R-based solutions for synthesizing archaeological survey data to assess changing land-use patterns in the Okanagan-Wenatchee National Forest, WA

Grant Snitker[1,2], Sean Bergin[3], Pete Cadena[4]

[1]ORISE / US Forest Service Southern Research Station; [2]University of Georgia, [3]Arizona State University; [4]US Forest Service, Okanogan-Wenatchee National Forest

Project Overview

Archaeological research has benefited from decades of site-specific projects, regional comparisons, and theory building from case studies. However, recent research themes concerning the emergence of complex social-ecological systems and long-term land-use legacies require a new approach to archaeological data. Large-scale syntheses of archaeological, paleoenvironmental, and geographical information provide an effective way forward to address these themes. In more concise terms— 'big questions' often require 'big data' to help answer them. The cultural resource management data collected by the US Forest Service (USFS) is one such 'big dataset' and represents an incredible investment in time, resources, and expertise. This poster presents the initial results of a pilot study to develop an R-based workflow to digitize, extract, and synthesize archaeological information across the entirety of the Cle Elum Ranger District, within Central Washington's Okanagan-Wenatchee National Forest. Our results indicate that synthesizing district-level archaeological data reveals patterns of land-use and survey coverage that were otherwise not recognizable. This work has the potential to not only strengthen this dataset's role in forest-wide cultural resource management, but also to reposition cultural resources as a valuable tool in creating knowledge and developing policy with direct influences on the health of human-environmental relationships in the future.

