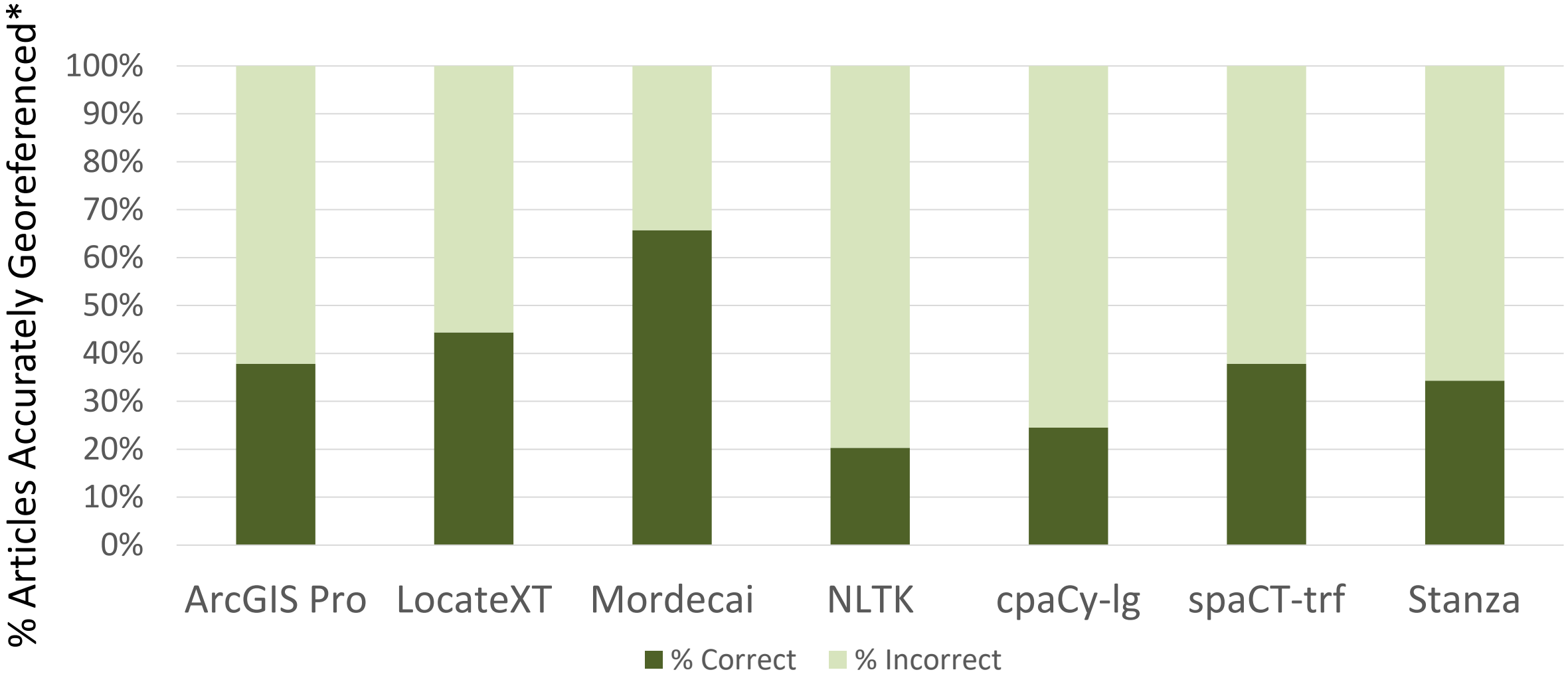
[1]Average annual precipitation from 1971 to 2000 from the 800-m resolution US Average Annual Precipitation layer, PRISM Group at Oregon State University (http://prism.oregonstate.edu, Accessed July 22, 2011).

*Examples from papers I have published. I'm part of the problem!

Article Geotagging

- Locations can be extracted from article text
- Pattern matching for coordinate values
  - JournalMap >80 coordinate formats
- Place-name recognition is challenging
  - Studies have many irrelevant place names
- ***The Challenge:*** develop approaches to identify and extract locations from journal articles based on place names.
  - Irrelevant place names
  - Uncommon/non-standard place names
  - Things that look like place names
    - E.g., “Canada Goose”

Performance of Natural Language Processing Tools for Article Geotagging



Godfrey, Kenyon, and Karl (*in prep*)

*Based on comparison to confirmed locations of 120 PLoS One articles.

