An Analysis of the PEARC 2018 Science Gateways Community Institute Hackathon: Lessons Learned

Linda Hayden, Alexander Nolte, Jim Herbsleb, Je'aime Powell, Ei Pa Pa Pe-Than, Mona Wong, Rajesh Kalyanam, Kevin Ellet, Sudhakar Pamidighantam, Kathy Traxler, Amy Cannon

Abstract

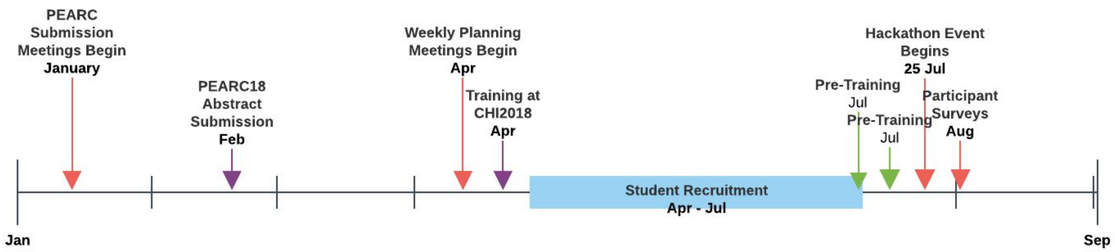
Science Gateways are web portals on which data, software, instruments, and computing resources provide users, and scientists streamlined workflows. These web portals allow scientist to focus on research with little to no need to utilize programmatic skills. Hackathons are time-bounded events in which collaborators intensely focus on a given subject or problem in an effort to generate solutions. Hackathon events provide opportunities for participants with diverse backgrounds, mixed specialties, and broad skill sets to interact in a manner that promotes disruptive solutions. Hackathons come in a variety of different forms, and serve many different purposes. These purposes include developing resources, infrastructure, practices, and culture for a community. The Science Gateways Community Institute team proposed a hackathon to introduce students attending the PEARC 2018 conference, to the concepts of Science Gateways. The aim of the hackathon was to stimulate interest in Science Gateways, while developing projects that were of value to researchers from different disciplines. We believed that PEARC would be an ideal venue for this type of event, due to the community of researchers, students, and practitioners from various disciplines that were interested and skilled in the high performance computing field attending the conference.

The Science Gateways Community Institute utilized a hackathon formatted event to provide students with an opportunity to collaborate and generate targeted skills while creating needed features for production web portals. The student participants included a group of 17 graduate and undergraduate students with predominately underrepresented minorities. The hackathon followed the PEARC 2018 conference held in Pittsburgh, PA. The science gateways involved during the event were MyGeoHub (Purdue University), Cosmic^2 (University of Michigan), SEAGrid (Indiana University), and SimCCS (Indiana University) with associated mentors serving as subject matter experts. This poster presents organizational methods and technologies used to coordinate the international planning team, lessons learned during the event, and deliverables presented by the student teams.

PROJECT DESCRIPTIONS

myGeoHub	https://mygeohub.org/ **Task #1**: Design and develop a Jupyter notebook to demonstrate interoperability with NCSA's BrownDog. **Task #2**: Design and develop a Jupyter notebook to collect and analyze social media (Twitter) activity during weather events **Knowledge**: Python & knowledge of JSON, Requests, pyLeaflet/Folium library
SEAGRID	https://seagrid.org/ **Task #1**: Build interface for existing applications Gromacsimum for gromacs within Seagrid **Task #2**: Build interface for existing applications Burai for QuantumEspresso within Seagrid **Knowledge**: JavaFX (for desktop), Python/Django (for web) To test a model interface using Beg.in software
COSMIC^2	https://cosmic2.sdsc.edu:8443/gateway/ **Task #1**: design and implement a client-based, image stack viewing app that will be launched by the gateway server. The app should allow user to create points on the individual image and return all points back to the server when user clicks on "Submit". Additionally if would be nice if the app allows the user to zoom, pan and display images in movie-mode. **Knowledge**: JavaScript (very helpful), HTML5 (helpful), Java (optional), cryo-electron microscopy
IU	https://geosurveyiu.scigap.org/ **Task #1**: Deploy mapping functions using JavaScript in a Web framework (Django) **Task #2**: Mapping Network layers of SimCCS sources, sinks and pipeline infrastructure **Task #3**: Georeferenced data ingestion for US, China and Australia **Task #4**: Selectable sets among the resources for specific scenarios

