

ARTICLE

Registered access: a ‘Triple-A’ approach

Stephanie OM Dyke^{*1}, Emily Kirby², Mahsa Shabani³, Adrian Thorogood¹, Kazuto Kato⁴ and Bartha M Knoppers¹

We propose a standard model for a novel data access tier – registered access – to facilitate access to data that cannot be published in open access archives owing to ethical and legal risk. Based on an analysis of applicable research ethics and other legal and administrative frameworks, we discuss the general characteristics of this Registered Access Model, which would comprise a three-stage approval process: Authentication, Attestation and Authorization. We are piloting registered access with the Demonstration Projects of the Global Alliance for Genomics and Health for which it may provide a suitable mechanism for access to certain data types and to different types of data users.

European Journal of Human Genetics (2016) 24, 1676–1680; doi:10.1038/ejhg.2016.115; published online 28 September 2016

INTRODUCTION

The past decade has witnessed an increase in international data sharing across biomedical research consortia spurred on by funders and journals to make research data available as rapidly as possible and forced in part by the need for extremely large data sets to detect patterns of health and disease.¹ The Global Alliance for Genomics and Health (Global Alliance²), an international coalition dedicated to improving human health by maximizing the potential of genomic medicine through effective and responsible data sharing founded on its *Framework for Responsible Sharing of Genomic and Health-Related Data*,³ is illustrative of this international drive.

Most public research data resources in genomics have both open and controlled access categories. While open access is typified by the HapMap⁴ and 1000 Genomes projects,⁵ controlled data access is used, for example, by the International Cancer Genome Consortium,⁶ with some data stored in the Database of Genotypes and Phenotypes⁷ or in the European Genome-phenome Archive.⁸ A controlled access system mandates review by a Data Access Compliance Office (DACO). Although the use of controlled access has been successful in providing greater access to data, plans for greater integration of data sets and informatics platforms for data-intensive science might well be thwarted in the absence of a more intermediary category that would allow easier access to some data hitherto categorized as ‘sensitive’ and thereby controlled without further qualification or nuance.

Within the Global Alliance, we are developing the concept of ‘registered access’, a novel data access tier that would fall between the now well-established ‘open access’ and ‘controlled access’ (also referred to as ‘managed access’) tiers.^{8–10} While not eliminating the need to control access to sensitive or identifiable data, our aim is to expand the currently binary open/controlled approach to protect the privacy of participants and patients and at the same time further the research to which they are contributing their data in a more proportionate manner. We are also focused on responding to the needs of the Global Alliance ‘Demonstration Projects’, scientific initiatives that are being

accelerated to demonstrate the value of data sharing, namely: the Beacon Project (<http://www.ga4gh.org/#/beacon>), Matchmaker Exchange¹¹ and the BRCA Challenge (<http://brcaexchange.org>). The need for an intermediate category of data and an intermediate data access tier stems from two main considerations. First, the controlled access mechanism is considered too onerous and lengthy a process for access to some types of data that are being shared and brought together by the Global Alliance Demonstration Projects, but that nonetheless do require a level of protection for reasons of privacy. Second, and along similar lines, the degree of oversight required of researchers using controlled access data sets is greater than we envisage would be justified within such a tier for researchers, clinicians and others who may need access to this registered access data. A new registered access tier offers the prospect of enabling rapid access for a wide range of users to all data shared in this way.

Several genomic projects and databases have made use of registration-based systems for access to data. These include the Asthma Gene Database, MedGene and PharmGKB,¹² projects participating in the Matchmaker Exchange project such as DECIPHER¹³ and PhenomeCentral¹⁴ and, more recently, the Simons Foundation Autism Research Initiative (<https://www.nextcode.com/ssc/>). Further development of such approaches to data access was recommended by experts participating in the National Human Genome Research Institute workshop on establishing a central resource of data from genome sequencing projects in 2012.¹⁵

The Registered Access Model that we describe here is based on our analysis of applicable research ethics and other legal and administrative frameworks. Its approval process would be considerably simplified compared with controlled access in that some of the multiple steps of the standard controlled access review procedure would either be streamlined or removed. These include, for example, undergoing additional scientific and ethics review. We thereby propose a three-stage approval process for registered access comprising an Authentication, Attestation and Authorization.

