SYNDERAI
Synthetic Data Examples – Realistic – using AI

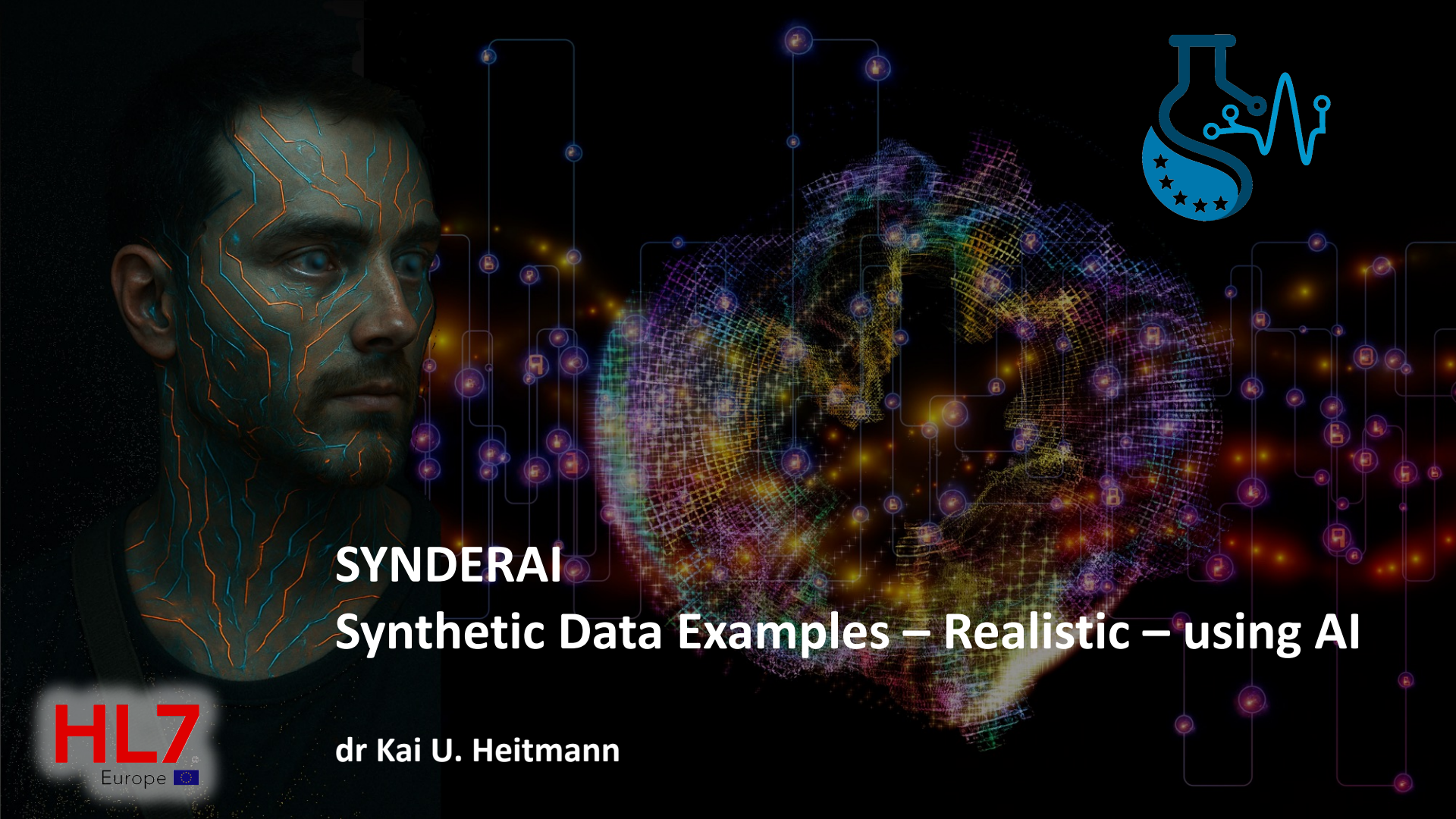
dr Kai U. Heitmann

Share EHR Xchange Format

Shaping the Future of FHIR® in Europe

HL7® Europe

Working Group Meeting
1-5 December 2025
Cologne, Germany



SYNDERAI

Synthetic Data Examples – Realistic – using AI

dr Kai U. Heitmann

HL7 Europe

SYNDERAI

- **Synthetic Data: Examples – Realistic – using AI (SYNDERAI)** – pronounced **/ˈsɪndəraɪ/**
- **xShare Project** Work Package
  - The xSHARE toolbox (D3.3) arranges tools that support proper implementation of the European EEHRxF.

SYNDERAI

- Synthetic Example Data is seen in
  - **testing** and **validation** (e.g. industry proofs, connect-a-thons etc.)
  - as well as in **education** and further **implementation support** for vendors

Share EHR Xchange Format

HL7 Europe

SYNDERAI Design

- SYNDERAI datasets are designed to
  - **represent realistic clinical scenarios**, including medications, allergies, problems, encounters, vital signs, depending on the covered use case,
  - **be conformant to HL7 FHIR Implementation Guides**, including IPS, EU Laboratory Report, and Hospital Discharge Summary,
  - **use realistic but not real patient data**, ensuring safety in both development and demonstration environments.

SYNDERAI Design

- Within xShare, SYNDERAI synthetic data is
  - used in architecture testbeds for **download, share, and visualize flows**, as seen at vi7eti.net,
  - prepared to support **IHE Connectathon/Plugathon test cases**
  - embedded in **documentation and walk-throughs** as examples of valid HL7 FHIR structures and content.

