We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300
Open access books available

170,000
International authors and editors

190M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Science Gateways and AI/ML: How Can Gateway Concepts and Solutions Meet the Needs in Data Science?

Sandra Gesing, Marlon Pierce, Suresh Marru, Michael Zentner, Kathryn Huff, Shannon Bradley, Sean B. Cleveland, Steven R. Brandt, Rajiv Ramnath, Kerk Kee, Maytal Dahan, Braulio M. Villegas Martínez, Wilmer Contreras Sepulveda and José J. Sánchez Mondragón

Abstract

Science gateways are a crucial component of critical infrastructure as they provide the means for users to focus on their topics and methods instead of the technical details of the infrastructure. They are defined as end-to-end solutions for accessing data, software, computing services, sensors, and equipment specific to the needs of a science or engineering discipline and their goal is to hide the complexity of the underlying infrastructure. Science gateways are often called Virtual Research Environments in Europe and Virtual Labs in Australasia; we consider these two terms to be synonymous with science gateways. Over the past decade, artificial intelligence (AI) and machine learning (ML) have found applications in many different fields in private industry, and private industry has reaped the benefits. Likewise, in the academic realm, large-scale data science applications have also learned to apply public high-performance computing resources to make use of this technology. However, academic and research science gateways have yet to fully adopt the tools of AI. There is an opportunity in the gateways space, both to increase the visibility and accessibility to AI/ML applications and to enable researchers and developers to advance the field of science gateway cyberinfrastructure itself. Harnessing AI/ML is recognized as a high priority by the science gateway community. It is, therefore, critical for the next generation of science gateways to adapt to support the AI/ML that is already transforming many scientific fields. The goal is to increase collaborations between the two fields and to ensure that gateway services are used and are valuable to the AI/ML community. This chapter presents state-of-the-art examples and areas of opportunity for the science gateways community to pursue in relation to AI/ML and some vision of where these new capabilities might impact science gateways and support scientific research.

Keywords: science gateways, virtual research environments, artificial intelligence, machine learning, collaboration

1. Introduction

Science gateways are end-to-end solutions for accessing data, software, computing services, and equipment specific to the needs of a science or engineering discipline. The goal of science gateways is to hide the complexity of the underlying research infrastructure and to enable scientists and educators to focus on their research and teaching—science gateways form one of the building blocks for critical infrastructure in research. Science gateways are often called Virtual Research Environments (VREs) in Europe and Virtual Labs (VLs) in Australasia [1]; we consider these two terms to be synonymous with science gateways. While quite a few research domains such as the life sciences, chemistry, and geospatial sciences have adapted the use of science gateways, there is still the need for a larger uptake in those domains and for broadening participation to further domains and user groups such as high-school students [2].

Some artificial intelligence/machine learning (AI/ML) research is supported by science gateways. Usually, this in the form of "software-as-a-service" for developed and trained AI/ML applications, for example, AI4Mars [3] on Zooniverse [4] that offers science gateways as a service for citizen sciences (see **Figure 1**).

Even though AI/ML and science gateways are both well-anchored in the high-performance computing (HPC) community, the field of science gateways still has low visibility to AI/ML application developers and users. A reason for the small degree of overlap of the AI community with the science gateway community includes that the AI/ML community has focused more on developer tools and languages than on intuitive and graphical interfaces. The trajectory of new concepts in academia is often first to develop effective methods, second to increase their efficiency and then, finally, to open them up to a wider community via considering usability. In the case of AI/ML it means to enhance the visibility via a set of capabilities that support AI/ML development and improve the uptake of science gateways in the AI/ML community. Furthermore, there is a need to advance the field of science gateway cyberinfrastructure itself.

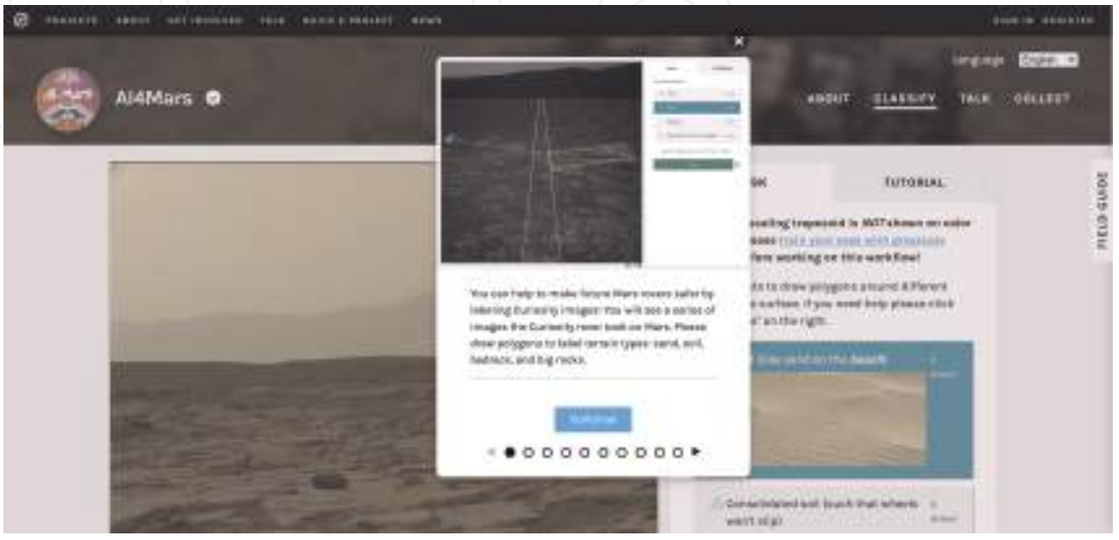


Figure 1.
The project AI4Mars uses citizen sciences to teach Mars rovers how to classify martian terrain.

These topics are critical for the next generation of science gateways and its community as AI/ML is already transforming many scientific fields.

One goal of science gateways is to facilitate collaborations between highly technical practitioners and less computer-savvy researchers, and thereby expand the reach and impact of science. In this case, it means to make AI/ML services accessible to the community. This chapter is part of the work to investigate the target groups in AI/ML research and to identify the opportunities and activities needed to achieve this goal and contribute significantly to critical infrastructure in research, teaching, and beyond.

