GS ID GLOBAL SOLUTIONS FOR INFECTIOUS DISEASES

White Paper

Sample Collections for the Detection of Unknown Pathogens

Section I – Background

Infectious disease researchers have always conducted field research and surveillance to collect specimens from humans and animals, and then brought them back to the lab for analysis. Having documented the source of the sample and the results of the analysis in a plethora of formats and information management systems, many samples are then frozen in repositories for future research. Freezers full of these samples can be seen lining the halls of research institutions around the world. We believe that the contents of these freezers represent an under-utilized asset of the infectious disease research community.

In almost every instance, the samples and associated data will have been collected for one primary research project. The process of collecting and categorizing samples can be time consuming and expensive, as are storing and maintaining the samples for future research. While the primary researcher will want to keep some of the samples around for future analysis, the value of many of the samples to the primary researcher often begins to decline with time. Meanwhile, other researchers are often seeking new samples that might provide insight to their own fields of research, whether the emergence of new infections, the evolution of known pathogens, or interesting immune responses that might lead to new drugs or vaccines. Rather than having to go out and collect new samples, a secondary researcher may be able to piggy-back on the existing repositories and collection effort of others.

The challenge is for the second researcher to find those useful samples. While samples are frequently shared between researchers, the majority of the initial interactions come from near-random connections: the chance reading of a paper, attending a presentation at a conference, or an introduction from a mutual colleague. We believe this process leaves too much to chance, and leaves valuable samples sitting in freezers around the world.

Therefore, we propose the creation of a web-based system to help connect researchers and repository owners around the world and share both data and specimens. The creation of such a system represents an opportunity to improve medical and public health research across the board. By providing a gateway to the samples and the data related to them, the system described here may be able to accelerate global research efforts, reinforce quality repository management, and potentially lead to better product development for the management and prevention of infectious diseases.

The three core components of the system are (1) a searchable system of data related to the samples; (2) a system to facilitate logistics for shipping samples for additional analysis; and (3) a network for connecting researchers to discuss current and future research efforts. The system should be simple, scalable, and searchable.

Section II – Landscape Assessment

Before proposing a new system, we wanted to determine if there were already existing system(s) to address some or all of the needs of infectious disease researchers. To do that, we conducted a thorough survey of the literature and a series of interviews with repository managers and infectious disease researchers around the world.

The unanimous conclusion of our interviews is that there is a tremendous demand for the proposed system. Despite the universal interest, we are currently not aware of any single system aimed at sharing data and samples for research on infectious diseases. While there is no single system, a few examples of data and specimen sharing systems did seem relevant.

Biobanking

For years, genetic researchers have been collecting and using specimens from longitudinal and population–based studies and storing them in repositories called biobanks. By searching specimens from individuals over time and across populations, researchers have started to uncover biomarkers for certain non-communicable diseases. Most industrialized countries and many US states have their own biobanking efforts. From the perspective of infectious disease researcher, many of the samples stored in these biobanks are not of use because of the ways that they are stored.

In addition to these individual biobanks, there are a number of efforts to bring together biobanks into networks. In the United States, for instance, the National Cancer Institute has developed the Early Detection Research Network (EDRN). In order to share samples, they developed ERNE, the EDRN Resource Network Exchange (ERNE). With both public and password-protected aspects, ERNE provides a searchable system to identify specimens of interest based on a variety of parameters. In Europe, similar biobanking networks include P3G (Public Population Project in Genomics) BBMRI (Biobanking and Biomolecular Resources Research Infrastructure), and PHOEBE (Promoting Harmonization of Epidemiological Biobanks in Europe).

Even while the specimens and biobanks may not be of use to most infectious disease researchers, these early efforts to share data and samples across institutions and across borders have a number of lessons for developing the infectious diseases network being discussed here. Some of the issues frequently raised include:

- Funding constraints for building and maintaining biobanks
- Lack of comprehensive inventory of relevant biobanks
- Needs for data standardization and harmonization
- Ethical issues relating to the use of samples and informed consent
- Intellectual property and patent issues

- **From Biobanks to Biomarkers**, Wellcome Trust, June 2006

- "'The biggest problem is the way one biobank does its collection, processing, and storage may not be the same as another biobank. And when researchers use tissue or samples from those two different biobanks, the results may not match because they weren't collected, processed, and stored in the same way.'"
- "Linking clinical information to these samples is also becoming a daunting task…The backlog has been in entering data from pathology reports, which are usually written by hand. 'Getting the information into the database in a way that is searchable has been a real challenge.'"

