PROJECT SUMMARY

Overview:

NSF cyberinfrastructure, including CMS, LIGO, OSG, and XSEDE, is undergoing a security transformation: a migration from X.509 user certificates to IETF-standard JSON Web Tokens (JWTs). This migration has facilitated a re-thinking of authentication and authorization among cyberinfrastructure providers: enabling federated authentication via InCommon/eduGAIN as a core capability, improving support for attribute, role, and capability-based authorization, and reducing reliance on prior identity-based authorization methods that created security and usability headaches. Achieving the benefits of a fundamentally new security credential ecosystem in our cyberinfrastructure, while avoiding the temptation to simply re-implement old X.509 methods using JWTs, requires leadership and coordination. The SciAuth project provides the needed leadership and coordination for this critical transformation through community engagement, coordinated adoption of community standards, integration with software cyberinfrastructure, security analysis and threat modeling, training, and workforce development. The project helps the community realize the benefits of an interoperable, capability-based ecosystem when transitioning between technologies, while maintaining the reliable and secure cyberinfrastructure upon which the scientific community depends.

Usable mechanisms for privilege management are critical for enabling productive scientific collaborations across a diverse and distributed scientific cyberinfrastructure ecosystem. The SciTokens project demonstrated that the use of JWTs with the IETF OAuth standard for privilege delegation provides a breakthrough for interoperable, least-privilege resource sharing in scientific collaborations. Now our challenge is to make that breakthrough technology usable by scientists across disciplines, project sizes, and software ecosystems by enabling coordinated deployments across cyberinfrastructures in active use today. The SciAuth project reunites members of the SciTokens team to address this next challenge.

Intellectual Merit:

The SciAuth project advances intellectual knowledge through adoption of Internet standards (JWT, OAuth, OIDC) for interoperable, least-privilege authorization across scientific cyberinfrastructure, in partnership with science projects (CMS, IceCube, LIGO, WLCG) and cyberinfrastructure providers (Fermilab, OSG, PATh, IRIS-HEP, XSEDE).

Broader Impacts:

The SciAuth project benefits society by supporting an enhanced infrastructure for research and education through the transition to a new security credential ecosystem and by supporting the development of a diverse, globally competitive STEM workforce through a fellows program that pairs students across the country with mentors from the project to collaborate on student-led projects on the topic of cyberinfrastructure security.

Project Description

Cyberinfrastructure is a key enabler for scientific collaborations today. Scientists interact with a variety of cyberinfrastructure components, distributed across campus and around the world, as an integral part of their day-to-day research activities. As cyberinfrastructure providers, our goal is for these interactions to be *seamless*, enabling the scientist to access computing, data, instruments, and software on the web, the desktop, and the command-line, in collaboration with fellow researchers near and far, without facing technical roadblocks that get in the way of scientific productivity.

Interoperable and usable cybersecurity mechanisms are essential to enabling the scientist's seamless use of cyberinfrastructure. Cybersecurity enables access to cyberinfrastructure for scientific collaborations, while protecting the underlying resources (sensitive data, powerful supercomputers, and unique instruments) from abuse. Scientists use security credentials (passwords, certificates, keys, and tokens) to log on to access these resources, and the usability of those credentials (how they are obtained, used, renewed, and recovered) and their interoperability (the ability to use them to access a variety of software and systems) are important factors for providing seamless access.

The goal of the "SciAuth" project is to improve the usability and interoperability of the security credentials that scientists use to access NSF cyberinfrastructure, thereby improving the productivity of the many scientific collaborations supported by NSF cyberinfrastructure. Our project does not propose a new credential mechanism for NSF cyberinfrastructure, but rather it provides community engagement, support for coordinated adoption of community standards, assistance with software integration, security analysis and threat modeling, training, and workforce development to enable improved interoperability and usability for security credentials across NSF cyberinfrastructure. We aim to help the community realize the benefits of an interoperable, capability-based ecosystem when transitioning between credential technologies. We have assembled a project team with decades of leadership in this area through our prior work with the CILogon [1,2], LIGO [3,4,5,6], OSG [7,8], SciTokens [9,10,11,12], TeraGrid [13,14,15], WLCG [16], and XSEDE [17,18,19] projects.

We propose this project at a critical time, when NSF cyberinfrastructure is undergoing a migration from using X.509 certificates for user authentication to using JSON Web Tokens (JWTs), a widely-adopted, IETF standard, interoperable, web-native credential format [20]. This migration began in 2018 when the Globus project ended its support for the Globus Toolkit and its underlying X.509-based Grid Security Infrastructure (GSI) [21]. That year, the NSF-funded SciTokens project demonstrated the feasibility of migrating from GSI certificates to JWTs, along with the security benefits of using JWTs to implement least-privilege, capability-based authorization in contrast to identity-based impersonation using GSI proxy certificates [22]. In 2019, the Worldwide LHC Computing Grid (WLCG) adopted the SciTokens model as part of the WLCG Common JWT Profiles [23]. In 2020, 11 projects in the Open Science Grid used JWTs with SciTokens software for over 115,000 scientific data transfers. Now, JWT pilot projects are underway in LIGO and XSEDE. OSG plans to retire GSI in 2022, and WLCG plans to begin removing X.509 user certificates from its infrastructure in 2023. Thus, now is a critical time for the SciAuth project to tie together and push forward these many community activities; SciAuth will make a major impact in ensuring these transitions have the necessary resources, training, and technology to complete successfully and on time.

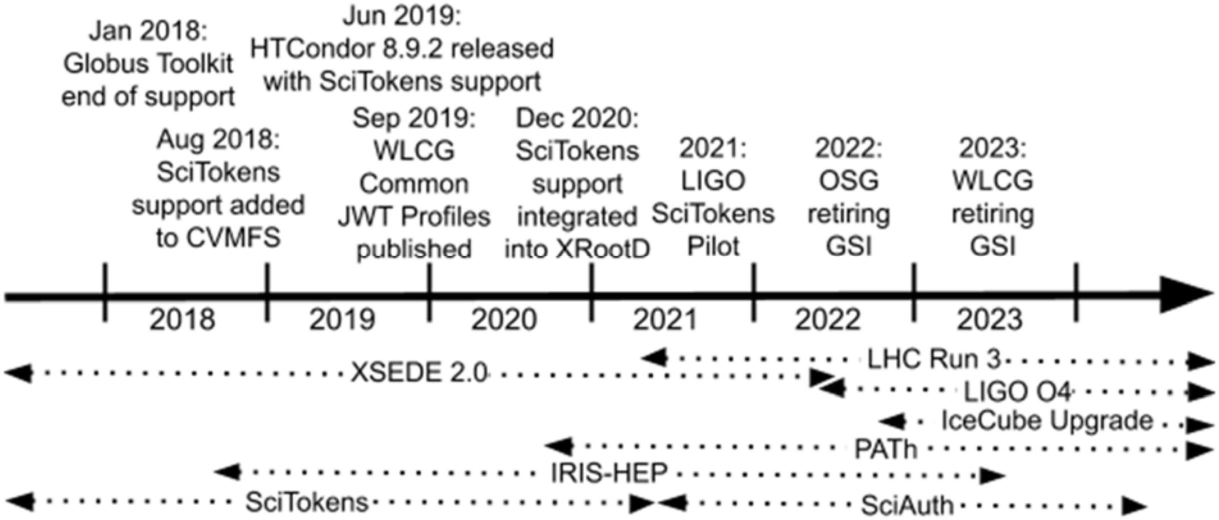

Jan 2018: Globus Toolkit end of support
Aug 2018: SciTokens support added to CVMFS
Jun 2019: HTCondor 8.9.2 released with SciTokens support
Sep 2019: WLCG Common JWT Profiles published
Dec 2020: SciTokens support integrated into XRootD
2021: LIGO SciTokens Pilot
2022: OSG retiring GSI
2023: WLCG retiring GSI
2018
2019
2020
2021
2022
2023
XSEDE 2.0
LHC Run 3
LIGO O4
IceCube Upgrade
PATh
IRIS-HEP
SciTokens
SciAuth

Figure 1. The migration from GSI to JWT credentials is proceeding over multiple years across multiple SciAuth project stakeholders and partner infrastructures.

Integrating a new security mechanism into production cyberinfrastructure takes years. Cyberinfrastructure providers must proceed with care to maintain interoperability across software components and resource providers and to provide stability for the many scientists that use the cyberinfrastructure each day. For example, OSG supports USATLAS and USCMS computing as part of WLCG, provides the CVMFS distributed filesystem for LIGO, and provides resources via XSEDE. In late 2020, OSG upgraded its CVMFS servers to support JWT authentication for LIGO's SciTokens pilot project, after months of coordinated effort across the projects by co-PIs Basney, Bockelman, Weitzel and others. While SciTokens has developed the software components and written the specifications to enable this ongoing transition, the need for leadership and coordination persists beyond the June 2021 end date of the SciTokens project. As illustrated in Figure 1, we propose the SciAuth project to take on those leadership and coordination activities beginning in July 2021, while the underlying SciTokens JWT technology is developed and maintained by the broader community.

If the SciAuth project is not funded, NSF cyberinfrastructure projects will still proceed independently with the migration from GSI to JWT credentials. However, it will be a lost opportunity for our ecosystem as there is a significant risk of a reimplementation of GSI instead of improving the conceptual foundations. Without SciAuth, we expect the transition will be uncoordinated and slower, and the risks of interoperability, security, and usability issues will be increased without the leadership that the SciAuth project can provide.

Intellectual Merit

The SciAuth project advances intellectual knowledge through adoption of Internet standards for interoperable, least-privilege authorization across scientific cyberinfrastructure, in partnership with science projects (CMS, IceCube, LIGO, WLCG) and cyberinfrastructure providers (Fermilab, OSG, PATh, IRIS-HEP, XSEDE), as illustrated in Figure 2. While JWTs are widely adopted for authorization for web applications, scientific cyberinfrastructure has unique requirements for interoperability and scalability that introduce novel JWT deployment scenarios. JWTs are most often used through a browser interface, which is extremely feature-rich compared to the terminals researchers work in, and there are significant usability challenges in navigating the browser / terminal divide. Scientific workflows utilize resources that are geographically distributed and operated by multiple providers with diverse security policies, requiring multiple token issuers that follow common profiles for token contents and distributed token verification. Unlike opaque tokens that are issued and verified by a single token server, JWTs support extensible token contents containing attributes about the issuer, the token subject, the strength of authentication, and specific authorizations for the token holder. Common profiles, such as the WLCG Common JWT Profiles, enable tokens to be used with common software implementations across distributed cyberinfrastructures, such as WLCG computing across the European Grid Infrastructure (EGI) and the US Open Science Grid (OSG).