Evaluating LLM Tools for Article Geotagging

JOURNALMAP
Research. Reimagined.

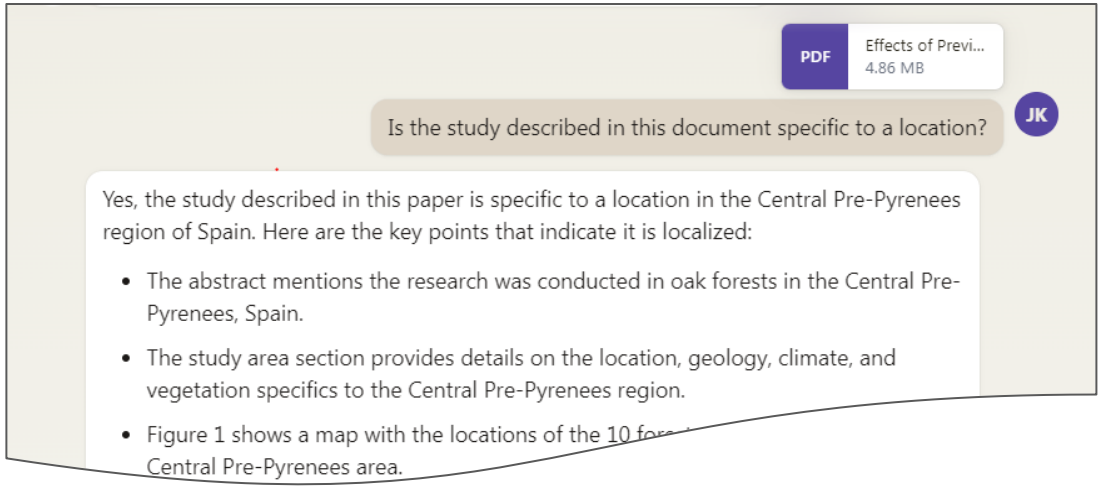
- 3 LLM tools
  - ChatGPT 3.5 (OpenAI)
  - Bard (Google)
  - Claude 2 (Anthropic)
- 120 PLoS One articles
  - Confirmed locations
- Prompt development
  - Uploaded PDFs
- Measured indicators
  - # articles returning locations
  - Average distance to confirmed location
  - Repeatability of LLM results

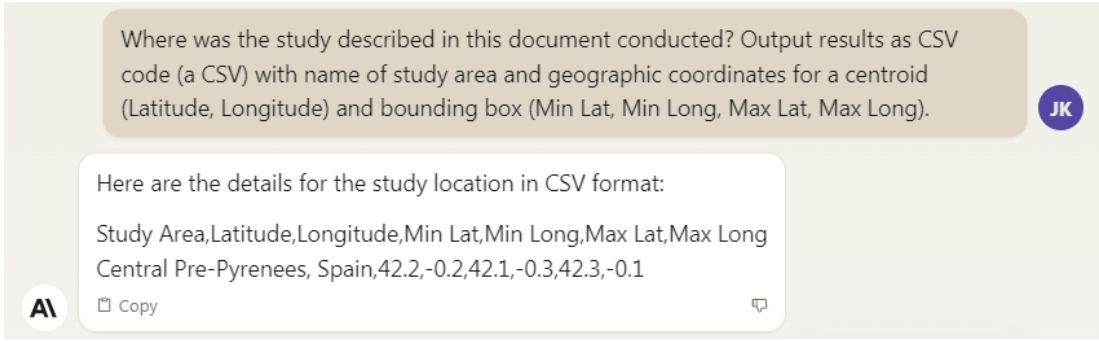
Can we use modern AI tools like ChatGPT to automate article geotagging?