For promoting collaborations between the science gateway community and the AI/ML community, it is important to target the three major groups: academic communities, funding agencies, and industry. Important actors in the science gateway community for increasing the collaboration are providers and developers of mature science gateway frameworks such as HUBzero [5], Apache Airavata [6], and Tapis [7]. Thus, in the area of academic communities, they can organize special tracks at Gateways conferences [8], which allow more conversations and exchanges of ideas in various science gateway communities including users of science gateways. Another possibility for widening the outreach is to hold webinars, panels, birds-of-a-feather sessions, and similar outreach efforts at conferences such as Practice & Experience in Advanced Research Computing (PEARC) [9], eScience [10], and Supercomputing [11] to promote the use of science gateways to research computing and interdisciplinary research communities, including AI/ML experts attracted by these conferences.

Another way to reach AI researchers is at their domain-specific conferences. Presentations at appropriate AI conferences would raise awareness of the concept of science gateways and would elucidate the opportunities to address pain points regarding usability and integration with complex computing and data infrastructures, for example, the International Conference on Machine Learning, Optimization and Data Science [12]. Papers could be "crowd-sourced" as previous papers to International Workshop on Science Gateways (IWSG) [13] or this manuscript.

Funding agencies and funded institutes/large projects around AI form another important target group for gathering requirements on usability and accessibility of AI/ML methods.

Examples of outreach to these groups include the promotion of science gateways to AI/ML award winners, especially at the various AI institutes [14] funded by National Science Foundation (NSF), and contacting cognizant program officers of AI/ML research-supporting programs to discuss roles for science gateways. While science gateways are well known in NSF directorates such as Office of Advanced Cyberinfrastructure (OAC) [15], program officers in other NSF directorates or other federal agencies might be less familiar with or not yet know about science gateways. In the 2019 update to the NSF National Artificial Intelligence Research and Development Strategic Plan [16], one of the major topics was the recognition of the gap between the capabilities of AI algorithms and the usability of AI systems by humans. The report states "Human-aware intelligent systems are needed that can interact intuitively with users and enable seamless machine–human collaborations." This is exactly the goal of science gateways and the fact that they are not mentioned as a potential solution emphasizes the point that science gateways are not yet well known in the AI community.

As a large producer and user of AI/ML concepts and technologies, industry is an important target group, especially projects that foster collaboration of funding agencies and companies, that is, a collaboration between Amazon and NSF that funds ten



Figure 2.
The Permafrost Discovery Gateway uses AI to provide access to pan-Arctic permafrost knowledge and information about the globe regarding pan-Arctic change.

research projects on fairness of AI [17]. The uptake of science gateways in industry is one of the goals some science gateway providers pursue for widening their community. Such adoption is also a measure of the sustainability achieved by science gateway frameworks. Meaningful connections include to introduce science gateways to technology providers, especially those in the space of cloud services such as Omnibond Systems who are already part of the science gateways community.

While there is interest especially from science gateway providers to form collaborations with academic communities, industry, and funding agencies, it is important to carefully select content to be presented to the target audiences. In order to achieve a wide outreach, we aim at answering an overarching question: how can Gateway concepts and solutions meet the needs of data science? In order to answer this question, the chapter is laid out with the following sections: first, we provide a brief background on AI, ML, and data-intensive computing in general. Second, we explain the terminologies, especially the difference between AI, ML, and data-intensive computing. Third, we explain what science gateways can do and describe their general capabilities. In this section, we also provide example gateways across different disciplines. Fourth, we discuss several science gateway opportunities for AI/ML research. Finally, we wrap up the chapter with the future outlook (**Figure 2**).

2. Background

Since its introduction [18], interest in AI has gone through several peaks and troughs. Generally, peaks have been driven by algorithmic progress coupled with the availability of appropriate computing resources and the troughs by their lack. This in particular has been true of brute force algorithms and shortcut modifications to those algorithms that prune brute force exploration. Contemporary with the notion of AI was the publication of the foundations of neural networks [19]. Although first appearing at a similar point in time, neural networks only began receiving widespread attention as a means for ML when software libraries and products became available to allow nonexperts to test the application of neural networks within their own research

domains, especially during the 1980s. After a great wave of progress, limitations of existing neural network topologies and training algorithms were realized, spawning decades of research into more effective ways to represent and compute upon neural networks.

One of the confounding factors for early progress on ML was the need for ML algorithms to be fed by large amounts of training data. In many domains, such training sets did not exist. However, one of the significant drivers of progress in this area has been the emergence of the Internet. The past three decades has seen the creation of massive training sets in the form of user actions on the Internet, immense corpi of content authored by people participating in the Internet, large libraries of images and video becoming widely available, and so forth. Another key force in creating training sets has been the advancement of massively parallel supercomputer resources (e.g., XSEDE [20], OSG [21]) that can compute large amounts of data using physics-based models, which subsequently can be used to produce ML-based approximate models to rapidly compute these same outputs. These decades have also created a paired interest in ML involving both scientific and commercial interests. Although this pairing creates a variety of potential ethical issues, it has driven progress in this field at a rate faster than most fields that lack this symbiosis. It has even seen commercial organizations making ML algorithms available to the public (e.g., see Abadi et al. [22]).

Based on this history, today AI, ML, and (more generally) data-intensive computing have been identified as high priorities for federally funded research. In both 2016 [23] and 2019 [16, 24], reports by the National Science & Technology Council, spanning two different presidential administrations. In response, comprehensive programs and funding priorities have been put forward by federal agencies including the NSF [25], DARPA [26], NIH [27], and DOE [28]. AI/ML methods are also high priorities for mission-driven science agencies such as National Aeronautics and Space Administration (NASA). National Institute and Standards and Technology (NIST) is leading efforts related to safety and benchmarking of AI/ML applications. Many government agencies are releasing large, curated data sets that can be used for training.

The applications of AI/ML research are scientifically promising, of strategic importance to national defense, and important for economic competitiveness. Research utilizing AI/ML is expanding in multiple specific domains, including big data and high-energy physics, astronomy, animal husbandry, agriculture (Ag), food security, climate change, and city infrastructure. The AI100 Project [29] and the Computing Community Consortium [30], as well as the National Science & Technology Council Reports, provide long-range overviews of AI/ML research challenges and opportunities, including their likely impacts on society as a whole.