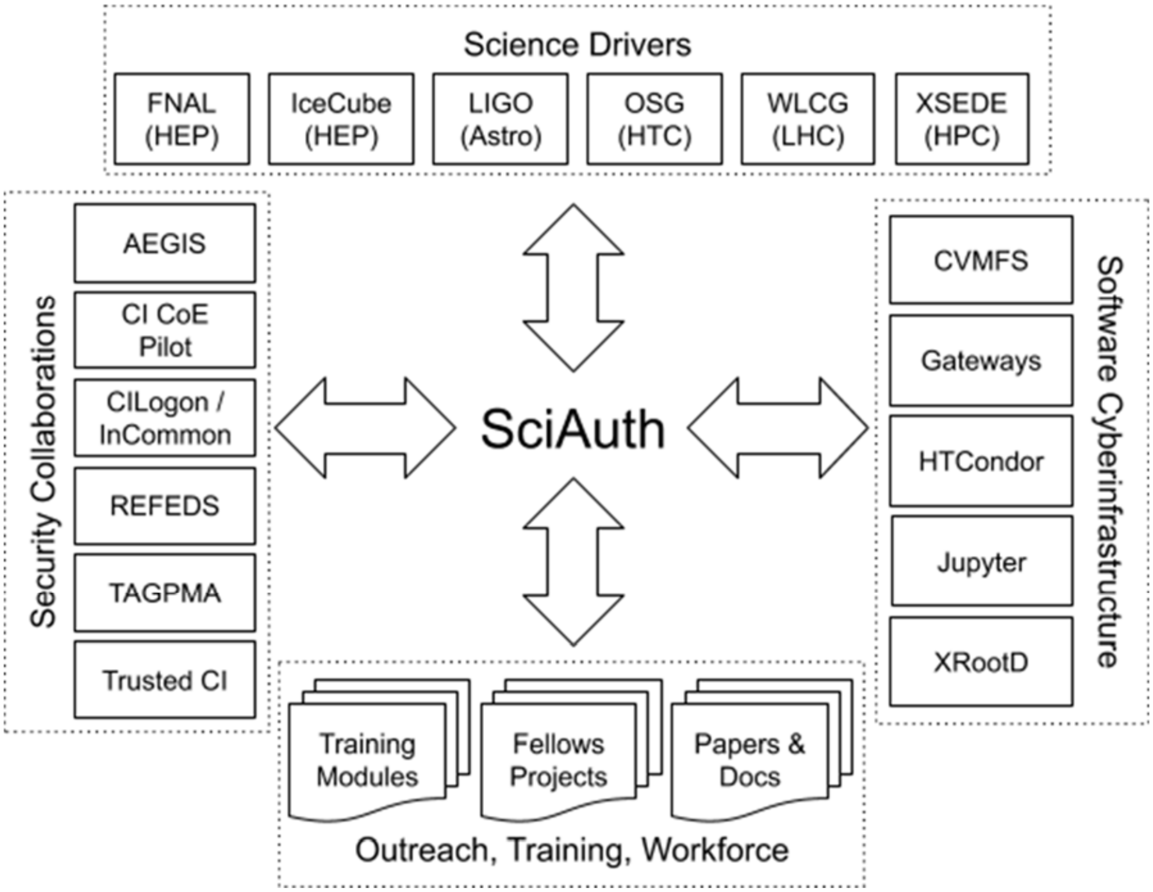


Figure 2. The SciAuth approach involves collaborations with science projects, software developers, and security groups along with outreach, training, and workforce development.

Federated identity plays an important role in enabling authorization across cyberinfrastructure. In our approach, scientists obtain authorization tokens by first authenticating with their campus identity provider (IdP), which is registered in the US InCommon federation [24] or the corresponding academic federation in the scientist's home country, interfederated through eduGAIN [25]. The campus IdP follows standards from the Research and Education FEDerations (REFEDS) group [26] to convey strength of authentication (e.g., to indicate the use of multi-factor authentication), strength of identity vetting (via standard levels of assurance), and attributes about the scientist's affiliations in an assertion to the token issuer. Based on these attributes and the policies set by the research collaboration, the token issuer provides JWTs to enable the scientist's access to cyberinfrastructure according to provider policies (e.g., requiring multi-factor authentication). The REFEDS Security Incident Response Trust Framework for Federated Identity (SIRTFI) [27] provides operational standards for handling security incidents in this distributed environment, including procedures for token revocation. Scientific cyberinfrastructure pushes the boundaries of large-scale federated identity and access management, and the SciAuth project enables innovation in NSF cyberinfrastructure security through adoption of new security credential standards in this collaborative, federated environment. Support for federated identity enables NSF cyberinfrastructure to benefit from campus adoption of multi-factor authentication, passwordless authentication (e.g., WebAuthn [28]), and other strong authentication methods, so science collaborations can focus on their unique authorization needs.

The transition from GSI credentials to JWTs improves the usability and security of the infrastructure for research and education, enabling research collaborations large and small to access computing resources. Usability is improved because JWTs are a lightweight standard format that is widely supported by current software languages, libraries, and frameworks, in contrast to GSI's X.509 proxy certificates, for which software support is ending or has already ended. Also, JWTs are managed by web-native standard protocols including OpenID Connect (OIDC) [29] and OAuth 2.0 [30], providing ease of token issuance, refresh, verification, and revocation. Security is improved through least-privilege authorization using the JWT "scope" claim [31], which specifies the capabilities for which the token holder is authorized (e.g., "storage.read:/protected/data"), in contrast to the identity-based authorization approach of GSI. These benefits have motivated projects including LIGO, OSG, and WLCG to begin their migration from X.509 proxy certificates to JWTs.

Broader Impacts

The SciAuth project benefits society by supporting an enhanced infrastructure for research and education and by supporting the development of a diverse, globally competitive STEM workforce through training and outreach activities, including a student fellows program, described below. The project's broader outreach activities also include presentations at practitioner conferences (PEARC, Gateways, JupyterCon, OSG All Hands) to increase awareness of project deliverables and adoption of community guidance.

The SciAuth fellows program, which is modeled on the IRIS-HEP fellows program [32], pairs students across the country with mentors to collaborate on student-led projects on the topic of cyberinfrastructure security. The project plans for 6 student fellows per year (18 total over the 3 year project duration), recruited nationally through organizations such as Women in CyberSecurity (WiCyS), the Minority Serving Cyberinfrastructure Consortium (MS-CC), and the CyberCorps: Scholarship for Service program, with an emphasis on recruiting participants from under-represented communities. The national recruitment efforts allow SciAuth to tap a broader pool of candidates than those at the SciAuth institutions and increase the likelihood of SciAuth broadening participation in computing. Each fellow proposes a research project (either selected from a list of examples published on the SciAuth web site or proposed by the fellow) related to the topic of authorization for scientific collaborations. Each fellow pursues their research project over a 12 week period with one of the project co-PIs (Basney, Bockelman, or Weitzel) assigned as a mentor. Fellows are selected via a competitive application process according to the strength of the proposed project, the academic preparation of the student, and the diversity of the applicants. Fellows will also meet regularly as a group to share their research progress and results. Fellows each receive a $1,000 stipend to support their research.

The SciAuth project also develops online training materials on the use of JWTs, OAuth, and OIDC with cyberinfrastructure. We will publish our training modules as Jupyter Notebooks using The Whole Tale (NSF DIBBS [33]) infrastructure for asynchronous learning and also provide synchronous training at events such as the OSG Summer School. Our training modules will also include videos and GitHub repositories following the Carpentries model. Co-PI Weitzel is a regular OSG Summer School and Carpentries instructor. Our training materials will supplement existing materials on use of JWTs in common programming languages and application frameworks by covering unique aspects of using JWTs with NSF cyberinfrastructure (i.e., integration with scientific applications, common cyberinfrastructure security policies, and science-specific JWT profiles such as the WLCG JWT Profiles). To support our training and other outreach efforts, the project will provide demonstration Docker containers, including a lightweight JWT issuer to enable developers to easily obtain tokens for testing purposes.

Science Drivers

The SciAuth project is motivated by the needs of the science communities who are migrating from X.509 user certificates to JWTs. In this section, we describe our partner science communities, their specific needs, and how we plan to help. Letters of Collaboration are also attached to this proposal. To facilitate sustainability of our science community collaborations, we contribute all documentation, software, and training deliverables to the relevant community maintainers, so their impact continues beyond the life of the SciAuth project.

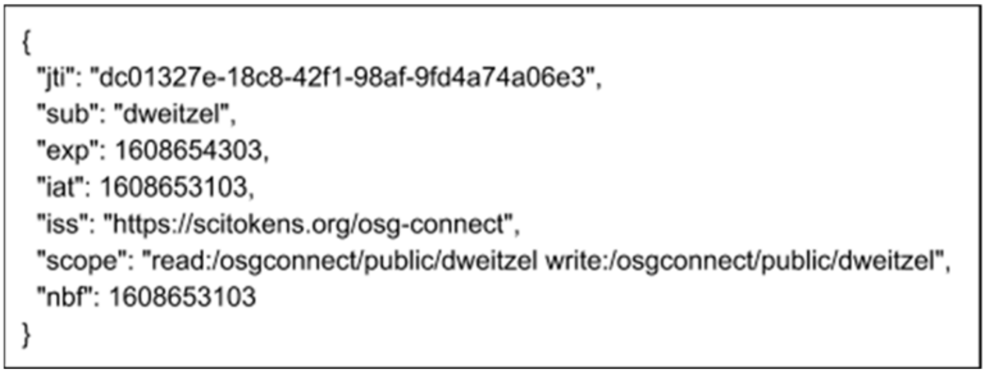
The NSF-funded Laser Interferometer Gravitational-Wave Observatory (LIGO) and associated LIGO Scientific Collaboration (LSC) are focused on the study of gravitational waves to explore the fundamental physics of gravity and gravitational wave science as a tool of astronomical discovery. LIGO was an original partner in the SciTokens project, and LIGO members are currently working with the co-PIs (Basney, Bockelman, and Weitzel) on a pilot project to adopt the SciTokens technology. Unique aspects of the LIGO deployment include: 1) use of

SciTokens for access to nonpublic scientific data (instrument frame files) in CVMFS (hosted by OSG), 2) use of HTCondor issued JWTs for authorized access to data, and 3) integration with federated identity via InCommon using group-based authorization using Internet2's COmanage and Grouper. LIGO's international collaborations with Virgo in Europe and KAGRA in Asia, along with a planned LIGO-India observatory, motivate the global adoption of federated identity and access management standards for LIGO's cyberinfrastructure. The current LIGO pilot includes a token issuer that authenticates scientists via InCommon, verifies authorizations in the LIGO LDAP directory (managed by COmanage and Grouper), and issues JWTs for authorized access to instrument frame files. While this pilot effort has demonstrated the functionality, significant effort is still required to address usability (e.g., provide needed command-line utilities, web portal interfaces, and documentation), interoperability (ensuring compatibility with OSG and other partners), and security (participating in risk assessment, threat modeling, and tabletop exercise activities), in collaboration with the SciAuth project. (See letter of collaboration from Stuart Anderson and Peter Couvares.)

