

# **Global Alliance for Genomics & Health: Automatable Discovery and Access Matrix (“ADA-M”)**

**- - Version 0.21 - -**

The Global Alliance for Genomics and Health (“[GA4GH](#)”) is an international, non-profit coalition of individuals and organizations working in healthcare, research, disease advocacy, life science, and information technology, dedicated to improving human health by maximizing the potential of genomic medicine through effective and responsible data sharing. Its mission is accomplished by “Task Teams” that work to “accelerate progress in human health by helping to establish a common framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data, and by catalysing data sharing projects that drive and demonstrate the value of data sharing.”

## **Automatable Discovery and Access**

The **Automatable Discovery and Access (“ADA”) Task Team** is working to create tools that can facilitate and increasingly automate the discovery and access of various resources (e.g., data, biological specimens, subjects, software). This involves issues of consent and other considerations that regulate the acceptable use of such resources. Core to the group’s work is development of an “**Automatable Discovery and Access Matrix (ADA-M)**” so that consent and other conditions of use can be represented in a standardised, computer-readable manner. This will be accompanied by an API specification and other support and training material.

Making use of ADA-M will reduce the costs and burdens of data stewardship, while increasing the efficiency of responsible data access (discovery and sharing). The matrix acts as a checklist that allows existing regulatory information to be clearly and consistently portrayed, and/or used as the basis for increasingly automated access procedures. In addition to “control statements” of increasing precision, the matrix includes opportunities to list specific terms and constraints. These are for factors for which automation may not be feasible, and where human interpretation of terms and constraints may be needed to ascertain which considerations are relevant as data discoverability and sharing permissions are decided. Over time, these details will also assist in identifying ways to improve the granularity of the matrix itself.

A sub-group of the ADA Task Team, led by Anthony Brookes, has developed v.0.2 of the ADA-M tool. This is now being circulated to all members of the ADA group, for piloting and validation. Next steps include identifying stakeholders and “alpha-testers” willing to further evaluate the matrix with resource stewards, not least by mapping their own data discovery and data access conditions and restrictions onto the matrix.

## **Scope**

ADA-M serves as a means to represent a series of fully independent, elemental variables (conditions, restrictions or limitations on data discovery, access and use), each of which can optionally be specified at increasing levels of granularity. As such, it aims to be a fully flexible and comprehensive

way to state any set of conditions to discover, and eventually access, resources comprising or annotated by data. It is formally pre-structured and designed to carry computable entries per line item in the matrix.

ADA-M does not aim to directly recapitulate the exact limitations found, for instance, in consent forms or similar governance documents. Instead, consents provided by subjects, and researcher- or institutionally-inspired conditions of use, should be able to inform the elements of the matrix. Aligning consent language, types of consent, and conditions of use to the ADA-M elements is therefore necessary for its effective use, and the viability of doing this mapping exercise will be evaluated as part of the initial testing and piloting of ADA-M.

## **Guidance notes for alpha testers**

ADA-M v.0.2 comprises 16 top-level control statements, grouped into three sections (Permissions, Meta-conditions, and Terms of agreement). Values that can be entered for these top-level control statements (line items) include "True" or "Untrue" or "Not Specified". No other value may be entered, and no top-level line item may be left blank. Entering "Not Specified" for any line item implies that no information is available or specified regarding that line item. Entering "True" for any top-level line item denotes that the most liberal of use conditions/consents applies for that control statement, and so nothing beyond this top-level setting need be stated.

Entering "Untrue" for any top-level line item means that a more restricted condition applies. One may stop there, or one may record more refined restriction conditions by completing increasingly precise sub-levels of each line item. Like the top-level, these sub-level line items can be stated only as "True" or "Untrue" or "Not Specified." However, at the final level, there is an "Optional list" which may be used to identify specific terms/constraints. Unlike the other levels where response options are only "True," "Untrue," or "Not Specified," these final levels provide a place to record explicit, detailed information. These details are intended for human interpretation. They are not presently intended to be used for automation of discovery and access processes. Instead, analysis of this information will help to identify variables that guide the development of more comprehensive versions of ADA-M in the future, which may support automation down to this final detailed level. (See below, *Example of a Completed Portion of ADA-M.*)

The intention of each of the 16 control statements is as follows:

### **Permissions:**

This comprises considerations generally covered by, or otherwise included in, consent forms. They represent aspects of control which are useful to know in resource discovery contexts, and also when seeking to actually access a given resource.

