Globus Integration with NIH's Researcher Auth Service (RAS) and the NIH's Common Fund Data Ecosystem (CFDE) Portal using OAuth, OIDC, and GA4GH Passports

Lee Liming – lliming@uchicago.edu

WoTBAn&Az 2022 - October 18, 2022

THE UNIVERSITY OF CHICAGO

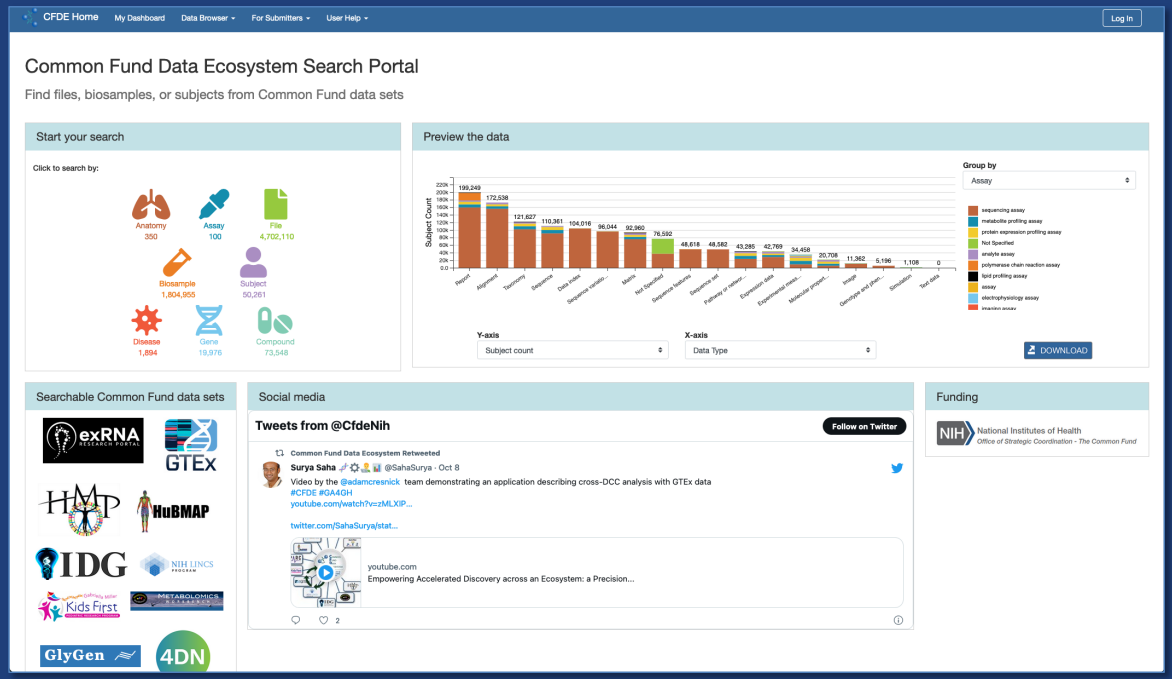
globus

NIH Common Fund Data Ecosystem (CFDE)

- **Standardize, simplify, and enhance researchers' experience with NIH's Common Fund data, much like…**
  - NASA EOS portal (Earth observational data)
  - USGS ScienceBase (geological data)
  - ACCESS (NSF supercomputing centers)
- **The Common Fund sponsors several dozen big programs (each with dozens of individual awards) that produce valuable medical data.**
  - Each program has one (or more!) **Data Coordinating Centers (DCCs)** that collect and manage the data from the program's research projects.
  - Each DCC has its own data portal (some have more than one), but it's challenging to navigate them all.

