Problem: Genome Data Held in Silos, Unshared, not Standardized for Exchange

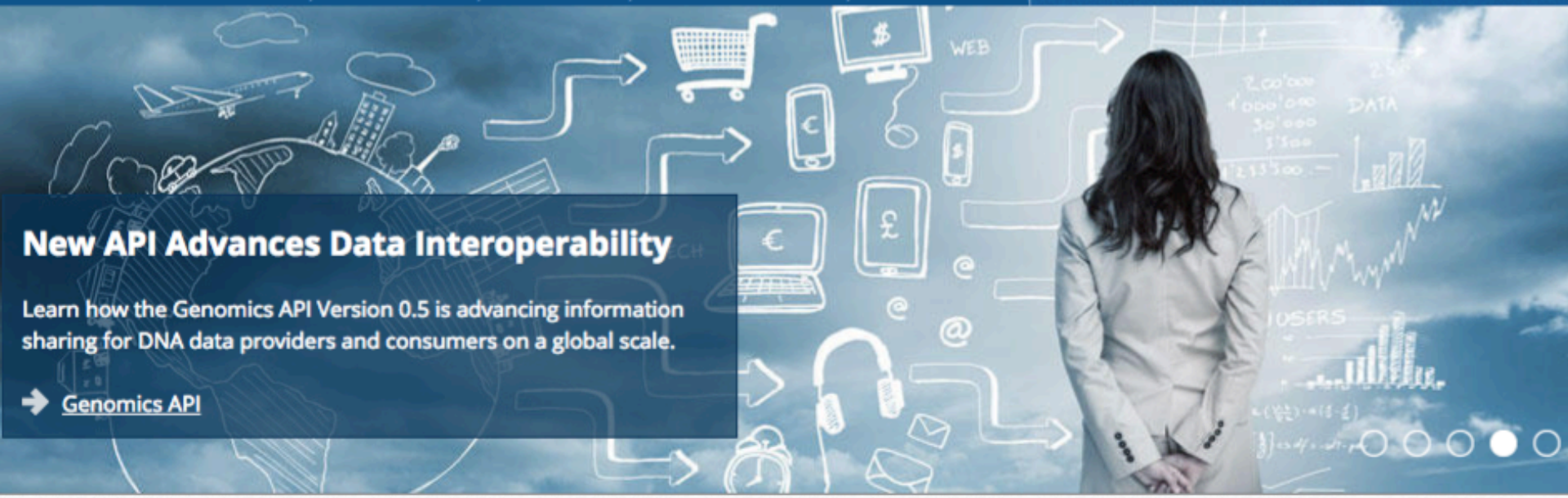
No one institute has enough on its own to make progress. Every researcher and clinician should be able to compare their genome data to others.

We need a public ledger for sharing

Global Alliance for Genomics & Health

Become a Member

ABOUT GLOBAL ALLIANCE | OUR WORK | MEMBERS | NEWS & EVENTS | CONTACT US



New API Advances Data Interoperability

Learn how the Genomics API Version 0.5 is advancing information sharing for DNA data providers and consumers on a global scale.

Genomics API

What is the Global Alliance?

The Global Alliance for Genomics and Health (Global Alliance) is an international coalition, dedicated to improving human health by maximizing the potential of genomic medicine through effective and responsible data sharing. The promise of genomic data to revolutionize biology and medicine depends critically on our ability to make comparisons

What is the Global Alliance doing?

Since its formation in 2013, the Global Alliance for Genomics and Health is leading the way to enable genomic and clinical data sharing. The Alliance's Working Groups are producing high-impact deliverables to ensure such responsible sharing is possible, such as developing a Framework for Data Sharing to guide governance and research and a

Who is involved?

The Global Alliance for Genomics and Health is an independent, non-governmental alliance, made up of hundreds of world-leading organizations and individuals from across the world. The Global Alliance is focused on bringing together a diverse set of key stakeholders across regions and sectors, including leaders in healthcare and research,

Enabling Responsible Sharing of Genomic and Clinical Data

- GA4GH Founded on June 5, 2013
  - More than 400 insitutional members
  - From more than 40 countries,
  - Approx. 1/3 are companies
- Mission: to enable rapid progress in biomedicine
- Strategy:
  - support major driver projects
  - create and maintain interoperability of technology platform standards
  - develop guidelines and harmonizing procedures for privacy and ethics in the international regulatory context
  - engage stakeholders across sectors to encourage the responsible and voluntary sharing of data and of methods

Cancer Global Data Sharing

Global Alliance for Genomics & Health

Cancer is driven by mutations in DNA. Precision treatment of cancer depends on knowledge of these mutations

Start with a Pilot:

-build a mechanism for recording cancer DNA mutations and clinical information from millions of cancer patient participants across the world

-Initially called the Actionable Cancer Genome Initiative

-Cancer is the right place to start. Once this is working, similar technology could be used to share DNA information for other diseases