"within any country/location" = specifies the geographical regions(s) where uses are permitted

"by all organisations" = specifies the category of organisation(s) for which uses are permitted

"by any category of person" = specifies the category of person(s) for which uses are permitted

"for any domain" = specifies the type(s), or domains of use which are permitted

### **Meta-Conditions:**

This comprises considerations that sit outside the realm of typical subject-level consent and terms of agreement.

“There are no other restrictions/limitations in force which are not herein specified” = provides a way to indicate whether additional usage constraints apply, not covered by this matrix

“No special evaluation required for access requests involving sensitive/restricted populations” = provides a way to indicate whether extra review procedures, typically manual, are mandated for proposed uses involving sensitive/restricted populations

“Identical consent permissions have been provided by all subjects” = provides a way to indicate whether consent and conditions differ between subjects whose records are included in a given resource (e.g., a dataset) without having to specify those differences on an individual subject basis

### **Terms of agreement**

This comprises considerations generally stipulated by researchers, institutions, regulations, or IRBs as a condition of granting access to, or continued use of a given resource. They may be useful to know in resource discovery contexts, but particularly assist when actually seeking to access a given resource.

“There are no requirements for any formal approval, contract or review conditions to be satisfied” = relates to the need or otherwise for some form of review, approval and/or contract condition, such as a legal agreement, IRB/IEC review, Data Access Committee (DAC) review, etc.

“There are no requirements regarding publication or disclosure of derived results” = relates to the need or otherwise for specific agreements regarding the publication and presentation of new findings and analyses as a result of access to, and use of a given resource

“There are no timeline restrictions” = relates to whether or not time constraints apply to when the resource may be accessed or made use of

“There are no requirements regarding data security measures” = relates to the need or otherwise for certain hardware, software, or other data security-related procedures to be in place in order to access and use the resource

“There are no requirements regarding withdrawal, destruction or return of any subject data” = relates to the need or otherwise that certain procedures will be followed regarding expunging provided information under particular conditions

“There are no restrictions regarding the linking of accessed records to other datasets” = relates to the acceptability or otherwise that accessed information may be integrated with, or linked to, other related information

“There is no possibility of recontacting data subjects” = relates to the whether or not, and under what circumstance(s), recontacting of subjects whose data or biospecimens are stored in a given resource may be permitted or required (such as for the management of incidental findings)

“There are no restrictions regarding intellectual property claims based on use of the accessed resource” = relates to the need or otherwise for an agreement as to how new intellectual property rights are allocated and controlled

“There are no requirements to report back regarding use of the accessed resources” = relates to the need or otherwise for some type of mandatory reporting to resource stewards, or others, regarding the use of the accessed resources

“No fees will be levied for access of the resources” = relates to whether or not financial fees or other charges will be levied to access and/or use a given resource

The matrix contains separate sections where the above answers can be entered to specify how a resource may be discovered or shared, as these two types of access are not equivalent. For example, a resource may be approved for discovery by anyone, but be restricted for use only by non-profit organisations. These ‘discovery’ and the ‘sharing’ sections of the matrix additionally both have separate sub-sections for ‘summary/aggregate’ data and for ‘record level’ data, enabling different usage conditions to be stated for these different forms of a dataset, in the discovery and/or the sharing context. The matrix will probably most commonly be used to state access conditions that apply to a whole datasets (i.e., a group of records), whereupon the information portrayed in the matrix can be managed as metadata for said resource. But there may be situations in which different research subjects have provided different consents/permissions, and the matrix allows for this by allowing the ‘Permissions’ section to be completed separately for each individual subject’s record in a given dataset. In contrast, the ‘Meta-Conditions’ and ‘Terms of agreement’ sections apply only to whole datasets, and so these are not filled in if completing the matrix for individual subjects. In situations where consents do differ between subjects and yet it is not practical to complete separate matrix columns for each one, then the fact that consents vary may be easily represented in one of the meta-condition items. Finally, any columns of the matrix that are not needed in any particular situation can be deleted and not used.

