Real-Time Medical Record Processing

Austin Bennett

BEAM SUMMIT NYC 2025

AZRA AI

“Identify patients in need of care faster, connect them to care sooner, and manage the whole patient journey, all in one end-to-end platform .”

www.azra-ai.com

What We Do

“Identify. Connect. Manage.”

Azra AI offers the industry’s only end-to-end platform for oncology and beyond that identifies patients in real-time, connects them to care sooner, and manages their entire cancer journey—all within a single solution. By automating workflows and utilizing advanced AI, we help oncology service lines reduce care delays, improve patient retention, enhance patient experience and outcomes, and boost net patient revenue.

Identify: Our advanced AI rapidly analyzes pathology and radiology reports, pinpointing cancer diagnoses and suspicious incidental findings in real-time. This ensures that every patient is identified promptly and no critical detail is missed.

Connect: Once identified, patients are immediately connected to care. Our system sends instant alerts to Navigators, enabling fast outreach and ensuring that patients receive timely support, which is crucial for better outcomes.

Manage: From suspicion through survivorship, Azra AI’s platform manages the entire patient journey within one seamless system. We empower oncology service lines with actionable insights derived from real-time analytics, driving growth and enhancing patient care. Additionally, our expert consulting services are available to support your team in effectively managing change and optimizing operations.

Outline:

1) Medical Standards/Tech

2) Use Case: Healthcare Translations

3) Future

HIPAA

https://www.hhs.gov/hipaa/index.html

HIPAA is a federal law that safeguards individuals' medical information by restricting who can look at and receive your health information, while also ensuring that your health records remain private and secure. It also provides you with the right to access your own health records and request corrections.


HITRUST®

Medical Standards

There are some contemporary ‘standards’ for Health/Medical.

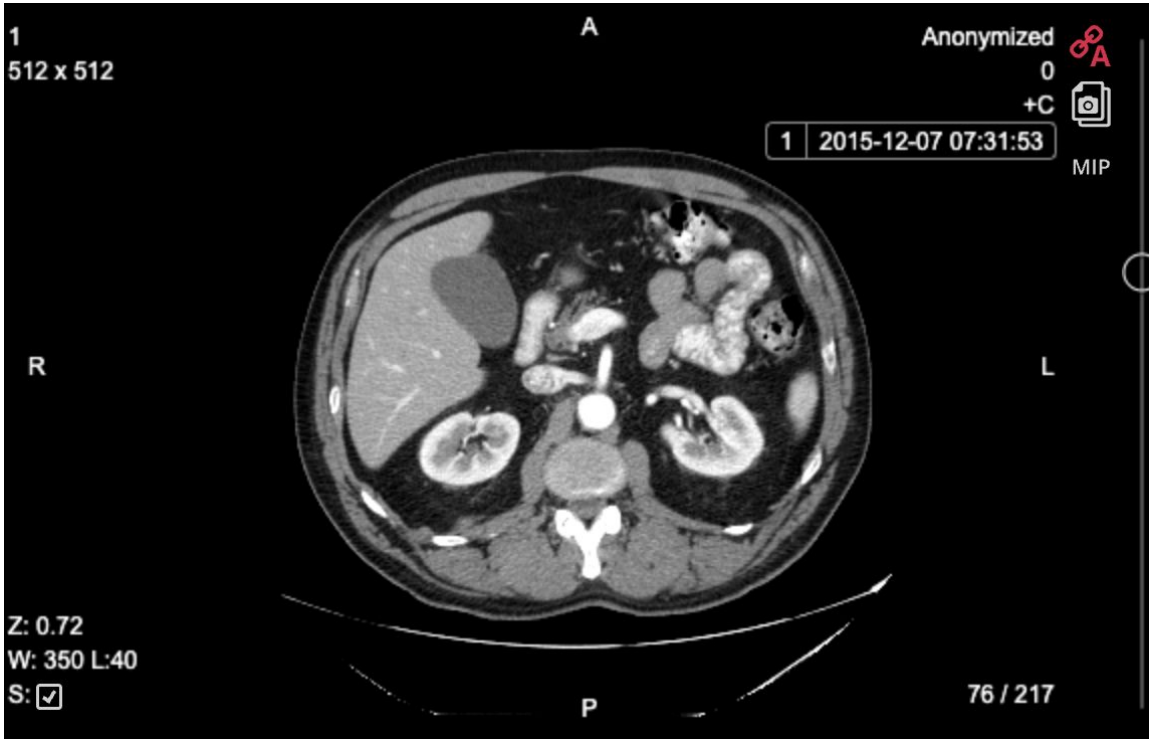
- DICOM
- Esp. Health Level 7
  - https://en.wikipedia.org/wiki/Health_Level_7
  - HL7v2
  - FHIR