  - **The Changing Face of Biobanks**, Genome Web Newsletter, 30 June 2009

While most of these biobanking efforts are focused on industrialized nations, there have been recent calls for the creation of large population-based studies across sub-Saharan Africa.[1] If these efforts are funded, they would obviously represent an excellent opportunity to add specimens and data to the proposed system.

Animal specimens

When not dealing with human specimens, there are a number of systems that have been built to help zoological researchers share specimens and data. The following are three examples:

- **VertNet** (www.vertnet.org) - VertNet is a searchable online database of vertebrate natural history collections. The current version shares records from 72 institutions and over 84 million records are included. According to the VertNet website, the collections include:
  - Specimens: Whole or parts of animals, including fossils, bones, teeth, pelts, feathers, skins and tissue; all provide clues about the evolution and natural history of species.
  - Associated Data: Field notes, photographs, video and sound recordings, geospatial data, climate data, gene sequences, and related scientific literature.

- **ZIMS** - The Zoological Information Management System has been developed by ISIS, the International Species Information System. ISIS is a non-profit organization that brings together professionals at over 800 zoos, aquariums, and related institutions in over 80 countries. ZIMS is a database system that serves as "a unified, global repository of knowledge on animal health and well-being." ZIMS allows ISIS member institutions to share information on the animals, husbandry, and veterinary care. The current ISIS system relies on standardized, locally maintained DOS-based databases with regularly scheduled submissions to

[1] Holmes MD, Dalal S, Volmink J, Adebamowo CA, Njelekela M, et al. (2010) Non-Communicable Diseases in Sub-Saharan Africa: The Case for Cohort Studies. PLoS Med 7(5): e1000244. doi:10.1371/journal.pmed.1000244.

- **GAINS** (www.GAINS.org) - GAINS was originally developed as the information management system for the Global Avian Influenza Network for Surveillance of wild birds. It provided a web-based open access database and mapping system for samples and test results from any species of bird. The database is now being expanded with funding from USAID's Emerging Pandemic Threats program to handle information regarding samples from any species and any infectious disease. The system is not designed for detailed sample repository inventory management.

Like the biobanks, these animal-oriented systems have some lessons for the proposed infectious disease system. For example, VertNet is clearly facing challenges of standardizing and integrating data across multiple organizations and networks. Recent articles have discussed cloud computing solutions[2], which may have implications for our proposed network. ZIMS also appears to have the potential to be an excellent example of access to relatively real-time data about animals and the results of analytical testing. However, the process of sharing results about animals is much easier than sharing health data about human populations. Likewise, using animal samples tend to be much easier that using human samples, without the need for informed consent of the animal, although permission of the owner of the animal must still be obtained, which in some cases may be limited due to risk of adverse repercussions. Given the importance of zoological origins of many emerging infectious diseases, the ability to connect with these animal oriented systems may provide a useful addition to a human sample network.

Clinical Trial Sample Management Systems

Many clinical trials are currently conducted over a large number of locations, including the collection and cataloguing of samples from across the less developed world. The management of these samples is often done through a system allowing the primary researcher to search across the various repositories, as well as consolidate the data. Most of these systems are used for a single study or series of related studies, and the data is not shared with those outside the study. The ability to integrate these repositories, their data, and their systems should be seen as a critical aspect of building the proposed system.

Conclusion of Landscape Analysis - While none of these systems meet all of the demands of the envisioned system for infectious diseases, they all represent good systems for us to learn from, and they may represent potential partners for future development. Since we were not able to discover an existing system, the following section outlines the requirements of a system that would facilitate the sharing of data and samples for infectious disease research.

---

[2] Constable H, Guralnick R, Wieczorek J, Spencer C, Peterson AT, et al. (2010) VertNet: A New Model for Biodiversity Data Sharing. PLoS Biol 8(2): e1000309. doi:10.1371/journal.pbio.1000309

Section III – Proposed system overview

At the front end of the proposed system, a web-accessible interface would allow users anywhere to search through a variety of available data on infectious diseases. The data would be linked directly to specific specimen collections around the world. Researchers would then be able to contact the specimen repositories directly and request access to samples. Ultimately, they would be able to obtain well-characterized specimens for their own further research, and then add back their own results to enhance the system going forward. In addition to serving as a data and specimen exchange, the system would allow for the development of a research community, sharing information and ideas, as well as proposing new research partnerships and programs.

