Research & Data Governance
Level 1-4

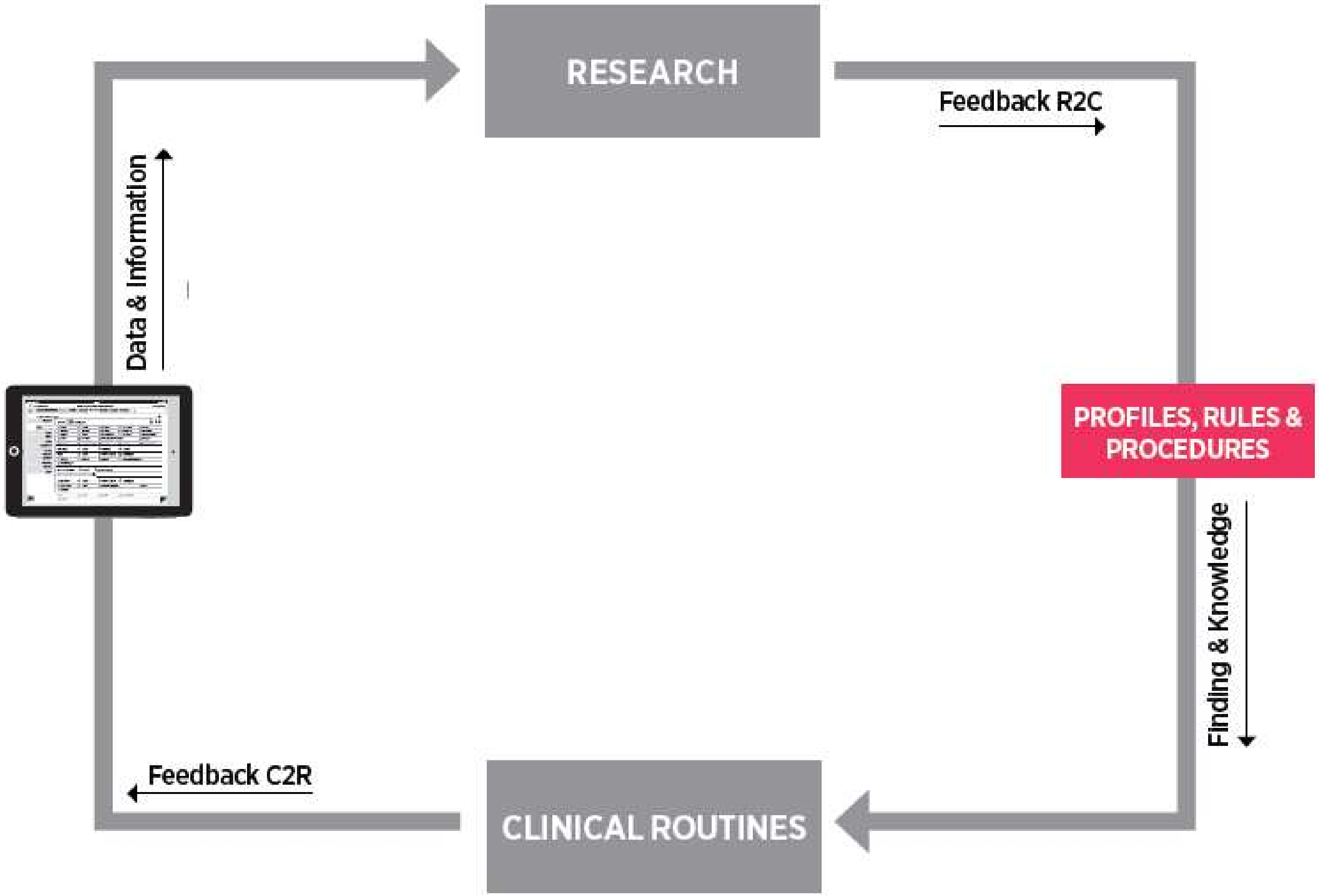
Stefan Hubeli, Dr. Hansjörg Lehner, PD Dr. Alexander Souza

