



# TOWARDS THE INTERNET OF FAIR DATA AND SERVICES

EASMA workshop - Brussels, May 23, 2019

Luiz Bonino  
[luiz.bonino@go-fair.org](mailto:luiz.bonino@go-fair.org)

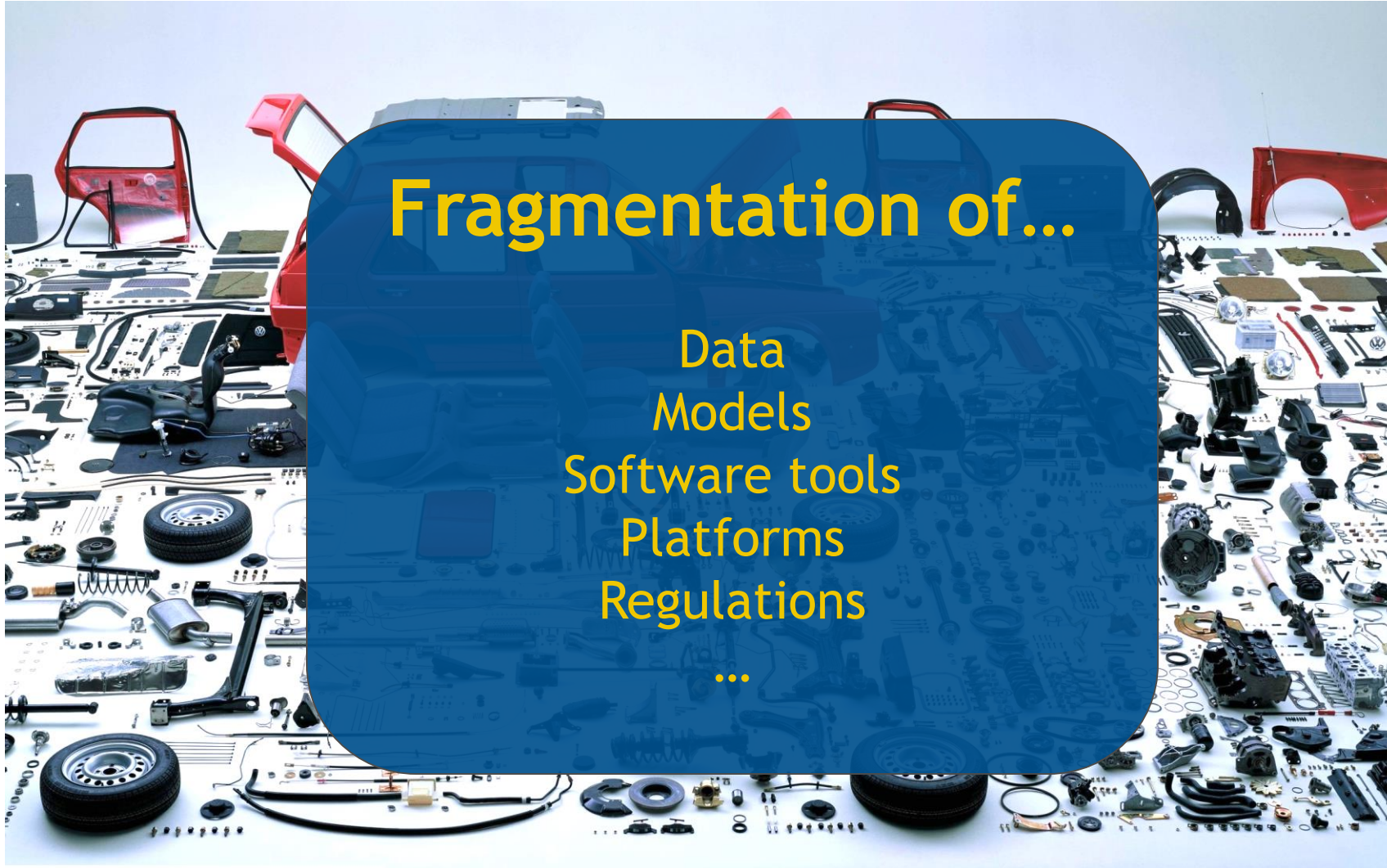
# THE DATA PROBLEM

# THE UNDERLYING PROBLEM





# THE UNDERLYING PROBLEM



# ENTERPRISE REALITY – HETEROGENEITY - CHALLENGES



How can I get to know what is available?

Where to get?

How to get?

How to integrate?

How to streamline it (re)use?



