NEON IdM Experiences

Working with the CI CoE Pilot to Solve Identity Management Challenges

September 9, 2019

Ryan Kiser, Terry Fleury, Christine Laney, Jeremy Sampson, & Susan Sons

For public distribution.

Abstract

This paper details experience from collaborative efforts between the NSF Cyberinfrastructure Center of Excellence (CI CoE) Pilot's Identity Management Working Group and staff from the National Ecological Observatory Network (NEON) to develop improvements to the NEON Data Portal as well as the products of these collaborative efforts.

Acknowledgements

The CI CoE Pilot team would like to thank Trusted CI for their co-sponsorship of the CI CoE Identity Management effort, for funding Terry Fleury to participate in this project, and for the use of templates and other materials used throughout. In addition, the CI CoE Pilot team would like to thank the NEON team for their work on this project.

This document is a product of the CI CoE Pilot Project in collaboration with NEON and Trusted CI. The CI CoE Pilot is supported by the National Science Foundation (NSF) under Grant Number OAC-1842042[1]. NEON is supported by the NSF under Grant Number DBI-1724433[2] Trusted CI is supported by the NSF under Grant Number ACI-1547272[3]. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

License

[1] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1842042
[2] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1724433
[3] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1547272

Table of Contents

Introduction

The National Science Foundation's National Ecological Observatory Network (NEON)[4] is a continental-scale observation facility operated by Battelle. NEON was designed by ecologists to collect long-term ecological data to better understand how U.S. ecosystems are changing. NEON data are provided freely and openly to the global research community. Operating at 81 field sites across the United States, with a planned operations timeline of 30 years, NEON collects intensive and standardized field, instrument, and remote sensing data to enable research at spatial and temporal scales not accessible to previous generations of ecologists.

NEON is based in Boulder, Colorado, with field research offices located across the US. Data are continuously streaming into headquarters from logging devices and mobile applications. In addition, hard drives containing data from airborne observations are routinely shipped to headquarters for ingest and processing. After processing is complete, data are published and made freely discoverable and accessible through NEON's data portal[5] and API.

A primary measure of success for NEON is the traceable use of NEON data in research, especially where it advances the state of ecological science. As such, NEON would like to better track how data are interacted with, whether by exploration and download through the data portal, the API, or third party data repositories that host some of NEON's data or metadata. A simple and secure sign in process which provides tangible workflow benefits may increase engagement with NEON's user community and allow NEON to better track research progress that leverages NEON outputs. These benefits should be accessible while still providing unauthenticated access to the data. Our primary interest in this project is to find a solution for user identity management that is secure, relatively easy to implement, and provides a modern experience for all users.

This project brought together Christine Laney (Principal Research Scientist, Battelle), Jeremy Sampson (Research Scientist, Battelle), Susan Sons (Identity Management Lead, CI CoE Pilot), Terry Fleury (Senior Research Programmer, Trusted CI) and Ryan Kiser (Senior Security Analyst, CI CoE Pilot and Trusted CI) to work collaboratively to identify and produce a solution over a period of one year.

[4] https://www.neonscience.org/
[5] https://data.neonscience.org/home

I. Personnel

1. NEON Team Members

Christine Laney is a Principal Research Scientist - Ecologist with Battelle, working on the NEON project. She has worked on numerous ecological monitoring and data synthesis initiatives. Her research area includes ecology, informatics, and programming, as well as social and practical aspects of data sharing for research purposes. She is currently the lead ecoinformaticist at NEON and guides the development of NEON's data portal.

Jeremy Sampson is a Research Scientist - Software Engineer at Battelle Memorial Institute currently focused on NEON. He has worked with Battelle since 2011 contributing to a wide range of software projects and is well versed in a variety of programming languages, frameworks, technology stacks, software engineering practices and design methodologies.

2. CI CoE Pilot and TrustedCI Team Members

Susan Sons serves as Identity Management team lead for the CI CoE Pilot, and Chief Security Analyst at Indiana University's Center for Applied Cybersecurity Research. Her work focuses on the security needs of Research and Development organizations, software security, control systems security, and running security teams for scientific cyberinfrastructure. Susan is a co-author of the Information Security Practice Principles[6].

Terry Fleury is a Senior Research Programmer at the National Center for Supercomputing Applications (NCSA). He has worked in the Cybersecurity Division since 2005 where he has assisted with the development of several open-source security projects including MyProxy, CILogon, and SWAMP. In his work with Trusted CI, he has performed several engagements assisting NSF projects with risk assessments and cybersecurity program development.

