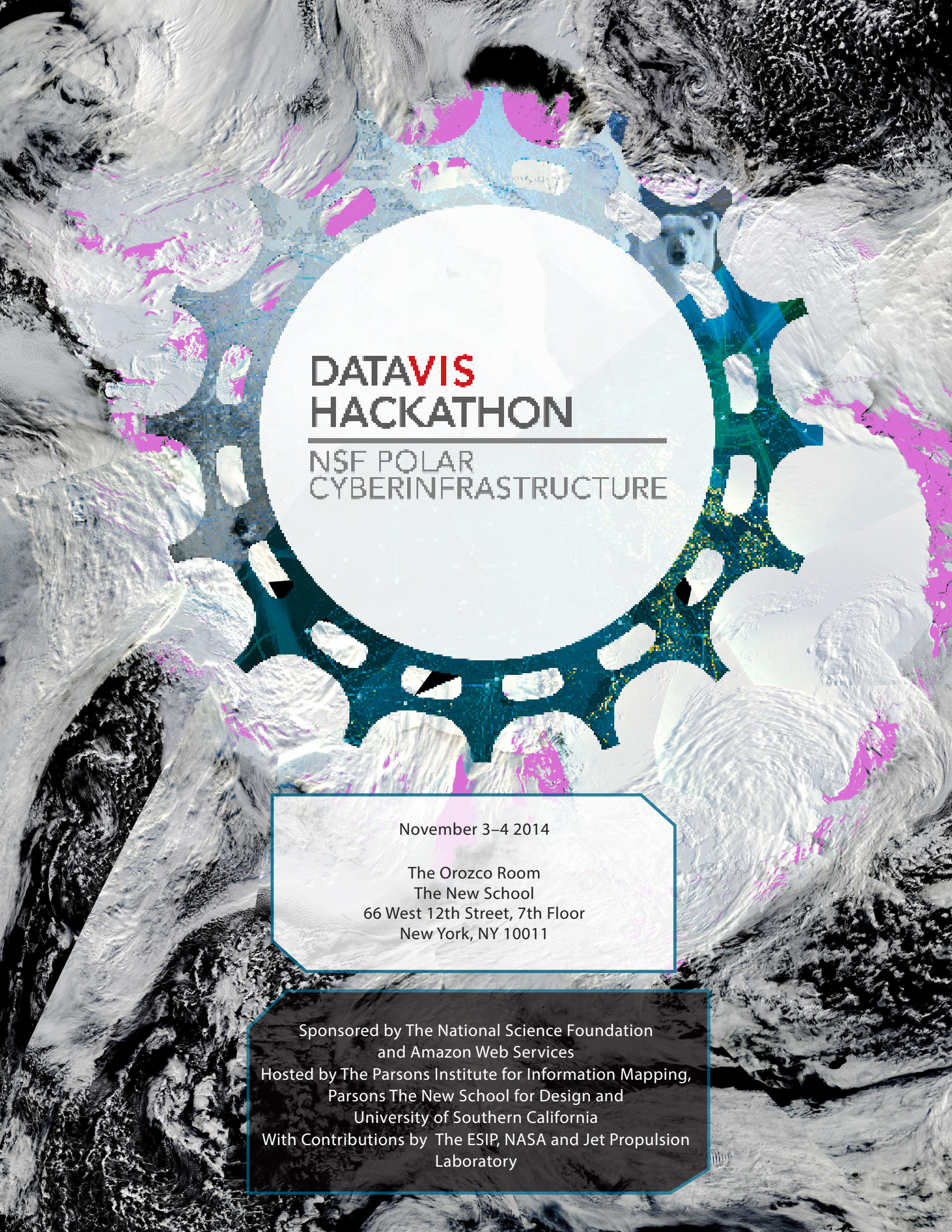


DATAVIS HACKATHON

NSF POLAR CYBERINFRASTRUCTURE

November 3–4 2014

The Orozco Room
The New School
66 West 12th Street, 7th Floor
New York, NY 10011

Sponsored by The National Science Foundation
and Amazon Web Services
Hosted by The Parsons Institute for Information Mapping,
Parsons The New School for Design and
University of Southern California
With Contributions by The ESIP, NASA and Jet Propulsion
Laboratory

This report may be cited as:
Mattmann, C., Ramirez, P., Mcgibbney, L. Pope, A., Wyngaard, J., 2015. Report on PolarViz Hackathon, NSF Polar Cyberinfrastructure. The New School, New york, 22pp. The hackathon and this report were co-funded by the National Science Foundation's Polar Cyberinfrastructure Program (Division of Polar Programs) and Division of Advanced Cyberinfrastructure under Awards PLR-1445624 and PLR-1348450. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation

REPORT ON

PolarViz Hackathon

The New School, New York
3-4 November 2014

OUTLINE

SPONSORS

HOSTS

CONTRIBUTORS

EXECUTIVE SUMMARY

HACKATHON BACKGOUND

Science data sets concerning the Polar regions are currently underutilized relative to their potential value - these data relate to U.S. and international maritime, oil and natural resources, and climate oriented concerns. It is exceedingly difficult to ask concise and scientific questions of this information since doing so involves a combination of high powered data processing; information retrieval, big data, and data visualization technologies to be orchestrated together and combined for many diverse end users. Improving the use and value of existing polar region data sets is crucial to understanding the polar regions variability over different timescales, with consequent associated benefits for society. One of the critical first steps towards realising such improvements involves fostering effective collaborations, communications and meetings between those community experts in data visualization; big data processing, information retrieval, and polar scientists.