Example proposed public record

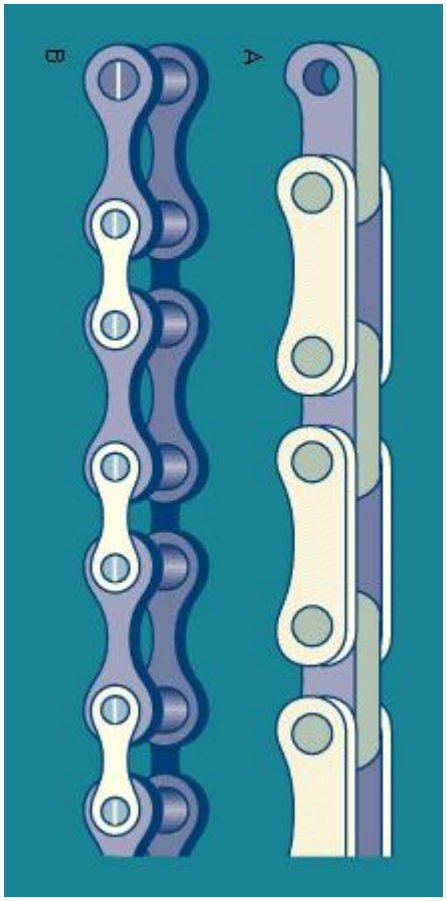
gene: BRAF
variant: V600E
Patient ID: 163a0083-26fa-4705-bcfb-d264c4cff796
Gender: Male
Ethnicity: White Caucasian
Age at Diagnosis: 57
Tumor Classification: non-small-cell lung carcinoma (MeSH D002289)
Tissue or organ of origin: Lung
Tumor morphology: Squamous (epidermoid)

Data Sharing Specifications

- Data open and available to all
- Ubiquitously accessible on the Internet
- Can scale to accept donations from 1000s of sources
- Not maintained by any central authority or tied to any single country, location or institution
- Not corruptible
- Protects participant privacy
- Stable design so that it may be used by many 3rd party application programs ("apps")

This is accomplished with a **Shared Public Ledger**

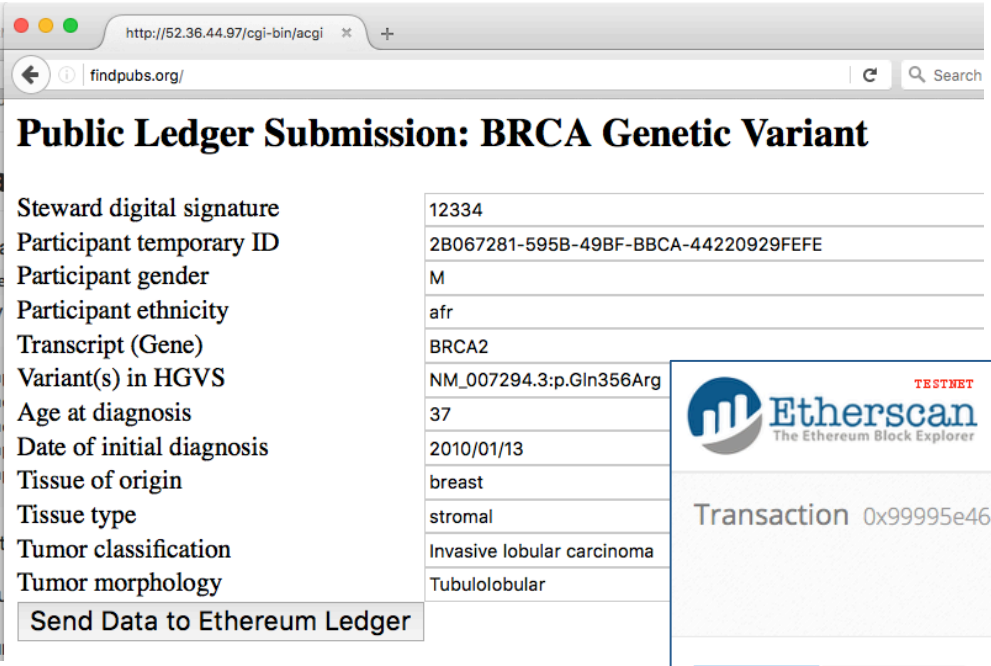
Ethereum: https://www.ethereum.org/foundation
Ripple: https://ripple.com/
Hyperledger: https://github.com/hyperledger/hyperledger
IBM Open BlockChain: www.**ibm**.com/**blockchain**/
MIT Enigma project
enigma.media.mit.edu
AirBnB (proposed): www.coindesk.com/**airbnb**-exec-use-**blockchain**/



