Privacy-Preserving Record Linkage: Ethico-Legal Considerations

Mark Phillips, Bartha Maria Knoppers, Dixie Baker and Petra Kaufmann - March 2018

1. Introduction and Background

Health data research can often benefit by enriching study participant data through reliable, convenient linkage to other data sets.[1] Where rare disease patients are concerned, it can take years to "connect the dots" in order to make an accurate diagnosis. A related difficulty in research is detecting when two records, that appear to be distinct participants, in fact relate to the same person, which risks distorting the analysis based on these data sets, since a given person can be counted two or more times. In at least some contexts, evidence suggests that participants would be comfortable with their data being linked.[2] Linkage aligns with the missions of both the International Rare Diseases Research Consortium (IRDiRC) and the Global Alliance for Genomics and Health (GA4GH), who have recently collaborated to explore the idea of linkage in their aims to increase health data sharing, including of reference databases.

But linkage also gives rise to ethico-legal considerations. This primer focuses on the impact of linkage on the interrelated issues of participant privacy, confidentiality, and data security, and makes recommendations about the design of privacy-preserving record linkage (PPRL) systems. Other legal issues are relevant to PPRL but fall outside of this primer's scope. A patent covering a particular linkage method, for example, may prevent that technique from being used without a costly license. This primer concentrates, however, on recommendations related to the implications of linkage on privacy.

This primer accompanies a primer on the technical considerations relevant to PPRL that arose out of the same joint IRDiRC/GA4GH initiative.

This primer's recommendations flow from principles including the need to reconcile the benefits of data sharing for health with privacy, as embodied in the general principles in the Global Alliance for Genomics and Health's *Framework for the Responsible Sharing of Genomic and Health-Related Data*[3]; the additional principle that any new development should aim to decrease, rather than expand, inequalities in access to health and healthcare; and the 2017 Organisation for Economic Co-operation and Development (OECD) recommendation that "international collaboration is essential to enable all countries to safely benefit from health data and to support the production of multi-country statistics, research and other uses of data that serve the public interest."[4] One co-author has previously noted that "[b]y catalyzing the right to benefit from science and recommending reinforced security and governance in order to facilitate data sharing, Big Data can be gathered for improved health research and clinical care."[5]

Given divergent laws across countries, as well as the shifting and uncertain content of data protection duties, this primer's recommendations do not provide instructions on how to achieve legal and regulatory compliance in every jurisdiction, but instead attempt to frame minimal requirements that appear necessary to begin to allow international linkage. This primer's focus is on highly influential laws in Europe and the United States, although it also attempts to incorporate in its analysis, likely in an incomplete way, relevant norms in other jurisdictions, when appropriate.

The sections that follow discuss (2) the process to generate linkage codes, (3) appraisals of the sensitivity of the data, (4) the desirability of delegating linkage and re-identification to distinct entities, (5) the feasibility of a

[1] For a recent example, see e.g. Mats G Hansson et al, "The risk of re-identification versus the need to identify individuals in rare disease research" (2017) *European Journal of Human Genetics*, doi:10.1038/ejhg.2016.52.
[2] In clinical trials in the Canadian province of Ontario, for example, see A.E. Hay et al., "Linkage of clinical trial and administrative data: a survey of cancer patient preferences", *Current Oncology*; **24**(3), https://doi.org/10.3747/co.24.3400.
[3] https://www.ga4gh.org/ga4ghtoolkit/regulatoryandethics/framework-for-responsible-sharing-genomic-and-health-related-data/
[4] OECD, OECD Recommendation on Health Data Governance (17 January 2017), https://www.oecd.org/els/health-systems/health-data-governance.htm at p. 14. For a discussion of access requirements to controlled access health data, see Paul R. Burton *et al.*, "Data Safe Havens in health research and healthcare" (2015) 31:20 *Bioinformatics* 3241.
[5] Bartha Maria Knoppers & Adrian Thorogood, "Ethics and Big Data in Health" (2017) 4 *Current Opinion in Systems Biology* 53.

distributed PPRL structure, (6) the return of results and, finally, (7) an overall summary of this primer's recommendations for PPRL systems.

