

The Future of the SCEC Community Information System

PRESENTER: **Edric Pauk**

OBJECTIVE

Our goal is to **adapt** the SCEC Community Information System (CIS) to meet the **evolving needs of the SCEC community**, providing **better insights** and improved **access to data and software** to support advances in **earthquake science**.

FUTURE FOCUS AREAS

1. **usage/impact metrics for software**
Research software developers, sponsors, and other decision makers want to determine **who uses** software, **what problems are solved**, and the **broad impact** of the resulting work. More awareness **increases the user base** and **thus expands impact** and brings collaborators to the project, improving its longevity.

2. **persistent dataset storage & DOIs**
Datasets that are **potentially useful** to one or more audiences will be **curated**, stored with **well-defined schemas**, and have **robust metadata**. Data tagged with permanent identifiers also require **long-term storage** to ensure their continued availability.

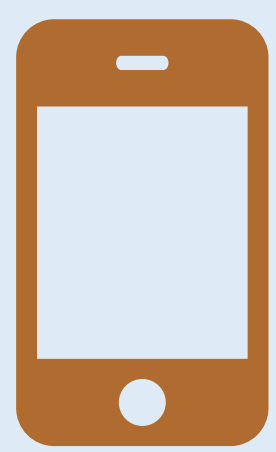
3. **enhanced data access tools**
Data access tools allow datasets to be **intelligently searched**, **visualized**, and useful **data subsets extracted**. Datasets that have a **large user base** and are **difficult to access** are **prime candidates** for new data access tools. The form of the tool (website, command line tool, API, etc.) depends upon the needs of the target audience.

CHALLENGES

1. **limited tracking data** from third-parties like GitHub, determining the **best impact metrics**, and the need for better **documentation, training**, and **user interfaces**.
2. **identifying datasets** that are ready for dissemination and locating **long-term storage** that is both persistent and cost effective.
3. varied **data formats**, and **identifying datasets** that would most benefit from new tooling.