¹Centre of Genomics and Policy, Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, QC, Canada; ²Public Population Project in Genomics and Society (P3G), Montreal, QC, Canada; ³University of Leuven, Leuven, Belgium; ⁴Department of Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka University, Osaka, Japan

*Correspondence: Dr S Dyke, Centre of Genomics and Policy, Department of Human Genetics, Faculty of Medicine, McGill University, 740 Dr Penfield Avenue, Montreal, Quebec, H3A 0G1, Canada. Tel: +1 514 398 6638; Fax: +1 514 398 8954; E-mail: stephanie.dyke@mcgill.ca

Received 18 April 2016; revised 28 June 2016; accepted 19 July 2016; published online 28 September 2016

LIMITATIONS TO CONTROLLED ACCESS

We start by considering the general criteria that are usually checked by Data Access Committees (DACs) and DACOs in the controlled access process and reflect on their impact on data access. These criteria are listed in Table 1 and require a combination of information provided by applicants (see Supplementary Table S1) and assurances provided by the applicants' host institutions, which assume legal liability for the applicants' use of controlled access data.

Different types of DACOs exist and may have varying roles, depending on their available resources, the area of expertise of members and the size and nature of the data resource they relate to. For instance, the Public Population Project in Genomics and Society offers DACO services that offer the creation of customized DACOs with the resources and policies required to ensure a complete review of applications for access to controlled data sets, in conformity with the goals and policies of the project, as well as the research participants' consents. However, in some cases, DACOs may operate on more limited resources and therefore encounter certain limitations to their controlled access review.¹⁶ Furthermore, some of the steps of controlled access review are associated with challenges, and they may not be necessary for all data access reviews.

In principle, given the non-exhaustible nature of data, it can be argued that a minimal set of criteria should be envisioned to foster more rapid access to and use of data sets. In this regard, depending on the sensitivity of the data, the necessity of reviewing the scientific merits of research proposals by DACs is questionable. Indeed, funding or research organizations are better positioned to carry out scientific review of research proposals. With the exception of a few large institutes, DACs are often operating on limited financial and human resources, rendering a thorough scientific review difficult if not impossible. Furthermore, in the absence of clearly delineated criteria and procedure for such reviews, the objectivity of decision making for data access could also be undermined.^{17,18}

The controlled access model can also serve to prevent controversial research uses through DAC review of research proposals.¹⁹ Culturally or politically sensitive topics are mentioned as conceivable yet not frequent examples of controversial research uses.¹⁶ One can claim such review falls within the scope of ethics review, a task outside the remit of DACs in general. DACs often refrain from adding another layer of ethics review, seeing it as a responsibility of the data users to satisfy the requirements for ethics approval.²⁰ To this end, DACs sometimes require an official ethics approval document from home

institutes,²¹ which have an effective role in ensuring research conducted in their facilities has received ethics approval from competent bodies. The scope of proposed data uses is also subject to review to ensure consistency with the data provider's objectives and policies and with the original consent of research participants.²² Reviewing this scope is not always straightforward. For example, DACs do not always have access to the consent forms that were used or sufficient resources to interpret them when needed.¹⁶ Alternatively, data-use limitations could be more explicitly stated in consent forms and articulated within ethics approvals for data collections. Ethics committees could have a role in controversial cases or when there is ambiguity. Consent-based conditions of data use could also be more clearly conveyed to data users with the use of standardized consent codes.²³

