Data Access Requests
Where can we standardize?

DURI & X-WS

Date: Tuesday, 15 November 2022
Time: 14:00 - 15:30 UTC
Meeting Chair(s): Jonathan Lawson & Melanie Courtot
Invited Google groups: DURI, REWS, Clin/Pheno, Ethical Provenance, Computable Cohorts, DACReS

Abstract
Currently, a researcher wishing to gain access to multiple datasets must submit individual data access requests (DAR) to each data access committee (DAC). This is a cumbersome system and acts as a barrier to data reuse. A central portal that issues the requests to all the relevant DACs would be much easier. However, there are a number of challenges to such a portal, least of all the huge investment that would be necessary in its creation and maintenance. Instead, the goal of this workshop is to decide on which parts of the overall data access request process are amenable to automation and/or standardization. For example, harmonized modular DAR forms that can be reused/reissued across platforms or streamlining the back-and-forth assessment process for DACs. In doing so, this session hopes to pick up on a number of discussions and efforts across GA4GH, including the Ethical Provenance Toolkit work including DACReS, and the Computable Cohorts DARathon. The purpose of this session is not to try to solve any or all of the challenges in this space, only to agree on a set of tasks that the group believe could be achievable in a medium timeframe. Finally, it is important to have input from both technical and policy experts across the data access ecosystem in order that any planned efforts are considered in the proper context.

Attendees: Sarion Bowers (Wellcome Sanger Institute), Romina Royo (BSC), Kurt Rodarmer (NIH), Soichi Ogishima (ToMMo, Tohoku Univ, GEM Japan), Marietjie Botes (University of Luxembourg, EU), BF Francis Ouellette, Aina Jene (EGA-CRG), Mallory Freeberg (EMBL-EBI, EGA), Andreas Bruns (GHGA, University Hospital Heidelberg), Christoph Schickhardt (German Cancer Research Center), Alma Karibo (Data Privacy Professional Health Research; AGC 54gene), Alaa Youssef (Organizational Readiness for AI and Ethics Research - Stanford School of Medicine), Shu 'Sue' Hui Chen (US National Institutes of Health (NIH)), Ray Krasinski (Philips Healthcare), Brendan Behan (Ontario Brain Institute)

	Agenda Item	Speaker	Time

1	DACReS Survey - short summary	Jonathan Lawson	15 min
2	Review example DAR forms - DUO as Data Use Limitations?	Discussion	25 min
3	Define objectives & open questions	Discussion	50 min

Links & Notes

ICGC-ARGO Data Access Compliance Office (DACO)
Resource Entitlement Management System GitHub & Live Demo page
Folder with example DAR application forms here.
Zoom recording
https://us02web.zoom.us/rec/share/BTTze9_kd_d29sVRWVgV9NV7KdzMJfYT05Rc0aCAj27tiWm_w_HTWB5uHsAmhv4x.c6OJkLwjT9JxMH1B

Open Questions

The definition of a DAC and its responsibilities differs between institutes and countries. Can a modular DAR address all or most of their concerns?
Do form inputs need to be computable?
How can we standardize the “submission form”? Could a researcher submit the same initial statement to numerous platforms.
It is unlikely that there will be a global standard for legal agreements, but how can we facilitate the process of creating each individual legal agreement?
Including DUO codes in a request makes it easier to match up with a DUL, but can you automate this entirely?
Basic information is a good starting point - who is applying, what data they need and what they plan to do with it.
Binary questions are helpful to exclude data sets quickly - e.g. “Are you going to do ancestry research?”
Designing and sharing forms. Can we agree on a format for exchange e.g. JSON instead of PDF?