Science gateways are widely known to bring advanced scientific capabilities to researchers in the form of data sets, HPC resources, and instruments such that researchers do not need to be experts in accessing those resources [31]. Today, we face another evolution in AI/ML, where it can be further democratized and advanced through the use of science gateways. This overview examines the opportunities for integrating AI/ML research with science gateway cyberinfrastructure, based on the extensive background information surveyed above.

3. Terminology

Following Stone et al. [30], we will distinguish AI, ML, and data-intensive computing as follows:

- AI is a branch of computer science that studies intelligence by synthesizing intelligence. AI is a broad field that encompasses subfields that include ML, autonomous systems, simulations of biologically based intelligence, and other fields.

- ML is a branch of AI that examines algorithms and methods by which computer programs can be taught to recognize patterns in data sets for purposes such as classification, recognition (i.e., facial, speech, and character), recommendation, surrogate models, and decision-making, among others.

- Data-intensive computing is a general term for computing that consumes large amounts of data, either streaming or static, as input, presenting challenging problems for scalably integrating storage, computing, and I/O. ML methods may be used in data-intensive computing.

This overview focuses on the requirements of ML methods that are being applied to a wide range of scientific data in diverse scientific fields in support of scientific research. ML holds the promise for scalably extracting information and knowledge (including scientific insights) from the large amounts of data generated by both experiments at all scales as well as scientific simulations, and fits well with the capabilities of science gateways today.

4. Science gateway infrastructure

Science gateways in general are noted for their ability to provide the following capabilities to support scientific research:

- Simplified access to research computing and storage resources.

- Ability to provide scientific software as a service.

- Ability to integrate diverse, distributed computing and data into a single platform.

- Ability to provide a range of scientific and engineering environments that support diverse stakeholder groups in a particular community.

- Ability to securely control access to resources and data.

- Support scientific collaboration through the sharing of access to results.

- Support for reproducibility of computational results.

In other words, we can consider science gateways as cyberinfrastructure environments to support Findable, Accessible, Interoperable, and Reusable (FAIR) research [32]. Many of the concerns and opportunities identified in [16, 23, 24] for the use of AI/ML research are FAIR challenges. The FAIR principles have created significant momentum in the research community recognizing a need to improve the quality of research by establishing common standards. However, bridging the principles with the research infrastructure remains a challenging task due to its diversity and domain-

specific nature of tasks. Science gateways provide an excellent opportunity to achieve FAIRness of research data and software.

4.1 Use cases in artificial intelligence

4.1.1 Physics

International collaboration is often basic for our scientific development, and these efforts are at the core of its infrastructure. Researchers are partnering with NSF and DOE to study how AI Frameworks can be leveraged in physics research. One such project, ML and FPGA Computing for Real-Time Application in Big-Data Physics Experiments as a science driver investigates the creation of a FAIR framework for AI [33]. Example is a project focusing on Inspired Artificial Intelligence in High-Energy Physics which builds on the successes of the last years with the Large Hadron Collider (LHC) and the combination of the Laser Interferometer Gravitational—wave Observatory (LIGO) and the Large Synoptic Survey Telescope (LSST) for Multi-Messenger Astrophysics by making artificial AI models and data more accessible and reusable with the goal to accelerate research and outperform current approaches [34]. Broad international collaborations as the Event Horizon Telescope (EHT) are excellent opportunities to introduce science gateways. This single-event global array of eight ground-based radio telescopes aimed at obtaining the image a black hole and its shadow could become a n international mainstay [35].

4.1.2 Photonics and quantum optics

AI is already taking part in the development of future technologies, as is the case of the quantum technologies. In this case, the integration of two communities, AI and photonics, have become complementary during the last decades, where ML protocols are being matched to photonic platforms giving rise to photonic neural network architecture [36]. The applications of AI, especially neural networks and ML in the field of quantum optics have also become prevalent for experimental setups used for classification and identification of light sources and quantum states by using these two approaches [37–39]. Such examples with computational properties have low-complexity and low-cost implementations promising quantum architecture that could apply underlying cyberinfrastructure enabling users to create their own workflows to run simulations codes. Those applications designed by the quantum community would generate tools with unprecedented capabilities available to researchers unfamiliar with the world of quantum optics and photonics.

4.1.3 Astronomy

The goal of the AGNet [40] project is to leverage AI to develop a novel interdisciplinary approach combining astronomy big data with ML tools to build a deep learning algorithm to estimate the masses of super-massive black holes. Measuring the masses *via* traditional methods is very expensive and such a new algorithm could transform the field of cosmology.

4.1.4 Agriculture

The connection between AI and Ag seems obvious. One applies big-data analytics, ML, and deep learning algorithms in geospatial information systems (GIS), satellite

data, lidar information, sensor data, and other tools and technologies to improve crops [41]. However, the connection is not as clear when discussing animal husbandry. An AI tool Project [42], Solving Dairy Cattle Genetic Improvement Challenges using Deep Learning, will use AI "to identify cattle that have the highest genetic potential for milk production and health status and make simplistic assumptions about the relationship between phenotypes and genotypes." Another interesting connection between AI and Ag is in the area of Food Security [43]. The Alan Turing Institute [18] is using ML and AI to leverage data and models of plant development, plant pathology, crop yields, and climate science to form a cohesive national crop modeling framework for the United Kingdom.

4.1.5 Climate change

The impact of climate change on the planet is much discussed in the news, but the ability to understand the true influence is limited by the ability to quantify how multiple factors work together to impact this change. The Permafrost Discovery Gateway [44] is using AI to assist with the management of ingesting large amounts of remote sensing data into machine and deep learning models which will ultimately provide "access to pan-Arctic permafrost knowledge, which can immediately inform the economy, security, and resilience of the Nation, the Arctic region, and the globe with respect to pan-Arctic change."

4.1.6 Urban planning

What do roads, sidewalks, parks, access to food and medical care, and even tree canopies providing shade on roads and sidewalks have to do with AI? Many studies have been done on food deserts and transit deserts, but by leveraging AI, researchers are able to look at all different types of neighborhood-scale infrastructure [45] (see **Figure 3**). By identifying infrastructure deserts, communities' ability to deal with them will be strengthened in the long term, particularly in low-income communities.