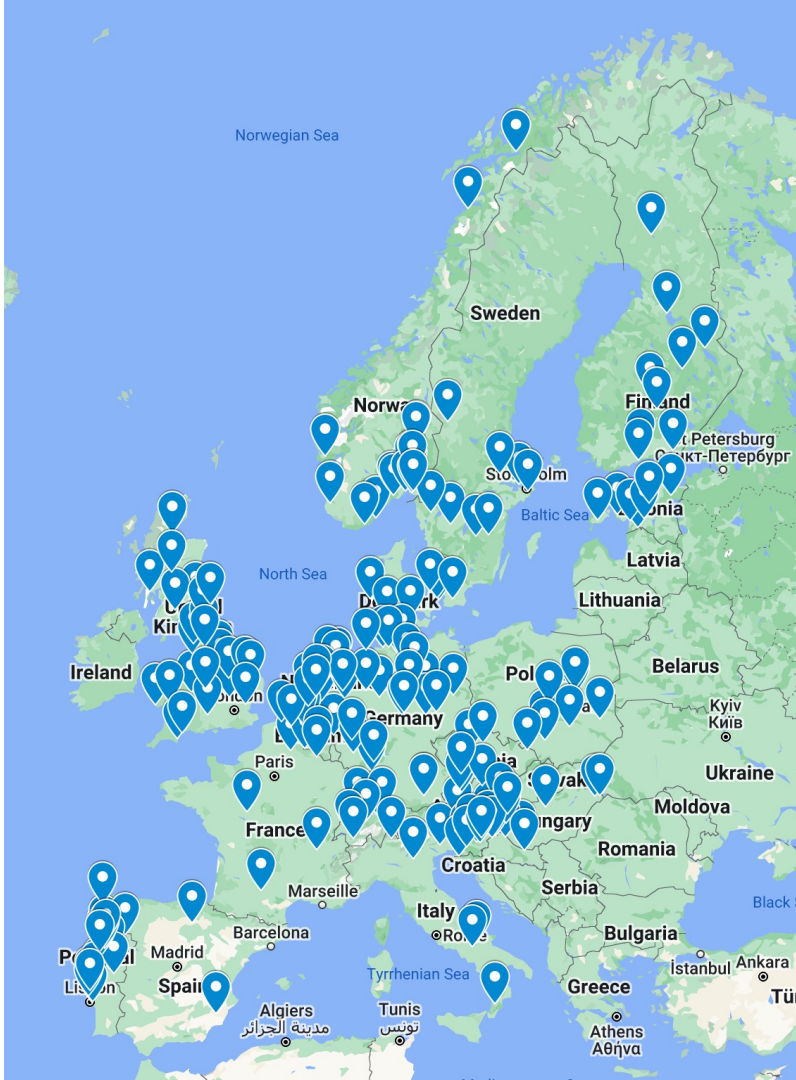
SYNDERAI Design

- This data enables xShare Adoption Sites and developers to
  - run the **Yellow Button tools** without privacy concerns
  - **simulate end-to-end workflows** with repeatable, traceable data
  - **demonstrate compliance** with technical and legal requirements.


CLICK

European Synthetic data

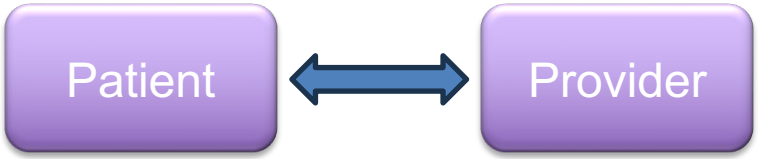
- Granular facts for Synthetic Example Data
  - …are subject to “real” medical background
  - …as if from or for “real” care
  - …but “invented” matching patient demographics
- European Population, Settings, Supplies, Locations, etc.

