Measuring Success: How Science Gateways Define Impact

Nancy Wilkins-Diehr
San Diego Supercomputer Center
University of California, San Diego
La Jolla, USA
https://orcid.org/0000-0002-3604-0325

Mark Miller
San Diego Supercomputer Center
University of California, San Diego
La Jolla, USA
https://orcid.org/0000-0001-9590-3728

Emre H Brookes.
Department of Biochemistry
University of Texas Health Science Center
San Antonio, USA
brookes@uthscsa.edu

Ritu Arora
Texas Advanced Computing Center
University of Texas at Austin
Austin, USA
rauta@tacc.utexas.edu

Amit Chourasia
San Diego Supercomputer Center
University of California, San Diego
La Jolla, USA
amit@sdsc.edu

Prasad Calyam
Department of Electrical Engineering & Computer Science
University of Missouri-Columbia
Columbia, USA
calyamp@missouri.edu

Douglas M. Jennewein
Information Technology Services
University of South Dakota
Vermillion, USA
https://orcid.org/0000-0002-8361-8361

Viswanath Nandigam
San Diego Supercomputer Center
University of California, San Diego
La Jolla, USA
vnandigam@ucsd.edu

M. Drew LaMar
Biology Department
College of William and Mary
Williamsburg, USA
mdlama@wm.edu

Sean B. Cleveland
Information Technology Services - Cyberinfrastructure
University of Hawaii
Honolulu, USA
https://orcid.org/0000-0002-7130-3434

Greg Newman
Natural Resource Ecology Laboratory
Colorado State University
Fort Collins, USA
gregory.newman@colostate.edu

Shaowen Wang
Department of Geography and Geographic Information Science
University of Illinois
Urbana, USA
shaowen@illinois.edu

Ilya Zaslavsky
San Diego Supercomputer Center
University of California, San Diego
La Jolla, USA
zaslavsk@sdsc.edu

Michael A. Cianfrocco
Department of Biological Chemistry & Life Sciences Institute
University of Michigan
Ann Arbor, USA
mcianfro@umich.edu

Kevin Ellett
Indiana Geological and Water Survey
Indiana University
Bloomington, USA
kmellett@indiana.edu

David Tarboton
Civil and Environmental Engineering
Utah State University
Logan, USA
dtarb@usu.edu

***Abstract*— Science gateways, also known as advanced web portals, virtual research environments, and more, have changed the face of research and scholarship over the last two decades. Scholars world-wide leverage science gateways for a wide variety of individual research endeavors spanning diverse scientific fields. Evaluating the value of a given gateway to its constituent community is critical in obtaining the financial and human resources to sustain gateway operations. Accordingly, those who run gateways must routinely measure and communicate impact. Just as gateways are varied, their success metrics vary as well. In this survey paper, a variety of different gateways briefly share their approaches.**

***Keywords*— science gateways, success metrics, measuring impact**

I. Introduction

The term Science Gateways was coined in 2004, to describe the portion of the NSF-funded TeraGrid program devoted to increasing accessibility of supercomputers to all scientists. In the ensuing years, the term has acquired a much broader meaning; it now serves as a descriptor for all advanced web interfaces used for research, education and scholarship. Launched in 2016, the Science Gateways Community Institute [1] helps researchers build and discover better, more functional gateways by documenting, developing, and disseminating best practices.

Maintaining a persistent online presence is critical to gateway success. This in turn requires funding to support the human and cyberinfrastructure that underlies each gateway. Accordingly, the business planning training offered by SGCI has been popular among gateway creators. Measuring and communicating impact is perhaps the single most important factor in determining whether a given gateway has sufficient funding (usually through research grant awards) to ensure a persistent online presence. Although critical, the definition of impact can vary widely among gateways due to the diversity of their goals and the needs of the communities they serve. In this paper, we survey impact measures used by a diverse set of gateways to document current best practices in impact measurement. Our goal is to provide a basis for evolution of the notion of impact for current and future gateway operators and funders.

II. GATEWAY EXPERIENCES

A. Ultrascan

The NIH and NSF supported UltraScan Gateway [2] supports the Analytical Ultracentrifugation (AUC) community. AUC experiments are performed in a wide range of disciplines, encompassing biomedical life and basic science applications in biophysics, biochemistry and molecular biology, as well as polymer chemistry, colloid, material and nanoparticle science. The gateway, in operation since 2006, allows users to manage and perform analyses of AUC experimental data. MPI-Parallel jobs are submitted through the gateway to various compute resources, including those allocated via XSEDE. Impact metrics typically reported include number of active institutions, users, and number of publications. For example, in 2018 we have data on 116 institutions and 140 investigators actively utilizing the gateway. These institutions are available on the gateway's website. The usage data is extracted from logs in our SQL database (MySQL). From 2010-2018, we have 158 publications registered. Publications are self-registered. We regularly ask users to submit publications through our website, but we have no guarantee that this is exhaustive.

