Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

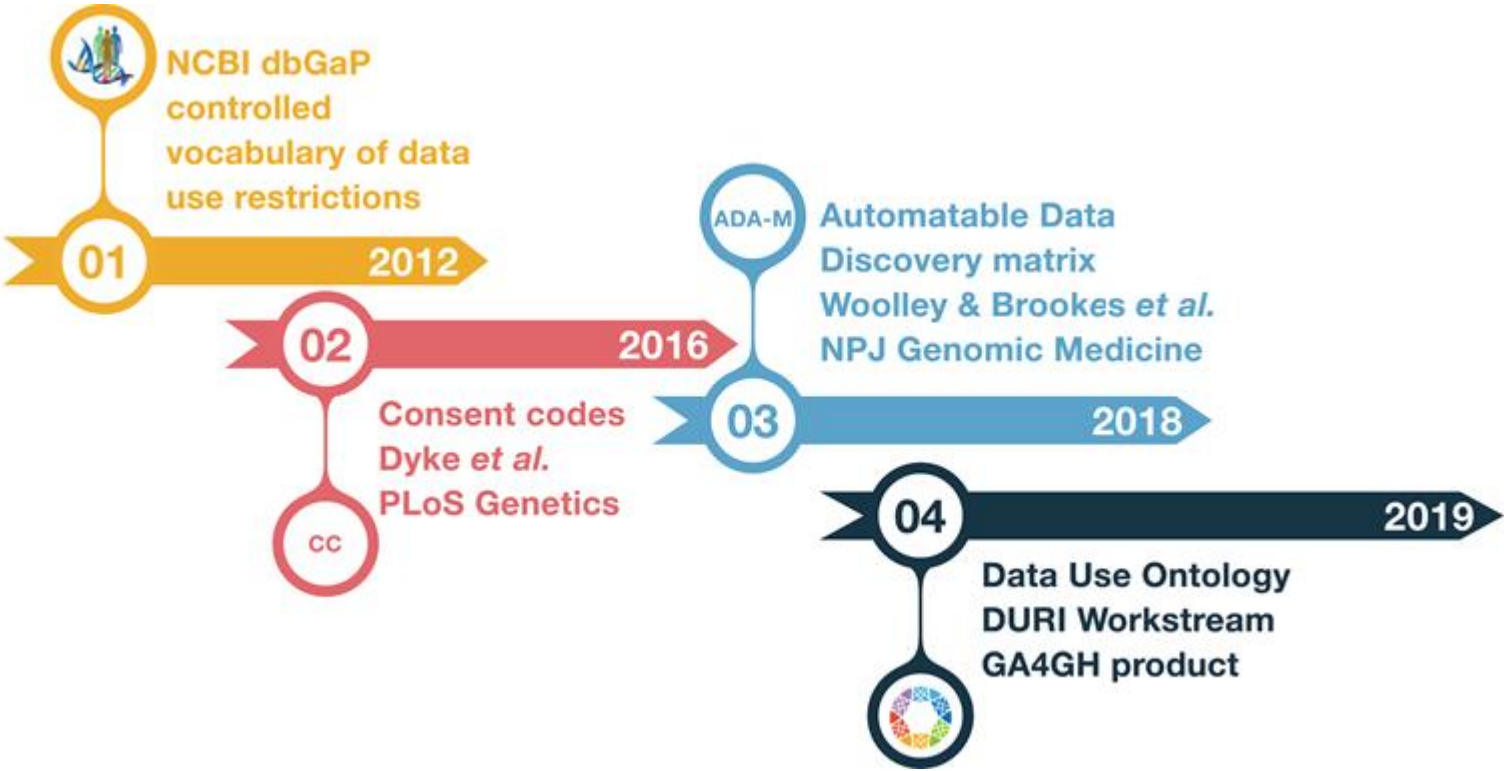
Automatable Discovery & Access Metadata
ADA-M …v2 [‘DUAD’]

GA4GH, October 2019

Anthony Brookes
Francis Jeanson, Melanie Courtot

Existing efforts

Global Alliance for Genomics & Health



Automatable Discovery & Access Matrix (ADA-M)

