Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

GIF AMA

GA4GH Implementation Forum: Ask Me Anything

March 11, 2025

Featured standard

BEACON

Developed by the GA4GH Discovery Work Stream, the Beacon API is a simple data discovery tool that allows someone to search genomic and biomedical data worldwide, helping researchers find new datasets without compromising dataset privacy or ownership.

Learn more here.

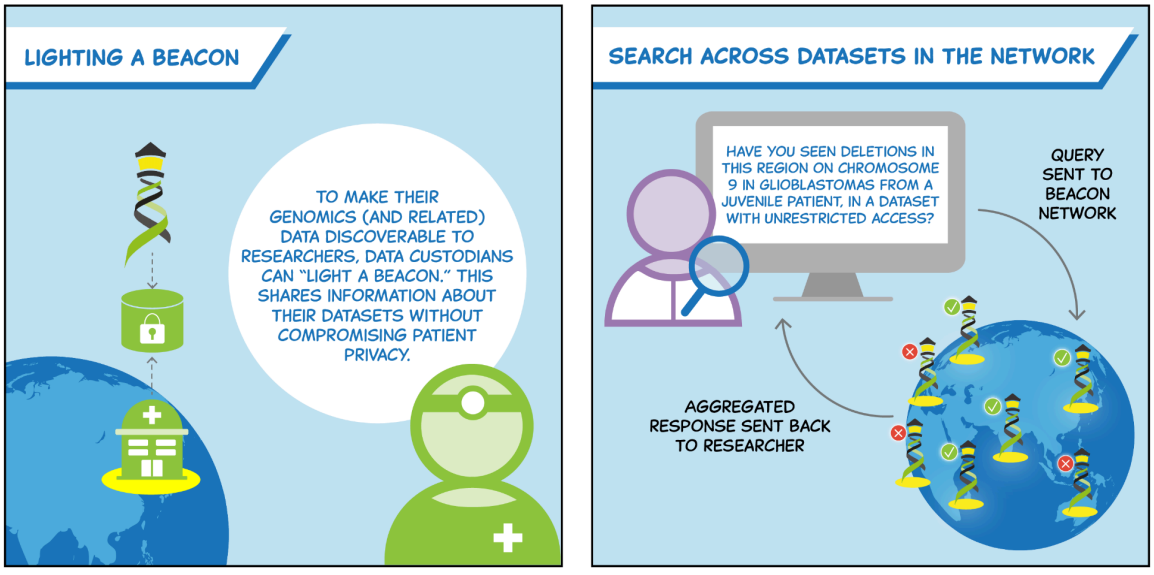


Image summary: *Beacon helps researchers discover useful information about data they can study while helping data custodians securely share insights from their data with the world.*

Meeting Overview

Title	GIF AMA: Beacon
Guests	Open to all, including • Beacon Product Team • Beacon Implementers • GA4GH Implementation Forum • Beacon-curious Implementers
Start time	Mar 11, 2025 11:00 AM EDT
End time	Mar 11, 2025 12:00 PM EDT
Location	Zoom
Description	This GIF AMA session will focus on the GA4GH Beacon standard, a framework for genomic data discovery that enables secure and controlled data sharing. Designed as an interactive discussion, this session will provide current and prospective implementers with direct access to the Beacon product team and experts who have deployed Beacon in diverse settings. Attendees will have the opportunity to ask technical and practical questions, learn from real-world implementations, and explore solutions to adoption challenges.

Agenda

	Item	Presenter	Time
1.0	Welcome	Beatrice	2’
1.1	Introduction to session and Beacon “experts” on the call	Mallory	3’
2.0	Current status of Beacon standard	Michael	5’
3.0	Q&A from attendees	Mallory (chair)	45’
4.0	Wrap-up and next steps	Mallory	4’
5.0	Close	Beatrice	1’

Mar 11, 2025 | [HOLD] Beacon GIF AMA

Attendees: Beatrice Amos, Mallory Freeberg, Michael Baudis, Senkei Umehara, Alaric snell-Pym, Bruce Hoff, Chris Boustred, Beatriz Ruz, Jaime Bachiller Hernandez, Eugenio Gaeta, Oriol Lopez-Doriga Sagales, Thomas Schaffter, Mireia Marin Ginestar, Justin Richardsson, Nicholas Owen, Dmitry Repchevvsky, Deepthi , Daniel Naro, Natassa Spiridou, David Alejandro, Sven Twardziok, Angela Page, Sasha Siegel, Ainhoa De La Torre, Fernanda Bittar, Jaime Delgado, Liina Nagirnaja, Peter Harrison, Taku, Ayton Meintjes, Jonathan Tedds, Radhika Mahajan, Liina Nagirnaja, Jordan Lerner-Ellis, Santo Diprima, Cameron MacPherson

