

RESEARCH COOPERATION: COVID-19

Blockchain-facilitated sharing to advance outbreak R&D

Technology may help overcome nontechnological barriers

By **Mark B. van der Waal**^{1,2}, **Carolina dos S. Ribeiro**^{1,3}, **Moses Ma**⁴, **George B. Haringhuizen**³, **Eric Claassen**¹, **Linda H. M. van de Burgwal**^{1,2}

Timely and widespread dissemination of resources and information related to pathogenic threats plays a critical role in outbreak recognition, research, containment, and mitigation (1, 2), as stakeholders from government, public health (PH), industry, and academia seek to implement interventions and develop vaccines, diagnostics, and drugs (3). But there are resistant barriers to sharing and cooperative research and development (R&D) in the context of epidemics, rooted in a lack of trust in confidentiality and reciprocity (4, 5), ambiguity over resource ownership (6), and conflicting public, private, and academic incentives (2–4, 6). Here, we suggest how recent advances in blockchain and related technologies can enable decentralized mechanisms to help break down these systemic and largely nontechnological barriers. These mechanisms resolve scalability, energy consumption, and security concerns of early blockchain models and may be applied to underpin and interconnect, rather than supersede or conflict with existing, well-established systems and practices for storing, sharing, and governing resources.

As opposed to centralized databases that are maintained by a single party, a blockchain involves an infrastructure of different parties (nodes), each maintaining an identical copy of a distributed ledger. Once time-stamped into the ledger, records cannot be altered or removed unnoticed, owing to cryptographic data-structuring. A one-way algorithm processes data into cryptographic identifiers (hash codes), which are unique for an input value, that is, the algorithm will have a different output if the input is altered in any way. There is no way to reconstruct underlying data content from a hash code. In a block-

chain, the hash code of the preceding record is included in the new record before “hashing” and time-stamping it, making the ledger evolve as a chained, time-stamped record-keeping system that is tamper-resistant by design: The hash of an altered ledger will deviate from the hash of the consensually verified ledger as maintained by the rest of the nodes. Hence, blockchains enable proof of the existence of specific data objects and their content at specific points in time while data itself may remain concealed. This distributed infrastructure offers a common and inviolable source of records that can be verified by (permitted) network entities, removing the necessity of having a mutually trusted, centralized intermediary for verification and record-keeping of exchanges.

BARRIERS TO SHARING

Outbreak R&D depends on access to pathogen samples, data, and information, which are shared through physical collections of microbial and viral cultures (biobanks), open-access or restricted genetic sequence databases, or ad hoc peer-to-peer exchanges, and often only after having been shared through scientific publishing or patenting. The following barriers hamper timely and widespread sharing through these systems.

Procedural delays

Rapid international cooperation during outbreaks is challenged by a lack of trust in reciprocity, with countries fearing unfair sharing of benefits arising from the use of their local resources by foreign parties. A prominent example arose in 2006, when the Indonesian government denied foreign access to H5N1 influenza samples because of concerns about the unaffordability of resulting vaccines (4). Such concerns underlie the Nagoya Protocol (NP) to the Convention on Biological Diversity (CBD), which stipulates that access to genetic resources must be preceded by consent from providing countries and (bilateral) agreements on access and benefit-sharing (ABS). Users are responsible for tracing rights holders to negotiate and obtain certificates and permits for any sample (5). Partial implementation, lack of transparency in national legislations, and divergent interpretations of rights and obligations under the NP

can delay this process (6) and thus, for example, obstruct the validation of diagnostics (7). The NP's central information system, the ABS Clearing-House, lacks a complete picture of national ABS conditions (5). Moreover, the commercial nature and prospects of R&D are hard to determine ex ante, complicating ABS negotiations. Reliable mechanisms for tracking resources and access to those resources across storage systems are lacking (8) but called for to (temporarily) suspend negotiations, rapidly share, and allow for formalizing intent retrospectively. If the NP's scope is expanded to include genetic sequence data (GSD)—as currently debated—free sharing and rapid exchanges or data risk additional obstruction (2, 5).

