



The ELIXIR Beacon in 2018: A driver project of GA4GH

Michael Baudis, University of Zurich | **SIB**

ELIXIR All Hands 2018, 4-7 June 2018, Berlin, Germany



Global Alliance
for Genomics & Health

www.elixir-europe.org

Beacon Project

An open web service that tests the willingness of international sites to share genetic data.



Beacon Network

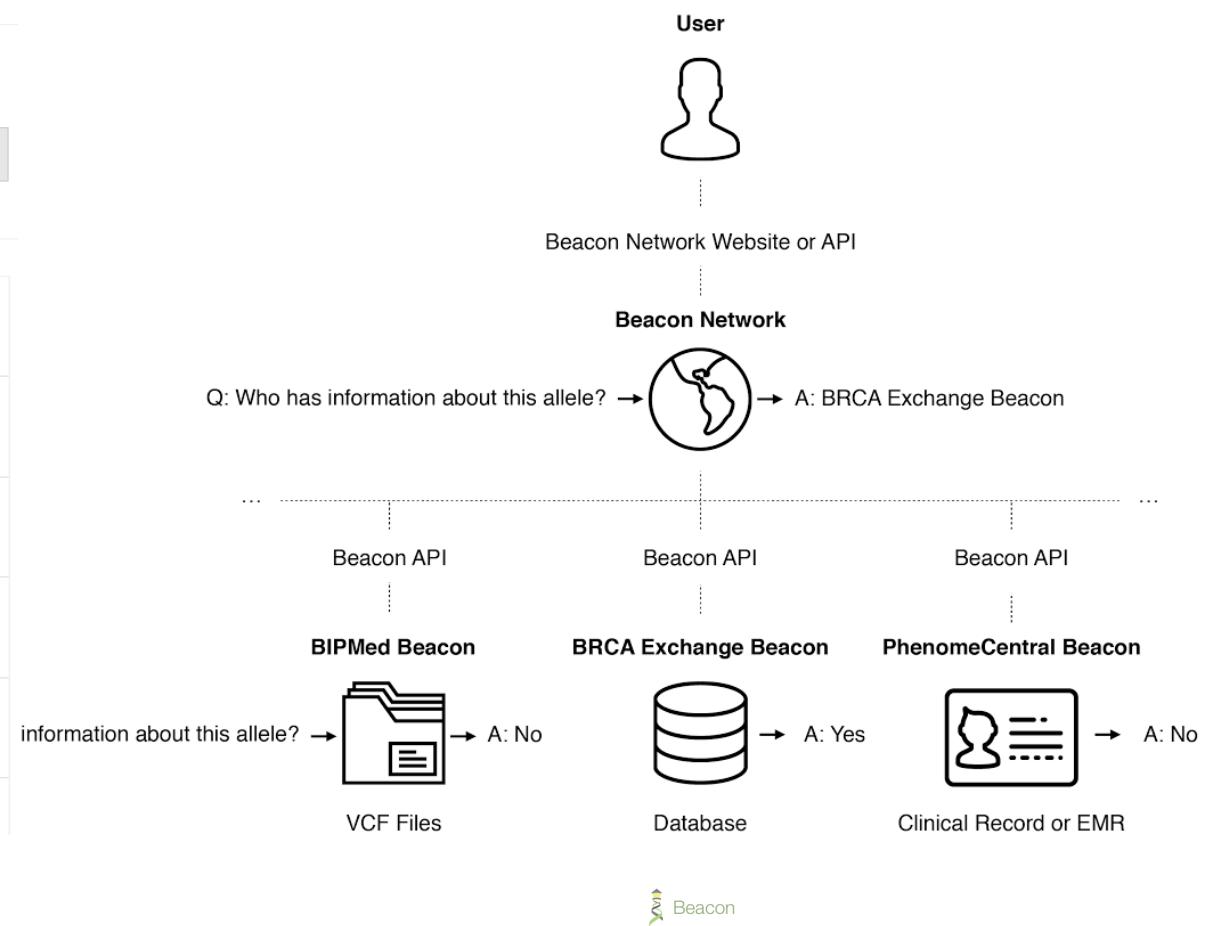
Search all beacons for allele

GRCh37 ▾ 10:118969015 C / CT | Search

Response All None
 Found 16
 Not Found 27
 Not Applicable 22

Organization All None
 AMPLab, UC Berkeley
 BGI
 BioReference Laborato...
 Brazilian Initiative on ...
 BRCA Exchange
 Broad Institute
 Centre for Genomic R...
 Centro Nacional de A...
 Curoverse
 EMBL European Biol...
 Global Alliance for G...
 Google
 Institute for Systems ...
 Instituto Nacional de ...

Resource	Hosted by	Status
BioReference	Hosted by BioReference Laboratories	Found
Catalogue of Somatic Mutations in Cancer	Hosted by Wellcome Trust Sanger Institute	Found
Cell Lines	Hosted by Wellcome Trust Sanger Institute	Found
Conglomerate	Hosted by Global Alliance for Genomics and Health	Found
COSMIC	Hosted by Wellcome Trust Sanger Institute	Found
dbGaP: Combined GRU Catalog and NHLBI Exome Seq...		Found