March 2019

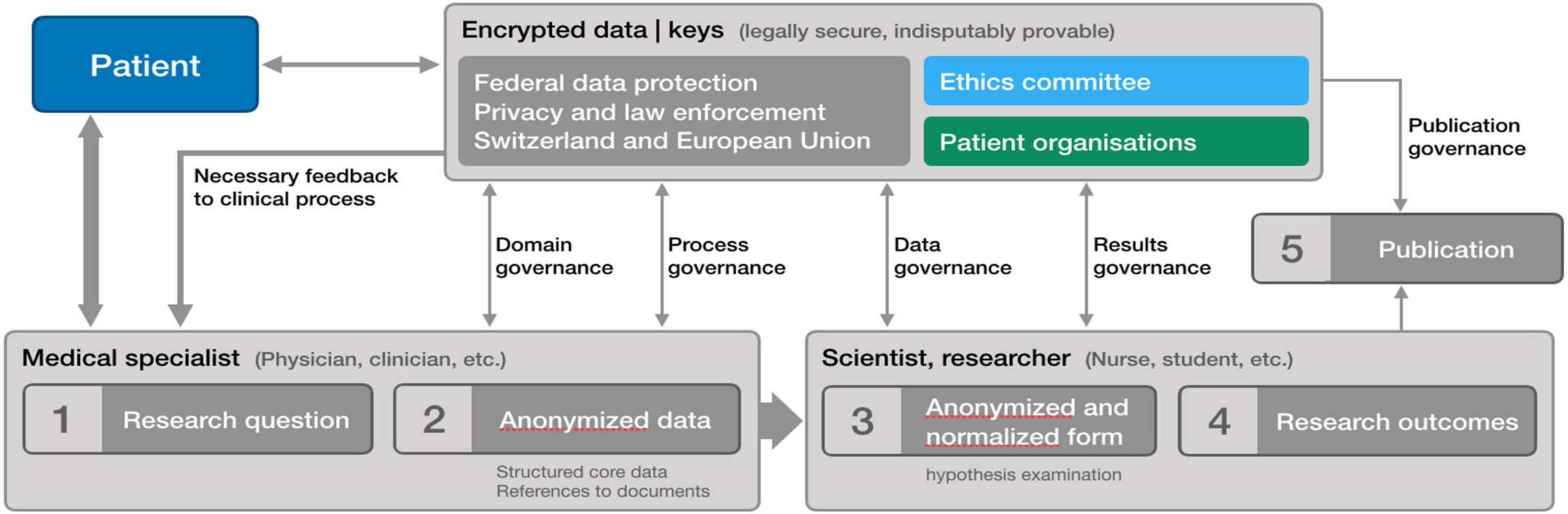
Agenda

- Status Data Governance & Processing
  - Research Process & Research Governance
  - Level 1 bis Level 4
- Appendix

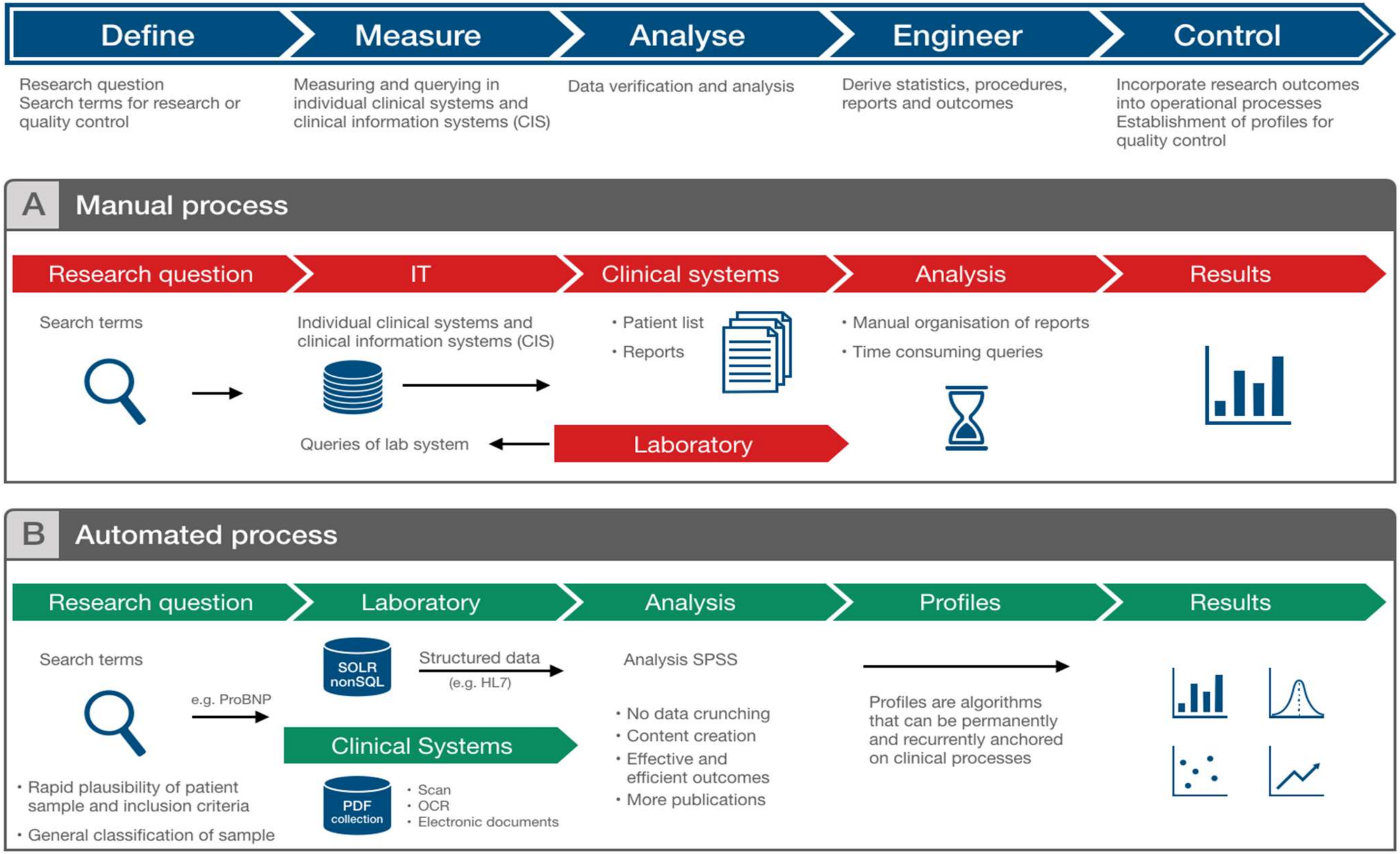
Feedback-Loop
Clinical Routines ←→Research