Secrecy and fragmented R&D

Timely sharing of data and information on emerging pathogens can be frustrated by individual (competitive) interests, reinforced by systemic incentives (2, 6). Researchers have an incentive to publish peer-reviewed papers and demonstrate scientific priority (2, 9). Preprint platforms and close interactions between publishers and the PH community accelerate dissemination timelines but can still delay sharing until raw data or materials have been analyzed and processed unilaterally into publishable formats. Governments and researchers lack trust in reciprocity for shared resources and especially for GSD, because reliable mechanisms to track access and use across (public and private) systems remain absent (8). Even in the presence of designated portals hosted by PH authorities, lack of trust in database security and confidentiality can keep researchers from sharing (6). Closed data hubs developed for fast sharing offer limited means for managing and monitoring access of individual resources on a case-by-case basis (9). For severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) sequences, a closed hub was created under the Global Initiative on Sharing All Influenza Data (GISAI) that controls access and prohibits redistribution. Commercial aspirations can also cause sharing delays, as patent incentives impede open dissemination before patent applications are drafted and submitted (6). Reluctance in sharing is further explained by data sensitivity. Countries may fear impaired trade and tourism, and criticism on the appropriateness of measures taken (6). Source tracing or data triangulation can unintentionally lead to the identification of affected regions or individuals (2, 10). Furthermore, actors risk infringing on ethical and legal frameworks (e.g., the European Union's General Data Protection Regulation), especially once outbreak emergencies and any data privacy exemptions have expired.

¹Vrije Universiteit Amsterdam, Athena Institute for Research on Innovation and Communication in Health and Life Sciences, Amsterdam, Netherlands; ²Trial Foundation, Maastricht, Netherlands; ³The Netherlands National Institute for Public Health and the Environment (RIVM), Center for Infectious Disease Control, Billroth, Netherlands; ⁴FutureLab, Mill Valley, CA, USA. Email: m.van.der.waal@vu.nl

Uncertain ownership rights

Competition between labs can lead to fragmentation of intellectual property rights (IPRs) over GSD-based inventions and to time-consuming legal procedures to determine who has priority for each claim (3). Uncertain ownership rights translate into uncertain accessibility and affordability of building-block resources, subsequently delaying investments by downstream developers (3). For Middle East respiratory syndrome–coronavirus (MERS-CoV), conflicts over ownership delayed sharing, leading to persistent knowledge gaps on viral origins and transmission dynamics and hampering the development of vaccines and treatments (11). Yet, IPRs remain an important incentive for necessary industry investment in high-risk R&D to develop and produce diagnostics, vaccines, and therapeutics (3).

BLOCKCHAIN TO OVERCOME BARRIERS

Blockchain could help address root causes by underpinning the outbreak R&D ecosystem as a common, privacy-preserving, inviolable, and verifiable layer for records of objects and identities (e.g., resources, individuals, and organizations), rules (e.g., access permissions and ABS provisions), and events (e.g., access and benefit-sharing). Some have expressed concern about the cost and sustainability of implementing blockchain systems, but advanced models have appeared that do not rely on energy-guzzling algorithms to operate the distributed ledger and assure the integrity of its records. For instance, the necessary software and servers to implement a blockchain network can be hosted by a consortium of known, reputable, and pre-appointed authority node operators (ANOs), and network access can be restricted to permitted entities (i.e., those registered in the system and holding the right permissions). Such a federated, permissioned network model offers superior scalability, sustainability, and options for confidentiality as compared to “permissionless” systems such as the Bitcoin or public Ethereum blockchains. Current open-source technologies exist that allow for integration with traditional database management systems and appear fit for cost-effective and compatible prototyping and implementation of an outbreak R&D blockchain infrastructure (ORBI). We discuss key concepts and features of a possible ORBI [elaborated on in the supplementary materials (SM)].

