



Global Alliance for Genomics and Health

Summary of Second Plenary Meeting

October 18, 2014, San Diego, CA, USA

The Global Alliance for Genomics and Health (Global Alliance) held its second major meeting on Saturday, October 18, 2014 in San Diego, California. The purpose of the plenary session was to convene the diverse community of Alliance members and stakeholders, share important progress made by the four Working Groups and projects being accelerated, and collectively discuss next steps for the Alliance to advance and scale up current efforts in genomic data sharing to improve human health.

This summary document includes a brief description of key themes, followed by a more comprehensive summary of each speaker's remarks, featuring goals, accomplishments, next steps, and participants' questions and feedback. In addition, the Appendix of this document includes brief descriptions of each Working Group's all-day planning meeting and the cross-Working Group meeting that took place in addition to the plenary session.

For more information, including presentation materials and videos, see:

<http://genomicsandhealth.org/news-events/events/october-18th-meeting-presentations>.

Key Themes

There was strong consistency in the messages and key issues raised by the presenters, panelists, and meeting participants. These included:

- **Need for Rapid and Effective Action:** The field is transforming rapidly, creating a vacuum that needs to be filled. The Alliance is focused on making quick progress and addressing the biggest challenges in this evolving environment.
- **Progress on Key Tools and Flagship Projects:** The Global Alliance developed key products over the past six months that were needed in the field, and is poised to deliver even more in 2015. By launching flagship projects such as the BRCA Challenge and Matchmaker Exchange, the Alliance is focused on demonstrating value and best practices.
- **Overcoming Technical *and* Regulatory Barriers:** Key tools now in place to tackle barriers to responsible data sharing include the GA4GH Genomics API and the Security Infrastructure. Policies around responsible sharing and use will continue to be challenging, and are a key focus of the Alliance, with the *Framework* serving as the policy foundation.
- **Need to Incentive Use and Handle Misuse:** The Global Alliance needs to actively promote and incentivize use of methods to share data and will identify the best approaches to minimizing misuse of data.
- **Engaging Diverse Stakeholders:** Diverse participation, both globally and sectorally, is vital for success, and a central focus as the Alliance builds and engages membership. The Alliance must also draw linkages not only to researchers and academics, but to the disease advocacy community, private sector, and regulators, acting as a global convener and hub for all involved.
- **Demonstrating Results:** Now that the Alliance has developed a number of key deliverables, we need to demonstrate the results of these actions and communicate about tools broadly.



Welcome Keith Yamamoto

Keith Yamamoto (University of California, San Francisco) opened the plenary meeting, framing the work of the Alliance as remarkable and inspiring within the broader field of genomics and health and at the “leading edge of precision medicine.”

Yamamoto described the Alliance as holding the promise to help spark a revolution across research, health, and healthcare. Yet, the gap between the data available and our understanding of its immense complexity is vast. The challenge confronting the Alliance is how to move through that information deluge, creating a “knowledge network” from the vast amount of data enabling precise diagnosis and treatment decisions for each individual, empowering further research, advancing clinical care, and informing patients and citizens.

Yamamoto outlined three types of reinforcing actions to continue to drive the mission of the Alliance.

First, the field requires significant additional data integration activities, including building the technical ability to integrate, store, and analyze genomic information.

Second, moving to a precision medicine approach will better serve the research and clinical communities. Yamamoto proposed a thought experiment: what if disease was classified by its genetic mechanism? One disease, such as breast cancer or diabetes, might be described by its multiple underpinning genomic mechanisms, just as one mechanism may be implicated in more than one disease.

Research and academic institutions in the Global Alliance network can advance this type of discovery, promote the development of more accurate disease classification, and advance diagnosis, therapy, and care. Operationalizing this research to health continuum would require four elements: an information commons and knowledge network, embracing the goal of precision medicine, building a layered knowledge network across disciplines, and drawing on network connections to carry innovation and discovery to new areas of research and patients groups.

Yamamoto described the power of a precision medicine approach in the case of Traumatic Brain Injury (TBI) in the U.S., where genomics is becoming an increasingly important part of diagnosis and treatment. Instead of the traditional one-size-fits all approach, some researchers are examining not just imaging and clinical data, but patient proteome and genome data and discovering new linkages between TBI expression and an individual’s genetic makeup.

Third, Yamamoto described the need to foster system integration not only between science disciplines, but all along the biomedical continuum, and among all stakeholders.

Yamamoto stressed that the Alliance is foundational on all three of these planes, and is at the lead of being able to develop the idea of precision medicine into practice.

More Information

- [Watch video of Keith Yamamoto’s presentation](#)
- [View Keith Yamamoto’s slides](#)



Global Alliance Progress and Goals

David Altshuler

David Altshuler (Broad Institute of MIT and Harvard), Chair of the Global Alliance Steering Committee, welcomed all attendees to the second plenary, and stressed the important moment for the Alliance. Altshuler reminded attendees that the field is in the midst of a data revolution, and the Alliance is addressing some of its greatest challenges: developing trusted routes for data sharing, effective managing of privacy issues, and respecting the diversity of views for how to drive efforts forward.

Altshuler stressed that the stakes are high, and that it will be up to the people in the room and involved in Alliance to make sure the field develops in the right way before we lose this window of opportunity. But he stressed that there is great potential to unleash a new era of innovation in health.

For over a year, the Global Alliance has been working to accelerate progress in human health by helping to establish a common framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data, and by catalyzing data sharing projects that drive and demonstrate the value of data sharing.

Altshuler discussed how the collaborative, iterative, and transparent model of the Alliance has a history in the technology space, but is newer to healthcare. Nevertheless, the progress of the Alliance has been substantial.