Medical Standards: DICOM

https://en.wikipedia.org/wiki/DICOM

https://www.dicomstandard.org/

Medical Standards: DICOM



Medical Standards: HL7

Health Level 7

- HL7v2
- FHIR

Medical Standards: HL7v2

```
MSH|^~\&|MegaReg|XYZHospC|SuperOE|XYZImgCtr|20060529090131-
0500||ADT^A01^ADT_A01|01052901|P|2.5
EVN||200605290901||||
PID|||56782445^^^UAReg^PI||KLEINSAMPLE^BARRY^Q^JR||19620910|M||2028-
9^^HL70005^RA99113^^XYZ|260 GOODWIN CREST DRIVE^^BIRMINGHAM^AL^35209^^M~NICKELL'S
PICKLES^10000 W 100TH AVE^BIRMINGHAM^AL^35200^^O|||||||0105I30001^^^99DEF^AN
PV1||I|W^389^1^UABH^^^^3||||12345^MORGAN^REX^J^^^MD^0010^UAMC^L||67890^GRAINGER^LUCY^X^^
^MD^0010^UAMC^L|MED|||||A0||13579^POTTER^SHERMAN^T^^^MD^0010^UAMC^L|||||||||||||||||||||
||||||200605290900
OBX|1|NM|^Body Height||1.80|m^Meter^ISO+|||||F
OBX|2|NM|^Body Weight||79|kg^Kilogram^ISO+|||||F
AL1|1||^ASPIRIN
DG1|1||786.50^CHEST PAIN, UNSPECIFIED^I9|||A
```

Medical Standards: HL7v2

- First introduced in 1989 [HL7 Founded 1987]
- Currently HL7v2.9

Medical Standards: HL7v2

… 2.3, … 2.8 … JUST ‘wider’ [MORE FIELDS]

Medical Standards: MLLP

LLP [Lower Level Protocol]

or

MLLP [Minimum LLP]

Transport for HL7, over TCP

https://github.com/GoogleCloudPlatform/mllp

Medical Standards: FHIR

Medical Standards: FHIR

Major Milestones in FHIR Standardization

Date	Version	Description
2011-08-18	-	The initial draft of FHIR, then known as Resources For Healthcare (RFH), was published on Grahame Grieve's blog in Australia[3][4]
2011-09-11	-	The standard was adopted by Health Level Seven International (HL7) as a work item[4]
2014-09-30	0.082	DSTU1 (First Draft Standard for Trial Use) official version published[5][6]
2015-10-24	1.0.2	DSTU2 (Second Draft Standard for Trial Use) official version published[5]
2019-10-24	3.0.1	STU3 (Third Standard for Trial Use)[5] included coverage of a variety of clinical workflows, a Resource Description Framework format, and a variety of other updates[7]
2019-10-30	4.0.1	Release 4 has the First Normative Content and Trial Use Developments[5][8]
2023-03-26	5.0.0	Release 5 [5][9]