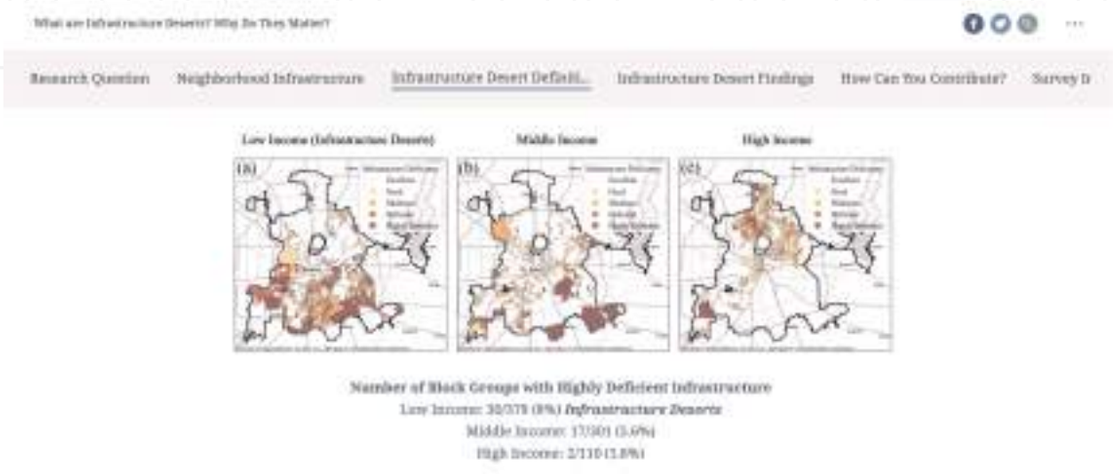


Figure 3.
The figure shows the different infrastructure deficiencies dependent on income in neighborhoods in Dallas.

4.1.7 Biology

Concerned with the detection of Regulatory Elements using GRO-seq data, the dREG science gateway [46] identifies the location of DNA sequence regions known as transcript regulation elements including promoters and enhancers—the critical components of the genetic regulatory programs of all organisms. The dREG computational code itself uses a support vector machines-based model trained by large-scale data. This science gateway democratizes the use of these sophisticated ML techniques to a wider community. The gateway interfaces with XSEDE compute infrastructures for seamlessly enabling access to compute intensive training and prediction phases. On the front-end of these ML models, user-friendly data visualization interfaces enable a wider community to interact with bigWig data, dREG signal predictions, and genome coordinates of peaks of transcriptions.

4.1.8 Humanities

Snow Vision uses image classification methods to identify pottery sherds created by Native Americans of the US Southeast. The Snow Vision science gateway [47] enables the humanities community to utilize ML-based matching algorithms to compare user uploaded sherd images to identify the original stamped designs from which their fragments descend. The gateway makes available the matching algorithms implemented by a deep-learning Point Cloud Library (PCL) for generating depth maps from 3D sherd image files and Caffe [48] deep learning toolkit.

5. Science gateways, quantum computing, and artificial intelligence

New paradigms in computing open up novel areas for exploring the potential of science gateways and AI/ML. A future direction is the emerging paradigm of quantum computing which uses the fundamental properties of quantum mechanics, such as superposition, entanglement, and interference. This domain strongly differs from classical computing, where one data qubit is equivalent to two classical bits of information. This feature is a promising solution for higher computational power within shorter calculation time, which is very useful for artificial neural networks and ML. At this stage, the latest frontier of computation relies on the hybrid development of these two areas in quantum computing and how both could support each other in the evolution of classification and clustering of big classical-to-quantum data. While the topic of AI and ML is well established in computer sciences, both are quite novel approaches on the science and technology side and not fully adapted yet. We analyzed where ML has developed better among those different fields and where it has moved to Quantum computation, more specifically, in the Quantum Machine Learning (QML) frontier. We have used Scopus data to show (after a normalization) in which of the sciences have become more active. A Scopus search on ML will provide us with the following fields list:

1. Computer Science

2. Engineering

3. Mathematics
4. Medicine
5. Physics and Astronomy
6. Biochemistry, Genetics and Molecular Biology
7. Decision Sciences
8. Materials Science
9. Social Sciences
10. Energy
11. Earth and Planetary Sciences
12. Environmental Science
13. Chemistry
14. Neuroscience
15. Business, Management and Accounting
16. Agricultural and Biological Sciences
17. Chemical Engineering
18. Multidisciplinary
19. Arts and Humanities
20. Health Professions
21. Pharmacology, Toxicology and Pharmaceutics
22. Immunology and Microbiolog
23. Psychology
24. Economics, Econometrics and Finance

Figure 4 provides those answers by displaying in blue the above fields list and in orange those in QML. The striking disparity on Physics and Astronomy (5), Materials Science (8) and Chemistry (13) can be explained from their expected larger processing demands. Besides, the actual growth of QML has happened in a much shorter timespan but pursues the very active growth shown in ML. Therefore, larger uptake of ML and QML is a fundamental need and can be improved by introducing science gateways.

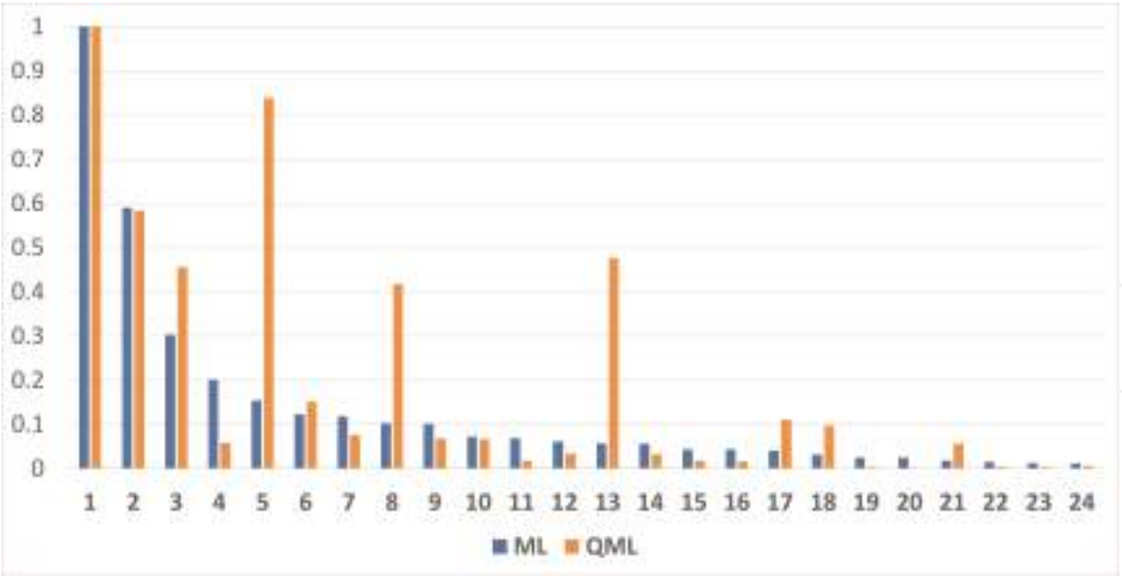


Figure 4.
Distributions of QML (orange bar chart) and ML (blue bar chart) into the different fields of Science and Technology. The uptake of QML on ML is described by normalization ratio of QML/ML of 139.6.