B. Interactive Parallelization Toolkit (IPT)

The main goals behind developing the NSF-funded IPT gateway were to lower the adoption barriers to IPT and to provide a self-paced learning environment for developing parallel programming skills [3]. While still under development, a prior version of IPT was operational for about six-months in 2018 and accessed by approximately 40 users, the majority of whom were taking a parallel programming course at the University of Texas at Austin.

The number of users (domain-experts and students) accessing IPT through the gateway is the key metric of success. However it is also important to measure how well users meet their learning goals. While the number of users is captured automatically, surveys are used to get the users' feedback on their learning experience. Surveys are conducted immediately after every training event, however, as the users are not obligated to take the surveys, the amount of data collected so far has not been statistically significant. Nonetheless, the objective and subjective responses obtained thus far through these surveys have indicated that the gateway is helping users in learning parallel programming concepts and parallelizing applications in a self-paced manner without spending time in software installations, and at a high-level of productivity. If 70% of the software/hardware resources available to the project are being utilized at a given time, the project team considers that as a success.

C. SeedMe: Stream, Encode, Explore and Disseminate My Experiments

SeedMe1 helps researchers to solve the "last mile" problem of sharing data and metadata as well as video generation from sets of images typically created as scientific visualization. The system provides researchers with a cloud-based service that may be used via a web browser, through web services, and by application clients and command line tools. Its also integrated with scientific applications Kepler, Vapor and VisIt. SeedMe1 has been in production since 2014 and currently has over 700 users who have uploaded over 150,000 items. The website receives approximately 200 active visitors a month.

The SeedMe2 [4] project is domain agnostic and aims to catalyze pre-publication research by improving and accelerating research productivity with rich and powerful data stewardship tools leading to fundamental advancements in science. SeedMe2 is a set of modular building blocks for creating data sharing and data management websites. It enables research teams to manage, share, search, visualize, and present their data in a web-based environment using an access-controlled, branded, and customizable website they own and control. An early research group has been heavily using the system since 2018 with 18 users that have shared over 17,000 items with an aggregate size of approximately 5GB. The project is currently transitioning from early access to production phase. Our success metric includes number of instances in production (for Research groups, Research consortiums, Science Gateways, Websites using our software, CI providers offering our software stack); number of users for each instance; number and size of data items in each instance; number of descriptions & discussions included in data items; usage of visualization and API features; and lastly subjective findings via user interviews and user surveys.

D. CyNeuro

CyNeuro science gateway [5] is being developed at the University of Missouri-Columbia since 2017 with the goal to advance cyberinfrastructure and software automation in neuroscience. The primary focus in the development efforts are to create openly-accessible software tools for data-intensive neuroscience research and education projects that benefit from

---

Wilkins-Diehr: NSF ACI-1547611
Miller: NSF ABI-19054444, NSF DBI-1759844, and NIH R01 GM126463
Brookes: NSF CHE-1265817, OAC-1740097, OAC-1912444 and NIH GM120600.
Arora NSF OAC-1642396
Chourasia: NSF OAC-1443083, ACI-1235505
Calyam: NSF OAC-1730655
Jennewein: NSF MRI 1626516
Nandigam: NSF EAR-1557484, EAR-1557319, EAR-1557330
LaMar: NSF DBI 1346584, DUE 1446269, DUE 1446258, DUE 1446284
Cleveland: NSF ACI-1557349
Newman: NSF ACI-1550463, OAC-1835574
Wang: NSF ACI-1047916
Zaslavsky NSF ICER-1639764, ICER-1639775, ICER-1639557
Cianfrocco: NSF ABI-1759826
Ellett: DOE DE-PI0000017 via West Virginia University Research Corporation
Tarboton: NSF ACI 1148453, ACI 1148090, OAC-1664061, OAC-1664018, OAC-1664119

advanced cyberinfrastructure resources (e.g., Neuroscience Gateway, JetStream) and technologies (e.g., JupyterHub, CIPRES). The approach to develop and refine capacity, capability and user support structures within CyNeuro have been through user surveys and mini-symposia of various stakeholders (e.g., researchers, resource providers, tool developers, data owners). CyNeuro hosts exemplar workflows and data sets that we are developing for exploring the potential of cyberinfrastructure and software automation in neuroscience. We are connecting to and leveraging local MU resources, as well as Neuroscience Gateway resources to scale research productivity, and develop large-scale training platforms. The ultimate purpose of CyNeuro is to support research and education use cases of neuroscientists (e.g., involving large-scale computation, image analysis) without expecting them to possess expert knowledge of programming and cyberinfrastructure configuration, which is beyond the repertoire of most neuroscience programs.