2. Generating Linkage Codes

We will use the term "codes" to refer to the information used to link records in any given system after steps have been taken to reduce the identifiability of the records.[6]

The simplest linkage approach is to assign each participant a universal identifier derived from a set of relatively immutable personal data. China's national identity card, for example, is an 18-digit number that combines a person's birthdate, birth location, and other data. Denmark and Estonia use 10- and 11-digit numbers, respectively, each of which encode birthdate, binary gender, a sequence number, and usually also a checksum.

This type of code, however, is unacceptable for international health data linkage. In the United States, for example, the *Health Insurance Portability and Accountability Act of 1996* (HIPAA) Privacy Rule, which governs the sharing and use of clinical data, requires that "means of record identification is not derived from or related to information about the individual".[7]

A more privacy-conscious approach to linkage is to consistently use a randomly generated code for each participant and associate that code with cryptographically hashed immutable personal identifiers. Since a random identifier has no intrinsic relationship to any personal data, an adversary cannot derive any personal data from it alone, nor even confirm the number itself when in possession of the relevant personal data, though the hash will still allow the system to detect records that refer to the same participant. This is the approach, for example, of the NIH GUID systems.[8] For a more complete discussion of the protections provided by cryptographic hashes, consult the accompanying technical primer.

This approach, however, is still not suitable for universal linkage given the strict regulation or outright prohibition of universal identifiers in various jurisdictions. *De facto* prohibition is the current state of the law at the federal level in the United States. In particular, despite the fact that HIPAA requires the Department of Health and Human Services (HHS) to create a standard, unique health identifier, due to privacy concerns, Congress has prohibited the use of federal funds to do so by maintaining language similar to the following in appropriations bills since 1999:

> None of the funds made available in this Act may be used to promulgate or adopt any final standard under section 1173(b) of the Social Security Act providing for, or providing for the assignment of, a unique health identifier for an individual (except in an individual's capacity as an employer or a health care provider), until legislation is enacted specifically approving the standard.

In the European Union, although the *General Data Protection Regulation* (GDPR), like the *Data Protection Directive* before it, empowers member states to adopt frameworks for an "identifier of general application"[9], few have done so. France's National Consultative Ethics Committee for Health and Life Sciences has in fact recommended the "prohibition of interconnection of databases designed for different purposes but with common identifiers."[10] The provision in the Regulation adds the condition that despite any national framework, such identifiers "shall be used only under appropriate safeguards for the rights and freedoms of the data subject pursuant to this Regulation." The existing interpretive guidance additionally suggests avoiding the use of the same code across different datasets.[11] An international system must therefore go beyond the NIH GUID approach by eliminating the need for a given identifier to be distributed to multiple data-holding sites. Existing PPRL systems of this variety include the Australia's Population Health Research Network, which uses a hashing technique known as

---

[6] The PPRL Task Team uses the term "coded" to describe records or other data whose direct and quasi identifiers have been removed and replaced with a re-identification code that is generated independently of the values of identity attributes. Coding requires that it is impossible to derive the participant's identity without access to the information associating the code with an individual. Coding, according to this definition, is a form of pseudonymisation, which refers generally to "personal data [that] can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person" (GDPR Article 4(5)).

[7] 45 CFR §164.514(*c*)(1).

[8] See e.g. NIH/NCATS GRDR, "Global Unique Identifier (GUID)", online: <ncats.nih.gov/grdr/guid>.

[9] Compare Article 8 of the Directive and Article 87 of the Regulation.

[10] National Consultative Ethics Committee for Health and Life Sciences, "Opinion n° 98: Biometrics, identifying data and human rights" (2007).

[11] Article 29 Working Party, Opinion 05/2014 on Anonymization Techniques [Opinion 05/2015].

Bloom filters,[12] and the European Network for Cancer research in Children and Adolescents' patient identity management system.[13]

Although hash values are intended to remain private, a PPRL system should still choose its hash function with care. Guidance around the world, including in the EU, favours a case-by-case evaluation of identification risks,[14] including caution when using a scheme to generate personal identifiers that has known weaknesses.[15]

This risk evaluation should extend to the safeguards applied to code storage and use. The HIPAA Privacy Rule, for example, sets out implementation specifications to prevent re-identification of the coded data.[16]

PPRL systems should exercise caution in including biometrics in the immutable personal data that is fed to the hash function, as was favoured by the Bio-PIN system.[17] Article 10 of the GDPR establishes that

> the processing of … biometric data for the purpose of uniquely identifying a natural person … shall be allowed only where strictly necessary, subject to appropriate safeguards for the rights and freedoms of the data subject, and only:
>
> (a) where authorised by Union or Member State law;
>
> (b) to protect the vital interests of the data subject or of another natural person; or
>
> (c) where such processing relates to data which are manifestly made public by the data subject.

