FHIR® IL by 8400 | State of Israel Ministry of Health | 8400 | JOINTElka | Israel Innovation Authority | Ministry of Digital Affairs DIGITAL ISRAEL | In Partnership with: Yad Hanadiv

Solution Architecture for JCPM Research Project

outburn

Table of Contents

Document History

Revision #	Date	Author	Main Changes
1	01/10/2021	Alex Baumberg Daniel Mechanik	Solution Architecture Answers added (Outburn part) Architecture Solution overview added. (Hospitals only)
2	03/10/2021	Hadassah	Solution Architecture Answers added (Hadassah part)

Solution Architecture Questions

1. What is your architectural approach (FHIR server, Façade, Asynchronous messaging, Hybrid)?
   FHIR server

You may find assistance or use the implementation approaches matrix and diagrams on **Appendix A** to help you decide.

The list of resources that will be exchanged within the processes is on **Appendix B**

a. Provide architecture diagrams including source, target and intermediary systems, application and storage components, security components. Indicate cloud/on-prem separation, if applicable. Indicate vendors and platform names. If custom development is required, indicate runtime platform/programming language.

b. Provide data flow diagrams for common scenarios.(attach a number beside each call in the sequence)

2. What is the total dataset size that you will be exposing via FHIR?

   Information will be shared upon FHIR specification completion

   Is this data originating from/should be copied to other organizational systems?

   No

3. If the data must be synchronized with other organizational systems - what is the acceptable synchronization delay?

   No

4. What is the size of a single record (in a business sense - might include several FHIR resources) that will be transferred?

   Information will be shared upon FHIR specification completion

   What's the number of records to be transferred per day/during peak load?

   N/A – no daily records transfer in current research project

5. Will the FHIR interface be exposed to multiple consumers?

   No

   What is the expected number of consumers?

   1

   What is the expected amount of concurrent requests during peak load?

   No concurrent requests expected

6. What infrastructure/platform will be used for FHIR server/façade/messaging?

   HAPI FHIR

   Tibco ESB

Provide vendor and system names. Is it already present in your organization or will be acquired/installed for the project?
HAPI FHIR will be installed for the project.
Tibco ESB is already in use as an organizational ESB

Does it natively support FHIR in client and/or server modes (i.e. FHIR client and/or FHIR server/facade is built into the platform) or it will require additional extensions/modules/custom development?
No, TIBCO ESB doesn't support FHIR

7. Where applicable - how scalability/availability/redundancy will be addressed?
   NA for research project

8. If FHIR façade/Server will be used - where and how data will be stored?
   FHIR data will be stored in Postgres DB according to the HAPI's JPA Data Model.

9. Where will the components of the solution be located (on-prem/cloud/hybrid)?
   If on a cloud , please describe which provider
   On-prem

10. How will the FHIR interface be monitored for quality & availability?
    Monitoring will be performed by the organizational monitoring server- TIBCO HAWK providing overall TIBCO ESB (FHIR Client) monitoring capabilities:
    - Infrastructure monitoring (e.g. Network & REST API failures)
    - Business Process monitoring (Syntax and Semantic exceptions)

11. How the interface will be secured (VPN, static IPs, TLS & certificates, etc.)? Will specialized security platforms/gateways be used for online/asynchronous schema validation?
    If yes, do they natively support FHIR?
    - TLS & certificates for FHIR Client (TIBCO ESB) and HAPI FHIR Server
    - TLS, certificates and schema validation (F5 reverse proxy) data exchange with partner organizations

12. For message based asynchronous communication –
    How data will be packaged (e.g. resources as individual files, FHIR bundle, custom envelope format - e.g. JSON array, bulk FHIR, etc.)?
    FHIR bundle (transaction type)

    Will space optimization (e.g. compression, BSON) be used?
    No

Is the selected infrastructure/platform compatible with the chosen format out of the box or additional customizations will be required?
Yes, additional customization will be required on the FHIR client side (TIBCO ESB)

13. Will FHIR resources conformance validation be performed and if yes - how it'll be done (online/ batch, what tools/infrastructure will be used)?

    Yes, HAPI FHIR server supports profile validation out of the box, and we plan to perform validation when posting resources to the server. Performance issues may arise when doing so on large number of resources. In this case we will reconsider profile validation usage.

14. Will code systems validation be performed and if yes - how it'll be done (online/ batch, what tools/infrastructure will be used)?
    Yes, this is an integral part of the profile validation.