High Energy Physics (HEP) provides a critical set of science communities for SciAuth, given their need for large-scale computing resources (billions of CPU hours per-year) and usage of distributed resources. Given their expense and scale, HEP experiments often organize into multi-institutional, global collaborations and want to leverage the computing resources available to each institution or country; this leads to a significant historical investment into federated infrastructure, especially for security. The largest umbrella infrastructure in HEP is the Worldwide LHC Computing Grid (WLCG), which coordinates the computing infrastructure for the four LHC experiments (and, more recently, other similarly-sized HEP collaborations outside the LHC). WLCG fills an important role of community organizer and effectively sets standards and best practices across many global infrastructures. The co-PIs have a long history of participation in WLCG working groups, including the Authorization Working Group which in 2019 produced the WLCG Common Token Profile. In SciAuth, we will partner with the WLCG on interoperability across infrastructures and coordinate timelines. Specific HEP science drivers will be CMS and IceCube. CMS operates a large, complex computing infrastructure, and we will help perform integrations between their CI and token technologies. CMS not only has many components to convert to token-based authorization, but it also is a heavy user of HTCondor, meaning that improvements done for CMS will have impact elsewhere. Due to the breadth of its infrastructure, CMS will serve as a case study for SciAuth in managing the transition to tokens. Similarly, SciAuth will partner with another experiment, IceCube, which has begun a technical transition to tokens [34]; SciAuth will work to ensure their tokens are interoperable across their computing resources and integrated with their HTCondor-based workflow system. (See letters of collaboration from Simone Campana, James Letts, and Benedikt Riedel.)

Another example from HEP is our planned collaborations with IRIS-HEP (an NSF-funded software institute) and CERN IT. Both have significant interest in helping the WLCG community make a timely transition to a token-based infrastructure. IRIS-HEP's ‘OSG-LHC’ area provides a software stack for the Open Science Grid and helps address the unique computing needs of the U.S. LHC community; this software and infrastructure were early adopters of using tokens for bulk data transfer between sites and have been promoting the use of tokens and HTTP for bulk transfers. Similarly, CERN IT develops a large-scale storage software service, EOS, that is used to manage hundreds of petabytes at CERN and elsewhere. Both IRIS-HEP and CERN IT rely on the common SciTokens library to implement token authorization and represent important science communities we will work with. SciAuth will work with them to identify additional requirements and features as part of their transition. (See letter of collaboration from Peter Elmer and Andreas-Joachim Peters.)

Our collaboration with Fermilab provides a valuable connection to Department of Energy science activities with a unique set of use cases and an aggressive deployment schedule. Fermilab is a CMS Tier-1 computing center, is the host laboratory for the Deep Underground Neutrino Experiment (DUNE), and is host to multiple smaller HEP collaborations that compute across OSG. Thus, OSG and WLCG interoperability are an important concern for Fermilab. While Fermilab has supported an international scientific user community for many years, the DUNE collaboration introduces a new scale of international use that is motivating Fermilab to pursue greater support for federated identity and access management. PI Basney is currently supporting a deployment effort at Fermilab to issue JWTs for HPC job submission and data access based on InCommon authentication and associated authorizations from the Fermilab user database (FERRY). Fermilab's need to support multiple smaller HEP collaborations is providing valuable input to WLCG JWT Profile discussions. For example, while it makes sense to deploy dedicated token issuers for the large ATLAS and CMS collaborations, it is more feasible to deploy a shared token issuer for Fermilab's smaller HEP collaborations, and we have been discussing modifications to the profiles in the WLCG Authorization Working Group to support the shared token issuer use case more effectively. (See letter of collaboration from Joseph Lykken.)



```
{
  "jti": "dc01327e-18c8-42f1-98af-9fd4a74a06e3",
  "sub": "dweitzel",
  "exp": 1608654303,
  "iat": 1608653103,
  "iss": "https://scitokens.org/osg-connect",
  "scope": "read:/osgconnect/public/dweitzel write:/osgconnect/public/dweitzel",
  "nbf": 1608653103
}
```

Figure 3. This example JWT from the OSG Connect service grants read and write access to the OSG Connect filesystem for HTCondor jobs.

The Open Science Grid (OSG) was an early adopter of SciTokens on their submission hosts. The OSG enables researchers in Bioinformatics, Chemistry, and many other disciplines to utilize distributed high throughput resources. The OSG serves as a testing ground for end users utilizing tokens. Feedback from users has improved the SciTokens XRootD integration and tooling for users to acquire tokens. In 2020 alone, OSG end-users have made 115,428 SciToken authenticated transfers transferring 23.3TB. (See letter of collaboration from Frank Würthwein.)

The science of cyberinfrastructure operations is also a driver for SciAuth. For example, cyberinfrastructure engineers use tokens to authenticate with OSG services, including access to Compute Entrypoints (CE) or when submit hosts are added to the OSG's Open Science Pool. A CE controls access to a resource’s compute resources, and its token has a special scope to access the resources. The pilot system used by OSG (glideinWMS [35]) authenticates with the CE using tokens generated by the organization requesting resources. The OSG also allows distributed submit hosts to utilize the OSG Open Science Pool, a collection of high throughput resources managed by OSG and made of opportunistic resources, allocated resources, and resources contributed by NSF CC* awardees. The submit hosts authenticate to the pool using tokens that are managed by OSG administrators. The token replaces the OSG’s shared password method for authenticating submit hosts, enabling OSG administrators to grant and revoke access to individual submit hosts without having to redistribute a shared pool password.

XSEDE is another valuable source of science drivers for SciAuth, working with XSEDE's Requirements Analysis and Capability Delivery (RACD) team on use cases and engineering plans for the migration to JWTs. The XSEDE Service Provider Forum provides stakeholders with diverse perspectives, representing HPC centers, academic cloud providers, and campus computing centers. XSEDE also facilitates our coordination with the AARC Engagement Group for Infrastructures (AEGIS) and The Americas Grid Policy Management Authority (TAGPMA), which we discuss below in our section on coordination via security collaboration groups. (See letter of collaboration from John Towns.)

Benefits to Scientific Applications and Users

The transition to JWTs for distributed authorization in NSF cyberinfrastructure benefits scientific applications and users by providing interoperable and secure access to the computing, data, instruments, and software that support scientific collaborations. Our project uses the following strategies to deliver these benefits: 1) enabling support for JWTs in common scientific software (CVMFS, HTCondor, Jupyter, pilot job frameworks, science gateways, XRootD) through developer training, documentation, and software integration, 2) participating in hackathons to achieve interoperability across common applications and platforms, 3) providing software demonstrations (e.g., a CMS demonstration using JWTs in a Jupyter notebook to launch an HTCondor simulation job and load results via XRootD), and 4) providing documentation and training directly to scientists on effective use of JWTs with NSF cyberinfrastructure.

Many of OSG’s technologies (such as the HTCondor-CE [36] and submit hosts mentioned above) are built on the HTCondor Software Suite (HTCSS) [37,38], a distributed computing project from UW-Madison that leverages High Throughput Computing (HTC). HTCSS, however, has a broader user base across science, industry, and other branches of government. The Partnership to Advance Throughput Computing (PATh) between the Center for High Throughput Computing (CHTC; the home of HTCSS) and the OSG provides support for HTCSS and has an explicit aim to support all of open science through the advancement of distributed HTC. HTCSS has the ability to authenticate tokens issued internally in the system; authenticate with externally-issued tokens such as used by the WLCG; and manage opaque tokens through OAuth 2.0, automatically renewing and acquiring tokens. However, work remains to support additional token acquisition workflows, better document integrating tokens with workflows, and provide training material for common identity providers. As SciAuth makes contributions to HTCSS in order to support specific science drivers, these improvements will be fed back into HTCSS to provide a wider impact for the SciAuth project and help sustain its project outputs. (See letter of collaboration from Miron Livny.)

An important token integration case is in data movement (particularly for the CMS and IceCube science drivers, which need to move data across infrastructures). As there are direct connections between sites that use different storage technologies, these transfers are where interoperability is most critical. SciAuth will continue to contribute to the reference C implementation for validation of JWTs using the SciTokens and WLCG Common Token Profile, the “scitokens-cpp” library (used by multiple storage technologies on the WLCG). We will partner with storage providers (such as CERN’s EOS team; see letter of collaboration from Andreas-Joachim-Peters) to ensure the needs of these core storage technologies are met.

One early adopter of token technologies on the distributed computing infrastructure was the CernVM File System (CVMFS [39]). CVMFS is a FUSE-based, distributed global filesystem used throughout the WLCG, OSG, and LIGO to distribute software, data, and container images. CVMFS achieves its scalability by distributing data as immutable objects over a multi-layered cache hierarchy; by having high hit ratios at each layer, it is able to support millions of I/O operations per second, and more than half a million cores throughout the globe use its repositories. As CVMFS utilizes HTTP(S) for several of its caching layers, it is a natural use case for authentication through tokens. The token is taken from the accessing process’s environment and used to authenticate cache access. CVMFS is used to distribute frame files for the LIGO use case described above and will serve as a SciAuth demonstration and example for token integrations with other HTTPS-based services.

