Presentation

- CoLaus cohort

- Collaboration with the CRG

- Docker deployment

- Next steps

Colaus cohort

- Currently we dont have access to genetic information

- For this reason we created fake genetic data to be beacon compliant

hypage	adiponectin	insulin	ht	diagtx1sbh	csdth_f	cafuse	waist	chol	jobtyp	wt	dginvtx1	age	cmatccd1_1	csdth_m	gene	variant
	6583	9	165		DROWING ON THE BOAT	1B3	106	7		87	CONTRACEPTIVE PIL	73		STROKE	FGB	C148T
	9238	19	182	PANCREAS CANCER	LUNG CANCER	1B3	75	5		76		58		UNKNOWN CARDIAC PROBLEM		
	2940	18	178	TESTICLE CANCER AND	COLON CANCER	1B3	78	5	MS	61		45		DON'T KNOW		
59	6180	4	186			1B3	91	6	ME	80		44	N02BA			
60	11442	8	167			1B3	100	7		89	ALCOHOL WEANING	55	A12AA	CAR ACCIDENT	FGB	C148T
	4801	20	159		HEART ATTACK	1B3	87	7	FM	83	HEADACHE	39		UNKNOWN INFECTION - OLD AGE		
	3617	11	156		STROKE	1B3	93	4	QW	66	ASTHMA	47				
	8606	6	175		BICYCLE ACCIDENT	1B3	70	7	EU	89		46				
	13856	20	178	BILATERAL KERATOCO	PANCREAS CANCER	4B6	80	4	MS	66	ALLERGY	48			FGB;SERPINF1	C148T;N47T;N60T
	12898	22	181			1B3	100	4	FM	61		56				
	3778	16	160		UNKNOWN	4B6	96	4	MS	72		49	N02BA	LUNG CANCER		
38	5189	6	175			1B3	89	5	QW	73		45				
14	5363	16	178		UNKNOWN ACCIDENT	1B3	97	5	PR	53		58		HEART ATTACK		
	4338	12	193		SUICIDE	1B3	70	6		74		37			FGB	C148T
40	6654	6	179		OLD AGE	1B3	122	5	IW	92		58	N02BA			
	6066	15	171		MENIERE'S DISEASE	>6	87	6	EU	68	AFFECTION CAVITY F	39				
	3335	12	176	GLUTEN ALLERGY	LUNG PROBLEM	4B6	99	4	ME	84		44				
	7156	15	169		DIABETES	4B6	80	5	MG	64		48		OLD AGE		
	20592	11	191			1B3	90	7	EU	71		71				
51	11368	6	164	TUMEUR AU CERVEAU	HEART ATTACK	1B3	105	6	ME	70		42		LEUKAEMIA		
	5185	11	167		ACCIDENT (HAVE A FALL IN THE FOREST)	4B6	109	4		60		63		HEART ATTACK	FGB	C148T
	22018	11	177		STROKE	4B6	75	5	PR	60		70		UNKNOWN		
35	5050	10	178		DON'T KNOW	1B3	97	7	ME	72		50				
27	11704	10	157		UNKNOWN CANCER	NONE	92	6	FM	102		69				
	15323	10	176			1B3	71	5		78		60	N02BA		FGB;SERPINF1	C148T;N47T;N60T
40	11930	11	175		COLON CANCER	1B3	99	4	QW	61	HEADACHE	45				
	10496	11	174		UNKNOWN CANCER	1B3	115	7	MS	69		55		CARDIAC INSUFFICIENCY		
59	5844	9	174	PANCREAS CANCER		1B3	80	6	FM	79	THROMBOSIS AND P	55		UNKNOWN		
		10	159	ABDOMINAL ANEURYMS AORTA		4B6	85	6		89	PAIN	60	A07EB		FGB	C148T

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS      ID         REF    ALT      QUAL FILTER INFO                              FORMAT      NA00001        NA00002        NA00003
20     14370    rs6054257  G      A        29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20     17330    .          T      A        3    q10    NS=3;DP=11;AF=0.017              GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3   0/0:41:3
20     1110696  rs6040355  A      G,T      67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2   2/2:35:4
20     1230237  .          T      .        47   PASS   NS=3;DP=13;AA=T                  GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20     1234567  microsat1  GTC    G,GTCT   50   PASS   NS=3;DP=9;AA=G                   GT:GQ:DP    0/1:35:4       0/2:17:2       1/1:40:3
```

Visit from the CRG

- Sabela de la torre and Frederic Haziza

- Presentation of the beacon API

- Docker deployment

Beacon API

SMARTBEAR
SwaggerHub™

← ga-4_gh_beacon_... ∨   1.0.0-rc1 ∨

Info

Tags

Servers

🔍 Search

default ⌃
GET   /
GET   /query
POST   /query

Schemas ⌃
SCHEMA Chromosome
SCHEMA Beacon
SCHEMA BeaconAlleleRequest
SCHEMA BeaconAlleleResponse
SCHEMA BeaconOrganization
SCHEMA BeaconDataset
SCHEMA BeaconDatasetAlleleRespon
SCHEMA BeaconError
SCHEMA KeyValuePair
SCHEMA DataUseConditions
SCHEMA AdamDataUse

Aa ☀ PUBLISHED ∨

Read Only