The U.S. National Science Foundation "DataViz Hackathon for Polar CyberInfrastructure" was designed to both bridge the cyberinfrastructure data visualization and polar science communities through informal collaboration, and to notably improve the use and value of polar region data sets through the adaptation of visualization tools for polar data.

HACKATHON IMPLEMENTATION

Forty polar scientists, data engineers, and design scholars, from a diverse set of institutions (see figure 1) gathered in New York in November for 2 days. Recognising the need for design expertise in visualisation problems it was hosted at The Parsons Institute for Information Mapping, personnel from here provided invaluable input from a third community not traditionally engaged by either polar science or cyberinfrastructure. Following introductory remarks and goal setting, by both the polar and cyberinfrastructure communities, lightning talks by session leaders served to guide participants into self forming hack subgroups. Following a pattern of discuss-hack-review-repeat, over the course of 4 half-day hack sessions in total, these groups practically tackled the issues and goals set forth. These sessions resulted in 42 commits to a communal github account created for the event.

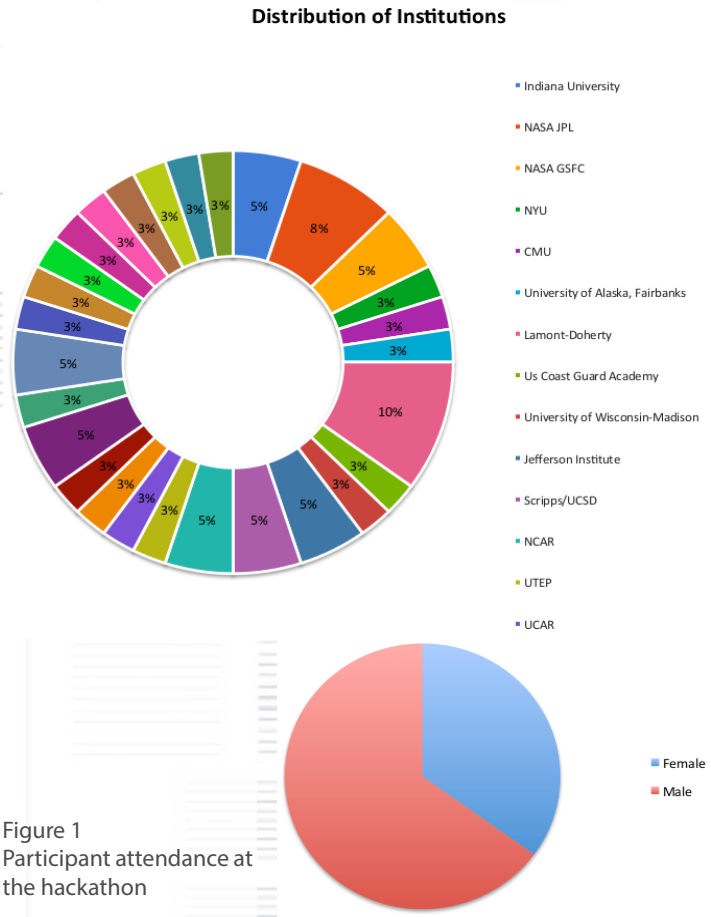


Figure 1
Participant attendance at the hackathon

SPONSORS HOSTS CONTRIBUTORS

HACKATHON CONCLUSIONS AND RECCOMMENDATIONS

Overwhelmingly participants indicated they would be likely attend another similar event, valuing most highly the trans-disciplinary interactions and networking opportunities, and most commonly lamenting the limited time in which to accomplish ambitious goals. Based on this feedback, and the practical outcomes achieved (code commits, skill and knowledge transfers, and new collaborations) it is recommended other similar event be held, particularly for this conjunction of domains.

Given the combined relative novelty of hosting a transdisciplinary, unconference, the organising committee learnt many practical lessons which should be of use to future event organisers.

Based on the transdisciplinary nature of the event, it is recommend that a greater effort be made, prior to the event or in introductions, to educate all participants as regards communicating with members of other domains and explaining some of the points of likely misunderstanding. For instance the science and engineering community here would have benefited from an overview of the nature of the design process, while the design and visualisation community would have found an explanation of the need for precision in science data helpful. In both cases such explanations would have reduced the time spent on unnecessary discussions.

In regards to the "unconference" nature of a hackathon, the use of a public hosting forums such as Github was highly successful in helping facilitate organisation prior to, during, and post the event, and it is therefore recommended. Participants also requested that future events be, longer (3 days at least), have more narrowly focused session goals, and that at least 1 joint hack session be held with all participants.

Finally, it is recommended that for a similarly data related event 2 primary recommendations are made. Firstly, that more time be spent prior to the event preparing data, preparing likely code applications, and soliciting and communally evaluating proposed sessions such that goals are well defined. Secondly that there be some dedicated input as regards data visualisation theory.