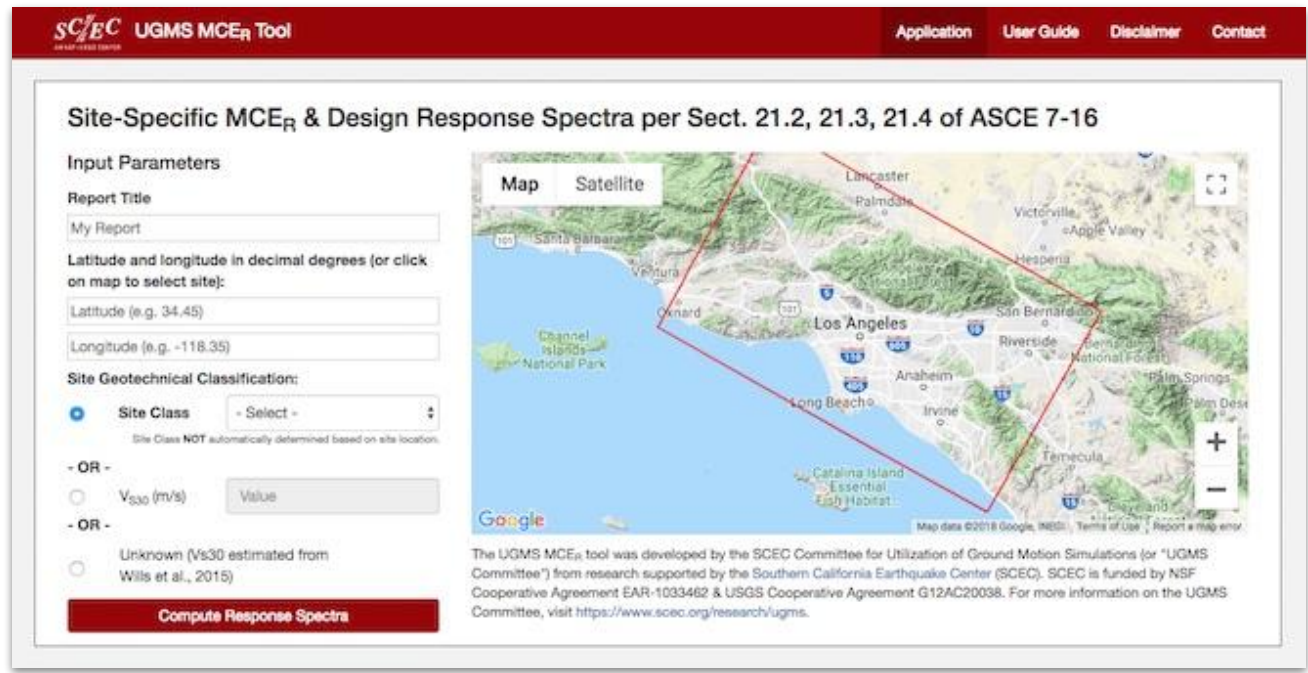
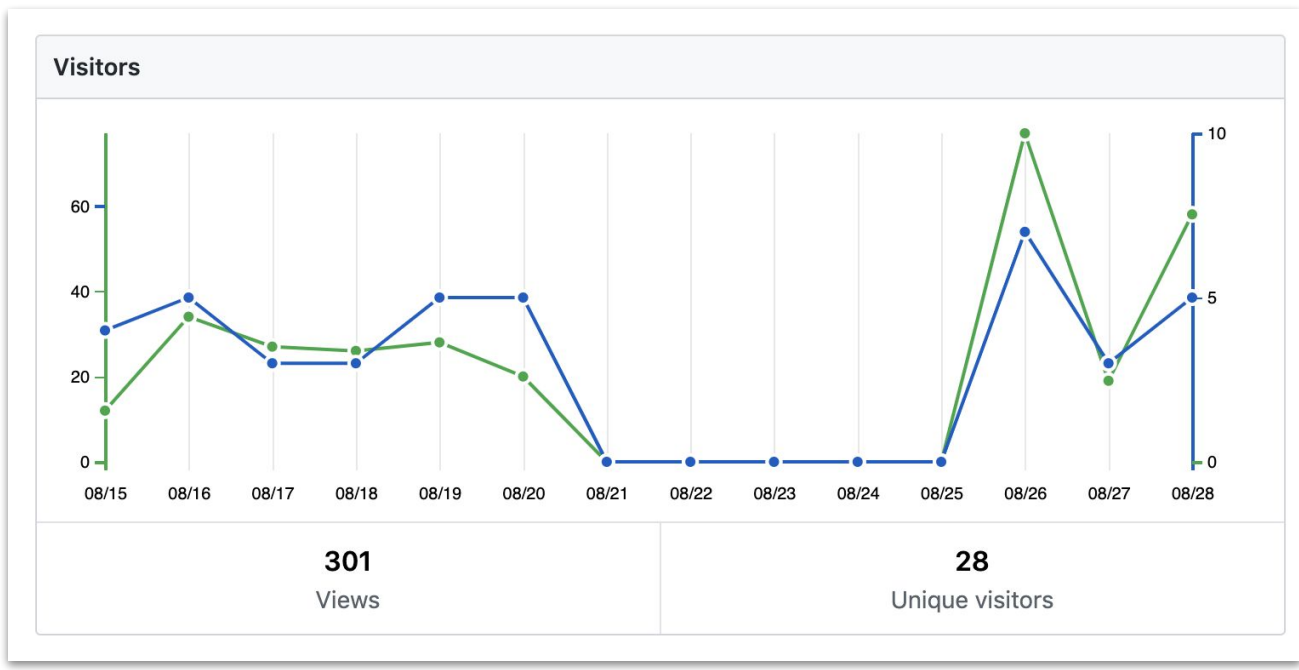
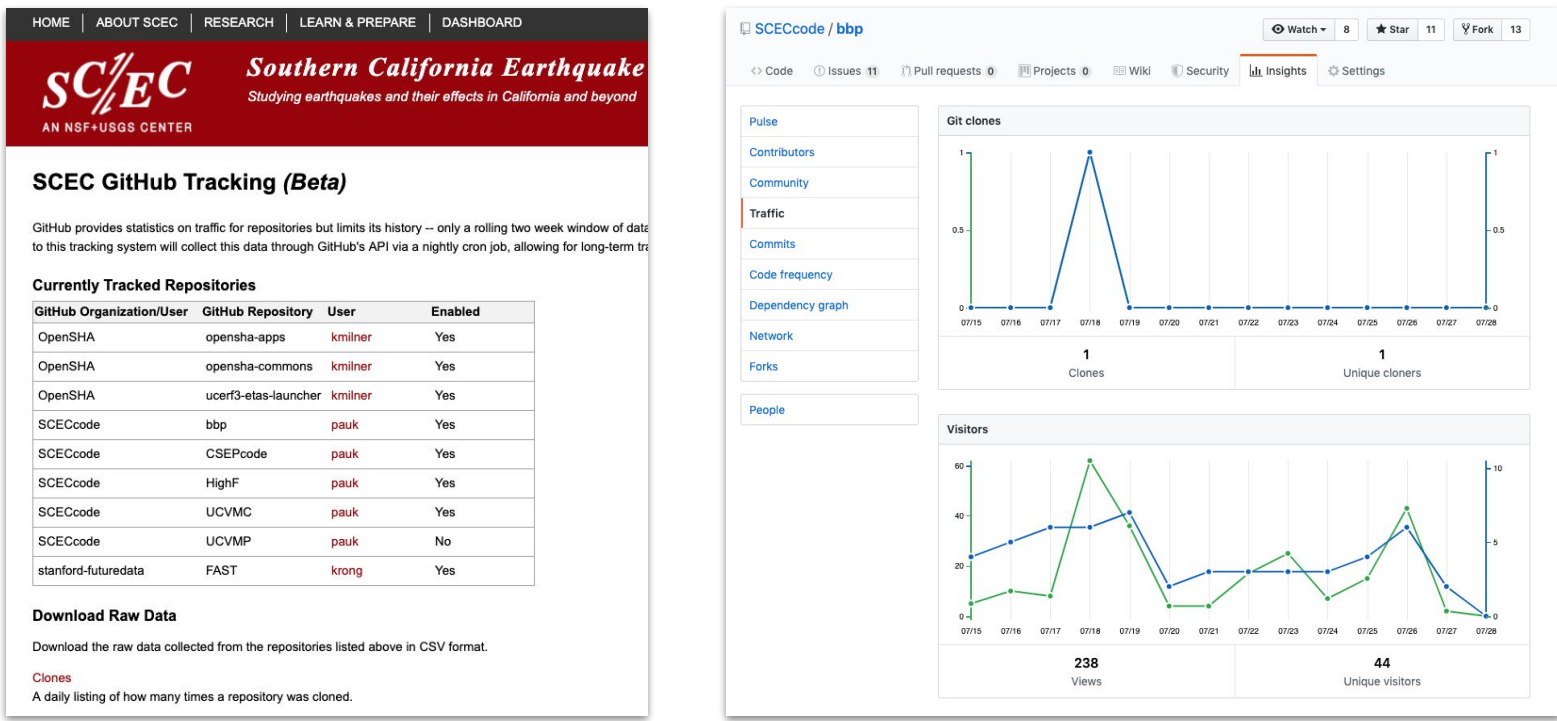
Help SCEC measure the impact of our research software, identify datasets, and develop data access tools for high-impact data delivery.



