A QUEUE-POLL EXTENSION

STANDARDISED, MONITORED, INDIRECT AND SECURE DATASHIELD ACCESS TO YOUR DATA

Julian Gruendner
07.11.2018
DataSHIELD Workshop 2018
• DataSHIELD does not require data to leave hospitals, but it requires requests to enter and responses to exit the hospital network.

• However:
  • application Provider IS NOT System Admin
  • firewalls are usually restrictive in a hospital setting
  • joining and exiting a data analysis network should be easy
Background

Problem

STRICT FIREWALLS of hospitals allow no direct access to opal server from outside network

Solution

QUEUE-POLL implementation to allow indirect access from outside
DS – Queue-Poll Extension

HOSPITAL NETWORK

Poll Service and Opal Server (VM)

- DS POLL GUI Server Container
- DS POLL Container: GUI Web Service and Running Poll Service
- datashield_docker_opal_net – docker network
- Opal Server Container
  - Data & Metadata Module
  - R Module
  - DataSHIELD
- Web Services
- Mongo DB
- R Server Container (RServe)

CLIENT

HTTPS

QUEUE

HTTPS

HTTPS

HTTPS
DS – Queue-Poll Extension

Poll Service and Opal Server (VM)

- DS POLL GUI Server Container
- Opal Server Container
  - Data & Metadata Module
  - R Module
  - DataSHIELD
- Web Services
- Mongo DB
- R Server Container (RServe)

HOSPITAL NETWORK

FIREWALL

HOSPITAL DMZ

CLIENT

HTTPS

QUEUE

HTTPS
How is extra Security provided?

• Queue only accepts Connections from authorized peers, configured via IP address

• Opal has built-in roles and rights management, which further restricts data access

• Data Provisioning only on official request
DS – Configurable Circle of Trusted IPs
The Installation Process

- **Requirements**: Ubuntu 18 or >

- **Fully automated installation** and provisioning process includes:
  - The Q-P extension
  - Opal and Datashield
  - Test data, which is automatically loaded into opal
  - An R server
Q-P Monitoring and Configuration

• User Interface for Configuring poll and Monitoring requests
  • Start-Stop Poll and configure queue endpoint
  • Live-Stream of incoming requests
  • Logfile export

Poll Service Control

Poll Status: RUNNING
Please insert your queue server host and port and your opal host and port - if left empty the default values will be used: Queue Server: queue_server:8001 Opal Server: datashield_opal:8443

Queue Server host and port
queuehost:80

Opal Server host and port
opalhost:80

START POLL SERVICE  STOP POLL SERVICE
Performance of Q-P

- Execution time with Q-P 2.4 times direct request time
- Impact of Q-P decreases with larger number of participants in dataset:
  - 5000 participants = 2.8 times direct request time
  - 50000 participants = 2.2 times direct request time
What is Next?

• Improvement of monitoring and logging
• Automated archiving of logfiles
• Field test and incorporation of feedback
• Improvement of setup scripts and establishing test pipeline
• (Analysis tools for logfiles)
• Create request approval mechanism
A quick live Demonstration