ELIXIR Beacon 2019-21 Deliverable D3.3

Project Title:	ELIXIR Beacon (2019-21)
Deliverable title:	Version 2 ELIXIR Beacon Network security best practice document to include Clinical Beacon / Beacons in a Biomedical setting
WP No.	3
WP Title	Security of ELIXIR Beacon and the ELIXIR Beacon Network
Contractual delivery date:	31 December 2020
Actual delivery date:	2nd March 2021
WP leads:	Dylan Spalding
Partner(s) contributing to this deliverable:	EMBL-EBI, ELIXIR-FI, ELIXIR-ES

Authors and Contributors:

Dylan Spalding, Mallory Freeberg, Thomas Keane (EMBL-EBI)

Jordi Rambla (ELIXIR-ES)

1. Executive Summary

A Beacon is a service that responds Yes or No to a particular allele query, with the intention of improving dataset discoverability while protecting participants' privacy. Beacons can be networked, and ensuring all Beacon nodes within a Beacon network follow the same security protocols is required to ensure the security and privacy of the individuals who supplied data to the Beacons, and to ensure the Beacon and network supply accurate and consistent results. As Beacon has been previously subject to re-identification attacks, and is both a GA4GH[1] standard and the ELIXIR Beacon Network[2] is an ELIXIR service, there is a high risk of reputational damage to both GA4GH and ELIXIR with a successful security breach. For deliverable D3.3 in the 2019-2021 ELIXIR Beacon Implementation Study, version 2 of best practice recommendations has been compiled which extends the version 1 best practice document[3] to include the requirement of Beacons in a clinical or biomedical setting.

2. ELIXIR Beacon Network Security Best Practice including clinical or biomedical Beacons

Introduction

To gain insight into prospective use-cases for the Beacon, a set of meetings with six cohorts (Table 1) interested in setting up a Beacon was held during the later part of 2020. During these meetings a standing agenda item was the particular security issues the cohorts saw around the use of their data within a Beacon. In addition to the continued development of the Beacon and Beacon Network Security Best practice, the requirements from these cohorts was taken into account to ensure that the resulting Best Practice document covers Beacons within a clinical or biomedical context as well as a research context. A summary of the security concerns and issues raised by the cohorts is given below, and where applicable each issue is linked to the relevant point in version 2 of the security best practice.

[1] https://www.ga4gh.org
[2] https://elixir-europe.org/about-us/commissioned-services/beacon-network-service
[3] https://docs.google.com/document/d/13K7bSTcjga0Z0JEey7XHJhrnnj60ffKopqvGG2XT0LE/edit#heading=h.yf4b8i96olrd

RD-Nexus[4]	ELIXIR Luxembourg[5]	RD-Connect Genome-phenome Analysis Platform[6]
Cancer Core Europe[7]	ELIXIR Italy[8]	Fundación Progreso y Salud[9]

Table 1: List of the six cohorts who participated in the use-case meetings.

Summary of the cohorts security concerns

One of the main issues identified was extending the query capabilities of Beacons, for example adding filtering options, and how these additional options may re-identify individuals, especially in a rare-disease context. Examples include filtering by phenotype - different cohorts had a diverse range of views on this. It is beyond the scope of this document to recommend particular strategies for re-identification mitigation due to the diverse range of consent and legal issues applicable to the different cohorts, but a general set of principles (points 1 to 3) were added to the security best practice to ensure the particular consent and legal issues relating to a particular cohort are considered prior to setting up a Beacon.

Another issue raised was differentiating between the technical security requirements and the data governance requirements. The technical security requirements can be advised on, and detailed in the best practice document, as they are consistent across the use-cases, but the re-identification issues are cohort dependant: for example rare disease participants are likely to be more easily re-identified than other participants, but it was also commented that some rare disease participants put data sharing ahead of the possibility of re-identification.

The use of synthetic data was discussed, and these datasets provide a good way of building trust for a cohort for the Beacon implementation. And such data can be used to allow testing, both with respect to gaining trust from prospective cohorts to testing new implementations before loading participants data (point 7).

For situations where the data is firewalled and a VPN or other solution cannot be found to access the firewalled data, another option is to run an aggregation pipeline within the firewall, and then to expose the aggregated data either via a VPN, or directly outside the firewall (points 10 and 11). This is similar to the solution to be implemented by CINECA WP4 [10] for analysis on cohort data. But this solution also allows the data owner to ensure they are happy with the data governance, and means that they do not have to worry about the actual

[4] https://www.rdnexus.com/
[5] https://elixir-luxembourg.org/
[6] https://rd-connect.eu/what-we-do/omics/gpap/
[7] https://www.cancercoreeurope.eu/
[8] https://elixir-italy.org/en/
[9] https://www.sspa.juntadeandalucia.es/fundacionprogresoysalud/
[10] https://github.com/CINECA-project/wp4-federated-joint-cohort-analysis/tree/master/4.2-federated-framework

running of the beacon. This method of operation would also help cohorts satisfy the ‘data minimisation’ principle under GDPR.
Outreach is crucial to ensuring that Beacons and the Bacon network are secure (point 4), as it allows an opportunity for education and training. While the technical solutions to security are well understood, guidance on techniques and tools available to ensure the Beacon respects the legal and consensual requirements of the cohorts was mentioned by the cohorts. Additionally it gives an opportunity to build trust in the standard, and any reference implementation, which is crucial to ensure continued uptake of the standard.