Date	Tag	Title
2018-01-24	v0.4.0	Beacon
2016-05-31	v0.3.0	Beacon

ELIXIR Beacon Project 2018



- WP1 - Development of Beacon API specification with the GA4GH
- WP2 - Supporting new ELIXIR Beacon development and deployment across Europe
- WP3 - Implementation of an ELIXIR Beacon Network & Registry
- WP4 - Security of ELIXIR Beacon and Beacon Network
- WP5 - Developing Indicators to establish Registry and Beacon as an Emerging ELIXIR Service
- WP6 - Strategic partnerships with national data cohorts and biobanks
- WP7 - Project Management and Communication

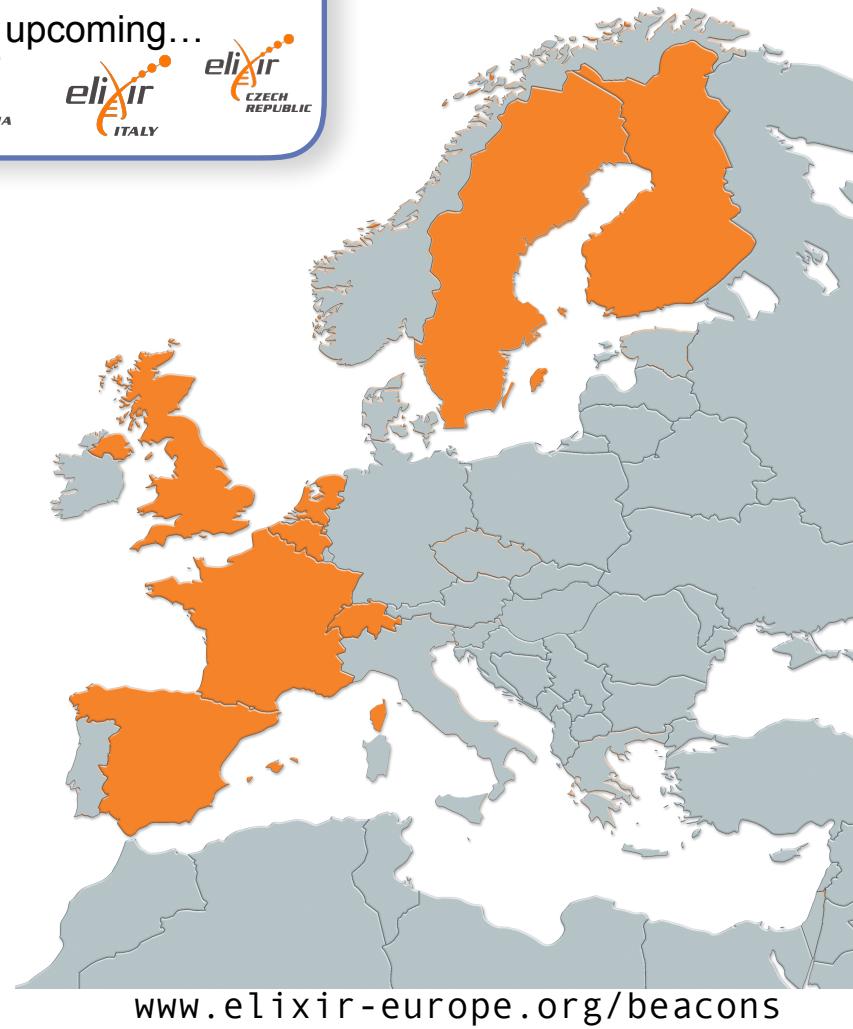
Node	Name of PI	Role	PMs	Other Contrib	Work Packages						
					1	2	3	4	5	6	7
FI	Ilkka Lappalainen	co-lead	5		X	X	X	X	X	X	X
ES	Jordi Rambla	lead	5		X	X	X				X
CH	Michael Baudis, (Heinz Stockinger)	co-lead	5		X					X	X
EBI	Dylan Spalding	co-lead	5		X	X	X	X			
SE	Niclas Jareborg	member	1			X					
BE	Yves Moreau	member	1				X				
FR	Macha Nikolski	member	1					X			
NL	Morris Swertz	member	1					X			
IT	David Horner	member	-								
FI	ELIXIR Compute Platform - Tommi Nyrönen				2						
TBD ⁶	ELIXIR Training Platform				1			X			
HUB	Serena Scollen, Susheel Varma	member coordinator	0.5		X				X	X	X