Ryan Kiser is a Senior Security Analyst at the Indiana University Center for Applied Cybersecurity Research, the CI-CoE Pilot, and Trusted CI. Ryan has worked on information security projects across a wide variety of domains including leading efforts to assess and improve the security of automotive engine systems, performing risk assessments for university central IT systems, and supporting researchers in efforts to adhere to regulated data requirements such as HIPAA, FISMA, and various CUI requirements.

[6] https://cacr.iu.edu/principles/

II. Choosing and Integrating a Solution

1. NEON's Resources and Constraints

NEON serves approximately 2000 users per month through the data portal and API. From internal logs, it appears that the API has seen substantially more use than the data portal; at times downloads through the API has been heavy enough to cause instabilities in data provisioning. There was no IP tracking on the data portal, and only IP tracking on API use. Very few users logged into the data portal's custom user account system, and no API interactions were associated with user accounts. Given the importance of tracking user interaction with NEON data to estimate the impact that NEON has on research, we combined the idea of providing a trusted third-party mechanism for users to log into the the data portal with allowing users to request API tokens. Interactions with tokens could be assumed to be trusted with large download requests, while non-authenticated interactions could be rate limited as needed. NEON's resources to develop a new IdM system were limited in funding and time to develop and implement.

2. The CI CoE Pilot Approach

Trusted CI[7] engagements with the NSF community have shown the value in developing a plan describing objectives, resources, deliverables, and risks. Such an engagement plan solves or prevents several potential problems. It provides a mechanism for project participants to track progress, a way to identify and prevent scope creep or drift, and an opportunity to identify other risks and how to mitigate them if they arise. The CI CoE Pilot has developed an engagement plan template for use when developing these plans. This template is based on an existing engagement plan template used by Trusted CI. This template was used as the basis for an engagement plan by the CI CoE Identity Management team and it is in use by other teams for their own distinct efforts to assist NEON.

The NEON and Pilot teams established recurring weekly calls. This proved beneficial for multiple reasons: it ensured that the NEON team had a regular opportunity to discuss issues with the Pilot team, it allowed the Pilot team the opportunity to assess progress and provide support, and it provided an opportunity to capture ongoing progress, challenges, and lessons learned in near real-time.

The original NEON Data Portal enabled users to log in with a username and password. See Figure 1 below for an example of the NEON Data Portal "Sign In" screen.

---

[7] https://trustedci.org

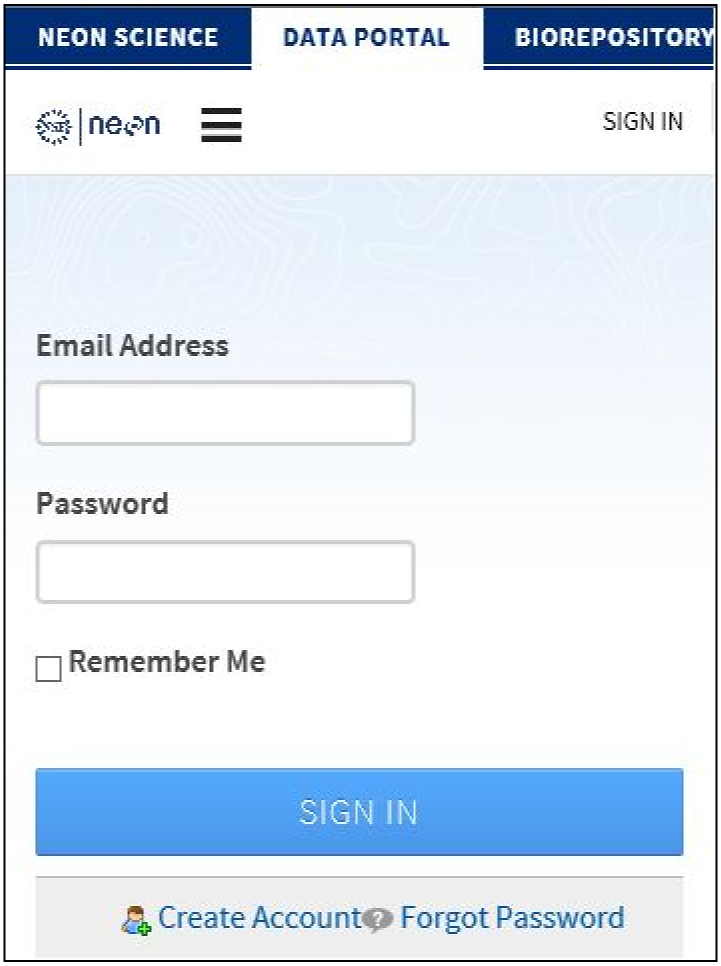

Figure 1. Original NEON Data Portal "Sign In" Screen

User account information is stored in a local datastore. A priority for the NEON team was to stop managing user credentials. One potential solution was to leverage external identity providers through the use of OAuth2 / OpenID Connect (OIDC). There were two main concerns with this approach. First, it was desirable to maintain the current database of users so users would not be required to re-register for the Data Portal. Second, consideration needed to be given as to which external identity provider(s) would be used. Google would be an easy choice as Google allows creation of OAuth2/OIDC client integrations for free, but users would be required to have a Google account. ORCID would also be a viable choice since users could create ORCID accounts for free and many researchers already have ORCID IDs due to journal requirements, but ORCID has an emphasis on scientific publications, so it may not be palatable for a general audience.