The key measures of success for CyNeuro include: (i) the number of new research/teaching tools and exemplars developed by CyNeuro developers/users, and their sustained usage by the community, (ii) development of a new undergraduate/graduate course sequence on "Cyberinfrastructure and Software Automation in Neuroscience" from biological/psychological sciences and computer science/engineering, and (iii) effective support of ongoing teacher and researcher training programs around neuroscience topics (e.g., NSF REU/RET, NIH R25). In particular, we are considering user-centered metrics such as ease-of-use, ease-of-setup, ease-of-self-service and increased options in terms of tools/resources to complete compute/data-intensive workflows of our users. Further, other expected outcomes to assess the level of success for developers/users enabled by the CyNeuro gateway include: sustainable software products (openly available via GitHub), and publishable learning outcomes through systematic evaluation studies in peer-reviewed venues.

E. *University of South Dakota Science Gateway*

The University of South Dakota (USD) Science Gateway is a campus gateway aimed at early on-boarding of newcomers to computational research, providing a streamlined graphical interface to the most commonly used applications for emerging computational researchers at USD and across South Dakota. In South Dakota where many aspects of cyberinfrastructure and computational research are still emerging there is not always a clear onramp to participation. The USD Science Gateway aims to address that by providing an accessible avenue to advanced computing on USD's Lawrence Supercomputer . With an initial focus on biology and bioinformatics applications, the Science Gateway will simplify user on-boarding to campus clusters, with a potential transition to XSEDE resources. To that end the gateway's goal is to increase productivity in groups unfamiliar with advanced digital resources and provide an accessible platform for learning scientific applications without the overhead of also learning the Unix command line and other intricacies of traditional High Performance Computing.

The gateway, first developed in 2018, is currently in early access with approximately 10 users in 2 research groups, incorporating federated authentication through CILogon and offering the QIIME and Mothur applications, with plans for incorporating applications from additional scientific domains next. In South Dakota one of the greatest challenges in delivering cyberinfrastructure is simply communicating its existence. The Science Gateway will serve as an awareness-building tool, demonstrating streamlined access to popular applications in campus research communities new to advanced computing. We envision the user community growing to several dozen users across multiple labs and institutions in South Dakota, with the number of groups new to advanced computing providing a key metric of the gateway's success. With a focus on lowering the bar for advanced computing, the Science Gateway will achieve success by increasing the number of researchers in South Dakota participating in computational research. In particular it will increase the number of new users to advanced computing, including classroom use, with awareness built by incorporating the gateway into training efforts through SD EPSCoR (NSF) and SD INBRE (NIH). We anticipate growth of approximately 10% by the end of the early access period (calendar year 2019).

F. *OpenTopography*

The NSF funded OpenTopography (OT) gateway [6] was initiated in 2009 to democratize access to Earth science oriented high-resolution topographic data, specifically lidar (light detection and ranging) and related compute intensive algorithms to process these data. The primary goals of OT are to enable fundamental discoveries and innovative applications with these data by streamlining data access and processing. OT utilizes cyberinfrastructure, including large-scale data management, high performance computing and service oriented architectures to provide efficient discovery, processing and visualization of large, high-resolution topographic datasets.

Since inception, 90,778 unique users have run 346,914 custom processing jobs via the OT portal, processing over 5.25 trillion lidar returns. An additional 407,064 jobs were invoked via APIs, either directly by users or via other gateways such as CyberGIS. OT has an international user community and they self-identify as being from academia (32.8%), commercial (8.8%), non-profit (5.6%), government (4.0%) and military (~1%) sectors. A conservative estimate identifies at least 290 peer-reviewed articles along with numerous these and other publications have been produced using OT gateway resources. These include academic works in Earth science, ecology, hydrology, geospatial and computer science, and engineering. OT data and tools are also routinely used in the classroom at the undergraduate and graduate level.