pwc

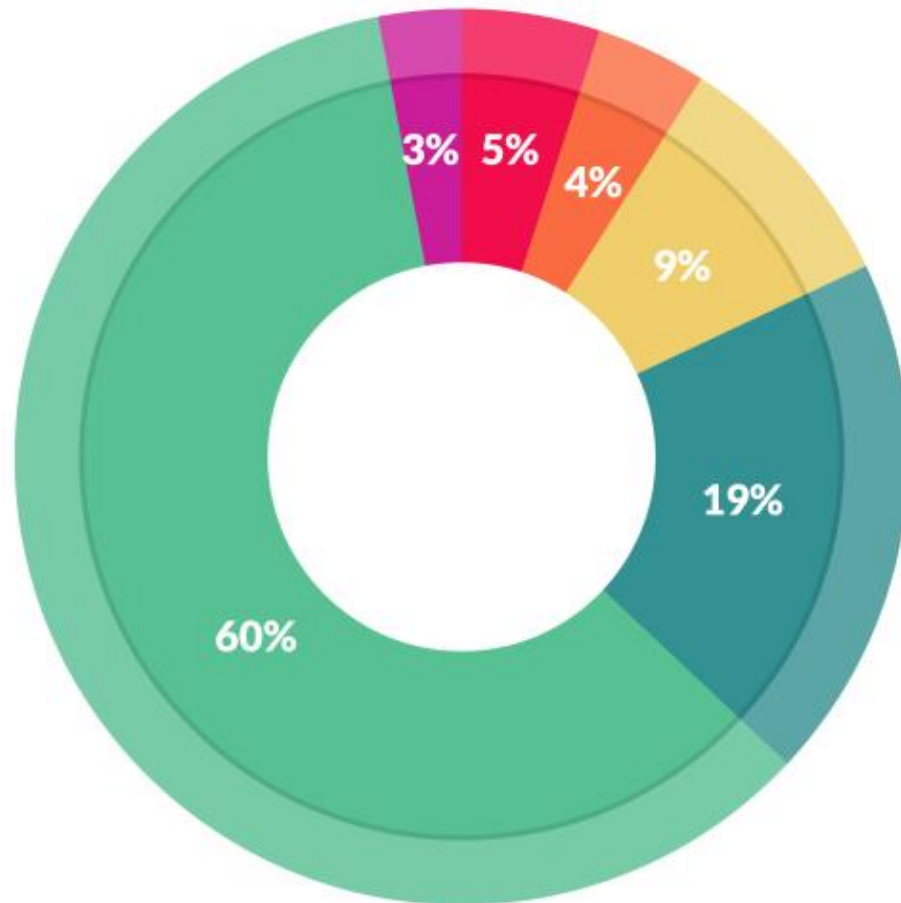
{ REST }



Integrated Enterprise



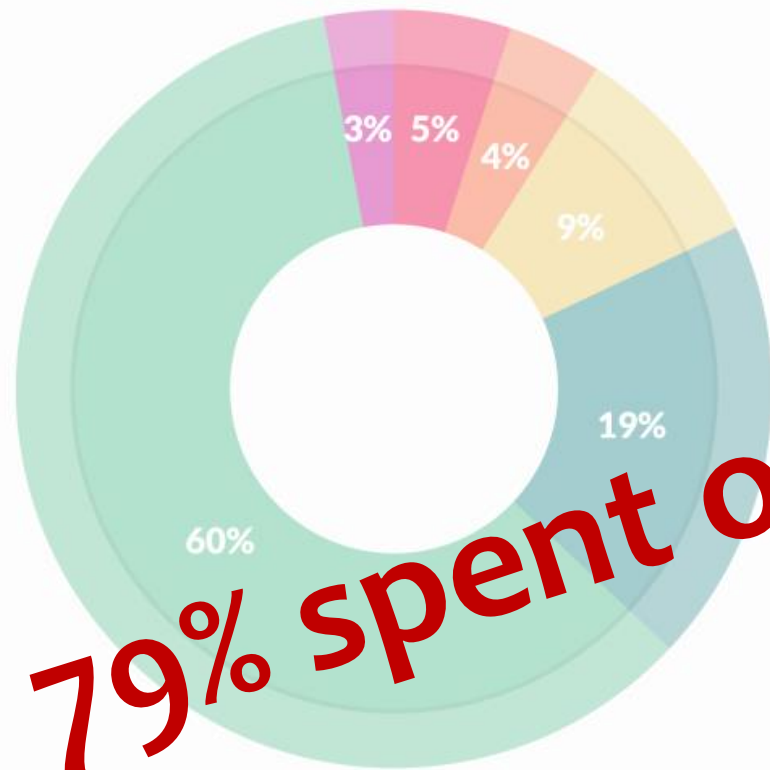
# DATA EXPERT EFFORT



## What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Source: Data Science Report 2016, CrowdFlower, 2016: [http://visit.crowdfunder.com/rs/416-ZBE-142/images/CrowdFlower\\_DataScienceReport\\_2016.pdf](http://visit.crowdfunder.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf)



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

**79% spent on data preparation**

Source: Data Science Report 2016, CrowdFlower, 2016: [http://visit.crowdfunder.com/rs/416-ZBE-142/images/CrowdFlower\\_DataScienceReport\\_2016.pdf](http://visit.crowdfunder.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf)

# ENTERPRISE REALITY – HETEROGENEITY – SOLUTION?

F  
I  
N  
D  
A  
B  
L  
E



# ENTERPRISE REALITY – HETEROGENEITY – SOLUTION?

F  
I  
N  
D  
A  
B  
L  
E

A  
C  
C  
E  
S  
S  
I  
B  
L  
E

# ENTERPRISE REALITY – HETEROGENEITY – SOLUTION?

F  
I  
N  
D  
A  
B  
L  
E

A  
C  
C  
E  
S  
S  
I  
B  
L  
E

I  
N  
T  
E  
R  
O  
P  
E  
R  
A  
B  
L  
E

# ENTERPRISE REALITY – HETEROGENEITY – SOLUTION?

F	A	I	R
I	C	N	E
N	C	T	U
D	E	E	S
A	S	R	A
B	S	O	B
L	I	P	L
E	B	E	E
	L	R	
	E	A	
		B	
		L	
		E	

# FAIR

I  
N  
D  
A  
B  
L  
E

C  
C  
E  
S  
S  
I  
B  
L  
E

N  
T  
E  
R  
O  
P  
E  
R  
A  
B  
L  
E

E  
U  
S  
A  
B  
L  
E



# FAIR PRINCIPLES

## Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier;
- F2. data are described with rich metadata;
- F3. metadata clearly and explicitly include the identifier of the data it describes;
- F4. (meta)data are registered or indexed in a searchable resource;

## Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles;
- I3. (meta)data include qualified references to other (meta)data;

## Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol;
  - A1.1 the protocol is open, free, and universally implementable;
  - A1.2. the protocol allows for an authentication and authorization procedure, where necessary;
- A2. metadata are accessible, even when the data are no longer available;

## Reusable:

- R1. (meta)data are richly described with a plurality of accurate and relevant attributes;
  - R1.1. (meta)data are released with a clear and accessible data usage license;
  - R1.2. (meta)data are associated with detailed provenance;
  - R1.3. (meta)data meet domain-relevant community standards;

<https://www.nature.com/articles/sdata201618>

# PERSONAL HEALTH TRAIN

# MAIN ELEMENTS



## Data Station:

- Publishes FAIR metadata
- Allows train to interact with data



## Data Gateway:

- Provides access and control to the data owner to its data located in different Data Stations