The pilot team began by developing a set of recommendations that reflected these observations as well as a plan for their implementation and captured them in the form of a whitepaper. Once complete, this was delivered to NEON and the Pilot team began efforts to support NEON with implementation.

Auth0[8] was selected for an initial test integration as it addressed both concerns. First, NEON would be able to import the current Data Portal user datastore into Auth0. Users could continue to use username/password authentication if they so desired. Second, Auth0 provides "social connections" to leverage external identity providers such as Google. The free tier of Auth0 allows for up to 7000 active users and two social connections, which should be sufficient given NEON's 2000+ users.

The next step was to modify the NEON Data Portal from authenticating against the local datastore, and instead use Auth0 as an OIDC provider. This was a major shift in authentication implementation for the portal, so a development instance was created so as not to impact active users of the portal. Auth0 provides OIDC client libraries for various programming languages and ample documentation on creating application integrations. Auth0 also provides a CLI tool for applying and updating tenant configurations programmatically. This assures that tenant configurations remain identical at all times. Three tenants were put in place in Auth0 to allow multiple purpose-built environments to maintain separate workspaces for integration testing, certification, and production. A small subset of the existing user database was exported to Auth0, and Google was selected as the initial "social connection", with an expectation to add CILogon[9] to leverage authentication with InCommon Federation identity providers. By connecting the portal to Auth0 using OIDC, common user attributes such as name and email address would be asserted by Auth0 regardless of the data source.

Once this initial integration proved sufficient, effort was allocated to implementing CILogon as a second "social connection". Auth0 allows for connection to any OAuth2/OIDC provider via a "Custom Social Connection" extension, but this does require some additional scripting to map claims from 3rd party identity providers to Auth0 user attributes. An example of the CILogon custom social connection is shown in Figure 2 below.