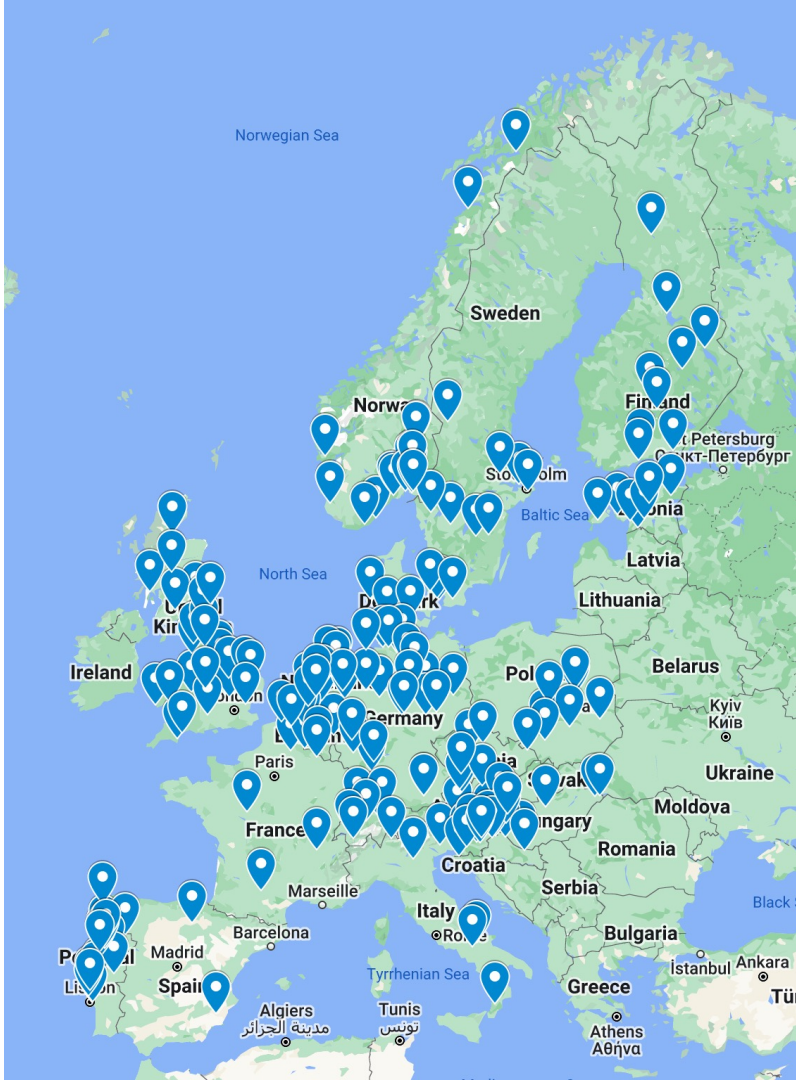


HL7 Europe

European Synthetic data

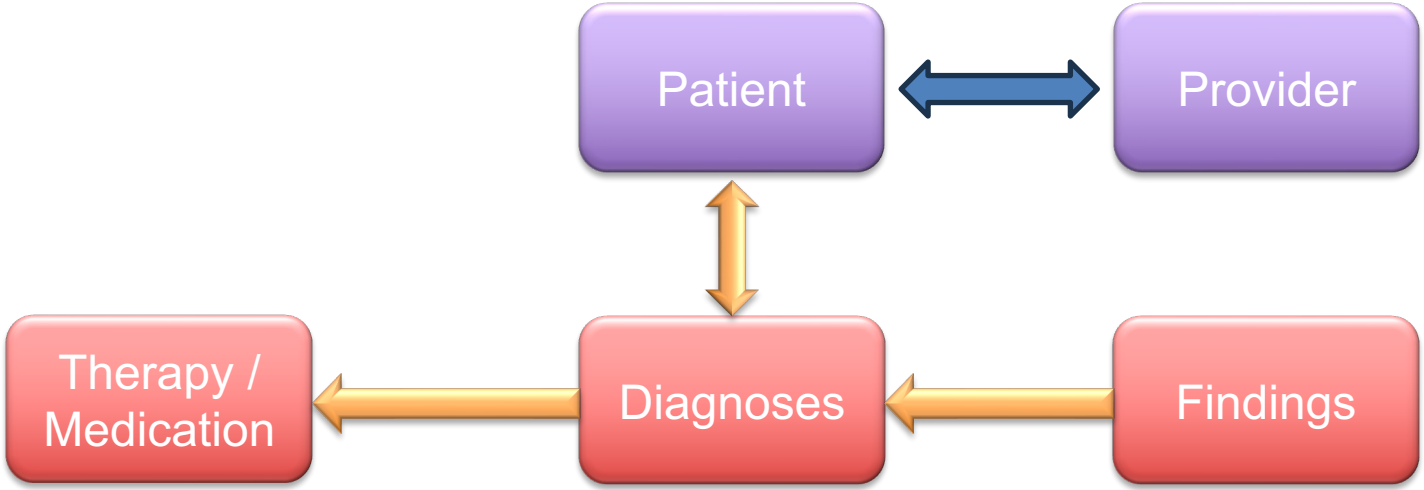
- SYNDERAI: using several sources of generated data, amalgamating it with additional localized data (for geo-location of the synthetic patients and providers)
- Geo-proximity





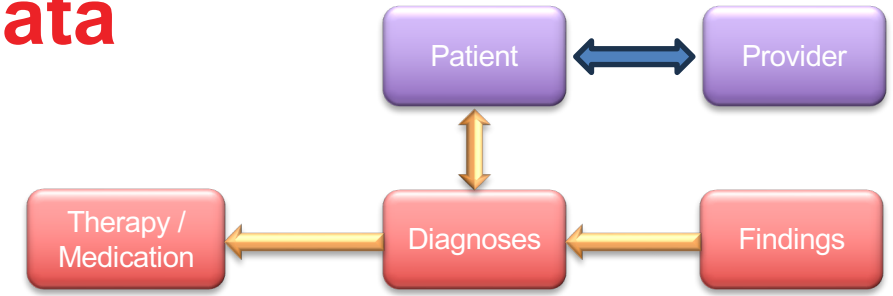
HL7 Europe

European Synthetic data



- Demographic stratification, mapping to a clinical “story” or even a Persona

European Synthetic data

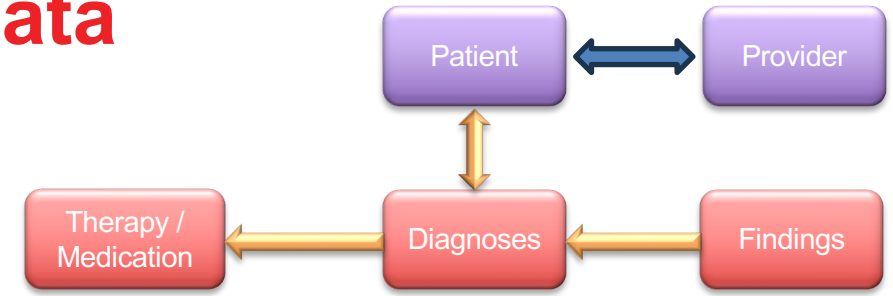


- Laboratory Report was first
- All others followed
- Example of using AI: generate the stratified, clinically expected, realistic reference ranges for lab values

Chemistry

Test	18 Nov 2024	Reference Range	Unit
Glucose [Mass/volume] in Blood	71.0	70 - 99	mg/dL
Urea nitrogen [Mass/volume] in Blood	16.2	7 - 26	mg/dL
Creatinine [Mass/volume] in Blood	**3.0 H**	0.9 - 1.3	mg/dL
Calcium [Mass/volume] in Blood	9.3	8.5 - 10.5	mg/dL
Sodium [Moles/volume] in Blood	143.3	135 - 145	mmol/L