Confirmed locations of 120 PLoS One articles used in Karl (2018) paper

LLM Article Geotagging Results

	% Articles Returning a Location	% Articles Accurately Located	Average Distance from Conf. Location
Bard	85%	64%	272km
ChatGPT	70%	69%	174km
Claude	**95%**	**95%**	**33km**

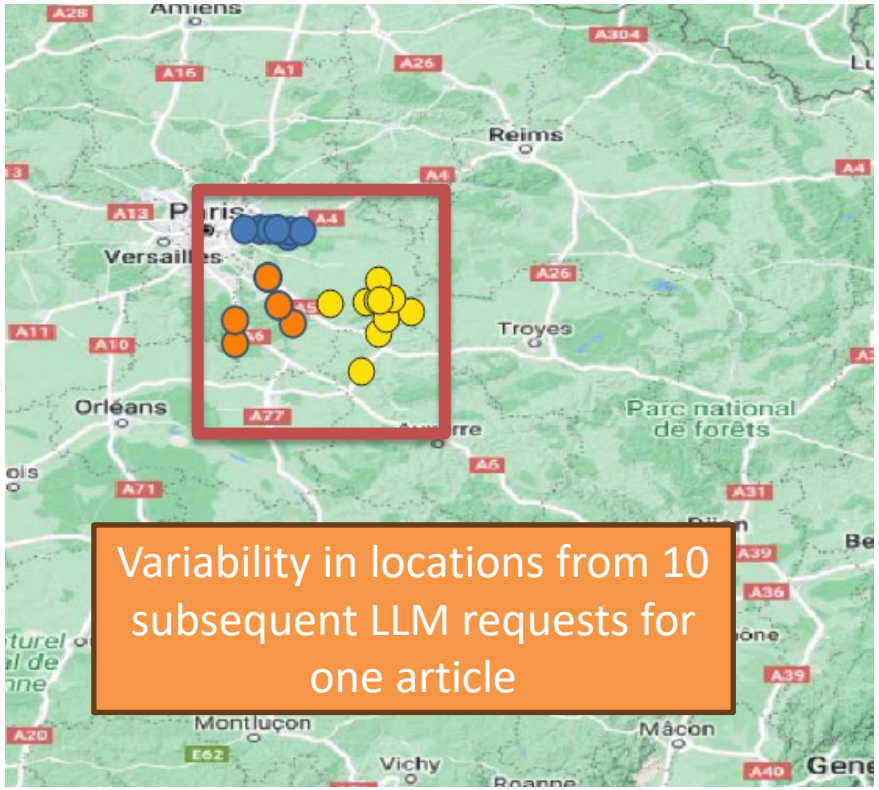




Example prompts and responses from Claude 2

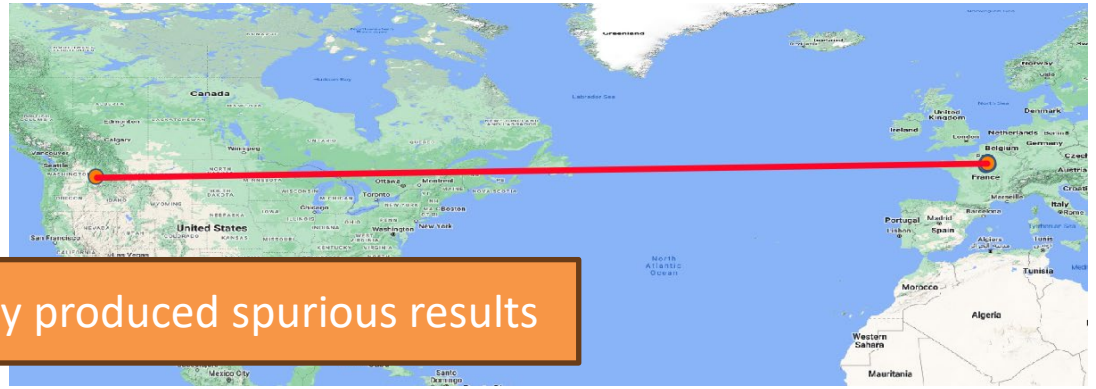
LLM Geotagging Repeatability

	Avg Bounding Width	Avg Bounding Height	% Returning Location	Avg # Success (out of 10)
ChatGPT 3.5	48.72km	147.56km	82%	7.5 (range 1-10)
Bard	639.73km	1800.00km	91%	10
Claude 2 Pro	7.67km	49.73km	100%	10