The Jupyter OAuthenticator [40] provides support for OAuth tokens in Jupyter notebooks, enabling users to authenticate to their notebooks and use tokens inside the notebook to access resources. Using the CILogon OAuthenticator module, notebooks can support InCommon federated authentication and obtain JWTs for use with HTCondor Python APIs and scientific filesystems like XRootD. The CILogon OAuthenticator module was originally developed for the LSST project [41]. While the use of OAuthenticator for basic authentication is well documented, the subsequent use of tokens for access to cyberinfrastructure resources inside the notebook is not. The SciAuth project will provide documentation and demonstration examples for this important use case.

Like Jupyter notebooks, science gateways provide web interfaces to cyberinfrastructure, where scientists authenticate to the gateway and then the gateway accesses resources on the scientist's behalf [42,43]. Many science gateways support OAuth tokens [44,45], so the goal of our engagement with the science gateway community (via the Gateways conference and online forums) will be to ensure that JWT profiles adopted by cyberinfrastructure resources providers are compatible with science gateway token management functionality.

While HTCondor is used to execute workloads for science collaborations like IceCube and CMS and the diverse set of end-users of the OSG, it also works in tandem with another component, GlideinWMS [46]. GlideinWMS instances interact with the wide set of resource providers on the OSG to request resources and start and configure HTCondor worker nodes (which is analogous to a component that starts VMs containing HTCondor worker nodes on multiple cloud infrastructures). GlideinWMS has historically only worked with X.509 certificates and is now adopting token authentication for the interaction between the central instances and sites as well as the HTCondor worker nodes and central managers. As SciAuth helps science drivers transition to tokens, we will work with this software provider to complete their transition and ensure any new HTCondor features are well coordinated with this underlying software.

Coordination via Security Collaboration Groups

The SciAuth project will work with multiple security collaboration groups to coordinate the adoption of JWTs across cyberinfrastructures.

The AARC Engagement Group for Infrastructures (AEGIS) [47] coordinates adoption of common identity and access management approaches across research infrastructures, including DARIAH-EU, EGI, ELIXIR, EUDAT, PRACE, WLCG, and XSEDE. PI Basney is a representative to AEGIS for XSEDE, and co-PI Bockelman is a representative to AEGIS for WLCG. The Authentication and Authorisation for Research and Collaboration (AARC) Blueprint Architecture [48] is a founding document for AEGIS, and this architecture documents a common cyberinfrastructure pattern of supporting federated authentication via InCommon and eduGAIN, connected to token issuers that use science community attributes to issue authorization tokens to scientific applications and services. As of December 2020, AEGIS has begun discussing JWT profiles for wider adoption across infrastructures in compliance with the Blueprint Architecture, making AEGIS an important collaboration group for SciAuth.

The Americas Grid Policy Management Authority (TAGPMA) [49] fosters cross-domain trust relationships for scientific computing environments across North, Central, and South America, in partnership with EUGridPMA and APGridPMA in Europe and Asia Pacific, respectively. TAGPMA chair Derek Simmel (see letter of collaboration) recently launched a series of workshops on token-based authentication and authorization for cyberinfrastructure. The first workshop [50], in December 2020, included sessions from WLCG, Globus, LIGO, XSEDE, and Fermilab, with presentations by co-PIs Basney, Bockelman, and Weitzel (who contributed to 4 out of 5 workshop sessions). This ongoing series of workshops provides a forum for discussing use cases, technical interoperability, and trust relationships. In particular, TAGPMA has a robust peer-review process for operators of credential services that will be critical for building trust as the community deploys JWT issuing services.

The Identity and Access Management (IAM) working group that is co-organized by the CI CoE Pilot [51] and Trusted CI [52] (see letters of collaboration from Ewa Deelman and Von Welch) provides a forum for sharing IAM practices across the NSF Major Facilities. This working group, along with Trusted CI's Large Facility Security Team (LFST), will be valuable outreach targets for the SciAuth project.

Lastly, given the importance of identity federations to enabling researcher access to cyberinfrastructure using their campus credentials, the US InCommon federation and the international Research and Education FEDerations (REFEDS) group are key venues for coordination and outreach for our project. For example, adoption of JWT identity tokens via OpenID Connect Federation [53] is an active area of work in InCommon and REFEDS that is relevant to our SciAuth goal of JWT adoption. Active participation in InCommon and REFEDS meetings and working groups will enable SciAuth project members to coordinate with campus IAM experts. PI Basney is already active in these groups through his work with CILogon.

Standard JWT Profiles

For the use of tokens, a key challenge in the distributed computing space versus significant web-based usage is the fact that the token issuer may be a distinct entity from the relying party (e.g., a web service that exposes an API authenticated via a token). This requires the token to have a standardized structure and an agreed-upon interpretation of the contents. In the web space, a number of the concerns are addressed by the OpenID Connect (OIDC) profile [29]; if a website wants to know a contact email in an identity token issued by Google, OIDC is the standard stating this should go into the claim named “email” and not “e-mail” or “emailAddress”.

SciTokens published the first token profile for capability-based authorization [54] and distributed verification [55]; this was taken as a starting point for the WLCG Authorization Working Group which published a more formal profile, the WLCG Common JWT Profile [23]. The WLCG profile provides for more fine-grained capabilities and authorization use cases as well as a complimentary group-based authorization mode.

During the standardization process, we took special care to ensure a single relying party could support both WLCG and SciTokens profiles; this was achievable because both profiles used the same underlying conceptual framework and tokens indicate the profile used. In SciAuth, we want to continue a process of “harmonization” with an end-goal of either merging the two efforts together or minimizing the differences, maximizing the interoperability between infrastructures.

Perhaps equally as important as a shared specification is the testing infrastructure to ensure the implementations are correct. With the WLCG, we will organize the testing of multiple implementations and deployments, help improve existing compliance suites to ensure interoperability, and continue the tradition of organizing “hackathons” to make targeted improvements to areas of the profile.

Results from Prior NSF Support: SciTokens

Basney is PI, Bockelman is co-PI, and Weitzel is Nebraska subaward PI of **NSF award #1738962**. **Budget:** $1,000,000. **Period:** July 2017 through June 2021. **Title:** "CICI: CE:

SciTokens: Capability-Based Secure Access to Remote Scientific Data." **Intellectual Merit:** SciTokens advances knowledge through use of distributed, least-privilege authorization in NSF cyberinfrastructure. Adoption of current Internet standards (i.e., JWT and OAuth) enables knowledge transfer of Internet security methods across sectors. **Broader Impacts:** The SciTokens software enables scientists to perform widely distributed computational science more reliably and securely. Integration with the widely-used HTCondor software and collaboration with OSG, LIGO, LSST, and XSEDE facilitates adoption by the wider scientific community. **Publications:** [9,10,11,12] **Research Products:** SciTokens open source software is publicly available on GitHub [56,57,58,59,60].

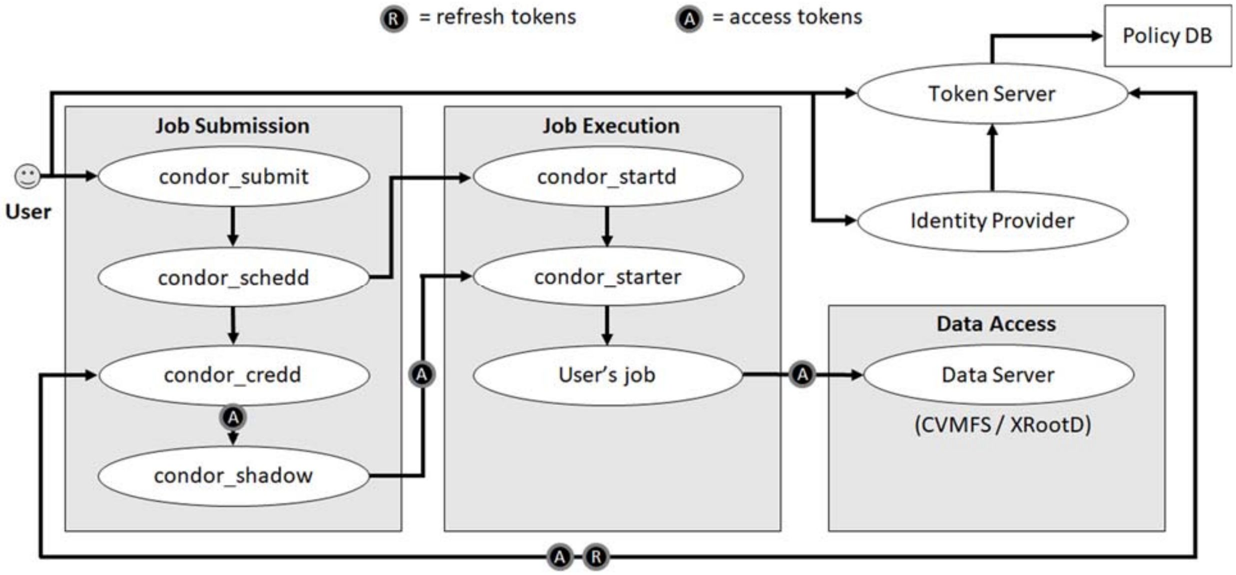


Figure 4. In this example of the SciTokens model, the user authenticates to a token server to issue tokens to HTCondor, which delegates the tokens to the user's job for data access.

The core technical innovation of the SciTokens project is the application of the JWT and OAuth standards to the needs of scientific cyberinfrastructure, where widely-distributed computing, data, instruments, and software services are harnessed for scientific workflows, requiring an authorization mechanism that itself is distributed (multiple credential issuers, distributed credential verification) and supports a wide range of security policies (credential strength, identity vetting, role-based and capability-based authorization). Typically, JWTs are used in a single web application, with a single token issuer and verifier. Also, typical OAuth deployment scenarios support only one or a few token issuers (e.g., Google, Globus), using opaque tokens that must be validated by a callback to the corresponding issuer. In contrast, SciTokens supports many token issuers, with signing keys, policies, and endpoint URLs published via OAuth 2.0 Authorization Server Metadata [61], using self-describing JWTs rather than opaque tokens, so the tokens can be independently verified by distributed services without requiring a callback to the token issuer. The use of JWTs with OAuth 2.0 is now a draft profile of the IETF OAuth working group [62]. The token refresh capability of OAuth 2.0 enables long-lived scientific workflows, and OAuth 2.0 Token Exchange [31] enables workflow systems to reduce token privileges, effectively implementing least-privilege delegation across the cyberinfrastructure ecosystem.