Research Governance



Research Process



Research Data Governance and Result Level 1 to 4

	Possible Question	Result, Feedback
Level 1 Availability of Patient Set (Sample)	Do we have any patients with e.g. • Steven Johnson, Lyell Syndrom? • Medications MTX and RA?	• *yes or no*
Level 2 # Number of Patients, Groups	Is there a significant (critical) sample size, e.g. • How many patients have lab values for Troponin, BNP and etc.?	• *number of patients*
Level 3 Question Focused Dataset (QFD)	Anonymized combination of a set of multi variables (study profiles, cohorts) filtered and aggregated • Full complete research QFD e.g. demographical, age , gender distribution	• *number of patients based on QFD*
Level 4 Complete QFD for research study, agreed through patient * patient consent is relevant	• Study monitoring • Studies on medication interaction • Intervention studies • Patient support studies •	• *patients based on QFD validated*

Research Governance: Outcome and Results

Iterata AG

We providing **System-Outcomes** to our Domain-Experts

Domain

We providing **Results and Findings** to the Public & Domain Community

Level 1

Availability of Patient Set (Sample)

Level 2

Number of Patients, Groups

Level 3

Question Focused Dataset (QFD)

Uptake

Processing

Use

Level 4

Complete QFD for research study, agreed through patient

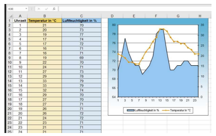
* patient consent is relevant

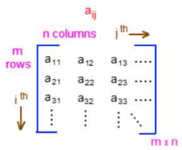
Requirements for Platform, Tools & Methods

	Platform & Tools	Methods
Level 1 Availability of Patient Set (Sample)	• Corpus of clinical documents needs to be available and searchable • Initial search profile for groundset agreed	• Anonymized full-text search • Aggregation queries
Level 2 # Number of Patients, Groups	• Structuring of patient- and document metadata: P4, hash, dates, timeline, gender, …	• Gathering and grouping • Anonymized aggregation • Statistics
Level 3 Question Focused Dataset (QFD)	• Details (except name) visible • Analysis on document metadata • Pseudonymized	• Gathering and grouping • Anonymized aggregation • Statistics
Level 4 Complete QFD for research study, agreed through patient * patient consent is relevant	• All details visible • Analysis on document metadata • and text	• Gathering and grouping • Anonymized aggregation • Statistics • Verification of results

Data addressability
- What does this mean and how to do it?

Data Tables, e.g. Excel





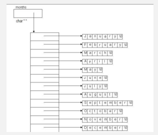
Algorithmic addressability

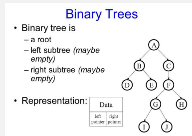
Why data addressability?

