*This abstract was submitted to and presented at CCRC International Conference 2022.*

# Global Health Data with Trust Initiative

Warren S. Gifford, PhD, David L. Turock, PhD, Clifton R. Lacy, MD; Great Plain Technologies, LLC, USA. Corresponding Author: Warren S. Gifford

***Goals of This Paper***

To describe the proposed Global Health Data with Trust Initiative and its benefits to:

- improve privacy and security of individual health data
- give individuals more control over their data
- make appropriate data available for personal health care and public health
- improve research, development, and education.

***Background***

Today, health data privacy laws often force individuals to sign away many rights to their health data to obtain health care, health insurance, and a variety of other services. Ironically, these same privacy laws then prevent organizations with health data from making it readily available for health care and public health research and education.

Data collection and analysis methods and practices have been inadequate to acquire the wide spectrum of information needed to study and address the broad and enduring impacts of COVID-19. Many research studies have included only small numbers of individuals and collected non-uniform data, making it difficult to derive broader aggregate findings[1].

Using conventional data privacy methods could create many risks, as evidenced by the steady stream of enormous personal data breach announcements. Data are stolen or changed, users are impersonated, fake users are created, systems are altered and overwhelmed.

These create many challenges for any health data system.

**SYSTEM DESCRIPTION**

The health screening process, described in the companion paper Global Health Screening and Tracking System Initiative, generates screening test results and uses individual and group historical results, to help determine if an alert should be created for COVID-19 or other conditions, Figure 1.

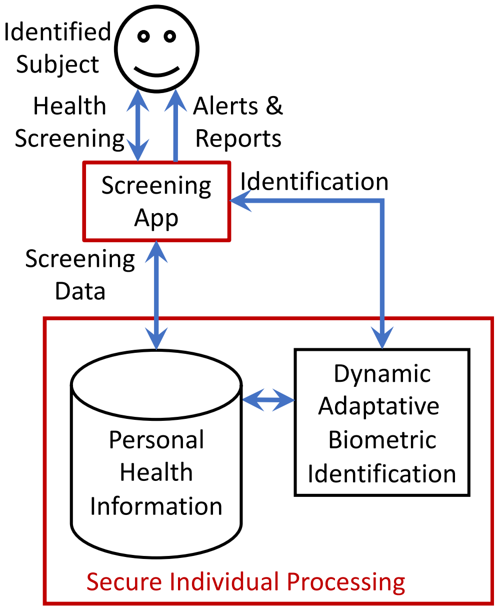
In this system, all personal data is owned by the subject, stored and processed using secure

[1] Groff, D., Sun, A., Ssentongo, A. E., Ba, D. M., Parsons, N., Poudel, G. R., Lekoubou, A., Oh, J. S., Ericson, J. E., Ssentongo, P. & Chinchilli, V. M., (2021, October, 18), Short-term and Long-term Rates of Post-Acute Sequelae of SARS-CoV-2 Infection A Systematic Review, JAMA Network Open. 2021;4(10):e2128568. https://doi.org:10.1001/jamanetworkopen.2021.28568

systems, devices, communications, and operations. Each subject's databases may be stored in a separate system, or in a large data system operating as individual virtual databases but effectively separated from other data. Data redundancy, integrity, data tracking and other essential functions are provided by the system.

**Figure 1**

*Individual Health Data Acquisition*


Identified Subject
Health Screening
Alerts & Reports
Screening App
Identification
Screening Data
Personal Health Information
Dynamic Adaptative Biometric Identification
Secure Individual Processing

***Secure Personal Identification***

Accurate subject identification is essential for privacy, security, and clinical use. Many different biometrics are used today for identification. Unfortunately, static biometrics might be simulated by a photograph, a computer-generated image, or other technique.

A distinctive feature of the identification system is use of biometrics generated as part of the health screening process. The sequence of tests, the nature and timing of the stimuli, and the variety of different features recorded provide a basis for a dynamic adaptive identity record.[2]

Changes in a subject's responses over time are an integral feature of the screening process, making it even harder to simulate a user's characteristics based on outdated information.

A key innovation is the ability of the system to adapt to changing conditions: changes in diseases and improvements in screening techniques and technologies. Thus, even as new hacking techniques emerge, such as using AI to simulate a subject's responses, the system can adapt its practices to counter those attempts. Note this method can be applied to any user of any aspect of the system, not just for health screening.