He noted that the Alliance itself has transitioned from a nascent coalition to a formal membership organization, with an established Constitution and composed of 141 exciting and unique organizational members in 23 countries to date. Altshuler also mentioned that diversity is vital to the Alliance's current and continued progress, and one of the priorities for the coalition is to actively to engage an even more diverse set of organizations and individuals in the coming months.

More Information

- [Watch video of David Altshuler's presentation](#)
- [View David Altshuler's slides](#)

Before wrapping his presentation, Altshuler looked to the future, and how a widespread willingness to share data for the greater good is needed to realize the benefits of the work. He looked to the cultural shifts and incentives that will be needed to realize greater sharing, while respecting privacy and security of individuals.

Altshuler closed by describing the Global Alliance as poised to deliver larger-scale products in the coming months, driven by the community's incredible passion and energy to advance human health.



Regulatory and Ethics Working Group

Bartha Knoppers and Kazuto Kato

Bartha Knoppers (McGill University) and Kazuto Kato (Osaka University), the Chair and Co-Chair of the Regulatory and Ethics Working Group (REWG), introduced the Working Group's mission to address the ethical, legal, and social implications of enabling responsible genomic and clinical data sharing. This includes preparing the overall policy *Framework* for data sharing, and developing forward-looking governance policies pursuant to the *Framework* on consent, privacy and security, and accountability.

Knoppers emphasized that progress in the past year has been substantial: the Regulatory and Ethics Working Group launched six active Task Teams, fostered broad international engagement, entered into discussions with multiple international consortia, authored numerous publications, and created a major policy document, the *Framework for Responsible Sharing of Genomic and Health-Related Data*.

The Regulatory and Ethics Working Group Chair described the six active Task Teams launched in 2014 that allow for the benefits of genomic research while protecting individual privacy.

- **Consent Task Team:** developing core elements of consent to enable responsible data sharing, including clauses relating to privacy, international research collaboration, and processes and methods for data storage. In 2014, the Team developed a draft set of Consent Tools in plain English and will develop a Consent Policy in the coming months.
- **Data Protection Regulation Task Team:** analyzing and rapidly responding to data protection regulation development occurring around the world, as it affects biomedical research and sharing of health-related data for research purposes.
- **Data Safe Havens Task Team:** developing policies for platforms for the secure and efficient exchange of clinical and genomic data. The Task Team has developed consensus positions for “Genomic and Clinical Data Sharing Policy Questions with Technology and Security Implications”.
- **Ethics Review Safe Harbor Task Team:** developing an international system that allows for mutual recognition of data access and ethics review. The Task Team is currently conducting a literature review and comparative analysis of ethics policies that will lead to an “ethics review safe harbor” to allow for mutual recognition of data access and ethics standards.
- **Framework Task Team:** successfully drafted and released the *Framework for Responsible Sharing of Genomic and Health-Related Data*, containing foundational data sharing principles and guidelines for research, guided by the human rights of all citizens to benefit from advances in science and its applications, and the right for contributors to be recognized therefore. The Framework has been translated into five languages to date (Arabic,

REWG Key Products and More Information

- [Framework for Responsible Sharing of Genomic and Health-Related Data](#)
- [Consent Tools](#)
- [Watch video of the REWG presentation](#)
- [View REWG presentation slides](#)



Japanese, Chinese, French and Spanish) and is published Open Access in *The HUGO Journal*.

- **Privacy and Security Policy Task Team:** developing a Privacy and Security Policy for genomic and clinical data sharing in line with the *Framework* and core Global Alliance policies and recommendations.

Currently, the REWG is establishing policy-specific Task Teams to draft content-specific policies guided by the *Framework*. This includes defining the key attributes of a data “safe haven” for secure and trusted genomic and clinical data sharing, a review of ethics review regimes, and developing a “points to consider” policy for establishing ethics review equivalency across intuitions. Knoppers further announced that two new policy-specific Task Teams, one on Accountability and the other on Privacy and Security, are in the process of being formed, based on need and strong member interest.

Knoppers concluded her presentation by explaining that the REWG is poised to continue advancing responsible data sharing in three avenues:

- 1) Supporting the creation of efficient and secure deposit, curation, and access mechanisms for available datasets;
- 2) Establishing a system to respect the rights of participants, researchers, and others; and
- 3) Exploring the possibility from a regulatory and ethical perspective of cross-border data linkage with health care providers and globally unique content-based identifiers for genomic research and medicine.

Questions from the plenary attendees included which policies and international developments the co-chairs are most closely watching. In response, the REWG Co-Chairs described concerns and policy changes regarding the protection of personal data, not only in the European Union, but in the United States, and around the world. Co-Chairs invited members of the Global Alliance to channel information on personal data protection policies and news to the REWG to ensure real-time engagement and response in this dynamic area.

REWG Next Steps

- Developing three Policies: a Consent Policy, a Privacy and Security Policy, and an Accountability Policy, all for release in 2015, to action the *Framework*
- Ethics Review Safe Harbor Task Team conducting review of ethics policies

Security Working Group

Paul Flicek and Dixie Baker

Paul Flicek (EMBL – European Bioinformatics Institute) and Dixie Baker (Genetic Alliance), Co-Chairs of the Security Working Group (SWG), described the need to lay a foundation for securing a data sharing commons, and to create a technology environment that provides individuals, researchers, clinicians, and other stakeholders assurance that data made available are shared, annotated, and interpreted only by appropriately authorized persons and entities in accordance with Global Alliance security and privacy policies. The Co-Chairs described their goal to help create a technology environment that provides such assurances, and the Global Alliance as a key enabler for this type of productive security ecosystem to emerge.