Objectives

- Compile and digitize decades of archaeological site and survey data from USFS Ranger District databases.
- Develop automated, scripted tools in R to extract meaningful information for these records that might not be documented in existing spatial USFS spatial databases.
- Customize a package in R to accommodate for the diversity of USFS site forms and non-USFS site forms.
- Leverage datasets created through R tools to evaluate district-wide site density, survey coverage, and changing land-use patterns.
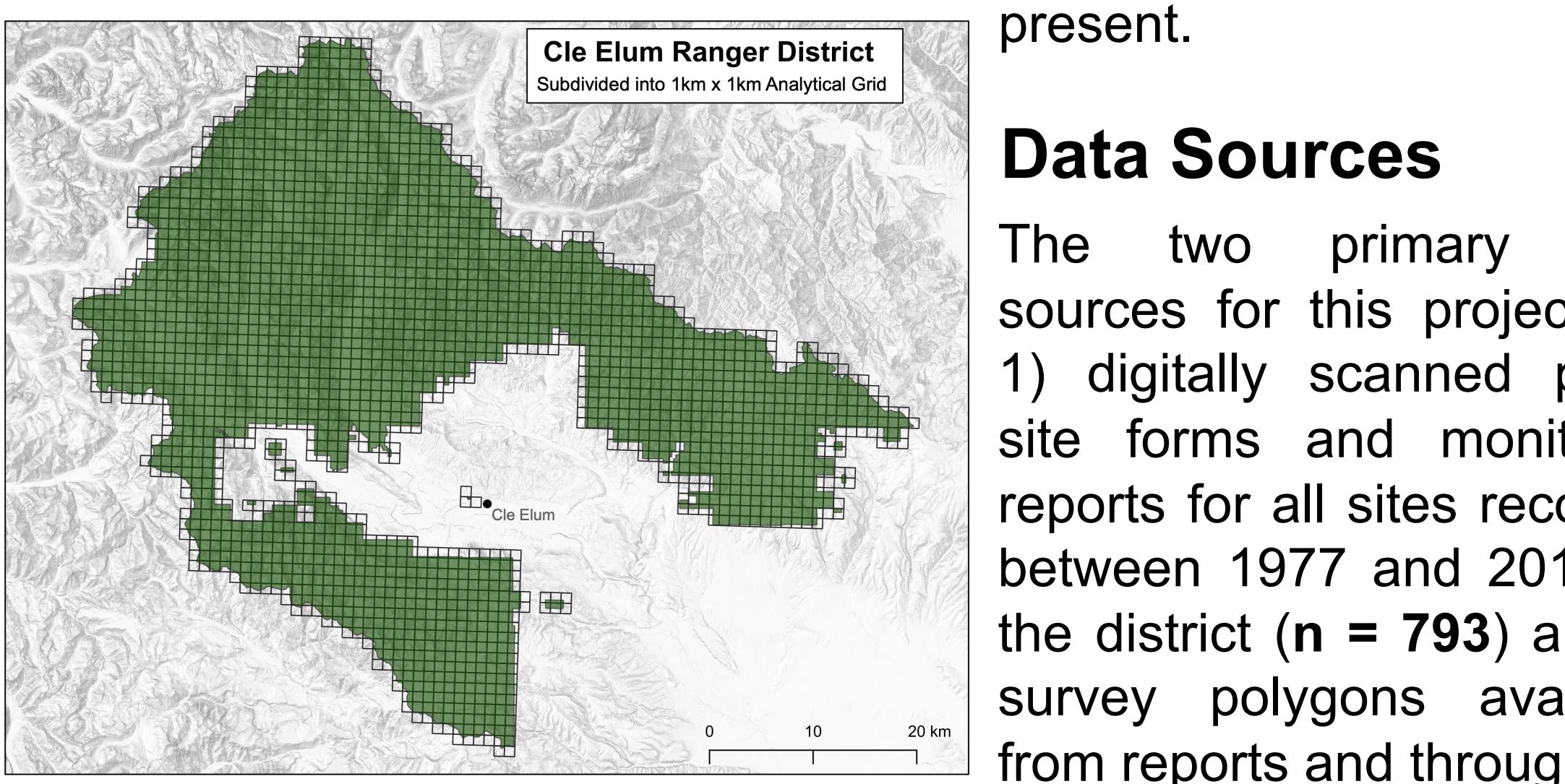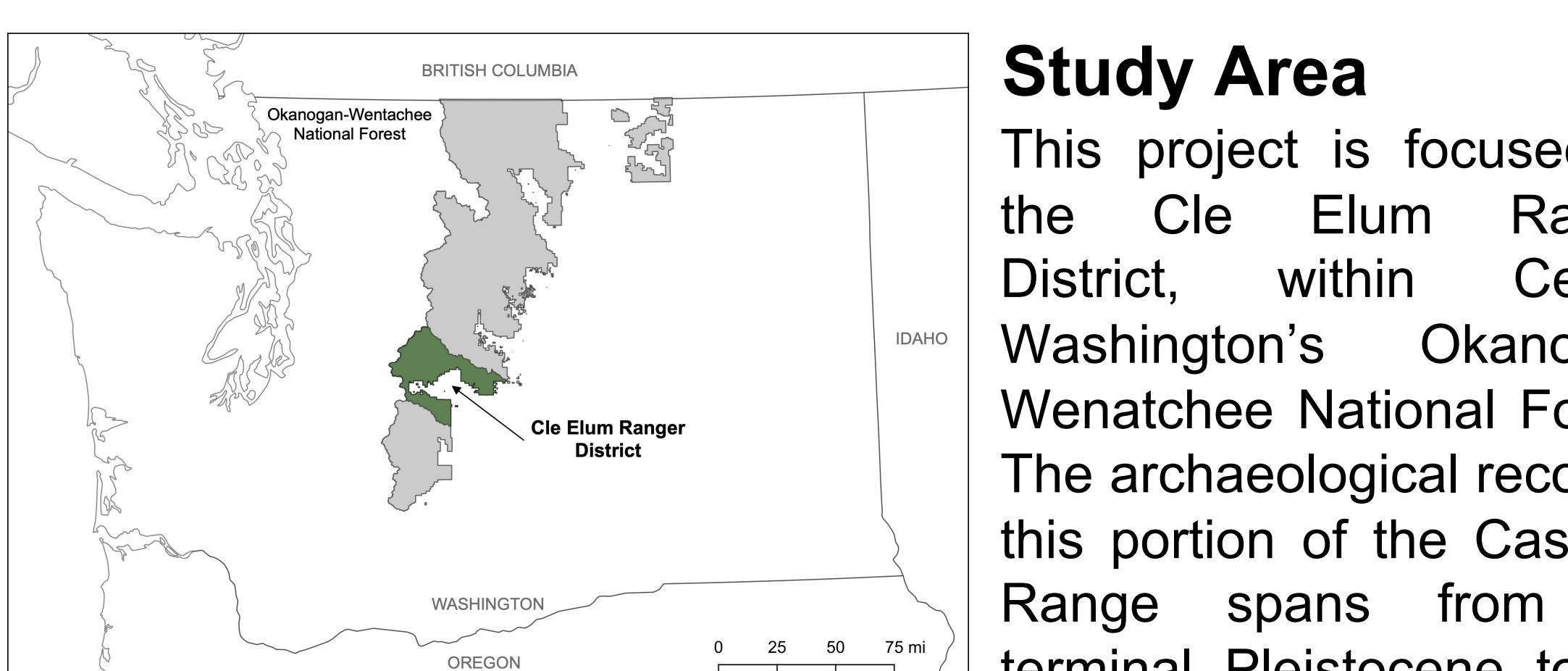
Study Area and Data Sources