This section is broken into four parts:

1) User profiles
2) Data related to samples
3) Sample logistics
4) Research networking

1. User profiles

Broadly, the users of this system fall into two categories: researchers and repository managers. It should be noted that many individuals will be both researchers and repository managers. Most researchers have their own collections, while remaining interested in other collections of interest. Likewise, most repository managers are conducting their own research and may be interested in using specimens from other repositories.

Researchers - For the purposes of this paper, researchers are the users who are looking at the available data and then looking to conduct further tests on available specimens. These researchers may be in academic institutions, the government, the private sector, or individual scientists. When they turn to the proposed system, they will be looking for the following information:

- *Well characterized samples* – The system must include accurate data about each sample, including demographic, geographic, and analytical data. The types of data are discussed below.
- *Appropriate samples* – Each researcher will have specific requirements for samples. Such requirements will include the anatomic site from which the sample was collected and the type of material (blood, tissue, etc.).
- *Sufficient quantities of samples* – In order to conduct additional tests on the samples, the researcher must know that there are sufficient quantities of the samples. To

- *Evidence of the samples being well maintained* – Most researchers will want to know that the samples have been properly handled and stored from the point of collection to their current location. Without visits to the repository itself, this may be difficult to assess from a distance, and a user rating system may help identify good and bad repositories over time.
- *Informed Consent* – Researchers will need to know that informed consent was obtained from donors allowing further analysis of the samples. In some cases, simply knowing that the collection effort received approval from an Institutional Review Board (IRB) or Institutional Animal Care and Use Committee (IACUC) will be sufficient; whereas others will require further review of the specific consent forms. Still others will want to know if specimens can be separated from the donor's identifiers ("scrambled") and will IRBs then approve their use by the researcher.
- *Access to the samples* – Researchers will be interested in gaining access to the samples as easily as possible. In almost all cases, contracts and legal agreements will be required to share the samples, as well as clear guidelines and policies for shipping samples.
- *Shipping of samples* – Researchers will want to know if there are any restrictions or challenges to shipping the specimens to different laboratories, especially in international studies.
- *Potential for future collaboration* – While the system will serve as a way to identify existing specimens, researchers may also be interested in identifying repository managers who are interested in collaborating and possibly building new collections.

If the researchers do conduct tests on the specimens, they will eventually be asked to share the results of their research. In addition, researchers may be interested in commenting or discussing the data with other researchers and repository managers. In this way, it is hoped that the researchers will enhance the value of the system going forward.

Repository managers - The repository managers are the users who maintain the specimen collections. While all collections are the result of an initial research project, they are eventually turned over to a repository manager. The repository managers will be required to post the data related to their specimens, as well as respond to inquiries and requests for samples.

Repository managers will only be drawn to the proposed system if they perceive value in sharing their samples with others. This value can be recognized through any number of ways, including acknowledgement in academic publications, participation in potential royalty stream, direct compensation for providing samples, or future collaboration efforts to help maintain ongoing operating costs. This recognition will need to compensate them for their previous collection efforts, as well as their ongoing maintenance of the repositories.

As their collections are always evolving, the proposed system may provide a unique opportunity for repository managers: When one repository manager decides to dispose of part or all of a collection, he may be able to advertise these surplus samples through the system, allowing researchers and other repository managers with different interests or financial resources to offer to take part or all of the unwanted samples. Through this process, the system would help minimize the destruction of potentially valuable samples for the entire research community.

2. Data related to samples

The primary interface for the system would be access to data regarding samples in the system. The following is meant to give an indication of the types of data that would be needed to make the system useful to researchers. The more data included, the more useful the system. A minimum set of data requirements should be established to add a collection to the system.

Reason for collecting - A general description of why the samples were initially collected, including what disease was initially targeted. Examples: clinical trial, diagnostic testing, surveillance study.

Specimen type and quantity -The type of sample(s) taken from each donor is probably the most basic information needed in the database. Examples would include blood, urine, tissue sample, or even entire animals, in the case of some zoonotic collections. The quantities of each type of sample will be important for other researchers. In most cases, conducting further testing will mean that the limited amounts of material will be depleted. Weighing the value of the new experiment versus the permanent loss of the sample will be an important factor in assessing the scientific proposals.