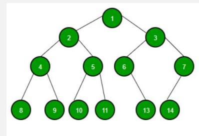
- efficiant traversings
- easy sortable
- efficiant searchable / findable

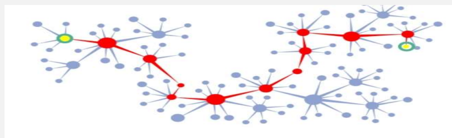
<u>Systematic addressability</u>

unstructured data chaos «Sauhaufen»	• *Not traversable* • *Not efficiant searchable and findable*
Lists	• *traversable* • *efficiant searchable and findable*
Trees	• *Branching lists*
Graphs	• *Loops* • *Polygon*





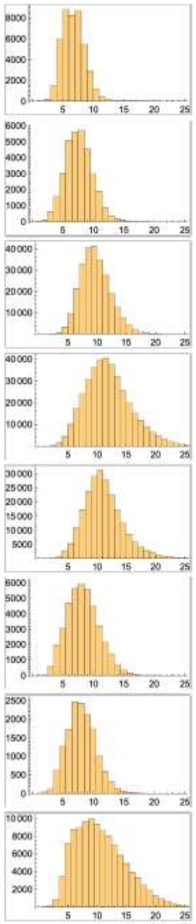




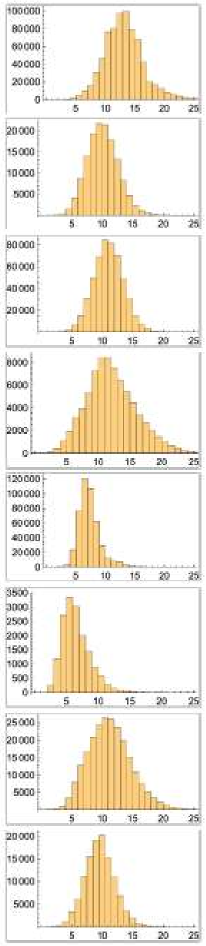
Data addressability
- Example languages

Systematic Searchable / Traversable

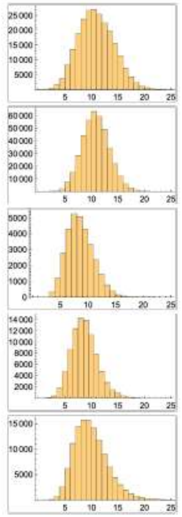
etc.



Arabic	43 768	6.24913	3.36028
Breton	32 733	7.64675	4.98327
Croatian	260 752	9.257	6.72885
Danish	376 468	11.4697	15.9237
Dutch	229 365	10.6407	11.3638
English	40 127	8.33284	7.00445
Esperanto	14 780	8.07463	5.6892
Faroese	108 508	10.6502	19.3565



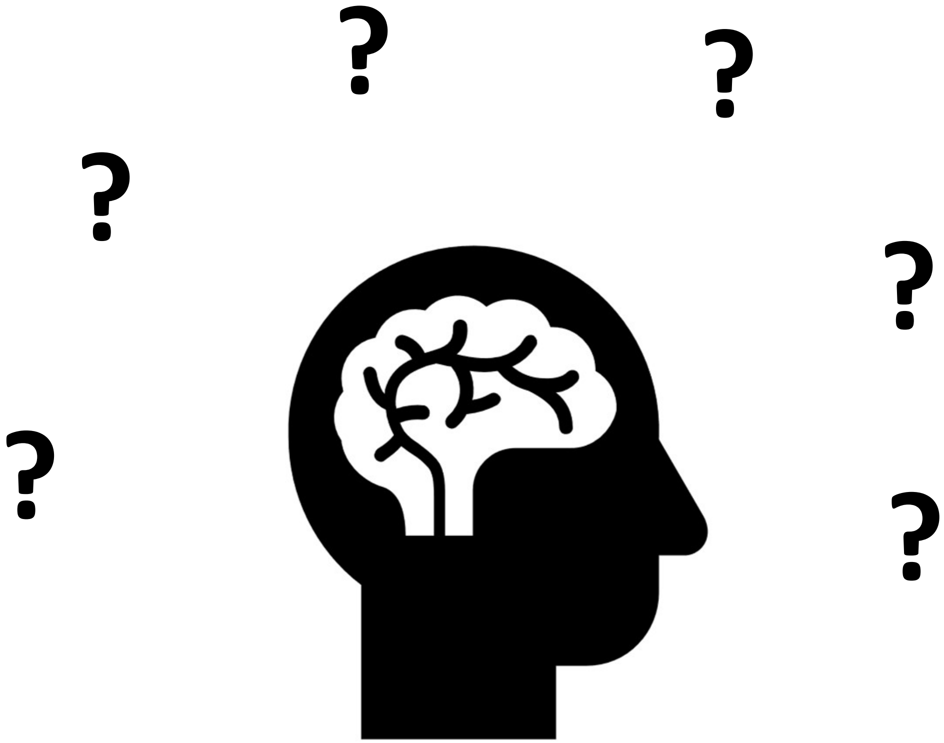
Finnish	728 498	12.6506	10.7855
French	139 657	9.60079	6.40334
Galician	515 385	10.5356	5.98442
German	76 155	11.6281	15.6155
Hebrew	455 264	6.75852	3.13379
Hindi	15 983	6.00325	4.73657
Hungarian	238 206	10.7419	12.1841
Italian	116 854	9.64161	5.878



Polish	234 907	10.4405	11.1556
Portuguese	407 581	10.3567	6.81621
Russian	31 801	8.10481	5.84458
Spanish	86 016	8.70562	5.96293
Swedish	121 430	9.48335	10.3674

etc.