***Requests and Approval for Data Reports***

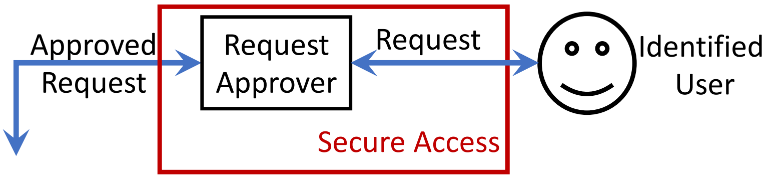
Generating a new data report requires a formal request for authorization, Figure 2. The request specifies the identity of the requester, the data requested, the appropriate processing to create the report to maintain security and privacy of the data, and terms of use such as sharing with others. Each request must be approved by an appropriate approval body.

Requests fall into two major classes: identified or de-identified data. For example, a health worker treating an individual needs specific data associated with that individual and probably wants it processed because they do not need or want the more granular level of detail collected. In contrast, many studies do not need data which identifies individuals; for example, broad statistics on the spread of COVID-19.

[2] Lacy, C. R., Gifford, W. S., & Turock, D. L., (2021, December, 30), system and methods for human identification of individuals with COVID-19 utilizing dynamic adaptive biometrics, New U.S. Utility Patent Application No. 17/565,668

**Figure 2**

*Report Request and Approval*



A primary goal is to assure that individuals cannot be re-identified based on the results not only of a particular request but also from the combination of previous requests and data readily available from other sources. This is challenging, and in general is an unsolved technical problem. We anticipate that approval organizations will be conservative in granting requests, and new methods will be developed to deter re-identification.

Standard formats and characteristics will be developed both for consistency of analysis and ease of assessing the potential for re-identification. Tools will be developed to assist the approval process and review process, such as scanning previous reports, looking for overlaps, frequency, quantization, etc. These may include factors such as statistics of numbers of subjects willing to respond to various types of requests.

Note that the approval process gets an analysis of the output of the request before it is released, so approvers can check for de-identification potential; however, the details may be hidden from the reviewers to prevent inappropriately viewing data. If a request is not approved, the approver may provide guidance to the requester, pointing to information already available, identifying specific concerns and suggesting how to overcome these concerns.

***Request Authorization***

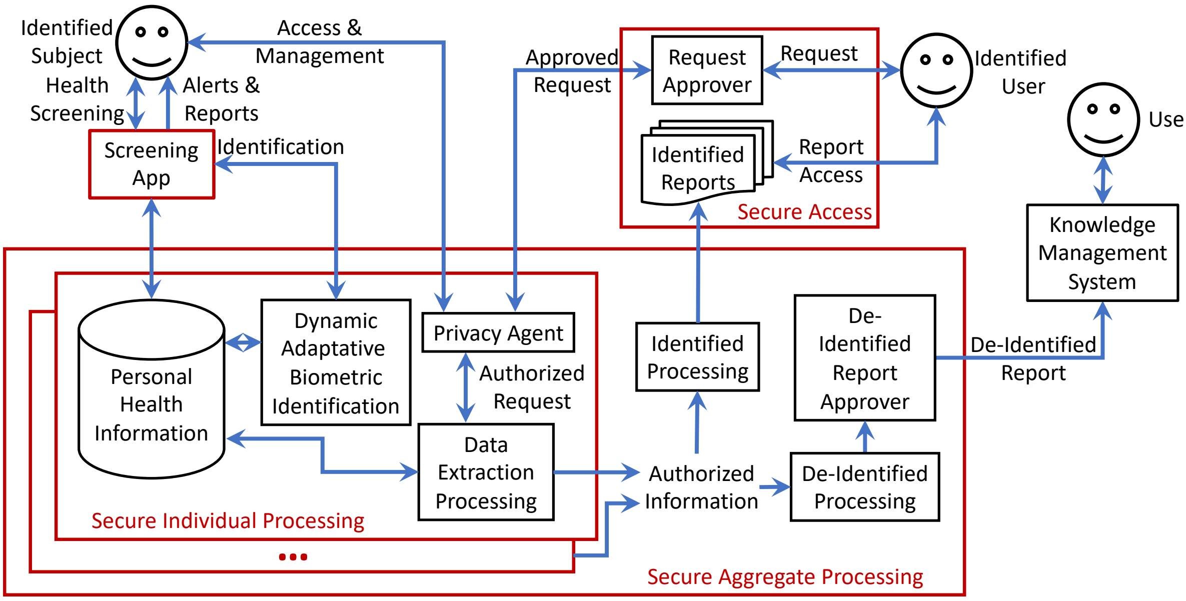
Each approved request for access to an individual's data requires authorization by the individual, either directly or indirectly. In cases such as a health worker requesting access to specific data, the individual may handle the request directly. Some individuals may delegate levels of authorization to other people, such as parents, spouse, or care givers. A Privacy Agent may act on the individual's behalf to provide authorizations, according to instructions by the individual, Figure 3.