The SciTokens project has demonstrated a better alternative to the deprecated Globus GSI, following principles of decentralization, least-privilege, open source, and open standards, resulting in wide adoption through direct software contributions by the project (e.g., in HTCondor, CVMFS, and XRootD as illustrated in Figure 4) as well as independent, interoperable implementations (e.g., by the dCache project [63]).

Related Work

Since the Globus project announced the end of support for the open source Globus Toolkit, members of the broader community (including EGI, OSG, and XSEDE) have been maintaining a fork called the Grid Community Toolkit (GCT) [64] to provide critical security updates while cyberinfrastructure providers and science projects plan to migrate from the Globus GSI. While the GCT has provided an essential stop-gap, it is not a long-term solution for the scientific community. OSG has announced that it will stop supporting GCT in 2022 as part of its migration to token-based authorization. Likewise, the XSEDE project (with its 11th year extension) is scheduled to end in 2022. Thus, the need for SciAuth project leadership will be critical over this period to assist with the community transition away from GSI credentials.

The Globus project recommends that the community migrate from GSI to Globus Auth [65]. In contrast to our open source JWT-based approach, Globus Auth uses OAuth with opaque tokens, requiring callbacks to the proprietary Globus Auth service for token verification. However, we believe that an open ecosystem of token issuers, with support for distributed token verification, is a better fit for NSF cyberinfrastructure and the broader scientific community.

We have verified Interoperability through recent hackathons for multiple independent implementations of the WLCG Common JWT Profiles. In particular, the Indigo IAM Service [66], used by CERN, provides tokens that interoperate with SciTokens software. Also, CILogon (a project led by PI Basney) now supports issuance of SciTokens and WLCG tokens in its open source, non-profit, subscription-supported IAM-as-a-Service offering for research collaborations. CILogon subscribers include Fermilab, LIGO, LSST, OSG, SCIMMA, and XSEDE.

Threat Model

Our work is guided by RFC 6819 (OAuth 2.0 Threat Model and Security Considerations) [67], which provides a comprehensive threat model for attacks on credentials and the resources they protect, and by RFC 8725 (JSON Web Token Best Current Practices) [68], which identifies threats and mitigations specific to JWTs. Figure 5 provides an Open Science Cyber Risk Profile [69,70] diagram illustrating credential threats in the context of access to scientific data. Key threats are credential exposure (because scientific computing involves use of widely distributed cyberinfrastructure with a variety of operators), granting of too much access (in case least-privilege delegation is not effectively adopted), and granting access to malicious clients (in case authorizations are improperly targeted or client credentials compromise enables malicious impersonation of trusted clients). SciAuth follows the SciTokens model using the JWT and OAuth mechanisms of restricted scope, audience, and lifetime, applied to the distributed science workflow use case, to address these threats. The SciAuth project plan includes publication of our detailed threat model for community consultation in the first quarter of the project.

The SciAuth project is also concerned with threats to JWT software implementations across NSF cyberinfrastructure. RFC 8725 notes the risk that "one kind of JWT can be confused for another" and strongly recommends defining JWT profiles for specific applications (in our case, cyberinfrastructure services). This concern motivates our desire to harmonize JWT profiles across SciTokens, WLCG, and others, to provide software developers clear guidance on secure processing of JWTs for scientific computing use cases. Our project plan also includes software security assessments, peer-reviews, and online training for software developers to address this concern.

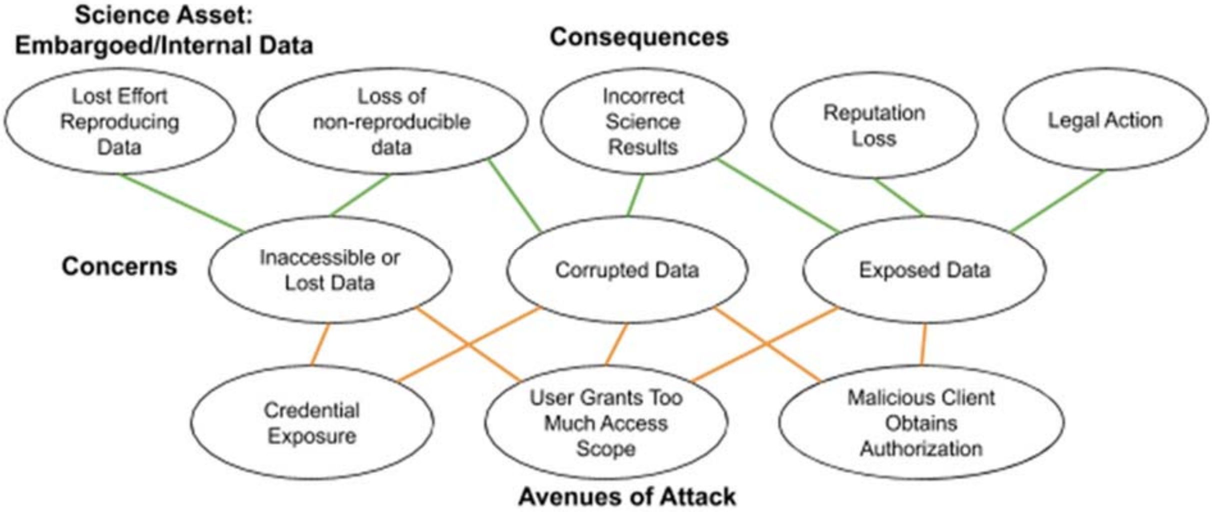

Science Asset:
Embargoed/Internal Data
Consequences
Lost Effort Reproducing Data
Loss of non-reproducible data
Incorrect Science Results
Reputation Loss
Legal Action
Concerns
Inaccessible or Lost Data
Corrupted Data
Exposed Data
Credential Exposure
User Grants Too Much Access Scope
Malicious Client Obtains Authorization
Avenues of Attack

Figure 5: This OSCRP diagram illustrates risks to scientific data due to credential attacks.

A robust incident response capability is also key to minimizing the impact of credential abuse. Effective coordination between (potentially widely distributed) credential issuers and relying parties is essential for containing a security incident. The REFEDS Security Incident Response Trust Framework for Federated Identity (SIRTFI) provides trusted contact points for incident coordination, guidance for logging security events to assist with incident investigations, and information sharing standards (e.g., use of the Traffic Light Protocol [71]). Our project plan includes incident response tabletop exercises to broaden community operational experience.

Sustainability Plan

The basis of our sustainability plan is open standards and open source implementations with broad community support. SciTokens has achieved sustained impact by contributing technology to long-lived projects and organizations, including CILogon, CVMFS, HTCondor, OSG, XRootD, XSEDE, and WLCG. Our experience confirms that support for JWTs in current programming languages and development platforms is a dramatic improvement over our two-decade long struggle with GSI proxy certificates and OpenSSL. With multiple interoperable open source implementations (CILogon, dCache, Indigo IAM, SciTokens), there is critical mass for sustainability of JWT mechanisms in the cyberinfrastructure community. As an example of our sustainability experience, we note that PI Basney's MyProxy credential management software was cited as a "clear success story for sustained software" in the report from the 2009 NSF-funded Workshop on Cyberinfrastructure Software Sustainability and Reusability [72].

The SciAuth project will contribute all deliverables to the appropriate project (e.g., Jupyter OAuthenticator) or community organization (e.g., AEGIS) for long-term sustainability. Any software that we produce will adopt the open source license of the home project, with the Apache 2.0 open source license as our default choice.

Solicitation Specific Review Criteria

In summary, the SciAuth project provides leadership and coordination for the NSF cyberinfrastructure community during a critical transition period, enabling improvements in interoperability and usability as the community migrates from X.509 user certificates to JWT authorization tokens. Our project is well aligned to the solicitation-specific review criteria:

- **Science-driven:** The SciAuth project is driven by the needs of research projects (CMS, IceCube, LIGO, WLCG), and cyberinfrastructure providers (Fermilab, OSG, PATh, IRIS-HEP, XSEDE). The project's broader impacts include workforce development through a fellows program that recruits students from under-represented communities.
- **Innovation:** The SciAuth project brings innovation to its target communities through the application of least-privilege, capability-based authorization to NSF cyberinfrastructure, a uniquely challenging target that brings together widely-distributed computing resources and services, in diverse operational environments.
- **Close collaborations among stakeholders:** The SciAuth project engages with cyberinfrastructure experts and domain scientists through community groups (AEGIS, CI CoE's IAM WG, InCommon, REFEDS, TAGPMA, Trusted CI's LFST) and practitioner conferences (PEARC, Gateways, JupyterCon, OSG All Hands Meeting).
- **Building on existing, recognized capabilities:** The SciAuth project builds on capabilities developed in the SciTokens project, being adopted now by research projects and cyberinfrastructure providers. By providing coordination and support for these projects' JWT adoption plans, the SciAuth project helps to more effectively leverage effort and investments across the projects to deliver an innovative and interoperable authorization ecosystem.
- **Project plans, and system and process architecture:** Our project plan (supplemental document) contains over 60 quarterly deliverables, with assigned leads, level of effort estimates, and associated metrics. The project's training modules include proof-of-concept Docker container demonstrations and Jupyter notebook tutorials to enable early adoption.
- **Sustained impact:** The SciAuth project integrates with ongoing activities in long-lived projects and organizations. The project's deliverables fill important gaps in planned community efforts, and these deliverables will provide benefits into the future as they are adopted by associated projects and community organizations.