Study Area

This project is focused on the Cle Elum Ranger District, within Central Washington's Okanagan-Wenatchee National Forest. The archaeological record in this portion of the Cascade Range spans from the terminal Pleistocene to the present.

Data Sources

The two primary data sources for this project are 1) digitally scanned paper site forms and monitoring reports for all sites recorded between 1977 and 2018 on the district (**n = 793**) and 2) survey polygons available from reports and through the Washington Department of Archaeology and Historic Preservation. Note that the study area is subdivided into 1km x 1km grid cells to facilitate spatial synthesis and to obscure exact site locations.
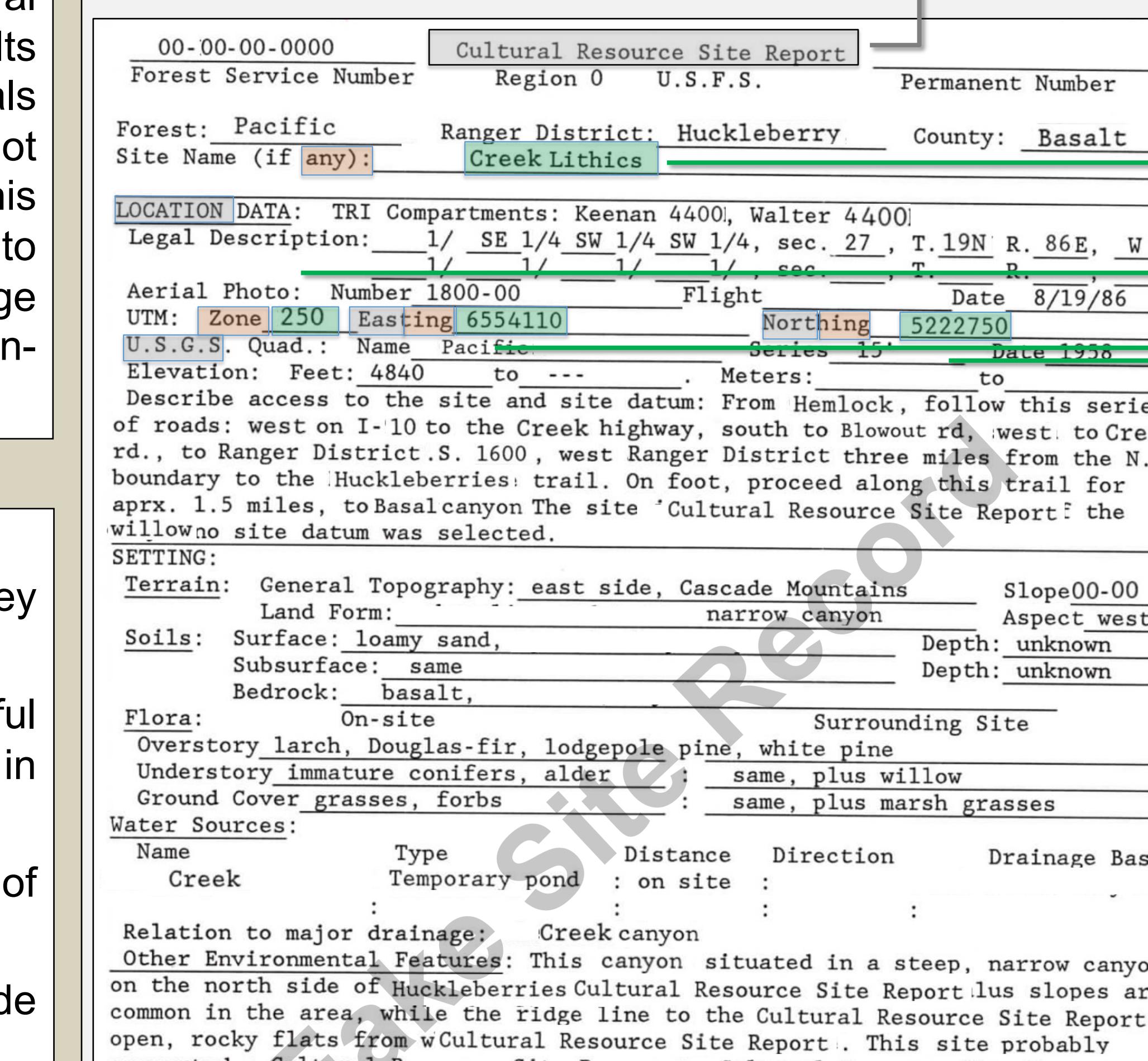
The Site Record Analyzer (SRP) Package for R