## **Alpha-Testing, Questions, and Feedback**

The ADA Task Team is hoping that many different groups, operating in different data/resource discovery and sharing settings, will consider whether or the not the ADA-M might be useful to them now or in the future. Any and all feedback about its suitability or otherwise will be gratefully received, and should be sent in the first instance to the group’s coordinator Emily Kirby [ekirby@p3g.org]. Questions about the ADA-M, and requests for help in piloting the matrix should also be directed to this same address, to be forwarded to members of the ADA team most able to assist. Individuals that might be interested in helping to develop an API to support the ADA-M should not hesitate to propose themselves as new members of the ADA Task Team.

### Example of a Completed Portion of ADA-M

The resources are permitted for use:	DISCOVERY	
	Aggregate/Summary Data	Record-level Data
within any country/location	Not Specified	Not Specified
<i>within specified countries/locations</i>	<i>*left blank*</i>	<i>*left blank*</i>
<i>within any country/location other than those specified</i>	<i>*left blank*</i>	<i>*left blank*</i>
by all organisations	Untrue	Untrue
by any non-profit organisations	True	True
<i>by specified non-profit organisations</i>	<i>*left blank*</i>	<i>*left blank*</i>
<i>by any non-profit organisations other than those specified</i>	<i>*left blank*</i>	<i>*left blank*</i>
by any profit organisations	Untrue	Untrue
<i>by specified profit organisations</i>	<i>*left blank*</i>	<i>*left blank*</i>
<i>by any profit organisation other than those specified</i>	List #1	List #1
by any category of person	Untrue	Not Specified
by any category of academic professional	True	Not Specified
<i>by specified categories of academic professional</i>	<i>*left blank*</i>	<i>*left blank*</i>
<i>by any category of academic professional other than those specified</i>	<i>*left blank*</i>	<i>*left blank*</i>
by any category of clinical professional	True	Not Specified
<i>by specified categories of clinical professional</i>	<i>*left blank*</i>	<i>*left blank*</i>
<i>by any category of clinical professional other than those specified</i>	<i>*left blank*</i>	<i>*left blank*</i>
by any category of profit-making professional	Untrue	Not Specified
<i>by specified categories of profit-making professional</i>	<i>*left blank*</i>	<i>*left blank*</i>
<i>by any category of profit-making professional other than those specified</i>	List #2	<i>*left blank*</i>
by any category of non-professional	Untrue	Not Specified
<i>by specified categories of non-professional</i>	<i>*left blank*</i>	<i>*left blank*</i>
<i>by any category of non-professional other than those specified</i>	<i>*left blank*</i>	<i>*left blank*</i>

This example completed section of ADA-M tells us that:

**For Aggregate/Summary Data** - it is unspecified whether or not aggregate/summary data (i.e., summative forms of data, not including subject-level information) can be used within any country or location. It can be used by any non-profit organization, but cannot be used by profit organizations included in List #1. It can be used by any category of academic professionals, by any category of clinical professionals, but only by profit-making professionals included in List #2. It cannot be used by any category of non-professionals, but there is no further information given as to whether or not there are specific categories of non-professionals which would be either allowed or disallowed.

**For Record-Level Data** - it is unspecified whether or not record-level data (i.e., forms of data comprising subject-level information) can be used within any country or location. It can be used by any non-profit organization, but cannot be used by profit organizations included in List #1. Information about use by academic professionals, clinical professionals, profit-making professionals, and non-professionals is not specified for record-level data.

*[nb. Within this figure, "List #1" and "List #2" are entered merely as short-hand abbreviations. This approach should not be taken when actually completing ADA-M. Instead, the relevant restrictions/permissions should all be listed and elaborated using free text within appropriate box in the matrix.]*