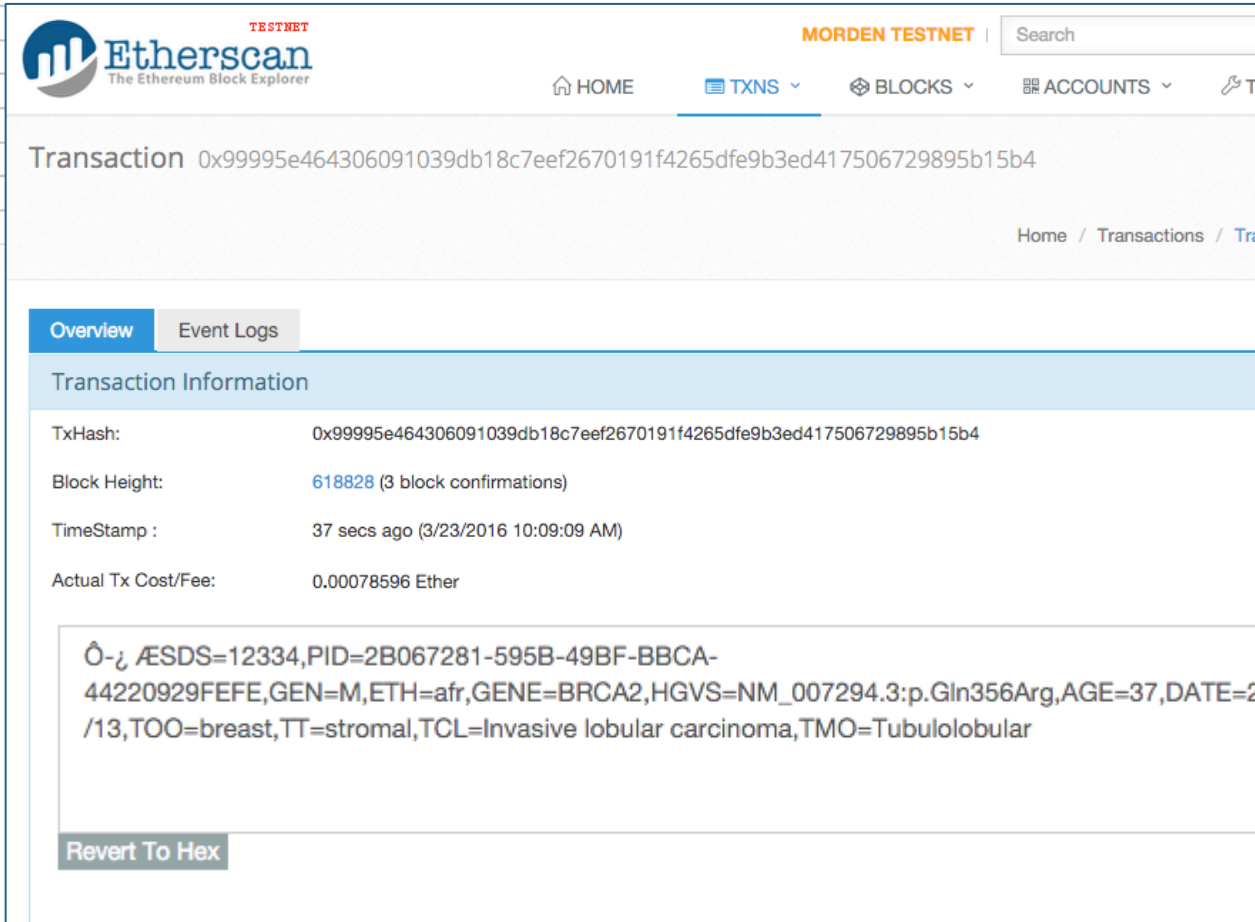
Simplest Shared Public Ledger

- Record of **transactions** over time
- Transaction is adding information to the database
  - Special case: New information marks previous information as out-of-date
- It is only possible to add more transactions, no transaction is ever erased or altered
- 1000s of copies of the ledger all over the world are kept in sync while additional transactions come in from multiple sources by miners
- A shared public ledger keeps track of data provenance, i.e. when and how data was entered and updated, so users have the reputation/ reliability information they need to filter out data they don't want

