## Developing a User Interface for Sharing Federated Genomic and Phenotypic Data Using the Beacon v2 protocol

Dionne So, Ricky Nguyen, Jessica Do, Zoe Kamarinos





## What is the Purpose?

### Human genomics research problem

- and **phenotypic data**, especially when it comes to data on rare cancer subtypes.
- strategies.





• A major obstacle in human genomics research is the paucity of accessible, well-annotated genomic

• There is currently no way to access aggregated data on high-risk children cancer, thus the availability of data is imperative for personalised medicines, inclusive diagnostics, prognostic and therapeutic

### **Beacon Network Protocol**

### How does it help solve the problem?

- GA4GH Beacon specification global standard
- Facilitate querying of aggregated data
- v1 and v2







## The v1 Beacon Network

- Originally, the Beacon protocol (version 0 and 1) allowed researchers to obtain genomic variant data; the ELIXIR Beacon network is an example of an UI that extends this.
- Information about the presence/absence of a given, specific, genomic mutation in a set of data, from patients of a given disease or the population in general.



"Do you have a chromosome 13 position 32,936,"







Cancer

Institute

'C' at 3 at 732?"	chr13 chr1 32,93 32,5	chr13 32,936,732	
		С	
no")			

## The v2 Beacon Network

- Accepted as GA4GH standard in April 2022.
- lacksquare

Beacon V2 Reference Implementation: a Toolkit to enable federated sharing of genomic and phenotypic data Get access > Manuel Rueda 🔤, Roberto Ariosa, Mauricio Moldes, Jordi Rambla

Bioinformatics, btac568, https://doi.org/10.1093/bioinformatics/btac568





Beacon v2 provides greater structured data responses to Beacon queries, introducing new features such as: annotated genomic variations, biosamples, data about individuals and study cohorts.

## **Beacon Network v2**

### The Model

The Beacon v2 default data model provides a set of schemas for common data entities with a focus on biomedical genomics.

The entities are defined as follows:

- Collections (Datasets and Cohorts)
- **Genomic variations**  $\bullet$
- Individuals lacksquare
- Biosamples lacksquare
- Analyses & Runs ullet



Children's

Cancer Institute



Beacon v2 Models entities and their relationships

## The Beacon Instance

### What is a Beacon?



### SIMPLE

Beacons can be implemented on top of any underlying variant or variant annotation datastore



### FEDERATED

Beacons can be lit and maintained by individual organizations, and assembled into a distributed network



### AGGREGATIVE

Beacon results can be aggregated across an entire populations to support de-identification.





### SECURABLE

Beacon access can be restricted using institutional security protocols, and authorization schemes can be implemented to respect consent conditions.



A beacon is a simple yet powerful genomics API that allows for federated data discovery and retrieval.



Beacon architecture

M. M. A. Aziz, R. Ghasemi, M. Waliullah, and N. Mohammed, "Aftermath of bustamante attack on genomic beacon service", BMC Medical Genomics, vol. 10, no. S2, 2017.

## The Beacon Network

The Beacon network is made up of federated data sources (beacons) that can be collectively searched with a single query. The network also contains two components: The Aggregator and the Registry.

The Aggregator

Aggregates return as a single query

	ĺ.

Beacon v2 instances







**The Registry** Keeps track of all the

Beacon instances

## **Current Issue with the Beacon Network**

- Currently most existing Beacon interfaces aggregate data from the v1 Beacon network.
- The current version (v2) greatly increased the complexity and richness of the metadata.
- Necessitates a more sophisticated user interface









## Very nested metadata





Cancer

Institute

	Description	Туре	Properties	Example	Enum
rents	TBD	array	eventAgeRange, eventCases, eventControls, eventDataTypes, eventDate, eventDiseases, eventEthnicities, eventEthnicities, eventGenders, eventLocations, eventLocations, eventNum, eventPhenotypes, eventSize, eventTimeline	NA	NA

One term from the Cohort Schema of the Beacon v2 model Taken from the GA4GH Beacon v2 Documentation

- Intuitive user interface
- Major advancement in human genomic data sharing in Australia
- Open source
- Enable querying the full catalogue of data by all members of the Beacon network.





## **Our Solution**

Creating a centralised and efficient method of sharing genomic and clinical data via a user interface fronting the Beacon Network v2.











- To be able to refine and search queries
- A dashboard to display aggregated results
- To be able to sort results based on a criteria



The User Interface

### HGP Beacon Network

The project is designed to enhance capability for securely and responsibly sharing human genome research data nationally and internationally, ensuring maximum value can be derived from these valuable assets. It is investigating best practice technologies that have been globally developed for the purposes of human genome data sharing, and deploying Australian first technologies in the form of a 'services toolbox' for improving FAIRness of genomic data at the organisations that hold most human genomes collected for research in Australia.