REGISTERED ACCESS AUTHENTICATION AND AUTHORIZATION

Bearing in mind these limitations in the context of controlled access, we propose that the review process in registered access would mainly be concerned with the qualifications of applicants for access to data. This level of review would require an assessment of the likely ethical and legal risks of data misuse (based on consent, identification risk and data sensitivity). For example, for the Beacon Project and Matchmaker Exchange, data uses may be constrained by the way in which, and how much, data can be queried. Therefore, several controlled access review criteria addressing ethics review and grouped under 'Ethics' in Supplementary Table S1 may no longer apply. Our Registered Access Model is also particularly suited to access to data resources where data are not 'distributed', thereby addressing concerns underlying the third category of review in controlled access: 'Security' (see Supplementary Table S1).

Registered access in health research can also draw guidance from national statistics institutes providing researchers with access to microdata. Their access processes reflect strict confidentiality requirements, as access is authorized through legislation rather than consent. Secure access to statistical microdata has been modeled along five dimensions: safe data, person (researcher), project, infrastructure and output.²⁴ A registered access process would focus primarily on ensuring the data user is trusted. By verifying that a data user is *bona fide*, one can to some extent impute that the project, security infrastructure and output will also be safe. In a data sharing context where the risk of identifying participants, or the sensitivity of the data, is low, additional review of these other dimensions may be redundant, or at least disproportionate. In other words, where the data are 'safe' and the data user is 'trusted', other aspects of secure access do not need to be heavily scrutinized.

A second observation from the access to microdata literature is the importance of the accountability relationship between data steward and researcher. National statistics organizations typically rely on statutory penalties and data access contracts to hold national researchers legally accountable. The enforceability of both comes into question when data are shared across borders. Administrative accountability – relying on host institutions to impose administrative sanctions for non-compliance – remains as a meaningful form of accountability. Host institutions may be held accountable by data stewards through reputation and contract to in turn ensure the quality and integrity of their activities. Trust and verification of the research, then, could largely be a process of verifying the host institution, through limiting registered access to a trusted network of institutions, or by scrutinizing the credentials of institutions on a case-by-case basis. In genomic research, host institution verification is likely to be

Table 1 List of criteria that are reviewed in controlled access data access by DACs and DACOs

<i>Controlled access review criteria</i>	
1	Applicants are affiliated to a recognized institution
2	Main applicant is qualified to undertake the proposed data analyses
3	Compatibility of proposed study with the database/data provider's objectives
4	Compatibility of proposed study with the consent requirements
5	Proposed study requires access to controlled access data sets
6	Required ethical obligations have been met
7	Scientific merit of proposed study (clarity, novelty and scientific excellence)
8	Applicants' privacy and confidentiality policy and security measures are adequate
9	Applicants' institution has approved and signed a Data Access Agreement

Abbreviations: DACs, Data Access Committees; DACOs, Data Access Compliance Office.

established on a case-by-case basis, as no formal accreditation bodies for 'safe institutions' or 'safe researchers' exist. In the controlled access model, this verification is reinforced with DACO review of researcher competence and local ethics compliance. In short, a robust method of institution verification or accreditation may be an important factor for a responsible registered access scheme. On the other hand, it raises questions about barriers that might ensue from increasing reliance on institutional affiliation.

Furthermore, registered access could be founded on a simple self-declaration system (ie, an attestation) for issues such as verifying compliance of the research with local ethical standards and procedures, adhesion to consent restrictions on the scope of data use and obligations not to reidentify anonymized data. Simple unilateral contractual commitments required of the data user can promote direct accountability of data users, without significantly increasing the arduousness of the access process for them. These commitments could easily cover much of the same requirements imposed through access agreements: that the data user agrees to comply with all applicable regulatory requirements (eg, has ethics approval if applicable); understands and respects data use limits; will not attempt to reidentify the participants; will take reasonable steps to protect the data from unauthorized access and delete it when the time period of approval has expired. This approach clarifies responsibilities and imposes a backstop of contractual accountability comparable to that found within a controlled access framework.