Trustful sharing

An ORBI would enable actors to anchor hashed records of their digital or physical resources to establish time-stamped proof of their existence, integrity, and (scientific) priority in the blockchain. Records themselves

would be kept in an “off-chain” repository (9) and include indexing metadata (i.e., fields that systematically describe the resource, for example, pathogenic properties, provenance, and ownership) to enable querying and analysis by permitted entities only. Records would also include hashes of and pointers to the underlying resources themselves, which could be stored in any existing storage service. Depending on the preferences of resource providers (e.g., desired level of confidentiality), these may be open-access repositories [e.g., of the International Nucleotide Sequence Database Collaboration (INSDC)] or restricted systems (e.g., private encrypted data vaults or semi-open platforms like GISAIID).

Data privacy and sensitivity concerns would be addressed through decentralized identity and access management: Only entities that can cryptographically authenticate with a decentralized identifier (DID) that meets the right conditions are granted permission to discover and/or access records and underlying resources. DIDs are globally unique identifiers that are registered on the

publishing (preprint) papers. PH centers could register raw epidemiological datasets before analyzing and processing into aggregated country-level reports, enabling integrated analyses by authorized entities or analysis support when centers are heavily burdened during a PH crisis. The mechanisms would offer actors fine-grained control over exposure, for example, enabling instant selective disclosure of sensitive data to supranational coordinating bodies only, offering a head start while countries prepare their official public response and measures.

As suggested by MiPasa, a recent multi-stakeholder initiative for coronavirus disease 2019 (COVID-19) surveillance, blockchain-facilitated sharing can feed into improved and accelerated analyses of PH data, a use case for which blockchain has also been considered by the Centers for Disease Control and Prevention in the United States on a national level. This use case can be extended to enhance resource sharing and collaboration among public, private, and academic actors throughout the outbreak R&D chain.



blockchain for all network entities (e.g., individuals, organizations, devices, resources, or any other digital or physical objects). DIDs contain no personally identifiable information, can point to external locations (e.g., storage services or other service end points), and enable universal authentication of identities and their attributes (e.g., qualifications, permissions, or other credentials). Required credentials or other access conditions can be controlled by resource providers to meet (confidentiality) requirements of any applicable ethical or legal (IPR) framework. Conditions would be deployed through smart contracts: blockchain-registered scripts that can trigger an action (e.g., grant access) on recording conditionally relevant events (e.g., authenticating with the required credentials) (9, 12). These mechanisms could incentivize actors to rapidly time-stamp records—especially when contributions by data collectors and repositories would become adopted into the norms for scientific attribution or claiming ownership of inventions. Next to records of samples and sequences, researchers could register analyzed data before writing and

Traceability, interoperability, defragmentation DIDs offer decentralized control over identity attributes and service end points, complementing and integrating key (centralized) tools for resource traceability—notably the INSDC’s accession number for sequences, digital object identifiers for publications, and the internationally recognized certificate of compliance (IRCC) for NP access permits. Existing identifiers could be attributed to a DID hosted in the common ORBI to establish stable links, addressing fragmentation and redundancy issues of the current system (8) and reducing administrative burden.

Paired with a time-stamped audit log, DIDs and smart contract-coordinated permissions would enable a reliable tracking system for both resources and access events across storage systems (8). Access interfaces can be offered for existing database management systems and their users who want to verify identities and permissions on the blockchain (12), allowing data to be stored as before but increasing monitoring options. Access events would be recorded to shape an immutable audit trail (i.e., who accesses what

and under which conditions). Such a shared identity and access management system enables secure interconnections between storage systems that are currently siloed or only integrated at national or regional levels (2, 8). Although unintended circulations outside the tracking system (e.g., offline) are hard to rule out completely, blockchain mechanisms offer to strengthen the chain of custody tool kit of existing systems. They offer verifiable records (e.g., all parties with unique access keys) should disputes arise and be resolved under any existing legal framework, reducing reluctance to share and bringing data resources within the scope of NP principles of fair ABS (8). Foul play would be further discouraged when disclosing audit trails becomes expected in GSD-based publishing and patenting.