Questions



Appendix
- Specific Domain Fields

Cardio Amgen

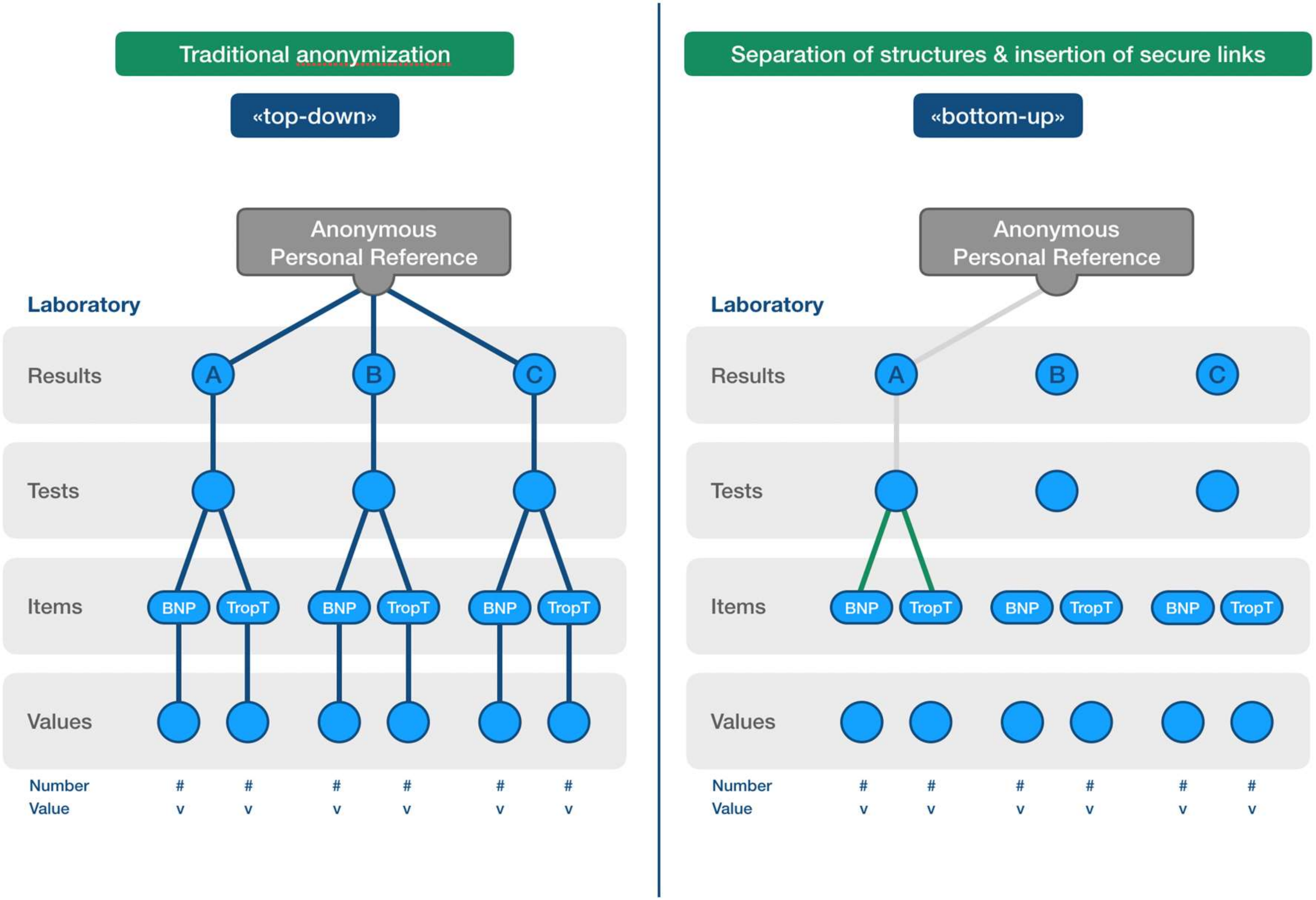
Research Data Governance and Result Level 1 to 4