- https://github.com/jsonform/jsonform
- https://gcanti.github.io/resources/json-schema-to-tcomb/playground/playground.html
- https://github.com/eclipsesource/jsonforms

Key Takeaways

- Guidance for drafting Data Access Request forms would be useful and give DACs “something to point to” to explain why they require particular information

- Guidance on the particulars of universal rules/requirements, such as how to make a PI accountable, would have the biggest impact
- There is considerable overlap between DAR forms, so identifying a core set of “universal” rules shouldn’t be impossible
- The MRCG already exists, how do we bridge this to a DAR
- There may be a link between DUO and DAR templates
- Structure DAR questions similar to DUO hierarchy
- What are the human readable questions that map to DUO?
- These need to be categorised as legal or ethical questions.
- Creating standardized text/rules/definitions needs to be first and before any technical solutions
- This project needs to connect DURI and REWS efforts
- Is there a parallel with the Universal MTA effort that is in widespread use?
- Need to clarify differences between Data Access Agreement, Data Access Requests, DAC etc.
- This can also create incentives for data sharing through adoption of DAR types supported by funding agencies see https://doi.org/10.1007/s11192-022-04361-2
- Standarizing upstream of DAR - at consent/collection level helps minimize these issues.

Minutes

JL: Presentation - *slides here*.

- Controlled Data Access
- Data access requests submitted to DAC, who review and approve/deny
- 2-3k DACs globally, but no governing body across them: individual management of requests
- Expected: some similarity
- Poll of 10 different DAR forms across GA4GH - US/UK/Canada/Australia/Finland
    - Who? What? Legal/Compliance
- Potential aspects that can be standardized.
- Not expecting to mint all of these standards
- Who are we benefiting?
- Do we think new DACs will seek advice?
- Do we think existing DACs will adjust their forms?

Poll:

First question from JL: What is beneficial about standardizing?

SB: Would help to understand if this is guidance or future tech or standards that could use a standard DAR process?
JL: We know that hopefully most folks are hoping not to build PDFs. GA4GH is not planning to build directly, more standardization. What should the actual human words in the DAR be? Not tech, more the questions?
SB: Recommendation of what a good DAR looks like?
JL: Yes, if there is agreement.
VV-F: Expectation as a researcher. We can rely on GA4GH - this is why we're asking, rather than how dare you ask me that. Gives some foundation for asking
SHC: Glad to hear this. More of a guidance - legalese. An ethical question - based on informed consent. Is it going to be a standard form - like a universal MTA? Institutions can sign onto, but we need some flexibility to add terms. If this is a template, or is it language guidance? At NIH - we want to make sure we apply CARE principles.
JL: Best practices. Back to MRCG - similar approach.
AB: We talk about standardize vs not, as if it's a trade off. The institutions know the difference. We need to make sure we follow the rule. The value comes out for those situations where a rule is needed, but a group is uncertain what it should be. An example, our DAR what info should be gather about PI to make they are accountable for. Don't really care about the detail, but allows smaller groups to running a DAC cost effectively, but meets community norms.
TN: From Finland - when we have support from biobank DACs. About 80% overlap between DAC questions. Definitely need to collect more of these forms. Some will be different. In Europe - identity vetting is from federated identity management. Provides legal affiliation. In US - legal official provides this.
MC: Same shared info. Bridge between MRCG and request submission. Multiple submissions would be easier. Could link with Passports & DUO. Interested in implementation. If there is a list of terms and attributes across DACs - We all aspire to have electronic DAC request systems. Who? What? Legal - these are definitely the starting point for formalising. This is step 1 and then step 2 is the technical implementation.
JL: Two tiers - creating text - boilerplate. PDF if someone wanted. On top of that fixed unique questions but agreed across countries. Need first before second.
MC: Agree - PDF best practice, but then use that to build a consensus form. Retrospective mapping and prospective mapping.
SB: I can see these two things being compatible. If you adopt A you can join B eventually, but need to keep them separate. Policy reasons for not joining.
MC: Overlap with REWS and DURI. A little bit of this is the best practice and then an implementation
SHC: I hear a lot of people want to hear automation. Informed consent language builder in some institutions - some are non-negotiable. Not sure how big a heavy lift that would be. Guidance that has been vetted by institutions worldwide would be useful. How could we make it into a standardized format - than a standard form. No need to read every single question. Make it