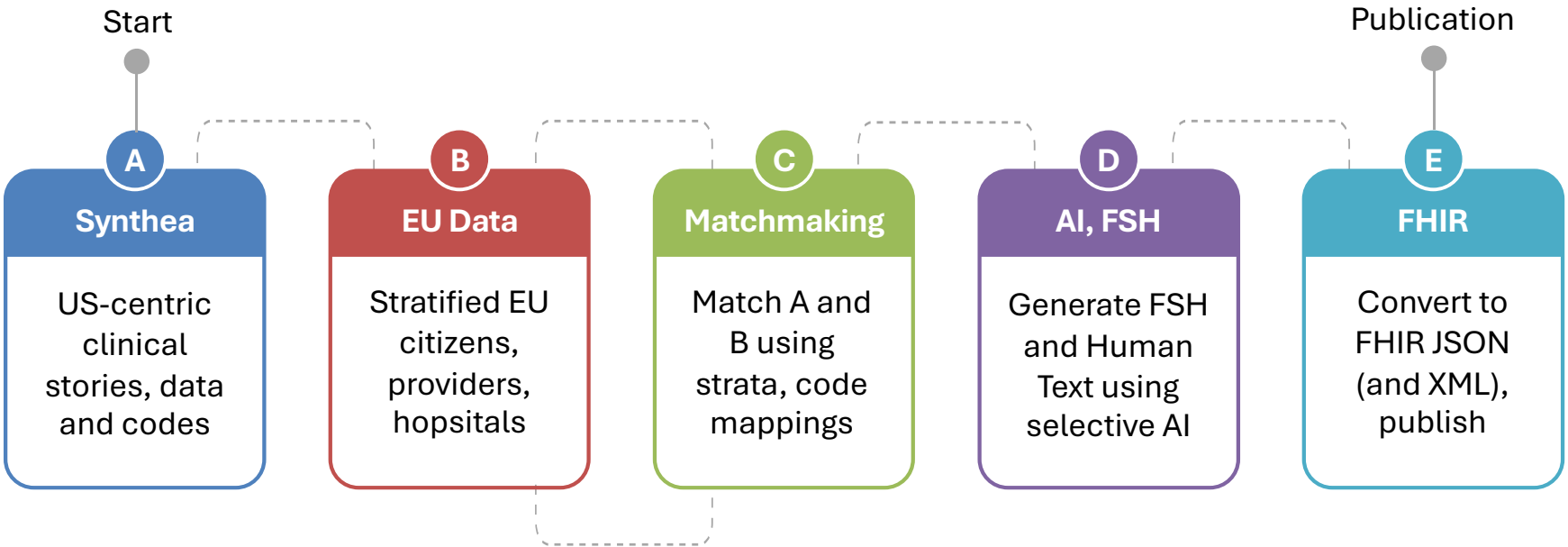
European Synthetic data



- Ongoing, further developments
  - More variations in clinical “stories”
  - Use Cases: Laboratory Report, Hospital Discharge Report, European (+ International) Patient Summary
  - Personas with consistent reports across Use Cases
  - Adding Imaging Report Synthetic Data in near future
- Combined with Example **Visualization**, see vi7eti.net
- Also on **zenodo** https://zenodo.org/records/16792934

Some details

SYNDERAI Synthetic Data Generation



SYNDERAI Synthetic Data Generation

A • Synthea (Mitre)  SyntheticMass MITRE

- US American set of synthetic data / clinical stories
- US coding systems such as RX Norm

B • European Citizen Dataset

- with addresses, contacts, closest primary provider / hospital
- Stratified with at least gender, age

C • Match Making

- Find matches of EU Citizen with Synthea set with similar age, same gender etc.