https://hl7.org/fhir/

https://en.wikipedia.org/wiki/Fast_Healthcare_Interoperability_Resources

Medical Standards: FHIR

Primarily REST / JSON … Though also supports RDF, XML

Healthcare Medical Systems

EMR / EHR

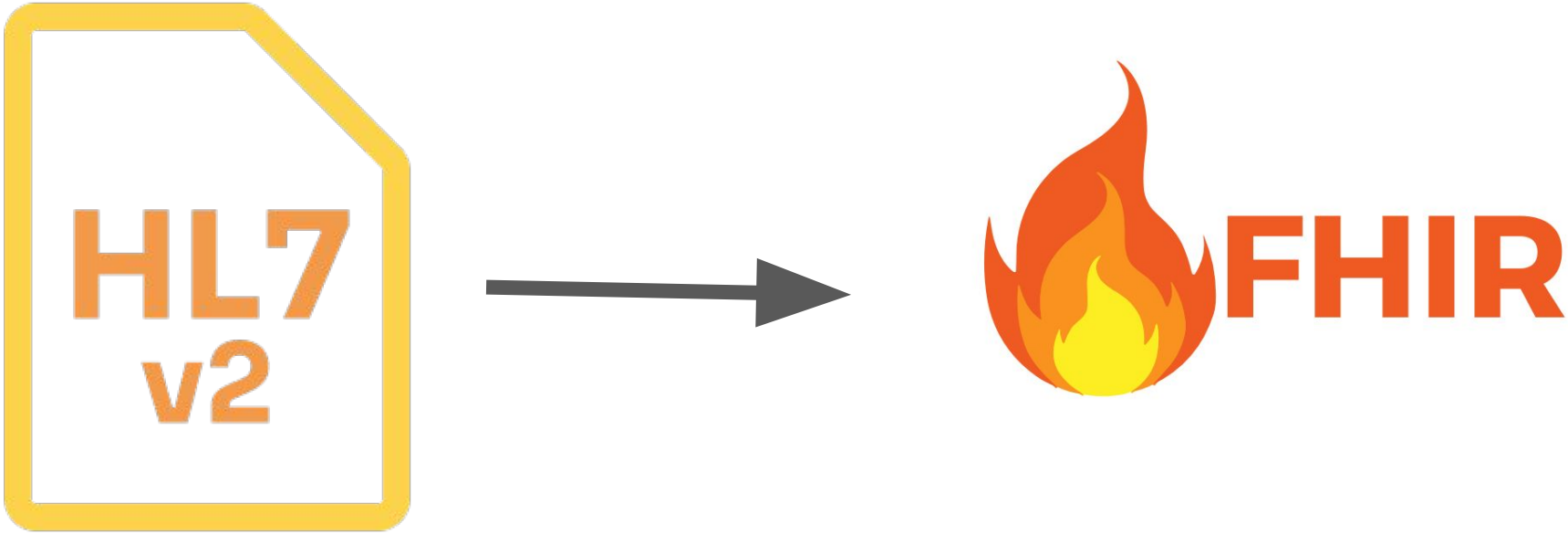
Electronic {**H**ealth/**M**edical } **R**ecord

Healthcare Medical Records

Even though there are standards, plenty of individual differences, deviations

Doing something

PROBLEM



Transformations

LOTs of ways to handle

How to limit what needs to be [re-]invented

Transformations

- Commercial Offerings
- https://github.com/LinuxForHealth/hl7v2-fhir-converter
  - Obviously missing message types
- https://github.com/microsoft/FHIR-Converter
  - Looked robust, was a candidate
- https://github.com/GoogleCloudPlatform/healthcare-data-harmonization-dataflow
  - And associated repos

GCP Healthcare API

- DICOM
- HL7v2
- FHIR
- Other useful stuff …

https://cloud.google.com/healthcare-api

“Harmonization”

https://github.com/GoogleCloudPlatform/healthcare-data-harmonization

And

https://github.com/GoogleCloudPlatform/healthcare-data-harmonization-dataflow

And

https://github.com/GoogleCloudPlatform/mllp

Transformations: Data Healthcare

Tools

- GCP Healthcare API
- Cloud Run [or other runtime for container]
- Dataflow
- PubSub [or other message queue]
- BigQuery [or other datastore if wanting to query/analyze]

Pub/Sub

- Rock Solid Message queue
  - Notable for ‘push’ subscriptions …

Cloud Run

MLLP && PubSub handler

https://github.com/GoogleCloudPlatform/mllp

https://cloud.google.com/blog/products/serverless/cloud-run-now-supports-multi-container-deployments

Mappings

- Whistle
- https://github.com/GoogleCloudPlatform/healthcare-data-harmonization

Whistle Data Transformation Language

Introduction

Whistle is a mapping language used for converting complex, nested data from one schema to another.

Whistle is a terse, efficient syntax to describe transformations of healthcare data, but is applicable to any domain.