6. The theoretical framework for science gateways in AI/ML research

AI/ML research faces following challenges that can be addressed by a theoretical framework for integration with science gateways. Each challenge is part of the research question how science gateways can improve AI/ML research and add aspects that are not well considered as part of the AI/ML research yet.

6.1 Availability

Providing software-as-a-service is a common mode of operation for science gateways. Many gateways are already providing trained ML methods to their user communities. In this sense, an ML application is just another piece of software that a gateway can provide, simplifying access and helping to ensure the software is up to date, is used in the correct way, and is installed on adequate resources.

High-performance AI/ML is a strategic priority, as it will be necessary for these methods to scale to support enormous data sets. Providing simplified access to HPC and other specialized resources is a strength of science gateways, particularly as the landscape of those resources evolves into Graphics Processing Unit (GPU), Field Programmable Gate Arrays (FPGA), and other environments tailored for efficient AI/ML. Various federal initiatives are promoting improved AI standards, access to benchmarked, open-source applications, and access to public data sets usable for training. Science gateways will provide access to these tools and data.

In science gateways, we typically consider AI/ML-based applications as having a target audience beyond that of the developer; that is, a developer has created an application and wants it to reach a larger user community. There is also an important scenario in which a researcher develops an AI/ML application to further their own research; the application itself may have no (or at least no perceived) broader use by others, at least in the initial phase. It is still essential that science gateways are available to support such research. Even if the software itself is never used again beyond

the original scope, its results may be published and thus must be auditable and reproducible by reviewers and future readers.

6.2 Validity, reproducibility, transparency

Software based on ML methods is very different from traditional scientific software in that the ML-based applications must be trained on data sets. Thus, one must separately validate any results obtained from an ML-based application. Changes in training data sets will give different results, so it is important to track not only the versions of a particular software used but also the version of the trained model and the data sets used in training, including any processes for cleaning or otherwise filtering the data.

Gateways, in their role as supporting scientific software-as-a-service, are already well-positioned to at least support the collection of version metadata needed to support reproducibility or at least the provenance of how a particular result was achieved.

More broadly, many ML applications should be understood as dataflow programs that combine well-known algorithms into specific applications. Experiments to improve the workflow and the validation of the AI methods are currently performed outside most gateways and are considered publication quality research in such journals as *Nature Methods*. Capturing these processes is an important enhancement that could easily be supported by science gateways.

Minimally, a provider of an ML-based application could at least publish the metadata about how a particular application was developed, trained, and validated, but gateways themselves can also support these processes directly. This would enable users to reproduce and inspect the application itself and the training data. Gateways could furthermore track the development of alternative pipelines and trainings by both the code authors and the interested members of the community. This would provide direct support and supplemental information to the methods publications by the original authors.

6.3 Privacy

Data privacy is the obverse aspect of trustworthiness, since many ML applications may work with sensitive data such as personal health information and proprietary data. Science gateways can be used to support privacy for data sets by limiting access to data through controlled and auditable user and programming interfaces. These gateways can operate within privacy protected environments that support The Health Insurance Portability and Accountability Act (HIPAA), Federal Information Security Modernization Act (FISMA), and other regulated data classifications. Further limitations, such as differential privacy, are open-research areas that can also be supported by gateways.

6.4 Trustworthiness, explainability, and uncertainty quantification

Trustworthiness, explainability, and uncertainty quantification are larger open problems in AI/ML research. Trustworthiness of scientific results obtained from AI/ML methods in scientific applications can be increased through the ability of science gateways to support reproducibility, auditability, and transparency in tracking how a particular application was developed, trained, and validated. Gateways can serve as a focal point where experimental results, computational results, and AI/ML-based models can be cross compared and validated. Explainability is an open-research area, as the results of many current methods (most notably artificial neural networks) cannot be understood by humans, even if we have full access to the software and

training data. However, as new, more explainable methods are developed, science gateways should be an important delivery mechanism. Science gateways may also be a strong vehicle for delivering codes that analyze AI/ML-based models for explainability. Such explainability and trustworthiness will be essential for R&D applications in regulated spaces such as clean energy systems and climate-mitigation technologies. Trustworthy and transparent workflows for AI/ML models will be necessary to apply those approaches to advanced nuclear energy and integrated energy systems technology R&D, a field which demands high standards of safety and performance. Finally, an ML-based application could be completely trustworthy and explainable and still give the wrong answer; more precisely, the answer has both known and unknown limits. Known limits are probabilities of correctness, and some AI/ML methods (such as reinforcement learning) benefit from continual training and supervision. Collecting feedback on correct versus incorrect outputs needs to be coordinated. AI/ML software deployed into gateways, as opposed to running in separate, isolated environments by each user, can track each methods' success rates.

6.5 Usability and user experience

Using gateways to enhance usability and the user (and researcher) experience has probably received the least amount of attention by the field. The application of AI/ML methods within gateways to enhance user experience (such as guided access, usage analytics, digital assistants) would enhance science gateways' capabilities and advance the field to a new level of maturity. In this space, we expect Javascript-based AI/ML solutions to become important for gateways. Already, there is a vibrant ecosystem of such frameworks, including: TensorFlow.js[1], ml5.js[2], Propel[3], Brain.js,[4] ml.js,[5] Neuro.js,[6] Synaptic,[7] and others. Worthy of mention, also, is ConvNetJS[8] which has some particularly user-friendly educational web applications. We note that AI has already begun to shape a lot of human/machine interfaces. Google search can know and categorize images from text. Voice assistants have become good at parsing queries and returning data correctly. We expect AI to transform Science Gateways themselves, as well as the way they work. "The main trick here is to allow humans to stay human. For decades computers were not exciting to use as they required us to change our ways."—Heilmann.[9] Gateways were always trying to help the humans stay human, AI/ML should enable them to succeed in new ways.