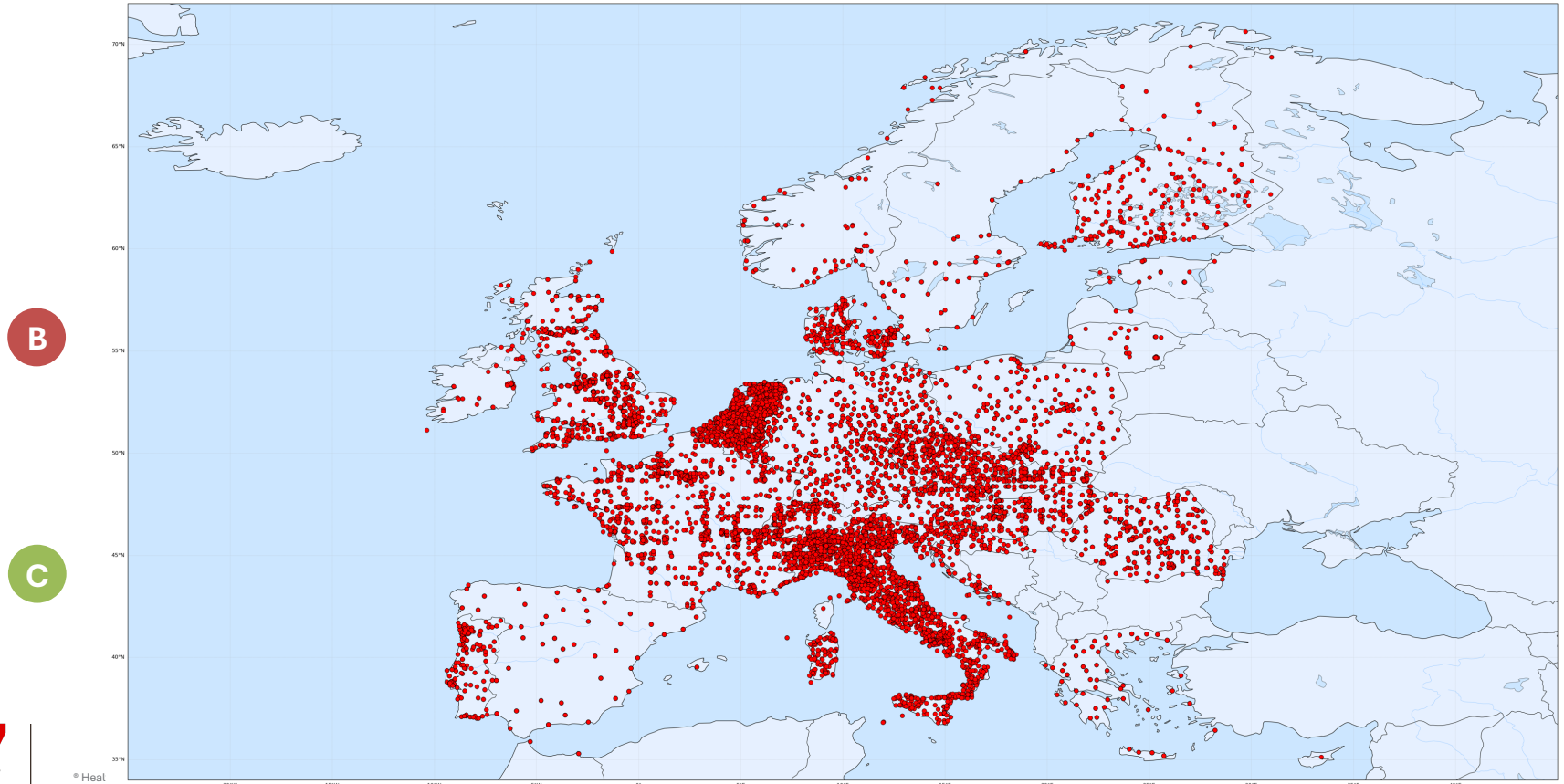
EU Citizen Dataset Examples

language	given	family	gender	birthdate	age	eci	street1	city	postcode	country
de	Pascal	Schönauer	male	1988-11-01	36	9198-808258-1	Ricardo-Feigl-Platz 54	Gerasdorf bei Wien	2201	Austria
de	Manuel	Steinwender	male	1947-03-08	78	5484-761690-8	Höflerweg 1/2	Ternitz	2630	Austria
de	Maya	Strauss	female	1947-03-13	78	9196-735357-3	Anja-Wechselberger-Ring 0	Oberndorf bei Salzburg	5110	Austria
de	Ela	Günther	female	1996-04-28	29	6941-486075-4	Florentina-Suppan-Straße 4/2	Waidhofen an der Ybbs	3340	Austria
de	Bianca	Schenk	female	1989-09-01	36	5269-211464-9	Diana-Pollak-Gasse 59	Bruck an der Mur	8600	Austria
de	Mira	Zöhrer	female	1996-09-10	29	8888-018881-3	Valentina-Steinböck-Platz 704	Frauenkirchen	7132	Austria
de	Beatrice	Bichler	female	1959-05-19	66	6185-234899-6	Theresa-Scheiber-Platz 35	Oberwälz	8740	Austria
de	Leontina	Sulzer	female	1963-09-29	62	9450-084540-2	Elvira-Haidinger-Gasse 281	Hartberg	8230	Austria
de	Elias	Gmeiner	male	1958-12-05	66	7939-365590-1	Brunnerstr. 34	Litschau	3874	Austria
de	Paul	Mitteregger	male	1964-09-08	61	8137-524402-7	Weningerstr. 6	Bad Ischl	4820	Austria
de	Elisabeth	Pfeiffer	female	1996-11-17	28	7613-597304-8	Anastasia-Reich-Platz 35	Wolkersdorf	2120	Austria
de	Nicolai	Bachmann	male	1944-09-19	81	3045-866369-4	Haumerweg 18	Neunkirchen	2620	Austria
de	Mark	Stoiber	male	1955-10-10	69	2158-820460-8	Fernando-Danninger-Gasse 996	Pulkau	3741	Austria
de	Johann	Fritsch	male	1968-05-26	57	9923-954485-3	Schwaigergasse 23	Schrattenthal	2020	Austria
de	Nick	Fellinger	male	1945-11-07	79	9234-036216-6	Schüllerweg 3	Dornbirn	6850	Austria
de	Simon	Fankhauser	male	2006-05-25	19	8972-442123-8	Moritz-Neuner-Platz 822	Marchegg	2294	Austria
de	Theo	Kurz	male	1976-09-28	49	8310-976287-8	Peter-Fuchs-Gasse 2	Melk	3390	Austria
de	Emma	Haller	female	2002-08-20	23	9549-097415-1	Kohlgasse 8/3	Zeltweg	8740	Austria
de	Aron	Petz	male	1951-04-22	74	3429-304328-6	Gasserstraße 25	Kufstein	6330	Austria
de	Aylin	Thalhammer	female	1970-12-21	54	6419-260781-6	Lisa-Marie-Luger-Gasse 506	Hartberg	8230	Austria

B

C

EU Citizen Dataset Geo Locations



HL7 Europe

SYNDERAI Synthetic Data Generation

D • Template-based Generation of synthetic examples instances
  - for EPS, LAB and HDR so far
  - AI use interspersed
  - Uses Twig Templates
  - Output as FSH

E • FHIR instance conversion from FSH (JSON+XML)

• Validation, Publication

TWIG Template

- Generates both the FSH and the human readable text
- Conditions and loops available

```
%%FSH%% {# tag required #}
Instance: {{ setInstance("Instance-AllergyIntolerance-" ~ instanceid) }}
InstanceOf: AllergyIntoleranceEuEps
Title: "AllergyIntolerance"
Description: "AllergyIntolerance"
Usage: #inline

* id = "{{ instanceid }}"

{% if allergyintolerance is empty %}

{# no known Allergy/Intolerance data, emit it and we are done #}

* code.coding[0] = $sct#716186003 "No known allergy (situation)"

* patient = Reference(urn:uuid:{{ patient.instanceid }}) "{{ patient.name }}"
```

```
{{ addHTML_tr() }}
{{ addHTML_td("No known allergies or intolerances") }}
{{ addHEAD_trend() -}}
```

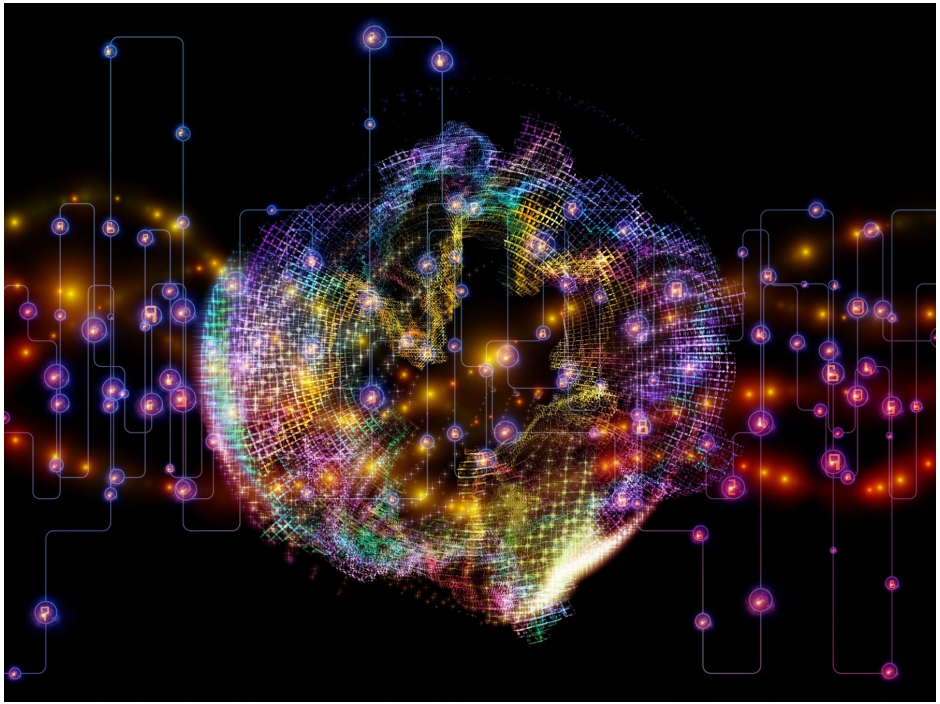
```
{# add category #}
{% if allergyintolerance.category is not empty %}
* category = #{{ allergyintolerance.category }}
{# HTML td 4: add human text for category/type #}
{{ addHTML_td(allergyintolerance.category) }}
{% else %}
{{ addHTML_td("?") }}
{% endif %}
```

```
{# populate .text #}
* text.status = #generated
* text.div = """
<div xmlns="http://www.w3.org/1999/xhtml">
<table class="hl7__ips">{{ emitHEAD() | raw }} {{ emitHTML() | raw }}</table>
</div>
"""
```