Appendix B

Architecture Solution Overview

The current document describes the high-level workflow architecture required for establishing data preparation, de-identification, and "model results" data exchange between partner organizations participating in the current research project.

Architectural considerations

Following considerations were taken into account during the architecture build-up:

1. **Patient privacy**: Source data used for the research should be anonymized according to the de-identification mask, approved by the "Helsinki data commission."
2. **Identified data access:** Proposed architecture is preventing unauthorized access to the patient-identified data.
3. **De-identified data export control:** De-identified data export is conditioned on organization review and approval.
4. **Export process initiation control:** The process initiation is controlled by the authorization mechanism (Loopback ACL)
5. **Inter-organizational data exchange:** Internal data exchange operations are allowed only and only using a secure communication protocol (TLS 1.2 + Authentication).
6. **External data exchange:** External data exchange process with partner organization is managed by "Secure Proxy Server" resized in the DMZ using secured REST API call.
7. **Development / Service Team data access:** Data access to the Physical Server keeping solution components is allowed by SSL VPN.

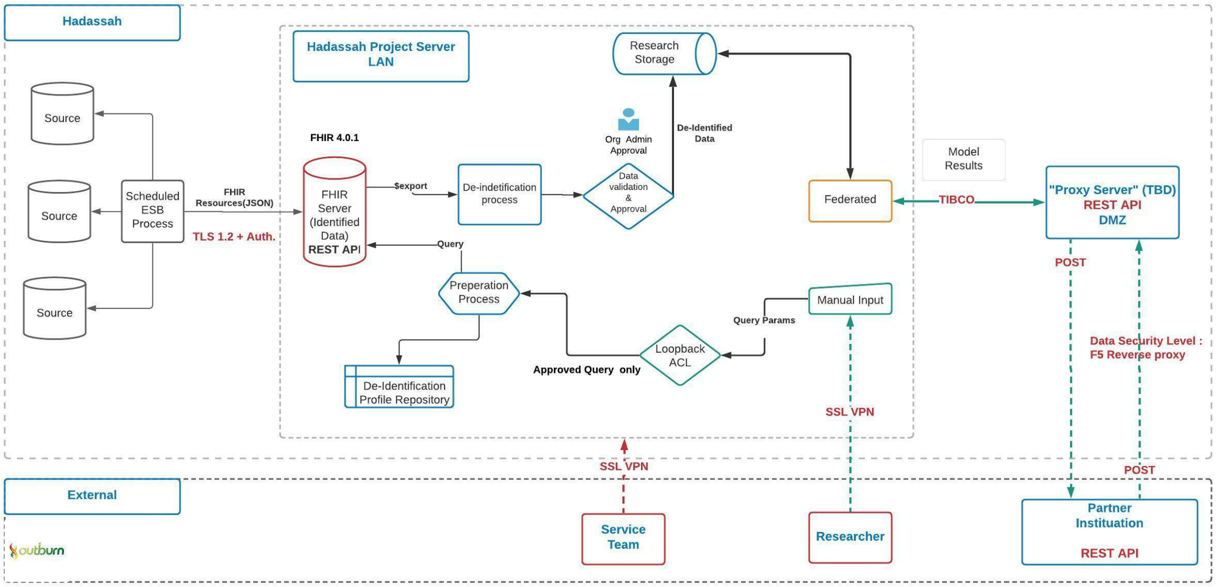
Main components and data flow

Vendors and Platforms used.

- **Deployment method**: On-Prem only
- **ESB**: Organizational ESB service (Hadassah – TIBCO)
- **On-prem physical server OS:** Linux Ubuntu 21.04
- **FHIR Server**: HAPI FHIR v5.4.1, PostgreSQL based
  - Final decision about FHIR Server vendor will be reconsidered upon specification completion
- **Federated Learning**: IBM
- **Data preparation and de-identification process**: Developed by Outburn. Deployment based on NodeJS and JSONata
- **Manual Input tool:** Developed by JCPM, JS based.
- **Researcher Authorization component:** Loopback ACL
- **Inter-organizational secure data exchange**: Tibco, F5 reverse proxy

Data flow

The diagram below represents the entire, high-level architecture in a single organization (approved by Hadassah)



1. **Source**: Organizational data sources of identified data.
2. **ESB**: Organizational ESB system responsible for native to FHIR transactions transformations according to the FHIR model specified for particular research.
3. **FHIR Server**: FHIR server instance installed on the physical project server connected to the organizational LAN

Identified data injection process: The data injection process is managed by a scheduled ESB process aimed at retrieving the organizational data from primary data sources, transforming it to FHIR based transactions, and sending it over to the FHIR server using secure REST API calls. FHIR server population by the identified data is a mandatory pre-requisite for the further processes described below.