The **Site Record Processor (SRP)** is a R package for extracting pertinent site information from archaeological site records stored as PDFs and outputting that information as a dataframe. The package is provided for the use or modification by other researchers and is available on Github.

Broadly, the **SRP** package loads site record PDFs form a directory selected by the user, converts the PDFs to images, and then uses the open-source optical character recognition (OCR) engine Tesseract to store each word in the site record as a vector, or list of words that can be parsed/extracted.

Since the formatting for site forms has changed over time, the type of site form is identified by the header or form specific formatting. Once the type of document has been identified, a new list of keywords is used to locate and record site information. The **SRP** package continues this process for each site; recording location information, period of use, site type or use, and the any other information that we wish to extract from the PDF site form.

Site Record Processor (SRP) Workflow

1. Identify the type of site form based on the document's header

2. Save information based on the words that come before and after the information we wish to record based on a "bookend" approach:

	Word 1	Word or Phrase	Word 2
Site Name:	any):	Creek Lithics	location
UTM Zone:	Zone	250	Easting
Easting:	Easting	6554110	Northing
Northing:	Northing	5222750	U.S.G.S.
Period:	possible):	prehistoric, historic	How
Site Use:	Type/Function/Use:	temporary camp	How

3. Record Information from the document to a database of all the scanned sites
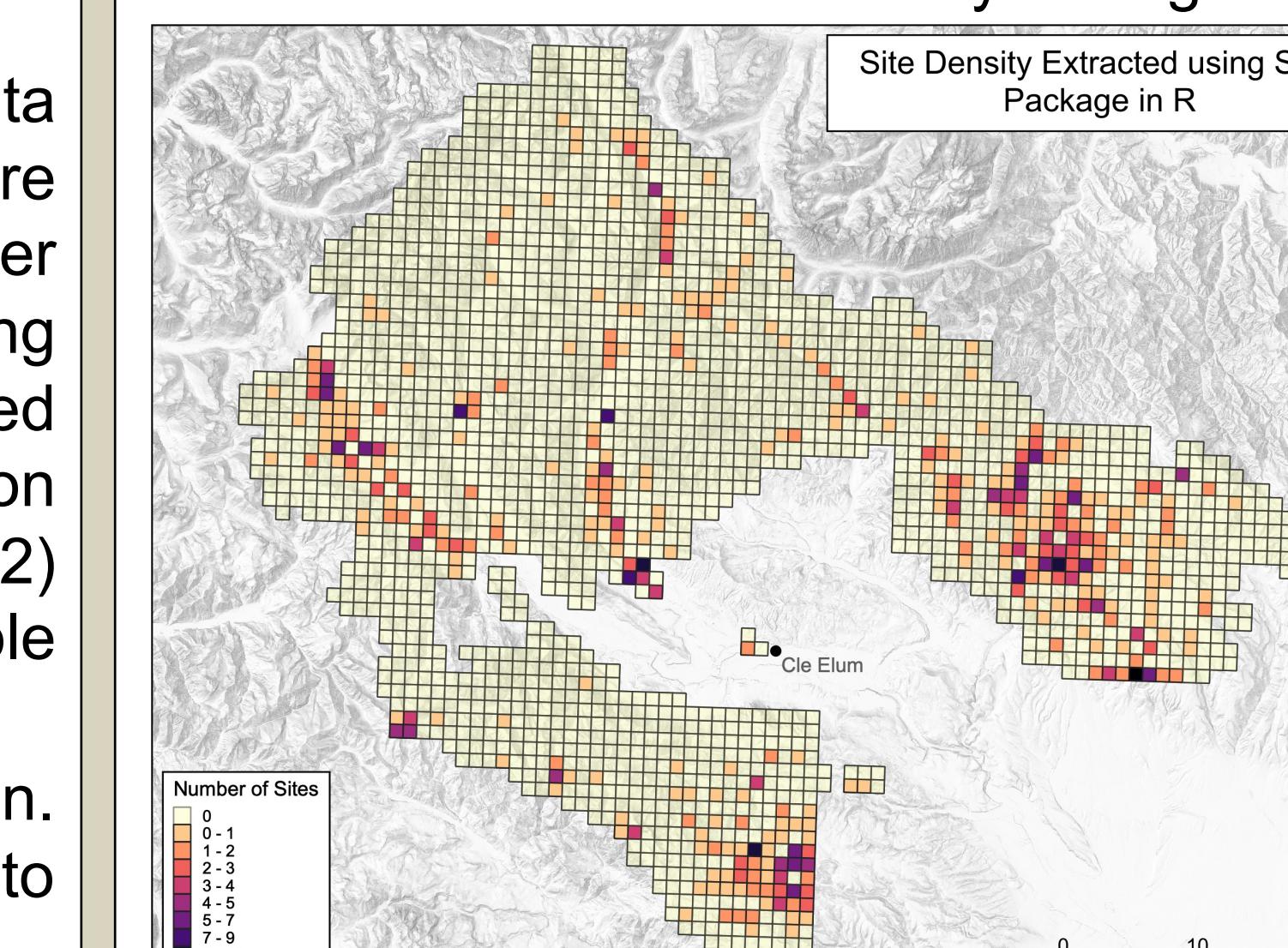
| Site Number	Site Name	Zone	Easting	Northing	Period	Use
00-00-00	Creek Lithics	250	6554110	5222750	Prehistoric, historic	Temporary camp

4. This process continues with each site record in the specified directory.
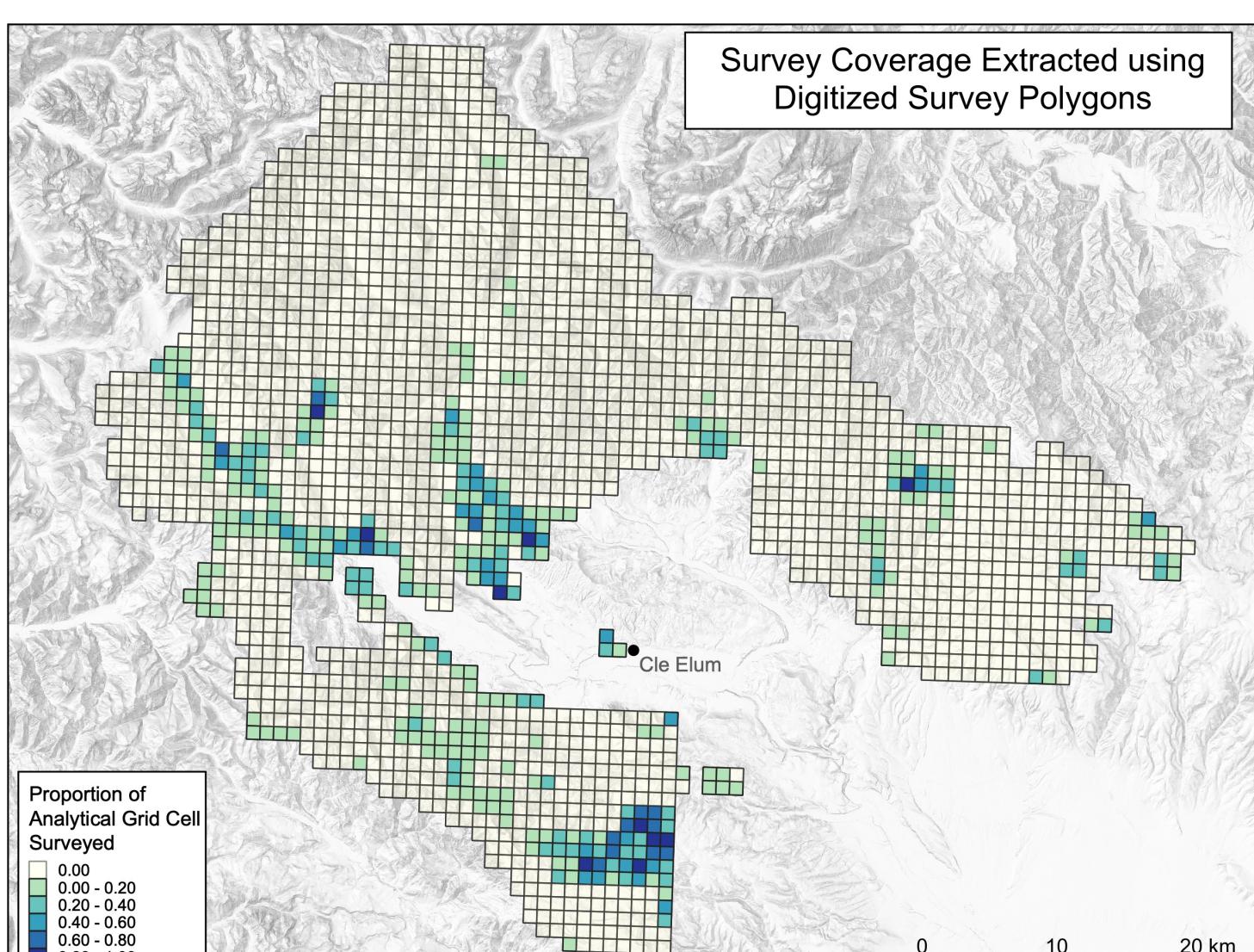
Using the SRP Package to Synthesize Data