The individual can check the validity of any request by querying the approval function. Special arrangements are made for emergency health situations where the user or their delegate is not available.

Each individual controls access to their data for different cases. For example, an individual might grant access to a health care worker for specific types of data over specified time periods and may limit the time access is available – for example, only heart and general health data for the last 3 years. Mental health data may be explicitly restricted. An individual might automatically grant some types of requests for approved de-identified processing, such as national and global statistics. Some requests may be by default outside the user's control, such as data necessary to comply with laws.

**Figure 3**

*Report Generation and Access*



In some cases, the privacy agent may contact the individual to determine what to do with a particular request. For example, the individual may be interested in programs for research on Long-COVID and want to see each request, the data requested, the research program, and how the information will be used, before deciding whether to authorize release or not.

Another dimension of data requests is the involvement of the individual. This can range from no involvement – for example, in general statistics – through releasing specific data that would not normally be released. Active research can include the individual taking some action, such as generating different types of data, changing diet, exercise or using medical items.

If a request is not authorized, an appropriate message is sent describing the reason. For example, the request may time out – that is, the individual has not responded within the time frame allowed in the request. The individual may choose not to answer, or may refuse to release the data, or may only approve release of partial data. Or the individual may not fit the categories specified in the data. Having these options can significantly improve the number of responses and, therefore, the statistical validity of studies.

***Report Generation and Access***

All processing is done using privacy and security systems, and only approved results are made available to the requester, Figure 3. A single request may include data from one or more than one individual. After each individual's system has provided a response, the responses are combined and further processed. Processing for identified reports is relatively straightforward since the release has been explicitly approved. For example, data for several individuals being treated by a health care worker.

For de-identified reports, the processing is likely to be more complex. Typical methods of de-identification include aggregation, quantization, and exclusion of small numbers. After this final level of processing, the report must be approved before it is released. This is similar to the request approval process, but now with the characteristics of actual data to review and compare with previous releases and other data sources.

### *Report Release and Use*

Each release of personal health information is tracked; for example, using techniques such as digital watermarks and explicit tracking. These facilitate tracking the subsequent use of the reports. For example, HIPAA procedures in the US require that each access to protected personal health information be recorded including the identification of the individual, the date and time, the purpose, and the identify and authorization of the person to access this information. In one case, this tracking represents over 30% of the total database volume.

Some of the reports are for public information. These need to be made widely available for consistency of use and to counter false information. One way to do this is to place all the information in a knowledge base system of released de-identified reports. This would help with tracking use of data after it is released as a report. The knowledge base should also incorporate research results, using the data to foster global collaboration on health research. This provides a unique opportunity to engage the public in learning about health and science.

## CONCLUSIONS

The COVID-19 global pandemic has clearly demonstrated the need for more frequent and more detailed health screening to allow each person to protect themselves and society. The pandemic has also clearly demonstrated the need for more detailed and extensive data analysis and research.

The companion paper Global Health Screening and Tracking Initiative describes the tools to perform this essential screening. The Global Health Data with Trust Initiative describes a mechanism to meet the simultaneous mandates of personal privacy and data research.

A quote from the CCRC Cyber Security Conference 2021 illustrates some of the challenges: "Companies don't have the resources, the skills, or the motivation to build resilient and resistant systems."

This is an ambitious proposal. Hopefully the combined benefits of personal health and privacy, balanced with social benefits, are sufficient to push for implementation. The initial stages will be focused on demonstrating technical and policy feasibility.

We are actively seeking partners to fund, build, and operate the initiative.