RAW STATS

- **"4-way conversations between domain scientists, data scientists, programmers and designers—so rare, so good."** - participant
- 40 participants
- 14 sessions
- 42 commits
- 2 days
- 1000 lines of committed code(mostly python)

SPONSORS HOSTS CONTRIBUTORS

OVERVIEW

The challenges facing society continue to grow in complexity. However, due to advances in technology our capacity to capture data across an enormous spectrum of parameters is also rapidly increasing. This data can be of great use in deriving solutions to the challenges faced if penetrating, revelatory, and predictive information can be extracted from it. In order to extract this information there is a growing need for innovative approaches to the fields of data analytics, synthesis and knowledge transfer.

Discipline specialisation allows for structured practices, standards, and authority to develop. But also constructs relatively fixed limits on method development and exploration. In consequence there is a growing move towards transdisciplinary collaborations, although there is as of yet little guidance on best practices for such. This hackathon was therefore foremost a result of a need for transdisciplinary collaboration, but also a theoretical exploration of the practicalities of doing so.

The Polar CyberInfrastructure program at the National Science Foundation is aimed at promoting exactly such collaborations. The program funds projects focusing on the polar aspect of (1) cost-effective transfer of data from remote field locations, (2) long-term sustainable curatorship, standardization, management and discovery of data and metadata, (3) visualization, manipulation, and analysis, particularly for understanding complexity, (4) access and interoperability across scientific disciplines, (5) effective use of HPC for direct and sustainable advances in current arctic research and (6) e-learning and educational tools based on cyberinfrastructure components. This event focused foremost on (3), the challenges of data visualization, manipulation, and analysis for understanding complexity.

OBJECTIVES

The hackathon was organised as an "unconference" focusing on participant involvement rather than presentations. It was advertised to the cyberinfrastructure and polar science communities via a number of relevant mailing lists (such as, the CRYOLIST, NASA Earth Science Data System Working Group's list, The Federation of Earth Science Information Partners lists, and certain Apache Software Foundation lists).

High level goals were set as:

- Assess the needs and challenges of, automated data curation, processing, dissemination, and visualisation, of polar datasets.
- Identify and evaluate relevant open source cyberinfrastructure technologies, techniques and tools, for meeting these needs and challenges.

SPONSORS

HOSTS

CONTRIBUTORS

NARRATIVE

As an "unconference" the format adopted was highly flexible. Dr. Chris Mattmann opened proceedings with a brief introduction of the hackathon agenda followed by Dr. Marco Tedesco who provided context as to the NSF's agenda, motivation and support for the Polar Science community and the Hackathon. After a keynote address by Dr. Curtis Lisle, the various proposed sessions were introduced. The workshop organizers specified that all sessions were to be proposed on Github by volunteer leads from the community prior to the event. This not only enabled provenance tracking but also provides continued opportunity for community comment and discovery. Over a coffee break attendees were free to review the sessions and question leads before assigning themselves to sessions of particular interest.

Typically hack session leads introduced their perceived goals to those who joined them, before discussions, planning, and implementation moved forward. With breaks for lunch, the day was wrapped up by brief summaries on progress, designs and discovered challenges from each lead. Sessions leaders were also open to determine whether the session should close or continue through until the next day.

Following a trend, day 2 also began with a keynote from Dr. Jer Thorp (adjunct Professor in New York University's ITP program) which treated the hackathon attendees to some of his award-winning software-based work which has been exhibited in Europe, Asia, North America, South America, including in the Museum of Modern Art in Manhattan. Following this participants were free to continue in the same sessions as previously, or as in some cases these had already concluded, join another session. While time was short, at the end of this concluding day session leads were again given time to report progress and conclusions.

A post event survey and the Github hosted session issues later facilitated all evaluations of the event. The following summarises some of the sessions held and their outcomes so as to provide a sense of what was practically accomplished over the two days.

All code outcomes can be found in the Github account under their issue numbers: github.com/NSF-Polar-Cyberinfrastructure/

ANTARCTIC METEOROLOGY RESEARCH CENTER AMRC DATASETS

Session Lead

Carol Costanza [SSUE 3]

Description

The AMRC at University of Wisconsin-Madison studies the weather in the Antarctic in two ways; Automatic Weather Station (AWS) data, and satellite composite imagery. For this workshop, a dataset has been prepared that contains five formats of AWS data and two formats of infrared satellite composites for one month - May, 2014.

Outcomes

- Creation of an appropriate data schema for the AWS data
- Creation of a solr searchable database using the above Schema

CRAWL AND PREPARE NSF ACADIS, NASA AMD AND NSIDC ARCTIC DATA EXPLORER DATASETS

Session Lead

Chris Mattmann [Issue-1]

Description

Building off of the NCEAS/open-science-codefest#26, this session continued data preparation and crawling of data archives identified as being relevant to the Polar Sciences. In particular cutting edge search technologies such as Apache Nutch and Apache Tika were used to crawl and extract resource metadata from NASA AMD, NSF ACADIS and NSIDC ADE databases with the goal of preparing data for visualisation.

Outcomes

- Around 10 participants were introduced to open source search and retrieval technologies, by the end of the session they were able to crawl and extract enhanced metadata from science products.
- It was discovered that the quality of scientific grade product metadata varies significantly between data archives
- It was identified that current tools (such as Apache Tika) need to be further developed to better handle scientific data due to its heterogeneity.

