



ELIXIR Webinar

ELIXIR Beacon project

Sabela de la Torre
Jordi Rambla



*[www.elixir-
europe.org](http://www.elixir-europe.org)*

Contents

- Introduction
- Beacon API
- Development
- Deployment
- Questions





Introduction



*[www.elixir-
europe.org](http://www.elixir-europe.org)*

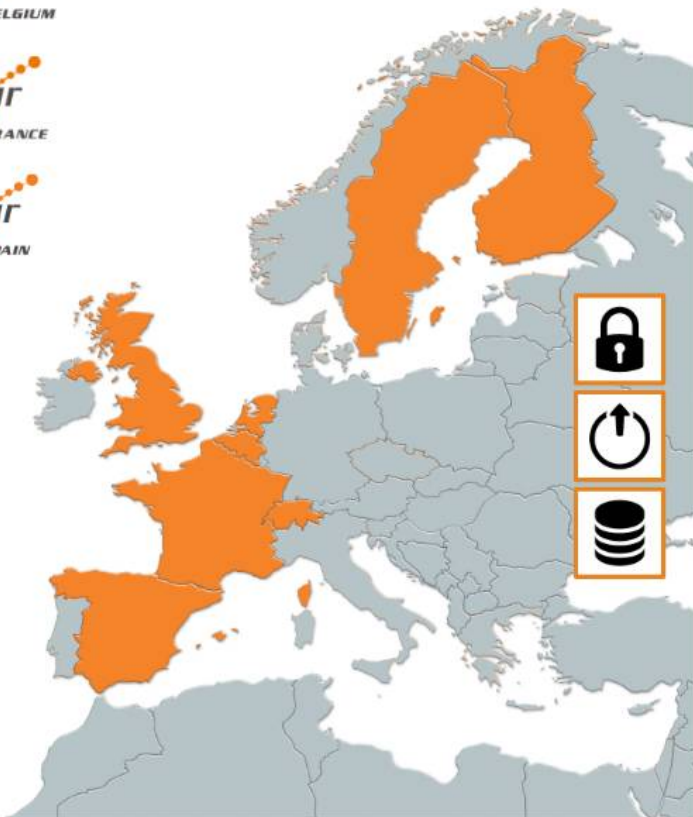
What is a beacon?



- A Beacon is a discovery service for genetic variants
- Developing tools to make data discoverable
- Aiding data sharing
- Increasing efficiency in data access process
- Answers questions like
 - Have you observed this nucleotide (e.g. C) at this genomic location (e.g. position 32,936,732 on chromosome 13)?



Elixir beacons



- Elixir Belgium
 - <http://elixirbe.esat.kuleuven.be/>
- Elixir Finland
 - 1st connected to the Elixir AAI
 - <http://elixir-beacon.csc.fi>
- Elixir France
 - <https://services.cbib.u-bordeaux.fr/beacon-web/>
- Elixir Switzerland
 - <http://beacon.arraymap.org/>
- Elixir Sweden
 - <https://swefreq.nbis.se/#/dataBeacon/>
- EGA beacon
 - <https://ega-archive.org/beacon>





Beacon API



www.elixir-europe.org

Before

- Beacon 0.1 (2014)
 - Really simple (2 records)
 - True/false response
- Beacon 0.2 (2015)
 - Complex (9 records)
 - True/false/overlap/null response
 - Datasets
 - Simple data use conditions
 - Self description
 - Not well adopted
 - Not polished enough



Now

- Beacon 0.3 (2016)
 - Simplified 0.2
 - Based on real needs
 - Improved support for datasets and cross-dataset queries
 - Modular and extensible
 - Various improvements to the data model
 - <https://github.com/ga4gh/beacon-team/releases/tag/v0.3.0>



Next

- Beacon 0.4 (in progress)
 - Support for complex variants
 - Improved data use conditions: consent codes and ADA-M
 - Documentation
 - Various minor improvements
 - Move away from avro → protobuf3
 - Test suite
 - Centralized error codes and messages



Future

- Stabilize the API: simplify and flexibilize
- 3 tier access: public, registered, controlled.
- Triple-A (registered): authentication, atestation and authorization
- Extend the query language to support data use
- Range queries
- Beacon Network API