We therefore propose that the registered access criteria shown in Table 2 should constitute the general and basic framework for the Registered Access Model.

Perhaps, the most challenging aspect of introducing an intermediate data access tier will therefore be defining suitable inclusion criteria for registered data users, which we envisage will include researchers in academia and industry, and different groups of clinical care professionals (eg, doctors, genetic counselors). It may be necessary to leave the specification of the required level of 'competence' to groups implementing registered access or to establish

a few standardized levels. A few key pieces of administrative information will be important for authentication processes. However, it is currently unclear what should be required to demonstrate *bona fide* researcher or clinical care professional status. Evidence of academic publication has typically been requested in controlled access review (Dyke *et al.* in press),²⁵ but a goal of registered access is to broaden access to a greater number of researchers as well as to provide quick and easy access to less sensitive data. Academic publication records, even minimal, preclude access by many including a large section of clinical professionals and students.

REGISTERED ACCESS ATTESTATION

Registered access would not involve the execution of a Data Transfer or Access Agreement (DAA) between data providers and data users as required in controlled access. In a context where most of the interactions in the data sharing environment are taking place online, the concept of registered access could provide an appropriate ground for the use of online agreements, setting the terms of use for data users who wish to access registered data. Until now, paper-based applications, meetings of DACs and DAAs have been the primary basis governing contractual relations between the data providers, users and their institutions. This can be an administratively heavy step for both the data user and institutions.

Clickwrap-type online agreements, for example, in the form of web-based agreements requiring the end-user to manifest consent by clicking an 'I agree' checkbox option at the end of a contract, are generally well documented and used for a variety of online transactions such as purchasing flights online. Although they constitute a legally binding agreement in many jurisdictions, with specific laws applicable to online contracts, their validity and enforceability may however vary from one jurisdiction to another and clickwrap agreements may not be enforceable in all countries. There are instances where these have been used to set conditions and terms of use for access to open access genomic databases.²⁶ For example, at the outset, the HapMap database used an open source data access policy in

Table 2 Proposed registered access criteria

<i>General criteria (for Authorization)</i>		<i>Information required (for example)</i>
Competence	Applicants are <i>bona fide</i> researchers/clinical care professionals	1a The applicant's name, title, position, affiliation, email address, institutional website and mailing address (for Authentication) 1b Any additional information required for authorization of researcher/clinical care professional <i>bona fides</i> 1c 'I am a <i>bona fide</i> researcher/clinical care professional...' (with a definition)
Ethics	IRB/REC requirements have been met Respect GA4GH data sharing framework Adhesion to consent restrictions on the scope of data use	2a 'I will comply with all ethical and legal regulatory requirements applicable in my institution and country/region in my use of the data' 2b 'My use of the data will be consistent with the GA4GH <i>Framework for Responsible Sharing of Genomic and Health-Related Data</i> (https://genomicsandhealth.org)' 2c 'I will only use the data for the purposes allowed by the provider. In particular, I will abide by any consent conditions'
Security/confidentiality	Will not reidentify data: respect privacy of participants Will not share data with others/keep it confidential Will not keep copies of the data for longer than permitted	3a 'I agree to forego any attempt to identify individuals represented in the dataset, except by prior written permission from the provider's sponsoring institution' 3b 'I will treat the data as confidential and I will not share it with others' 3c 'I will delete all copies of the data when the permission period has expired'

Abbreviations: GA4GH, Global Alliance for Genomics and Health; IRB, Institutional Review Board; REC, research ethics committee.

a clickwrap format.^{27,28} Essentially, the HapMap project used a clickwrap licence agreement until all of the data was placed in the public domain, at which point the agreement was abandoned as a requirement for access. While not yet the standard medium for DAAs, these types of online agreements could arguably allow for a more balanced approach to access agreements by creating rapid, open and efficient access to data. They may also help in providing users with clear and upfront instructions on the use of the data. Standard registered access Attestation statements are listed in Table 2 (see 1c–3c) as an example of conditions that would form the core of a registered access agreement. This simple form of agreement would be strengthened by more detailed terms and conditions available on the website that registers the user's attestation.