Level	Possible Question	Result, Feedback
Level 1 Availability of Patient Set (Sample)	Do we have any patients with e.g. • KHK? • Statine?	• *yes or no*
Level 2 # Number of Patients, Groups	Is there a significant (critical) sample size, e.g. • How many patients have lab values for Troponin, BNP and etc.?	• *number of patients*
Level 3 Question Focused Dataset (QFD)	Anonymized combination of a set of multi variables (study profiles, cohorts) filtered and aggregated • Full complete research QFD e.g. demographical, age , gender distribution	• *number of patients based on QFD*
Level 4 Complete QFD for research study, agreed through patient * patient consent is relevant	• Study monitoring • Studies on medication interaction • Intervention studies • Patient support studies •	• *patients based on QFD validated*

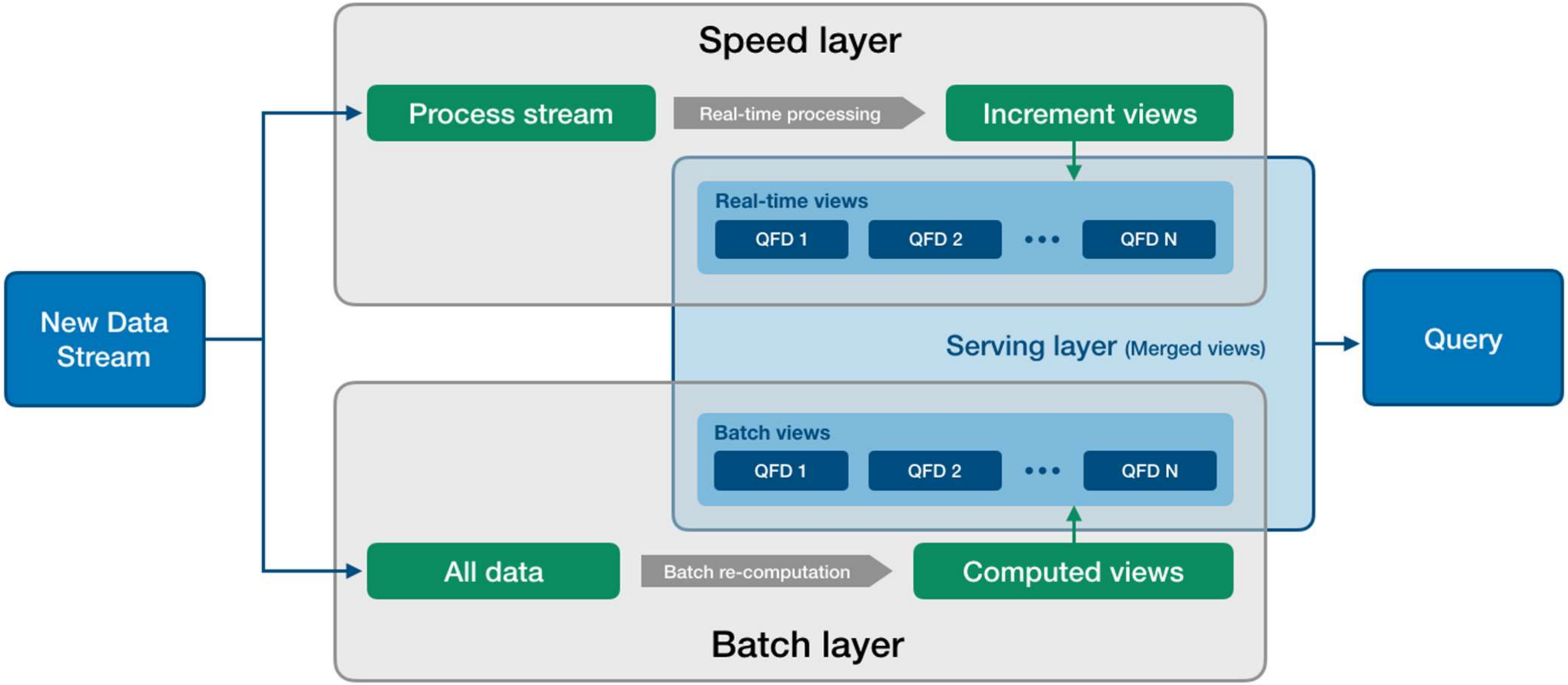
Appendix

- Anonymization / Deanonymization
- 2-Speed IT
- 4 Layer
- SEMAT
- Data View
- Stakeholder, Algorithms, Access