1. Evaluating USFS District-wide Site Density and Survey Coverage

Information for each site was extracted and compiled from digitized, paper site forms using the SRP package in R (v. 3.5 or above). Additionally, survey coverage and extent data were complied from digitized reports or through the Washington DAHP WISAARD online repository. Site locations extracted using SRP were used aggregate sites into 1 km x 1km grid cells to evaluate overall site density throughout the Cle Elum Ranger District (**left**). Since site records were recorded over a period of 40 years, the availability/accuracy of GPS units, map projections, and other factors can influence site location accuracy. For this reason, only the latest coordinates extracted by SRP were used. Additionally, survey polygons coverage per 1km x 1km grid cell were evaluated to generate a map of survey coverage for the Cle Elum Ranger District (**right**).
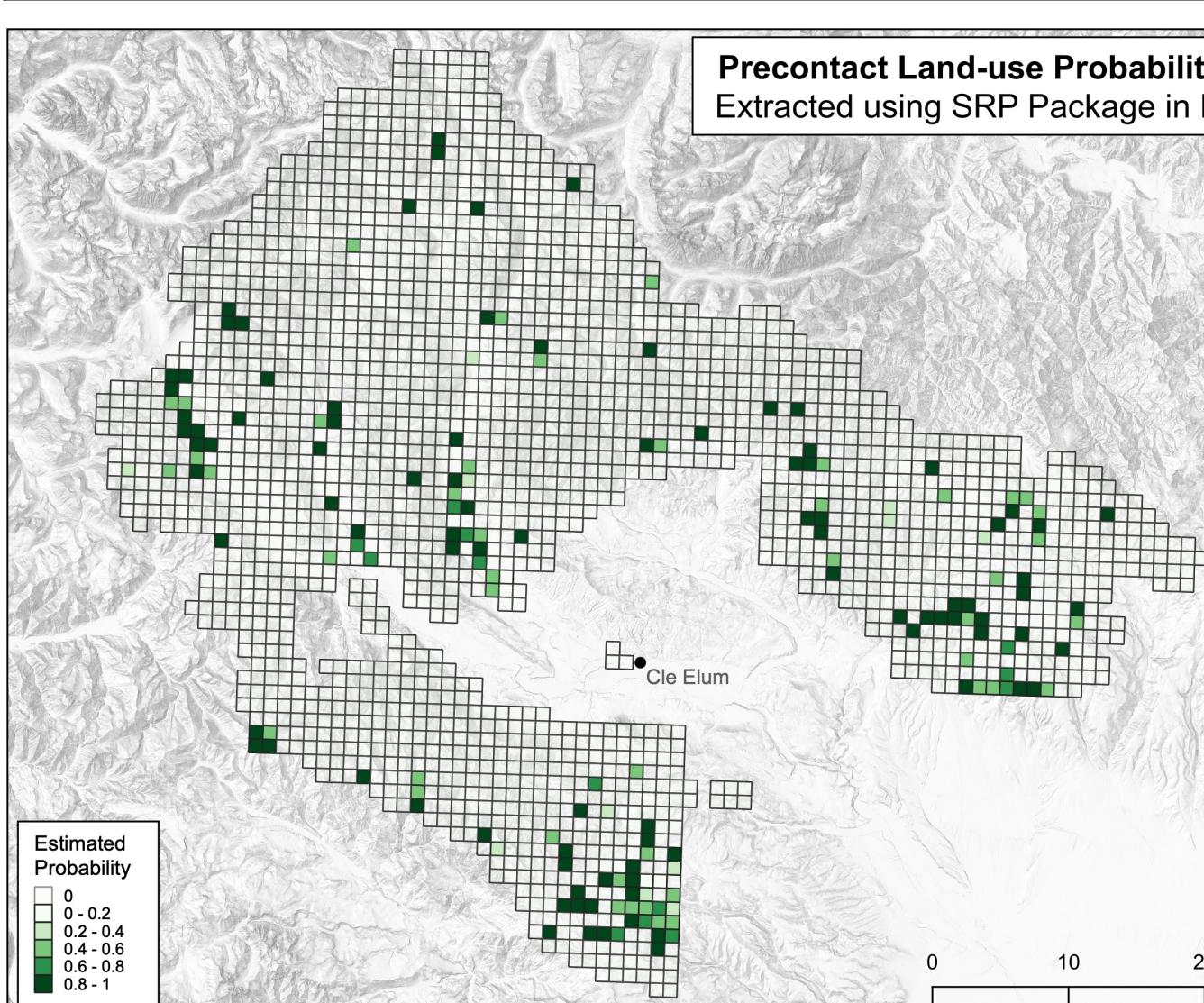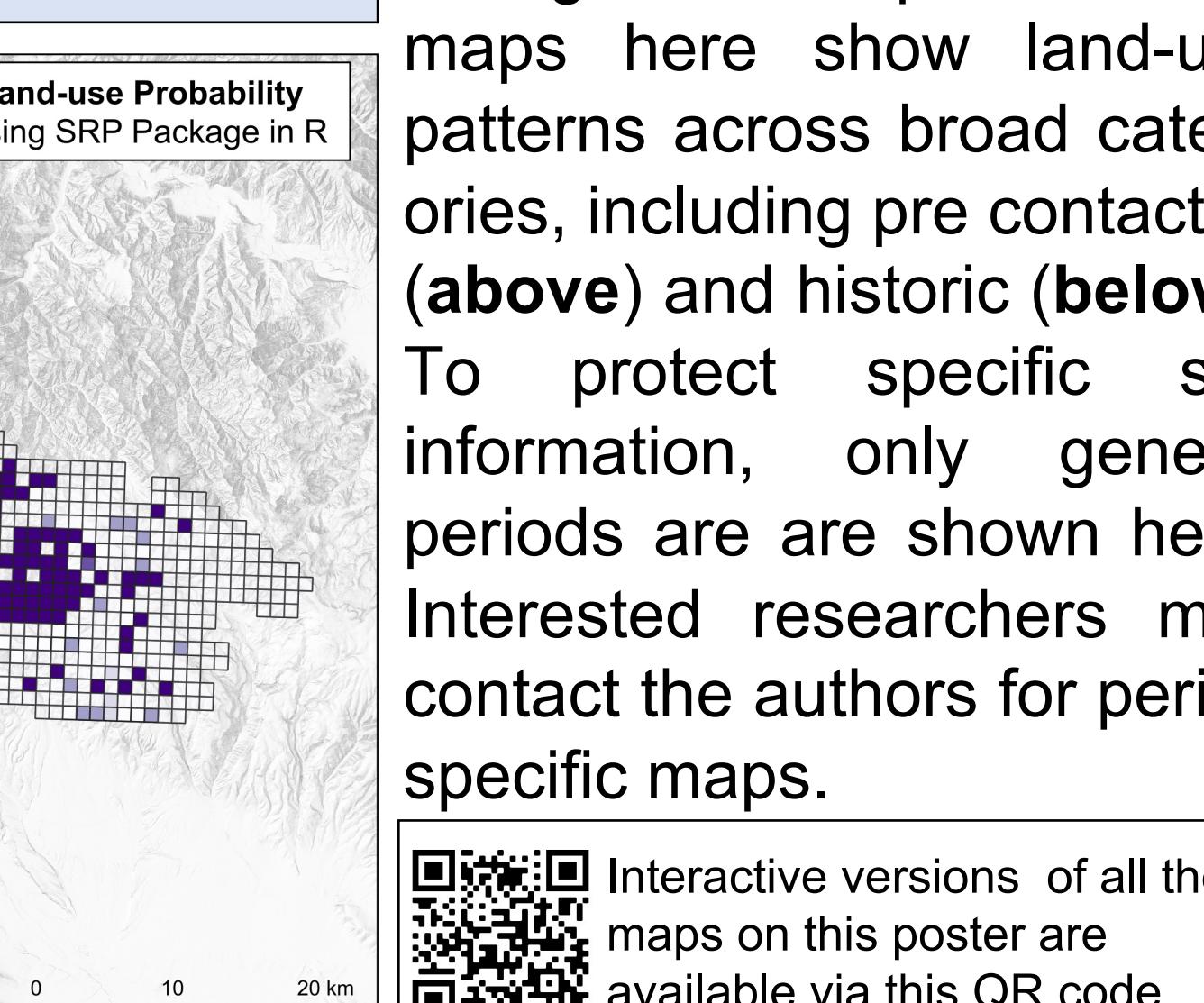