Sign Up

Children's Cancer Institute



Get in touch











### Edit/Combine query

Start building a query by selecting a filter above

### SUMMARY

### RESULTS

Query Options	Id ↑↓	Data Type	Institution 1	Runs ↑↓	······ ^↓	↑↓	****** 1
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	login	4 min ago	Safari	Colombo	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Safari	Colombo	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf:	is.stimepagedemo.com
	generic_event	4 min ago	Safari	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf:	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Colombo	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com
	generic_event	4 min ago	Chrome	Unawatuna	Sri Lanka	1C5CHFA_enUA719UA720&sxsrf=	is.stimepagedemo.com





		Register Beacon	Documentation Logout
Cohorts	Runs	Genomic Variants	Individuals

## **User Interface** FEATURES



### Query bar

Displays the query terms as chips when the query option is selected.





Tabular representation of the filtered information that can be further sorted or

paginated.



### **Dropdown Menus**

Specialised query terms under each query option, which allow for a more specified query.



single or double term based

on an Id or Label.





### **Results table**



### Summary bar

A total count for the results returned by each query option selected.

### **Input Menus**

Prompts the user for a



### Chips

Visually represents the querying options selected.







### As a researcher:

### I want to find all paediatric patients (age < 21 yrs)

HGP Beacon Network			nalyses		Biosamples		)	
Edit/Combine o	query							
Sex	Dis	seases	Id	I	AgeRar	nge		
SUMMARY	RESULTS							
Query Options	Id	ţ↑	Diseases	- Ascending Descending	AgeRange	↑↓	Sex	
1000	generic_event		4 min ago	Order	me	U	nawatuna	
INDIVIDUALS	login		4 min ago		Safari	С	olombo	
	generic_event		4 min ago		Chrome	U	nawatuna	
	generic_event		4 min ago		Chrome	U	nawatuna	
	generic_event		4 min ago		Safari	С	olombo	
	generic_event		4 min ago		Chrome	U	nawatuna	
	generic_event		4 min ago		Chrome	U	nawatuna	
	generic_event		4 min ago		Chrome	U	nawatuna	
	generic_event		4 min ago		Safari	U	nawatuna	
	generic_event		4 min ago		Chrome	U	nawatuna	
	generic_event		4 min ago		Chrome	U	nawatuna	
	generic_event		4 min ago		Chrome	U	nawatuna	
	generic_event		4 min ago		Chrome	С	olombo	
	generic_event		4 min ago		Chrome	U	nawatuna	



				F	Registe	er Beacon	Documentation	Logout		
5	Runs Genomic				emic Va	ariants	Individ	uals		
						🗌 Kary	votopicSex	(+) (+)		
						🗹 Dise	ases	•	AgeOfOnset	
						Ethn	licity	$\oplus$	DiseaseCode	
î↓	*****	ţ1		t↓		Expo	osures	$\oplus$	FamilyHistory	
	Sri Lanka	a	1C5CHFA_enUA7	19UA720&sxsrf=	is.stir	Geo	graphicOrigin	Ð	Notes	
	Sri Lanka	a	1C5CHFA_enUA7	19UA720&sxsrf=	is.stir	🔽 Id		÷	Severity	
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf=	is.stir	🗌 Info		÷	Stage	
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf=	is.stir	🗌 Inter	rventionsOrProcedure	es 🕂		
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf=	is.stir	🗌 Mea	sures	$(\pm)$	Clear all filters	
	Sri Lanka	a	1C5CHFA_enUA7	19UA720&sxsrf=	is.stir	Pedi	grees	(+)		
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf:	is.stir			æ	From 0 - to 20 -	
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf=	is.stir		lotypicreatures	Ð		
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf=	is.stin	Trea	tments	$(\pm)$		
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf=	is.stir	Clear all	filters	8		
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf=	is.stim	epagedemo.cc	om			
	Sri Lanka	3	1C5CHFA_enUA7	19UA720&sxsrf=	is.stim	epagedemo.co	om			



×

### As a researcher:

I want to know how many patients have at least two tumours that have undergone WGS