In addition to the built-in functionality, the engine can be extended with plugins which can provide native transformations, extra features, integration with external services, and otherwise extend the engine functionality.

Beam & Dataflow

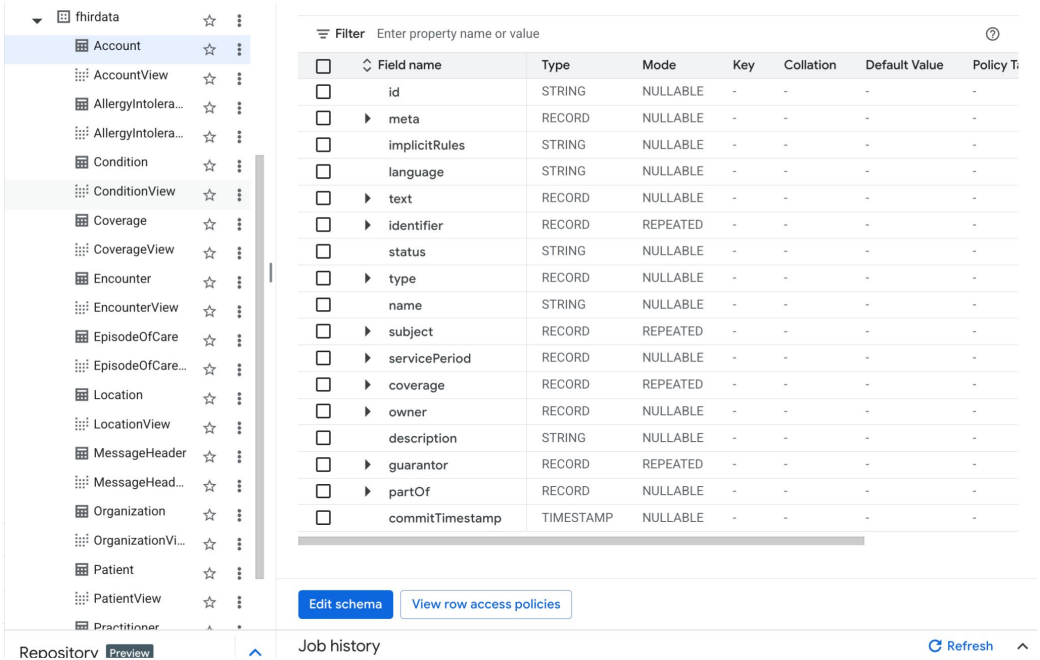
- BEAM SUMMIT!
- HIPAA Compliant

Dataflow Template

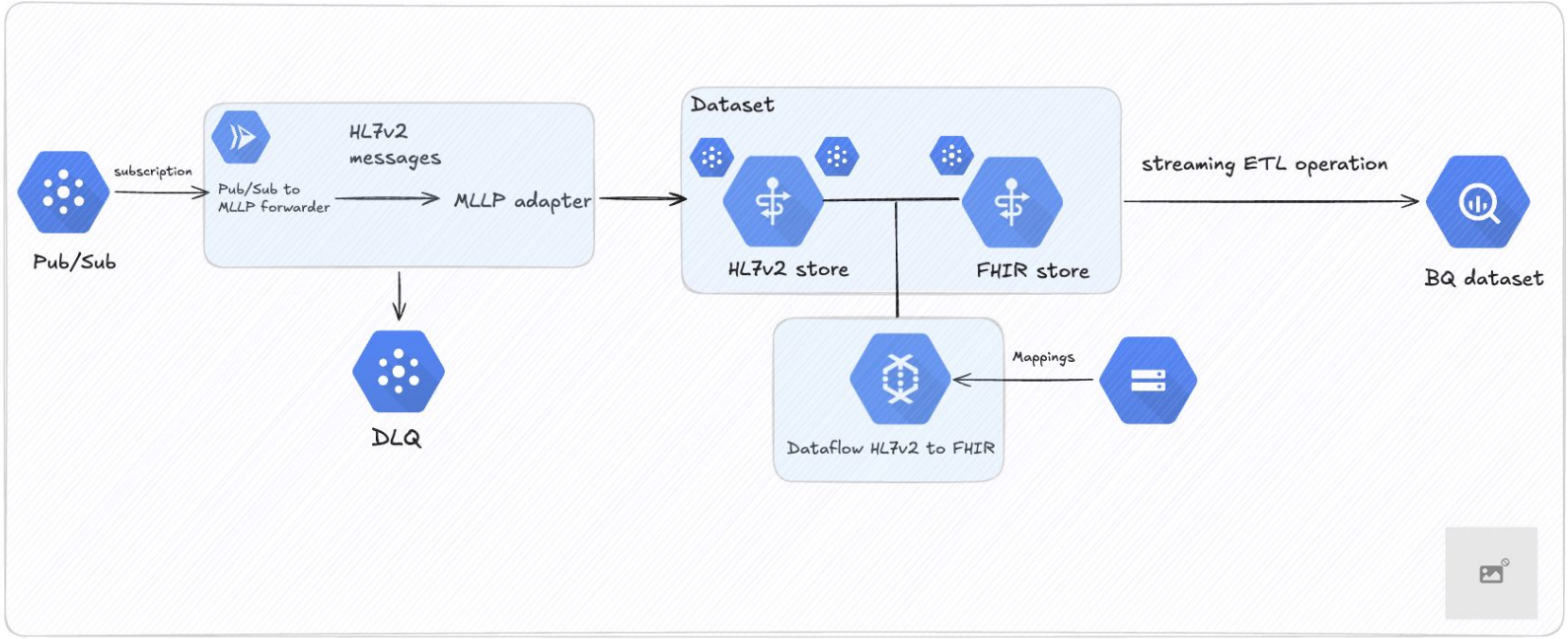
https://cloud.google.com/dataflow/docs/guides/templates/using-flex-templates