Demographic data - Demographic data relating to the donor is critical information. Some of this may be simple (species, age, gender, ethnicity/race), whereas more complex demographic data may be related to the primary study. For instance, parental information for genetic studies, sexual orientation or activity levels for sexually transmitted diseases, or travel history. While some of the most basic data fields will be common to all samples, there will be a number of fields that are unique to each specimen or study.

Geographic and temporal data - Geographic information about where the samples were collected and temporal data about when they were collected. While GPS systems are increasingly used to geo-locate the origins of many new sample collection activities, many old samples may have more basic data, e.g. collected in Thailand in 1995. In addition, many collections will have samples taken from the same donor over multiple time points. Importantly, these longitudinal collections can allow researchers to study factors before and after infection.

Laboratory data - In almost all instances, the samples will have been processed for some sort of analytical results (e.g. infected / uninfected). The tests will vary widely

depending on the type of sample collected and the purpose of the study. Assuming samples are eventually used by a variety of researchers, the number of test results on any given sample may increase over time.

IRB-approved Informed Consent Forms - In order for a third party to use the specimens, consent forms from the original donors will be needed. Indicating consent forms are available would be a minimum.

Storage location information - Data related to the location of a sample. At the most basic level, this data needs to reflect the current location of a sample, from the institution and geographic location down to the freezer, shelf, box, and vial level. As samples are often divided into aliquots, storage information needs to be tracked for each aliquot of each sample, including those shipped to other institutions. In addition, some researchers will want to ensure the storage history of the sample: starting with the geographic and temporal data of the initial collection, any movement or relocation of the sample may generate new data, especially regarding temperature variations, as certain factors will depend upon a sample being kept frozen, for instance. Some of this information may be useful to the secondary researcher, whereas some of the other data (such as freezer location) will only be useful to the primary research and the repository management system.

Data related to institutions / repositories - Repositories are collections of collections. These may vary in size from one collection to thousands, with the size of each collection also varying. The size of a repository can be measured in a number of ways, from number of samples to number of freezers. In some cases, the freezers may be measured in square feet/meters or storage area, if they are walk-in. In general, the size of a collection may be of interest to other researchers. However, in many cases, it will be the smallest, most remote collections that are of most interest.

Storage condition information (freezers, back-ups, security) - When evaluating the quality of a repository, researcher will be interested in knowing how the collection is managed, and how secure the samples are. For instance, given the fragility of many samples, knowing whether there are generator back-ups for failed electric systems may be critical. Knowing whether there has ever been a power failure during the life of a sample will also impact its value. As the system links to sample collections in more remote and less reliable locations, the value of this data increases. Some repositories may even be able to indicate temperature variance over the life of the samples.

Institutional relationships - In many cases, initial collection efforts are the results of joint efforts between multiple organizations. Understanding the relationship between the parties and the current ownership of samples can be important. As individual researchers move location to location and from institution to institution, the ownership and even location of samples can be difficult to track. Clear indication of the individual and institution with final authority over the use of the samples will be critical. For instance, contact information for the principal investigator (P.I) on the original study should be included, as well as direct contact information at the current repository.

Approval process - Some sense of the approval process at the individual repository should be indicated. Potentially, the ability to submit a Scientific Justification letter (concept sheet) or download sample documentation might be useful.

Contact Information – Contact information for each collection or repository should be included.

Potential additional resources - The results of the original study and any additional studies may have been shared with other systems outside of the repositories. Being able to link these external resources with the samples themselves would be of tremendous value. The following are just a few examples.

- ***Pub Med*** – PubMed is the online library of biomedical research publications, maintained by the National Institute of Health (NIH). Most publications are already tagged by author, date, and key-words. Some of the metadata related to the individual collections and the results of the analytical data is often incorporated in these papers. Linking papers to collections should be a basic function, and extracting data from papers would be considered an excellent improvement, if possible.
- ***GenBank*** – An open-access database of all publically available nucleotide sequences and their protein translations. Historic sequences have been added, and all new sequences generated by NIH funding are required to be posted to GenBank. Sequences are linked to researchers and publications, as well as key-words. Linking sequences to actual samples would be a tremendous resource.
- ***ClinicalTrials.gov*** – A registry of federally and privately supported clinical trials in the United States and around the world. This data would give an indication of ongoing collection efforts, as well as upcoming trials. Obviously, many of the other data (e.g. analytical results) would not be available until after the trial. This connection would provide researchers an indication of potential opportunities for future research and new sources of samples.