Meeting Recording
Slides

- AMA slides
- Beacon slides

Info

- Beacon project website: genomebeacons.org
- A few FAQ: genomebeacons.org/FAQ/
- Beacon API documentation docs.genomebeacons.org/
- Beacon OpenAPI example at progenetix.org : progenetix.org/OpenAPI/
- Beacon Variant Scouts site with some request example genomebeacons.org/variant-query-types/generated/requestProfiles_g_variant/

Action items

- [ ] Complete feedback survey on GIF AMA event
- [ ] Make resources available linked from this document (slides, video recording, Q&As)
- [ ] Join GA4GH Slack channel #discovery_ws to ask additional Beacon-related questions
- [ ] Use sli.do for future AMAs ;)
- [ ] Add Yatish's answers to the Q&A

Introduction and Updates to Beacon

- History
  - Development began in 2014
  - beacon-network.org aggregator created by DNAStack
  - 2019 - ELIXIR Beacon Network
- A beacon network federates genome variant queries across databases that support Beacon API

- Beacon v2
    - represents a simple put powerful genomics API for federated data discovery and retrieval
    - the ability to do variation queries were refined
    - Relies heavily on filters
- Beacon development is closely aligned to other standards development (e.g. CNV)
- Beacon = protocol
- beacon = implementation of Beacon
- Beacon is not only for discovery, but also delivery
- Beacon Scouts
    - Groups of people working in different areas focused on future directions for Beacon (e.g. Beacon variation scout is focused on interplay with VRS)

Questions and Answers

Note: This is a community-driven resource, and your contributions help make it better! If you spot a mistake, see an opportunity to add a resource, or have more details to share, please jump in and make changes. We welcome and encourage your input!

Q: Given the flexibility of Beacon, how far can you stretch the data models and implementation before it stops *being* Beacon?

A: Beacon has been specifically developed to be flexible and will refrain from becoming super specific. If you are using Beacon for something other than variant queries and biosamples, then making it your own is encouraged. However, those using Beacon for that purpose should aim to align it as much as possible with the default model. Something else to consider is that Beacon is a discovery tool, so allowing too much querying and detail might cross the threshold into analysis.

Q: In the beacon2-pi-api github I don't see a container or folder associated with the frontend. Is there a standard frontend that is recommended to use for this implementation?

A: We have one implementation that you can grab from the reference implementation: https://github.com/EGA-archive/beacon2-ri-api. We are developing a new one for production that will be released, tentatively, in summer 2025.

Q: Do you have any idea or plan on how Beacons v2 will be integrated into EHDS?

A: Nothing formal yet. Building blocks have been developed for other components , and may be built for Beacon in the future, but it's not been formally pursued and is something that ELIXIR should follow up on.

Q: In Beacon v2.0, does the filtering mechanism work exactly the same way for all query types (e.g., genomic variants, individuals, biosamples) or are there specific considerations when querying genomic variants?

A: It is consistent across all query types. But you need the filtering terms correctly populated. The question you are having is a bit generic but a filtering term only applies to its scope. You need to make sure the filtering term you are using has the correct scope. I don't know if this answers your question, but please, don't hesitate to reach me out by mail or GitHub if you want more specific answers.

Q: I'm implementing a Beacon using the v2 Production Implementation and have many synthetic data in VCF and Phenopacket (JSON) format. Is there a way to convert Phenopacket data to a Beacon-friendly format used in Beacon v2 Production Implementation?

A: We are currently finalizing Phenopacket to BFF (Beacon-Friendly Format) mapping and will have it linked to the Beacon v2 Production Implementation. Please check in with the repo in a couple of weeks or so. There is also this tool.

Q: Does Convert-Pheno work with Beacon v2?

A: -Yes, it does work for the "individuals" entity. I am the author of the tool (M. Rueda). Please contact us (GitHub/e-mail) if you encounter any issues or have suggestions to improve the completeness of the mapping.