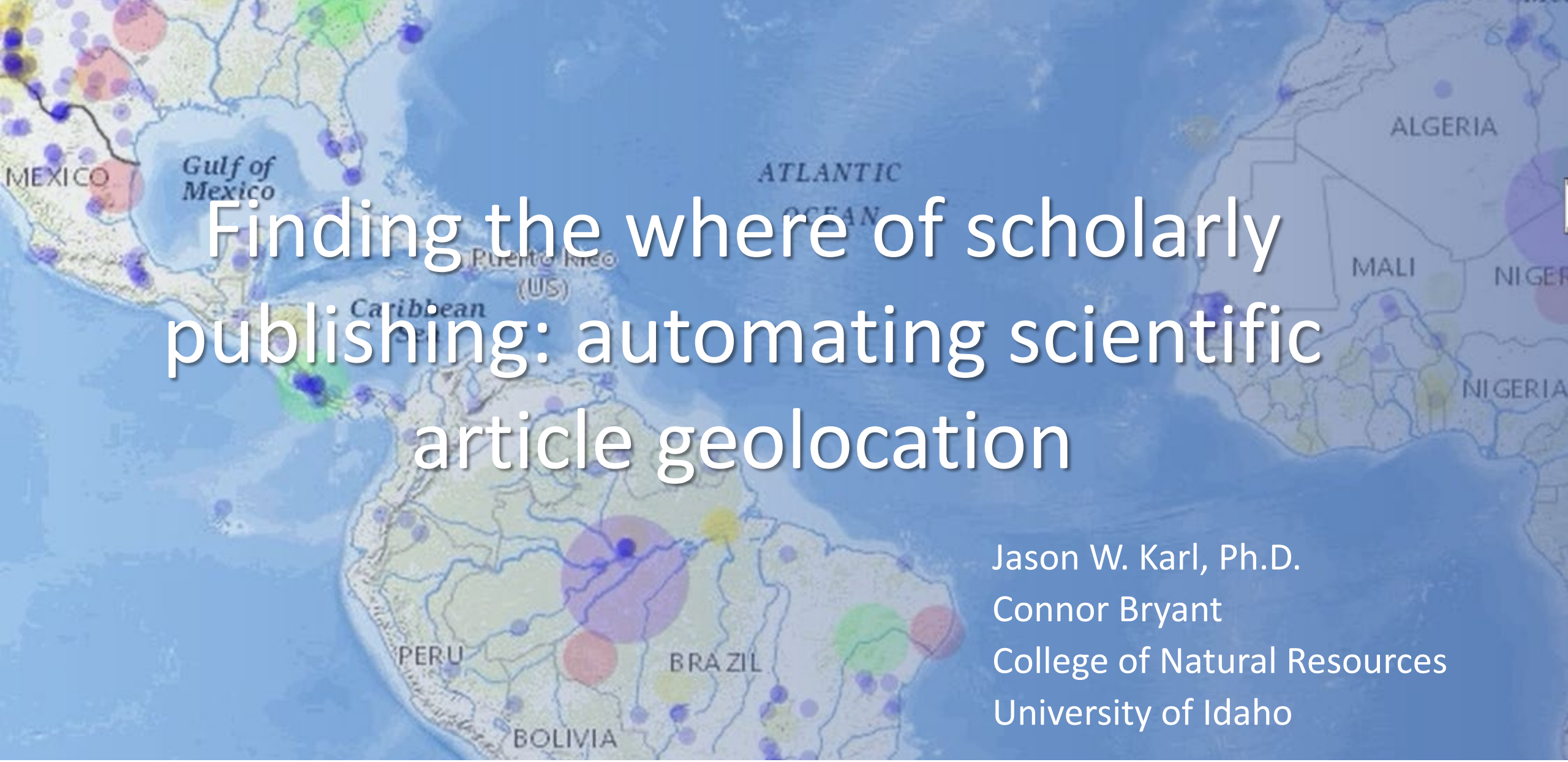


Variability in locations from 10 subsequent LLM requests for one article

- LLMs are generative AI tools
  - May produce different results to the same prompt
- Locations requested 10 times for a set of 10 articles
- Generally, variability was minor for repeat requests
- Claude 2 performed the best



Bard occasionally produced spurious results



Finding the where of scholarly publishing: automating scientific article geolocation

Jason W. Karl, Ph.D.
Connor Bryant
College of Natural Resources
University of Idaho

JOURNALMAP
Research. Reimagined.

University of Idaho