Registered access could provide an interesting case for the implementation of such agreements. For instance, an efficient mechanism of clickwrap agreement enforcement when a breach or misuse is discovered is denial of access to the database by the user who has been identified and authorized.²⁷ A feature that would further enhance registered access would be to limit registration for 1 year, so as to renew authorization annually.

The registered access Authorization process would include verifying that the Attestation has been completed. Depending on the other elements requested, we envisage an officer rather than a committee would be responsible for a formal rather than a substantive review for Authorization, with referral to a controlled access review process if applicants fall outside standard registration criteria.

CONCLUSIONS

Improving access to health-related data must involve a careful calibration of protections, bearing in mind the public benefits of health research and indeed the rights of scientists and citizens alike to participate in, and to benefit from, scientific research.^{29,30}

Registered access is likely to be suitable as a mechanism for access to data types that are less sensitive, low risk data, such as non-stigmatizing health-related data from non-vulnerable individuals who would expect, or have consented to, data sharing for the purposes envisaged.³¹ It could also be a valuable tool to provide tiered access to different types of data users, including researchers and clinicians, and for access to multiple data sets as well as to facilitate data discovery. We aim to develop the Registered Access Model further through implementation and customization with the Global Alliance Demonstration Projects and, in particular, attention to the requirements for its clinical use.

Although not the primary aim, formalising our understanding of registered access may also contribute to improving and streamlining the controlled access process, if only by reducing pressure on DACOs and the controlled access system. Most importantly, in providing clarity to ethics governance bodies and other research partners, thus enabling this novel data access tier, projects for which as a lesser degree of data access review is warranted will be able to benefit from registered access.

CONFLICT OF INTEREST

As part of its International Policy interoperability and data Access Clearinghouse (IPAC), the Public Population Project in Genomics and Society (P3G) offers DACO services to the research community.

ACKNOWLEDGEMENTS

We thank Niklas Blomberg and Ilkka Lappalainen for comments on the manuscript and members of the GA4GH Beacon Project, Matchmaker

Exchange and BRCA Challenge for helpful discussion of this work. SD is supported by the Canadian Institutes of Health Research (Grants EP1-120608; EP2-120609), Genome Quebec, Genome Canada, the Government of Canada and the Ministère de l'Économie, Innovation et Exportation du Québec (Can-SHARE Grant 141210), and the Canada Research Chair in Law and Medicine. MS is funded by the IRO funding of University of Leuven. BK would like to acknowledge the funding support of the Canada Research Chairs Program. Funding for this research was also provided by Autism Speaks (MSSNG project).

AUTHOR CONTRIBUTIONS

SD and BK developed the concept of registered access. SD conceived of and conducted the ethical and legal research. EK, MS and AT participated in the research. SD wrote the manuscript with contributions from all other authors.