## Train:

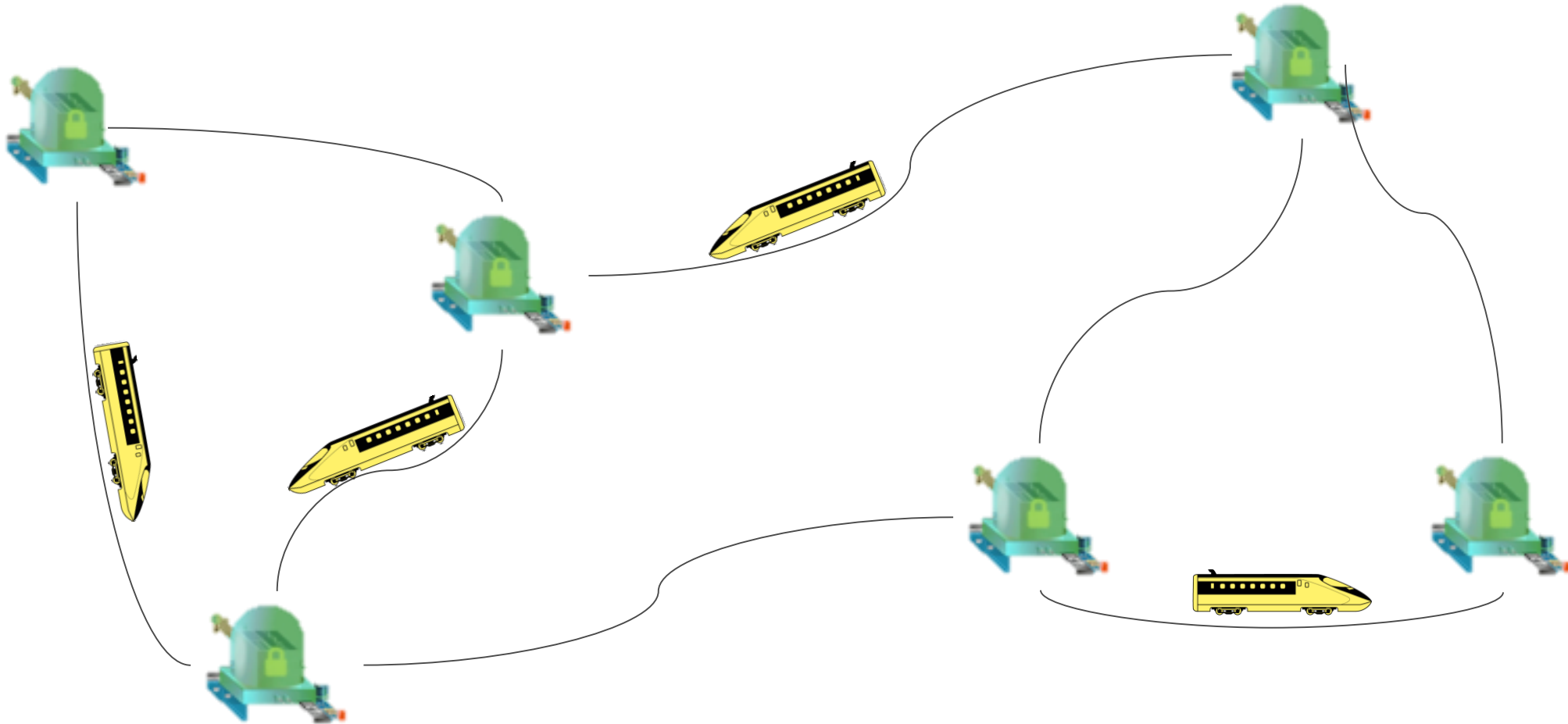
- Interacts with data (process, integrate, analyze, move, ...)



## Tracking System:

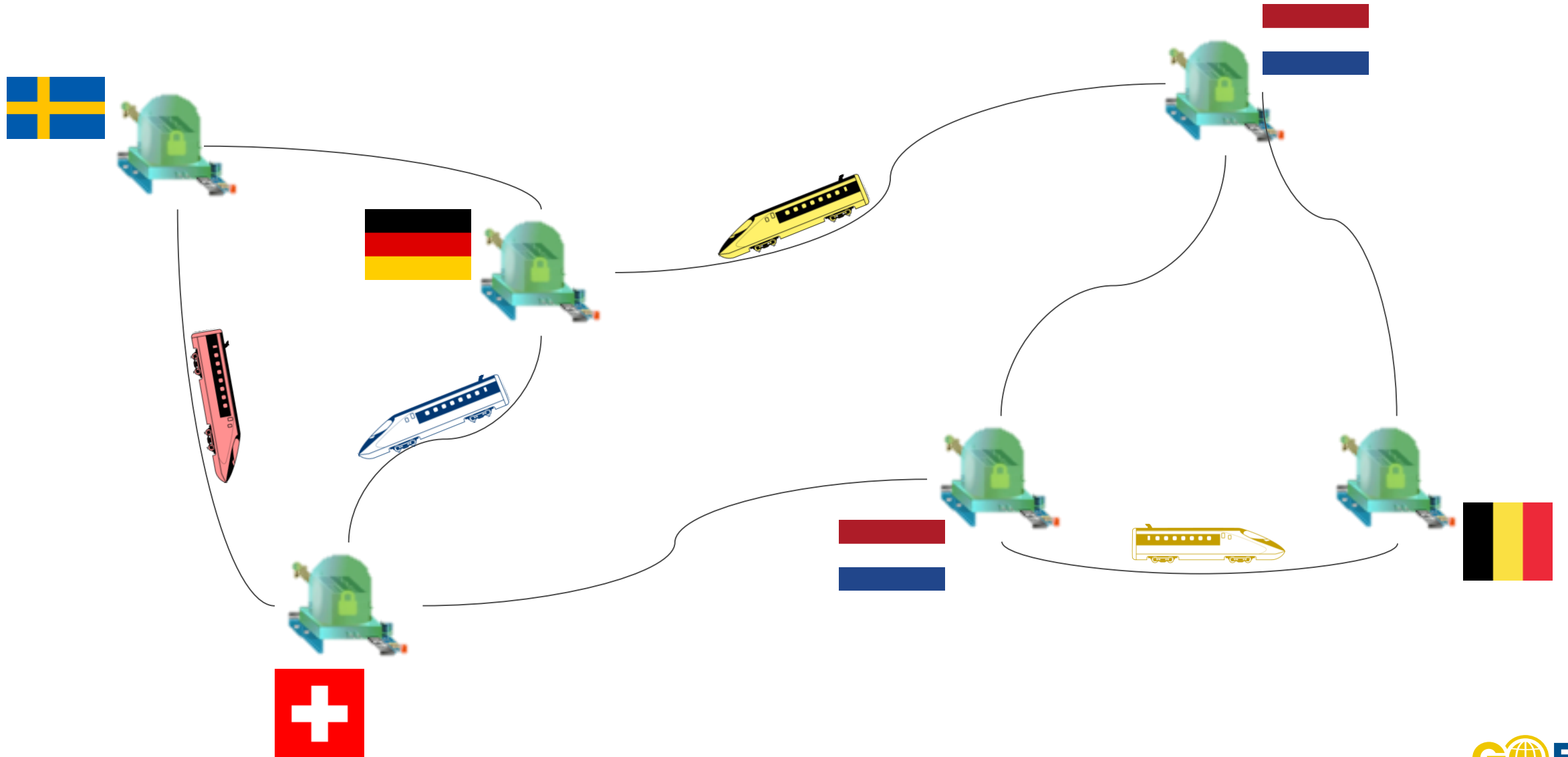
- The validation, routing and transport infrastructure