6.6 AI for gateway cyberinfrastructure

In addition to opportunities to leverage AI/ML in research, there are also benefits to adopting these technologies in the underlying cyberinfrastructure powering the

[1] https://www.tensorflow.org/js

[2] https://ml5js.org/

[3] https://stackshare.io/propel

[4] https://github.com/BrainJS/brain.js

[5] https://github.com/mljs/ml

[6] https://neuro.js.org/

[7] http://caza.la/synaptic/#/

[8] https://cs.stanford.edu/people/karpathy/convnetjs/

[9] https://www.infoq.com/news/2018/11/human-interfaces-ai/

gateways. For, instance, there are opportunities for centralizing sets of services that could be leveraged for small/short computational jobs (such as classifying an item) that could be provided by frameworks such as Tapis [49], Airavata [50], HUBzero [51], or even commercial cloud services (AWS lambda etc.) with potential to support a hosted catalog of AI/ML functions that could be leveraged by existing and new gateways. Gateways could leverage these lambda-like functions for AI integrations both for AI/ML research as well as incorporating some of these tools into the way the gateway is managed and delivers functionality—recommendations, analyzing gateway data/metrics, and classification of user jobs/workflows that could make gateway operations more efficient and useful to end users/researchers. Further, there is the potential for gateways to leverage AI to enhance usability and accessibility through

- Chatbots—gateways could support customers in real-time and also help reduce help service costs as leveraging AI instead of manual support. These AI chatbots could learn from researcher responses and offer better support with every day and iteration. Additionally, this can also translate to more sustainability as operations run with less staff focus on this and the ability to focus more on gateway functions.

- Accessibility compliance—automated scanning can provide gateway accessibility solutions that will audit gateway interfaces for accessibility. This allows gateway developers to always have the gateway Americans with Disabilities Act (ADA) and Web Content Accessibility Guidelines (WCAG) compliant.

- Better search with natural language processing (NLP)—a large element that can impact usability is search. When users perform searches on gateway, they are looking for something specific. Using semantic search, can make their experience more user-friendly and rewarding and allow them to utilize the gateway better and continue to use it, increasing retention.

- AI research assistants and personalized user experiences—AI-powered assistants are becoming increasingly popular in the world of e-commerce, and there is large potential for these AI-powered virtual assistants to assist gateway users and helping them out in their research journey.

- Sentiment analysis of user correspondence—AI-driven sentiment analysis tools can aid in precisely understanding how researchers feel about services and features. Such tools can analyze researcher correspondence and comments to provide a precise overview of the likeability of the current gateway. This data can help gateway managers/developers improve offerings, add new features, remove unwanted features, and offer a better user experience, leading to increased research outputs, retention, and potential growth and sustainability.

- Cybersecurity—AI-driven pattern recognition of bad actor behaviors can analyze system access and activities to aid in identification of compromised accounts or API security flaws. This type of data and recommendations can assist gateway and infrastructure providers in identifying and addressing security vulnerabilities leading to better protection of this advanced computing and data resources and research intellectual property. We go into more detail in the next section for this topic.

Overall, AI will be revolutionary in the way gateways can be developed/managed/protected and how users will interact with the gateway. Investing in and developing AI-powered tools is a concrete area that will lead to improvements in gateway usability and functionality and doing so in centralized ways can push the entire research community forward.

7. Cybersecurity and critical infrastructure protection

The primary danger posed by the use of Science Gateways comes from the "community account" model on which these applications are based. Some of the science gateway frameworks create on purpose a single account through which they schedule all the jobs by a group of users of the web-facing interface, for example, [6] allowing for accessing HPC resources nationwide in the USA.

Because it is automated, such systems must submit their jobs (typically) without going through two-factor authentication systems (although the gateway infrastructure could, in principle use a two-factor system on the web-facing side). HUBzero has implemented such modules and allows for re-use of login credentials such as Google accounts only if the two-factor system is used at least once during the first login to the system.

Science gateway frameworks are differently designed in regard to security: some only run a limited set of commands related to moving files and a science code, others such as Jupyter Notebooks take code from a web-facing interface and run it on the target system.

Most of the science codes developed in research domains are often written without any kind of security in mind. AI networks themselves, depending on the details of their training, could, in principle, contain vulnerabilities that would be very hard to identify. Individual applications are likely to have numerous vulnerabilities and a clever hacker could provide input parameter files that trigger buffer overrun attacks, etc. In addition, there is the small (but nonvanishing) chance that the science gateway framework itself might be hacked in some way.

Some solutions help mitigating these issues, for example, singularity with AppArmor on the target system. AppArmor provides kernel-level protection against arbitrary sorts of unwanted access. AI is a promising solution *via* pattern recognition of bad actor behaviors supporting to identify compromised accounts or API security flaws.

8. Outlook

Outreach to different target groups will be a crucial topic to establish new science gateways in addition to the ones already used in the AI/ML community. Zooniverse is a great example of a science gateway for collaborating on AI methods. It is one of the frameworks we plan to reach out to. The academic community is well reachable at conferences, and thus, presenting about science gateways and AI at a diverse set of conferences can connect the communities. Furthermore, federal funding programs for AI/ML research and development are important influencers in this space. NSF, NASA, Advanced Research Project Agency-Energy (ARPA-E), Office of Nuclear Energy (DOE-NE) and Office of Energy Efficiency & Renewable Energy (DOE-EERE) all have solicitations associated with AI/ML applications. The last three of these agencies

are in the energy space. Existing science gateways in the AI area are good starting points to analyze the uptake of such gateways by the community and to identify existing pain points using such solutions. Partnering with industry is a promising way to accelerate the collaboration between the science gateways and the AI/ML community. The goal is not only to increase the uptake of science gateways for AI methods but also to increase the uptake of AI methods for science gateway infrastructures. Both fields can benefit from each other and accelerate science *via* combining methods for ML with usability of science gateways. Integrating science gateways into teaching will further enhance the knowledge in the community and train students on using and potentially developing science gateways and/or AI/ML concepts and methods and, hence, train the next generation of users and developers. With AI being such an important field in academia and industry, this is a crucial step for highly needed workforce development. This combination of AI/ML and science gateways creates the next generation of critical infrastructure for research and teaching.