The major product being developed by the SWG is the Global Alliance Security Infrastructure, Version 1.0. The objectives of the Security Infrastructure are to manage five types of genomic data security risks: 1) unauthorized disclosure of stakeholder data, 2) unauthorized access or use of stakeholder data, 3) corruption or destruction of data, 4) disruption or degradation of services supporting availability and access to data, and 5) that inappropriate actions result in security incidents that diminish participation in the Global Alliance ecosystem.

The Security Infrastructure provides recommendations for security technology infrastructure to support Global Alliance privacy and security policies, Baker announced that Version 1.0 of the document has been published on the Global Alliance website. The Infrastructure is a “living document,” and as such, has been published with a link to enable submission of comments.

Looking to the future, the Co-Chairs described four primary areas of focus for the SWG. The first is establishing a collaborative effort for security incident reporting. Second, the SWG plans to develop standards for consent management. Third, the Working Group looks to test the proposed Security Infrastructure against existing and emerging sharing activities, particularly drawing on the expertise and insights of other working groups. Finally, the group will provide cross-cutting advice to other Alliance Working Groups and initiatives to ensure harmonized approaches to responsible data sharing. Three new SWG Task Teams were also proposed: Incident Response, Software Security, and Cloud Security.

In response to a question about the amount and type of data that can be released before a privacy breach, Flicek discussed reframing the concern from creating uniqueness from a data breach to doing harm. Whereas creating uniqueness is a technical concern, the question of when a data breach crosses over into harm is a policy one that should be further discussed. When asked a question on incident handling, the SWG Co-Chairs described the rapidly changing area of application security and, from a technical perspective, that such measures must be embedded in any data sharing approach.

SWG Key Product and More Information

- [Security Infrastructure Version 1.0](#)
- [Watch video of the Working Group's presentation](#)
- [View the presentation slides](#)

SWG Next Steps

- Establish three new Task Teams for Incident Response, Software Security, and Cloud Security
- Solicit feedback on the Security Infrastructure, Version 1.0
- Provide security support to other GA4GH Working Groups and projects



Data Working Group

David Haussler and Richard Durbin

David Haussler (University of California, Santa Cruz) and Richard Durbin (Wellcome Trust Sanger Institute) introduced the Data Working Group (DWG), describing the work of the group to overcome siloed data in incompatible systems and break down barriers to collaboration.

A global, accessible, and collaborative way of working defines the broader mode of operation of the DWG, which utilizes an open source software development environment at <https://github.com/ga4gh>. All individuals are welcome to participate and decisions are made nimbly by those most active in the open source development process.

In the past year, the DWG has launched five task teams to overcome key barriers to the technical sharing of genomic data:

- **File Formats Task Team:** custodians of the SAM/BAM, CRAM and VCF/BCF file formats, responsible for standards definition, evolution, and reference implementations. A current goal of the Team is to promote adoption of CRAM for data-sharing. The CRAM format saves approximately 30% disk space over BAM with lossless compression, but more importantly enables much greater future savings through managed data “smoothing”..
- **Genomic Reads Task Team:** providing Application Programming Interfaces (APIs) to interoperably store, process, explore and share DNA sequence reads across multiple organizations and on multiple platforms. Lead developers of the GA4GH version 0.5 API. Next steps include releasing a reference implementation of the 0.5 API, developing browser integrations (3 are currently underway), and developing batch processing integrations (currently for Spark and Google Dataflow).
- **Reference Variation Task Team:** developing and standardizing ways in which genomes are represented and compared, and how genetic variants are reported. Basic variant reporting relative to the current human genome reference is already featured in the version 0.5 API. The longer-term goal is to develop with others a game-changing human genetic reference that includes all common human genetic variation and provides a stable and consistent method of representing both simple and complex genetic variation.
- **Metadata Task Team:** focused on “everything but the sequence,” this Task Team is building representations and exchange formats for the non-genetic and non-clinical information that comes along with genomes. The Task Team is preparing to turn to ontology work with the Clinical Working Group, model refinement, and validation and data transfer tools.
- **Benchmarking Task Team:** creating test suites and methods to measure the accuracy and efficiency of genomics software. This Task Team has made dramatic progress since launching in June. They are actively developing benchmarking tools that include a variant comparison engine and working with leading groups to host data.

DWG Key Product and More Information

- [Genomics API](#)
- [Watch video of the Working Group's presentation](#)
- [View the presentation slides](#)



The Co-Chairs also announced the Working Group will incubate four new task teams, building on the success of current efforts:

- **RNA and Gene Expression Task Team:** will provide APIs to interoperably store, process, explore and share RNA sequence reads, computed transcript structures, and their expression levels
- **Genome Annotation Task Team:** will develop standardized ways to represent information associated with particular regions of the genome such as presence of functional units or disease-associated variants
- **Genotype2Phenotype Association Task Team:** will formalize the language and methods used to represent different kinds of genotype-phenotype associations and how confident we are in our assessment of these associations
- **Containers and Workflows Task Team:** will concentrate on methods that increase the portability of code and the reproducibility of computational analysis

The Co-Chairs further described an exciting idea that was discussed at their group workshop the preceding day: Globally Unique Content-based Identifiers or Digests. Haussler raised the idea of producing concrete technology so that any genome dataset in the world can have an abstract identifier that is 1) unique, 2) privacy preserving, 3) not centrally assigned, 4) independent of the computational representation of the data, and 5) unforgeably linked to the content of that dataset. This identifier system, based on a cryptographic hashing method, would be vital for verification, de-duplication, and for auditable tracking of reproducible analysis and inference. The ideas extend to other data types beyond genome sequence data.

When asked about the DWG's engagement with research funders, the Co-Chairs responded that the group is engaging additional funders in several countries who are eager to participate in the Group's activities. David Altshuler added that the Alliance Secretariat is working to secure funding to support these advancements, and asked the Alliance community to put the group in touch with additional funding agencies.