Anonymization / Depersonalisation

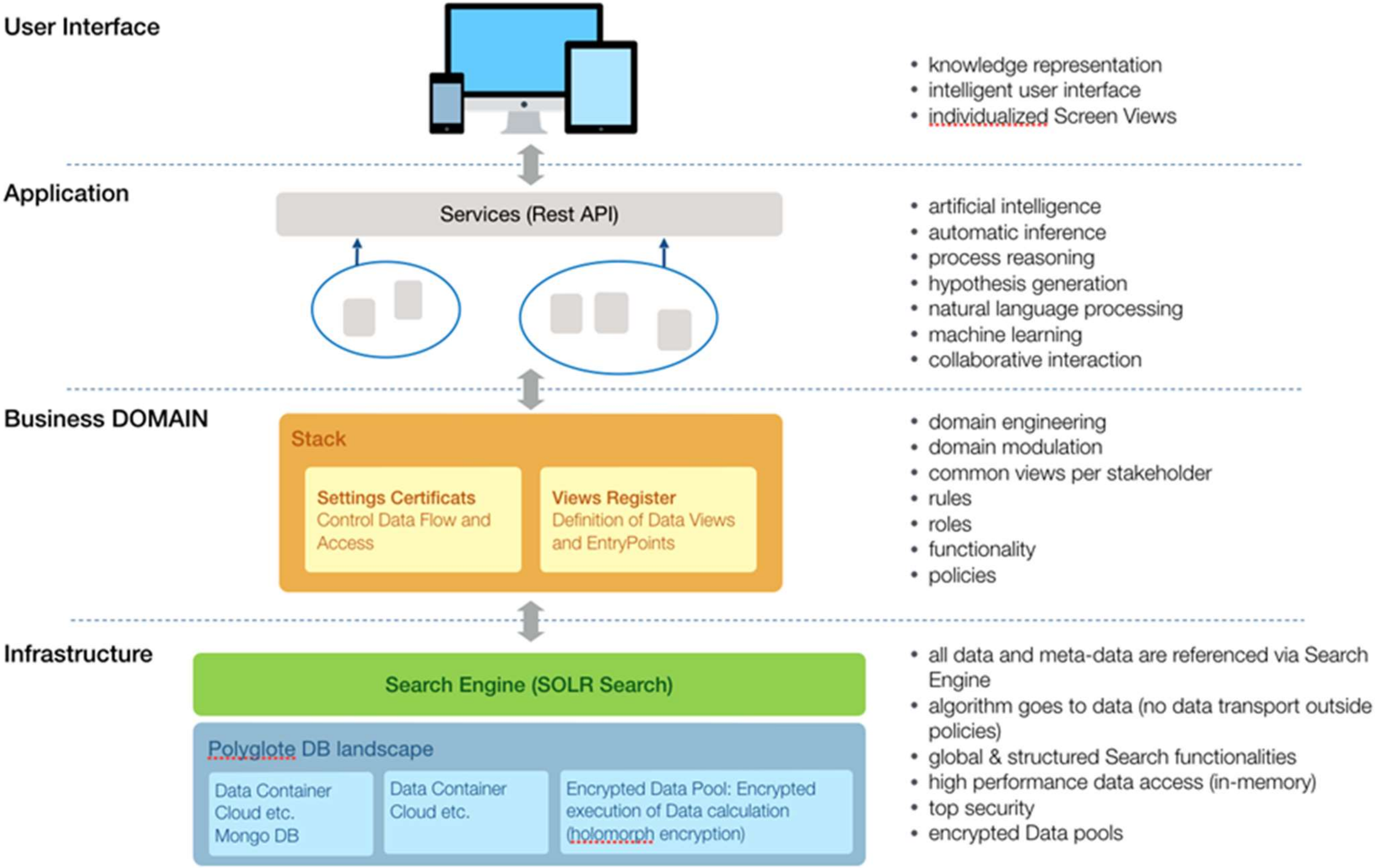


2-Speed IT

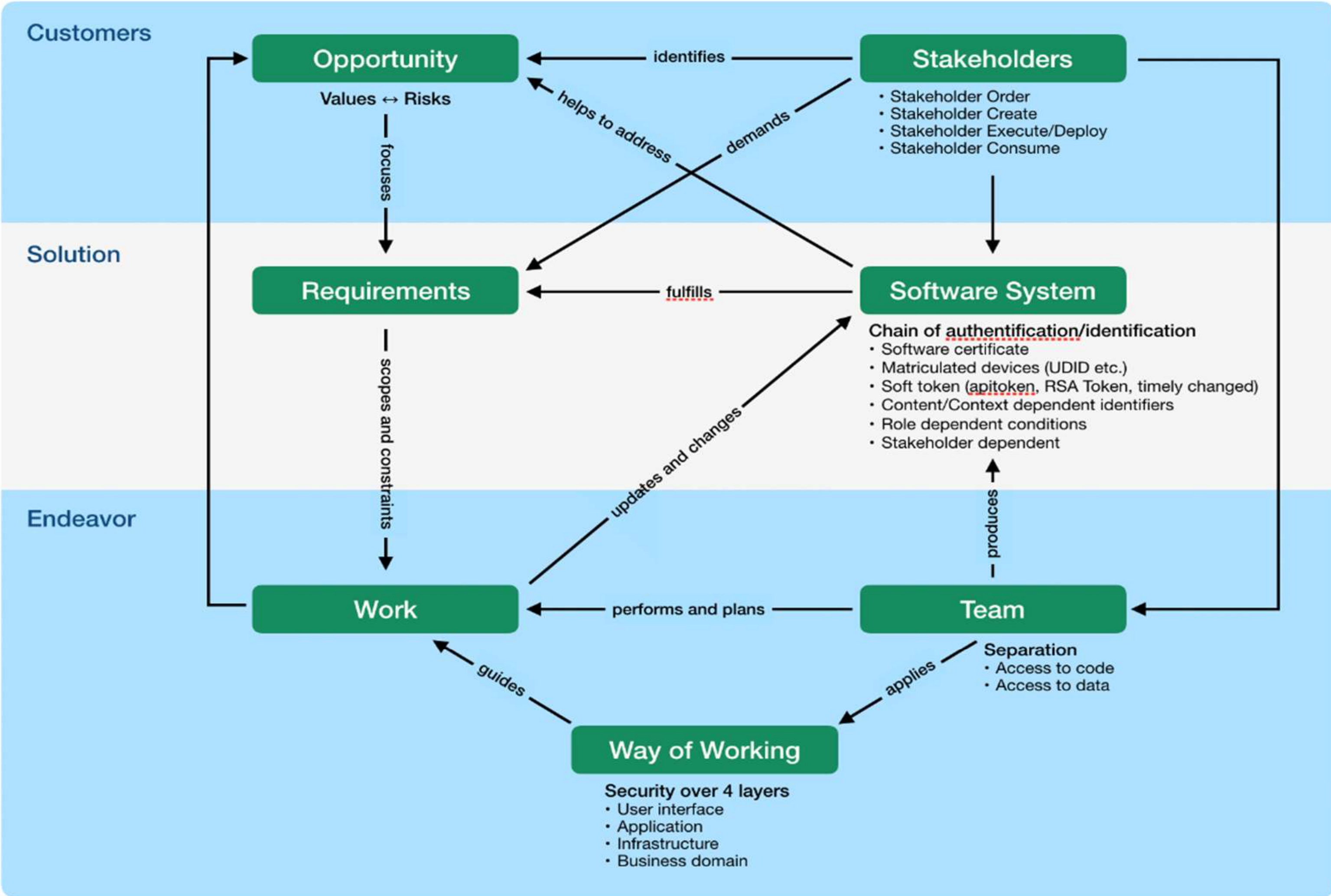


QFD = Question Focused Datasets

4 Layer Model



SEMAT



Customers
Solution
Endeavor
Opportunity
Values ↔ Risks
Stakeholders
• Stakeholder Order
• Stakeholder Create
• Stakeholder Execute/Deploy
• Stakeholder Consume
identifies
helps to address
demands
focuses
Requirements
fulfills
Software System
Chain of authentification/identification
• Software certificate
• Matriculated devices (UDID etc.)
• Soft token (apitoken, RSA Token, timely changed)
• Content/Context dependent identifiers
• Role dependent conditions
• Stakeholder dependent
scopes and constraints
updates and changes
produces
Work
performs and plans
Team
Separation
• Access to code
• Access to data
guides
applies
Way of Working
Security over 4 layers
• User interface
• Application
• Infrastructure
• Business domain

Data View

A **Data-View**

Two Speed IT
- lambda architecture

B **System Architecture**

Four layers
- User Interface
- Application
- Business DOMAIN
- Infrastructure

Implementation

A. Incoming, Referencing
Development of new data stream

B. «Poison cabinet» of all data, Storage of original data / links
HL7 lab files → Poison cabinet
PDF report → Link to original data from archive

C. «Ur-soup»: Super QFD (Anonymisation)
Separation of personal data according to message type, items

D. QFD Troponin Research Study: first derivative of Super QFD
Parameter: Troponin Study Criterion

Each data point has:
- Reference to avatar
- Reference to HL7 (Message type, item)
- Reference to avatar linked to person, «zero knowledge proof» concept

Stakeholder Algorithms Access

	Stakeholder	Algorithms			
1	Top	**ORDER** Orders query and task Adopts leadership control, adherence to governance			
2	Create		**CREATE** Creates/develops query and algorithm according to order of top Stakeholder		
3	a) Execute b) Deploy			**EXECUTE** Triggers execution or determines distribution of algorithm **DEPLOY** Given legitimated stakeholders access to algorithm	
4	Consume				**CONSUME** Receives set of results