3. Specimen logistics system

A successful search through the database would lead a researcher to the set (or sets) of specimens of interest. Once the system makes a connection between the researcher and the repository manager, the following steps would need to be followed in order to obtain specimens:

- **Scientific justification** - Sample collections represent limited resources that get depleted with use (both in volume and in quality if repeated thawing/freezing is required for aliquotting). In most cases, repository managers will require a written scientific justification for the sample request, often called a concept sheet.
- **Legal Agreements** – If the request for specimens is approved by the repository manager, the two parties will need to enter into a variety of agreements, depending on the proposed use of the samples. While each of these documents would need to be negotiated between the two parties individually, the system may

    - Material Transfer Agreements (MTAs)
    - Collaborative Research and Development Agreements (CRaDAs)
    - Intellectual Property / Licensing Agreements
    - Review of Institutional Review Board (IRB) Approval / Consent Forms
- **Shipping and handling** – Having identified the samples and approved the request, the repository managers would begin the process of pulling and aliquoting the requested samples. The samples would then be packaged and shipped to the researcher at appropriate temperature (ice, dry ice, etc.). The shipping of infectious material can be a complex and expensive process. Improper shipment may make the samples useless to the receiving researcher or even any further research. Providing clear guidelines on how infectious material should be shipped would enhance the value of the system.
- **Results** – Depending on the agreement, the results of the research would be shared with the P.I., repository manager and possibly the rest of the research community. Preferably, the results would be uploaded into the proposed system, enhancing its value going forward. Providing an outlet for published and unpublished results about the same samples would be very useful, although most results are not currently shared back with the repository owners, unless published.
- **Future recognition –** Depending on the agreements, the results may be shared more broadly with the research community through publications and presentations. Likewise, any intellectual property or royalty streams would be shared going forward. Unfortunately, many of the repositories indicated that they have not received the proper recognition, in their eyes, for their contribution of samples to certain published papers and results. Providing guidelines and industry standards would help. Likewise, a ratings system for the researchers could be integrated into the system, allowing the repositories some feedback mechanism.

4. Research networking community

In addition to the specimen related data and the specimen sharing, the ability for researchers and repository managers to network would be an important aspect. Researchers should be allowed to comment and discuss data and specimens, or even research elsewhere. Likewise, the ability to discuss future collaboration efforts should be included, as well as post upcoming conferences and events. Finally, there may be ways to advertise and stir up potential interest (and additional funding sources) for new collection efforts or research opportunities.

Section IV – Limitations / concerns

Dispersed data may not be easy to collect - Even within the same institution, it is common for different collections to have stored their data in separate and not-necessarily compatible databases, as well as many different physical locations. As collections grow older, the knowledge of the collection and their storage information is often maintained as institutional or personal knowledge – in order to track down the right specimen, you may have to locate the original researcher or the scientist currently responsible for the collection, sometimes not the same person. Many institutions have repeatedly tried unsuccessfully to gather the information related to their own collections (what is in our freezers and where). Gathering the relevant information from a large number of institutions is a major endeavor that should not be underestimated.

Limited resources at repositories - The initial data entry and development process may entail a significant amount of work on the part of the repositories and their managers. In most cases, their time and resources are already extremely limited. In order to build critical mass in the system, it may be necessary to compensate the repositories for their time to enter the initial data. Once the system is built and the value recognized, additional institutions will hopefully see the value of sharing their data and samples.

Limited samples - While the data can easily be shared with others without diminishing its value, the specimens themselves are a limited resource. There is only so much clinical material that has been collected. The repository manager will have to use personal judgment to determine whether the proposed research is of sufficient value to diminish the limited supplies. In most cases, interested researchers will need to present a brief scientific justification for the use of the samples. Ultimately, the determination of sharing the samples will reside with the individual researchers and their institutions.

Repositories not always acknowledged - In the course of our interviews and based on our experience, we are aware of repositories that have shared their specimens with interested researchers, only to have their collaboration not mentioned in publications. Establishing guidelines for public acknowledgement and co-authorship on papers should be included in the system. Indeed, through a review system, it may be possible to allow individual researchers to rate the collaboration efforts of their partners.