Our success metrics includes a number of wide-ranging factors including growth in users and usage, growth in our data holdings via growing number of partners from government and industry, national and international agencies, growing number of software collaborations including algorithm development, and finally publications enabled by OT. User metrics and usage analytics were an important component of the gateway design from an early onset and has proven to be vital for both gateway design and optimization, as well as developing partnerships with data providers who value access to OT analytics. Additional success metrics include being an important community facilitator. As part of OT education and outreach activities, we have co-organized and taught 29 short courses on high resolution technologies, processing, and applications, reaching hundreds of students, faculty, and professionals. These courses are fully subscribed and teaching materials are popular and freely available via the OT gateway.

Since its founding, the OT gateway has established itself as a mature and highly visible resource that supports a broad interdisciplinary user base in academia and beyond. The impact of OT, a collaboration between the University of California San Diego, UNAVCO and Arizona State University spans numerous disciplines across the earth and natural sciences, remote sensing, computer science, and geographic information science.

G. *CIPRES - CyberInfrastructure for Phylogenetic RESearch*

Our intent in creating CIPRES [7] was to make it possible for researchers around the world to access the compute resources needed to conduct phylogenetic relevant research. The need for CIPRES arose when the growing computational requirements for analysing a wealth of new DNA data outstripped the local compute capabilities of most researchers in the field. The primary goal was to provide easy access to the most important community codes on large HPC clusters that are adequate for most analyses. The secondary goal was to create a gateway software platform that would be robust in heavy use and that could be adopted by gateway creators in support any scientific field. CIPRES has been in operation for 10 years, and is fully mature as a gateway.

Our success metrics are aligned with the primary intent stated above. Currently, user-supplied information is combined with data from the XSEDE and CIPRES databases to track the number of user-submitted jobs, submission successes, number of users (new and returning), rate of user turnover, and amount of compute resources consumed per unit time. We no longer track user visits, number of users on the site, or number of user accounts, because historically these correlated only weakly with our intent. Through the CIPRES database we track users' country and institution, which tells us who CIPRES is supporting, to what extent we are supporting leading edge research at prestigious institutions, research at average or under-resourced institutions, training at teaching institutions. We are concerned about user experience, and track user sentiment about CIPRES features and toolkit through an annual survey. Ultimately, we believe the value of CIPRES is not that it allows users to "run jobs", but that enables creation of new knowledge. Publications are our key metric for this goal. We use a combination of Web of Science, Google Scholar, and self-reporting by users to appraise the number of publications CIPRES supports each year. Finally, we evaluate the success of the software package as a platform for gateway creation by tracking how many implementations of the CIPRES Workbench Framework are in use worldwide.

H. *QUBES*

The Quantitative Undergraduate Biology Education (QUBES) gateway [8] was launched in 2014 to address challenges in quantitative biology education. QUBES serves a community of about 7,000 math and biology educators, professional societies and education projects with an infrastructure that supports the community through communication, sharing of resources, access to software tools, and professional development. The QUBES community involves a range of post-secondary institutions, including both two- and four- year colleges from across the country. We have partnered with professional societies and education projects to promote their activities and resources, and we have connected with the discipline based education community to encourage the use of assessment in reform efforts.

In addition to building basic services for the community, we have also engaged participants in professional practices including the adoption and use of Open Education Resources (OERs) as well as the incorporation of evidence-based pedagogical approaches. We support discussion and exploration of emerging educational areas of emphasis such as developing inclusive and equitable learning environments, and incorporating new disciplinary practices such as Data Science. Our services to the community include infrastructure to share educational resources as OERs, journal clubs, websites for conferences and workshops, training and support for using software tools in the classroom, and communication about upcoming opportunities.

Our success metrics are largely focused on community activities. In particular, the number of participants in our professional development programming, and the demographics of those participants in terms of institution types and geographic distribution. Another important metric for us is the number of partner projects. These are projects designed to support quantitative biology education driven by professional societies or individuals with NSF funding. We monitor both the number of projects on QUBES, as well as the number of times QUBES is written into projects. Our partners and participants in professional development produce OERs and we monitor both the generation of these materials and their use by the broader community.