HL7 Europe

AI in SYNDERAI

- “Low dose“ AI application
  - appropriate dosage for a specific medication
  - Stratified normal lab value ranges based on patient‘s conditions
  - conclusion per lab report, based on all prior synthetic lab reports



Human Text in SYNDERAI

- Human reable text with AI assistance
  - especially for the Hopsital Discharge Reports (HDR).
  - In reality, HDR sections typically contain text along with granular data such as codes or measurements for medication, results, diagnoses, etc.
- For consistency in the example generating algorithms, a **Instance Short Hand** (ISH) notation was added to the tooling
  - allows concise description of instance contents using same mechanisms as for FHIR instance generation as for other artifacts

ISH

- **Instance Short Hand (ISH)**
- Conversion of human or AI invented text to FHIR examples

```
# HDR for Yascha Schulze with open fracture of left zygomatic arch
patient
  birthdate 1996-09-24
  given Yascha
  family Schulze
  localid K5463847500
  gender male
  postcode 12053
  city Berlin
  street Biebricher Straße 3
  country DE
  phone +49 152 865746356
  nameset de
  latitude 52.481903
  longitude 13.425926
# closest hospital will be dectected by SYNDERAI
encounter
  start 2025-06-06T21:45:00Z
  end 2025-06-08T11:00:00Z
  reason* $sct#16664831000119108 Open fracture of left zygomatic arch (disorder)
  service* $sct#1362046005 Computed tomography of maxillofacial area (procedure)
# the provider person in the hopital as primary contact person there
provider
  prefix dr dr
  given Jochen
  family Bein
  email jochbein@viviantes.de
  phone +49 30 234276-13
section*
  type admissionevaluation
  code $loinc#67852-4 Hospital Admission evaluation note
  title Admission evaluation
  text """The patient, a 29-year-old male, was admitted following a physical altercation
```

Story Telling

A smaller set of personas were invented based on the described synthetic foundation.

In selected cases, stories were defined for Hospital Discharge Reports that were completed by matching data.

SYNDERAI Synthetic Data Generation / Use Policy

- **Definition**: **Synthetic data** refers to data that are *entirely artificially generated* and not derived from any identifiable individual; synthetic datasets may be:
  - **Algorithmically generated** using simulation tools (e.g., Synthea, internal model generators)
  - **Statistically modelled** from aggregated or anonymized population parameters
  - **Augmented** with fictional identifiers, timestamps, and locations

No original patient data is used, and no record linkage to real systems is possible

SYNDERAI Synthetic Data Generation / Use Policy

- **Generation Process**
  - Data are generated by designated software components (FHIR Device resources identified in Provenance)
  - Generation models are validated for structural correctness (FHIR Shorthand / JSON) and semantic plausibility (code system integrity, terminology consistency)

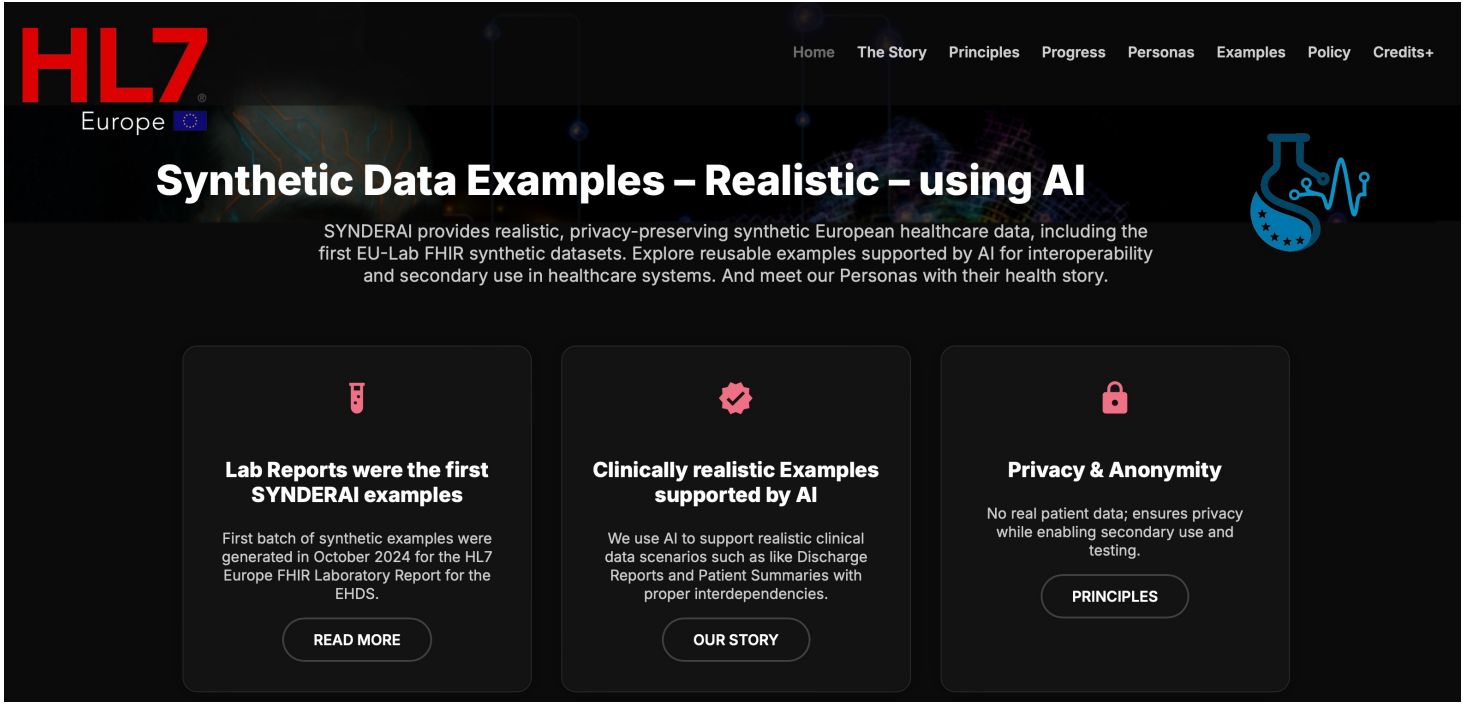
SYNDERAI Synthetic Data Generation / Use Policy

- **Generation Process**
  - Data are labelled with:
  - `meta.security` codes `HTEST` and/or `TRAIN`
  - A `meta.tag = { system: "https://synderai.net/fhir/tags", code: "synthetic" }`
  - A Provenance record referencing this policy URL

…must not be used for clinical decision-making or any production environment.