										Register Beacon	Docun	nentation   Logout		
HGP Beacon Network		Analyses	Biosamples		Cohorts	Ru	<b>)</b> Ins	Ge	nomic Variants		Individuals			
Edit/Combine query		ery		🗹 Biosample Status	Ð	Biosample Statu	S		Biosampled	•		3 LibraryStrategy		<ul> <li>KaryotopicSex</li> <li>Sex</li> </ul>
	Id	WGS	Status	Collection Date	Ð	abnormal sample			IndividualId	(+		WGS		Diseases
SUMMA	ARY	RESULTS	Ascending Descending	Collection Moment	<b>(</b> +)		oply		Info	Œ		Ар	ply	Ethnicity
Quer	ry Options	IndividualId	Order WGS	<ul> <li>Diagnostic Markers</li> <li>Histological Diagnosis</li> </ul>	( <del>)</del>	our	ion ↑↓		LibraryLayout LibrarySection	( <del>-</del>				Exposures     GeographicOrigin
	1000 Cohorts	generic_event	4 min ago	🗌 Id	÷	na Sri Lank	a		LibrarySource	Œ				✓ Id
	1000 INDIVIDUALS	login generic_event	4 min ago 4 min ago	<ul> <li>Measurements</li> <li>Obtention Procedure</li> </ul>	( <del>+</del> )	Sri Lank	a		LibraryStrategy Platform	(+				
		generic_event generic_event	4 min ago 4 min ago	Pathological Stage	Ð	na Sri Lank Sri Lank	а		PlatformModel	(				
		generic_event	4 min ago 4 min ago	Phenotypic Features     Sample Origin Type	<ul> <li>(+)</li> <li>(+)</li></ul>	na Sri Lank	a		runDate	(				Pedigrees
		generic_event generic_event	4 min ago 4 min ago	V Tumour Grade	•	na Sri Lank	а	10	5CHFA_enUA719UA720&sxsi	rf= is stimepagedemo.com rf= is.stimepagedemo.com				<ul> <li>PhenotypicFeatures</li> <li>Treatments</li> </ul>
		generic_event generic_event	4 min ago 4 min ago	✓ Tumour Progression	Ð	na Sri Lank na Sri Lank	а	1C	5CHFA_enUA719UA720&sxsi 5CHFA_enUA719UA720&sxsi	f= is.stimepagedemo.com f= is.stimepagedemo.com				Clear all filters
► Children' Cancer	's +	generic_event generic_event	4 min ago 4 min ago	Clear all filters Chrome	Colombo	o Sri Lank	a	1C 1C	5CHFA_enUA719UA720&sxsi 5CHFA_enUA719UA720&sxsi	f= is.stimepagedemo.com f= is.stimepagedemo.com				
Institute														





User selects filters in dropdown menus that will make up their query



User decides on which schema they want to query by







## **USER PROCESS**



Request is sent to API and information relevant to their query is returned

User provides any extra query information to input menus



- Underlying motivation for stems from the limited cohort of available cancer data that each institute has alone.
- Bolster the range of patient data through the updated UI that supports the new v2 Beacon network.
- Provide quick and easy access to query a network of extensive cancer data.
- Facilitate comprehensive cancer research and enable confidence to make statistical inferences when profiling tumors.



## End Vision



Beacon v2 Networks

Kamile Taouk (6), Angela Lin (6), Marie Wong-Erasmus (6), Mark Cowley (6), Tiffany Boughtwood (3), Jeff Christiansen (1), Joe Copty (7), Shyamsundar Ravishankar (7), Kylie Davies (2), Matthew Downton (9), Kelsey Druken (9), Ben Evans (9), Clara Gaff (8), Andrew Gilbert (5), Christina Hall (1), Matthew Hobbs (7), Oliver Hofmann (11), Jessica Holliday (1), Warren Kaplan (9), Ross Koufariotis (10), Sarah Kummerfeld (7), Conrad Leonard (10), Andrew Lonie (1), Heath Marks (2), Siobhann McCafferty (4), David Monro (9), Andrew Patterson (11), John Pearson (10), Bernard Pope (1), Shyamsundar Ravishankar (7), Florian Reisinger (11), Andrew Robinson (9), Victor San Kho Lin (11), John Scullen (2), Marion Shadbolt (1), Andrew Treloar (4), Jingbo Wang (9), Scott Wood (10), Mustafa Syed (6).

1. Australian BioCommons 2. Australian Access Federation (AAF) 3. Australian Genomics 4. Australian Research Data Commons (ARDC) 5. Bioplatforms Australia 6. Children's Cancer Institute (CCI) and the Zero Childhood Cancer Program (ZERO) 7. Garvan Institute of Medical Research 8. Melbourne Genomics Health Alliance 9. National Computational Infrastructure (NCI) 10. QIMR Berghofer Medical Research Institute (QIMRB) 11. The University of Melbourne Centre for Cancer Research (UMCCR).











## Special Thanks to









# Thanks!

Any questions?

For further inquiries email: <u>ktaouk@ccia.org.au</u>