```
 19    get:
 20      description: Get information about the beacon
 21      operationId: getBeacon
 22      responses:
 23        '200':
 24          description: successful operation
 25          content:
 26            application/json:
 27              schema:
 28                $ref: '#/components/schemas/Beacon'
 29  /query:
 30    get:
 31      description: Get response to a beacon query for allele information.
 32      operationId: getBeaconAlleleResponse
 33      parameters:
 34        - name: referenceName
 35          description: 'Reference name (chromosome). Accepting values 1-22, X, Y.'
 36          in: query
 37          required: true
 38          schema:
 39            $ref: '#/components/schemas/Chromosome'
 40        - name: start
 41          description: |
 42            Precise start coordinate position, allele locus (0-based).
 43            * start only:
 44              - for single positions, e.g. the start of a specified sequence alteration where the
                    size is given through the specified alternateBases
 45              - typical use are queries for SNV and small InDels
 46              - the use of "start" without an "end" parameter requires the use of "referenceBases"
 47            * start and end:
 48              - special use case for exactly determined structural changes
 49          in: query
 50          required: false
 51          schema:
 52            type: integer
 53            format: int64
 54            minimum: 0
 55        - name: startMin
 56          description: |
 57            Minimum start coordinate
 58            * startMin + startMax + endMin + endMax
 59              - for querying imprecise positions (e.g. identifying all structural variants starting
                    anywhere between startMin <-> startMax, and ending anywhere between endMin <-> endMax
 60              - single or double sided precise matches can be achieved by setting startMin = startMax
                    XOR endMin = endMax
```

Docker deployment
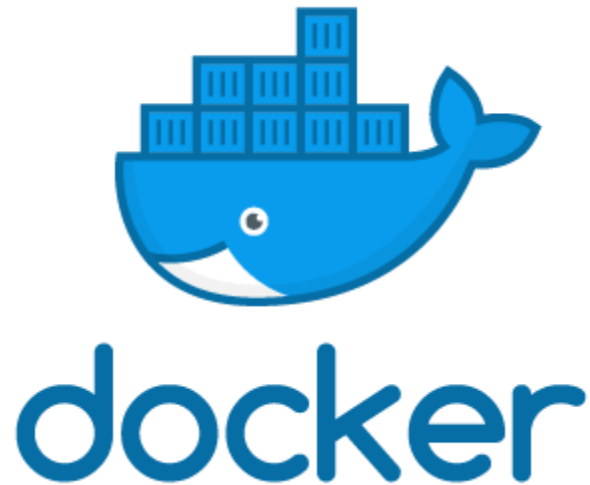
[ga4gh-beacon](#)/[beacon-elixir](#)

Deploy with Docker

We create 2 images: one for the beacon, and one for a database, preloading some test data. Run the following command in the current directory:

```
docker-compose build db
docker-compose build beacon
```

You can now instanciate the images into 2 containers. The docker-compose file contains the settings to create the network to attach the containers to, and a volume to store the data for the database. We are ready to boot up a test system, with:

```
docker-compose up -d
```

You are now ready to query the beacon on `localhost` (port 9075). For example:

- localhost:9075/elixirbeacon/v1/beacon/
- localhost:9075/elixirbeacon/v1/beacon/query?
  referenceName=Y&start=2655179&referenceBases=G&alternateBases=A&assemblyId=GRCh37&datasetIds=1000geno
  mes
- localhost:9075/elixirbeacon/v1/beacon/query?
  referenceName=Y&start=2655179&referenceBases=G&alternateBases=A&assemblyId=GRCh37&datasetIds=1000geno
  mes&includeDatasetResponses=HIT

The `-d` flag runs the containers *detached*, ie we get the prompt back. You can check the logs with:
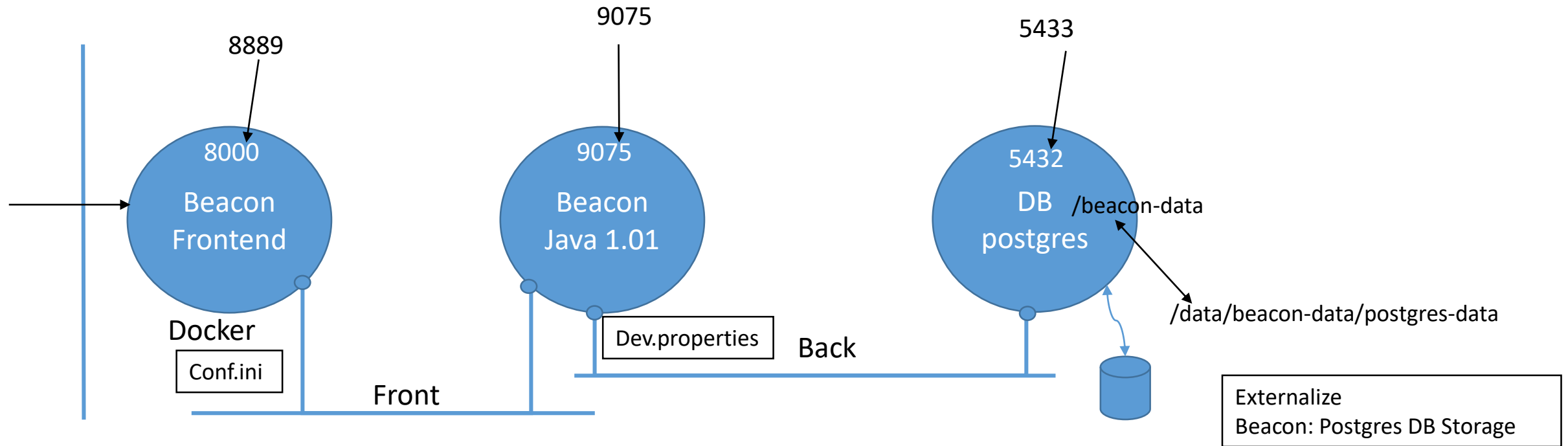
```
docker-compose logs -f
```

Tear down the system and remove the database volume, with:

```
docker-compose down -v
```

Docker deployment on HES-SO Server

Assembly ID ← ref genome
Reference Name ←7
Reference Base ←A
Alternate Base ←T
Variant ← INS DEL DUP SNP

VCF format

8889

9075

5433

8000

Beacon Frontend

9075

Beacon Java 1.01

5432

DB postgres

/beacon-data

/data/beacon-data/postgres-data

Docker

Conf.ini

Dev.properties

Back

Front

Externalize
Beacon: Postgres DB Storage

Sample file in the DB

```
1  datasetId;chromosome;position;variantId;reference;alternate;end;svType;svLength;variantCount;callCount;sampleCount;frequency;sampleMatching
2  1;Y;2655179;rs11575897;G;A;;SNP;;22;1233;1233;0.0178;22
3  1;Y;2655470;.;A;C;;SNP;;5;1233;1233;0.0041;5
4  1;Y;2655753;.;A;T;;SNP;;1;1233;1233;0.0008;1
5  1;Y;2655799;.;A;G;;SNP;;1;1233;1233;0.0008;1
6  1;Y;2655988;.;A;G;;SNP;;1;1233;1233;0.0008;1
7  1;Y;2655993;.;C;G;;SNP;;1;1233;1233;0.0008;1
8  1;Y;2656125;.;C;T;;SNP;;1;1233;1233;0.0008;1
9  1;Y;2656126;.;G;C;;SNP;;14;1233;1233;0.0114;14
10 1;Y;2656275;.;G;A;;SNP;;2;1233;1233;0.0016;2
11 1;Y;2656676;.;A;G;;SNP;;1;1233;1233;0.0008;1
12 1;Y;2657175;rs2534636;C;T;;SNP;;89;1233;1233;0.0722;89
```

Next steps

- Modify CoLaus with genetic information

- Load the data in the postgre DB

- Update the backend with access levels

  - Open: No authentication (verification of identity) or authorization (verification of access rights).

  - Registered: Authentication of the user is performed. For example it is required that the user is bona-fide researcher.

  - Controlled: Both authentication and authorization checks are performed to ensure that the user has been granted to access the data they are querying.

- Test the API

- Make available the CoLaus dataset to WP1, WP3, WP5

Useful information about the Beacon project:

- This is some Specification: https://github.com/ga4gh-beacon/specification
  - Swagger parser (paste the yaml in it): https://editor.swagger.io/
  - Latest release: v1.0.1
  - Ongoing work: v1.1.0 (branch develop)

- Reference implementation (Java & Postgres, also Docker image available): https://github.com/ga4gh-beacon/beacon-elixir
  - Latest release: v1.0.3, compliant with version v1.0.1 of the specification
  - Ongoing work: v1.1.0, compliant with version v1.1.0 of the spec (branch v1.1.0_develop)

- EGA Beacon. Sample queries:
  - Info endpoint: https://test-beacon-api.ega-archive.org/
  - Query endpoint (+ filters): https://test-beacon-api.ega-archive.org/query?referenceName=19&start=53077410&assemblyId=GRCh37&referenceBases=T&alternateBases=C&includeDatasetResponses=all&filters=ega.dataset.technology:3
  - Listing **access levels**: https://test-beacon-api.ega-archive.org/access_levels
    - Notice that the Beacon should omit all the fields marked as NOT_SUPPORTED from the response.
  - Listing filtering terms available: https://test-beacon-api.ega-archive.org/filtering_terms
  - New query type BeaconGenomicRegion: https://test-beacon-api.ega-archive.org/genomic_region?referenceName=10&start=100006353&end=100006356&assemblyId=GRCh37&includeDatasetResponses=HIT
  - New query type BeaconGenomicSNP: https://test-beacon-api.ega-archive.org/genomic_snp?referenceName=10&start=100000055&referenceBases=C&alternateBases=G&assemblyId=GRCh37&includeDatasetResponses=HIT