Acknowledgements

The authors would like to acknowledge SGCI and NSF as funding body for SGCI's funding under award 1547611. B.M. Villegas Martínez would like to thank finantial support of the Mexican National Council on Science and Technology (CONACyT) through the postdoctoral scholarship CVU 549550. Wilmer Contreras Sepulveda would like to thank CONACyT for the academic scholarship to develop master studies.

IntechOpen

DOI: http://dx.doi.org/10.5772/intechopen.110144

Author details

Sandra Gesing[1*], Marlon Pierce[2], Suresh Marru[2], Michael Zentner[3], Kathryn Huff[4], Shannon Bradley[5], Sean B. Cleveland[6], Steven R. Brandt[7], Rajiv Ramnath[8], Kerk Kee[9], Maytal Dahan[10], Braulio M. Villegas Martínez[11], Wilmer Contreras Sepulveda[11] and José J. Sánchez Mondragón[11]

1 Applied R&D, University of Illinois Systems Discovery Partners Institute, US

2 Cyberinfrastructure Integration Research Center, Indiana University, US

3 Sustainable Scientific Software Division, San Diego Supercomputer Center, US

4 Department of Nuclear, Plasma and Radiological, University of Illinois at Urbana-Champaign, US

5 Innovative Software and Data Analysis, University of Illinois at Urbana-Champaign, US

6 Information Technology Services Cyberinfrastructure, University of Hawai'i, US

7 Computer Science and Engineering, Louisiana State University, US

8 Computer Science and Engineering, The Ohio State University, US

9 College of Media and Communication, Texas Tech University, US

10 Advanced Computing Interfaces, Texas Advanced Computing Center, US

11 Optics Department, National Institute of Astrophysics, Optics and Electronics, Mexico

*Address all correspondence to: sgesing@uillinois.edu

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

References

[1] RDA VRE-IG. https://www.rd-alliance.org/groups/vre-ig.html. [Online; Accessed: 09 March 2021]

[2] Gesing S, Brandt S, Bradley S, Potkewitz M, Kee K, Whysel N, Perri M, Cleveland S, Rugg A, Smith J. A vision for science gateways: bridging the gap and broadening the outreach. In: Practice and Experience in Advanced Research Computing, PEARC '21, New York, NY, USA, 2021. Boston, MA, USA, Association for Computing Machinery

[3] AI4Mars. https://www.zooniverse.org/projects/hiro-ono/ai4mars. [Online; Accessed: 09 March 2021]

[4] Zooniverse. https://www.zooniverse.org/. [Online; Accessed: 09 March 2021]

[5] Gesing S, Zentner M, Clark S, Stirm C, Haley B. Hubzero®: novel concepts applied to established computing infrastructures to address communities' needs. In: Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning), PEARC '19, New York, NY, USA, 2019. Chicago, IL, USA, Association for Computing Machinery

[6] Pierce M, Marru S, Abeysinghe E, Pamidighantam S, Christie M, Wannipurage D. Supporting science gateways using apache airavata and scigap services. In: Proceedings of the Practice and Experience on Advanced Research Computing, PEARC '18, New York, NY, USA, 2018. Pittsburgh, PA, USA, Association for Computing Machinery

[7] Padhy S, Jamthe A, Cleveland SB, Smith JA, Stubbs J, Garcia C, Packard M, Terry S, Looney J, Cardone R, Dahan M, Jacobs GA. Building tapis v3 streams api support for real-time streaming data event-driven workflows. In: Practice and Experience in Advanced Research Computing, PEARC '21, New York, NY, USA, 2021. Boston, MA, USA, Association for Computing Machinery

[8] Gateways Conferences. https://sciencegateways.org/engage/annual-conference. [Online; Accessed: 09 March 2021]

[9] PEARC Conference. https://pearc.acm.org/. [Online; Accessed 09 March 2021]

[10] eScience Conference. https://escience2021.org/. [Online; Accessed: 09 March 2021]

[11] SC Conference Series - The International Conference for High Performance Computing, Networking, Storage, and Analysis. https://supercomputing.org/. [Online; Accessed: 09 March 2021]

[12] The 8th International Online & Onsite Conference on Machine Learning, Optimization, and Data Science. https://lod2022.icas.cc/. [Online; Accessed: 09 March 2021]

[13] Atkinson M, Gesing S. IWSG 2018 - International Workshop on Science Gateways 2018. CEUR Workshop Proceedings: IWSG; 2019

[14] NSF AI Institutes. https://www.nsf.gov/pubs/2020/nsf20604/nsf20604.htm. [Online; Accessed: 09 March 2021]

[15] NSF OAC. https://www.nsf.gov/cise/oac/about.jsp. [Online; Accessed 09 March 2021]

[16] The National Artificial Intelligence Research and Development Strategic Plan - 2019 update. https://www.nitrd.

gov/pubs/National-AI-RD-Strategy-2019.pdf. [Online; Accessed: 09 March 2021]

[17] 3 questions about the Amazon–National Science Foundation collaboration on fairness in AI. https://www.amazon.science/3-questions-about-the-amazon-national-science-foundation-collaboration-on-fairness-in-ai. [Online; Accessed: 09 March 2021]

[18] Turing AM. Computing Machinery and Intelligence. Netherlands, Dordrecht: Springer; 2009. pp. 23-65

[19] McCulloch W, Pitts W. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics. 1943;**5**(4): 115-133

[20] Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, et al. Xsede: accelerating scientific discovery. Computing in Science & Engineering. 2014;**16**(05):62-74

[21] Pordes R, Petravick D, Kramer B, Olson D, Livny M, Roy A, et al. The open science grid. Journal of Physics: Conference Series. 2007;**78**:012057

[22] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, OSDI'16, pp. 265–283, USA, 2016. Savannah, GA, USA, USENIX Association

[23] The National Artificial Intelligence Research and Development Strategic Plan 2016. https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf. [Online; Accessed: 09 March 2021]

[24] The National Artificial Intelligence Research and Development Progress Report. https://www.whitehouse.gov/wp-content/uploads/2019/11/AI-Research-and-Development-Progress-Report-2016-2019.pdf. [Online; accessed 09-03-2021]