Ethereum Cancer KnowLedger Pilot

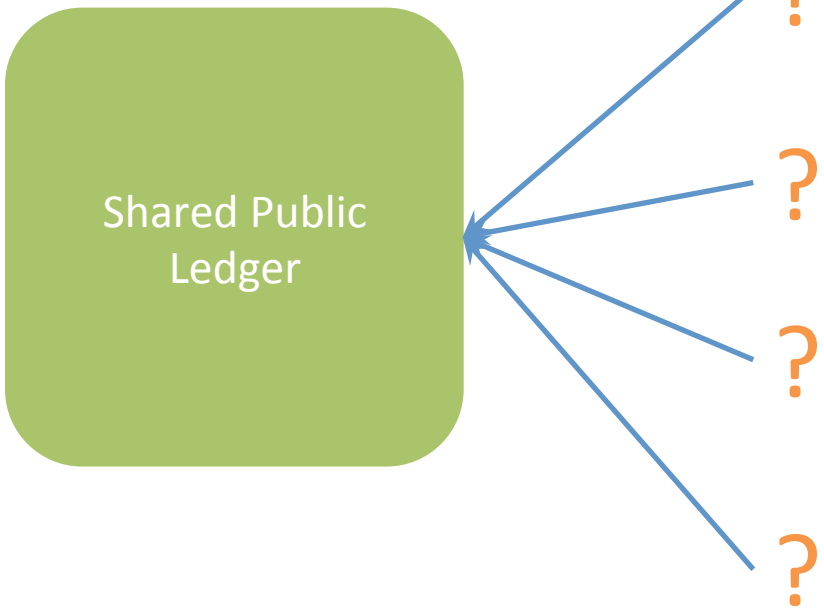


Website at findpubs.org



See:
https://github.com/maximilianh/acgi

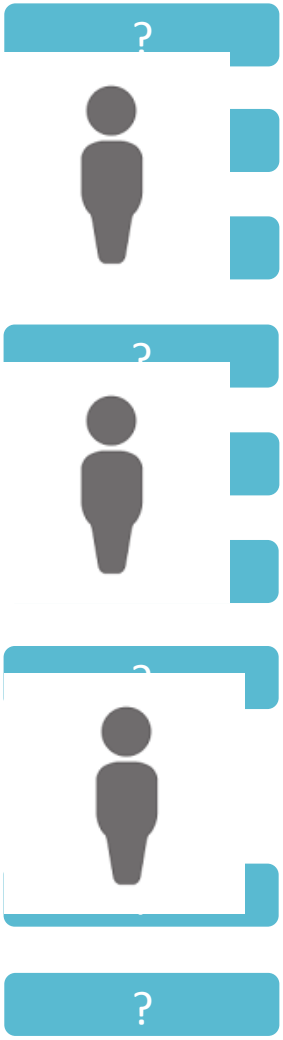
Who will use the shared public ledger?



Data Users are:

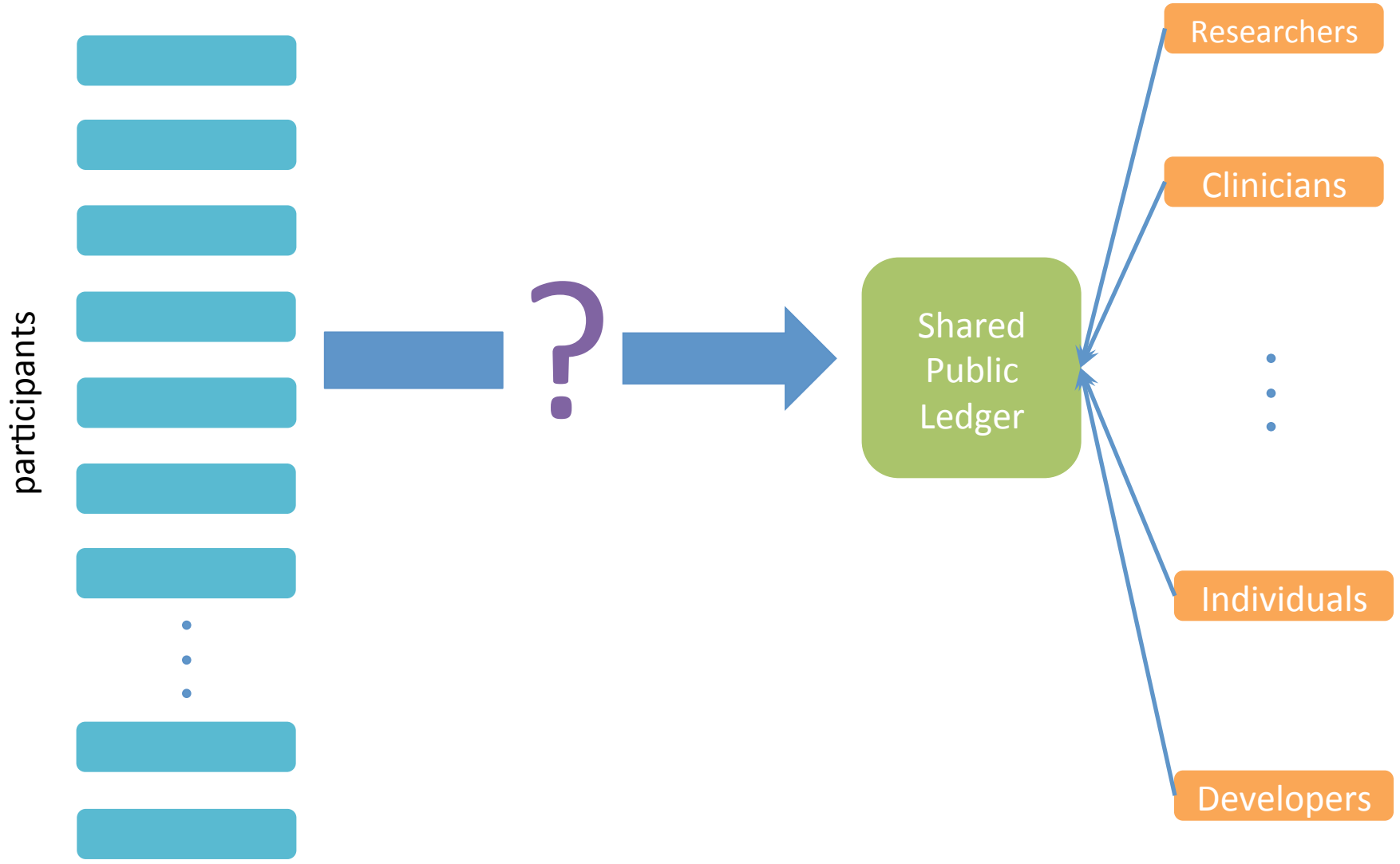
- Professional Researchers
- Citizen Scientists
- Clinicians
- Developers of molecular dx, drugs, decision analysis tools
- Payers
- Patients/participants
- Regulatory agencies and treatment guideline organizations

Where do the data come from?