Development



www.elixir-europe.org

Development

- Application
 - Java 8
 - Spring Boot framework
 - Hibernate
- Testing
 - JUnit
 - DBUnit

```
▼ src/main/resources
  ► META-INF
    application-dev.properties
    application-dev.yml
    application-test.properties
    application-test.yml
    git.properties
    log4j.xml
```

```
▼ ELIXIR Repo - elixir-beacon [boot] [human-data-beacon master]
  ▼ src/main/java
    ► org.ega_archive.elixirbeacon
    ► org.ega_archive.elixirbeacon.config
    ► org.ega_archive.elixirbeacon.constant
    ▼ org.ega_archive.elixirbeacon.controller
      ► ElixirBeaconBaseController.java
      ► ElixirBeaconController.java
    ► org.ega_archive.elixirbeacon.convert
    ▼ org.ega_archive.elixirbeacon.dto
      ► Beacon.java
      ► BeaconAlleleRequest.java
      ► BeaconAlleleResponse.java
      ► Dataset.java
      ► DatasetAlleleResponse.java
      ► Error.java
      ► Organization.java
    ▼ org.ega_archive.elixirbeacon.model.elixirbeacon
      ► BeaconData.java
      ► BeaconDataPK.java
      ► BeaconDataset.java
    ▼ org.ega_archive.elixirbeacon.properties
      ► SampleRequests.java
    ▼ org.ega_archive.elixirbeacon.repository.elixirbeacon
      ► BeaconDataRepository.java
      ► BeaconDatasetRepository.java
    ▼ org.ega_archive.elixirbeacon.service
      ► ElixirBeaconService.java
      ► ElixirBeaconServiceImpl.java
```



Customizing the application

```
server.port=9075
server.servlet-path=/v03
server.context-path=/elixirbeacon
...
datasource.elixirbeacon.url=jdbc\:postgresql\://127.0.0.1\:5432/elixir_beacon_dev
datasource.elixirbeacon.username=YOUR_USER
datasource.elixirbeacon.password=YOUR_PASSWORD
...
###Beacon information###
elixirbeacon.beacon.api=0.3
elixirbeacon.beacon.id=elixir-demo-beacon
elixirbeacon.beacon.name=Elixir Demo Beacon
elixirbeacon.beacon.homepage=https://ega-archive.org/elixir_demo_beacon/
elixirbeacon.beacon.alternative.url=https://ega-archive.org/elixir_demo_beacon_web/
elixirbeacon.beacon.description>Lorem ipsum
#Time zone: UTC
elixirbeacon.beacon.creation.date=2015-06-01
elixirbeacon.beacon.last.edition.date=
###Organization information###
elixirbeacon.organization.id=EGA
elixirbeacon.organization.name=European Genome-Phenome Archive (EGA)
elixirbeacon.organization.description=The European Genome-phenome Archive (EGA)...
elixirbeacon.organization.address=
elixirbeacon.organization.logo=https\://ega-archive.org/images/logo.png
elixirbeacon.organization.contact=mailto\:beacon.ega@crg.eu
elixirbeacon.organization.welcome.url=https\://ega-archive.org/
```

application-XX.properties

```
#properties
#sample #1
querySamples:
  assembly-id-1: GRCh37
  position-1: 6689
  reference-name-1: 17
  alternate-bases-1:
  dataset-ids-1:
#sample #2
  assembly-id-2: GRCh37
  position-2: 1040026
  reference-name-2: 1
  alternate-bases-2:
  dataset-ids-2: EGAD00001000740
#sample #3
  assembly-id-3: GRCh37
  position-3: 1040026
  reference-name-3: 1
  alternate-bases-3: C
  dataset-ids-3: EGAD00001000740
```

application-XX.yml



Data persistence

- PostgreSQL

```
CREATE TABLE beacon_dataset_table (  
    id character varying(50) NOT NULL PRIMARY KEY,  
    description character varying(800),  
    access_type character varying(10),  
    reference_genome character varying(50),  
    size integer  
);  
  
CREATE TABLE beacon_data_table  
(  
    id serial NOT NULL PRIMARY KEY,  
    dataset_id character varying(50) NOT NULL REFERENCES beacon_dataset_table(id),  
    chromosome character varying(2) NOT NULL,  
    "position" integer NOT NULL,  
    alternate character varying(100) NOT NULL,  
    UNIQUE (dataset_id, chromosome, "position", alternate)  
);  
  
CREATE OR REPLACE VIEW beacon_dataset AS  
    SELECT bdat.id,  
           bdat.description,  
           bdat.access_type,  
           bdat.reference_genome,  
           bdat.size  
    FROM beacon_dataset_table bdat  
    WHERE (bdat.access_type::text = ANY (ARRAY['PUBLIC'::character varying::text,  
                                              'REGISTERED'::character varying::text,  
                                              'CONTROLLED'::character varying::text]))  
    AND bdat.size > 0 AND bdat.reference_genome::text <> ''::text;  
  
CREATE OR REPLACE VIEW beacon_data AS  
    SELECT bd.dataset_id,  
           bd.chromosome,  
           bd."position",  
           bd.alternate,  
           ebdat.reference_genome  
    FROM beacon_data_table bd  
    INNER JOIN beacon_dataset ebdat ON bd.dataset_id::text = ebdat.id::text;
```