When Alliance members at the plenary raised that the rare disease community is also considering the development of globally unique identifiers, the distinction between identifiers for individuals and identifiers for datasets was discussed (the digest concept is the latter) and it was proposed to review the cross-working group implications of these efforts and report out on this issue in the coming months.

In response to a final question of how available the tools of the DWG are to the broader global community, DWG Co-Chairs explained that these tools are available to anyone who can access the Internet, but acknowledged that development of tools must be combined with education and outreach to increase contributions and uptake. Once the reference implementation is complete, Haussler stated that the DWG will be in a stronger position to reach out actively and broadly to potential users.



DWG Next Steps

- An open source reference implementation for the Genomics API in 2015
- Tools including a benchmarking suite and variant comparison engine in 2015
- Fully launching four new Task Teams: RNA and Gene Expression, Genome Annotation, Genotype2Phenotype Association, and Containers and Workflows
- Exploring the concept of dataset Digests with the other Working Groups and outside organizations

Clinical Working Group

Kathryn North

Kathryn North (Murdoch Childrens Research Institute) began by noting that the Clinical Working Group (CWG) is driven by one question: how do we represent phenotypic data and link it to genotypic information? The goal of the CWG is to address both the research and clinical use of genomic data, utilizing an approach that is physician-oriented, researcher-focused, and patient-centered.

North described the major activities of the CWG, including the current mapping of initiatives that promote data sharing. In addition to working closely with the BRCA Challenge and the Matchmaker Exchange, the CWG hosts the following four task teams.

- **Catalogue of Activities Task Team:** mapping current initiatives that promote or enable data sharing, starting with Mendelian Genetic Disorders and turning next to cancer, because of the immediate therapeutic implications, and an eHealth catalogue of activities.
- **Genomic Data and Electronic Health Records Task Team:** working to better integrate genomic medicine and research into electronic health records. The CWG is currently mapping existing and emerging initiatives with the Global Genomic Medicine Collaboration (G2MC) and others.
- **Clinical Cancer Genome Task Team:** working to harmonize the many sequencing efforts in the cancer community by promoting standards and best practices that support clinical decision-making. Near-term goals of this Task Team include working closely with existing initiatives.
- **Phenotype Ontology Task Team:** bringing together international efforts to develop and promote standard language and tools for recording patient clinical phenotypes for diagnostics and translational research.

CWG Key Product and More Information

- [Catalogue of activities for Mendelian Genetic Disorders](#)
- [Watch video of the Working Group's presentation](#)
- [View the presentation slides](#)

In response to a question about the next steps for linking the CWG's cutting-edge genomics work to electronic health records, North responded a likely path would be selecting a few demonstration projects, sparking a lively discussion about the need to involve major private electronic health



records providers (several of which had already been invited to participate). Alliance leadership agreed that while the CWG will remain technology agnostic, the participation of the private sector is of critical importance.

A second area of discussion centered on a question of integrating non-human model organisms into the work of the CWG. North agreed to explore this important idea, which was also raised in the context of the DWG's Metadata Task Team.

Plenary attendees also raised questions about better integrating disease into the CWG's approach. In response to a question about the place of infectious disease in the CWG, the Alliance agreed to convene a group to explore the idea of an Infectious Disease Working Group or Task Team in the next months. Those interested in exploring this idea were invited to volunteer or suggest others who should be involved.

Finally, in response to a question about how the CWG plans to engage the patient community into the future, North looked to explore ways to encourage the patient community to be more active in responsible sequencing and data sharing.

CWG Next Steps

- Catalogue of Activities Task Team is developing a catalogue of eHealth activities
- Expanding phenotype ontologies
- Developing Demonstration Projects to incorporate genomic data into Electronic Medical Records across multiple sites
- Exploring integrating infectious disease, the private sector, the patient community, and non-human model organisms into activities



BRCA Challenge

John Burn and Stephen Chanock

Sir John Burn (Newcastle University) and Stephen Chanock (National Cancer Institute), BRCA Challenge Steering Committee Co-Chairs, introduced the newly launched BRCA Challenge, the mission of which is to translate the rapid expansion of sequencing capacity into useful knowledge and, in particular, learn how to rapidly interpret variant data to generate clinical utility.

The Steering Committee Co-Chairs described the BRCA Challenge as a vanguard effort to aggregate BRCA1 and BRCA2 data in order to understand variation and its impact on human health, while demonstrating how to approach large datasets for other disease areas of study. Since its recent formation, the BRCA Challenge has formed a Steering Committee and is formalizing strategic goals, a structure, key deliverables, and an aggressive timeline for progress.

The BRCA Challenge announced its first steps will be ensuring that three major datasets (ClinVar, LOVD, and UMD) are queryable. The group then plans to expand to include other datasets (like ENIGMA, CIMBA), and seek out additional sources of data to add to the Challenge as well.

BRCA Challenge More Information

- [Watch video of the presentation](#)
- [View the presentation slides](#)

Burn and Chanock proposed releasing the three deliverables:

1. Population based assessment of allele frequencies of variants using available sequencing resources;
2. Federated data collection of Pathogenic Variants for BRCA1/BRCA2, both building a structure for data sharing and establishing a common terminology; and
3. (In the long term) Improve and refine information for use in clinical interpretation of mutations

The Co-Chairs concluded that advancing the BRCA Challenge, an effort that has captured the imagination of many in the room and around the world, in the longer term requires sustaining infrastructure, retaining expert leadership, and providing opportunity for grants and contracts to advance knowledge.

BRCA Challenge Next Steps

- Steering Committee laying out goals, deliverables, and a timeline for advancement
- Ensure three major datasets are able to be queried in 2015, with further expansion

Matchmaker Exchange

Heidi Rehm

Heidi Rehm (Harvard Medical School) discussed Matchmaker Exchange, one of the key projects that the Global Alliance is working to accelerate. Matchmaker Exchange is already working to bring together cases with overlapping phenotype and candidate genes both to gain research insights to return information to patients.