- Ultimately, all data comes from individual **participants** who wish to share their genetic and clinical information for research or improvement of medicine

How do data enter the ledger?



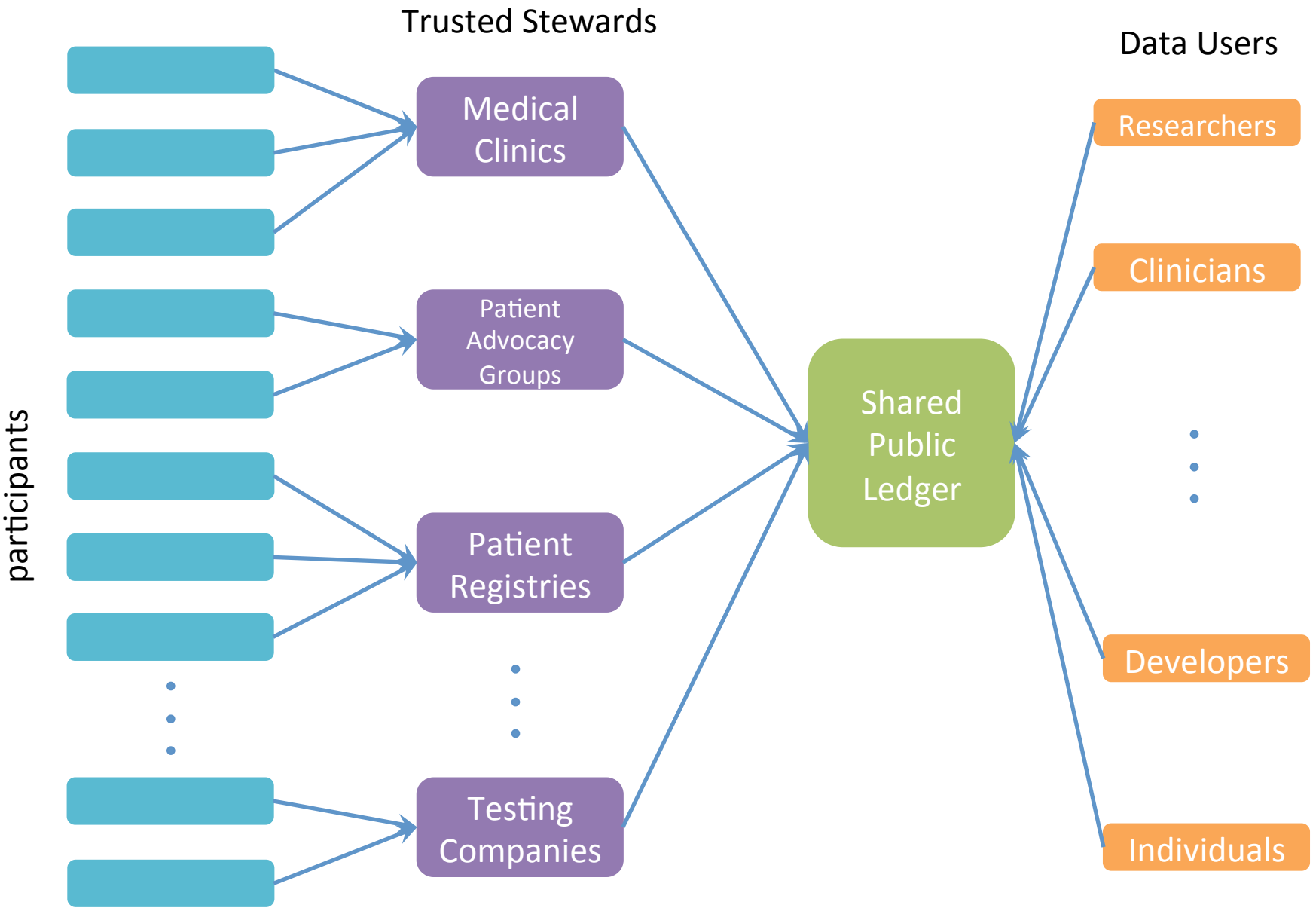
A participant works with a **Trusted Steward** to add their information to the ledger

Possible Trusted Stewards

- Medical research institutions (e.g. GENIE institutions)
- Hospitals and clinics
- Patient registry services and clinical trial recruitment organizations
- Patient agency advocate groups – possibly providing service to allow patients to maintain agency over data AND participate in clinical trials (e.g. Sage Trust and Genetic Alliance)
- Genetic testing companies

All trusted stewards use the same software (provided by GA4GH) to add information to the ledger

System is designed to support thousands of stewards globally


Trusted Stewards
Data Users
participants
Medical Clinics
Patient Advocacy Groups
Patient Registries
Testing Companies
Shared Public Ledger
Researchers
Clinicians
Developers
Individuals

What goes into the public ledger and what stays with the steward?