EXPLORING DATA WITH TIME AS A 4TH DIMENSION

Session Lead

Allen Pope [Issue 78]

Description

This session explored solutions to a very real "problem" faced by many polar scientists - how to explore and interact with a netcdf "cube" or data where each variable is defined by 4 dimensions (x, y, z, and time). The group (netcdf users) wanted a platform which allowed them to select which variables/dimensions to show, sliders to go through the data, and editable color bars. Export of figures and video was also a plus, but the most important task was seen to be visualization and data exploration.

Outcomes

- Two highly effective software packages were identified and demo-ed by the group using some ocean model output data. These are Panoply (http://www.giss.nasa.gov/tools/panoply/) and Vapor (https://www.vapor.ucar.edu/). Panoply is a very fast, intuitive tool for visualizing netcdf data. Vapor is a more complex but ultimately more powerful tool for data rendering and visualization.
- Exposure to both tools was of significant interest to the group with some group members subsequently making use of such for their research.

POLAR DATA ANALYTICS AS A SERVICE (PDAAAS)

Session Lead

Lewis John McGibbney [Issue-15]

Description

This session tackled the identification and then provisioning of useful software components for the emerging Polar Scientist in a packaged manner. This was an attempt to ease use and eradicate complex installation procedures driving science and reducing technical barriers.

Outcomes

- A Vagrant image was produced cf. https://github.com/NSF-Polar-Cyberinfrastructure/issue-15 which packages many of the most popular scientific software libraries into an isolated, clean environment 'out-of-the-box' completely removing installation and configuration.
- Participants noted that real time data streams and sources of documentation / intelligence from Polar regions is also something, which if packaged into a similar environment, could come in extremely useful. An example would be a local server which hosts incoming streams of real time shipping information for Polar regions.

ICE CORE (AND ICE CORE ARCHIVE) VISUALIZATION

Session Lead

S. McKenzie Skiles [Issue 79]

Description

The national climate data center hosts an ice core archive (http://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/ice-core), with an interactive map AND google earth map that allows you see information on the core and then link back to the data either via text file or download from FTP. While already an incredibly useful service, an archive that is more user friendly, and perhaps also shows a visual summary of data, would make for easier data exploration and selection.

There has been some work done on ice core visualization (http://earthobservatory.nasa.gov/Features/Paleoclimatology_IceCores/). While these are useful, there could be more exciting ways to visualize individual core data.

Outcomes

- Borrowing from the WebGL Globe, this session was able to release a prototype ice core visualisation now available on github (here).
- This work could be taken further relatively easily by using Cesium to achieve greater functionality

OPEN-SOURCE POLAR DATA WORKFLOWS WITH TANGELOHUB

Session Lead
Curtis Lisle [Issue 42]

Description
TangeloHub is an open-source, web-hosted data science workflow processing and analysis application funded by NSF's Biology division through the ArborWorkflows project. Source code repositories can be found at https://github.com/Kitware/tangelohub and http://www.arborworkflows.com. TangeloHub allows users to create and run multi-step workflows interactively through a web interface. Quick visualizations and dataset management are built into the software system, as well. Individual steps in a workflow can be implemented in either the Python or R languages, and a web-based editing environment is included so collaborators can upload, share, and interact over both code and datasets.

During this session, TangeloHub was introduced along with a use case developed prior to the hackathon on one of the Polar Datasets (the AWS weather station data). A scatterplot below was created interactively using TangeloHub operating on the AWS data. The plot shows the humidity measurements for all stations as a function of the station's vertical elevation. A detailed demonstration of how to create custom data analyses and invoke them through the web interface was conducted. Participants reviewing TangeloHub's capabilities and interacted with a sample biogeographic application where different species observation points were georeferenced and displayed using both a hierarchical, phylogenetic tree and their locations on an interactive Google map background.

Outcomes
- Participants were able to get hands on experience with a TangeloHub instance to process datasets themselves.
- Curtis Lisle of KnowledgVis and Justin Paul-Peters of Indiana U collaborated to create a Tangelo-based application that renders MODIS imagery and reads AWS weather station information from a live database (see Figure 2) Tangelo (http://www.tangelohub.org/tangelo/) is a web framework that combines simple Javascript-based webpages with simple, python-based computation capability.