Use of biometrics also prevents the data from being considered de-identified according to the HIPAA Privacy Rule's Safe Harbor[18] standard, which prohibits the presence of biometric data in a de-identified record, and may attract regulatory rules in other privacy laws in any number of other jurisdictions, such as the US states of Texas,[19] Washington State,[20] and Illinois.[21] The latter law, for example, regulates the use of biometric identifiers generally including "a retina or iris scan, fingerprint, voiceprint, or scan of hand or face geometry"[22]. Caution is also merited due to the positions taken by, for example, the National Research Council, which describes biometric identification, for example, as "inherently fallible".[23] Experts distinguish identification from authentication, and emphasize that even more serious "problems arise when systems begin using biometrics for authentication."[24]

A final consideration when designing a PPRL system is to build-in robust support for the irrevocable right of participants to withdraw their consent any time after giving it. This is a standard feature of the fields of research ethics, medical ethics, and data protection. For a system to comply with the GDPR, it must "be as easy to withdraw as to give the consent."[25]

But withdrawal poses specific technical difficulties for linkage systems, notably when using standardized sets of immutable personal data, hashing functions in general, and Bloom filters in particular. To these difficulties are

---

[12] See e.g. Randall SM, Ferrante AM, Boyd JH, Semmens JB. (2014). Privacy-preserving record linkage on large real world datasets. *Journal of Biomedical Informatics* 50, 205-212. ; Schnell R, Bachteler T, Reiher R. (2009) Privacy-preserving record linkage using Bloom filters. *BMC Medical Informatics and Decision Making* 9(41), doi: 10.1186/1472-6947-9-41.
[13] Michael NITZLNADER a, 1 and Günter SCHREIER, "Patient Identity Management for Secondary Use of Biomedical Research Data in a Distributed Computing Environment" *eHealth2014*, doi:10.3233/978-1-61499-397-1-211.
[14] *Supra* note 9 at 23.
[15] Opinion 05/2015, *supra* advocates putting further safeguards in place "if the range of input values … are known they can be replayed through the hash function in order to derive the correct value for a particular record" (at 20) or if the output values "are rotated in a way that "might trigger the occurrence of patterns, partially reducing the intended benefits" (at 22). Note however, that it seems to be more permissive than HIPAA in that "pseudonymisation can be independent of the initial value … or it can be derived from the original values of an attribute or set of attributes" (at 20).
[16] 45 CFR §164.514(*c*).
[17] JJ Nietfeld, J Sugarman & J-E Litton, "The Bio-PIN: A Concept to Improve Biobanking" (April 2011) 11:4 *Nat. Rev. Cancer* 303–08.
[18] Not to be confused with the US/EU data-sharing scheme of the same name which the EU Court of Justice's *Schrems* decision determined to provide inadequate protection to Europeans' personal data.
[19] Texas Business & Commerce Code §503.001 (Capture or Use of Biometric Identifier).
[20] *An Act Relating to biometric identifiers* (Engrossed Substitute House Bill 1493), Washington Laws of 2017, chapter 299.
[21] *Biometric Information Privacy Act*, Illinois Compiled Statutes, chapter 740, number 14.
[22] *Ibid* at 4/1.
[23] National Research Council, *Biometric Recognition: Challenges and Opportunities* (Washington, DC: National Academies, 2010).
[24] Steve Riley, "It's Me, and Here's My Proof: Why Identity and Authentication Must Remain Distinct" (2006), online: <technet.microsoft.com/en-us/library/cc512578.aspx>.
[25] Article 7(3).

added those that are inherent in complex, decentralized personal data-sharing environments in which multiple entities transfer, link, and aggregate existing data sets among one another. Retroactive withdrawal is not required, and is often not possible, but prospective withdrawal must be supported by any entity that processes personal health data.[26] If a participant who has opted to withdraw their consent later re-enters another component of the system, its design should ensure that none of their other previous data will "re-emerge". These difficulties appear to be surmountable, but a PPRL framework should pay specific and thoughtful consideration to the way it implements participants' right to withdraw or limit their consent.