4. **De-identification profile repository** keeps de-identification profile (mask) used for data de-identification approved by the organizational "Helsinki data committee." The profile repository is capable of storing multiple de-identification profiles per research.
5. **Data preparation process** aimed to retrieve a research population defined by the de-identification profile stored in the profile repository from the FHIR Server using FHIR RESTful patient-level Queries for further de-identification.
6. **De-identification process** aimed to apply **a** de-identification mask on the identified data previously retrieved from the FHIR server for further approval by the organization.

7. **Data validation and approval components** allow access to de-identified data stored in ndjson format by the organization for review and approval. The de-identified data review and approval is a mandatory workflow sub-process for the further data population into the Research Storage.
8. **Research Storage** is keeping de-identified ndjson files used as a source of the data for the Federated learning server.
9. **Federated Learning.** The Federated machine learning technique trains an algorithm across multiple organizations holding data samples retrieved from research storage. As a result, the "Model Results" exchange between partner organizations participating is needed in the scope of particular research. Data exchange between different organizations is performed by a secure Proxy server installed in the DMZ of the organizational network.
10. **Proxy server** used for a secure "Model Results" data exchange between Federated Learning machines located at partner organizations. The data exchange process is managed through secure REST API calls.
11. **" Manual Input" component** used for the entire data preparation and de-identification process initiation by the researcher. The component is accessible by secure communication over the SSL VPN and requires authorization through the Loopback ACL component. (username, password). The Loopback ACL component is keeping user credentials linked to the specific research(s) and aimed to control a data export process initiation for authorized researchers only having permission to get access to particular research only. The manual input uses methods exposed by the REST services of the data preparation component and providing the following capabilities to the researcher:
    a. **Initiate export**
    b. **Refresh population (true/false).** The current method is providing capabilities to initiate an export for the patient population and referenced clinical data stored by the FHIR server as following:
        i. **true:** the most recent population injected into the FHIR server by the scheduled ESB process will be used.
        ii. **false**: recurrent data export using population used at previous run.
    c. **Get the number of patients used for the current run.**
    d. **Check export status.**

