Healthcare Data Lake

Kendal, Joseph
University of Bristol
jk17246@bristol.ac.uk

Sherred, Jago
University of Bristol
j.sherred.2019@bristol.ac.uk

Benson, Luke
University of Bristol
wr19606@bristol.ac.uk

Liu, Anna
University of Bristol
gf19916@bristol.ac.uk

Cismaru, Armand
University of Bristol
fz19792@bristol.ac.uk

November 7, 2020

Abstract

Digital healthcare provided by the NHS in England typically operates in silos. GPs have electronic systems to manage patient care which are distinct from hospital systems which are distinct from the ambulance service, 111, mental health services etc. Each data owner has a wealth of data that, if combined, would generate a more valuable resource than it does in isolation. While there are solutions to integrate this data for direct care purposes, there is no centralised solution to use this data to inform future care or service provisioning. This project is designed to explore the benefits of cloud technologies to produce a prototype secure, scalable health data storage platform that can underpin local healthcare analytics.

1 Overview

1.1 Client

1.2 Domain

1.3 Project

1.4 Vision

2 Requirements

2.1 Stakeholders

Primary stakeholder Philip Harfield at Bristol, North Somerset and South Gloucestershire CCG (BNSSG). Philip Harfield is our client, and this software is being developed for him at BNSSG. BNSSG require a piece of software to store healthcare data from various sources to inform clinical and strategic decisions. This means that BNSSG is our primary stakeholder.

Additional stakeholders This software will provide services to a number of local healthcare organisations such as NHS trusts and the Healthier Together STP and such all these additional users are secondary stakeholders to this project. These organisations will need to be able to provide healthcare information to the software which will need to be able to load and store the data for future analysis.

2.2 User stories

API

HL7 FHIR

OpenAPI

Data lake

Data catalogue

ETL & console

3 Personal Data, Privacy, Security and Ethics Management

3.1 GDPR

3.2 Security

3.3 Ethics

4 Architecture

Overview We propose a modular, (cloud)platform-independent solution that offers high scalability and performance at a low cost for maintenance, development and deployment. The key to achieving this is leveraging the practises of infrastructure-as-code (IaC), serverless architectures and open standards. Therefore, development expense is focused on delivering the most value, flexibility and ease-of-use to the clients.

5 Development Testing

6 Release Testing

7 OO Design & UML