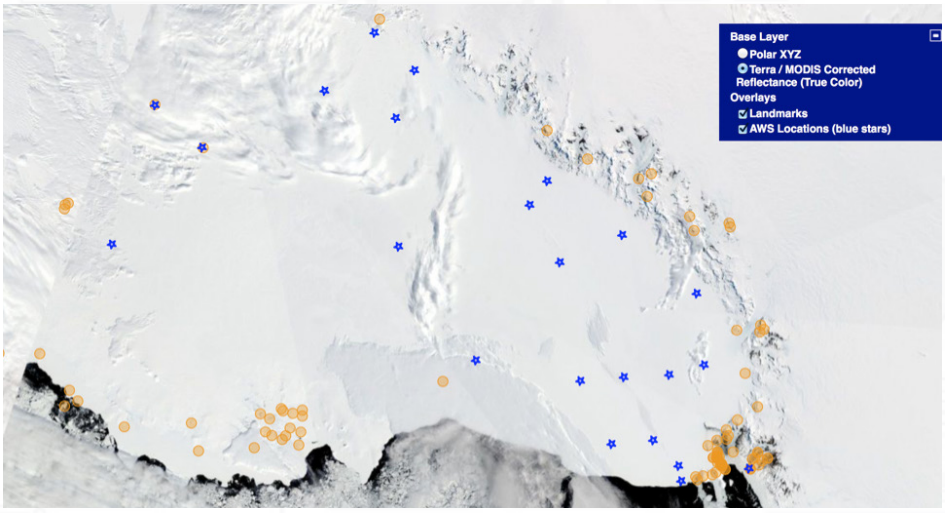


Figure 2
Screenshot of the prototype showing the cloud layer from MODIS, a set of static points of interest (tan circles), and the AWS weather station locations (blue stars). Below is a screenshot showing the editing of a workflow pipeline in TangeloHub:

NASA NEAR REAL-TIME POLAR IMAGERY SERVICES

Session Lead
Jeffrey Schmaltz [Issue 43]

Description

NASA's Global Imagery Browse Services (GIBS, http://earthdata.nasa.gov/gibs) was developed to provide highly responsive, scalable, and expandable imagery services using Open Geospatial Consortium (OGC) Web Map Tile Service (WMTS) standards. Currently, there are more than a dozen MODIS imagery products available in polar stereographic projections for each pole, including four daily one kilometer 11 micron thermal infrared band images during all seasons. Imagery back to mid-2013 is currently available and reprocessing of imagery from the entire MODIS record is underway and community input is being solicited on recommendations for additional imagery layers from MODIS and other NASA instruments. I would like to give a brief demo of the imagery layers and access methods.

Outcomes

- A demonstration was given of the imagery layers and access methods of GIBs
- ysis. These are notoriously difficult to configure and install.
- AVagrantimagewasproducedcf.https://github.com/NSF-Polar-Cyberinfrastructure/issue-15whichpackagesmany of the most popular scientific software libraries into an isolated, clean environment'out-of-the-box'completely removing installation and configuration.
- Participants noted that real time data streams and sources of documentation /intelligence from Polar regions is also something, which if packaged into a similar environment, could come in extremely useful. An example would be a local server which hosts incoming streams of real time shipping information for Polar regions.

GISCUBE, OPEN SOURCE WEB-BASED GEOPROCESSING AND VISUALIZATION APPLICATION

Session Lead
Maziyar Boustani [Issue 11]

Description

GISCube (https://github.com/MBoustani/GISCube) is an open source web-based GIS application that provides variety of GIS capabilities such as data management, metadata extraction, visualization as well geospatial tools. In this session, GISCube was demoed for processing as well as visualizations some polar data.

Outcomes

- This session introduced GISCube to participants, particularly as an easy access open-source alternative to ArchGIS.
- GISCube lacks supporting geospatial data in different projections for visualization
- There are interests on netCDF tools such as clipping netCDF with shapefile.

"HIGHER DIMENSIONAL" POLAR DATA VISUALIZATION

Session Lead
Alex Boghosian [Issue 50]

Description

Data visualization need not stop at a 2d image, especially since the data themselves are geospatial. It is possible to print various polar datasets in 3D, but there many challenges in doing so which may prevent 3D printing's potential being embraced. These challenges include having to grapple with obtuse architectural modeling software, adjusting the resolution and scale of the data to suit the printer, physical limitations of the printer itself, and changing the format of the data to suit that of the printer, amongst others. This session discussed how to make 3D printing easy for scientists to use, if 3D printing could be used for polar research purposes beyond data visualization and if so what tools would be need for this to happen, and is it possible to make interesting and accurate models for people to download and print on their own.

Along similar lines, it is possible to create an interesting "installation" of a these datasets using light or sound. This could be a very effective and fun alternative for exploring data that are not inherently visual. There is a lot of potential for creating visualizations that go beyond the computer screen.

Figure 3
3D printed model of Greenland made prior to the hackathon using the discussed approaches. The snow/ice layer is removable.

Outcomes

- New ideas for data to print were developed, this included making a model of a brine channel.
- Exploring ways of making the brine channel model revealed limitations. For instance the brine channel .stl file existed but was impossible to print as it was very small, and consisted of a cube of floating pieces. To print one, much of the dataset was first deleted, the remaining data scaled up, and supports were added so it would not break while printing.
- A new workflow for smoothing and printing topography data was developed. This resulted in a model of Jakobshavn bed topography.

BUILD A VISUALIZATION, TELL A STORY

Session Lead
Allen Pope [Issue 82]

Description

The goal of this session was to identify important polar science topics that had not been well-explained to the public and attempt to address these shortcomings through an innovative data-driven visualization. In the first session, the group broke into two groups: the first tackled public perception of Arctic vs. Antarctic sea ice, while the second discussed rapid and projected changes in West Antarctica.