Rehm described how conceptually, matchmaking occurs when a submitter queries all API-linked databases with gene candidates, disease name, and the submitter's. If a match occurs, the depositor and the requestor are both notified and details of case are shared manually to begin follow-up studies to validate the match.

Rehm announced two major developments with the project. First, the development of API Version 1.0, which is already forming an interface between several distinct genomic databases. Second, Rehm announced the creation and launch of a new website www.matchmakerexchange.org which will form the basis upon which the group communicates with the broader community of interest.

Matchmaker Exchange Key Products and More Information

- New website at www.matchmakerexchange.org
- [Watch video of the presentation](#)
- [View the presentation slides](#)

She noted that these accomplishments are being accelerated through close collaboration with each of the Alliance Working Groups: the DWG providing support for the API development, the REWG providing feedback on consent protocols, the SWG providing guidance on query authentication, and the CWG sharing expertise with phenotyping for matchmaking.

To build on this initial progress, the Matchmaker Exchange project plans to finalize the API in conjunction with the GA4GH Data Working Group and to develop guidance for groups without a database wishing to choose a site for data deposition and matchmaking support.

Matchmaker Next Steps

- Finalize API for database queries
- Develop guidance for groups without a database seeking matchmaking support

Beacon Project

Steve Sherry

Steve Sherry (National Center for Biotechnology Information) presented the accomplishments of the Beacon Project, a project first proposed at the Alliance's March 2014 plenary meeting as a way to demonstrate the leadership, institutional permission, and technical ability to share data. Beacons also show where data is present, and where gaps in mapping may need to be addressed.

Sherry noted that since the initial proposal in March 2014, fifteen beacons are now active around the world, with more to come. The beacons themselves are an API query that indicates if a simple allele is present in an affiliated institution's data repository, and the group announced it will soon release a Beacon of Beacons, in which a single beacon query searches each database in a federation of repositories. Sherry issued a call to the Alliance community to join in the project to link additional dozens and hundreds of data repositories to the effort. The Beacon Project is also collaborating with data repositories to finely tune beacon queries to adhere to individual repository concerns about data sensitivity or consent requirements.



Overall, the Beacon Project has established three tiers of beacon queries, based on the strength of an affiliation between users and data repositories. The most basic relationship is with an anonymous user who can gain access to genomic information that can be provided without privacy concerns. Sherry mentioned that users who provide a verifiable name and institutional affiliation may access additional genomic information that may be scientifically useful, and in the final case, a user may enter into a binding agreement for access to one or more specific datasets, including the commitment to abide by clear and robust privacy standards, and gain full access to the relevant data.

Beacon Key Products and More Information

- [Beacon project database](#)
- [Watch video of the presentation](#)
- [View the presentation slides](#)

Plenary attendees raised several questions about protections for data and the possibility of donor identifiability in the case of forensic or law enforcement queries and other scenarios. Currently, beacons only point users to aggregated information from a group of studies, and not to individual level data. However, Alliance leaders agreed that even providing aggregate information is delicate, requiring guidelines for what types of aggregated data can be released, its implications, and the risks of identification.

Beacon Next Steps

- Project team will finalize and release a Beacon of Beacons in the near future
- Global Alliance will convene a group to further develop guidelines for data donor privacy protection



Panel on Enablers and Challenges to Data Sharing

The final session during the plenary meeting was a panel discussion moderated by **John Mattison (Kaiser Permanente)** and featuring panelists **David Glazer (Google)**, **Brad Margus (A-T Children's Project)**, **Johan den Dunnen (Leiden University)**, **Cindy Bell (Genome Canada)**, and **Lana Skirboll (Sanofi)**. These diverse panelists were selected because of their different vantage points throughout the world of genomics and their alternative perspectives on big, new opportunities to advance the Global Alliance.

John Mattison, as moderator, led off the discussion noting that while genomic complexity is vast, putting datasets together as the previous presentations described allows us to understand incredibly complex interactions in a more meaningful way. Mattison noted that privacy concerns are still of critical importance to this field, but over time, social norms may evolve and shift the value proposition in favor of increased sharing.

Lana Skirboll, who spent her career at NIH before entering the private sector at Sanofi, raised a number of key issues for the field. She described industry as being in the middle of a transition, increasingly recognizing the importance of data sharing and open innovation through increased engagement with public-private partnerships, working with academics, and others. Skirboll mentioned that governments are also transitioning, as they begin to realize that data are an economic asset, restricting movement across borders, and that these are political developments to pay attention to and engage in the context of trade treaties and other means.

Skirboll also raised the issue of law enforcement access to data, mentioned in an earlier presentation, and believes that access is inevitable and suggested this community focus on identifying the risks and defining informed consent policies. Another emerging area of data protection is for the data emerging from the exploding field of real-time, personal sensors. Finally, in this shifting world, Skirboll recommended that the Global Alliance consider working more closely with regulators to ensure approaches are informed and aligned.

David Glazer described what he called a "thought-experiment" to illustrate key areas of focus for the Global Alliance. Glazer asked: If the genomics and health community wanted to put together a database of 100,000 genomes that was useful and available in a matter of months, what he called a Tree of Life for researchers, what would be required?

Glazer noted that the first area to tackle would be the technical sharing of data, and this would likely require a federated project to connect data repositories and quickly get to scale. Data hosts would form the branches of the structure, filled in with sets of data provided by many data collectors and contributors. Contributors could include highly mobilized individuals and groups, like Autism Speaks or Brad Margus and the A-T Children's Project, whose energy and resources would be vital. But Glazer emphasized that this is already technically doable.

