DUAD: Towards Machine Readable Data Access Conditions

By members of the GA4GH Data Use and Research Identity (DURI) working group

Live version here: PDF

Problem

Data assets, databases, and other share-worthy health research resources typically do not have access conditions attached to them. This makes it difficult for owners or custodians to determine who can access the resource and under what conditions.

Solution

Our goal is to create a common human and machine readable access model that links data resources to access terms and conditions. This model is based on approved GA4GH standards to bring together semantic and syntactic aspects of consent and data use conditions. We propose to call this model **DUAD** for **Data Use Access & Discovery**.

More specifically, DUAD will reference the formal access terms of the GA4GH Data Use Ontology (DUO) and borrows structure and logic principles that underpin ADA-M (Automated Data Access Matrix) to optionally apply additional conditions. **DUAD will not require changes to either of these existing standards, but merely references/exploits them.**

Technically, DUAD is being devised as a *SchemaBlocks* based protocol that data custodians can use to define and implement their own consent and data use conditions in unambiguous form, to then assign to their data resources. DUAD will allow a flexible use such that data use terms can simply implement DUO terms in standalone blocks or can define more complex blocks that include terms with added conditions. DUAD thereby leverages the ontological richness and clarity of DUO combined with the versatile restriction/permission logic and contextual nuances borrowed from ADA-M. The modular design allows users to employ whichever consent and control elements best fit their use case.

The current version (below) is a work in progress.

Please send comments and suggestions to Francis Jeanson (fjeanson@yahoo.com), Anthony Brookes (ajb97@leicester.ac.uk) and Melanie Courtot (mcourtot@ebi.ac.uk). We would also welcome use cases implementing DUAD, and are happy to support their encoding.

Example Applications

Application 1: DUAD based access documents may be created for datasets as a means for institutions to strengthen their internal data governance.

Application 2: A DUAD based document may facilitate semi or fully automated access by requesting parties. In particular, we believe DUAD could be used in consort with GA4GH Passports such that

Passport VISAs, would explicitly refer to a DUAD document and authorize an individual to access the DUAD related resource(s). That is, by being awarded the VISA the individual's profile would match the DUAD requestor profile, and the individual would have agreed to comply with any further DUAD access conditions.

Application 3: A DUAD based document could serve as the basis from which is generated consent form language with respect to the terms of use for a participant's data. This would help ensure that patient's see how their data is intended for use in a manner consistent with intended sharing of that data.

DUAD Schema Block (under development)

```
SchemaBlock:
  version:
    type: string
  components:
    schemas:
      AccessProfile:
        type: object
        properties:
          name:
            type: string
          accessProfileVersion:
            type: string
          accessProfileReferences:
            description: 'Publications, URLs, DOIs
            for the resource.'
            type: array
            items:
              type: string
          accessProfileCreateDate:
            type: string
            example: '2017-01-17T20:33:40Z'
          accessProfileUpdates:
            description: 'Dates at which this access
            profile was updated.'
            type: array
            items:
              type: string
          resourceName:
            type: string
          resourceReferences:
            description: 'Publications, URLs, DOIs
            for the resource.'
            type: array
            items:
              type: string
          resourceDescription:
            type: string
          resourceDataLevel:
            type: string
            enum:
              - UNKNOWN
              - DATABASE
              - METADATA
```

We start by defining general information about a DUAD document including a name, version, any external references, dates of creation and updates.

Next we define information about the resources (data, document, record, etc.) that a DUAD document is created for. This includes the name of the resource, any references about that resource (publications about a dataset), a general text description of the resource, what kind of resource this is (resource data level), the contacts (based on Contact schema), and the organisation that is custodian or owner of this resource.

```
        - SUMMARISED
        - DATASET
        - RECORDSET
        - RECORD
        - RECORDFIELD
    resourceContactNames:
      type: array
      items:
        $ref: '#/components/schemas/Contact'
    resourceContactOrganisations:
      type: array
      items:
        type: string
    sharingMode:
      type: string
      enum:
        - UNKNOWN
        - DISCOVERY
        - ACCESS
        - DISCOVERY_AND_ACCESS
    permissionMode:
      type: string
      enum:
        - UNKNOWN
        - ALL_TERMS_PERMITTED_BY_DEFAULT
        - ALL_TERMS_FORBIDDEN_BY_DEFAULT

Contact:
  type: object
  properties:
    name:
      type: string
    email:
      type: string

RestrictionClass:
  type: object
  properties:
    restriction_rule:
      type: string
      enum:
        - UNKNOWN
        - NO_CONSTRAINTS
        - CONSTRAINTS
        - FORBIDDEN
    restriction_object:
      $ref: ./OntologyClass
    constraints_details:
      type: string

TermClass:
  type: object
  properties:
    data_use_class:
      $ref: ./OntologyClass
    restriction_class:
      $ref: ./RestrictionClass
```

sharingMode is used to state if the resource can only be discovered by others (only metadata information is available) or if this resource is accessible only, or if both are possible.

permissionMode is used to define the overall default interpretation of a DUAD description as either implying that All Terms not explicitly defined are permitted OR All Terms not explicitly defined are forbidden (see Terms below).

This is the contact class used by resourceContactNames (see above) and may be used by future objects.

RestrictionClass is used by TermClass (below) to define **either or both** a restriction_rule and a restriction_object. A restriction_rule serves as the 'operation' to apply on the term's data_use_class (see below). Options include: Unknown, No Constraints, Constraints, or Forbidden.

An additional restriction_object from DUO data use requirements can also be assigned if suitable. For instance: 'timelimit on use'.

A constraint_details description can be added if additional clarifications are required. For instance: '6 months' might be a detail required for a timelimit on use.

A TermClass object defines a DUO data_use_class term that meets a sharing condition required by the resource. For instance: 'publication required'

The corresponding restriction_class (see above) is selected to add a restriction rule or object to be applied on the DUO term. For instance: 'CONSTRAINTS' on 'publication required' will cue a reader to lookup the contraints_details that might include a list of recognized journals for publication

```
Terms:
  description: 'List of permitted terms of use
  including general terms, profile of users,
  purposes of use, etc.'
  type: array
  items:
    $ref: ./TermClass
```

that are acceptable.

Terms simply regroups all access terms (TermClass objects) into a list (array). Hence many terms can be applied to a DUAD document for a particular dataset/resource of interest.