References Cited

[1] Jim Basney, Heather Flanagan, Terry Fleury, Jeff Gaynor, Scott Koranda, and Benn Oshrin. CILogon: Enabling Federated Identity and Access Management for Scientific Collaborations. In Proceedings of the International Symposium on Grids and Clouds (ISGC), PoS(ISGC2019)031, 2019. https://doi.org/10.22323/1.351.0031

[2] Jim Basney, Terry Fleury, and Jeff Gaynor, "CILogon: A Federated X.509 Certification Authority for CyberInfrastructure Logon," Concurrency and Computation: Practice and Experience, Volume 26, Issue 13, pages 2225-2239, September 2014. https://doi.org/10.1002/cpe.3265

[3] Jim Basney and Scott Koranda, "A Study of Three Approaches to International Identity Federation for the LIGO Project," July 2013. https://hdl.handle.net/2022/16760

[4] Jim Basney and Scott Koranda, "InCommon Membership in eduGAIN: the LIGO Perspective," May 2013. https://hdl.handle.net/2022/16690

[5] Jim Basney, Scott Koranda, and Von Welch, "An Analysis of the Benefits and Risks to LIGO When Participating in Identity Federations," October 2011. https://dcc.ligo.org/LIGO-G1100964/public

[6] Weitzel D, Bockelman B, Brown D, Couvares P, Würthwein F, Fajardo E. Data Access for LIGO on the OSG. Proceedings of Practice and Experience in Advanced Research Computing (PEARC), July 2017. https://doi.org/10.1145/3093338.3093363

[7] Ruth Pordes, Don Petravick, Bill Kramer, Doug Olson, Miron Livny, Alain Roy, Paul Avery, Kent Blackburn, Torre Wenaus, and Frank Würthwein. (2007). "The Open Science Grid", J. Phys. Conf. Ser. 78, 012057. https://doi.org/10.1088/1742-6596/78/1/012057

[8] Altunay M, Avery P, Blackburn K, Bockelman B, Ernst M, Fraser D, Quick R, Gardner R, Goasguen S, Levshina T, Livny M, McGee J, Olson D, Pordes R, Potekhin M, Rana A, Roy A, Sehgal C, Sfiligoi I, Wurthwein F. A Science Driven Production Cyberinfrastructure—the Open Science Grid. Journal of Grid Computing, 2011. https://doi.org/10.1007/s10723-010-9176-6

[9] Alex Withers, Brian Bockelman, Derek Weitzel, Duncan A. Brown, Jeff Gaynor, Jim Basney, Todd Tannenbaum, Zach Miller, “SciTokens: Capability-Based Secure Access to Remote Scientific Data”, PEARC ‘18: Practice and Experience in Advanced Research Computing, July 2018, Pittsburgh, PA, USA. https://doi.org/10.1145/3219104.3219135

[10] You Alex Gao, Jim Basney, and Alex Withers. 2020. SciTokens SSH: Token-based Authentication for Remote Login to Scientific Computing Environments. In Practice and Experience in Advanced Research Computing (PEARC ’20), July 26-30, 2020, Portland, OR, USA. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3311790.3399613

[11] Alex Withers, Brian Bockelman, Derek Weitzel, Duncan Brown, Jason Patton, Jeff Gaynor, Jim Basney, Todd Tannenbaum, You Alex Gao, and Zach Miller. 2019. SciTokens: Demonstrating Capability-Based Access to Remote Scientific Data using HTCondor. In Practice and Experience in Advanced Research Computing (PEARC ‘19), July 28-August 1, 2019, Chicago, IL, USA. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3332186.3333258

[12] Derek Weitzel, Brian Bockelman, Jim Basney, Todd Tannenbaum, Zach Miller, and Jeff Gaynor. Capability-Based Authorization for HEP. In 23rd International Conference on

Computing in High Energy and Nuclear Physics (CHEP 2018), July 9-13, 2018, Sofia, Bulgaria. https://doi.org/10.1051/epjconf/201921404014

[13] Jim Basney and Jeff Gaynor, "An OAuth Service for Issuing Certificates to Science Gateways for TeraGrid Users," TeraGrid Conference, July 18-21, 2011, Salt Lake City, UT. https://doi.org/10.1145/2016741.2016776

[14] Jim Basney, Terry Fleury, and Von Welch, "Federated Login to TeraGrid," 9th Symposium on Identity and Trust on the Internet (IDtrust 2010), Gaithersburg, MD, April 2010. https://doi.org/10.1145/1750389.1750391

[15] Jim Basney, Stuart Martin, JP Navarro, Marlon Pierce, Tom Scavo, Leif Strand, Tom Uram, Nancy Wilkins-Diehr, Wenjun Wu, and Choonhan Youn, "The Problem Solving Environments of TeraGrid, Science Gateways, and the Intersection of the Two," Fourth IEEE International Conference on eScience, December 2008, pages 725-734. https://doi.org/10.1016/j.ijhcs.2005.04.017

[16] Paul Millar, Andrea Ceccanti, Fabrizio Furano, Dmitry Litvintsev, Brian Paul Bockelman, & Alessandra Forti. (2019, November). Third-party transfers in WLCG using HTTP. Presented at the 24th International Conference on Computing in High Energy & Nuclear Physics, Zenodo. http://doi.org/10.5281/zenodo.3599313

[17] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, Nancy Wilkins-Diehr, "XSEDE: Accelerating Scientific Discovery", Computing in Science & Engineering, vol.16, no. 5, pp. 62-74, Sept.-Oct. 2014, https://doi.org/10.1109/MCSE.2014.80

[18] Lee Liming, Jim Basney, John Paul Navarro, and Shava Smallen. 2020. Use Case Methodology in XSEDE System Integration. In Practice and Experience in Advanced Research Computing (PEARC ’20), July 26–30, 2020, Portland, OR, USA. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3311790.3399622

[19] Jim Basney, Rion Dooley, Jeff Gaynor, Thejaka Amila Kanewala, Suresh Marru, Marlon Pierce, and Joe Stubbs, "Integrating Science Gateways with XSEDE Security: A Survey of Credential Management Approaches," XSEDE Conference, July 2014, Atlanta, GA. https://doi.org/10.1145/2616498.2616559

[20] M. Jones, J. Bradley, and N. Sakimura. 2015. JSON Web Token (JWT). RFC 7519. https://doi.org/10.17487/RFC7519

[21] I. Foster. Support for open source Globus Toolkit will end as of January 2018. https://github.com/globus/globus-toolkit/blob/globus_6_branch/support-changes.md

[22] S. Tuecke, V. Welch, D. Engert, L. Pearlman, M. Thompson. 2004. Internet X.509 Public Key Infrastructure (PKI) Proxy Certificate Profile. RFC 3820. https://doi.org/10.17487/RFC3820

[23] Altunay, Mine; Bockelman, Brian; Ceccanti, Andrea; Cornwall, Linda; Crawford, Matt; Crooks, David; Dack, Thomas; Dykstra, David; Groep, David; Igoumenos, Ioannis; Jouvin, Michel; Keeble, Oliver; Kelsy, David; Lassnig, Mario; Liampotis, Nicolas; Litmaath, Maarten; McNab, Andrew; Millar, Paul; Sallé, Mischa; Short, Hannah; Teheran, Jeny; Wartel, Romain. WLCG Common JWT Profiles (Version 1.0). Zenodo. September 25, 2019. https://doi.org/10.5281/zenodo.3460258

[24] InCommon Federation. https://www.incommon.org/federation/

[25] eduGAIN interfederation service. https://edugain.org/

[26] REFEDS (the Research and Education FEDerations group). https://refeds.org/

[27] REFEDS. Security Incident Response Trust Framework for Federated Identity (SIRTFI) v1.0. 2016. https://refeds.org/sirtfi

[28] D. Balfanz, A. Czeskis, J. Hodges, J.C. Jones, M.B. Jones, A. Kumar, A. Liao, R. Lindemann, and E. Lundberg. Web Authentication: An API for accessing Public Key Credentials. W3C Recommendation, March 2019. https://www.w3.org/TR/webauthn/

[29] N. Sakimura, J. Bradley, M. Jones, B. de Medeiros and C. Mortimore, OpenID Connect Core 1.0, OpenID Foundation, November, 2014. https://openid.net/specs/openid-connect-core-1_0.html

[30] D. Hardt. 2012. The OAuth 2.0 Authorization Framework. RFC 6749. https://doi.org/10.17487/RFC6749

[31] M. Jones, A. Nadalin, B. Campbell, J. Bradley, and C. Mortimore. 2020. OAuth 2.0 Token Exchange. RFC 8693. https://doi.org/10.17487/RFC8693

[32] IRIS-HEP Fellows Program. https://iris-hep.org/fellows

[33] Adam Brinckman, Kyle Chard, Niall Gaffney, Mihael Hategan, Matthew B. Jones, Kacper Kowalik, Sivakumar Kulasekaran, Bertram Ludäscher, Bryce D. Mecum, Jarek Nabrzyski, Victoria Stodden, Ian J. Taylor, Matthew J. Turk, Kandace Turner. (2017). Computing environments for reproducibility: Capturing the ‘‘Whole Tale’’, Future Generation Computer Systems https://doi.org/10.1016/j.future.2017.12.029

[34] David Schultz. (2019, November). A New Authorization System for IceCube Applications. Presented at the 24th International Conference on Computing in High Energy & Nuclear Physics, Zenodo. http://doi.org/10.5281/zenodo.3599241

[35] Igor Sfiligoi, Daniel C. Bradley, Burt Holzman, Parag Mhashilkar, Sanjay Padhi, and Frank Wurthwein. 2009. The Pilot Way to Grid Resources Using glideinWMS. In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering - Volume 02 (CSIE '09). IEEE Computer Society, USA, 428–432. https://doi.org/10.1109/CSIE.2009.950

[36] Bockelman B, Livny M, Lin B, Prelz F. Principles, technologies, and time: The translational journey of the HTCondor-CE. Journal of computational science. 2020. https://doi.org/10.1016/j.jocs.2020.101213

[37] Douglas Thain, Todd Tannenbaum, and Miron Livny, "Distributed Computing in Practice: The Condor Experience" Concurrency and Computation: Practice and Experience, Vol. 17, No. 2-4, pages 323-356, February-April, 2005. https://doi.org/10.1002/cpe.938

[38] HTCondor Software Suite (HTCSS). https://research.cs.wisc.edu/htcondor/

[39] Weitzel D, Bockelman B, Dykstra D, Blomer J, Meusel R. Accessing Data Federations with CVMFS. Journal of Physics: Conference Series. 2017 October; 898:062044-. https://doi.org/10.1088/1742-6596/898/6/062044