What is more challenging, Glazer argued, is the policy, access, and consent rules that govern the sharing and use of this information. Glazer concluded that in his opinion, the best way forward on the policy side is to develop a portable consent process where data donors can choose to make their data available to all qualified researchers, without the researchers needing to apply for access to every "branch of the tree" individually.

Brad Margus shared his family's own struggle with disease, as two of Margus's young sons were diagnosed with ataxia-telangiectasia (A-T) a rare and devastatingly degenerative disease. Margus has since founded the A-T Children's Project and devoted himself to coordinating and seeking funds



for research on A-T. Recent advances in data sharing helped Margus to identify individuals whose genomic information may hold a key for understanding more about this disease. Furthermore, Margus has begun raising funds to perform additional sequencing, another avenue that has only been available in recent years. From the perspective of a leading disease advocate, Margus reiterated the importance of standardized, flexible consent forms to aid in the responsible collection and use of genomic data to fight disease.

Cindy Bell described her perspective as a funder in the field of genomics and health. Echoing Keith Yamamoto's opening remarks, Bell described the genomics and health community as being at an inflection point, where additional impact must be demonstrated. Bell praised the incredible energy of Global Alliance contributors, recognizing the voluntary efforts that have driven much of the coalition's progress to date.

Yet Bell acknowledged that volunteerism, while a powerful demonstration of commitment, is not sustainable in the long run and that targeted investments are vital. Bell described her organization's own launching of a pilot project to support the objectives of the Global Alliance in Canada and encouraged similar investment. Remarking on the shifting field, Bell concluded that costs have shifted in recent years from sequencing to data management and infrastructure, and noted that funders are increasingly supportive of public-private partnerships to achieve impact.

Johan den Dunnen, the final panelist to offer remarks, spoke next from the perspective of database management, as the founder of the Leiden Open Variation Database. He stressed that DNA diagnostics, DNA knowledge, is based on sharing information on genes, variants and phenotypes, and emphasized sharing. Years ago, surprised and frustrated by the lack of data sharing, den Dunnen embarked on a remarkable personal project to create what is now LOVD, an open source database of DNA variants. Den Dunnen described a culture that is shifting, but yet still largely hasn't overcome resistance to uploading data or attracting funding for this essential work. He suggested making sharing data obligatory by law with associated payment to host and curate the data.

After these brief panel remarks, a question from the audience about how the Global Alliance might successfully handle instances of data misuse and regulate the field sparked a lively discussion about responsibilities and priorities. Alliance Steering Committee Members, including Altshuler and Knoppers, agreed this is an issue of great significance for the future of the Alliance, and agreed that this would be a future focus of the effort to discuss how best to handle these issues as a group.



Closing Remarks

Martin Bobrow

Martin Bobrow (University of Cambridge, Emeritus) delivered closing remarks at the plenary meeting, summarizing the productivity of the coalition and energetic mood of the meeting. Bobrow described the Alliance as both a philosophical discussion and social movement for responsible data sharing, connected to an even broader network of researchers, clinicians, patient's advocates, and more.

Ultimately, Bobrow said, the Alliance will succeed or fail not on its philosophy, but rather on its ability to produce transformative ideas and products that are both high quality and highly relevant. To have this impact, the challenge for the Alliance is choosing the most impactful areas of genomics and data sharing to focus on in the coming months, amidst so many promising avenues. Many new and exciting ideas were shared at the plenary meeting today, and members of the Alliance will be taking up a number of them.

Bobrow pointed out that one of the successes of the Global Alliance is that so much has been achieved to date through an essentially voluntary effort, proof that this work is highly important and interesting to this field. Although, to have transformative impact over time, this work must be sustained by funders.

Bobrow concluded by thanking the attendees, participants, and Alliance leadership, and invited all to the next plenary meeting in 2015.



Appendix: Working Group and Cross-Cutting Projects Meeting Summaries

Clinical Working Group Sunday, October 19, 2014

Working Group Progress

The Clinical Working Group continues to evolve. The Catalogue of Activities for Mendelian Genetic Disorders represents a major deliverable in 2014, on which several subsequent Work Products will be modelled. The following Task Teams reported on their work and next steps:

Task and Project Team Work

Phenotype Ontology Task Team

- Will assemble a multidisciplinary group to promote and advance the creation of phenotype ontologies in the areas of cancer and common disease.

Clinical Cancer Genome Task Team

- Will convene a group of leaders involved in known cancer data sharing pilots and efforts in order to share best practices and assess whether there is opportunity to collaborate.
- Will create a list of activities (Catalogue of Activities) for patient reported outcomes and crowd sourcing.

eHealth Task Team

- Many tasks were identified and assigned to move forward including:
 - o Define and prioritize various use cases;
 - o Define challenges and opportunities for analytics/interpretations when operating across data types;
 - o Create a Catalogue of Activities specifically for eHealth;
 - o Provide guidance for low concordance of exomes;
 - o Identify issues for pedigree data collection;
 - o Establish open source / crowd-sourced commons;
 - o Identify data storage issues;
 - o Establish methods to identify conflicting interpretations;
 - o Identify best practices for family history data collection; and
 - o Summarize and track the evolution of patent policy issues.

Future Goals and Points of Agreement

The CWG is looking to engage the broader interested community and solicit advice and feedback in San Diego. A few recommendations raised in San Diego include: creating a website similar to the DWG Github to showcase current work and more easily allow others to participate, formalizing a communication strategy (e.g. regular newsletters, greater engagement with other parts of the world such, etc.), formalizing cross communication among Working Groups, convening meetings in other parts of the world, and generating greater engagement with clinicians and scientist. These comments reflected both an awareness on behalf of the CWG as well as the desire on behalf of the interested community to see that the CWG is more transparent and more accessible.