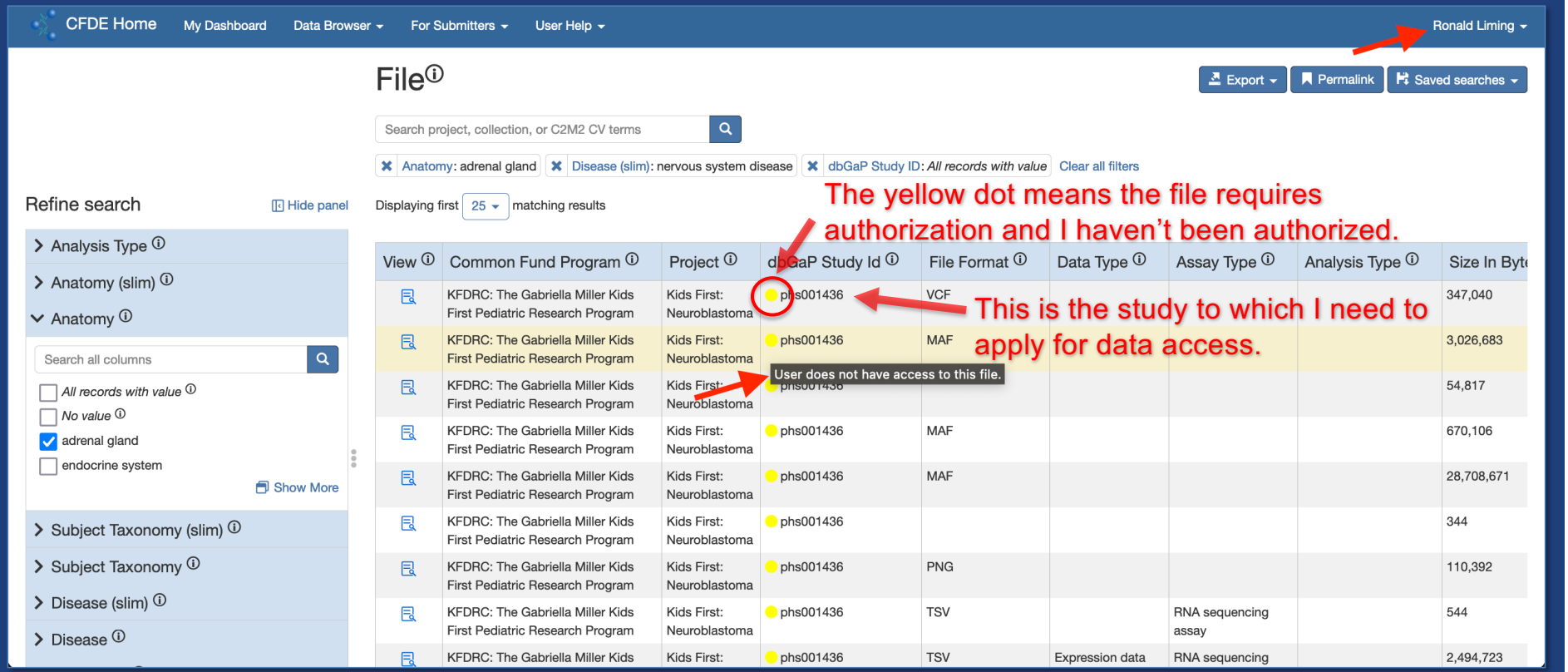
NIH Common Fund Data Ecosystem (CFDE) Portal – app.nih-cfde.org

- **Discover NIH Common Fund data relevant to a specific research problem**
- **Goal: Inform researchers about their NIH data access authorizations during discovery process**



University of Maryland School of Medicine · UC Davis · USC Viterbi Information Sciences Institute · UC San Diego · The University of Chicago · Icahn School of Medicine at Mount Sinai

Inform researchers about their authorizations…



Authorizing access to human subject data

- **Privacy for human subject data is very important!**
  - Penalties for improper access are severe
  - Specific policies are complex
- **Access is often authorized by data access committees (DACs)**
  - DACs accept and review researcher applications against defined data access policies
  - DAC decisions must be implemented by data access systems
- **There are many DACs and there are many data access systems**
  - DACs and data access systems *are not* tightly integrated
  - Researcher applications (like the CFDE Portal) may not be closely affiliated with the DACs *or* with the data access systems!
  - Thus, *access to authorization data* is a key capability