Public Ledger

- Participant's genetic variants in selected genes
- ~1 dozen broad, non-identifying clinical features
- Steward's identity and contact info
- Random numerical ID for participant

Steward

- Participant personal identifying information and staged consent
- Participant's extended clinical and genetic info
  - Possibly identifying
- Participant's instructions for sharing addl info with qualified researchers w/o recontact
- Participant's instructions for recontact

Example: participant -> public ledger

- Participant visits doctor at a medical clinic
- Doctor orders genetic test
- Doctor suggests data donation through steward (possibly her own institution)
- Test results come back

Test Results

multiplexed ICE COLD PCR™

Transgenomic®
Advancing Personalized Medicine

MX-ICP *EGFR* Test Report

(CONFIDENTIAL)

PHYSICIAN	SPECIMEN	PATIENT
Physician's Name:	Specimen Type:	Patient's Name:
Hospital/Institution:	Collection Date:	Date of Birth:
Mailing Address:	Receive Date:	Patient ID:
Fax Number:	Report Date:	Gender:
		Requisition #: HT000

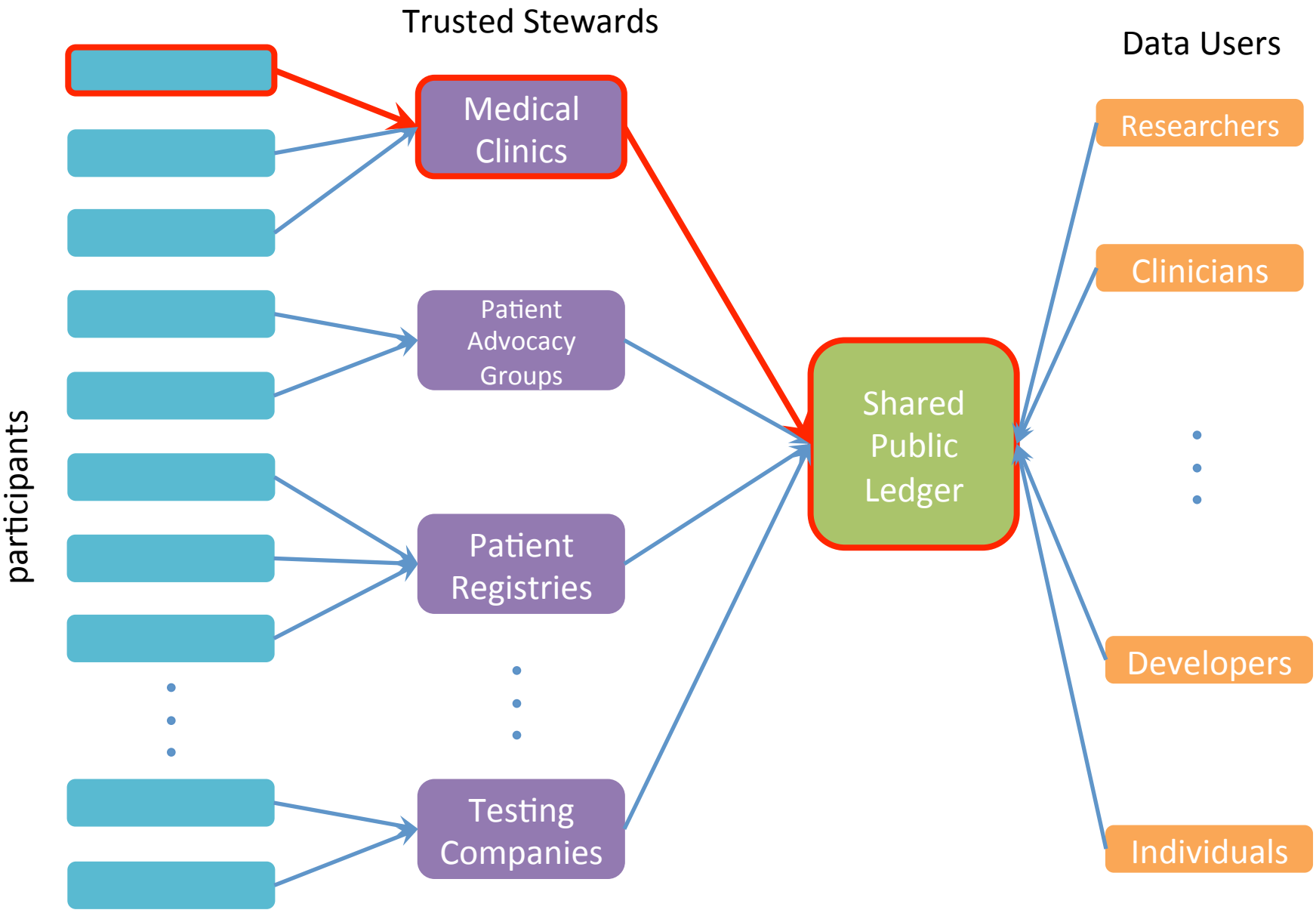
TEST RESULTS

Positive. The T790M mutation in *EGFR* exon 20 was detected.