TIMELINE



POST EVENT SURVEY RESULTS

[*Note: The codes are for the Cosmic^2 (Cx), MyGeoHub (Gx) and SimSSC (Sx) gateways.*]

1. Most interviewees stated that they came to the hackathon to learn about new technologies and that they had little or limited prior experience using. Also, every interviewee stated that they perceived the hackathon have been a great learning experience. Here are some quotes:
   Previous experiences: "*It was my first time seeing these technologies*" (C2), "*I was familiar with Python but not with NLTK*" (G1), "*I knew a little bit about HTML and JS*" (S1)
   Motivation to participate: "*my goal was to learn new languages […] and now I see how they work*" (C3)
   Perceived learning: "*I was surprised by how much I learned*" (C2), "*I am more confident [using those technologies] now than I was at first*" (C3), "*I think I have got the basics [of the technologies] down now but I plan to learn more*" (S2)
2. Most interviewees stated that they would like to continue working on their project. Some also mentioned a general interest in gateways:
   Q: Are you now more interested in Science Gateways. A: "*Definitely! [My mentor] opened my eyes to it*" (C2), "*It was pretty cool to know and interesting to figure out […] I am excited to continue learning more about [the gateway]*" (G1), "*I would be interested to work on another Gateway*" (S3)
3. Despite all of the interviewees stating that they learned a lot during the hackathon, they all rated their experiences in various programming languages higher before (M = 3.09, SD = 0.62) than after the hackathon (M = 2.47, SD = 0.73). All scales we used were 1 to 5. Initial explanation for this would be that the participants felt they already knew a lot before the hackathon and realized that they still have a lot to learn when putting their knowledge to the test during the hackathon. This however is only an assumption that needs to be underpinned by interviewee statements.
   All participants were satisfied with the outcome of the hackathon (M = 3.98, SD = 0.91).

COLLABORATION TOOLS

- **Documentation**
  - *Google Drive*
  - *Trello*
  - *GitHub*
  - *YouTube*
- **Communications**
  - *Slack*
  - *Zoom*
  - *SMS Messaging*
  - *Twitter*
- **Scheduling**
  - *Doodle*
  - *Google Calendar*

CONCLUSIONS AND OUTCOMES

- The event took place over the course of 24 hours at the *Wyndham Grand Pittsburgh Downtown* from July 25th-26th. There were eighteen (18) student participants that formed four (4) groups with associated mentors. The event concluded with a presentation session that was streamed live and allowed viewer participation. Various awards were given throughout the event based on games, work, social media interactions, and final deliverables.
- Critical to the success of the SGCI Hackathon was the participation of several Science Gateways including subject matter experts as both members of the planning committee, and as group mentors.
- The use of both Science Gateway directed tasks, in combination with group based deliverables added structure to the event. Initial deliverables included a "one slide, one minute" project description and presentation by the participants early in the event. Final deliverables included oral presentations with associated powerpoint and live demos. Final source code and presentations were distributed through a group generated GitHub repository. The combination of both presentations and live demos produced increased levels of both hard, and soft skills in the participants.

Future Hackathon Implementations

- Time should be alloted at the beginning of the hackathon, before allowing coding, to formulate ideas and assign team tasks.
  - This will allow possible critical tasks and an assurance of purpose to form.
- Group Mentors need to attend the event.
  - Remote mentoring inhibits both team building, and can limit communication. The result was a less productive work dynamic.
- Mentors and teams should meet before the event to understand expectations, develop strategies and assess technical capabilities of participants.
  - This allows mentors to provide any needed guidance, and or needed training prior to an event.
- Create defined checkpoints during the night.
  - Overnight work can suffer due to exhaustion of the participants. Including checkpoints assure critical tasks are accomplished, as well as provide opportunities for error correction and or mental resets.

SPONSORS

This project is funded by the National Science Foundation under award number ACI-1547611. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Website: https://sciencegateways.org/engage/hackathon