I. *'Ike Wai Gateway*

The 'Ike Wai Gateway[9], 'Ike means knowledge and Wai means water in Hawaiʻian, launched in 2018 to support research in hydrology and water management by providing data and tools to address questions of water sustainability in the state of Hawaiʻi. The gateway currently supports 60 University of Hawaiʻi (UH) researchers from UH Manoa and UH Hilo. The gateway also supports a community stakeholder decision support tool developed in collaboration with the USGS to support groundwater recharge simulations for the island of Oʻahu. One measure of gateway success is the number and quality of annotated data-sets/products and tools produced and made available. These products are developed by researchers and community members who see value in contributing products to the gateway and quality products require time to produce, verify and annotate properly to support FAIR data principles. Success is also measured by the use of the data and tools to support sustainable water management disseminated through papers, reports, policy and management actions; tracked through a combination of usage analytics and direct outreach engagements with stakeholders and community members to gather feedback and gauge impact. The total number of users of the gateway is a measure of success, however growth in the number of users who are external to the university would demonstrate that the local and broader community stakeholders are finding value in the gateway's products and services.

J. CitSci.org

CitSci.org is a global citizen science support platform and cyberinfrastructure that advances the utility, impacts, and outcomes of field-based citizen science projects [10]. This gateway currently supports more than 760 projects ranging from those that monitor water quality to maple syrup productivity and wildlife populations to invasive species as well as thousands of volunteers that have contributed close to one million scientific measurements. CitSci.org is unique in that it is fully customizable; it allows projects to 'create their own citizen science projects' in a do-it-yourself (DIY) approach. Projects can define what they wish to measure, document how they measure it, and build customized datasheets for real-time data entry online and via mobile applications. This saves hundreds of projects the costs and hassles associated with creating their own web frontend and backend systems as well as their own custom mobile apps. CitSci.org also provides an integrated suite of volunteer management capabilities plus data exploration and visualization tools to empower people to create their own visualizations of trends, relationships, and comparisons; and has been integrated with collaborative conservation systems for co-created citizen science and collaborative mental modeling. Tools also exist for volunteer communications, alerts/ notifications, bulk uploading of legacy datasets, and download of data.

For us and our science gateway, we define success as supporting and guiding citizen science projects towards meeting their own goals and objectives and generating their desired impacts and outcomes. We aim to amplify the impacts of our projects such that they create positive social-ecological-economic change for the communities in which projects are taking place. Towards this end, we aim to be change-makers by supporting projects that result in positive changes for socio-ecological systems and the quality of life for people, ecosystems, and communities. To measure success, we try to look beyond the more obvious metrics of numbers of projects, numbers of volunteers, and numbers of observations to more specifically look at harder to reach and measure metrics including percent active projects, number of peer-reviewed publications supported, number of decision-making policies informed, and the degree to which participants are being impacted through metrics of scientific literacy, behavior, attitude, and knowledge gains. A separate article [J-2] summarizes the analysis of 134 case studies to look at intent for use in decision-making.

K. CyberGIS

CyberGIS represents the new generation of geospatial information science and systems (GIS) based on advanced computing and cyberinfrastructure while the CyberGIS Science Gateway was originated as TeraGrid GIScience Gateway that is one of the earliest science gateways operated based on NSF advanced cyberinfrastructure resources [12] (Wang et al. 2016). The specific goals of our gateway are to democratize access to advanced cyberinfrastructure for enabling geospatial discovery and innovation. The gateway has gone through multiple generations of research and development cycles ranging from various back-end, middleware, and front-end technologies to user environments, and has served several thousand users in diverse domains including agriculture, bioenergy, emergency management, geography and spatial sciences, geosciences, and public health. Our success metrics relate to our stated goals in the following ways: 1) solving major scientific problems; 2) enabling broad research advances; 3) supporting education and training; and 4) engaging the public to appreciate the power and value of advanced cyberinfrastructure and cyberGIS. Accordingly, we measure major progress of related science teams, the number and variety (e.g., journal papers, dissertations, data, and software) of publications, the number of students and courses/tutorials, and the number and type of general users engaged in public outreach activities.

L. Data Discovery Studio

Data Discovery Studio (DDStudio, DataDiscoveryStudio.org) enables resource discovery in the Earth Sciences, helping users to find data, models, and other types of resources from many different repositories and resource collections used by geoscientists. Besides searching the unified catalog, users can contribute missing data resources, edit and improve metadata, organize found data into shareable resource collections, and further explore datasets using Jupyter notebooks. DDStudio has been developed through two NSF EarthCube-funded Building Blocks projects. The CINERGI (Community Inventory of EarthCube Resources for Geoscience Interoperability) project created an automated metadata augmentation pipeline that uses text analytics and several geoscience ontologies to extract additional metadata elements and add them to dataset records. These include keywords reflecting: characteristics of resources such as measured parameters, equipment used, geospatial features analyzed, scientific domains, and geospatial processes studied; missing spatial and temporal extents; and organization identifiers. Development of the initial metadata pipeline was followed by the EarthCube Data Discovery Hub project, which expanded the inventory to over 1.6 million data resources from 40+ geoscience data repositories and from community contributions, and enabled geospatial and temporal filtering, creation of collections, and dataset exploration capability.