INTERPRETATION

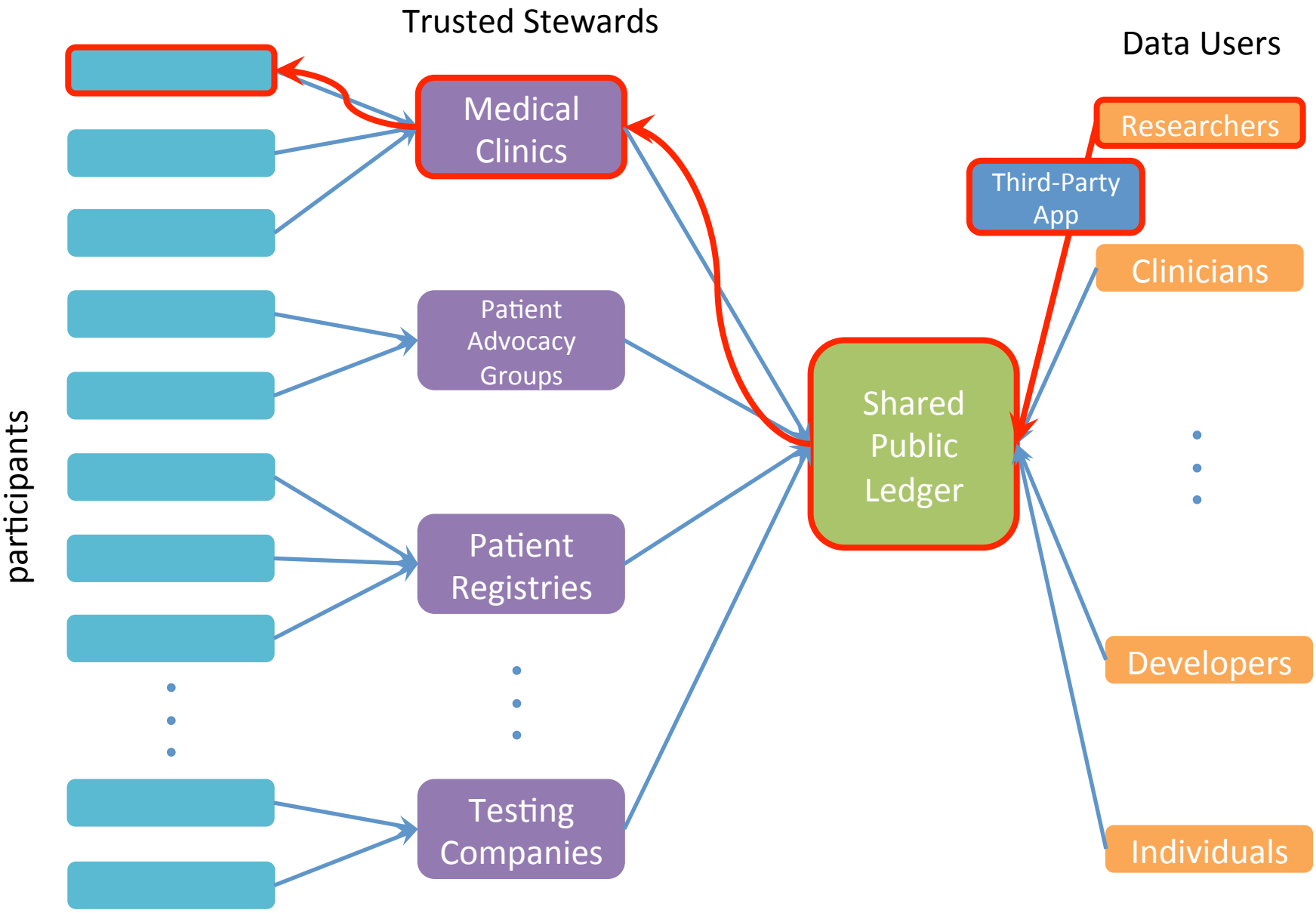
National Comprehensive Cancer Network (NCCN) guidelines for non-small cell lung carcinoma (NSCLC) endorse broader molecular profiling in individuals not responding to first-line therapy to determine cause for acquired resistance prior

- Participant visits steward
- Steward records from the participant:
  - personal information
  - test results
  - additional genetic and clinical data
  - consent to donate to database
  - additional sharing and recontact preferences
- Steward appends participant's publicly sharable genetic/clinical data to public ledger


Trusted Stewards
Data Users
participants
Medical Clinics
Patient Advocacy Groups
Patient Registries
Testing Companies
Shared Public Ledger
Researchers
Clinicians
Developers
Individuals

Recontact

- Data user discovers mutations of interest by using 3rd party app on the public ledger and wants more information about the participant that provided in the ledger
- Data user contacts participant's steward
- Steward has info about under what circumstances the participant will share additional data or agrees to be recontacted
- As appropriate, steward will
  - supply additional information to data user or
  - set up contact between user and participant