Outcomes

- The sea ice group found that what started as a seemingly simple question was actually too broad an issue to address within the wider Arctic systems science lens. Potential datasets discussed included ice cover satellite measurements, Arctic shipping behavior, geopolitical tensions, and resource availability. While an interesting discussion, no convergence was reached and the group disbanded after one session. However, the group did identify that there is a gap in the availability of data services and documentation what would lead researchers to the information they need. The group came to the conclusion that **if the data were easily available and described correctly, it would be possible for social scientists to layer their stories atop the data efficiently.**
- The second, "West Antarctica" group spent a significant amount of time discussing the interesting science behind West Antarctic Ice Sheet change and converged on the idea of using map-based data with a system of tutorials and smart pop-ups to show users different feedback mechanisms in glacier behavior.
- Other potential design ideas included using story maps, or instead the cinematic scrolling stories becoming increasingly popular across many online feature stories and blogs. These were seen as more straightforward but less nuanced.
- Design expertise was helpful when present, but intermittent, which may have hampered the group's progress.
- Ultimately the group produced a wireframe for a potential tool and identified some key and complementary datasets which could be used to explain dynamic and dramatic ice sheet change. The importance of including simple model results was discussed in order to pedagogically highlight the relationships between cause and effect (and not just correlation). This wireframe was followed up with post-hackathon discussions by a few of the participants. While there was significant interest, it was identified that at the present time such an endeavour was still prohibitively expensive as envisioned.

CONCLUSIONS AND RECOMMENDATIONS

To fully leverage the energy and collaborations generated throughout the event, an apparent realization has unearthed that a number of social and practical barriers need be considered. Commentary in the following section provides some contextualised examples of challenges encountered before making a brief summary of recommendations. It is hoped that the commentary will better equip future organisers of events both within the NSF and beyond.

PARTICIPANT EXPERIENCE

A series of post-event questionnaires (completed by thirty-two participants at the event, and numerous others after it), surveyed participants experience of the hackathon. From the responses obtained the following data points were most prominent and indicative.

- 94% of participants achieved their ideal outcome fairly or very well, with the most important outcomes described as networking, learning about new tools, and developing ideas for further work.
- Overall the vast majority of respondents indicated they were likely or very likely to attend a similar event.
- The networking opportunity presented was highly appreciated.
- The transdisciplinary interactions were rated as challenging yet highly beneficial.
- While satisfied with the duration participants indicated that they would have appreciated more time (often so as to avoid missing co-incident sessions, and in order to accomplish set out session goals).
- Where dissatisfaction was expressed it was largely attributed to the fact that little was practically accomplished, often due to long discussions, although these did culminate in ideas. A minority also indicated that the goals set for some sessions were unrealistic given the time allocated

HUMAN FACTORS

As an "unconference" there were a number of well known group dynamics challenges that might possibly have been minimised if anticipated and group leads, particularly, prepared. Examples of such include:

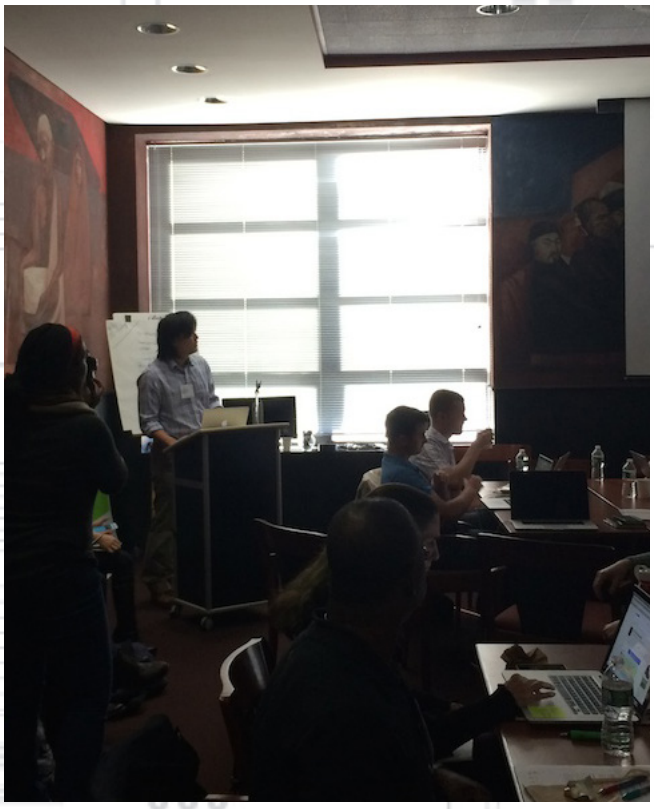
- Some participants excluded themselves from discussions due to a perceived lack of personal capacity to contribute to the topic at hand.
- Conversations were at times dominated by strong personalities.
- The natural tendency for participants to engage primarily with others of similar background was not beneficial.
- Discussions could diverge from the session's designated topic and objectives.

SPONSORS HOSTS CONTRIBUTORS

COMMUNICATION

Unsurprisingly the differences in background resulted in communication difficulties. This effect manifest in a number of ways.

- The science and engineering community struggled to translate their science understanding of data into non-scientific terms for fear of losing information. This is where the design feedback from Parsons faculty was particularly valued as their questions and comments provided different perspectives, leading to better defined objectives, narratives, and audiences for the proposed visualization.
- The science and engineering community was notably unfamiliar with the unstructured, exploratory, and iterative design process. Where the design community advised beginning with understanding, observing and defining the problem to be addressed (most importantly by defining the audience), many participants began with how the final visualization should look and prefered to instead have the design set before implementing anything. At times this resulted in an inability of groups to come to consensus and resulted in dispersion of participants.