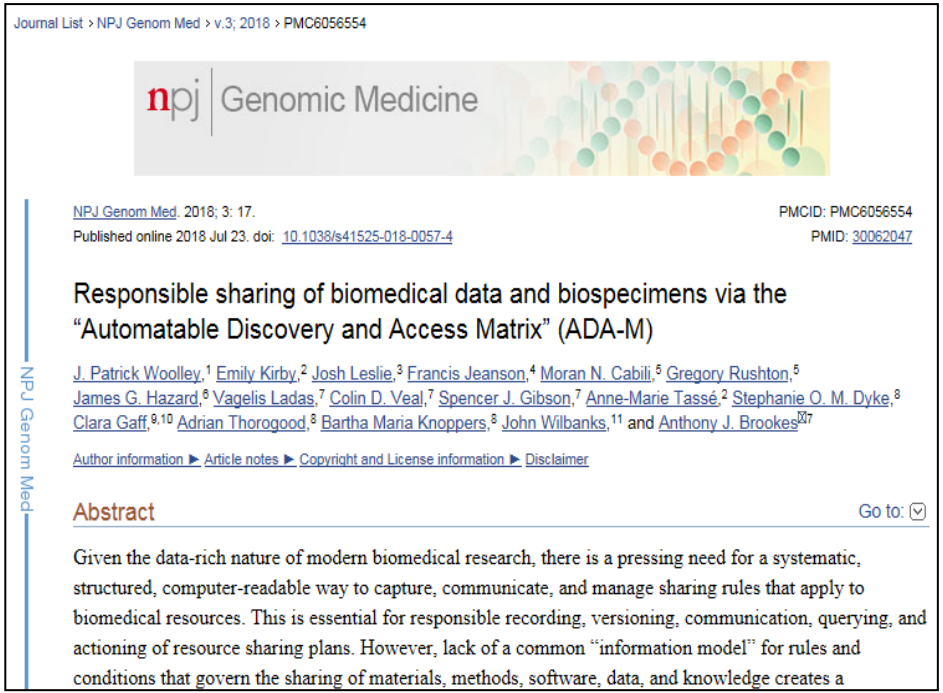
Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

IRDiRC
INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

Co-Chairs: **Anthony Brookes; John Wilbanks**
Coordinator: Emily Kirby

Vartika Agrawal; Alessia Baretta; Moran Cabili; Knox Carey; Virginia Chavis; Ed Conley; Primavera De Filippi; Stephanie Dyke; Luca Emili; David Fajgenbaum; Clara Gaff; Chiara Garattini; Jack Goldblatt; Peter Goodhand; Elli G. Gourna; Tudor Groza; Reece Hart; James Hazard; **Francis Jeanson**; Anneliene Jonker; Petra Kaufmann; Alastair Kent; Warren Kibbe; Vagelis Ladas; Bartha Maria Knoppers; Ray Krasinski; Lilian Lau; Josh Leslie; David Lloyd; Hanns Lochmuller; Alex Mankovich; Mauricio Moldes Quaresma; Francis Ouellette; Anthony Philippakis; Barbara Prainsack; Jordi Rambla; Fabien Richard; Olaf Riess; Gregory Rushton; Andrea Saltzman; Mahsa Shabani; Anne Marie Tassé; Adrian Thorogood; Mihaela Ulieru; Susan Wallace; **J. Patrick Woolley**; Tomoyuki Yamada; Andreas Zankl.

+ Substantial contributions from IMI(EMIF), P3G and Intel.



Journal List › NPJ Genom Med › v.3; 2018 › PMC6056554

npj | Genomic Medicine

NPJ Genom Med. 2018; 3: 17.
Published online 2018 Jul 23. doi: 10.1038/s41525-018-0057-4

PMCID: PMC6056554
PMID: 30062047

Responsible sharing of biomedical data and biospecimens via the "Automatable Discovery and Access Matrix" (ADA-M)

J. Patrick Woolley,[1] Emily Kirby,[2] Josh Leslie,[3] Francis Jeanson,[4] Moran N. Cabili,[5] Gregory Rushton,[5] James G. Hazard,[6] Vagelis Ladas,[7] Colin D. Veal,[7] Spencer J. Gibson,[7] Anne-Marie Tassé,[2] Stephanie O. M. Dyke,[8] Clara Gaff,[9,10] Adrian Thorogood,[8] Bartha Maria Knoppers,[8] John Wilbanks,[11] and Anthony J. Brookes[7]

Author information ▶ Article notes ▶ Copyright and License information ▶ Disclaimer

NPJ Genom Med

Abstract

Go to:

Given the data-rich nature of modern biomedical research, there is a pressing need for a systematic, structured, computer-readable way to capture, communicate, and manage sharing rules that apply to biomedical resources. This is essential for responsible recording, versioning, communication, querying, and actioning of resource sharing plans. However, lack of a common "information model" for rules and conditions that govern the sharing of materials, methods, software, data, and knowledge creates a

+ API
+ Support software

Consent design

? Standardisation ?

Legal/policy requirements

? Interpretation ?

Consent acquisition

? Precision ?

Institution & researcher requirements

Potential Uses

ADA-M Goals

Machine readable structure (‘metadata/license’) to increasingly automate discovery and sharing of data

Products

1. Design a **flexible** standard, for **all** relevant data use parameters, **ATOMIC** at **‘Goldilocks’ granularity**
2. Create standard formats (e.g., JSON, key-value text), support documents, software, API

Support for

1. Retrospective: transfer any and all existing use conditions to a standard representation
2. Prospective: capture evolving use conditions in new/longitudinal studies to a standard representation.