REST API

- Endpoint: / (root)

```
{
  "id" : "elixir-demo-beacon",
  "name" : "Elixir Demo Beacon",
  "apiVersion" : "0.3",
  "organization" : {
    "id" : "EGA",
    "name" : "European Genome-Phenome Archive (EGA)",
    "description" : "The European Genome-phenome Archive (EGA) is a service for permanent archiving and sharing of all types of personally identifiable genomic data.",
    "address" : "",
    "welcomeUrl" : "https://ega-archive.org/",
    "contactUrl" : "mailto:beacon.ega@crg.eu",
    "logoUrl" : "https://ega-archive.org/images/logo.png",
    "info" : null
  },
  "description" : "This <a href='\"http://ga4gh.org/#/beacon\"'>Beacon</a> is based on the GA4GH Beacon <a href='\"https://github.com/ga4gh/schemas/blob/master/elixir-demo-beacon\"'>Beacon</a>.",
  "version" : "v03",
  "welcomeUrl" : "https://ega-archive.org/elixir_demo_beacon/",
  "alternativeUrl" : "https://ega-archive.org/elixir_demo_beacon_web/",
  "createDateTime" : "2015-06-01T00:00.000Z",
  "updateDateTime" : null,
  "datasets" : [ {
    "id" : "EGAD00001000740",
    "name" : null,
    "description" : "Low-coverage whole genome sequencing; variant calling, genotype calling and phasing",
    "assemblyId" : "grch37",
    "createDateTime" : null,
    "updateDateTime" : null,
    "version" : null,
    "variantCount" : 41586925,
    "callCount" : null,
    "sampleCount" : null,
    "externalUrl" : null,
    "info" : {
      "accessType" : "PUBLIC",
      "authorized" : "true"
    }
  } ],
  {
    "id" : "EGAD00001000741",
    "name" : null,
    "description" : "Low-coverage whole genome sequencing; variant calling, genotype calling and phasing",
    "assemblyId" : "arch37",
  }
}
```



REST API

- Endpoint: /query

- i.e. /query?

- start=6688&assemblyId=GRCh37&referenceName=17&includeDatasetResponses=true

```
{
  "beaconId" : "elixir-demo-beacon",
  "exists" : true,
  "error" : null,
  "alleleRequest" : {
    "alternateBases" : null,
    "referenceBases" : null,
    "referenceName" : "17",
    "start" : 6688,
    "assemblyId" : "GRCh37",
    "datasetIds" : null,
    "includeDatasetResponses" : true
  },
  "datasetAlleleResponses" : [ {
    "datasetId" : "EGAD00001000740",
    "exists" : true,
    "error" : null,
    "frequency" : null,
    "variantCount" : null,
    "callCount" : null,
    "sampleCount" : null,
    "note" : "OK",
    "externalUrl" : null,
    "info" : null
  }, {
    "datasetId" : "EGAD00001000741",
    "exists" : true,
    "error" : null,
    "frequency" : null,
    "variantCount" : null,
    "callCount" : null,
    "sampleCount" : null,
    "note" : "OK",
    "externalUrl" : null,
    "info" : null
  } ]
}
```





Deployment



[www.elixir-
europe.org](http://www.elixir-europe.org)

Deployment

- Backend:
 - <https://github.com/elixirhub/human-data-beacon>
 - PostgreSQL database
 - Compile and package into a JAR (maven & Java)
 - Deploy the JAR
- Frontend:
 - <https://github.com/elixir-europe/human-data-beacon-ui>
 - Use NodeJS to compile the AngularJS application
 - Use a web server (i.e. Apache) to serve this application
- Or use the docker images