SPONSORS

HOSTS

CONTRIBUTORS

STRUCTURE

In relation to the event's structure tailoring was required to accommodate both the transdisciplinary, and "un-conference" community involved, aspects of the event.

- It is inevitable that session leads will propose projects or challenges related to their own work. This can be advantageous as these people fully understand the issue but can also potentially lead to the whole event being overwhelmed with "pet projects". Even when self admitted this is not necessiarily conducive to the goals of fully open cross-domain fertilisation of ideas and expertise. The event organiser is encouraged to facilitate a balance in this regard.
- A number of the sessions were overwhelmed by a need for the software specialists to do significant coding and compiling work before the design and science members were able to contribute. By facilitating greater discussion prior to the event, greater technical preparation, such as code and data readiness, should be possible so as to enable more rapid progress at the event.
- In some sessions there was an imbalance in the number of representatives of the various expertise domains present (Issue 77 was an example of this where no visualisation expert was present and the session was thus later abandoned). This is inevitable at an event where participants are free to choose their tracks. While from a productivity perspective this is undesirable, the result is not wholly negative. Firstly, such instances nevertheless fostered good discussion and networking (results participants rated highly and something that the lead of Issue 77 indicated as an outcome). And secondly because of this freedom of movement, once a session reached an end point for lack of expertise, participants simply moved on to another session where they could continue contributing and learning. The trade-off in terms of retaining the flexible freedom of movement for participants is therefore still considered a advantageous approach.
- In contrast to these challenges, the use of Github, not just as a public code repository but as the as an event organising platform, was highly successful. By managing the event as a project with issues to be tracked in a public forum, progress and discussion could be easily followed and contributed to by everyone (such as the session proposals). It is recommended that such a platform be used again, however, even greater value could be found if session or daily commits had been encouraged, or if someone had been assigned to document the event in Github.
- The occurrence of co-incident sessions on the same or very similar topics was a common criticism. While a challenge common to any parallelised event there are a number of structural changes that could be made to alleviate this. The event duration might be increased, the event's focus might be narrowed, pre-event discussion could be used to single out only one or two challenges to be addressed in each hack period (a few hours, a day, or even the whole event).

SPONSORS HOSTS CONTRIBUTORS

PURSUING AN EFFECTIVE TRANSDISCIPLINARY HACKATHON

In light of the above, the survey, and organisers experience, the following recommendations are made for future organisers of similar events.

- Acknowledge the human challenges of an "unconference" and take time to discuss potential group dynamic problems at the beginning, both with session leads and the complete group.
- Given the apparent lack of knowledge of the design process, on the part of the science and engineering communities, educate those participants briefly in the welcoming session as to the iterative nature of such and the need to first explore the what, why, where, and who questions of the problem being tackled.
- So as to increase productivity and reduce fruitless discussion time costs, educate all participants on communicating across domains and utilising visualisation. This could be done in a brief introductory session, or perhaps via printed guidelines.
- Prior to the event greater development and participant interaction would serve to better define session objectives and goals.
- Prior to the event a larger quantity of data needs to be made readily accessible. This may include not just acquiring remote access but also crawling, and even some preformatting processing. Prepared data could also be used to pose a challenge for a session.
- Session leads should be directed to develop a task follow up plan if appropriate.
- Everyone should be encouraged to contribute progress, throughout the process, to the public forum (Github in this case). In some cases this might require an introduction to that forum as different domains may have no experience with it.
- If possible it was clear that a longer event would have been appreciated, extending to 3 full days is recommended.
- In order to address the request that more expertise be present to take more ideas through from concept to realization, we do not simply recommend more participants be invited, alought this could easily also be advantageous. But rather suggest that either the organisers define the sessions prior to invitations being made, or a pre-invite survey be made, regarding data challenges to be addressed, within the intended participant source communities. With either approach the intention would be to facilitate the decisions around who should be invited more precisely so as to have an optimal ratio of appropriate skills.
- Requests were made for more designer/user experience focused sessions. It is unclear what the motivation behind this was but it is recommended that the goals of the event might be more clearly defined after initial interest pre-event discussions.
- A session that would involve all participants should be held.

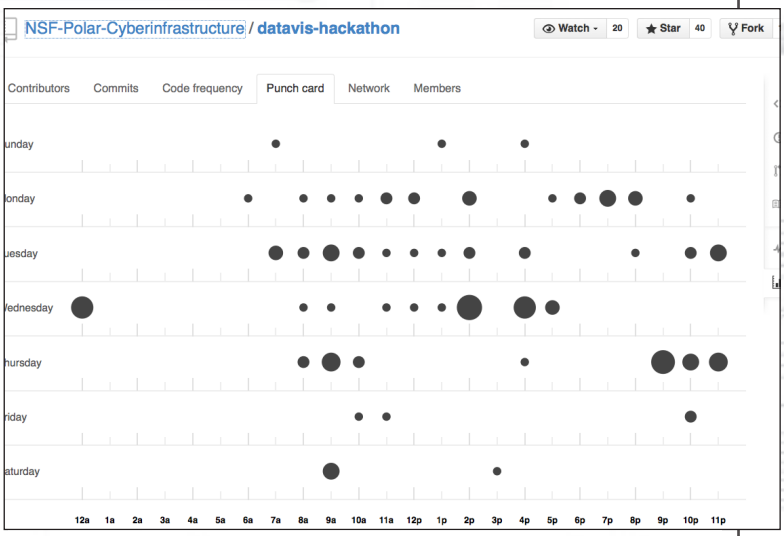


Figure 4
Github plot of commits by day and time for the datavis-hackathon repository.