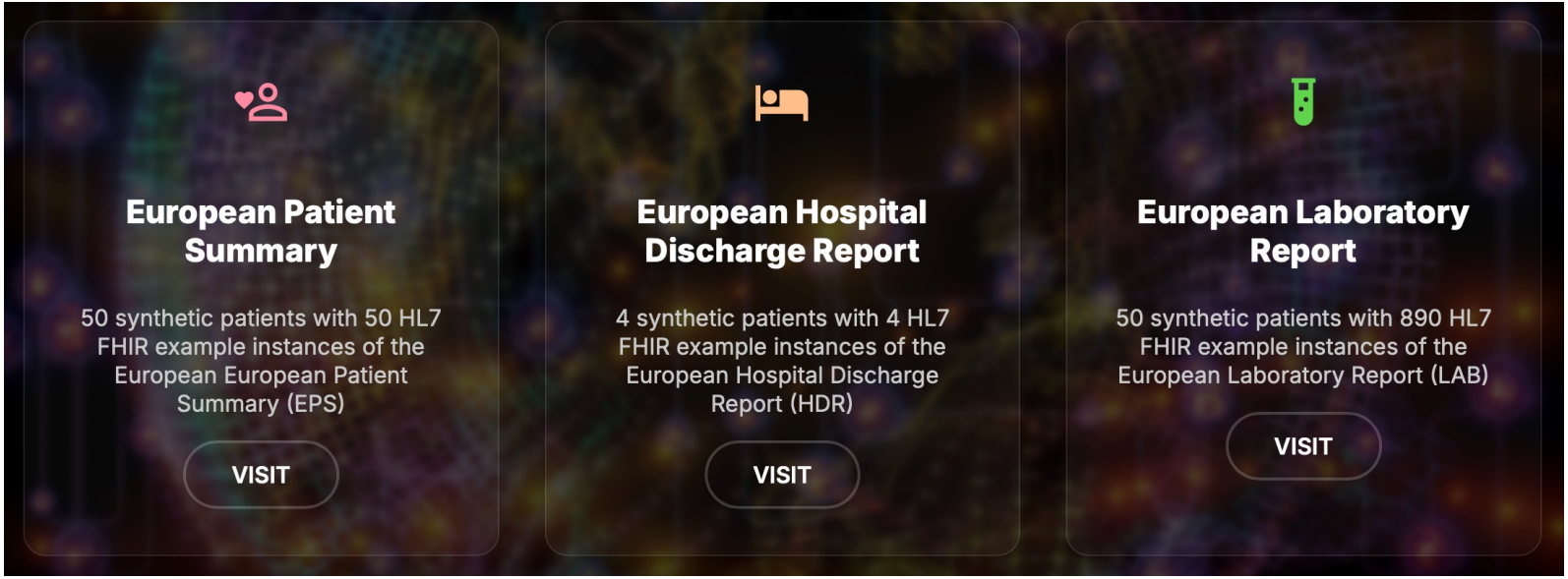
Publications and Timeline

SYNDERAI Website synderai.net / GitHub

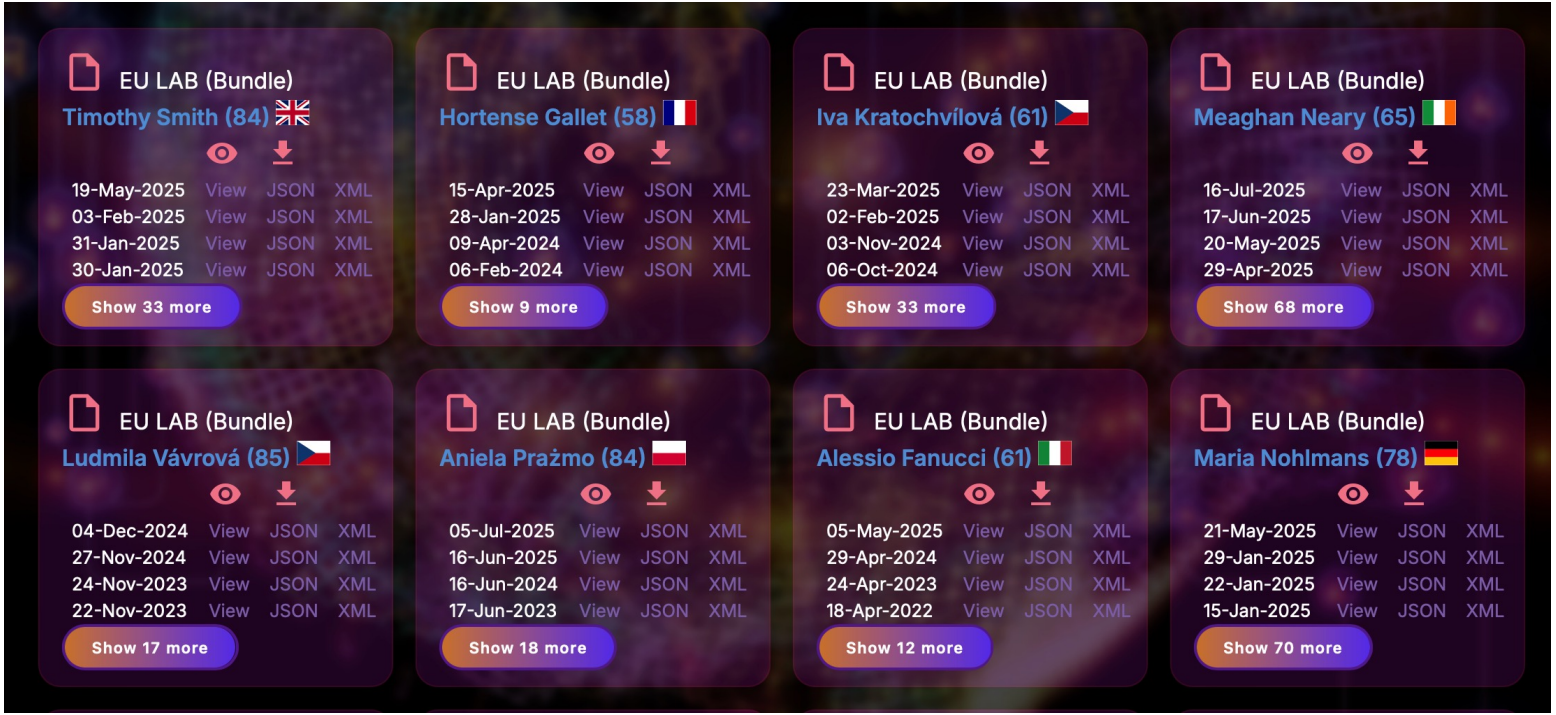


Synthetic Data Example Categories

- 50 EPS, 890 LAB and 4 HDR so far



Synthetic Data Example List for EU LAB



Viewing

SYNDERAI is combined with Example **Visualization**, supported also by the Gravitate Health project, see vi7eti.net

Laboratory Report

Patient

Name: **Smith**, , Timothy,
DOB: 10-SEP-1941 (Age: 84)
Gender: male
Address: 9 Jeffrey orchard
NW1 North Joshuaville (United Kingdom)
ID: 8029-862360-6 (ECI)

Report

Date: 19-MAY-2025

Laboratory

dr **Ample**, Ex
Laboratoire Central Européenne
Boulevard du Jardin Botanique 32
1000 Brussels (Belgium)

Requested by

Evelina Children's Hospital
SE1 7 London (United Kingdom)

Specimen

Collected: 19-MAY-2025

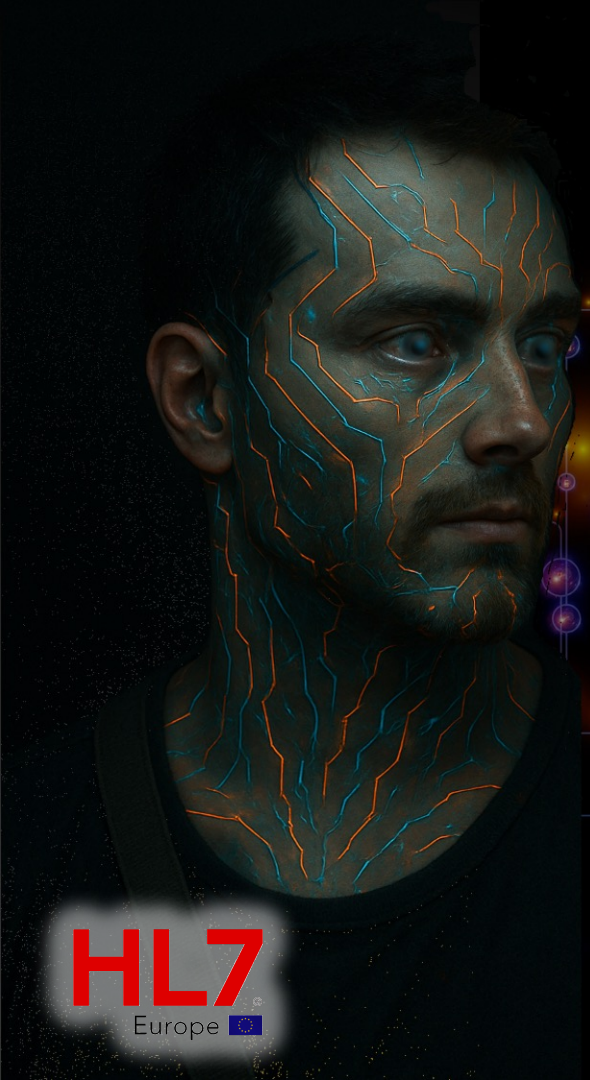
Chemistry

Test	19-MAY-2025	Reference Range	Unit
Hemoglobin A1c/Hemoglobin.total in Blood	5.5	4.5 - 6.4	%
Glucose [Mass/volume] in Blood	131.1	70 - 140	mg/dL
Urea nitrogen [Mass/volume] in Blood	19.6	7 - 25	mg/dL
Creatinine [Mass/volume] in Blood	0.7	0.6 - 1.3	mg/dL
Calcium [Mass/volume] in Blood	9.1	8.5 - 10.5	mg/dL
Sodium [Moles/volume] in Blood	143.9	135 - 145	mmol/L
Potassium [Moles/volume] in Blood	4.7	3.5 - 5.1	mmol/L
Chloride [Moles/volume] in Blood	102.9	98 - 107	mmol/L