Profile Completion Rule Set for ADA-M v1.0

RULE #1

At least one ‘Permissions’ or ‘Terms’ category must have a value

RULE #2

Values must be entered for required HEADER fields:

- Metadata name [metadataName(v1)]
- Metadata version [metadataVersion(v1)]
- Resource name [resourceName (v2)]

And likewise for META-CONDITIONS field:

- Mode of sharing [sharingMode(v2)]
- Interpretation rule if multiple Obligatory permissions are specified [multipleObligationsRule(v1)]
  but ONLY IF >1 [Obligatory] string is present across all Permissions & Terms controlled value and free text fields

RULE #3

Values must comply with permitted values for HEADER fields:

- Metadata name [metadataName(v1)]
- Resource data level [resourceDataLevel(v1)]

And likewise for PERMISSIONS fields:

- use within countries/locations [country(v2)]
- use by organisations [organisation(v2)]
  - use by non-profit organisations [nonProfitOrganisation(v2)]
  - use by profit organisations [profitOrganisation(v2)] . . . etc

THE RESOURCE/DATA ARE PERMITTED FOR USE:

within any country/location

within specified countries/locations

by all organisations

by any non-profit organisations

by specified non-profit organisations

by any profit organisations

by specified profit organisations ***. . .etc***

use for research purposes

for specified categories of research

use for methods development

for specified categories of methods development research

use as reference or control material

for specified categories of control or reference use

use for biomedical research

for specified categories of biomedical research

use for research concerning fundamental biology

for specified categories of fundamental biology research

use for research concerning genetics

for specified categories of genetics research ***. . .etc***

use for clinical purposes

for specified categories of clinically related use

use for clinical decision support

for specified categories of clinical decision support

use for clinical purposes concerning any disease

“Unrestricted”

or

“Unrestricted[Obligatory]”

or

”Limited”

or

”Limited[Obligatory]”

or

”Forbidden”

TERMS OF AGREEMENT:

There are no requirements for any formal approval, contract or review conditions to be satisfied

Formal approval, contract or review conditions are to be met, as specified

There are no requirements regarding publication or disclosure of results, or included references or acknowledgements

Publication or disclosure of derived results is subject to restrictions, as specified

There are no timeline restrictions

The period of access has time limitations, as specified

There are no requirements regarding data security measures

User must have adequate data security measures, as specified

There are no requirements regarding withdrawal, destruction or return of any subject data

Some subject data must be withdrawn, destroyed or returned, as specified

There are no restrictions regarding the linking of accessed records to other datasets relevant to subject identifiability

Accessed records may only be linked to other datasets, as specified

There is no possibility of recontacting data subjects

Subject recontact may occur in certain circumstances, as specified

Subject recontact must occur in certain circumstances, as specified

There are no restrictions regarding intellectual property claims based on use of the accessed resource

Options for intellectual property claims based on use of the accessed resources are limited, as specified

There are no requirements to report back regarding use of the accessed resources

Reporting on use of the accessed resources may be required, as specified *. . .etc*

META-CONDITIONS:

There are no other use restrictions/limitations in force which are not herein specified

Other permissions/limitations may apply as specified

"True"