Appendix A

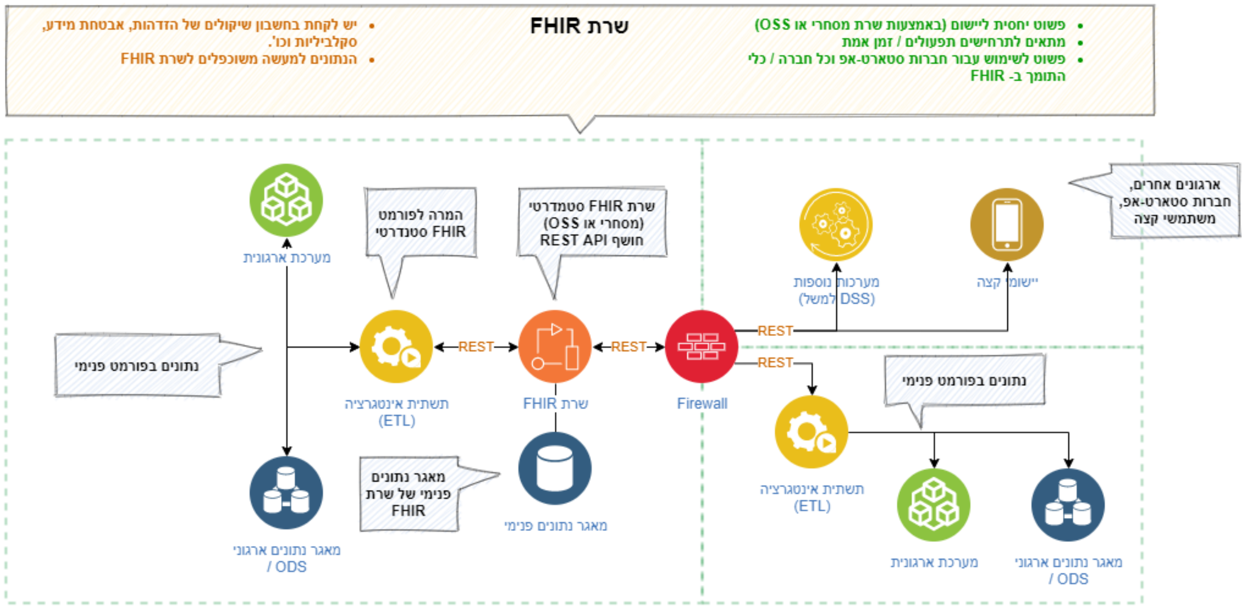
Implementation approaches

	FHIR server	FHIR Façade	Messaging
Legend: - Green: well suited - Orange: partially suited - Red - Ill suited	online request/response style communication with the data persisted in the FHIR server itself and replicated in/out to other systems as necessary	online request/response style communication without persisting the data, but rather dynamically translating online requests to/from FHIR and forwarding them to other systems	asynchronous/batch communication, exchanging FHIR payload via message bus/queue/file shares/כספות
Large dataset used by / originated in other organizational systems that do not support FHIR	🟥	🟩	🟩
Small dataset / dataset dedicated for specific task and not used by other systems	🟩	🟥	🟥
Business needs call for online interaction	🟧	🟩	🟥
Data must be in sync with other systems in near-real time	🟧	🟩	🟧
System must support high number of concurrent requests	🟩	🟧	🟩
Time to market & solution complexity	🟩	🟧	🟩
Business needs require advanced functionality on the server side (e.g. search)	🟩	🟧	🟥

Large volumes of data must be transferred			

Implementation approaches diagrams

1. Approach #1 - using FHIR server

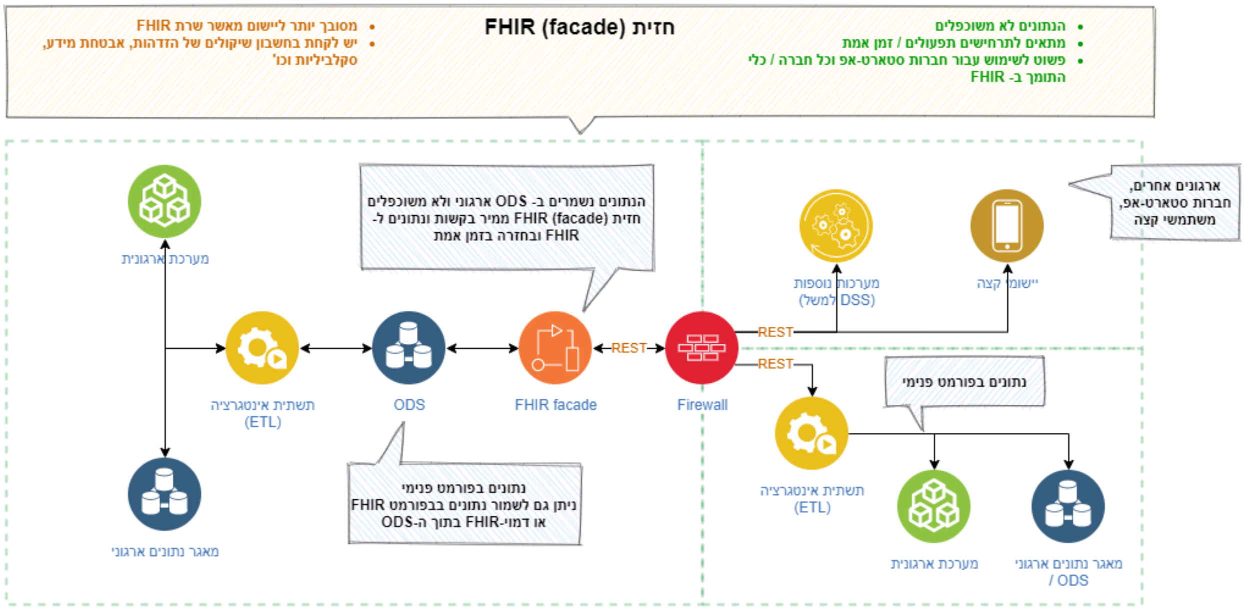


שרת FHIR

- פשוט יחסית ליישום (באמצעות שרת מסחרי או OSS)
- מתאים לתרחישים תפעולים / זמן אמת
- פשוט לשימוש עבור חברות סטארט-אפ וכל חברה / כלי התומך ב- FHIR

- יש לקחת בחשבון שיקולים של הזדהות, אבטחת מידע, סקלביליות וכו'.
- הנתונים למעשה משוכפלים לשרת FHIR