# ALGORITHMS TO DATA - TRAINS TO STATIONS

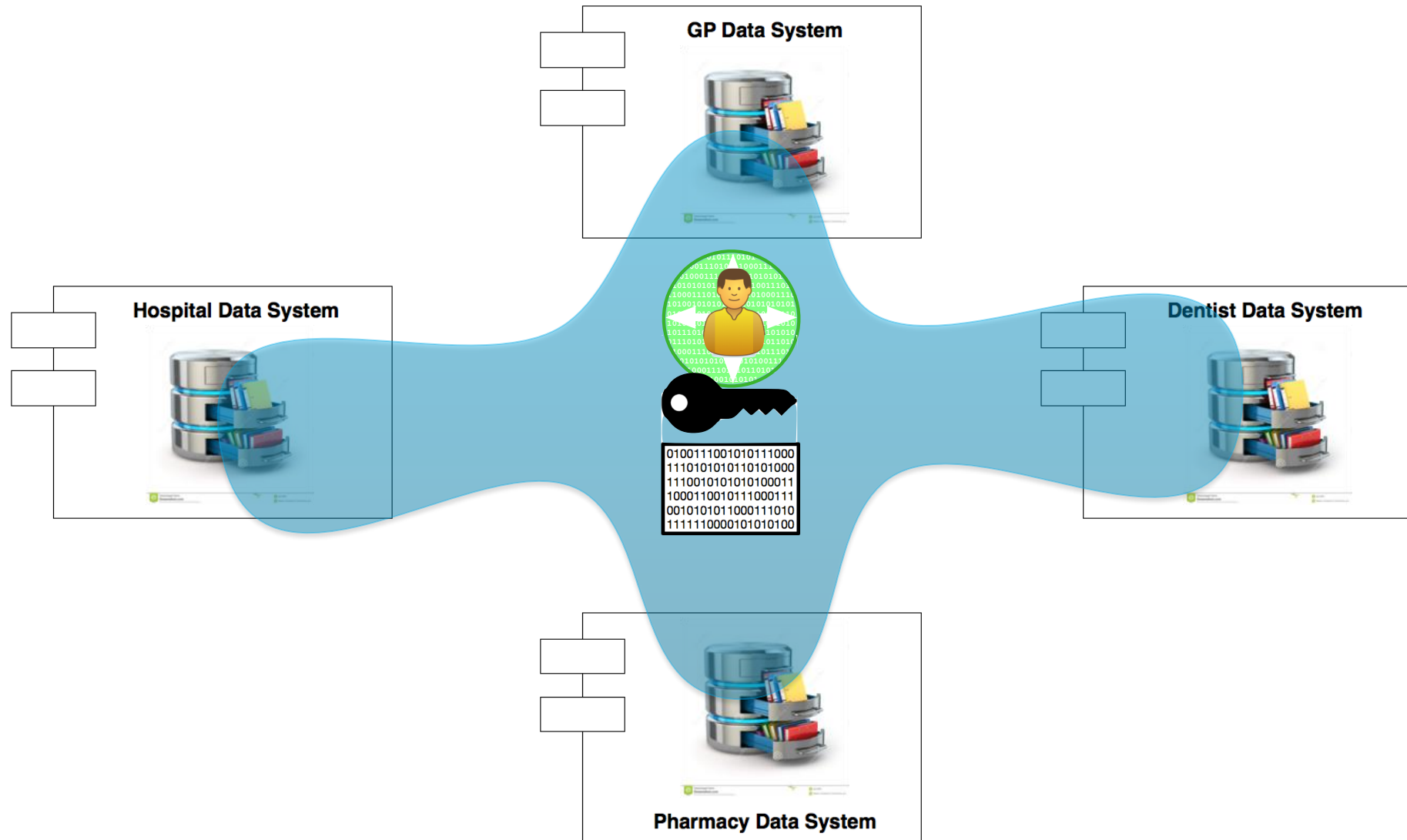




# ALGORITHMS TO DATA - TRAINS TO STATIONS



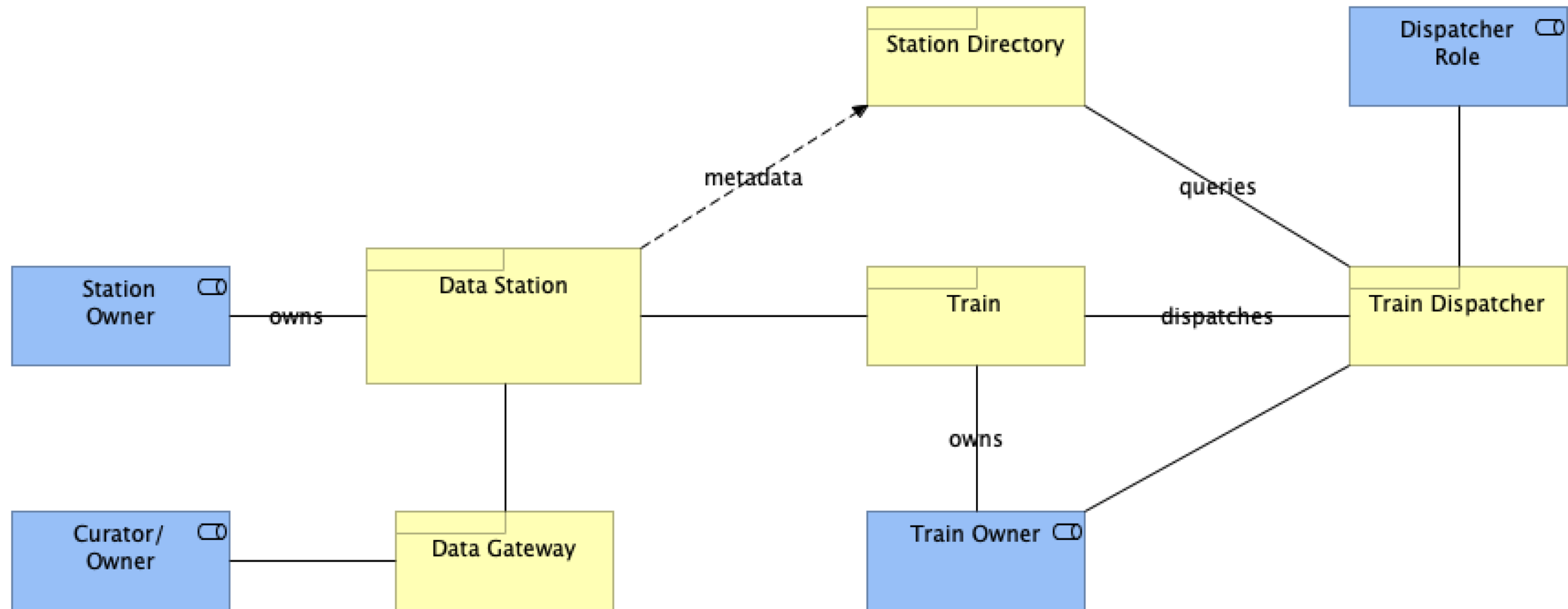
# DISTRIBUTED PERSONAL/INDIVIDUAL DATA



# PERSONAL HEALTH TRAIN KEY CHARACTERISTICS

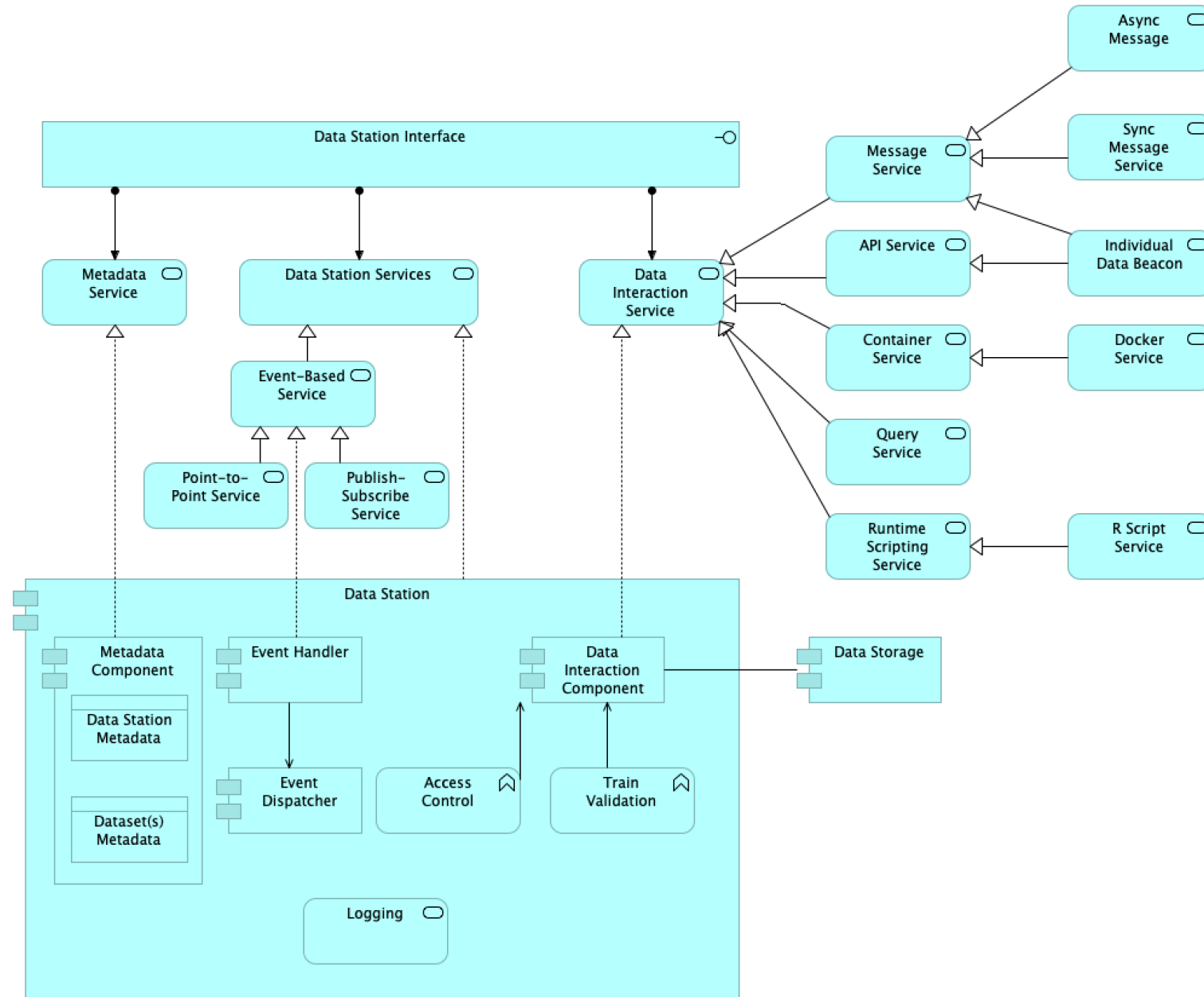
- Large-scale: potentially all citizens, health providers and associated entities;
- Internet/Web-like open environment;
- Heterogeneous - data, implementations, analysis;
- Multi-purpose - research, care, administrative, ...;
- ...