3. Sensitivity of the Data

PPRL-related data should be treated as sensitive.

Whether a PPRL system actually links data, or simply makes it convenient to do so, it increases centralization of the data. Centralize systems dramatically increase both the risk of a breach, since adversaries' motivation to access the data increases exponentially with the degree to which it is comprehensive, as well as the potential harm of a breach, given the much larger quantity of data compromised. Centralized or centralizable global repositories will become subject to attempts at mass surveillance by public and private actors, of which the public has become increasingly aware. Because any system's "security must be evaluated not based on how it works, but on how it fails" when it encounters a person intent on subverting the system,[27]

PPRL systems with the capacity to lead toward centralization of medical and health data should be treated as sensitive. The design of an international PPRL system must therefore carefully analyze the risks, benefits, and available safeguards based on a variety of threats.

Such analysis is often required to meet legal obligations. Personal data held for a research purpose in the Canadian province of British Columbia, for example, can only be linked in accordance with the law if "any data linkage is not harmful to the individuals who are the subjects … and benefits … are clearly in the public interest."[28] Article 27 of the GDPR requires that an entity developing an international PPRL system carry out a privacy impact assessment, and Article 28 would likely also require that it consult the relevant supervisory authority.

A related point to consider is the risk that data aggregation through linkage may transform data that was not reasonably foreseeably identifiable into potentially identifiable data. Linkage should thus either require that continued non-identifiability is proven, using measures like *k*-anonymity or *l*-diversity[29], or that sufficient alternative safeguards be put in place.

4. Separating the Linkage Entity

Data protection law generally establishes a principle of proportionality, or "data minimization", according to which the least amount of personal data should be processed that is necessary to achieve the purpose. In a system with multiple users, such as the kind of PPRL system this paper is examining, this principle aligns with the information security principle of "least privilege", according to which each entity in a system should be provided with only the minimum amount of information and resources necessary for its legitimate purposes.

For this reason, a PPRL system that tasks a single entity with the responsibility for linkage, even one that is deemed "trustworthy" should be avoided. This is especially the case when the linkage entity is also one of the data holders. In a large-scale system, distributing decision making such that no single entity has controls on its own of the complete dataset is preferable, such as by using approaches including secure multi-party computation (SMC) and federated linkage.

For this reason, in circumstances where the PPRL system must support the ability to re-identify participants when necessary, this capacity to re-identify should be delegated to a distinct entity with which other actors in the system

[26] Additionally, varying interpretations exist of the storage and processing activities that fall within one or the other of these categories, and that should therefore be subject to withdrawal.
[27] Bruce Schneier, "A National ID Card Wouldn't Make Us Safer" (1 April 2004) *Minneapolis Star Tribune*.
[28] See *Freedom of Information and Protection of Privacy Act*, RSBC 1996, c 165, ss 35(1)(*b*), s. 36.1(1); *E-Health Act*, SBC 2008, c 38, s 14(2.1)(*d*); *Personal Information Protection Act*, SBC 2003, c 63, s 21(1)(*c*).
[29] K. Stark, J. Eder, K. Zatloukal. Priority-based k-anonymity accomplished by weighted generalisation structures. Data Warehousing and Knowledge Discovery, Proceedings. 2006;4081:394-404. PubMed PMID: WOS:000241158200038.

may need to collaborate for the purposes of re-identification. This is the approach adopted by the EUPID system.[30] Even if re-identification may be the task of a specific entity, re-contact should be carried out, when required, by the entity that was initially in contact with the participant: in other words, generally the physician, not the researcher.

Under EU data protection law, the linkage entity is likely to be considered a joint controller along with the other entities in the system, rather than a mere processor acting only according to a controller's instructions, as the linkage entity will not process its data solely on the instructions of any other single entity. Article 11(2) of the GDPR, however, relieves a controller of duties that would otherwise apply, including fulfillment of the right of access, in cases where the controller is unable to identify the data it holds, which will be the case for the linkage entity, even if it remains personal data.

5. Distributed Structure

The legal principle of minimisation which, as noted earlier, aligns with the information security principle of least privilege in favouring decentralization of personal data processing whenever the same benefits can still be achieved.