Next Steps

Potential next steps that were discussed included:

- Creating a somatic equivalent to the BRCA Challenge (i.e. BRAF);
- Defining clinically actionable genomes; and,
- Returning incidental germline findings.

Before deciding next steps in these areas, the CWG will investigate what efforts currently exist and then determine whether to create a new deliverable/project/task team.

Data Working Group

Friday, October 17, 2014

Working Group Progress

In 2014, the Data Working Group released the GA4GH API version 0.5 and the API continues to improve and evolve. There is broad enthusiasm for moving both existing and new the initiatives forward. The following Task and Project Teams reported on their work and next steps:

Task and Project Team Work

Reads and Reference Variation Task Team

- A graph data model is a work-in-progress; a v0.1 graph API using this data model to be defined.

Benchmarking Task Team

- Developing tools to benchmark performance of human SNP and small indel calls.
- Developed a document describing definitions of SNP/indel performance metrics and how to stratify performance by variant type, sequence context, and functional region.
- To develop a variant comparison engine over the next 3 months, which will have the ability to do vcf-vcf and vcf-database comparisons.

File Formats Task Team

- Official GA4GH adopted file formats: BAM/CRAM/VCF/BCF.
- Moving forward, promote adoption of CRAM.

Metadata Task Team

- Progress moving forward and aligning several data elements as well as API development.
- Next step will be to apply the metadata specifications to driving biological use cases (e.g., BRCA, or cancer in general).

Beacon Project

- New NCBI Beacon service offers responses **with Public, Controlled and Full levels** of access. Properties of answer may depend on relationship between user and repository.
- Next steps: further develop the API and attract new beacons.



RNA and Gene Expression Task Team (NEW)

- provides APIs to interoperably store, process, explore and share RNA sequence reads, computed transcript structures, and their expression levels

Genome Annotation Task Team (NEW)

- Develops standardized ways to represent information associated with particular regions of the genome such as presence of functional units (coding exons, regulatory elements, etc.) or disease-associated variants.

Genotype2Phenotype Association Task Team (NEW)

- Formalizes the language and methods we use to represent different kinds of genotype-phenotype association and how confident we are in our assessment of these associations.

Containers and Workflows Task Team (NEW)

- Concentrates on methods that increase the portability of code and the reproducibility of computational analyses

Future Goals and Points of Agreement

The concept of globally unique content-based identifiers or digests was defined. Any version of any genome sequence dataset (or other large dataset) in the world can have an abstract identifier that is:

1. unique for that dataset version (no “copy” of that sequence+metadata at any other location at any time in the future will ever have a different identifier, and no two different versions will ever “collide” by accidentally getting the same identifier)
2. content-dependent (the same, i.e. semantically identical, genetic sequence content stored in a different format will have the same content-based identifier)
3. privacy preserving (given only the identifier you can’t determine anything significant about the genome the data are derived from unless you are allowed to use that identifier to retrieve the genome itself)
4. not assigned by any central authority (each medical center can generate their own ids)
5. unforgeable (if I send you Jill’s genome with Sarah’s id, you will know immediately that something is wrong)

Next Steps

- Consensus on the addition of four new Task Teams (listed above) in addition to the existing five Task Teams.
- Overlap with key projects is likely to increase.

Regulatory and Ethics Working Group

Sunday, October 19, 2014

Working Group Progress

In 2014, the Regulatory and Ethics Working Group published the *Framework for Responsible Sharing of Genomic and Health-Related Data*. It is referenced in the Global Alliance Constitution. The Working Group also published Consent Tools (see below).

Task and Project Team Work

Framework Task Team

- The *Framework* has been translated into 5 languages to date (Chinese, Japanese, French, Spanish, and Arabic) and is currently being translated into Greek and German. It is available open access on the Global Alliance website and in *The HUGO Journal*.



Data Safe Havens Task Team

- Developed consensus position statements for “Genomic and Clinical Data Sharing Policy Questions with Technology and Security Implications.”
- Findings have been relayed to the Consent Task Team and the Privacy and Security Policy Task Team.

Consent Task Team

- In concert with P3G-IPAC, three Consent Tools have been prepared for international data sharing (legacy consent; consent clauses; consent template); all require adaptation according to local social, cultural and legal specifications.
- Currently developing a Consent Policy in line with the Policy template in Appendix 2 of the *Framework*.

Ethics Safe Harbor Task Team

- Developing an “ethics review safe harbor” that would allow for mutual recognition of data access and ethics review.

Data Protection Regulation Task Team

- Keeping track and responding to global data protection regulation developments and actively working with the Beacon Project group on sensitive data issues and a privacy test.

Privacy and Security Policy Task Team

- Developing and Privacy and Security Policy that operationalizes the Foundational Principles and Core Elements of the *Framework* by offering tools, benchmarks and best practices to guide and inform responsible data sharing and governance processes as it relates to data privacy and security.

Future Goals and Points of Agreement

The Consent Task Team is also currently discussing with GSK representations/Harvard MRCT participants how the Consent Tools can inform consents for clinical trials and industry standard consent forms. Additionally, the Consent Task Team is looking at the concept of **machine-readable consents** (noting that HL7 is working on machine readable consent directive, as is the World Economic Forum).

In the future, the Data Protection Regulation Task Team plans to explore genomic cloud computing and data protection issues.

Noting that a paper has been written by REWG Executive Committee member Paul Burton and colleagues (currently undergoing peer review) that explores the topic of “data safe havens”, it was agreed that a “data safe haven” is a place where data can be stored and accessed by all types of groups, that can be trusted by all parties, and that hosts genomic and clinical data, in both open and controlled formats.