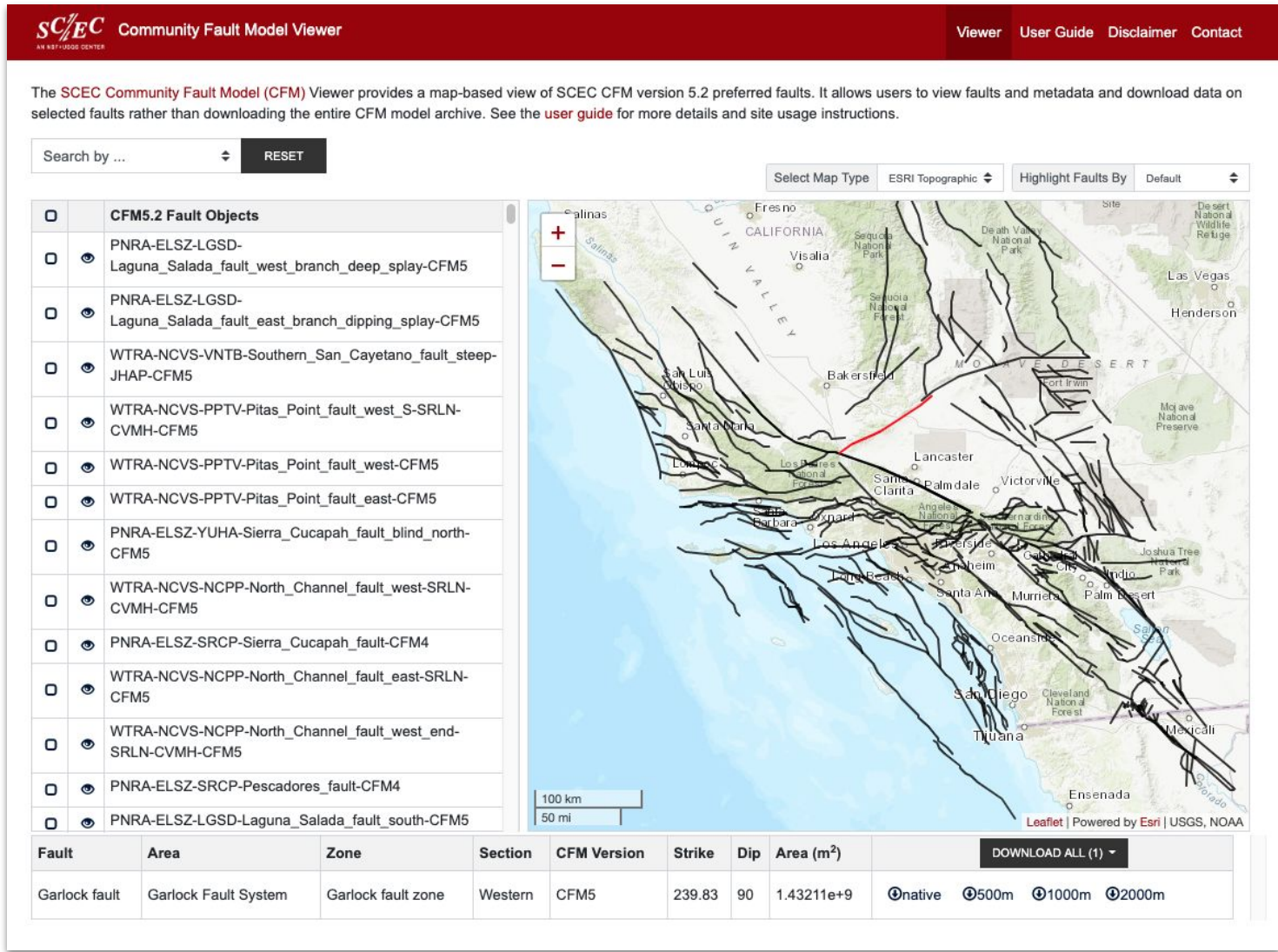
Scan to take our short survey online!