- 1 Kosseim P, Dove ES, Baggaley C *et al*: Building a data sharing model for global genomic research. *Genome Biol* 2014; **15**: 430.
- 2 The Global Alliance for Genomics and Health: A Federated Data Ecosystem for sharing genomic and clinical information. *Science* 2016; **352**: 1278–1280.
- 3 Knoppers BM: Framework for responsible sharing of genomic and health-related data. *HUGO J* 2014; **8**: 3.
- 4 International HapMap Consortium: The International HapMap Project. *Nature* 2003; **426**: 789–796.
- 5 The 1000 Genomes Project Consortium: A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–1073.
- 6 International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A *et al*: International network of cancer genome projects. *Nature* 2010; **464**: 993–998.
- 7 Mailman MD, Feolo M, Jin Y *et al*: The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007; **39**: 1181–1186.
- 8 Lappalainen I, Almeida-King J, Kumanduri V *et al*: The European Genome-phenome Archive of human data consented for biomedical research. *Nat Genet* 2015; **47**: 692–695.
- 9 Toronto International Data Release Workshop A, Birney E, Hudson TJ *et al*: Prepublication data sharing. *Nature* 2009; **461**: 168–170.
- 10 Ramos EM, Din-Lovinescu C, Bookman EB *et al*: A mechanism for controlled access to GWAS data: experience of the GAIN Data Access Committee. *Am J Hum Genet* 2013; **92**: 479–488.
- 11 Philippakis AA, Azzariti DR, Beltran S *et al*: The Matchmaker Exchange: a platform for rare disease discovery. *Hum Mutat* 2015; **36**: 915–921.
- 12 Frodsham AJ, Higgins JP: Online genetic databases informing human genome epidemiology. *BMC Med Res Methodol* 2007; **7**: 31.
- 13 Firth HV, Richards SM, Bevan AP *et al*: DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 2009; **84**: 524–533.
- 14 Buske OJ, Girdea M, Dumitriu S *et al*: PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum Mutat* 2015; **36**: 931–940.
- 15 NHGRI: *Workshop on Establishing a Central Resource of Data from Genome Sequencing Projects*. NHGRI: Bethesda, MD, USA, 2012.
- 16 Shabani M, Borry P: 'You want the right amount of oversight': interviews with data access committee members and experts on genomic data access. *Genet Med* 2016; **18**: 892–897.
- 17 Shabani M, Knoppers BM, Borry P: From the principles of genomic data sharing to the practices of data access committees. *EMBO Mol Med* 2015; **7**: 507–509.
- 18 Shabani M, Dyke SO, Joly Y, Borry P: Controlled access under review: improving the governance of genomic data access. *PLoS Biol* 2015; **13**: e1002339.
- 19 Expert Advisory Group on Data Access: *Governance of Data Access*. Wellcome Trust: London, UK, 2015.
- 20 Joly Y, Dove ES, Knoppers BM, Bobrow M, Chalmers D: Data sharing in the post-genomic world: the experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO). *PLoS Comput Biol* 2012; **8**: e1002549.
- 21 US National Institutes of Health: *Genomic Data Sharing Policy*. NIH: Bethesda, MD, USA, 2014.
- 22 Kaye J, Hawkins N: Data sharing policy design for consortia: challenges for sustainability. *Genome Med* 2014; **6**: 4.
- 23 Dyke SO, Philippakis AA, Rambla De Argila J *et al*: Consent Codes: upholding standard data use conditions. *PLoS Genet* 2016; **12**: e1005772.
- 24 OECD Expert Group for International Collaboration on Microdata Access: *Executive Summary of the Final Report*. OECD: Paris, 2014.
- 25 Dyke SO, Saunier KM, Pastinen T *et al*: Evolving Data Access Policy: the Canadian Context. *FACETS (in press)*.

- 26 Pereira S, Gibbs RA, McGuire AL: Open access data sharing in genomic research. *Genes* 2014; **5**: 739–747.
- 27 Gitter DM: Resolving the open source paradox in biotechnology: a proposal for a revised open source policy for publicly funded genomic databases. *Hous L Rev* 2006–07; **43**: 1475.
- 28 Singh KK: *Intellectual Property Protection to Bioinformatics and Genomic Databases and Open Source Analogy to Biotechnology*. India: Springer, 2015, pp 169–193.
- 29 Knoppers BM, Harris JR, Budin-Ljosne I, Dove ES: A human rights approach to an international code of conduct for genomic and clinical data sharing. *Hum Genet* 2014; **133**: 895–903.
- 30 UNESCO: *The Right to Enjoy the Benefits of Scientific Progress and its Applications*, UNESCO: Paris, 2009.
- 31 Dyke SOM, Dove ES, Knoppers BM: Sharing health-related data: a privacy test? *Genomic Med* 2016; **1**: 16024.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

© The Author(s) 2016

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)