Conclusion

By interviewing the six prospective cohorts regarding security issues with respect to Beacons, the security best practice document for Beacons and the Beacon Network has been updated to ensure it is relevant to clinical or biomedical cohorts. The focus of the principles has moved from being purely technical to giving some level of guidance on the processes required to ensure any prospective Beacon conforms to the ethical and legal issues of that particular cohort.

Appendix I

Beacon and Beacon Network Security Best Practice

Version 2

Data Governance

1. Approval of the controller of the data that is loaded to the Beacon (derived or otherwise) must be obtained prior to loading data to a Beacon.
2. The data controller must explicitly set out the level of data access that the Beacon will support.
3. Data controllers should be aware of techniques to minimise the risk of re-identification, such as aggregating the response away from the detail (e.g. exon response instead of individual variant, higher level ontology term for phenotypes), or aggregating the data within a Beacon from multiple subjects.
4. Maintainers of Beacons should attend training or outreach events provided by the ELIXIR Beacon Network to ensure that their Beacon is as secure as possible.
5. A risk analysis and data impact assessment should be performed prior to running a Beacon to ensure the Beacon conforms with the consent or legal basis of the participants data, ensuring that the Beacon corresponds to GDPR best practice..

6. In line with GDPR recommendations, each Beacon or node within a Beacon Network should hold minimal data. Only data that is used in this node should be made accessible.
7. Where possible a new implementation of a Beacon should be tested using synthetic data to ensure the Beacon works as expected before loading real participants data to the Beacon.

Deployment and operation

8. All queries and responses within and without a Beacon Network must be sent over HTTPS, where encryption is provided by the underlying Transport Layer Security[11,12] (TLS). Where possible, Version 1.3 should be utilised.
9. User authentications and authorisations required for registered and controlled access within the ELIXIR Beacon Network should be obtained through ELIXIR AAI[13], through which alternative GA4GH AAI[14] compatible Identity providers can be linked.
10. A Beacon may be installed within a private network for internal use. It is recommended that such Beacons are still notified to the ELIXIR Beacon group so the maintainers can be informed of security issues.
11. To allow access to data derived from firewalled data, access should be granted by either:
    a. Setting up a Beacon within the firewalled network which contains minimal data (see 3), and set this Beacon to respond only to queries from a trusted server outside the firewall or on the same Virtual Private Network (VPN), or
    b. Aggregating the data in the public Beacon in such a way that it conforms to the data governance rules applicable to the data.
12. The data required for each Beacon should have redundant copy as backup, or the data should be able to be regenerated from another source. In the case of data breach, the backup data should be restored.
13. Rate limiting and other mitigation measures should be applied to each Beacon node to reduce the risk of DDoS attack.
14. Software security patches should be applied once they become available to reduce the risk of exploitation, vulnerabilities database, such as the Common Vulnerabilities and Exposures database (CVE[15]) or National Vulnerability Database (NVD[16]) or Github security alerts to learn of new vulnerabilities.
15. The GA4GH data breach response protocol, as defined by the GA4GH Data Security [17] workstream, should be followed and extended for the ELIXIR Beacon Network when data breach incident is discovered.

---

[11] Transport Layer Security 1.2 (RFC5246)
[12] Transport Layer Security 1.3 (RFC8446)
[13] https://elixir-europe.org/services/compute/aai
[14] https://github.com/ga4gh/data-security/blob/master/AAI/AAIConnectProfile.md
[15] https://cve.mitre.org/
[16] https://nvd.nist.gov/
[17] https://www.ga4gh.org/genomic-data-toolkit/data-security-toolkit/

16. Data breaches occurring in Beacons attached to the ELIXIR BEacon network must be reported to the central Beacon Network security mailing list: beacon-security@elixir-europe.org
17. Each node should assign a specific person or persons as security point of contact for the node.
18. Each node must inform the Beacon Network Registry when their status[18] is updated. The Registry should periodically (no longer than every hour) update the service catalogue.
19. Beacon Aggregator nodes must be registered with the applicable registry.

3. Adjustments Made

Deliverable submitted ca. 2 month late.

[18] Status can be online, off-line, or changing the access requirements to the Beacon.