or

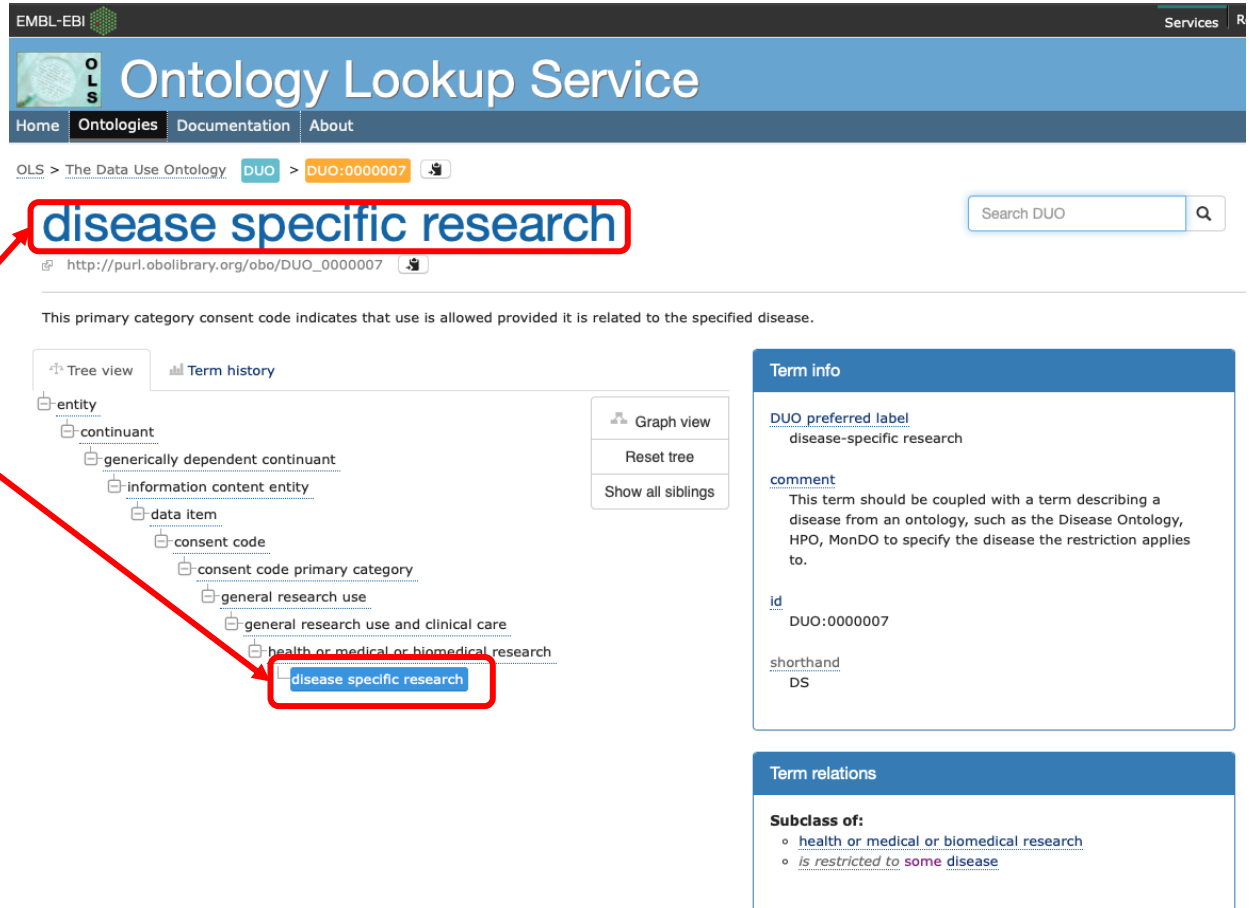
"Untrue"

Utility – now and going forward

1. Communication
   Reporting/displaying the ‘profile’ of a dataset’s accessibility
2. Enhance discovery
   Add as extra information to answer ‘discovery’ queries (e.g. Beacon)
3. Automated data sharing
   Align requests to ADA-M profiles, and share if conditions are met
4. Join forces with DUO

http://purl.obolibrary.org/obo/DUO_0000007

Global Alliance for Genomics & Health



Data Use Ontology

Utility – now and going forward

1. Communication
   Reporting/displaying the ‘profile’ of a dataset’s accessibility
2. Enhance discovery
   Add as extra information to answer ‘discovery’ queries (e.g. Beacon)
3. Automated data sharing
   Align requests to ADA-M profiles, and share if conditions are met
4. Join forces with DUO, to:

Combining Initiatives with DUO

1. ADA-M is based on non-directional CONCEPTS for condition/terms of use

   EACH REQUIRES A DEFINITIVE NAME/CODE AND DEFINITION WHICH COULD BE REVIEWED AND STANDARDISED WITH DUO

2. ADA-M allows a limited set of top-level responses
   ("Unrestricted", "Limited", "Forbidden", "Obligatory", "True", "False")

   EACH REQUIRES A DEFINITIVE NAME/CODE AND DEFINITION WHICH COULD BE REVIEWED AND STANDARDISED WITH DUO

3. ADA-M allows a more granular responses, currently all free text

   BASED ON REAL-WORLD NEEDS, MANY OF WHICH COULD BE BASED ON OTHER ONTOLOGIES AGREED WITH DUO

ADAM-2 (DUAD)

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

```
resourceDataLevel:
            type: string
            enum:
              - UNKNOWN
              - DATABASE
              - METADATA
              - 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
```

ADAM-2 (DUAD)

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

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

ADAM-2 (DUAD) …Simple Example

```
terms: [
        term: {
                data_use_class: DUO ‘time limit of use’,
                restriction_class: {
                        restriction_rule: NO_CONSTRAINTS
                }
        },
        term: {
                data_use_class: DUO ‘publication required’,
                restriction_class: {
                        restriction_rule: CONSTRAINTS,
                        restriction_details: ‘PNAS, Nature, ICJM, Frontiers’
                }
        }
]
```

Building on existing efforts

Global Alliance for Genomics & Health