GA4GH AAI and Passport - www.ga4gh.org

- **GA4GH AAI defines GA4GH claims and visas**
  - *Visas* are **authorization assertions** signed by a *visa issuer* that can be communicated throughout the system; signature remains with visa
  - Visa contains signed JWT which may be used as an OAuth access token with data access systems
- **GA4GH Passport defines a JSON data structure containing a person's visas**
  - Applications may request and obtain a person's passport
  - Visas in the passport inform the application about the person's authorizations
  - The application may use the embedded OAuth access tokens to access data on the person's behalf

GA4GH AAI and Passport - www.ga4gh.org

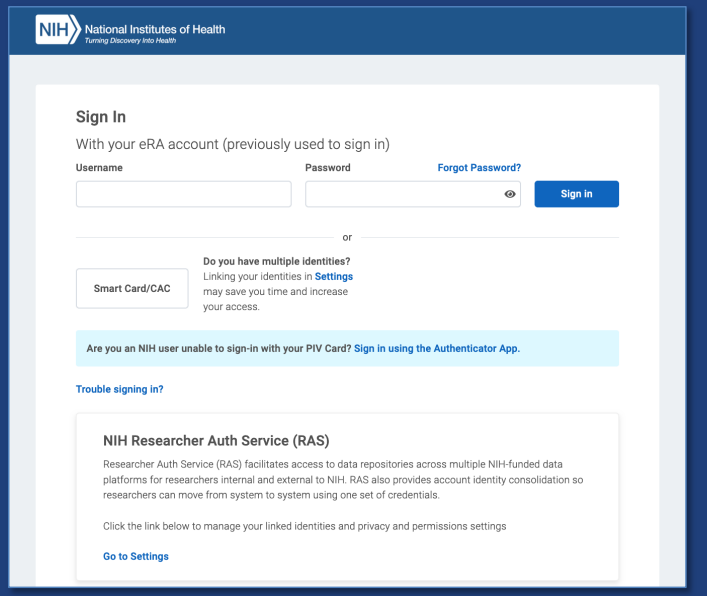
Global Alliance for Genomics & Health

- **GA4GH AAI defines an *OIDC broker* service**
  - OIDC broker services obtain visas from issuers (process is not specified)
  - OIDC broker services authenticate researcher identities
  - OIDC broker services generate and deliver passports for authenticated individuals to applications
- **OIDC brokers leverage OpenID Connect 1.0**
  - OIDC Auth Code flow (confidential client)
  - OIDC Implicit flow (ID Token response type)
  - During the OIDC flow, application requests a well-known OAuth scope for passport access (`ga4gh_passport_v1`)
  - The authenticated individual's passport appears as a claim in the `/userinfo` response (assuming passport access scope is present)

Key implication: Because most OIDC servers only accept the access tokens they themselves issue, in order to obtain a researcher's passport, *the researcher must authenticate with the OIDC broker.*

NIH Researcher Auth Service (RAS) - auth.nih.gov/docs/RAS/

- **NIH's RAS is a GA4GH OIDC broker**
  - Obtains visas from dbGaP (Database of Genotypes and Phenotypes)
  - Visas originate from dbGaP data access committees (DACs)
- **Goal: RAS authorizes access to NIH controlled-access data**
  - NIH aims to integrate its many data access systems with RAS
  - NIH aims to integrate RAS with other authorization systems
- **RAS's OIDC service enables multiple authentication methods**
  - eRA Commons, NIH smartcard/authenticator, Login.gov, InCommon IDPs*, social auth*
  - Identity linking is supported
  - Clients can be configured to enable distinct set of login methods



Approach: CFDE Portal and Globus

- **CFDE Portal uses several Globus APIs and services**
  - **Auth** for portal logins (with existing institution accounts)
  - **Groups** for portal permissions (data submissions & curation)
  - **Transfer** and **Flows** for data submissions
  - These require Globus-issued access tokens
- **Requiring researchers to authenticate twice (once with Globus for portal, once with RAS for passport) was rejected as bad UX**
- **Globus already supports use of other OIDC auth services**
  - Could we add NIH RAS as an OIDC auth provider in Globus and pass the researcher's passport information through to the CFDE Portal?