# IFDS - MAIN COMPONENTS

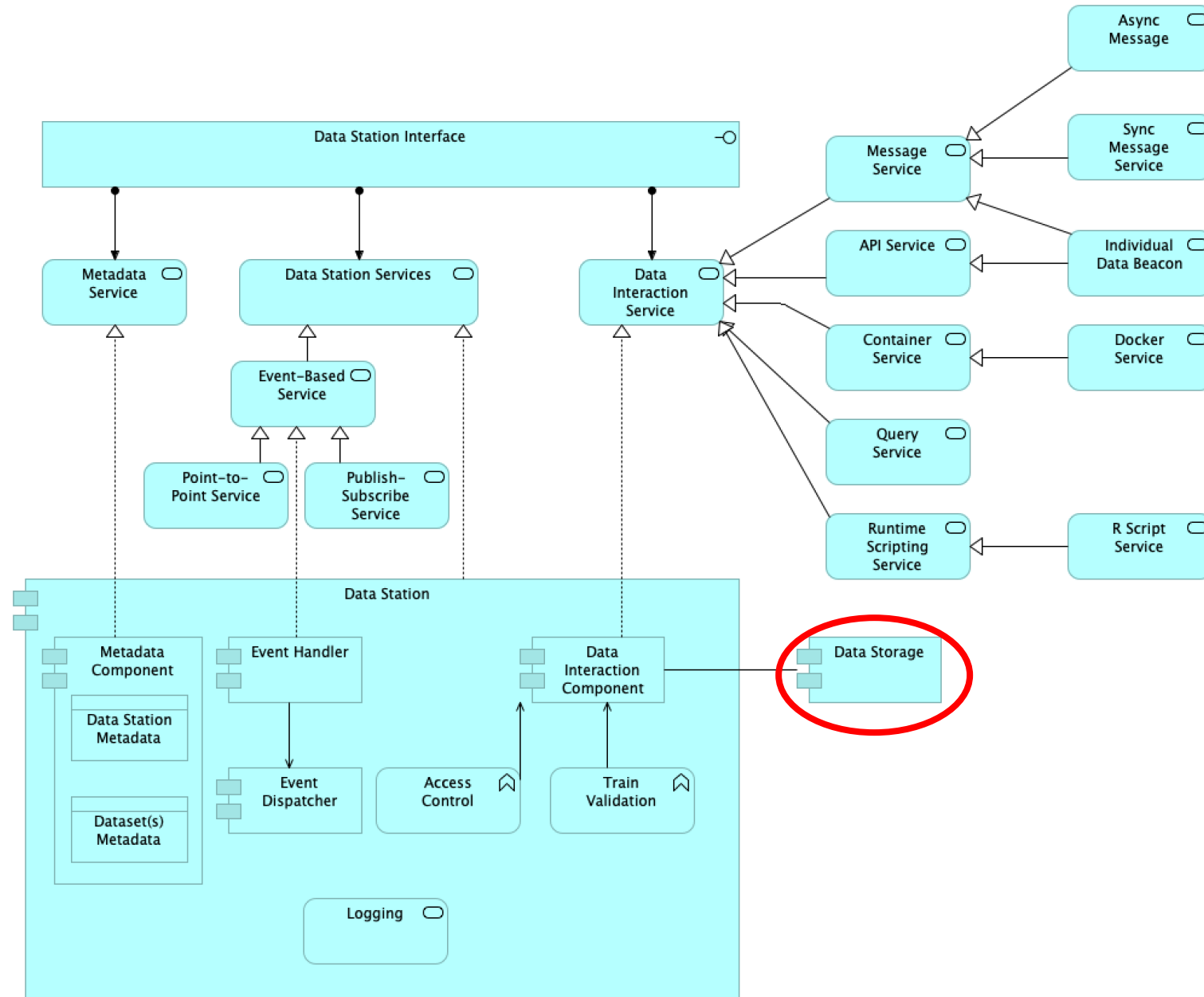




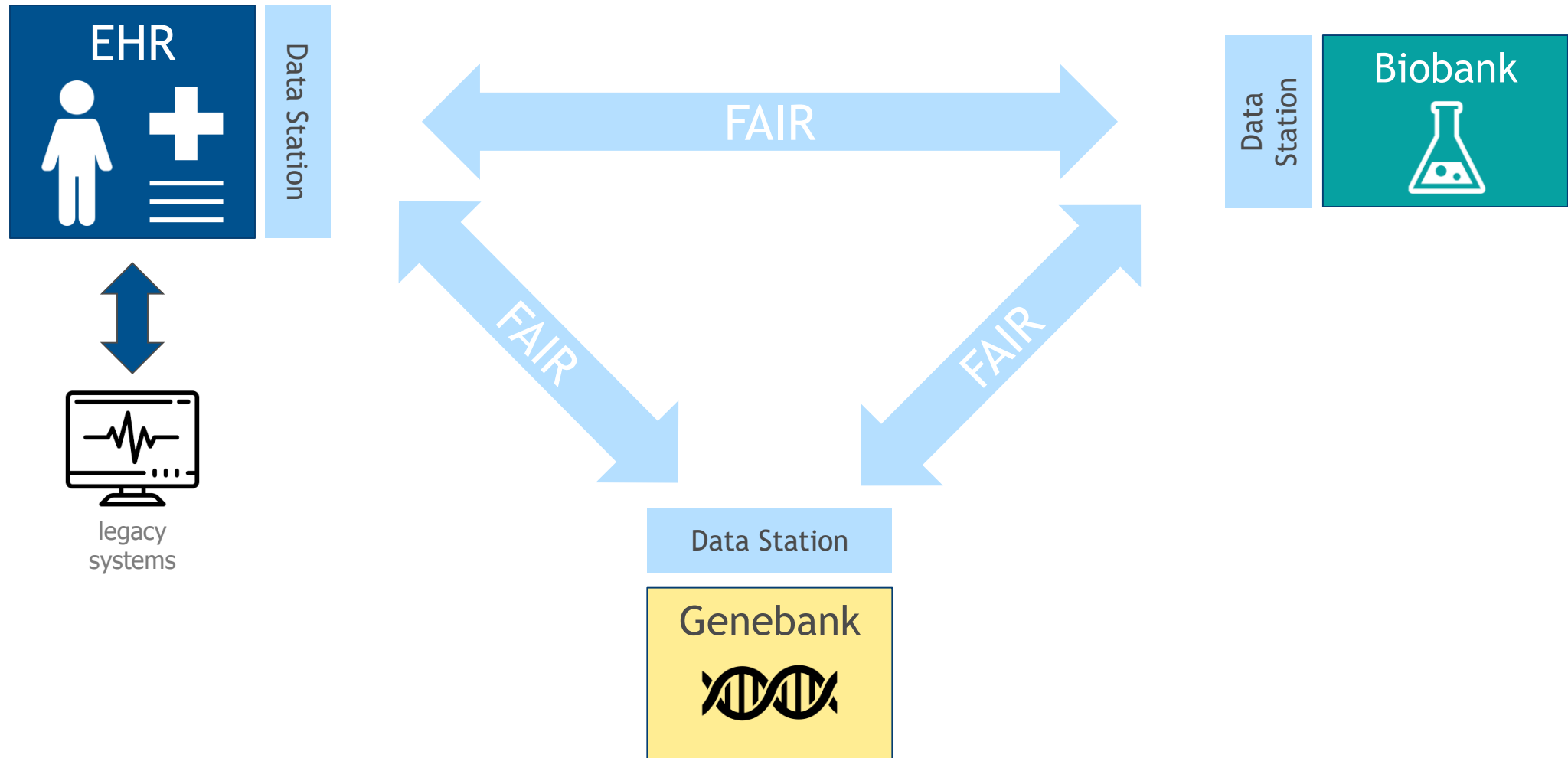
# IFDS - DATA STATION



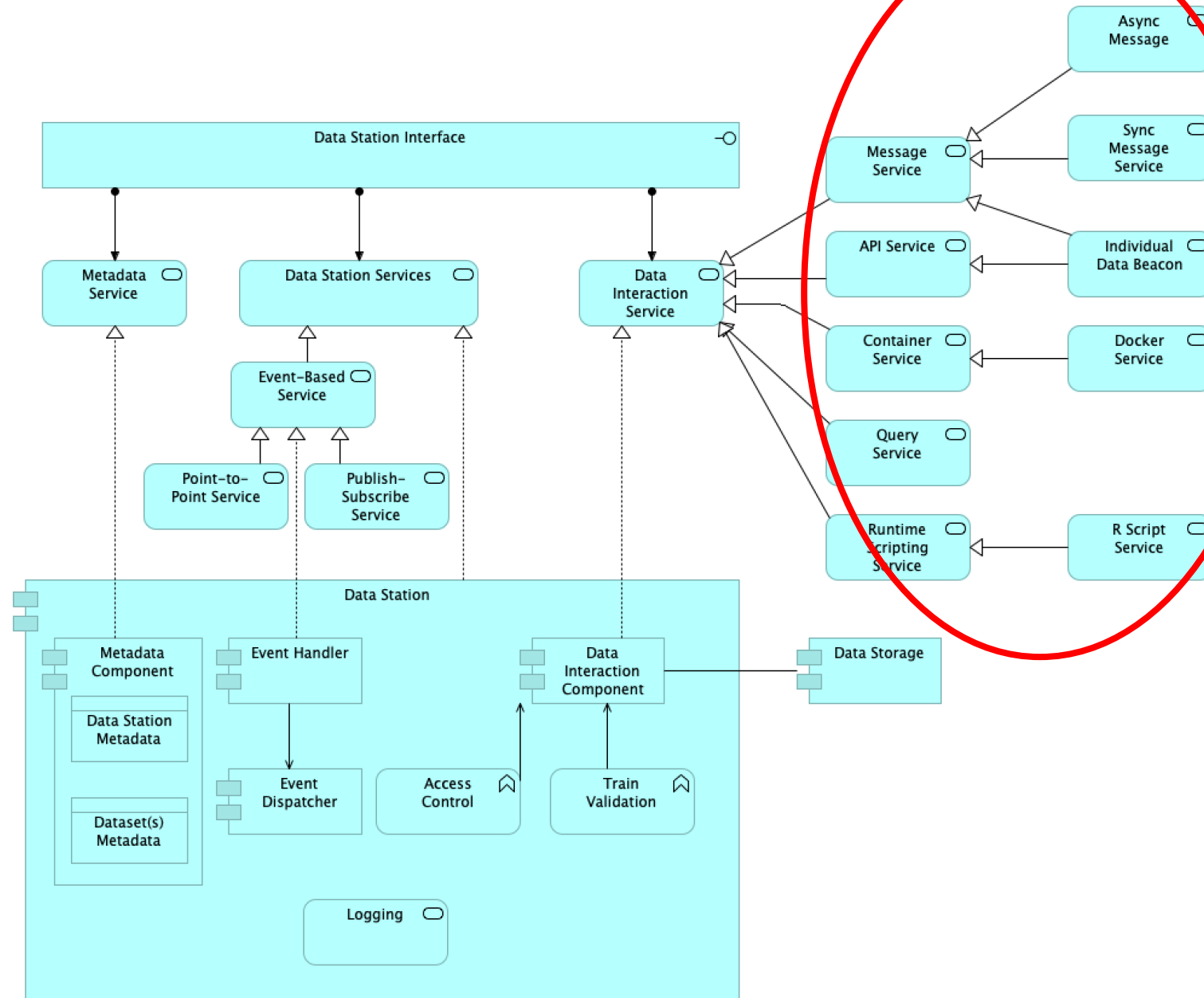
# IFDS - DATA STATION



# IFDS - DATA STATION - EXTENDING EXISTING SOLUTIONS

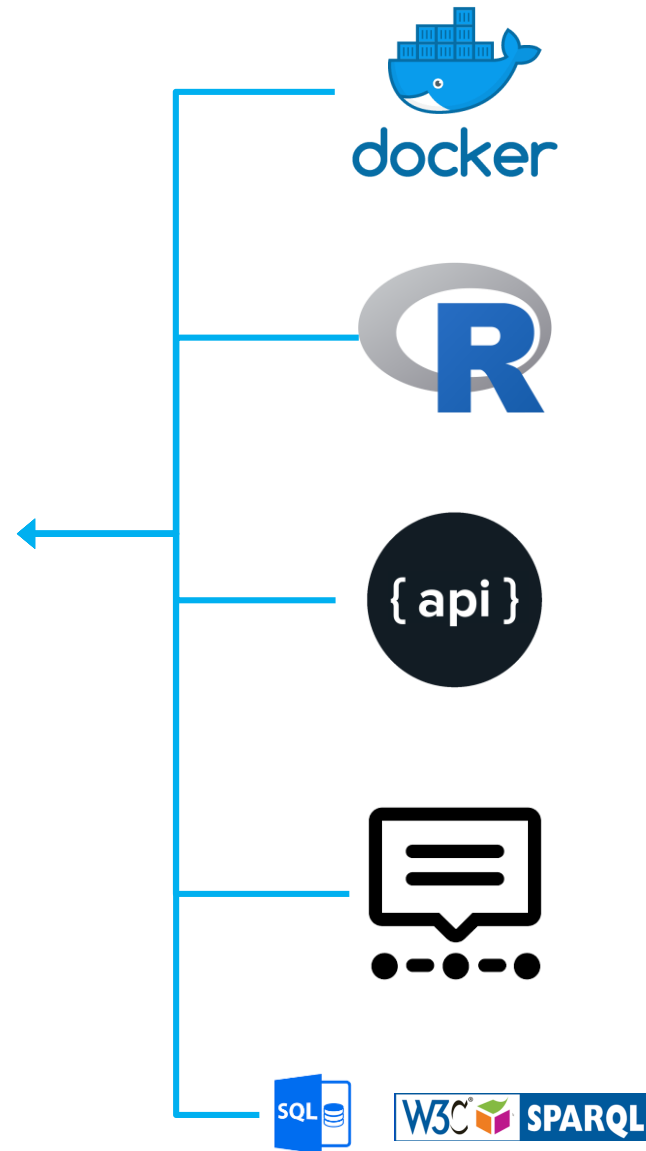