Node work package participation, as in 2018 project plan



2018 Progress

- Release of v.0.4 specification
 - basis for **GA4GH Beacon v.1.0**
 - undergoing GA4GH certification (first for procedure)
- Building flexible **Beacon Network(s)**
- Beacon paper under review in *Nature Biotechnology*
- Post-2018 planning, documented in **5-year plan**
- Widening Node participation & additional use cases:
 - ELIXIR Norway joined Implementation Study, to light Beacon(s) against **Marine data**
 - Widen this out to more use cases? Plants, other...



Security and GDPR



- Requirements analysis document, outlining **re-identification mitigation**
 - Individual Beacon and network of Beacons
- GA4GH Data **Security questionnaire** completed
- GDPR workshop in Brussels:
 - Raw data subject to recitals 33 (scientific research) and 34 (genetic data)
 - GDPR not applicable to anonymous data (recital 26)
- Genetic data kept anonymous by re-identification mitigation techniques
- Access restrictions (**tiered Beacon access**) can be adapted for each node
 - Accounts for local legal interpretations



This forward looking Beacon interface implements additional, planned features beyond the current GA4GH specifications.

Info

Query

Dataset

tcga

Reference name*

9

Genome Assembly*

GRCh38 / hg38

Start min Position*

19,500,000

Start max Position

21,975,098

End min Position

21,967,753

End max Position

24,500,000

Alt. Base(s)*

DEL

Bio-ontology

icdot:c50.9: (4065)

Beacon Implementations

- implementing existing resources with Beacon protocol
- e.g. TCGA cancer variants (structural and SNV)

SNV Example

DGV Example

CNV Example

Beacon Response

- quantitative (counts for variants, callsets and samples)
- Handover to authentication system for data retrieval
- **no exposure** of data beyond standard Beacon response and additional pointer to matched data

Prototyping Query Extensions

- testing e.g. bio-metadata queries using ontology terms

Dataset	Assembly	Chro	Start Range	End Range	Pos	Ref Alt	Bio Query
tcga	hg38	9	19,500,000 21,975,098	21,967,753 24,500,000		DEL	icdot:c50.9

Variants	Calls	Samples	f alleles	Response Context
54				JSON
54				UCSC
54			0.0243	Handover



This Beacon implementation is developed by the Computational Oncogenomics Group at the University of Zurich, with support from the SIB Technology group and ELIXIR.