NIH RAS integration in Globus

1. **Add NIH RAS as a Globus authentication provider**
   - Use RAS's GA4GH OIDC service endpoint
   - eRA Commons and NIH smartcard/authentication login options enabled
2. **Add `ras_passport` OAuth scope in Globus**
   - Globus must whitelist OIDC client for scope (RAS team must approve all clients)
   - When scope is requested and when authenticating with RAS, add the `ga4gh_passport_v1` scope to the RAS OIDC flow
   - On successful authentication, Globus calls RAS's `/userinfo` endpoint, obtains passport, removes all signatures from visas, and stores remaining claims in session cache
3. **Add passport info to Globus's OAuth token introspection response**
   - Because of Globus's identity linking, the `/userinfo` response isn't guaranteed to return RAS identity info
   - Globus's token introspection response already has a `session_info` section for authentication events (time, IDP, identity, REFEDs MFA claims, etc.)

Globus token introspection with RAS claims


```
{
 "token_type": "Bearer",
 "scope": "email profile openid ras_passport",
 "username": "cfdetestuser151@era.nih.gov",
 "exp": 1665942264,
 "iat": 1665769464,
 "sub": "879c4a48-5975-4b35-bfc9-107c7e1ce2de",
 [...]
 "session_info": {
   "session_id": "388bbd74-265b-40ba-a880-c0434d4ee191",
   "authentications": {
     "879c4a48-5975-4b35-bfc9-107c7e1ce2de": {
       "acr": "https://stsstg.nih.gov/assurance/aal/1 https://stsstg.nih.gov/assurance/ial/1",
       "amr": null,
       "idp": "a12f6103-9e70-4f0d-bfb2-ecec75410c22",
       "auth_time": 1665769463,
       "custom_claims": {
         "cfde_ga4gh_passport_v1": [
           {
             "exp": 1665812663,
             "iat": 1665769463,
             "iss": "https://stsstg.nih.gov",
             [...]
             "ras_dbgap_permissions": [
               {
                 "role": "pi",
                 "phs_id": "phs000710",
                 "version": "v1",
                 "expiration": 1672549200,
                 "consent_name": "Unrestricted",
                 "consent_group": "c99",
                 "participant_set": "p1"
               }
             ] } ] }}}}}
```


Passport info is valid until…

Set of authorization assertions (JWTs/signatures removed!)

CFDE Portal user experience

- **Researchers may login to CFDE Portal using any login provider available in Globus**
- **Personalized dbGaP authorization features are enabled if researcher uses NIH Research Auth Service (RAS)**
- **Automatic portal logout when passport expires (12 hrs)**

Future directions for CFDE's RAS integration

- **Ease-of-use improvements**
  - Interactively offer to enable dbGaP features (with a RAS login) if not already enabled in the session
  - Prompt to reauthenticate with RAS when passport expires
- **As RAS expands its login options, consider enabling *all of them* in the Globus integration, but restrict CFDE Portal to the RAS options**
  - All CFDE Portal logins would use RAS via Globus (so passport would always be available), but researchers could still authenticate with InCommon IDPs, Login.gov, eRA Commons, etc.

Future directions for Globus's RAS integration

- **Use RAS visas to control access to data via Globus**
  - Data access could require a specific visa (from a specific issuer?)
  - Prompt to authenticate with RAS when visa is needed
  - Similar to the protected data features Globus has already
- **Globus should manage passport lifecycle internally**
  - Use RAS's refresh token to automatically refresh passport
- **Improve integration with RAS OIDC service**
  - How to handle RAS-asserted identities that overlap with other OIDC providers (CILogon, Globus, Google, etc.)?
  - Can GA4GH OIDC brokers accept passport-scoped access tokens issued by other OIDC servers? What assurances are needed?
- **Support other GA4GH OIDC brokers (e.g., ELIXIR)**

lliming@uchicago.edu

https://www.ga4gh.org
https://auth.nih.gov/docs/RAS
https://docs.globus.org/api/auth
support@globus.org