easier for investigator and institution. How can we facilitate that? Universal MTA - worldwide sign-on. Easy to execute. No back and forth. How can this can be built into DACs?
JL: Two DACs agree on a same set of questions - if one completes the form - it can trust the information from both. Eventually - Melanie's DAC could trust mine and vice versa. Similar to a single IRB review.
SHC: Has to be in alignment. Or even a similar review, less process.
FP: NCI DAC - Thinking about a DAC is doing a review - if you have 100 requests - any study that meets that DUL - you can grant anything less stringent. Reduces DAC burden and requester burden.
JL: Visionary approach. Have toyed with this idea at the Broad. More about how a DAC operates than a DAR question.
FP: If questions are grouped together - ancestry research? yes/no? Etc. Auto excludes studies. Do you intend to combine similar diseases? HMB/GRU limited etc. DAR form connects to DUL.
JL: Allow access to data post approval. But is a more broad and open sharing. DACs may shy away from that. Vaso/Ted? Good question for DACReS group.
VR: Details about the DACReS meeting - toolkits. December 2nd 11am EST. Nuance to this. Procedural benefits to batch processing, but there is also responsible ethics governance. We assume this was explicit in the consent form, but that is not always the case. Machine assisted interpretation, how do we represent the patient consent?
AB: Comments back - tension. Two things - tools to build their own DACs. Here are things you can use. There is an interest in policy standardization. Trying to do both things. But fundamentally different goals. The tool should be the first priority, but should eventually lead into interoperability. BUt not sure how to achieve that.
TN: Automation vs standard vocabularies. 1+MG - we have the DUO standard. Enables allowed Data Use Limitations to be encoded, but what are the questions we need to ask to codify that? The questions could be implemented in underlying technical structure.
JL: What we are saying is that - you could standardize the questions, so you know you're asking the same questions. Engage researchers separately, but ask the same questions. Not trying to get lawyers from every country to agree.
SG: Two categories of Q. Ethical or Legal requirement - must ask. Are you happy for it to be a commercial entity? Important. Tackle the first set of questions first.
JL: Follow up to this to submit your DAR forms. An actual TXT or PDF would be better.
AY: I agree the IRB/DAC line is critical. If we can come up with a framework that a legal/ethics might request, we can come up with a template. I came across this in interviews - number one barrier is for researchers to convince the compliance officers. Matrix of risks by legal.
YJ: I have a concern with this exercise. I have seen this compilation of DAA - I have published this myself. Will be painful given number of them. I would pick reviews on what has been out there. Not recommend exhaustive gathering access forms. I think this will be duplicating. Metareview.
JL: Two distinct things. DAA (and aliases). I agree this has been covered, DACReS/REWS folks are working on. This is more which questions go in the PDF from the researcher.

YJ: Often integrated in a single document. Less trouble in jurisdiction compatibility. Still check what's out there first. Don't request from GA4GH, as it may not get much response. But going to a technical methods are a good idea.
JL: I agree - point well taken. Agree on DAR form questions, but as we engage you may also need to be an advocate for adopting this.
MC: +1 Yann. The less work the better. More of call for people to share documentation or source of information. DAA has been covered. Try to look at how these meet in the middle. We require a lay summary on our DAR.
JL: Summary. DAA legal work - good to build on. As we think about the DAR form itself, how to handle various ones of those in a technical fashion. A lot of potential in standardizing the DAR form questions. Facilitating researchers for them to apply to multiple DACs with one form, maybe even in one place. This is talking with DACs about making things easier for researchers.
TD: Developers of data sharing platforms. Members of funding agencies - how DAC are organised. Own proposition on standardised DAR - oriented more at how we can use this information to create incentives for data sharing. Funding agencies can extract information from standard DARs.
MC: Reminds me about monitoring. A dashboard of what type of usage is happening. Possible use case?
TD: Insight into data usages would be very valuable. Where DACs are less formal across Europe. My institute has no DAC - I decide on it myself. If you would be collecting data usage - to let DAR go to different DACs depending on purpose and expertise, but then you must standardize data usages you might expect. This could be tricky.
SB: Where we get into DAC whether they should be doing ethical review. We view that as not the DAC function. Goes back to original consent. At Sanger most consent forms are standard - limited. Trying to avoid bespoke DAR form - as we have standardised upstream of DAR.
JL: Point well taken. Start this exercise. Primary issues that DACReS group came up with. How ethically involved is the DAC. Spectrum. Check DUL up to second IRB. The questions that are more ethical will spark questions/discussions.
SB: You can tackle some of those things upstream.