[40] JupyterHub OAuthenticator. https://github.com/jupyterhub/oauthenticator

[41] Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST). https://lsst.org/

[42] Jim Basney, Von Welch, and Nancy Wilkins-Diehr, "TeraGrid Science Gateway AAAA Model: Implementation and Lessons Learned," TeraGrid Conference, August 2-5, 2010, Pittsburgh, PA. https://doi.org/10.1145/1838574.1838576

[43] Von Welch, Jim Barlow, James Basney, Doru Marcusiu, Nancy Wilkins-Diehr, "A AAAA

model to support science gateways with community accounts," Concurrency and Computation: Practice and Experience, Volume 19, Issue 6, March 2007. https://doi.org/10.1002/cpe.1081

[44] Isuru Ranawaka, Suresh Marru, Juleen Graham, Aarushi Bisht, Jim Basney, Terry Fleury, Jeff Gaynor, Dimuthu Wannipurage, Marcus Christie, Alexandru Mahmoud, Enis Afgan, and Marlon Pierce. 2020. Custos: Security Middleware for Science Gateways. In Practice and Experience in Advanced Research Computing (PEARC ’20), July 26–30, 2020, Portland, OR, USA. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3311790.3396635

[45] Jim Basney, Rion Dooley, Jeff Gaynor, Suresh Marru, and Marlon Pierce, "Distributed Web Security for Science Gateways," Gateway Computing Environments Workshop (GCE11), November 17, 2011, Seattle, WA. https://doi.org/10.1145/2110486.2110489

[46] Antonio Pérez-Calero Yzquierdo, Brian Paul Bockelman, Diego Davila Foyo, Kenyi Hurtado Anampa, Todor Trendafilov Ivanov, Farrukh Aftab Khan, Amjad Kotobi, Krista Larson, James Letts, Marco Mascheroni and David Mason for the CMS Collaboration. "Exploring GlideinWMS and HTCondor scalability frontiers for an expanding CMS Global Pool," 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018), EPJ Web of Conferences 214, 03002 (2019). https://doi.org/10.1051/epjconf/201921403002

[47] AARC Engagement Group for Infrastructures (AEGIS). https://aarc-community.org/about/aegis/

[48] Nicolas Liampotis (ed.), "AARC Blueprint Architecture 2019", AARC-G045, November 2019. https://doi.org/10.5281/zenodo.3672785

[49] The Americas Grid Policy Management Authority (TAGPMA). http://www.tagpma.org/

[50] TAGPMA Workshop on Token-Based Authentication and Authorization (WoTBAN&AZ 2020). December 2020. https://indico.rnp.br/event/33/

[51] Ewa Deelman, Anirban Mandal, Valerio Pascucci, Susan Sons, Jane Wyngaard, Charles Vardeman, Steve Petruzza, Ilya Baldin, Laura Christopherson, Ryan Mitchell, Loic Pottier, Mats Rynge, Erik Scott, Karan Vahi, Marina Kogan, Jasmine Mann, Tom Gulbransen, Daniel Allen, David Barlow, Santiago Bonarrigo, Chris Clark, Leslie Goldman, Tristan Goulden, Phil Harvey, David Hulsander, Steve Jacobs, Christine Laney, Ivan Lobo-Padilla, Jeremy Sampson, John Staarmann, and Steve Stone, "Cyberinfrastructure Center of Excellence Pilot: Connecting Large Facilities Cyberinfrastructure," 2019 15th International Conference on eScience (eScience), San Diego, CA, USA, 2019, pp. 449-457. https://doi.org/10.1109/eScience.2019.00058

[52] Andrew Adams, Kay Avila, Jim Basney, Dana Brunson, Robert Cowles, Jeannette Dopheide, Terry Fleury, Elisa Heymann, Florence Hudson, Craig Jackson, Ryan Kiser, Mark Krenz, Jim Marsteller, Barton P. Miller, Sean Piesert, Scott Russell, Susan Sons, Von Welch, and John Zage. 2019. Trusted CI Experiences in Cybersecurity and Service to Open Science. In PEARC'19: Practice and Experience in Advanced Research Computing, July 28-August 1, 2019, Chicago, IL, USA. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3332186.3340601

[53] R. Hedberg (Ed.), M. Jones, A. Solberg, S. Gulliksson, and J. Bradley. "OpenID Connect Federation 1.0 - draft 13", December 2020.

https://openid.net/specs/openid-connect-federation-1_0.html

[54] SciToken Claims and Scopes Language. https://scitokens.org/technical_docs/Claims

[55] SciToken Verification. https://scitokens.org/technical_docs/Verification

[56] Derek Weitzel and Brian Bockelman. (2020, September 22). SciTokens Python Library (Version v1.2.3). Zenodo. http://doi.org/10.5281/zenodo.4042163

[57] Jeff Gaynor, Jim Basney, and Venkat Yekkirala. (2020, August 10). SciTokens Java (Version v1.2.1). Zenodo. http://doi.org/10.5281/zenodo.3978461

[58] Derek Weitzel. (2018, March 22). SciTokens Nginx (Version v1.0). Zenodo. http://doi.org/10.5281/zenodo.1205539

[59] Derek Weitzel, Brian P Bockelman, and Matyas Selmeci. (2020, June 24). SciTokens CPP Library (Version v0.5.1). Zenodo. http://doi.org/10.5281/zenodo.3906550

[60] Brian Bockelman, Derek Weitzel, and Matyas Selmeci. (2020, August 6). SciTokens for XRootD (Version v1.2.2). Zenodo. http://doi.org/10.5281/zenodo.3974653

[61] M. Jones, N. Sakimura, and J. Bradley. 2018. OAuth 2.0 Authorization Server Metadata. RFC 8414. https://doi.org/10.17487/RFC8414

[62] V. Bertocci. 2020. JSON Web Token (JWT) Profile for OAuth 2.0 Access Tokens. https://datatracker.ietf.org/doc/draft-ietf-oauth-access-token-jwt/

[63] Tigran Mkrtchyan, Olufemi Adeyemi, Patrick Fuhrmann, Vincent Garonne, Dmitry Litvintsev, Paul Millar, Albert Rossi, Marina Sahakyan, Jürgen Starek, and Sibel Yasar. "dCache - storage for advanced scientific use cases and beyond," 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018), EPJ Web of Conferences 214, 04042 (2019). https://doi.org/10.1051/epjconf/201921404042

[64] Grid Community Toolkit (GCT). https://gridcf.org/

[65] S. Tuecke, R. Ananthakrishnan, K. Chard, M. Lidman, B. McCollam, S. Rosen, and I. Foster. 2016. Globus Auth: A research identity and access management platform. In 2016 IEEE 12th International Conference on e-Science (e-Science). 203–212. https://doi.org/10.1109/eScience.2016.7870901

[66] A. Ceccanti, M. Hardt, B. Wegh, A.P. Millar, M. Caberletti, E. Vianello, and S. Licehammer. 2017. "The INDIGO-Datacloud Authentication and Authorization Infrastructure." Journal of Physics: Conference Series, Volume 898. https://doi.org/10.1088/1742-6596/898/10/102016

[67] T. Lodderstedt (Ed.), M. McGloin, and P. Hunt. 2013. OAuth 2.0 Threat Model and Security Considerations. RFC 6819. https://doi.org/10.17487/RFC6819

[68] Y. Sheffer, D. Hardt, and M. Jones. 2020. JSON Web Token Best Current Practices. RFC 8725. https://doi.org/10.17487/RFC8725

[69] Peisert, Sean, Von Welch, Andrew Adams, RuthAnne Bevier, Michael Dopheide, Rich LeDuc, Pascal Meunier, Steve Schwab, and Karen Stocks. Open Science Cyber Risk Profile (OSCRP), Version 1.3. August 2020. https://hdl.handle.net/2022/21259

[70] S. Peisert and V. Welch, "The Open Science Cyber Risk Profile: The Rosetta Stone for Open Science and Cybersecurity," in IEEE Security & Privacy, vol. 15, no. 5, pp. 94-95, 2017. https://doi.org/10.1109/MSP.2017.3681058

[71] TRAFFIC LIGHT PROTOCOL (TLP): FIRST Standards Definitions and Usage Guidance — Version 1.0. https://www.first.org/tlp/

[72] Craig Stewart, Guy Almes, and Bradley Wheeler (eds.). 2010. Cyberinfrastructure

Software Sustainability and Reusability: Report from an NSF-funded workshop.
https://hdl.handle.net/2022/6701

Data Management Plan

Introduction

This document outlines the data management plan for the proposal "**CICI:UCSS:SciAuth: Deploying Interoperable and Usable Authorization Tokens to Enable Scientific Collaborations**", PI Basney.

The primary outputs of the SciAuth project are technical papers, presentations, training modules, and software contributions to external projects.

Roles and responsibilities

Dr. Basney (PI) has overall responsibility for management of project data. Dr. Basney will ensure that project participants at the University of Illinois properly implement this data management plan. Dr. Basney delegates responsibility to each subaward PI for management of project data stored at subawardee institutions. The co-PIs will be responsible to ensure the primary outputs collected at their institutions are transferred to the University of Illinois by the end of the project.

Types of data

The project manages the following types of data:

- **Technical reports** or papers and their raw data will be made available through their respective journals and corresponding institutional repositories. Preference will be given to open access journals.
- **Training modules** (including Jupyter Notebooks) will be published via The Whole Tale (NSF DIBBS) infrastructure (https://wholetale.org/).
- **Software** will be publicly accessible under Open Source licenses from commercial software hosting providers (e.g. GitHub).

Policies for access/sharing and appropriate protection/privacy

All data products will be public and freely available via the project web site.

Policies for re-use, re-distribution, and production of derivatives

Technical documents and training modules will be licensed under the Creative Commons 3.0 license (Attribution / Non-Commercial / Share-Alike)[1]. All software will be licensed under an

[1] http://creativecommons.org/licenses/by-nc-sa/3.0/us/

Open Source license. Where practical, we will use the Apache Software License 2.0[2]; where possible, the Contributions to larger code bases that are *incompatible* with these licenses will default to the preferred license of that code base. All papers resulting from this work will be copyrighted and licensed according to the journal where they are published.