Beacon query => Handover Handle => Authentication => Data Retrieval

Beacon+ example implementation using public somatic variation data

Beacon+

This is an implementation of a Beacon "handover" concept, in which a Beacon query response additionally delivers representation of the query results (i.e. callsets, biosamples, metadata ...), which can then be accessed after authentication ("yes"|"no") or quantitative ("n matches") Beacon response from a data delivery mechanism.

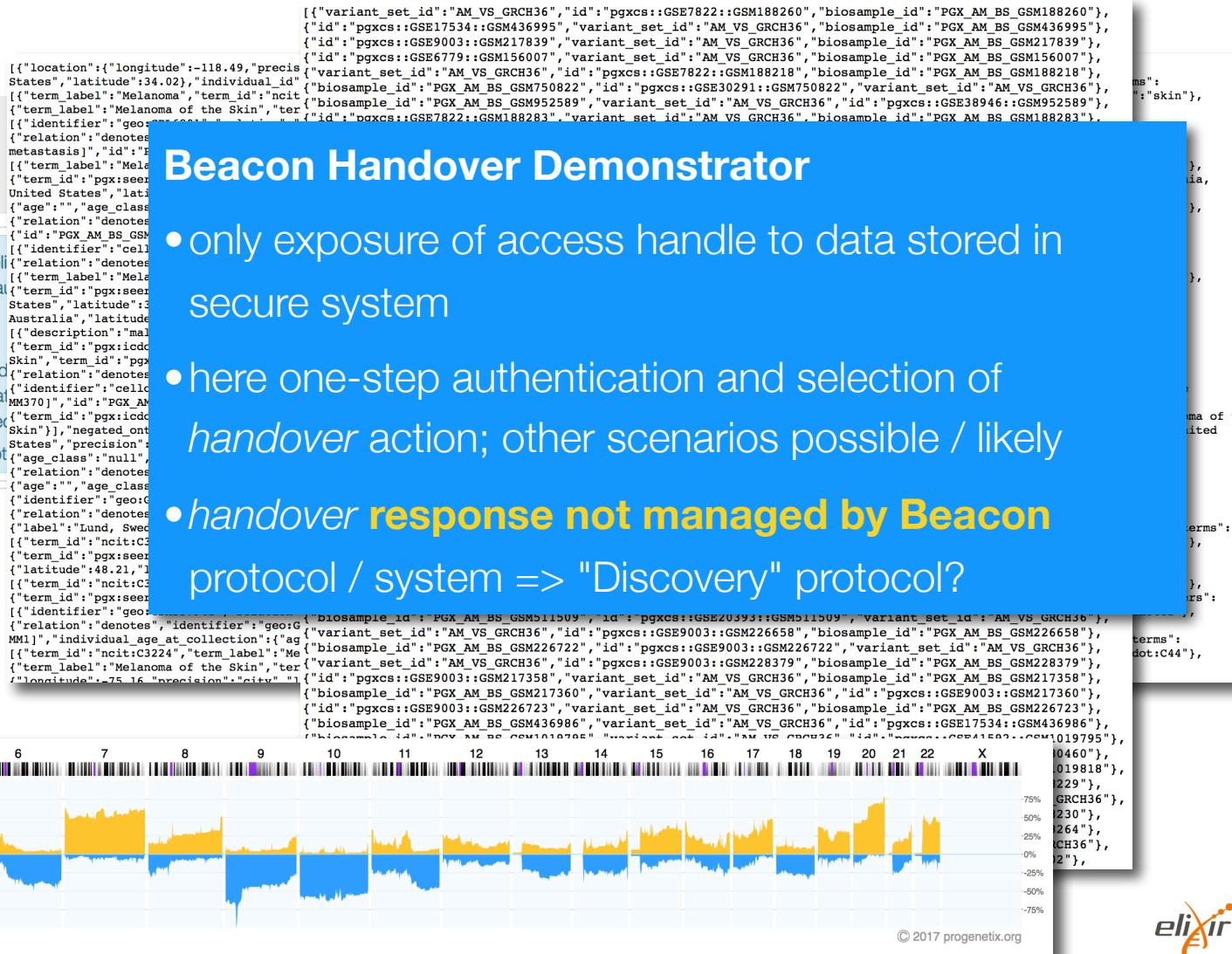
The current implementation exemplifies some possible scenarios:

- providing a histogram of regional gain/loss frequencies (DUP, DEL) for samples with structural variation of interest
- returning data of the associated callsets which matched the Beacon query (this is for feature demonstration)
- returning the metadata (diagnoses etc.) of the biosamples from which the matching callsets were derived

This demonstrator does not implement authentication procedures yet; login & password fields can be left empty.

Handover Action	<input checked="" type="checkbox"/> Plot DUP/DEL histogram <input type="checkbox"/> Export Callset Data <input type="checkbox"/> Export Biosample Data
Login	Login Password <input type="password"/> Process Data
Password	
arrayMap	
progenetix	
Global Alliance	for Genomics & Health

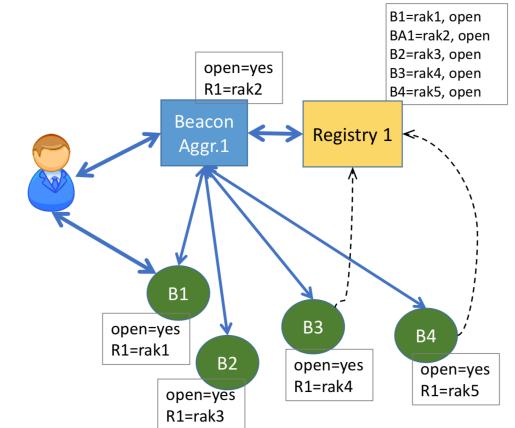
171 samples





Towards a shining future...

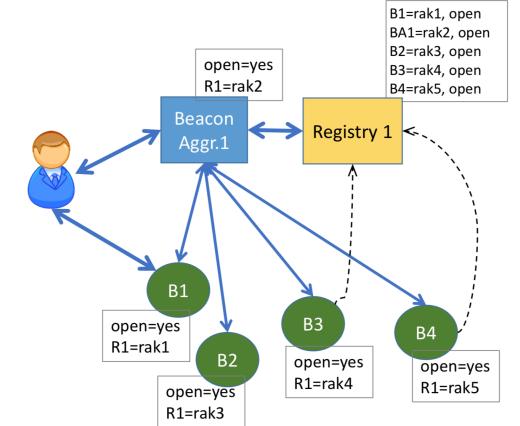
- extensive/complete representation of genome **variant types** in query
 - close coordination with **GA4GH::GKS** and **GA4GH::Discovery** work streams
- providing a tested model for layered **registered access**
 - ELIXIR AAI
- implementing **Beacon network(s)** throughout ELIXIR
 - open protocols for extension and external implementations
- extending Beacon query protocol (**metadata...**)
 - keeping "aggregate response" model
- Beacon queries as entry points for data delivery, using "**handoff**" scenarios





Towards a shining future... "The" Genome Query API

- extensive/complete representation of genome **variant types** in query
 - close coordination with **GA4GH::GKS** and **GA4GH::Discovery** work streams
- providing a tested model for layered **registered access**
 - ELIXIR AAI
- implementing **Beacon network(s)** throughout ELIXIR
 - open protocols for extension and external implementations
- extending Beacon query protocol (**metadata...**)
 - keeping "aggregate response" model
- Beacon queries as entry points for data delivery, using "**handoff**" scenarios



Reminder: Beacon WorkShop Thursday a.m.

Thank You!



... and many more



Global Alliance
for Genomics & Health