2. Approach #2: Using a FHIR Façade



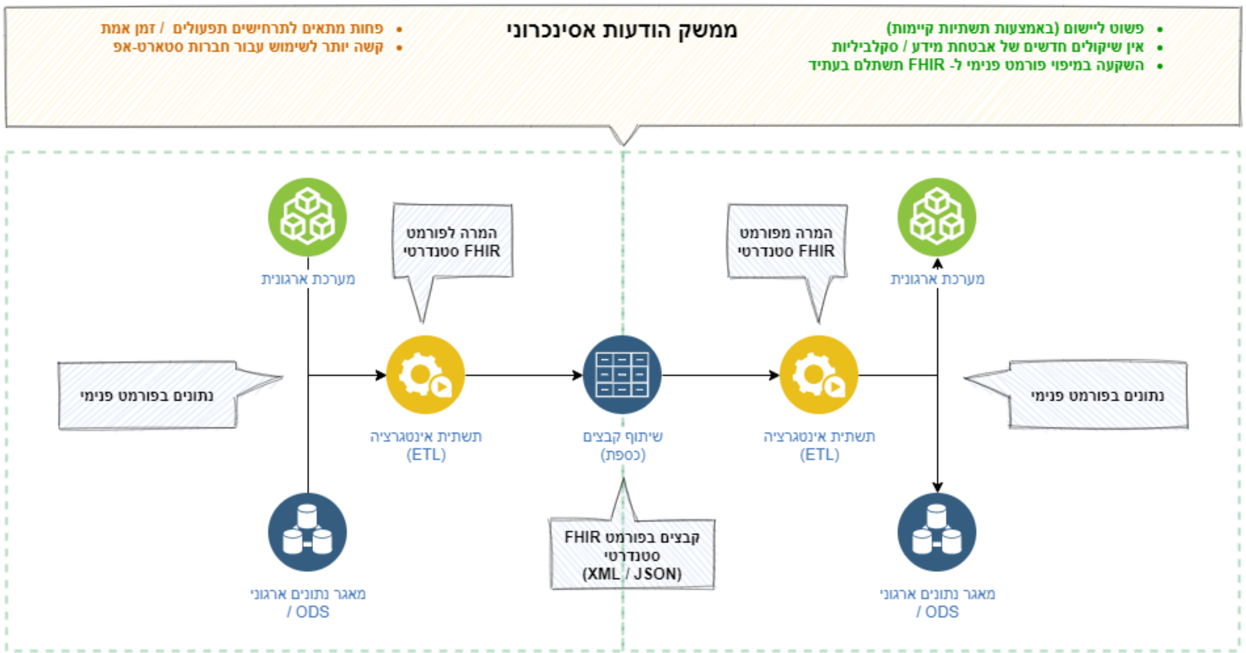
FHIR (facade) חזית

- הנתונים לא משוכפלים
- מתאים לתרחישים תפעולים / זמן אמת
- פשוט לשימוש עבור חברות סטארט-אפ וכל חברה / כלי התומך ב- FHIR

- מסובך יותר ליישום מאשר שרת FHIR
- יש לקחת בחשבון שיקולים של הזדהות, אבטחת מידע, סקלביליות וכו'

מערכת ארגונית
תשתית אינטגרציה (ETL)
ODS
FHIR facade
REST
Firewall
מאגר נתונים ארגוני
הנתונים נשמרים ב- ODS ארגוני ולא משוכפלים חזית (facade) FHIR ממיר בקשות ונתונים ל- FHIR ובחזרה בזמן אמת
נתונים בפורמט פנימי ניתן גם לשמור נתונים בבפורמט FHIR או דמוי-FHIR בתוך ה-ODS
ארגונים אחרים, חברות סטארט-אפ, משתמשי קצה
מערכות נוספות (למשל DSS)
יישומי קצה
REST
REST
נתונים בפורמט פנימי
תשתית אינטגרציה (ETL)
מערכת ארגונית
מאגר נתונים ארגוני / ODS

3. Approach # 3: Using Asynchronous messaging



ממשק הודעות אסינכרוני

- פשוט ליישום (באמצעות תשתיות קיימות)
- אין שיקולים חדשים של אבטחת מידע / סקלביליות
- השקעה במיפוי פורמט פנימי ל- FHIR תשתלם בעתיד

- פחות מתאים לתרחישים תפעולים / זמן אמת
- קשה יותר לשימוש עבור חברות סטארט-אפ