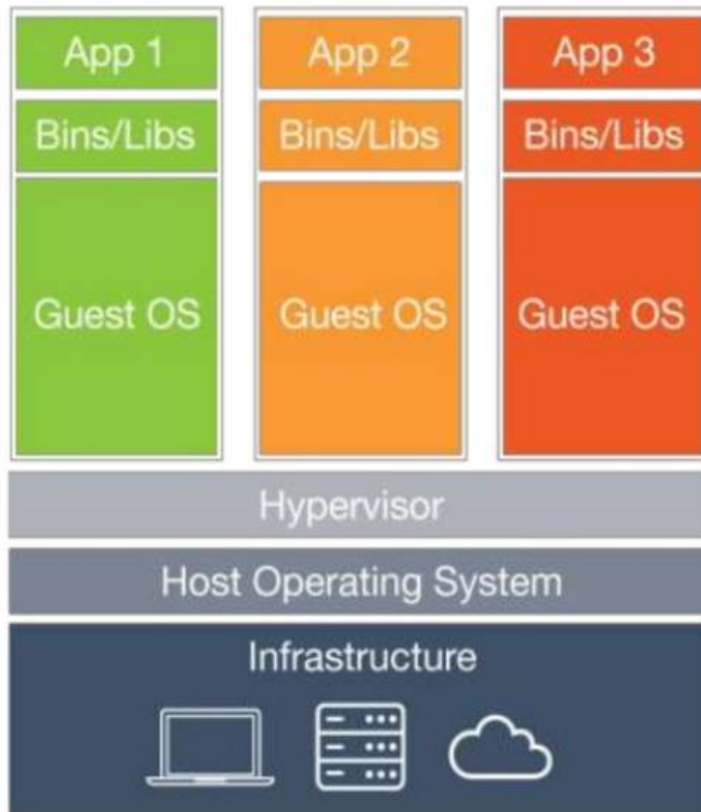
What is docker?



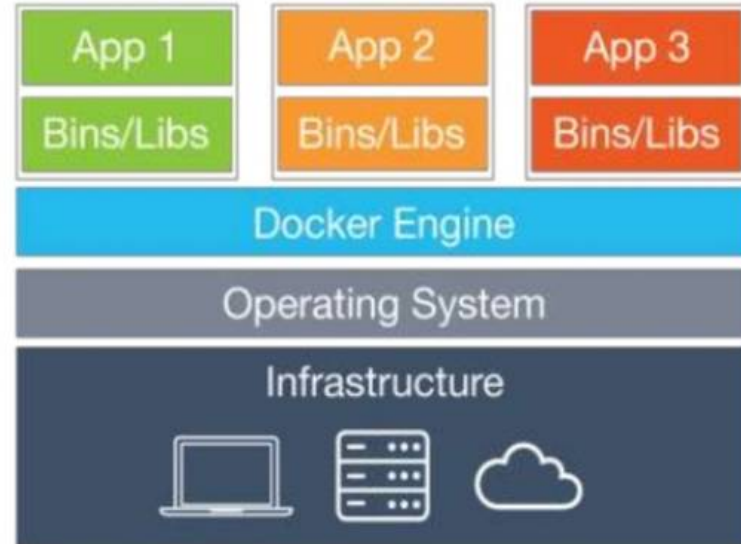
- What
 - Tool designed to make it easier to create, deploy, and run applications
- Why
 - To eliminate “works on my machine” problems for code development
- How
 - By providing a uniformed wrapper around a software package: «Build, Ship and Run Any App, Anywhere»



Virtual Machine vs Docker



VM uses a hypervisor to create hard resource boundaries between running instances.



Docker Engine shares resources, allow containers to be much more lightweight

ELIXIR Beacon Docker images

- 2 options:
 - Engine + database: <https://hub.docker.com/r/egacrg/beacon/>
 - Engine + database + UI: <https://github.com/elixir-europe/human-data-beacon-ui>

Elixir Beacon



By use of this Beacon Service, I agree to forego any attempt to re-identify individuals represented in Beacon Service Replies, except where expressly authorized by law or by a written prior permission from the respective DAC. ([more details](#))

Query available datasets

Select a dataset: EGAD00000000028 (34114 variants) ▾

Reference genome:

Chromosome: 5 ▾

Position:

Allele:

This Beacon is based on the GA4GH Beacon . Please, keep in mind that Indel queries are not supported yet.

Note: The EGA archives a large number of datasets, some of which are **Publicly** available. If you have an account on this website, you can access to **Registered** datasets. To access **Controlled** datasets can be done by contacting the relevant Data Access Committee (DAC), whose details are displayed on the Dataset description page under "Who controls access to this dataset" (click on the Dataset ID to go to the Dataset page). Once you have access to a dataset, you will be able to query it at EGA Beacon.

Dataset ID ▾	Short title ▾	Access type ▾
<input type="text"/>	<input type="text"/>	<input type="text"/>
EGAD00000000028	dataset_description	PUBLIC





Questions?



www.elixir-europe.org

Past