Specimens and national interests - We are aware of a number of institutions and countries that are not interested in shipping their specimens outside their borders. They would rather see the scientists come to their country for work, or to rely on scientists on site. In essence, this approach mandates some tech transfer of skills to labs located in less developed countries. The system could make clear the policies of certain countries / institutions, as well as making suggestions for potential solutions. Ultimately, the development of skilled labs throughout the developing world will lead to more and better research.

Intellectual property - While many of these specimens may be used for pure research purposes, in some cases the hope would be to lead to eventual product development, whether diagnostics, drugs, or vaccines. As with the development of any product, there will be intellectual property issues and potential royalty agreements. While the proposed system cannot address all possible scenarios, we suggest that it include some guidelines for intellectual property. Given that many of the specimens will be shipped across international borders and regions, conflicting IP climates may be involved.

Data results not linked together to samples - Ideally, a researcher would like to know all of the test and studies that an individual specimen has been included in. The results of all of the analytical tests would be stored and accessible based on an identification number. However, most repositories are not able to provide that information. Specimens are often shipped to other locations for the analytical work, and the results may not be shared, or shared only in aggregate. For animal derived samples, results are often linked to the individual animal rather than the sample. A sophisticated system that would allow researchers to track the results of all tests done on an individual as well as the individual sample would be a tremendous innovation and extremely useful to researchers.

Database development issues - While we are not experts in software or database development, we recognize that there will be significant programming issues to address in developing the proposed system. While some repositories may still maintain paper records of their collection, most are using a variety of software systems to maintain the data, including Microsoft Excel, Microsoft Access, SQL databases, and SAS systems. In addition, there are a number of repository management-specific software systems, including Freezerworks, LIMS, and others. Finally, it is not uncommon to come across customized versions of any of the above. In building a new system, a better understanding of the software systems in use will be necessary, as well as the challenges of integrating them into a single, searchable system.

Section V – Recommendations / Next steps

1. **Expand scope beyond emerging infectious diseases –** The initial interest of this project was to identify collections of interest to researchers in emerging infectious diseases. We believe the scope of the project should be increased to include repositories for all infectious diseases. In order to identify a sample with an unknown pathogen, the sample will need to have been tested against a number of known pathogens. The results of those tests (and access to those samples) will be not only be useful for researchers beyond the narrow field of EIDs, but also expand possibilities for discoveries related to EID's by allowing for retrospective studies into possible misdiagnoses and cross-reactions as new test methods are developed. We do not believe that the design of the system would be significantly different to incorporate the broader field of all infectious diseases, and we believe the value of the system would increase with the incorporation of more data and more repositories. Hence, we assume that the system should be built for all infectious disease researchers, not just the field of EIDs.

2. **Expand regions of interest** – The initial project had an explicit interest in collections in South-East Asia (particularly Cambodia, Laos, Thailand, and Vietnam) and in sub-Saharan Africa (particularly Kenya, Tanzania, and Uganda). Just as we propose expanding beyond *emerging* infectious diseases, we believe that the proposed system presents an opportunity to share data and samples from all over the world. We do not believe that the design of the system would be significantly different to incorporate collections from all over the world, and we believe the value of the system would increase with the incorporation of more data and more repositories.

3. **Build database first –** We strongly recommend that the database of samples and their results be built first. From the perspective of a researcher, the value of the system will be based on the quality and quantity of data included. From the perspective of the repository, the value of the systems will be based on the number of researchers using the system. In order to attract researchers, the first step should be to incorporate as much data as possible. We believe there are two routes to build the database:
    a. **Contract key repositories to provide data** – By targeting a few repositories of interest, the database could be built by asking the repository managers to collaborate in building the new database. This may entail hiring teams to go to the repositories or to pay the repositories themselves to add the data to the system. This approach would ensure the highest quality data from each repository is incorporated, as well as becoming rapidly familiar with the process of data acquisition and sharing and any limitations. The system could also be supplemented by adding a scaled down version of the second route below to "prime the pump" with large

b. **Assemble data from publicly available databases** – Alternatively, a system could be built which utilizes information already made public through PubMed, Gen Bank, ZIMS, GAINS and other sources. This approach would require software engineers to build a system, without requiring the time from individual repositories. However, the approach would need to ensure that the new system was an improvement on the existing system(s). This approach may not create new data, but rather just a new platform for accessing existing data. Also, this process does not include any of the researchers or repositories, limiting the opportunity for feedback during the development process.