DDStudio success metrics ultimately reflect advancing cross-disciplinary geoscience, as enabled by comprehensive data discovery across multiple repositories, using spatio-temporal, full-text and faceted search over semantically-enhanced metadata, improved ability to interpret and re-use resource descriptions, create collections, and launch Jupyter notebooks to visualize or analyze the registered resources. Translating metadata records into schema.org markup and their subsequent indexing by Google results in better visibility of geoscience data to the general public. Further, DDStudio search has been embedded in other systems (e.g., ModelMyWatershed) and on web sites. At the time of this writing, the DataDiscoveryStudio.org site has been public and active for about 9 months; in that time, the site has had 1.1K visitors in 2.6K sessions with 6.3K page views. In addition to autogenerated metrics such as the number of portal users, count of registered resources and metadata collections, number and types of metadata enhancements, and the size of the underlying integrated geoscience ontology, internal project metrics quantify search improvements through focused testing among groups of researchers from several domains.

M. COSMIC²

The COSMIC2 science gateway (cosmic-cryoem.org) provides HPC access to the structural biology community, allowing users to determine macromolecular structures using cryo-electron microscopy (cryo-EM) data [13]. Cryo-EM is the fastest-growing field of structural biology, where new users are now able to prepare samples and collect high-quality data at both regional and national cryo-EM facilities. Using these instruments, users collect thousands of movie files that can total up to 10-20 terabytes per project. We expect that there are 50-100 users around the United States who have limited to no HPC background or infrastructure available to them, although this number can be much more as cryo-EM becomes a go-to technique for structural biologists around the country. We believe that the role of the COSMIC2 gateway is to provide software and data management, allowing users to leverage cutting-edge algorithms with minimal effort for job submission.

The success of the COSMIC2 gateway will be evaluated by metrics related to user number as well as the number of cryo-EM structures calculated by the gateway. Considering that the gateway is focused on serving cryo-EM users who have limited access to HPC resources or are new the field, we will track specific numbers related to how many new users are enabled by our gateway in order to move their projects forward. This will require onboarding surveys, tracking of user job submissions, and final structures (if any) determined for a given user. We will also be tracking overall user number, but we believe an emphasis on new users will help to determine whether the gateway has led to growth in the field as to compared to helping existing users. Beyond tracking user number, we believe that another metric is the number of cryo-EM structures deposited in the Electron Microscopy Data Bank, as well as the number of atomic coordinates that are published in the Protein Data Bank. As we consider published cryo-EM structures and atomic models to be the goal of cryo-EM projects, tracking the number of these that cite our gateway will be critical for measuring success.

Based on these metrics, we believe our gateway will be functioning effectively if we are able to 1) attract new users (or users without HPC infrastructure) and 2) ensure that these new users are publishing and depositing their cryo-EM structures.

N. SimCCS

The SimCCS Gateway (simccs.org) provides novel decision-support capabilities for evaluating carbon capture, utilization and storage technologies (CCS) for mitigating greenhouse gas emissions to the atmosphere. Developed in 2018 by members of the U.S.-China Clean Energy Research Center [14], the gateway supports decision making by integrating applications in operations research, geographical information systems, carbon capture engineering, pipeline infrastructure design and reservoir performance prediction. Users are able to produce integrated CCS system designs for problems ranging from single facilities to large, regional networks involving multiple CO2 sources and geologic sinks. By harnessing the power of high-performance computing resources, users in the research and policy communities can investigate how CCS may play a meaningful role in mitigating climate change. Users in the commercial sector are able to run large ensembles of experiments to evaluate financial risk and find optimal investment solutions for implementing CCS technologies.

Although early in the development cycle, SimCCS Gateway has attracted users in the energy and policy research communities across the U.S., China and Australia. Feedback from multiple webinars and workshops convened in the first year of gateway operation has led to accelerated development of four core applications used to build, solve and analyze complex optimization problems for designing integrated CCS infrastructure solutions. Early success metrics have centered on growth in user numbers and developing an international user base in the energy research and policy communities. Such metrics are inherently aligned with the developer's goals of achieving high impact scholarship and innovative educational opportunities for students at the postgraduate level. A surprising outcome to date has been the overwhelmingly positive response by the commercial sector to the novel capabilities of the gateway. Our vision of future success includes the extent to which we can grow the user base from private-sector industries. Since CCS technologies require very significant investments to achieve large-scale reductions in greenhouse gas emissions, our goal is for SimCCS gateway to enable industry to move forward in implementing low carbon solutions.