Q: I want to implement Beacon in a search API. Is it possible to discover/deliver other types of data that are not variants?

A: In principle, Beacon is in no way bound to variants. The Beacon standard is fundamentally a framework/design principle for an API. On top of that we provide some optional data models. You have to decide what your network wants to use for the data model. There's also two objectives you can meet with beacon 1) make things findable 2) use it as the first step in the sharing process (optional).

Q: Our scientists are looking for statistically significant connections between phenotype and genotype data, and are asking for ways to find the frequencies of matches rather than counts. I'm still working with them on what they need, but it might be something like if a request to /api/individuals specifies a filter AND some genotype criteria, we have a way to return how many individual match both (as the current count granularity returns), and ALSO how many just match the phenotype filter and not the genotype constraint, and vice versa, and how many people match neither - effectively a matrix of how those two boolean constraints are represented in the data.We are wondering if this should be a new granularity, more detailed than "count" but less than "records" - has anybody else considered this kind of use case?

What are your thoughts? Is there interest in collaborating on a standard representation for this kind of thing? Or is it really out of scope for beacon and we should just make our own private extension to the protocol?

A: The data model is open. In principle what you can do, is define your own menial metadata model for responders. You can also mirror your data to a server where essential parts can be queried. In Beacon v1, we had the handover concept which allows you to embed in your responses handover links that might point towards VCF files for example, it's flexible.
A: This sounds like the "aggregate" granularity that is being discussed in the corresponding scout

Q: I am looking at ways to federate our genomics discovery across a sparse university and hospital infrastructure where analysis and data are held in so many different places. Would having a private Beacon instance help solve this issue?

A: In Spain, we have a Beacon Network for sharing data amongst hospitals, which adapts to the hospital's way of working and needs. Even if it's still in an early stage, I believe it's one of the possible solutions to share data, as the potential of it is huge.
A: This is what the Beacon Network is for.It's a kind of virtual beacon that is backed by a set of beacons and aggregate responses. Regarding the security, Beacon doesn't impose specific security, but usually OpenID Connect is implied.

Q: There are many security considerations to take into account when implementing a Beacon. For groups with limited resources, how easy is it to configure all those security parameters for Beacon v2?

A: Using the concept of granularity i.e. a gatekeeper that tells you where queries are allowed to come in from, and blocking those queries that come from places that aren't approved.. There are a lot of factors to security - technical and data privacy issues. You can implement firewalls for the technical aspect. For individual privacy think about what is the data you're querying. Don't point your beacon to the individual data, point it to a safe form of data. You can also add noise to the data. Beacon doesn't specify how to do any of that. It's all down to you.

Q: Are there any aggregator or network implementations that people can go look at for inspiration

A: We can show demos, but some of them are private so we can't let you play with them. In terms of public beacons we are building a registry of beacon implementations. This implementation registry will be launched in the next few months. Beacon network is an ELIXIR project that you can check out here, it's completely open source. One challenge is

that most beacons are not data rich in their data delivery. We have an R file that you can use through Progenetix that's a beacon aggregator. While doing this we found problems with variant delivery. The Bento platform also offers a user interface enabling querying over a network of Beacon nodes.

Q: Is beacon being used by clinical laboratories for day to day work? Once a variant is identified by another group, what mechanisms are used to find out more about the patient and their phenotype to help support a variant classification?

A: We developed a private network for local hospitals for querying rare disease cases, which was simply based on genotype counts. The clinician could then connect with the respective hospital for in-detail phenotype information. You can also have a look at the Matchmaker Exchange program that uses Beacon Network in rare disease context as well (variant-level matching, classification etc) https://www.matchmakerexchange.org/. Other examples include:

- ERDERA - they have a Beacon spec that you can copy and adapt.
- DECIPHER - used in clinics in the NHS (UK) is already part of Matchmaker Exchange and is planning on implementing a Beacon over their database.
- RD-Connect, which is the central platform for variants in ERDERA has implemented Beacon V2 and is in Matchmaker exchange.

Q: How does my beacon connect to other beacons?

A: The aggregation of beacons' responses is done by Beacon Network. The query is universal, so it's done the same way as you would to a single beacon, but you will need to install the software. The BSC has implemented in collaboration with us here: https://github.com/elixir-europe/beacon-network-docker

Q: What's the best channel to ask beacon v2 implementor questions in Slack?

A: GA4GH Slack channel #discovery_ws