BigQuery

- Works great for LOTs
- Worth digging into Iceberg Storage Layer



DIAGRAM


Pub/Sub
subscription
Pub/Sub to
MLLP forwarder
HL7v2
messages
MLLP adapter
DLQ
Dataset
HL7v2 store
FHIR store
streaming ETL operation
BQ dataset
Mappings
Dataflow HL7v2 to FHIR

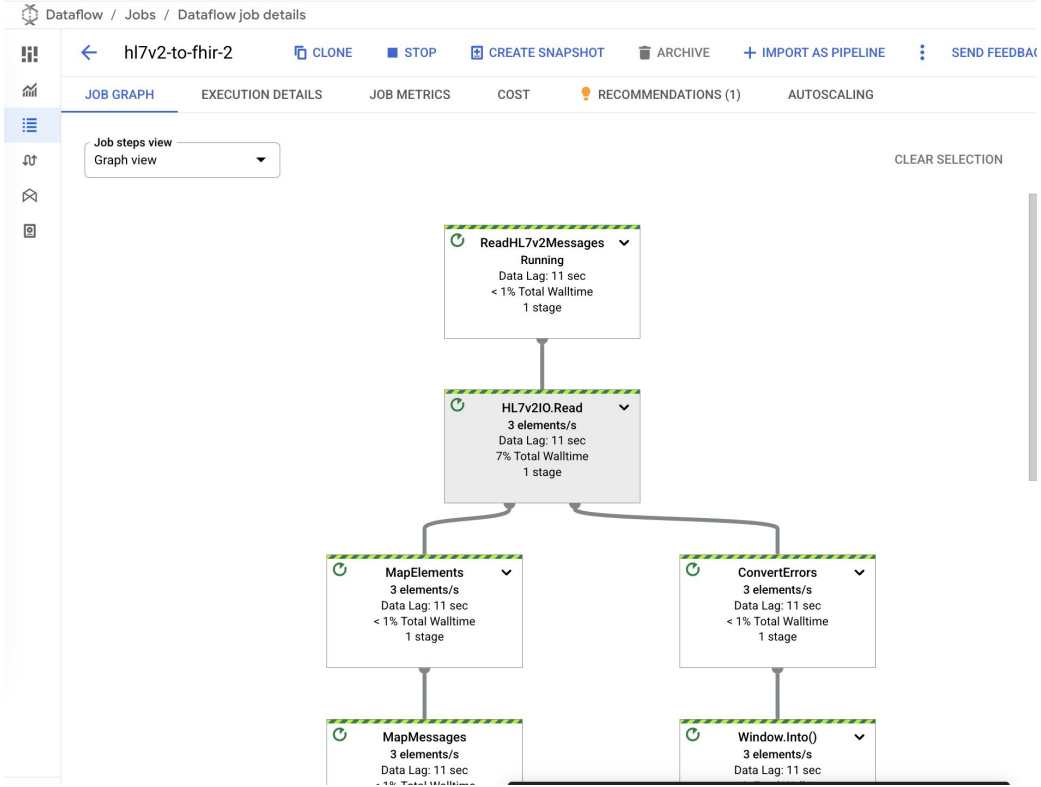
Deviating from the Standard

```
HL7 Application Error: failed to parse the HL7v2 data with the store's schema package: could not match data to field of ". [DG1]"
    could not match field repetitions of "DG1.4 [ST]"
        field instance 0 did not match components
            could not match data to "DG1.4 [ST]"
                could not match data to "DG1.4 [ST]"
                    too many sub-elements (2 sub-elems but only 0 sub-nodes)
in array index 2
in group member "DG1"
in group member "ADT_A08"
```

```
HL7 Application Error: failed to parse the HL7v2 data with the store's schema package: could not match data to field of ". [GT1]"
    could not match field repetitions of "GT1.49 [ST]"
        field instance 0 did not match components
            could not match data to "GT1.49 [ST]"
                could not match data to "GT1.49 [ST]"
                    too many sub-elements (2 sub-elems but only 0 sub-nodes)
in array index 0
in group member "GT1"
in group member "ADT_A08"
```

```
HL7 Application Error: failed to parse the HL7v2 data with the store's schema package: could not match data to field of ". [PV2]"
    could not match field repetitions of "PV2.3 [CWE]"
        field instance 0 did not match components
            could not match data to field of "PV2.3 [CWE]"
                could not match data to "CWE.2 [ST]"
                    too many sub-elements (2 sub-elems but only 0 sub-nodes)
in group member "PV2"
in group member "ADT_A08"
```