References

Thijs Devriendt: This is the paper I wrote about this topic before:
https://link.springer.com/article/10.1007/s11192-022-04361-2

Summary

- Interest in standardizing to
    - Have shared understanding of the info to be submitted
    - Bridge between MRCG and actual request submission
    - Have shared form when multiple biobanks (eg Japan, FInland) requests
    - Standard implementation to link with passport
- Concerns about
    - Legalese
    - Implementation or best practices?
    - Trying to do 2 things at once
    - Conflating DARs and DAAs
-

Action Log & Objectives

1. Who would like to be active in this?
2. Request DAR form
3. Follow up meetings
4. Folks who could dedicate leadership to this
    a. Compiling agreements
5.

Zoom Chat Log

14:05:34 From Fabio Liberante To Everyone:
https://docs.google.com/document/d/18crzlhs4-eAME91q2q3skqLmWrB5EF3HX8ekDm2EEqw/edit?usp=sharing

14:12:15 From Shu Hui Chen To Everyone:
just wanted to make sure that the "standard" is a guide as there are regs that affect the terms
14:14:26 From Kurt Rodarmer To Everyone:
The question (I think) was about whether the standard was around technology or guidance
14:14:39 From Marietjie Botes To Everyone:
Standardising DAC forms will save negotiation time and legal costs, saving funding money to be spent on research. Technology can be applied to allow for a system that can fairly "personalise" such a form ….
14:15:14 From vasiliki rahimzadeh To Everyone:
But those implementation considerations is something the DACReS group take on through consultative discussions and research we recently undertook in the global survey
14:16:01 From vasiliki rahimzadeh To Everyone:
^*could take on
14:16:08 From Melanie Courtot To Everyone:
If there is benefit we'd want to formalize these words - similar to the DUO work, which already includes DAR relevant terms
14:17:31 From Spencer Gibson To Everyone:
I agree having a standard list questions (or at least a minimum set) will allow requesters ti find out the information before making a formal request. That will allow ethics bodies in institutions to formalise their responses to minimise any inconstancies between differing applications between applications from the same institutions between differing projects
14:19:04 From Soichi Ogishima To Everyone:
I agree.
We have standardized request forms and developed a common web form for 14 major biobanks in Japan. This is because it is burdensome for researchers to fill out different request forms for different biobanks. We provide a system to search data across biobanks, which is another reason why requests across biobanks are increasing.It would be ridiculous for a researcher to fill in a different form for each biobank for a similar request. We have developed the minimum common elements required for a request. At this time, for example, we organized the similar terms "principal investigator" and "researcher representative" as synonymous. We are maintaining a kind of ontology. It is now midnight (Japan time), so I am going to bed, but I would very much like to contribute to this activity from GEM Japan.
14:19:24 From Shu Hui Chen To Everyone:
Is the idea to create something similar to the the Universal MTA that can be signed on by institutions or guidance and allowable for flexibility?
14:20:00 From Pinar Alper To Everyone:
Form elements that help categorise the data user and their particular use will be beneficial. Especially for institutions/DACs establishing a form for the first time.
14:20:41 From Stacey Donnelly To Everyone:

I think people will appreciate templates and guidance on best practices.
14:22:02 From Shu 'Sue' Hui Chen To Everyone:
I'm pretty sure that had a draft example of DAC Charters at one point that can help with setting up de novo DACs.
14:23:41 From Mousumi Ghosh, NCI, NIH To Everyone:
I'm fairly new here, so please excuse the question. Do we need to ask each requester all the same questions? Why not ask only the questions based on the constraints in datasets they are requesting?
14:26:16 From vasiliki rahimzadeh To Everyone:
If one of the concerns is bespoke data use restrictions, might it be useful to embed a mechanism for community or crowdsourced contribution of data use terms that can be built into a standard form?
14:26:20 From Sarion Bowers To Everyone:
The purpose, as I see it, is to try and harmonise and present good practice for DACs as there's currently a multiude of practice and no consistency on what constitutes a DAR that provides due diligence while maximising access to data.
14:27:11 From vasiliki rahimzadeh To Everyone:
^+1 to Sarion
14:27:39 From Shu 'Sue' Hui Chen To Everyone:
@Mousumi, there are information needed to capture that is required because the Data Access Request is essentially part of an agreement (contract) between the institutions for sharing the controlled-access data as well as enough information that can allow the Data Access Committee to assess the ethical appropriateness of use of the data.
14:29:33 From Soichi Ogishima To Everyone:
Great discussion. I will check the discussion later. Good night. :-)
14:29:56 From Mousumi Ghosh, NCI, NIH To Everyone:
@sue: yes-maybe a two tier system. Once the first set of criteria is fulfilled, maybe a more specific DAR form that is generated based on specific restrictions in dataset
14:31:38 From Jonathan Lawson To Everyone:
@Mousumi - our work with the NHGRI DAC led us to develop a dynamic DAR form that asks a base set of questions, and then populates others based on the DULs
14:33:51 From Melanie Courtot To Everyone:
My DAC may want review due to legal requirements - yes @Shu
14:34:02 From Spencer Gibson To Everyone:
I apologise if this is old news, but is there any potential benefit in have some standardised terms with agreed and clear definitions, even beyond the questions, to ensure that terminology as well as the questions are standardised and interpreted consistently (I think Johnathan just said something similar though)
14:34:10 From Melanie Courtot To Everyone:
But at least they know they're getting the right thing in the right place
14:34:18 From ying huang To Everyone:

Thanks, Sue! Totally agree with your comments. my two cents, when it comes to automation, the upper string structures/rules are very important for the build clear logics into the flow. Also, could we utilize the consent groups to direct structured DAR template?
14:34:28 From Yann Joly To Everyone:
There is an existing harmonized european data access agreement we could build from that: https://ega-archive.org/submission/dac/documentation
14:34:43 From Mousumi Ghosh, NCI, NIH To Everyone:
@Jon glad to hear; I guess that is what I am going for. Are there any specific requirements in the system for that to happen?
14:37:14 From Melanie Courtot To Everyone:
@Freddie - you could even grant prospective access to *new* datasets being generated *after* their request
14:37:55 From Sarion Bowers To Everyone:
Just to be clear - the "harmonised" DAA is the result of comparing a number of DAAs (and is actually based on the example DAA also provided at that link). It's useful to see the outcome of comparisons but I think the word harmonised can be misleading in as much as it's not authorised or approved by any European body.
14:38:01 From Melanie Courtot To Everyone:
Like that description @Jonathan “data use profile of a researcher”
14:38:53 From Shu 'Sue' Hui Chen To Everyone:
I think what Freddie mean is that it could be be built in to allow that many DARs could be grouped so only one action can be done to cover multiple actions.
14:39:54 From Melanie Courtot To Everyone:
Lindsay: Lindsay.smith@ga4gh.org
14:40:18 From Shu 'Sue' Hui Chen To Everyone:
+1 Melanie re “approve new datasets that have the same DUL after the initial approval”
14:40:37 From Lindsay Smith To Everyone:
Thanks Melanie! To all - please don't hesitate to reach out via email
14:42:08 From vasiliki rahimzadeh To Everyone:
GA4GH DACReS Working Group Meeting
Please join on Friday December 2nd @ 8am PT / 11am ET / 4pm GMT for a DACReS Working Group Meeting to debrief on outputs from the REWS/DURI workshop as well as discuss next steps for further deve

Please join on Friday December 2nd @ 8am PT / 11am ET / 4pm GMT for a DACReS Working Group Meeting to debrief on outputs from the REWS/DURI workshop as well as discuss next steps for further developing the DACReS toolkit.
14:42:09 From vasiliki rahimzadeh To Everyone:

https://us02web.zoom.us/j/4426465151?pwd=VXp4djBjRmN4YWw4cGNZUkQrSGlJdz09
14:42:26 From ying huang To Everyone:

When we thinking about standralize the DAR review, we could start with standralized DAR template (modules) recommended by GA4GH, by that, the reviewers at different institute would not have to have the SAME review process/system, rather, will receive the SAME DARs.
14:42:33 From Yann Joly To Everyone:
Sarion, RE: DAA it is promoted by both EGA and the EU (given that it stems from one of there project. However for legal reasons they will not formally authorize/approve which is to be expected. I'm mostly suggesting it as a starting point as I prefer to build on exisitng ressource rather than duplicating.
14:42:47 From Shu 'Sue' Hui Chen To Everyone:
In addition to Informed Consent, IRB/EB decisions on data sharing terms/restrictions are very important in addition to the institution compliance officer
14:45:46 From vasiliki rahimzadeh To Everyone:
Totally agree Sue, some of my other empirical work shows however that IRBs frequently lack members with expertise in data management/privacy needed to review data security specs and use terms.
14:46:43 From vasiliki rahimzadeh To Everyone:
Thinking ahead, the coordination between these two ethics oversight bodies (IRBs + DACs) is really critical
14:47:11 From Adrian Thorogood To Everyone:
is this a standard DAR form, or a standard "electronic Data Access Request form template"? where the fields actually requested will be populated based on the datasets requested? This seems like it could be a more comprehensive but also flexible approach.
14:52:10 From Francis Ouellette To Everyone:
+1 for @yann
14:53:31 From Melanie Courtot To Everyone:

https://docs.google.com/document/d/18crzlhs4-eAME91q2q3skqLmWrB5EF3HX8ekDm2EEqw/edit#
14:53:42 From Fabio Liberante To Everyone:
Yes - permissions are open
14:54:09 From Alexander Bernier To Everyone:
Thijs Devriendt (on the call) had done a great set of interviews with DAC members and DAC-adjacent staff - might be able to speak well to the state of the literature
14:54:16 From Thijs Devriendt To Everyone:
I think a DAR form with common categories as Jonathan showed in the beginning would be very helpful for various purposes :) I personally do not think that legal terms, like those collected in DAAs, need to be integrated into this.
14:54:32 From Jonathan Lawson To Everyone:
+1 Thijs
14:54:55 From Maili Raven-Adams To Everyone:

In the Ethical Provenance group we are looking at DAA's so maybe we can all work together in tandem

14:55:45 From Melanie Courtot To Everyone:

Not legal terms (I'm hoping we can abstract from legal ;)) but eg our DAR requires IRB approval, which I believe would also appear in the DAA - so there is overlap between both?

14:57:48 From vasiliki rahimzadeh To Everyone:

Please share your insights if you are able @Thijs!

15:01:06 From vasiliki rahimzadeh To Everyone:

+1 Sarion

15:01:31 From Shu 'Sue' Hui Chen To Everyone:

+Sarion

15:01:51 From Thijs Devriendt To Everyone:

This is the paper I wrote about this topic before: https://link.springer.com/article/10.1007/s11192-022-04361-2

15:03:02 From vasiliki rahimzadeh To Everyone:

Merci Thijs!

15:03:05 From Melanie Courtot To Everyone:

Thanks @thijs - added to the minutes under 'References' section

15:04:03 From Yann Joly To Everyone:

This is the one we wrote compiling agreement from IHEC : https://www.nature.com/articles/s41597-019-0310-4#Sec2

15:04:06 From Freddie Pruitt To Everyone:

Those questions would be based off the DULs

15:05:44 From Melanie Courtot To Everyone:

Feel free to email Fabio to join: Fabio Liberante <fabio.liberante@ga4gh.org>

15:06:36 From Sarion Bowers To Everyone:

Sorry was pre-emptively waving goodbye!

15:06:48 From vasiliki rahimzadeh To Everyone:

Merci Jonathan!

15:06:51 From Brendan Behan (Ontario Brain Institute) To Everyone:

Thank you!

15:06:54 From Shu 'Sue' Hui Chen To Everyone:

Thanks everyone. Looking forward to "seeing" you in other sessions!