Using the SRP Package to Synthesize Data

2. Estimated Land-use Patterns by Archaeological Period

Information about each site's artifact assemblages, diagnostics, and the archaeological period assigned by the recording archaeologist are compiled by the **SRP** package. Each site is assigned a weighted probability that it belongs to an archaeological period based on its diagnostic artifacts, overall assemblage, and dating methods recorded in the site form. Probabilities are then aggregated up into a 1km x 1km grid cells across the Cle Elum Ranger District, so that each cell had an overall probability that its composition dates to each archaeological period. For example, a grid cell that had a single site with a Cascade Point (8000 - 5000 BP) has a 1.00 probability of dating to the Middle Archaic and a 0.00 probability of dating to other periods. The maps here show land-use patterns across broad categories, including pre contact (**above**) and historic (**below**). To protect specific site information, only general periods are are shown here. Interested researchers may contact the authors for period specific maps.

Pre-Contact Periods	Date Ranges
Late Archaic	2000 BP - AD 1720
Middle Archaic	5000 - 2000 BP
Early Archaic	8000 - 5000 BP
Paleoarchaic	11,000 - 8000 BP
Pre-Clovis / Clovis	15,000 - 11,000 BP

Historic Periods	Date Ranges
Post-War Recreation	AD 1945 - 1970
CCC / Great Depression / WWII	AD 1930 - 1945
Commercial Period (Mining/Logging/Trapping)	AD 1880 - 1930
Contact / Euro-American Exploration / Fur Trapping	AD 1720 - 1880

Interactive versions of all the maps on this poster are available via this QR code.

Future Work

- Expand functionality of the SRP package to accommodate for additional site record forms formats.
- Extract and compile additional information from site record forms, such as environmental descriptions, site dimensions, and artifact descriptions.
- Deploy SRP package to examine site records on other districts within the Okanagan-Wenatchee National Forest or other National Forests.

Download and Test SRP Package for R

Download and test the Site Record Analyzer (SRP) Package for R via Github here: **https://github.com/seanbergin/srp** or in R:

```
1 library(devtools)
2 devtools::install_github("seanbergin/srp")
```

Contact

Get in touch with questions or comments

grant.snitker@uga.edu
sean.bergin@asu.edu
pete.cadena@usda.gov

https://grantsnitker.github.io/
https://asu.academia.edu/SeanBergin