[25] Artificial Intelligence (AI) at NSF. https://nsf.gov/cise/ai.jsp. [Online; Accessed: 09 March 2021]

[26] DARPA AI Next Campaign. https://www.darpa.mil/work-with-us/ai-next-campaign. [Online; Accessed: 09-03-2021]

[27] NIH Artificial Intelligence, Machine Learning, and Deep Learning. https://www.nibib.nih.gov/research-funding/machine-learning. [Online; Accessed: 09 March 2021]

[28] DOE Artificial Intelligence and Technology Office. https://www.energy.gov/science-innovation/artificial-intelligence-and-technology-office. [Online; Accessed: 09 March 2021]

[29] Gil Y, Selman B. A 20-Year Community Roadmap for Artificial Intelligence Research in the US. Computing Community Consortium (CCC) and Association for the Advancement of Artificial Intelligence (AAAI). 2019. arXiv preprint arXiv: 1908.02624. Available at: https://cra.org/ccc/resources/workshopreports/

[30] Stone P, Brooks R, Brynjolfsson E, Calo R, Etzioni O, Hager G, et al. Artificial Intelligence and Life in 2030. One hundred year study on artificial intelligence: Report of the 2015-2016 Study Panel. Stanford University, Stanford, CA, p. 2016, 2016

[31] Lawrence KA, Zentner M, Wilkins-Diehr N, Wernert JA, Pierce M, Marru S, et al. Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community. Concurrency and

Computation: Practice and Experience. 2015;**27**(16):4252-4268

[32] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016 Mar 15;**3**:160018. DOI: 10.1038/sdata.2016.18. Erratum in: Sci Data. 2019 Mar 19;6(1):6. PMID: 26978244; PMCID: PMC4792175. Available from: https://pubmed.ncbi.nlm.nih.gov/26978244/

[33] Project at the intersecion of AI and physics. http://www.ncsa.illinois.edu/news/story/\doe_awards_2.2m_to_project_at_the_intersection_of_ai_and_high_energy_physic. [Online; Accessed: 09 March 2021]

[34] Collaborative Research: Frameworks: Machine learning and FPGA computing for real-time applications in big-data physics experiments. https://figshare.com/articles/poster/\Collaborative_Research_Frameworks_Machine_Learning_and_FPGA_computing_for_real-time_applications_in_big-data_physics_experiments/11803764. [Online; Accessed: 09 March 2021]

[35] Akiyama K, Alberdi A, Alef W, Asada K, Azulay R, Baczko A-K, et al. First M87 event horizon telescope results. IV. Imaging the central supermassive black hole. The Astrophysical Journal Letters. 2019; **875**(1):L4

[36] Shi B, Calabretta N, Stabile R. Numerical simulation of an inp photonic integrated cross-connect for deep neural networks on chip. Applied Sciences. 2020;**10**(2):474. DOI: 10.3390/app10020474

[37] Ahmed S, Sánchez Muñoz C, Nori F, Kockum AF. Classification and reconstruction of optical quantum states with deep neural networks. Physical Review Research. 2021;**3**: 033278

[38] Bhusal N, Lohani S, You C, Hong M, Fabre J, Zhao P, et al. Spatial mode correction of single photons using machine learning. Advanced Quantum Technologies. 2021;**4**(3):2000103

[39] You C, Quiroz-Juárez MA, Lambert A, Bhusal N, Dong C, Perez-Leija A, et al. Identification of light sources using machine learning. Applied Physics Reviews. 2020;**7**(2): 021404

[40] AGNet. http://www.ncsa.illinois.edu/about/fellows_awardees/\agnet_weighing_black_holes_with_deep_learning. [Online; Accessed: 09 March 2021]

[41] Agriculture. https://calendars.illinois.edu/detail/6824/33401360. [Online; Accessed: 09 March 2021]

[42] Animal husbandry. https://ai.ncsa.illinois.edu/research/funded-projects/. [Online; Accessed: 09 March 2021]

[43] Food security. https://science-council.food.gov.uk/sites/default/\files/appofadvancedanalyticsfinalreport.pdf. [Online; Accessed: 09 March 2021]

[44] Liljedahl AK, Jones BM, Brubaker M, Budden AE, Cervenec JM, Grosse G, Jones MB, Marini L, McHenry K, Moss J, Morin PJ, Nitze I, Soliman A, Wind G, Witharana C. Permafrost discovery gateway: a web platform to enable discovery and knowledge-generation of permafrost big imagery products. In: AGU Fall Meeting 2019, San Francisco, USA, December 2019. AGU

[45] Infrastructure Deserts. https://storymaps.arcgis.com/stories/\7b8c

2029e30749449a3a692451e51ddf. [Online; Accessed: 09 March 2021]

[46] Wang Z, Christie MA, Abeysinghe E, Chu T, Marru S, Pierce M, and Danko CG. Building a science gateway for processing and modeling sequencing data via apache airavata. In: Proceedings of the Practice and Experience on Advanced Research Computing, pp. 1–7. 2018

[47] Zhou J, Smith K, Wilsbacher G, Sagona P, Reddy D, and Torkian B. Building science gateways for humanities. In: Practice and Experience in Advanced Research Computing, pp. 327–332. 2020

[48] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia, pp. 675–678, 2014

[49] Cleveland SB, Jamthe A, Padhy S, Stubbs J, Packard M, Looney J, et al. Tapis api development with python: best practices in scientific rest api implementation: Experience implementing a distributed stream api. In: Practice and Experience in Advanced Research Computing, PEARC '20, pp. 181–187, New York, NY, USA, 2020. Portland, OR, USA, Association for Computing Machinery

[50] Marru S, Gunathilake L, Herath C, Tangchaisin P, Pierce M, Mattmann C, et al. Apache airavata: a framework for distributed applications and computational workflows. In: Proceedings of the 2011 ACM Workshop on Gateway Computing Environments, GCE '11, pp. 21–28, New York, NY, USA, 2011. Seattle, Washington, USA, Association for Computing Machinery

[51] David Benham and Sandra Gesing. Hubzero© goes onescienceplace: The next community-driven steps for providing software-as-a-service. In: 2019 15th International Conference on eScience (eScience), pp. 642–643, 2019