Trusted Stewards
Data Users
participants
Medical Clinics
Patient Advocacy Groups
Patient Registries
Testing Companies
Shared Public Ledger
Researchers
Third-Party App
Clinicians
Developers
Individuals

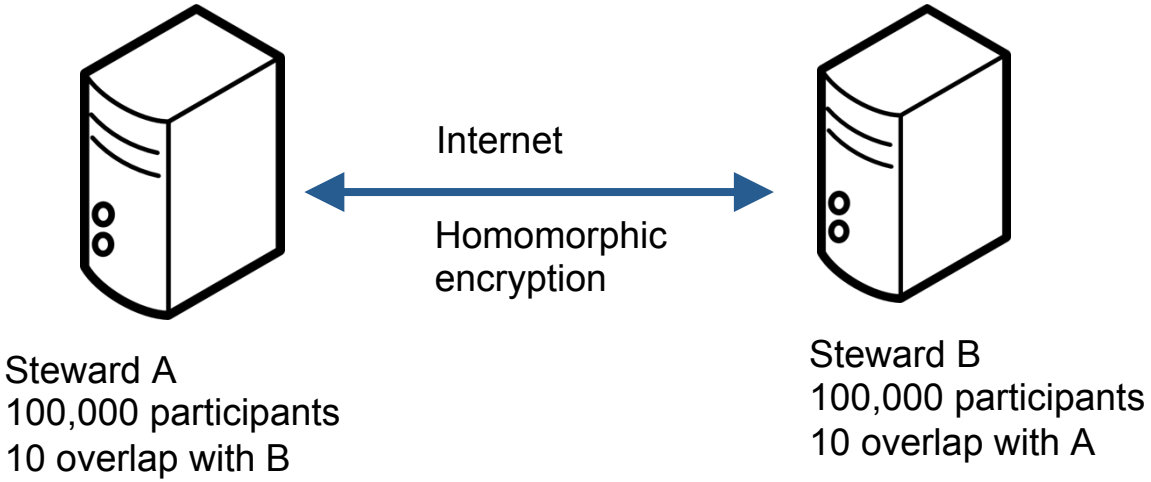
How do two stewards know if they have a participant in common?

- On the public ledger, a participant is identified with a random number
- Each steward securely stores personal identifiable information for their participants; only they can associate participant with a public random number
- Personal identifiable information is compared between the two stewards without revealing any information except which participants they have in common by a cryptographic trick (secure multiparty computation)

To make it comparable, personal information could be collected by the NIH NDAR GUID standard

	A	B	C	D	E	F	G	H	I	J
1	ID	FIRSTNAME	LASTNAME	MOB	DOB	YOB	COB	SEX	SUBJECTHAS	USEEXISTINGGUID
2	1	CLIFTON	BALLARD	7	6	1939	CENTRAL CIT	M	N	N
3	2	GREG	HERNDON	11	12	2009	GROJEC	M	N	N
4	3	MARIO	SOUSA	7	23	1946	PLYMOUTH	F	N	N
5	4	MARTIN	QUEEN	1	28	1974	BEAUSEJOUF	F	N	N
6	5	JESSE	ALEXIS	4	29	2004	RISING SUN	F	N	N
7	6	LEE	MOSER	11	29	1970	NAVODARI	M	N	N
8	7	RICKY	TRENT	1	6	1961	ALLISON	M	N	N
9	8	HECTOR	PECK	2	28	1971	OCEAN ISLE I	M	N	N
10	9	ROBERTO	SANFORD	8	25	1934	HOEDSPRUIT	F	N	N
11	10	ALVIN	EMERSON	7	8	1984	SPELLE	F	N	N
12	11	SETH	CHOWDHUR	8	3	1949	MARKGRONI	F	N	N

Demo: secure multiparty computation to compute private set intersection



Neither server sees the personal information on the other. Only the checksum identifying the participants in common is visible, nothing else

Runtime: 10 seconds over a transatlantic link, single CPU

Implementation and experiments by Max Haeussler

Who can be a steward?

- Any entity that:
  - Has a legitimate permanent contact
  - Follows the rules
  - Has enough participants to prevent “participant reidentification by steward” (small stewards can be anonymously pooled)

- All stewards will have Internet ratings; these can be available on a ledger (e.g. like AirBnB); users can filter out unreliable steward data

Who pays for all this?

- System can be designed and implemented for a few million dollars; long term maintenance is the only issue
- Governments or philanthropies (possibly associated with hospitals, patient advocacy groups, etc.) could supply general funding
- "Taxes" on genetic tests could provide revenue
- Stewards can be motivated to secure data donations either by altruism or commissions from data users
- 3rd party app developers can charge for use of their tools or sell advertising to support their efforts and to support the public ledger

Summary

- A completely decentralized, public database is possible, while still protecting privacy
- We need this because “trust is local”
- No single state government or private organization can/should own or control all the world’s genetic data
- Once launched, a shared public ledger grows and is maintained organically by the global community because it benefits them, much like the Internet itself