[8] https://auth0.com/docs/login
[9] https://www.cilogon.org/oidc

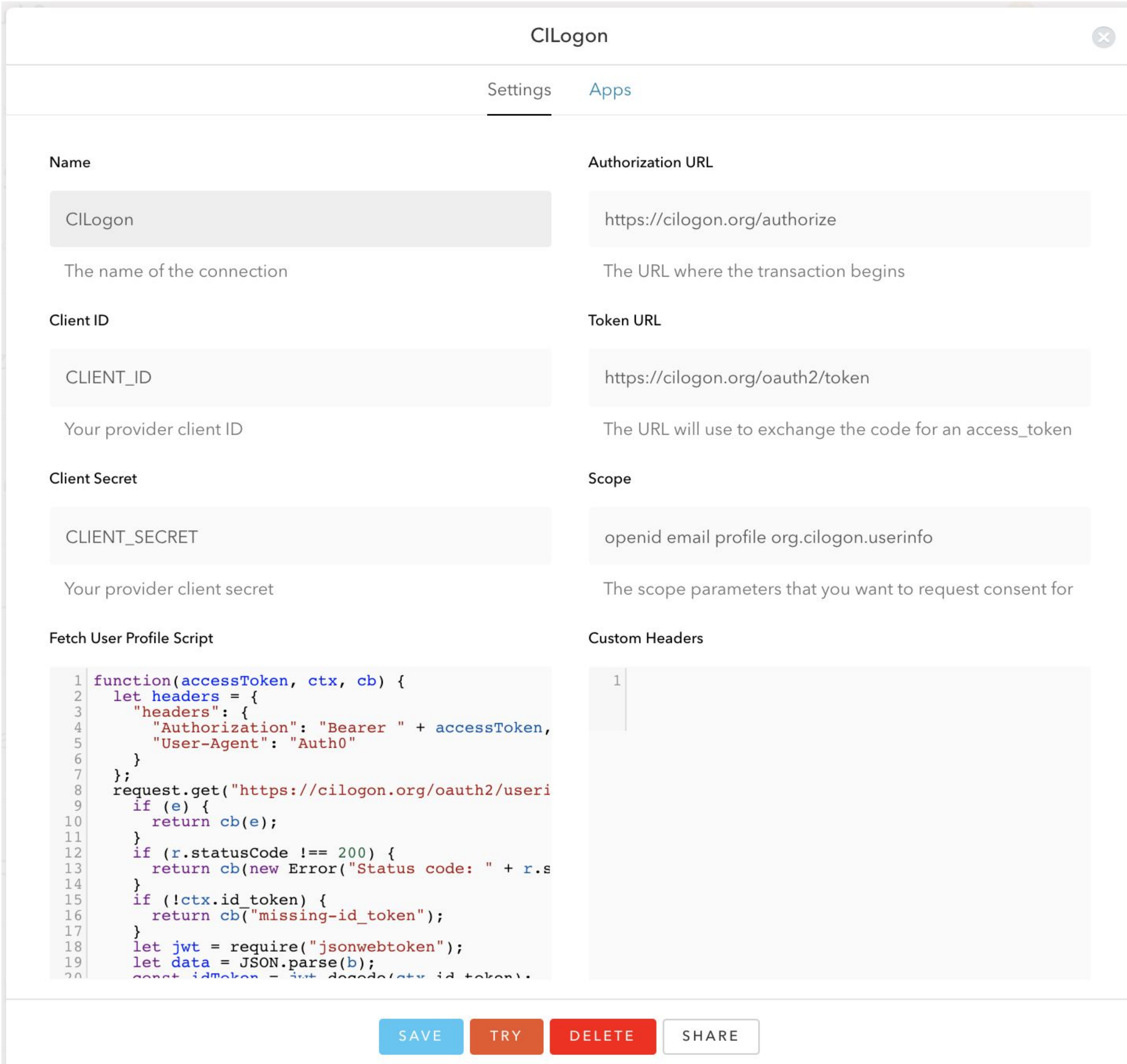

Figure 2. Auth0 Custom Social Connection Configuration for CILogon

Jeremy consulted with Terry for details on registering a new CILogon OIDC client for use as a "custom social connection" in Auth0. Terry provided guidance on which scopes to request as well as the claims returned by CILogon for InCommon Federated Identity Providers. Jeremy was then able to write the necessary custom script to map attributes appropriately for Auth0 user objects, and also created a custom UI login box that built upon the existing Auth0 version for the development NEON Data Portal. See Figure 3 for an image of the new Auth0 login screen which allows a user to log in using either their old username and password, or one of the "social connections" Google or CILogon.

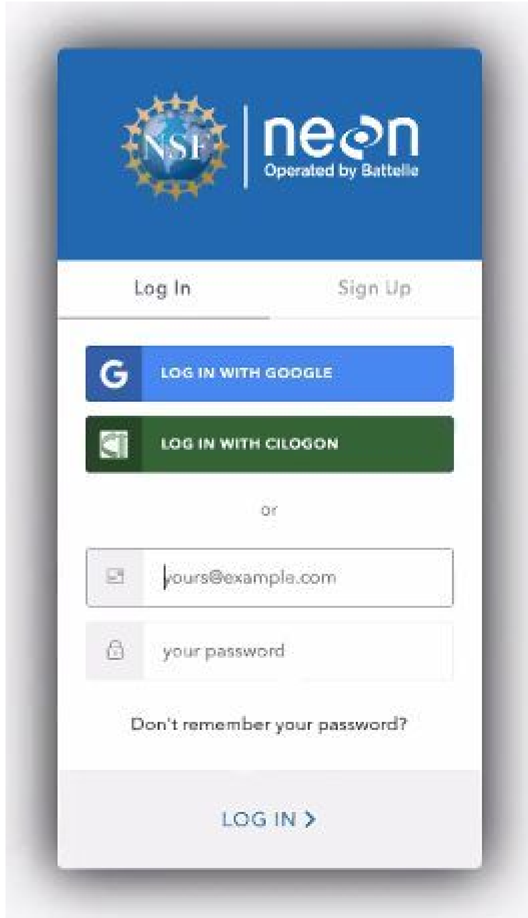

Figure 3. Auth0 Login Screen for the NEON Data Portal

III. Looking Ahead

1. Results of this Engagement

The NEON user management system now allows users to create and manage accounts associated with credentials that they already have through a third party system; for many academic researchers, this includes their University account credentials via the integration with CILogon. In addition, NEON no longer needs to manage as many user accounts, and will have enriched logs, due to the implementation of Auth0.

During this activity, NEON also took the opportunity to modernize the user account web pages and start providing more opportunities for the community to engage with NEON, such as providing ways to add Twitter handles and ORCID IDs, as well as data product and field site preferences to account information.

2. Future Plans

In the future, NEON would like to further leverage Auth0 and CILogon. This would include migrating the handling of all user accounts (including staff enterprise accounts) to Auth0 and using more of the resources that Auth0 provides. NEON anticipates supporting a richer engagement with its user community, particularly through providing benefits such as signing up or opting out of both generic and custom newsletters about the NEON program and about data products based on research or location preferences, respectively.