Carbon dioxide, total [Moles/volume] in Blood	26.5	22 - 29	mmol/L
Cholesterol [Mass/volume] in Serum or Plasma	**244.9 H**	0 - 200	mg/dL
Triglyceride [Mass/volume] in Serum or Plasma	**490.0 H**	0 - 199	mg/dL
Cholesterol in LDL [Mass/volume] in Serum or Plasma by Direct assay	119.1	0 - 130	mg/dL
Cholesterol in HDL [Mass/volume] in Serum or Plasma	**27.8 L**	40 - 100	mg/dL

Timeline 2025/2026

- Examples for 1,000 synthetic patients start of Q1 2026
- Investigating necessity for 25,000 records, including EPS and IPS output, LAB and derived HDR for Q2 2026
- Important landmark: xShare EHR Xchange Format ends in 2026
- Will be continued after xShare ends

Summary



Summary

SYNDERAI

Synthetic Data Examples – Realistic – using AI

- European Use Cases: Laboratory Report, Hospital Discharge Report, European (+ International) Patient Summary, soon Imaging
- **1,000+ Synthetic Lab Report instances with multiple reports per patient over time**
- **1,000+ Synthetic European Patient Summaries**
- "low dose" AI application for selected areas
- Part of the xShare Project, see synderai.net
- combined with Example **Visualization**, supported also by the Gravitate Health project, see vi7eti.net

HL7 Europe

Thank you!

SYNDERAI
Synthetic Data Examples – Realistic – using AI

dr Kai U. Heitmann • kai.heitmann@hl7europe.org