The tasks of linkage and re-identification, which the previous section suggested should be separated from the data-holding entities, should therefore also be distributed among multiple entities, according to the considerations laid out in the technical primer, whenever this is feasible, as it is generally likely to be. In this way, a regional linkage or re-identification entity might take the primary responsibility for data contributors or data users in their region.

Such a distributed structure, however, may risk inconsistency with the HIPAA Privacy Rule's Safe Harbor. Data sets that have been de-identified by removing its list of prescribed identifiers may only retain a "unique identifying number, characteristic, or code" under limited conditions, and only to allow the information to be re-identified by the same entity that produced it. One of the restrictions imposed by the Safe Harbor is that this entity shall not "disclose the code … for any other purpose".[31] Unless this entity can be described as engaging in distributed linkage for the sole purpose of allowing it to re-identify the data, the Safe Harbor will not allow sharing a code this way, which is likely to be a challenge.

6. Restrictions on Cross-Border Personal Data Transfer and Localisation Laws

Communication between distributed entities aimed at identifying matching participants, even when using PPRL techniques, will nonetheless generally be considered by the relevant legal frameworks to entail the disclosure of personal data. When this data crosses a jurisdictional boundary, it is also an international transfer of personal data.

From the GDPR's perspective, disclosure and transfer each must meet distinct regulatory conditions to be legal. Justification for personal data transfer may be satisfied either when an EU Commission adequacy decision applies to a legal framework to which the receiving entity is bound, or by alternative measures including the use of standard contractual clauses approved by the Commission. Transfer must be documented according to the record-keeping requirements described in Article 25 and comply with other GDPR requirements related to transfers.

Another set of restrictions on transfer are data localization laws that have begun to emerge in a variety of jurisdictions and that amount to outright prohibitions on personal data leaving the jurisdiction.

China proposed one such law in 2017.[32] The application of the proposed law remains unclear with respect to key questions, including the entities to which the law would apply as well as the specific types of data that cannot be transferred outside the country. The proposed law does not, however, generally impose an absolute ban on cross-border transfer even on those entities to which it applies, and is instead attentive to factors such as whether the transfer affects fewer than 500,000 data subjects, whether the data subject provided implied consent to a transfer they initiated, and whether the transfer received security review approval from the government.

---

[30] Michael Nitzlnader & Günter Schreier, "Patient Identity Management for Secondary Use of Biomedical Research Data in a Distributed Computing Environment" in A. Hörbst et al. (Eds.), *eHealth2014 – Health Informatics Meets eHealth* (IOS Press, 2014), doi:10.3233/978-1-61499-397-1-211.

[31] 5 CFR §164.514(c)(2).

[32] See e.g. Hogan Lovells, "China's draft data localisation measures open for comment", https://f.datasrvr.com/fr1/717/55906/Chinas_draft_data_localisation_measures_open_for_comment.pdf; Hogan Lovells, "China's Revised Draft Data Localisation Measures", http://www.hldataprotection.com/2017/05/articles/international-eu-privacy/chinas-revised-draft-data-localisation-measures/

A quite different type of localization law entered into force in Russia in 2015.[33] This type of law poses far less difficulty for a PPRL system, since although it requires that Russian citizens' personal data must be stored on infrastructure physically located within Russia at all times, it does not prohibit the data from being transferred outside the country, so long as another copy of the data remains in Russia.

Canadian federal policy[34] and provincial laws[35] have opted to restrict the exportation of personal data effectively held by the government, which are also less likely to impact this system.

The type of law most likely to affect a PPRL system, however, at least an international one, are those similar to Australia's *Personally Controlled Electronic Health Record Provision*, which prohibits personal health data from leaving Australia in some situations.[36]

7. Return of Results

An international PPRL system should take a flexible approach to the return of results.

The difficulty in this area is that while participant re-identification is sometimes prohibited, as is sometimes the case under HIPAA, in other policy frameworks the ability to return results is mandatory.[37]

A PPRL system should thus be designed to accommodate both possibilities. The most obvious approach would be to design the PPRL system so that each regional entity has the option of incorporating support for re-identification or not, according to its needs. Care must be taken to appropriately account for the potential that a participant has their data held simultaneously by multiple regional entities, some of which allow re-identification while others prohibit it, to ensure that the overall system can simultaneously comply with the contrasting normative regimes.