Facilitating compliance

Smart contracts would be applied to automate identification and authorization processes, accelerating, easing, and reducing transaction costs of compliance procedures. For instance, contracts could generate (and record) a unique access key for network entities on signing for the required ABS provisions, or trigger ABS obligations (e.g., payment) on recording actual access. This would enable users to demonstrate and assert compliance for both public and protected resources without the current administrative burden, substantially reducing sharing timelines. Blockchain prohibits unilateral changes to deployed smart contracts, clarifying and enforcing permissions, rights, and obligations for network entities. With the DIDs and audit log, the system could rebuild trust in agreements being upheld, incentivizing the input of resources.

Though smart contracts would allow for bilateral terms and conditions, a lack of alignment and harmonization in ABS provisions would impede the efficiency of an ORBL. Progress by governments and PH authorities on defining the scope, alignment, and harmonization of governance structures, and especially legal global frameworks, thus remains crucial (1, 5). An ORBL offers to facilitate policy implementation and promote compliance by translating best practices—such as the standardized material transfer agreements for research and commercial use under the World Health Organization's (WHO's) Pandemic Influenza Preparedness (PIP) Framework—into a certified library of smart contract templates, along with user interface components to modify the values of prespecified template attributes. In the Indonesian H5N1 case, such a system could have assisted in granting prompt access for entities involved in a noncommercial response while triggering conditional ABS provisions for any commercial follow-up.

Mapping R&D contributors

Blockchain could further contribute to trust and reciprocity by mapping contributors and their agreements throughout the outbreak R&D chain, avoiding time-consuming procedures for clarifying ownership such as those that were needed during the MERS-CoV emergency (11). R&D records could be stored in a repository that is optimized for directed acyclic graphs, which allows related records to be linked, capturing the evolution of R&D branches over time. A similar mechanism is applied by GitHub and finds support in recent literature (13). The audit log would affirm appropriate links and rightful contributions, and foul play could be further discouraged by algorithmically identifying probable links based on record metadata (probabilistic graphical modeling). Graphs may even assist in consolidating IPRs over ensuing inventions when smart contracts that define how to equitably distribute ownership among contributors are properly designed, certified, and offered in the system as configurable templates. These could coordinate auditable distribution of arising benefits (e.g., royalties) to all contributors—from those who register samples to those committing evidence of scientific value and/or patentability, and all stakeholders in between. In response to SARS, aggregating all fair contributors into a single patent-holding consortium (a patent pool) could have reduced risks for licensees and accelerated follow-on R&D (3). R&D graphs could thus support complex multistakeholder networks such as the WHO's R&D Blueprint and the Coalition for Epidemic Preparedness Innovations (CEPI) in prioritizing R&D while respecting individual ownership, by recording public and private contributions that can be accounted for retrospectively.

THOUGHTS ON IMPLEMENTATION

Key concepts we have discussed have been explored in recent efforts (9, 12, 13) and fit with existing open-source technologies (see SM). However, designing and implementing an ORBL-like system raises sociopolitical, legal, and technical issues that need effective resolution. Political willingness and involvement of stakeholders at the global governance level (e.g., WHO, Food and Agriculture Organization of the United Nations, World Organisation for Animal Health, World Intellectual Property Organization, and CBD) will be essential for aligning with existing (legal) frameworks and procedures and for coordinating pilots demonstrating system functioning in (simulated) practice. Adopting a multistakeholder governance model analogous to the Global Health Security Agenda, embodied by a dedicated steering group (SG) that includes a fair, global representa-

tion of acknowledged stakeholders, seems promising (see SM). An SG could oversee the appointment of ANOs and facilitate in-system design, implementation, and promotion through technical and policy working groups. Standardization of key enabling technologies (e.g., through the International Organization for Standardization, World Wide Web Consortium, and Institute of Electrical and Electronics Engineers) and interfaces with existing storage systems (e.g., INSDC, GISAID, and COMPARE) will determine success and sustainability, as will intuitive user clients and graphical user interfaces (2). Increased restrictions on sharing through strengthened access control could emerge but seem unlikely because this may conflict with legal obligations under the International Health Regulations and principles of cooperation, transparency, and openness. Finally, blockchain is not a panacea. Efforts to address market failures and regional capacity building to improve R&D are essential for long-term preparedness (14, 15). ■