On the subject of identifiers, there is some confusion about what the various types of identifiers are in concept and practice (e.g. UUID, GUID, ORCID), including from a regulatory and ethical perspective. Therefore, the REWG supports the idea of working with the other Working Groups to develop a glossary or short document that explains what each of these terms means in concept and practice, and how they may impact on the work of the Global Alliance. Not only would this document better frame the discussions currently being had within projects, Working Groups,

Task Teams, and Member Organizations, but it would also aid the genomic and clinical data sharing community more broadly. Indeed, as identifiers (especially universal identifiers) are usually assigned



by some agency or consortium with governmental or international oversight, this may be a space where the Alliance wants to set leading standards for identifiers for genomic and health-related data. To do so, it is necessary to speak the same language and understand what these various terms mean.

Next Steps

Future ideas for possible development include:

- Exploration of the **categories for data access** (i.e. public/open, registered/user-identified, controlled/user signed agreement) and **unique identifiers** (e.g. UUIDs, GUIDs).
- Creating a new “**Accountability Policy Task Team**” in late 2014 that will develop an Accountability Policy, in line with the Policy template in Appendix 2 of the *Framework*. The Policy will explore the characteristics of accountability within the Global Alliance (e.g., identifying relevant stakeholders, framing violations of various types as properly being handled by organizations, and speaking about consequences of data misuse through community norms). Tiers of accountability as well as scenarios and appropriate responses to data misuses may be explored and addressed in the Policy.

Security Working Group **Sunday, October 19, 2014**

Working Group Progress

The major product of the group has been the Security Infrastructure document (to be released as v1.0, with an open link for comments). Other Task Teams are currently developing a need for SWG input.

Task and Project Team Work

The Working Group decided to establish three new Task Teams:

Incident Response Task Team (NEW)

Software Security Task Team (NEW)

Cloud Security (NEW)

Future Goals and Points of Agreement

The Global Alliance currently cannot accept responsibility for security oversight (corresponding to Control Objectives 4 and 5 in the Security Infrastructure). Two proposed models were proposed to address this:

- Financial Services Information Sharing and Analysis Center (FS-ISAC, <https://www.fsisac.com/>) – a member-owned, non-profit entity that serves as the global financial industry’s go-to resource for cyber and physical threat intelligence analysis and recommendations, and anonymous information sharing
- Security Incident Response Trust Framework for Federated Identity (Sir-T-Fi, <https://www.terena.org/mail-archives/sirtfi/pdfMAM3MjGx3A.pdf>) -- a draft document developed by Security for Collaborating Infrastructures group of the Interoperable Global Trust Federation (IGTF) that specifies requirements and metadata for managing security threats and incident response in the absence of overarching governance or uniform policy
- Other possibilities mentioned were the College of American Pathologists (CAP) lab-certification program, and the International Standards Organization (ISO) 27000 series.

Response to and mitigation of security breaches needs to be developed, including:

- Controlled message template response
- Begin defining use cases for realistic breaches



A technology solution for consent management also needs to be developed. Toward this objective, Eve Maler, Chair of the Katara User Managed Access (UMA) Work Group, presented ongoing work to define a profile of the OAuth 2.0 authorization standard (IETF RFC 6749) that enables individuals to authorize access to resources that they own.

Going forward, the SWG will launch three new Task Teams focusing on Software Security, Incident Response, and Cloud Security. In addition, interaction with the Task Teams of the other GA4GH Working Groups will increase. The interaction will be achieved in several ways:

- Embedding at least 2 SWG members within Task Teams requiring SWG support.
- Inviting Task team members to the bi-weekly SWG calls.
- Participating in other GA4GH Working Groups. Very Large projects to be tackled by SWG as whole if necessary.

Next Steps

In the near future, form the three new Task Teams, drawing from the SWG Interest Group as appropriate.

Cross-Cutting Projects Session Sunday, October 19, 2014

Progress

The four current Working Groups are catalyzing key collaborative projects that aim to share real-world data. The Project Teams move their work forward autonomously, with varying levels of coordination support and oversight from the Global Alliance, and drawing on expertise from the Working Groups as required. It is a considerable achievement that these projects have been initialized and are moving forward through uncharted territory as we continue to define relationships and support mechanisms between key projects and the Global Alliance.

Working Group Alignment

Matchmaker Exchange

- Uses a single API to interface with many similar databases
- Guidance is requested from **REWG** on proposed informed consent for different tiers (“Levels”) of matchmaking: no consent for Gene + high-level phenotype; consent required for sharing variants, genomic files and detailed phenotype.
- Guidance is requested from **SWG** on query authentication and verification steps, security requirements on the exchange.
- From **CWG**, need to engage expertise on phenotyping for matchmaking, will need improved specificity and scoring for phenotype queries as datasets enlarge. A potential quick win for cancer communities would be to do similar for somatic mutations, which are open source, less concern about re-identification.

Beacon Project

- Started as a simple test of the willingness of international sites to share data with a “yes” or “no” answer to querying whether the “beacon” has variant X.
- “Beacon of Beacons” is being developed (**DWG**); it aggregates beacons and answers the question does *any* Beacon have genomes with a given allele at this position.
- There are a growing number of beacons and synergy between **Matchmaker Exchange** and Beacon Project and potential for greater alignment should be explored.



- Input requested from **REWG** and **SWG** on proposed data access levels and the option to register and identify users (identity credentials/identity proofing); aspiration is a trusted repository.

BRCA Challenge

- Aims to translate the rapid expansion of sequencing capacity into useful knowledge and learn how to rapidly interpret variant data to generate clinical utility.
- Global outreach and engagement will be important.
- API and ELSI issues to be developed and addressed by **DWG** and **REWG**, respectively.

Next Steps

Working Group Coordinators will play a connecting role in ensuring that key project needs are addressed in a timely manner.