8. Conclusion and Summary of Recommendations

This primer recommends incorporating the following features and guidelines into any policy that guides the design and operation of a PPRL system:

1. **Use a trusted cryptographic hashing method to enable linkage.** A variety of existing approaches do just this, as discussed in the accompanying technical primer. The identifying data fields that may or must serve as the inputs to the hashing algorithm should be chosen carefully, including according to the considerations discussed earlier.

2. **Do not distribute a single, unique ID.** While this would allow institutions to link with one another without external assistance, which may be desirable in some cases, such a system's vulnerability expands by an order of magnitude and the approach is inconsistent with the law of a number of jurisdictions.

3. **Avoid detecting duplicate patients using biometric data.** Biometric data is subject to numerous legal restrictions and prohibitions internationally, particularly when used to identify an individual.

4. **Support participant withdrawal.** Participants should be able to withdraw their consent to use and retention of their data throughout the distributed system. None of this data should "re-emerge" if the participant later re-enters the system.

5. **Perform a data security and privacy impact assessment.** Taking note of all the circumstances including the sensitivity of the health data that will conceivably be linked by the system.

---

[33] Vera Shaftan, "Russian data protection authority explains data localization law; says cross-border transfer still permitted" (4 August 2015) *Data Protection Report*, http://www.dataprotectionreport.com/2015/08/russian-data-protection-authority-explains-data-localization-law-says-cross-border-transfer-still-permitted/

[34] Alice McGregor, "Canada wants to keep federal data within national borders", https://thestack.com/cloud/2016/08/03/canada-wants-to-keep-federal-data-within-national-borders/

[35] See *Personal Information International Disclosure Protection Act*, Statutes of Nova Scotia, 2006 ch. 3; *Freedom of Information and Protection of Privacy Act*, Revised Statutes of British Columbia, 1996, ch. 165, as amended.

[36] *Personally Controlled Electronic Health Records Act 2012*, No. 63, 2012, s. 77 ("Requirement not to hold or take records outside Australia").

[37] See e.g. the following position statement of the American Society of Human Genetics: J.R. Botkin et al. "Points to Consider: Ethical, Legal, and Psychosocial Implications of Genetic Testing in Children and Adolescents" (2015) 97 *American Journal of Human Genetics* 6 at 9. Similarly, consent to the return of results is mandatory to participate in Genomics England's 100,000 Genomes Project. See Genomics England, "The 100,000 Genomes Project Protocol v4", https://doi.org/10.6084/m9.figshare.4530893.v4 2017.

6. **Determine the appropriate approach to consent.** The authors' views diverge on how participants should consent to having their data enter the linkage system. No service should be set up so as to depend on consent to linkage if opting out of that service is likely to be a disproportionate hardship, in order to prevent the system from becoming *de facto* mandatory. The system should support specific, opt-in consent, since many jurisdictions will specifically require it. In other jurisdictions, it may be legally permissible to have the intensity of consent fall along a spectrum from broad to specific specific depending on circumstances. Some authors preferred advance notification with the opportunity to opt-out, such as to improve the quality of healthcare delivery.

7. **Limit identifiability.** Demonstrate that it remains the case that it is not reasonably foreseeable that the data is re-identifiable post-linkage using either mathematical calculations or equivalent alternative safeguards.

8. **Support different approaches to re-identification among participating entities.** The system should be consistent with conflicting rules that in some contexts prohibit re-identification, while in others make return of results mandatory.

9. **Distribute key duties within the system among distinct, trusted entities.** Separate the entities responsible for linkage and re-identification from entities that hold detailed participant data, and support regional distribution and federation of these tasks.

A comprehensive policy that frames the obligations of international PPRL systems might effectively be set out in an international Code of Conduct for health data sharing, an endeavour that is attracting a flurry of interest[38] given the increased incentives on sectors of data controllers to develop these tools pursuant to the GDPR. Adherence to an appropriately approved Code of Conduct provides a controller with additional evidence of legal compliance.

Although the authors are not aware of any existing system that meets all of the requirements outlined above, no insurmountable technical barriers exist that prevent one from being established. From a policy point of view, such a system would further the mission of both GA4GH and IRDiRC, as described at the outset, and may find application beyond the specificities of rare disease research.

[38] See e.g. Mark Phillips et al., "Concretizing the Cloud", forthcoming.