Identifiers.org protected data feature

EMBL's European Bioinformatics Institute (EMBL-EBI)

Renato Caminha Juaçaba Neto rjuacaba@ebi.ac.uk

Identifiers.org & EMBL-EBI

EMBL-EBI

- identifiers.org provides two services to its users
  - **Curated registry** of online data collections
  - **Resolver** of online locations for entries in data collections
- Maintained and hosted by the European Bioinformatics Institute
  - World leading source of public biomolecular data
  - Part of the European Molecular Biology Laboratory (EMBL) Europe's flagship laboratory for the life sciences
- Registry curation
  - Registry is maintained up to date to the best of our abilities
  - New entries are created on demand

Persistent identifiers (PIDs) and our resolver

- Global PIDs rely on collection identifiers and local entry identifiers
  - Each collection as a unique prefix (e.g. uniprot, biostudies, drugcentral, etc)
  - Assuming each entry in a collection has a persistent local ID (e.g. accession numbers)
  - PIDs are compact identifiers composed of *prefix* and *local ID* [1]
- Our resolver
  - Our registry include URL patterns for each data collection
  - From that, our resolver finds updated online locations for any PIDs
  - URLs identifiers.org/<PID> always redirect to a valid online location

    for example: identifiers.org/uniprot:P12345

[1] https://doi.org/10.1093/bioinformatics/btaa864

Protected data feature

- Some collections are **not** openly accessible directly
- Authentication is sometimes required to
  - Determinate access level
  - Constraint and identify access
- Current use cases
  - Sensitive information
  - Costly data access
- New identifiers.org feature: Protected resources
  - Continued lightweight and inclusive design

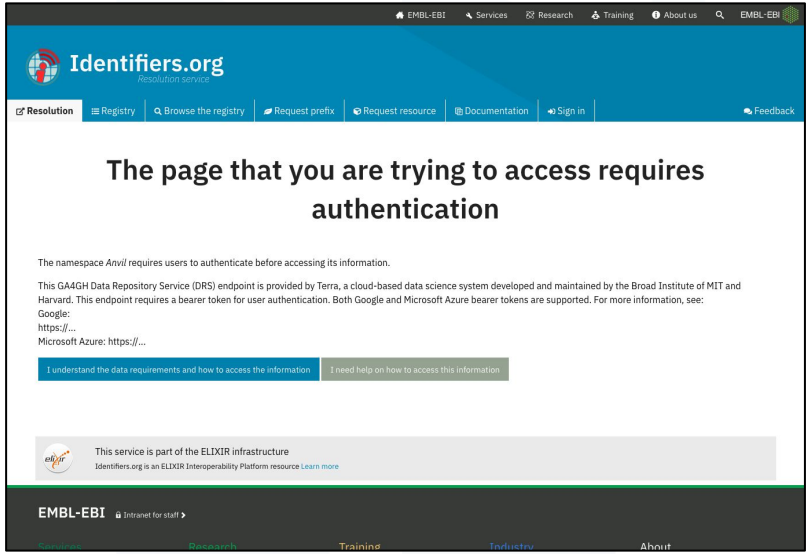
Protected resources in identifiers.org

- Extra information added to these protected entries
  - *High level description* on how to access entries
  - *URL* for more information on the process
- Collections available in different versions
  - Some versions may be protected and others not
  - Supported via multiple online resources per collection
- Optional: Landing page for these resources
  - Data providers may not have a good landing page
  - We provide an landing page in these cases
  - This page includes the help information from the previous point

Unauthorized

This server could not verify that you are authorized to access the document requested. Either you supplied the wrong credentials (e.g., bad password), or your browser doesn't understand how to supply the credentials required.

Apache Server at data.terra.bio Port 80



Future prospects

- This is a first step in supporting sensitive information
- There are issues
  - Not very machine actionable
  - Not much metadata is available
- We hope to evolve this as
  new guidelines and metadata schemas are available