REFERENCES

1. NSF DataViz Hackathon for Polar CyberInfrastructure, http://nsf-polar-cyberinfrastructure.github.io/data-vis-hackathon/, Accessed: November 2014.

SPONSORS

HOSTS

CONTRIBUTORS

APPENDECES

ORGANIZING COMMITTEE

Dr. Chris Mattmann
University of Southern California and Jet Propulsion Laboratory, California Institute of Technology

Dr. Annie Bryant Burgess
University of Southern California

Dr. Suzanne Carbotte
Columbia University

Dr. Bruce Caron
New Media Research Institute

Dr. Patrick Driscoll
Aalborg University, Denmark

Mr. Christopher Goranson
Presidential Innovation Fellow (GSA)

Mr. Aaron Hill: Parsons
The New School for Design

Dr. Daniel Katz
National Science Foundation

Dr. Martin Lehmann
Aalborg University, Denmark

Dr. Alan Maceachren
Penn State University

Dr. Jonathan Pundsack
University of Minnesota Polar Geospatial Center

Dr. Marco Tedesco
National Science Foundation

Mr. Joel Towers: Parsons
The New School for Design

Dr. Saskia Van Manen
Open University

Dr. Allen Pope
NSIDC

Mr. Jihoon Kang
Parsons the New School for Design

ACKOWLEDGEMENTS

This hackathon would not have been possible without the help of many individuals and organizations. In particular we wish to thank Dr. Marco Tedesco for providing the initial funding. We wish to thank the Organizing Committee (http://nsf-polar-cyberinfrastructure.github.io/datavis-hackathon/#committee) for its tireless work and telecons in preparation for the event. In addition, special thanks are due to Amazon Web Services and Ms. Traci Truthkoski for providing AWS servers for the event. We thank Mr. Paul Ramirez, Mr. Maziyar Boustani and Dr. Lewis John McGibbney for help with logistics at the event. In addition, we thank Katie Wanner and Jihoon Kang for their help and support and logistics from New School. Finally thanks to Dr. Jane Wyngaard for excellent support in putting together the workshop report.

SPONSORS HOSTS CONTRIBUTORS

PARTICIPANT LIST

INDRANI DAS Lamont-Doherty Earth Observing Laboratory, Columbia University

S. MCKENZIE SKILES University of California Los Angeles (UCLA)

DAVID REAGAN Indiana University

LEWIS JOHN MCGIBBNEY NASA Jet Propulsion Laboratory

JEFFREY SCHMALTZ NASA Goddard Space Flight Center

RYAN BOLLER NASA Goddard Space Flight Center

TYLER PALSULICH New York University (NYU)

CHALALAI CHAIHIRUNKARN Carnegie Mellon University

WENWEN LI Arizona State University

JEREMIAH DABNEY University of Alaska, Fairbanks

MATTHEW SAVOIE National Snow and Ice Data Center (NSIDC)/University of Colorado, Boulder

JUSTIN PAUL-PETERS Indiana University

JOHN MORTON Lamont-Doherty Earth Observing Laboratory, Columbia University

SUZANNE CARBOTTE Lamont-Doherty Earth Observing Laboratory, Columbia University

CHRISTOPHER SWEENEY US Coast Guard Academy

CAROL COSTANZA University of Wisconsin-Madison

AARON PRESNALL Jefferson Institute

STEPHEN DIGGS Scripps/University of California San Diego

KANCHANA WELAGEDARA Computer Society of Sri Lanka/Apache Software Foundation

SPONSORS HOSTS CONTRIBUTORS

ZORAN HRNCIC Jefferson Institute

MIA BENNETT University of California Los Angeles (UCLA)

ALLEN POPE National Snow and Ice Data Center (NSIDC)/University of Colorado, Boulder

MAZIYAR BOUSTANI NASA Jet Propulsion Laboratory

SCOTT PEARSE National Center for Atmospheric Research (NCAR)

RACHEL OBBARD Dartmouth University

LAURA KEHRL University of Washington APL Polar Science Center

MARTIN LEHMANN University of Aalborg

SASKIA VAN MANEN Open University, UK/Parsons Institute for Information Mapping

JESSE JOHNSON University of Montana

CURTIS LISLE, KnowledgeVis, Inc.

YUAN HO, Unidata/UCAR

BRUCE CARON, New Media Studio/University of California Santa Barbara

GEETHA RATNAM, Scripps/University of California San Diego

ERIC NIENHOUSE, National Center for Atmospheric Research (NCAR)

CHRISTINE LANEY, University of Texas El Paso (UTEP)

ALEX BOGHOSIAN, Lamont-Doherty Earth Observing Laboratory, Columbia University

PAUL RAMIREZ, NASA Jet Propulsion Laboratory

JUSTIN FIELDS, Rails Dog

ZHONG LIU, George Mason University

SPONSORS

HOSTS

CONTRIBUTORS