Dataflow


Dataflow / Jobs / Dataflow job details
hl7v2-to-fhir-2
CLONE
STOP
CREATE SNAPSHOT
ARCHIVE
IMPORT AS PIPELINE
SEND FEEDBAC
JOB GRAPH
EXECUTION DETAILS
JOB METRICS
COST
RECOMMENDATIONS (1)
AUTOSCALING
Job steps view
Graph view
CLEAR SELECTION
ReadHL7v2Messages
Running
Data Lag: 11 sec
< 1% Total Walltime
1 stage
HL7v2IO.Read
3 elements/s
Data Lag: 11 sec
7% Total Walltime
1 stage
MapElements
3 elements/s
Data Lag: 11 sec
< 1% Total Walltime
1 stage
ConvertErrors
3 elements/s
Data Lag: 11 sec
< 1% Total Walltime
1 stage
MapMessages
3 elements/s
Data Lag: 11 sec
Window.Into()
3 elements/s
Data Lag: 11 sec

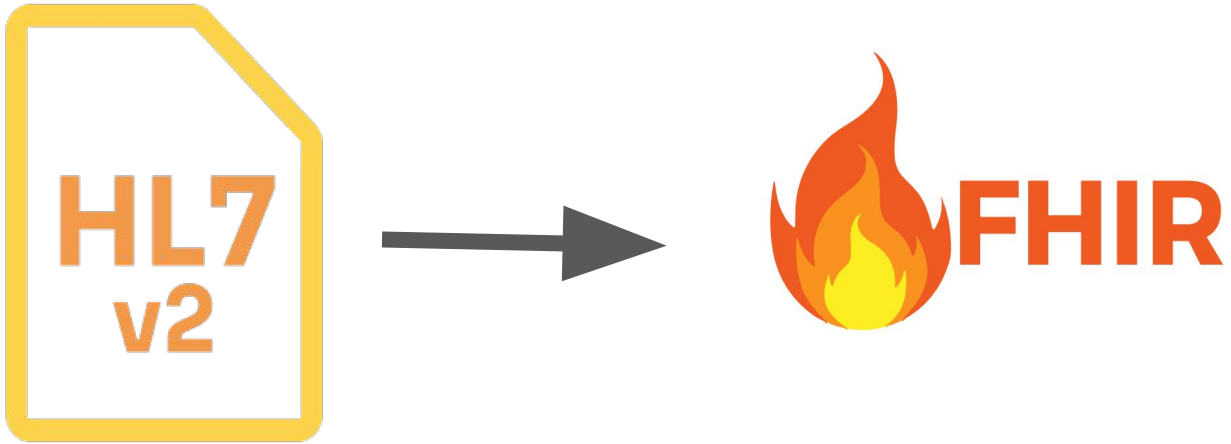
Future

Future Work

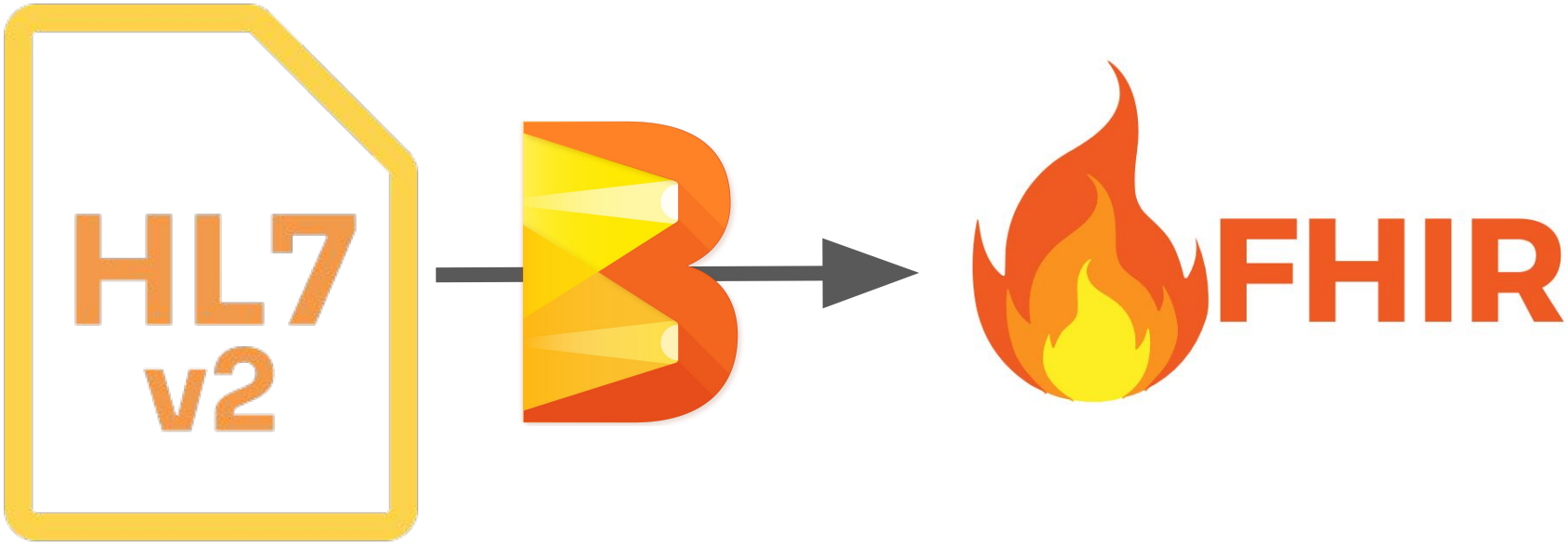
- Many of the mentioned components *could* be superfluous [or simplified]

Simplification

- HL7 -> FHIR is ‘just’ some data transformations
- Then ALL Beam?



Simplification



Beam MOAR!?

- Are there other common translations?
- EDIFact to ANSI X12
- https://en.wikipedia.org/wiki/X12_EDIFACT_Mapping

Beam

- What all is Beam in the business of?
  - **HOW**
  - [Maybe] WHAT?
    - Business Logic != Compute
    - Business Logic [ex: translations] have immense value
    - Some Data Transformation logic is NOT individual company/business-specific
      - GENERAL/COMMON TRANSFORMS!
        - CORE industry-specific translations?

Thanks!
Questions?

Austin Bennett

Thanks!
QUESTIONS?

austin@apache.org

austin.bennett@azra-ai.com

linkedin.com/in/austinbennett

BEAM SUMMIT NYC 2025