O. CUAHSI HydroShare

HydroShare is a domain specific data and model repository operated by the Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI) to advance hydrologic science by enabling researchers to more easily share data, model and workflow products resulting from their research and to create and support reproducibility of the results reported in scientific publications. HydroShare was launched in June 2015 and as of March 1, 2019 has 3650 user accounts and supports the mandate for open data that is findable, accessible, interoperable and reusable (FAIR). HydroShare is comprised of two sets of functionality: (1) a repository for users to share and publish data and models, collectively referred to as resources, in a variety of formats, and (2) web application tools that can act on content in HydroShare for computational and visual analysis. Together these serve as a platform for collaboration and gateway for computation.

HydroShare metrics are ultimately targeted at quantifying the impact of HydroShare on advancing hydrologic knowledge. This is a challenge as impact and advances in knowledge are hard to quantify. We measure the number and type of users, how long they have had their account and how active they are. New users with an account created in an adjustable reporting period (e.g. last 30 or 180 days) quantify how the user pool is broadening and capabilities and practices that HydroShare supports are penetrating into the field (750 new users for 180-day period that ended 3/1/2019). Returning users are defined as users who were active (logged in and used the gateway) in the most recent reporting period but created their account prior to its start (450 returning users for 180 -ay period that ended 3/1/2019). This metric quantifies sustained adoption. We also track the number of resources (8784 total as of 5/10/2019). A resource is the discrete unit of digital content within HydroShare and may include observation data, model and analysis results, and models and scripts. Resources are framed as social objects that can be published, annotated, discovered and accessed and serve as the basis for collaboration [15]. Resources may be private while they are still under construction and have incomplete metadata. Private resources may be shared with other users or groups while being collaboratively developed, but ultimately we like to see resources transition to public or permanently published (immutable) and track the status (private 5203, discoverable 646, public 2730, and permanently published 205 as of 5/10/19) and strive to incentivize publication. One incentive is that permanently published resources are assigned a citable and trackable digital object identifier (DOI).

App launches for different apps (any web-based tool that connects to HydroShare through the API) help us know what users are doing and what apps are having an impact. One of the motivations for HydroShare is reuse of the data shared and published. We participate in schema.org to enable discovery of public content through systems harvesting this metadata (most notably Google data) and enable linkages between Google Data and Google Scholar that are part of the Google information ecosystem. We also participate in the Clarivate Data Citation Index (DCI). The DOI for permanently published HydroShare resources is registered with DCI to enable tracking of papers indexed in the Web of Science that cite data in HydroShare. While citation does not actually imply reuse and there can be reuse without citation, this is the best metric we have to date that quantifies impact on the advances in knowledge that have come about through HydroShare.

III. SUMMARY AND INSIGHTS ACROSS GATEWAYS

Clearly metrics of success can vary considerably across science gateways. While user counts, citizen science observations and publication counts are often a common theme, gateway leaders are clearly thinking much more deeply about what constitutes success.

Not just the number of users, but the types of users, the activities of users and the satisfaction of users are very important. Some gateways study user behavior to improve usability, others conduct surveys. Many gateways are interested in attracting new users - sometimes from particular communities, so capturing relevant information when users set up accounts will be important. Many gateways also track the behavior of users - what tools do they use; what content do they access, create or share; how many jobs do they run; how often do they return; what does user turnover look like, are users more productive? Citizen science user impacts include the degree to which participants are being impacted through metrics of scientific literacy, behavior, attitude, and knowledge gains. Gateways designed for use by students look at how well students meet their learning goals.

Some gateways look more in a more detailed way at the usage, for example many may be looking for usage patterns that reflect the full-featured offerings of the gateway. Are users using datasets across multiple repositories or reuse data in new ways, is the gateway used to drive policy and management decisions? Some of this requires direct outreach to users in order to track. The Citizen Science Alliance for example does a small number of labor-intensive case studies to better understand the impact on policy that just a few of the hundreds of science teams using their framework are having.

Gateways also look at what their users produce - publications, but also educational materials, reports, datasets and software products related to the gateway, as well as how these are shared with and used by others. In addition to asking users report citations, some gateways are proactively using the Clarivate Data Citation Index, the Web of Science and Google Scholar. Collaborations and partnerships enabled by the gateway are still another measure.