REFERENCES AND NOTES

1. S. Moon et al., *Lancet* **386**, 2204 (2015).
2. F. M. Aarestrup, M. G. Koopmans, *Trends Microbiol.* **24**, 241 (2016).
3. J. H. M. Simon, E. Claassen, C. E. Correa, A. D. M. E. Osterhaus, *Bull. World Health Organ.* **83**, 707 (2005).
4. D. P. Fidler, *PLoS Med.* **7**, e1000247 (2010).
5. C. dos S. Ribeiro et al., *Science* **362**, 404 (2018).
6. C. dos S. Ribeiro et al., *PLoS ONE* **13**, e0195885 (2018).
7. M. Koopmans, K. de Lamballerie, T. Jaensich, ZIK Alliance Consortium, *Lancet Infect. Dis.* **19**, e59 (2019).
8. F. Rohden et al., "Combined study on digital sequence information (DSI) in public and private databases and traceability" (Publication CBD/DS/AHTEG/2020/1/4, Convention on Biological Diversity, 2020); www.cbd.int/doc/c/18/1/6/793/574b14ca40cb6468/479584/ds-ah-teg-2020-01-04-en.pdf.
9. J. L. B. Cisneros, F. M. Aarestrup, O. Lund, *Blockchain Healthc. Today* **1**, bhtly1.17 (2018).
10. C. L. Simpson et al., *Int. J. Environ. Res. Public Health* **11**, 8383 (2014).
11. S. J. N. McNabb et al., *Lancet Respir. Med.* **2**, 436 (2014).
12. M. Steichen, B. Fiz, R. Norvill, W. Shbar, R. Slatk, "Blockchain-based, decentralized access control for IPFS," paper presented at the 2018 IEEE International Conference on Internet of Things (Things) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Halifax, Nova Scotia, Canada, 2018, pp. 1499–1506.
13. D. R. Wong et al., *Nat. Commun.* **10**, 917 (2019).
14. WHO, "An R&D blueprint for action to prevent epidemics: Plan of action" (Publication WHO/EMP/PH/2016.02, WHO, 2016); www.who.int/blueprint/what/improving-coordination/workstream_5_document_on_financing.pdf.
15. L. A. Reparant et al., *Science* **346**, 433 (2014).

ACKNOWLEDGMENTS



We acknowledge [1] (10/20), K. Hamilton Duffy, N. Klomp, J. Laros, M. Kroon, J. Flach, R. van der Waal, and anonymous referees for discussion and feedback. M.B.W. and C.S.R. contributed equally to this work. M.B.W., L.H.M.B., and E.C. codevelop blockchain-based solutions for clinical trials (Triall). C.S.R. and G.B.H. codevelop a European platform for detecting and analyzing outbreaks (COMPARE). M.M. is the applicant of a patent on managing IPRs using blockchain.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6492/719/suppl/DC1
10.1126/science.aba1355

Science

Blockchain-facilitated sharing to advance outbreak R&D

Mark B. van der Waal, Carolina dos S. Ribeiro, , George B. Haringhuizen,  and Linda H. M. van de Burgwal

Science **368** (6492), 719-721.
DOI: 10.1126/science.aba1355

ARTICLE TOOLS	http://science.sciencemag.org/content/368/6492/719
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2020/05/13/368.6492.719.DC1
RELATED CONTENT	http://stm.sciencemag.org/content/scitransmed/12/541/eabb5883.full http://stm.sciencemag.org/content/scitransmed/12/534/eabb1469.full http://stm.sciencemag.org/content/scitransmed/11/499/eaat0360.full http://stm.sciencemag.org/content/scitransmed/9/396/eaal3653.full
REFERENCES	This article cites 11 articles, 2 of which you can access for free http://science.sciencemag.org/content/368/6492/719#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Downloaded from <http://science.sciencemag.org/> on May 14, 2020

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2020, American Association for the Advancement of Science