# IFDS - DATA STATION

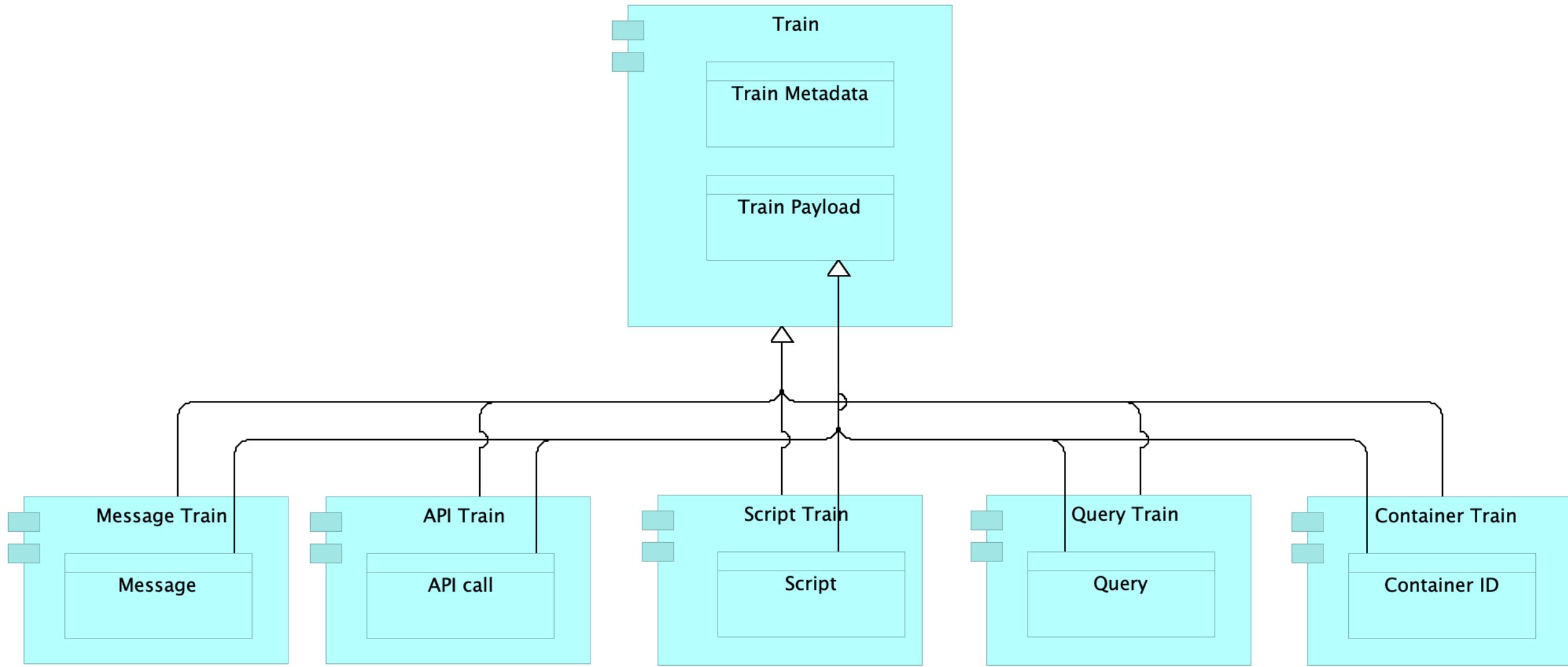




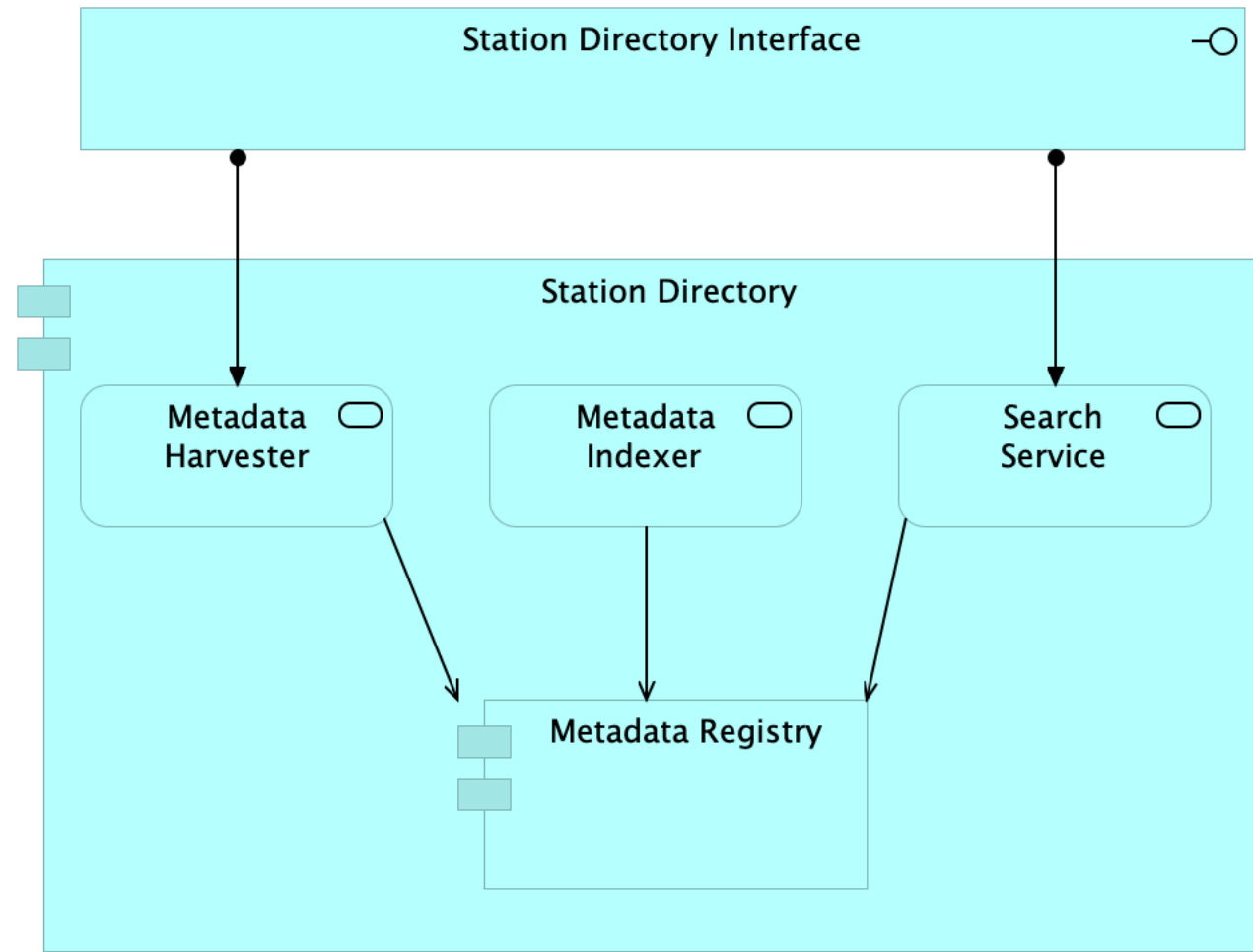
# IFDS - DATA STATION



# IFDS - DATA TRAIN



# IFDS - STATION DIRECTORY



THANK YOU!!!



## Luiz Bonino

International Technology Coordinator – GO FAIR  
Associate Professor BioSemantics – LUMC

E-mail: [luiz.bonino@go-fair.org](mailto:luiz.bonino@go-fair.org)

Skype: [luizolavobonino](https://www.skype.com/people/luizolavobonino)

Web: [www.go-fair.org](http://www.go-fair.org)