Data storage and preservation of access

All reports, presentations, manuscripts, and other documents that record research outputs generated under this project will be deposited in IDEALS (https://www.ideals.illinois.edu/), the Illinois Digital Environment for Access to Learning and Scholarship. All datasets and accompanying documentation generated under this project will be deposited in the Illinois Data Bank (https://databank.illinois.edu/), the file-based repository for research data at the University of Illinois at Urbana-Champaign. Both repositories are optimized for their respective content types and support robust indexing and stable access.

[2] http://www.apache.org/licenses/LICENSE-2.0.html

Project Plan

The overall goal of the project is to improve the usability and interoperability of the security credentials that scientists use to access NSF cyberinfrastructure, thereby improving the productivity of the many scientific collaborations supported by NSF cyberinfrastructure. The project undertakes the following activities in pursuit of this goal:

- **Outreach**: Community engagement through security collaboration groups (AEGIS, REFEDS, TAGPMA, Trusted CI) and scientific practitioner conferences (PEARC, Gateways, JupyterCon, OSG All Hands)
- **Standards**: Support for coordinated adoption of community standards, in partnership with science projects (CMS, LIGO, WLCG) and cyberinfrastructure providers (Fermilab, OSG, PATh, IRIS-HEP, XSEDE)
- **Software**: Assistance with software integration, enabling support for JWTs in common scientific software (CVMFS, HTCondor, Jupyter, pilot job frameworks, science gateways, XRootD), contributed back to the home project
- **Security**: Collaborative security assessments, tabletop exercises, and threat modeling to build confidence in and experience with the operational security of the JWT mechanism
- **Training**: Providing synchronous and asynchronous training modules on the use of JWTs, OAuth, and OIDC with NSF cyberinfrastructure
- **Fellows**: A workforce development program that pairs students across the with mentors from the project to collaborate on student-led projects on the topic of cyberinfrastructure security

PI Basney, along with co-PIs Bockelman and Weitzel, will collaborate to achieve project goals according to the schedule of project milestones and deliverables outlined below.

Project Organization

The project coordinates progress via a weekly all hands call. The team meets for longer-term planning and evaluation at an annual day-long all hands meeting. PI Basney is responsible for overall project management and - as necessary - budget reallocation, with co-PIs Bockelman and Weitzel responsible for work at Morgridge and Nebraska (respectively). The project uses Google Drive for document sharing and GitHub for revision control. Internal and public email lists are hosted by Google Groups, and the project web site is hosted by GitHub Sites.

Metrics

The following 4 project success metrics follow from our project goal to improve the usability and interoperability of the security credentials that scientists use to access NSF cyberinfrastructure:

1. **Metric:** Number of user communities that have adopted authorization tokens, tracked as both number of NSF projects (awardees) and non-NSF projects.
   Target: 25 by end of Year 2.
   Explanation: We track adoption as our key usability metric, based on our experience that scientists will adopt security mechanisms that enable their productivity and reject security mechanisms that get in the way. We specifically target the 20+ NSF Major Facilities for adoption through outreach via the CI CoE Pilot's IAM working group and Trusted CI's Large Facility Security Team and via direct partnerships (e.g., CMS and LIGO).

2. **Metric:** Percentage of infrastructures that have migrated from X.509 user certificates to authorization tokens.
   Target: 100% by end of Year 3.
   Explanation: Our project aims to enable infrastructures currently using X.509 user certificates to successfully complete their planned migrations to authorization tokens, on schedule and with minimal disruption to scientific productivity. Our target infrastructure list is CMS, LIGO, OSG, XSEDE, and WLCG, but we will also include any additional TAGPMA relying parties and Grid Community Toolkit user communities in this metric that we identify via outreach.

3. **Metric:** Number of interoperable implementations of authorization tokens in NSF software cyberinfrastructure.
   Target: 24 by end of Year 2.
   Explanation: Our training modules and other outreach activities will enable developers to integrate authorization tokens into their implementations and enable cyberinfrastructure operators to deploy them. Our periodic hackathons will enable interoperability. Our target of 24 gives us an aggressive goal, building on the 6 interoperable implementations currently known (CVMFS, dCache, HTCondor, Indigo IAM, SciTokens, and XRootd).

4. **Metric:** Number of students impacted via our fellows program, our training modules, and our other outreach activities.
   Target: 36 each year.
   Explanation: Our broader impacts are focused on workforce development. In addition to 6 fellows each year, we expect at least another 30 students to attend synchronous training sessions and/or use our asynchronous training modules each year. In addition to the number of students, we will track student demographics to measure our effectiveness at reaching a diverse group of students.

Milestones and Deliverables

The following table provides quarterly deliverables for each of the above activity areas, with the addition of a Management area for project meetings and NSF reports. For each deliverable, we identify the site lead, an effort estimate (in person weeks), and associated metrics and Key Performance Indicators (KPIs). Effort estimates match budgeted project effort across the 3 project sites, confirming the feasibility of the project plan.

Quarter	Area	Lead	Effort	Deliverable	Metrics/KPIs
Y1Q1	Fellows	Illinois	1	Select Y1 Fellows	# of applications received, applicant diversity
Y1Q1	Security	Illinois	1	Publish threat model	feedback received
Y1Q1	Software	Nebraska	2	Release Docker container for lightweight token issuer	downloads, bug reports, mailing list posts
Y1Q1	Standards	Morgridge	2	Convene working group on alignment of SciTokens and WLCG profiles	working group membership, representation
Y1Q1	Management	Illinois	0.2	Launch project website, email lists, etc.	traffic, updates
Y1Q1	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y1Q2	Training	Nebraska	2	Develop online training module: JWT Basics for CI Developers	# of people trained
Y1Q2	Fellows	All	4	Mentor Y1 Fellows	# of fellows, diversity
Y1Q2	Software	Morgridge	3	Work with CMS to integrate tokens with HTCondor, Jupyter, XrootD	downloads, bug reports, mailing list posts
Y1Q2	Standards	Morgridge	3	Publish JWT Profile for CI	# of authors, adoption
Y1Q2	Security	Illinois	2	Security assessment of JWT implementations for CI	participants, findings
Y1Q2	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y1Q3	Fellows	All	1	Y1 Fellows Workshop: Sharing Project Results	# of fellows, diversity
Y1Q3	Outreach	All	2	Submit paper to PEARC22	paper accepted
Y1Q3	Security	Illinois	1	Token issuer peer-review (TAGPMA)	participants, findings
Y1Q3	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y1Q4	Outreach	All	1	Present paper at PEARC22	citations, attendees, feedback
Y1Q4	Training	Nebraska	2	Develop online training module: OAuth Basics for CI Developers	# of people trained

Y1Q4	Software	Nebraska	2	Demonstrate JWTs for user-managed file shares (scopes / permissions)	downloads, bug reports, mailing list posts
Y1Q4	Management	All	0.2	Submit Annual Report to NSF	review by NSF Program Officer
Y1Q4	Security	Illinois	2	Tabletop exercise: refresh token compromise	participants, findings
Y1Q4	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y2Q1	Training	Nebraska	2	Develop online training module: OAuth Device Flow for CI Developers	# of people trained
Y2Q1	Software	Nebraska	2	OSG milestone: complete migration from Grid Community Toolkit	downloads, bug reports, mailing list posts
Y2Q1	Security	Illinois	2	Publish operational guidelines for token issuers	adoption, feedback
Y2Q1	Fellows	Illinois	1	Select Y2 Fellows	# of applications received, applicant diversity
Y2Q1	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y2Q2	Outreach	Illinois	1	Present at Gateways Conference	attendees, feedback
Y2Q2	Security	Illinois	2	Tabletop exercise: token issuer compromise	participants, findings
Y2Q2	Fellows	All	4	Mentor Y2 Fellows	# of fellows, diversity
Y2Q2	Standards	Morgridge	1	CI JWT Hackathon & Interop Fest	participants, implementations
Y2Q2	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y2Q3	Outreach	All	2	Submit paper to PEARC23	paper accepted
Y2Q3	Training	Nebraska	2	Develop online training module: OAuth Token Exchange for CI Developers	# of people trained
Y2Q3	Fellows	All	1	Y2 Fellows Workshop: Sharing Project Results	# of fellows, diversity
Y2Q3	Security	Illinois	1	Token issuer peer-review (TAGPMA)	participants, findings
Y2Q3	Software	Morgridge	2	Address science driver using tokens (agile approach)	downloads, bug reports, mailing list posts
Y2Q3	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y2Q4	Outreach	All	1	Present paper at PEARC23	citations

Y2Q4	Security	Morgridge	2	Tabletop exercise: submit node compromise	participants, findings
Y2Q4	Standards	Morgridge	2	Revised JWT Profile for CI	# of authors, adoption
Y2Q4	Management	All	0.2	Submit Annual Report to NSF	review by NSF Program Officer
Y2Q4	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y3Q1	Training	Nebraska	2	Develop online training module: OIDC for CI Developers	# of people trained
Y3Q1	Software	Nebraska	2	Address science driver using tokens (agile approach)	downloads, bug reports, mailing list posts
Y3Q1	Fellows	Illinois	1	Select Y3 Fellows	# of applications received, applicant diversity
Y3Q1	Security	Illinois	2	Tabletop exercise: identity provider compromise (SIRTFI)	participants, findings
Y3Q1	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y3Q2	Outreach	Morgridge	1	Present at JupyterCon	attendees, feedback
Y3Q2	Fellows	All	4	Mentor Y3 Fellows	# of fellows, diversity
Y3Q2	Standards	Morgridge	1	CI JWT Hackathon & Interop Fest	participants, implementations
Y3Q2	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y3Q3	Fellows	All	1	Y3 Fellows Workshop: Sharing Project Results	# of fellows, diversity
Y3Q3	Training	Nebraska	2	Develop online training module: Credential Management for Researchers	# of people trained
Y3Q3	Outreach	All	2	Submit paper to PEARC24	paper accepted
Y3Q3	Security	Illinois	1	Token issuer peer-review (TAGPMA)	participants, findings
Y3Q3	Software	Morgridge	2	WLCG milestone: migration from X.509 to tokens	downloads, bug reports, mailing list posts
Y3Q3	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time
Y3Q4	Management	All	0.2	Submit Final Report to NSF	review by NSF Program Officer
Y3Q4	Outreach	All	1	Present paper at PEARC24	citations
Y3Q4	Management	All	1	Project meetings, website updates, coordination	project deliverables achieved on-time