Answer 3 questions!

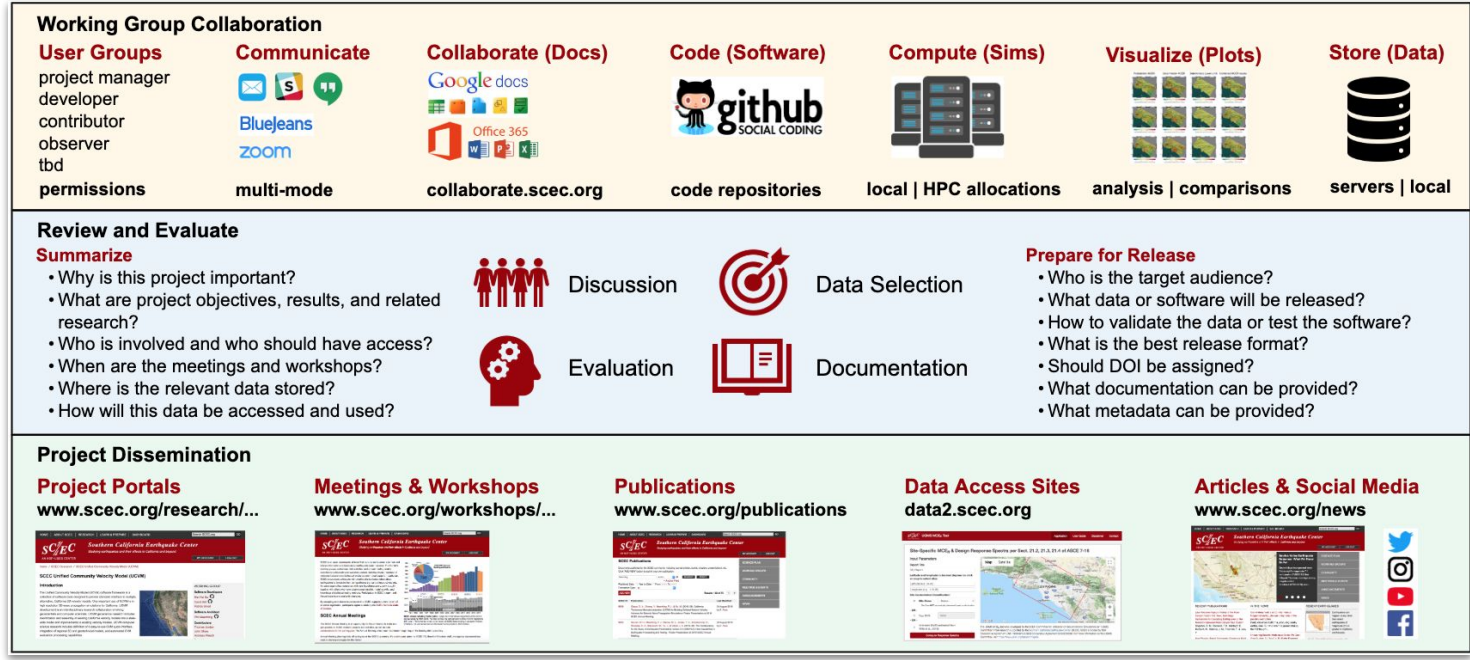
- 1) What metrics are important to measure the impact of research software?
- 2) How do you currently store datasets and ensure they contain robust metadata?
- 3) What datasets do you need access to right now that are unavailable or too cumbersome to obtain?



```
-bash-4.2$ ./plot_cross_section.py -b 34.0,-118.75 -u 34.0,-117.75 \
-h 100 -v 10 -d vs -c cvms -a s -s 0 -e 4000 -o cross.png
Using parameters:
lat1 = 34.0
data_type = vs
lat2 = 34.0
cvm_selected = cvms
ending_depth = 4000
horizontal_spacing = 100
color = s
lon1 = -118.75
vertical_spacing = 10
lon2 = -117.75
starting_depth = 0
Retrieving data. Please wait..
```



Fault	Area	Zone	Section	CFM Version	Strike	Dip	Area (m ²)	Download All (B)
North Channel fault-west	Western Transverse Ranges Area	North Channel-Ventura fault system	North Channel	CFM5	276.271	35.564	1.17709e+9	@raw @100m @200m
Santa Cruz fault	Peninsular Ranges Area	Elmore-Laguna-Santa Cruz fault zone	Vista Obispo	CFM5	305.867	89.4168	8.1769e+7	@raw @100m @200m
Pitas Point fault-east	Western Transverse Ranges Area	North Channel-Ventura fault system	Pitas Point-Ventura	CFM5	276.193	43.588	1.24809e+9	@raw @100m @200m
Garlock fault	Garlock Fault System	Garlock fault zone	Western	CFM5	239.83	90	1.43211e+9	@raw @500m @1000m @2000m
Pitas Point fault-west	Western Transverse Ranges Area	North Channel-Ventura fault system	Pitas Point-Ventura	CFM5	null	null		@raw @500m @1000m @2000m
Pitas Point fault-west-S	Western Transverse Ranges Area	North Channel-Ventura fault system	Pitas Point-Ventura	CFM5	null	null		@raw @500m @100m @2000m



Edric Pauk¹



Tran T. Huynh¹



Mei-Hui Su¹



Philip J. Maechling¹

¹ University of Southern California