The important thing for gateway designers is to think about all of these goals and establish the means to collect, track, publicize and adapt metrics over time. These deeper success metrics are inexorably linked to the deeper goals of the PI in partnership with their funding entity. The definition of success is also fluid, and its constituents represent a fluid set of needs. As the community evolves and technologies change, the metrics may change as well..

IV. FUTURE WORK

The use of success metrics among science gateway developers is a vibrant topic that will never be fully covered. Future work includes inclusion of an international perspective, planned for the International Workshop on Science Gateways 2019 [16] and a subsequent submission of expanded content for the next gateway-themed special journal issue. In order to allow continued expansion on this topic, SGCI is considering augmenting the gateway catalog [17] to support the inclusion of success metrics. This would provide a central repository for new gateway designers to peruse to learn about the variety of success metrics in use by others.

ACKNOWLEDGMENT

The lead author would like to acknowledge the extraordinary interest of SGCI clients in serving as co-authors for this work.

REFERENCES

[1] K. Lawrence., et al. "Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community," Concurrency and Computation: Practice and Experience 27, No. 16 (2015): 4252-4268.

[2] B. Demeler, et al. "UltraScan-III version 3.5: a comprehensive data analysis software package for analytical ultracentrifugation experiments." (2017).

[3] IPT Gateway, website accessed on April 3, 2019: https://ipt.tacc.cloud/

[4] Extensible data sharing powered by SeedMe2 building blocks website accessed on May 13, 2019 https://dibbs.seedme.org

[5] P. Calyam, S. Nair, "Science Gateway Development to aid Cyber and Software Automation for Neuroscience Researchers and Educators", 13th Gateway Computing Environments Conference (Gateways), 2018; CyNeuro - http://cyneuro.org

[6] S. Krishnan, et al.., 2011, May. OpenTopography: a services oriented architecture for community access to LIDAR topography. In Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications (p. 7). ACM.

[7] Miller, Mark A., Wayne Pfeiffer, and Terri Schwartz. "Creating the CIPRES Science Gateway for inference of large phylogenetic trees." 2010 gateway computing environments workshop (GCE). Ieee, 2010.

[8] Donovan, S., et al. 2015 QUBES: a community focused on supporting teaching and learning in quantitative biology. Letters in Biomathematics, Vol 2(1); 46-55, doi: 10.1080/23737867.2015.1049969

[9] S.B. Cleveland,. J. Geis, G.A. Jacobs, 2018. "The `Ike Wai Gateway- A Science Gateway For The Water Future of Hawai`i" Proceedings of Science Gateways 2018, Austin TX, USA September 2018 (SGC18) DOI:https://doi.org/10.6084/m9.figshare.7152464.v2

[10] [G. Newman, J. Graham, A. Crall, and M. Laituri. The art and science of multi-scale citizen science support. Ecological Informatics, (3-4):217–227, 2011.

[11] [G. Newman et al., Leveraging the power of place in citizen science for effective conservation decision making, Biological Conservation, Volume 208, 2017, Pages 55-64, ISSN 0006-3207, https://doi.org/10.1016/j.biocon.2016.07.019.

[12] [K-1] S. Wang,, Y. Liu, and A. Padmanabhan, (2016) "Open CyberGIS Software for Geospatial Research and Education in the Big Data Era". SoftwareX, 5: 1-5

[13] Cianfrocco MA, Wong M, Youn C, Wagner R, Leschziner AE. 2017. COSMIC²: A Science Gateway for Cryo-Electron Microscopy Structure Determination. Proceedings of Pearc'17 Conference, New Orleans, LA, USA, July 9-13, 2017.

[14] Y. Wang, S.Pamidighantam, S. Yaw, E. Abeysinghe, S. Marru, M. Christie, K. Ellett, M. Pierce, and R. Middleton. 2018. A New Science Gateway to Provide Decision Support on Carbon Capture and Storage Technologies: Proceedings of Pearc'18 Conference, Pittsburgh, PA, USA, July 22–26, 2018, DOI: 10.1145/3219104.3229244.

[15] Horsburgh, J. S., M. M. Morsy, A. M. Castronova, J. L. Goodall, T. Gan, H. Yi, M. J. Stealey and D. G. Tarboton, (2016), "HydroShare: Sharing Diverse Environmental Data Types and Models as Social Objects with Application to the Hydrology Domain," JAWRA Journal of the American Water Resources Association, 52(4): 873-889, 10.1111/1752-1688.12363.

[16] http://iwsg2019.eu accessed May 15, 2019.

[17] http://catalog.sciencegateways.org accessed May 15, 2019.