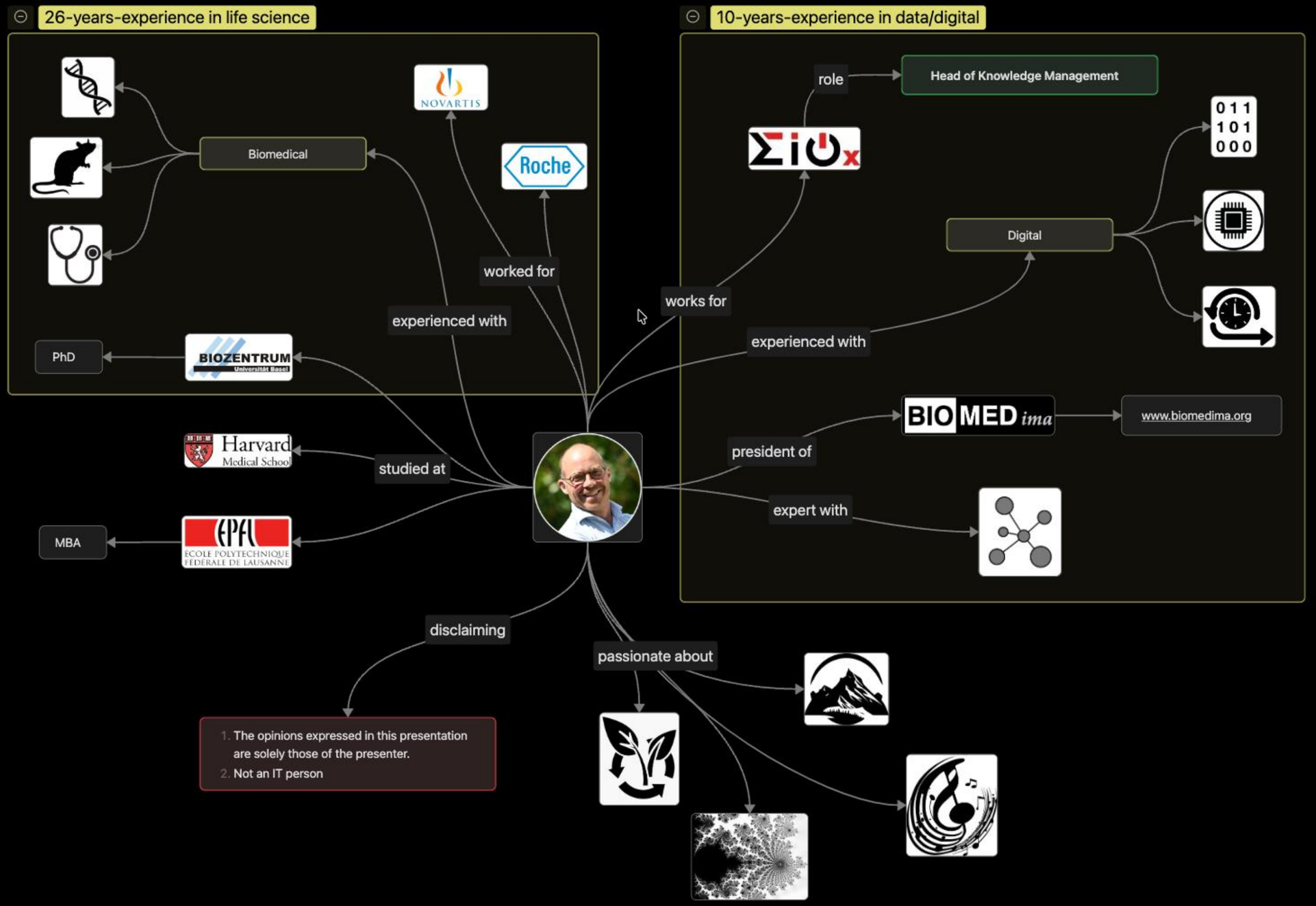
Intelligent Data For Life Sciences

Beyond the Hype, how Tailored Graph-AI Solutions Change the Game

Cedric Berger, PhD, MBA

2025-11-20

connecteddatalondon


26-years-experience in life science
Biomedical
NOVARTIS
Roche
worked for
experienced with
PhD
BIOZENTRUM
Universität Basel
Harvard
Medical School
studied at
MBA
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
10-years-experience in data/digital
role
Head of Knowledge Management
works for
Digital
experienced with
BIOMED ima
www.biomedima.org
president of
expert with
disclaiming
1. The opinions expressed in this presentation are solely those of the presenter.
2. Not an IT person
passionate about

Life Science Industry

(One) definition: Human (and ideally humane) care, public and private sectors, diagnosing, treating, ideally curing pathologies using drugs and devices.

Global Market

Billions
of people with chronic diseases

$ 10.55t
(2024 market)

Dedicated global, regional and national
regulators and laws

Drug Dev.

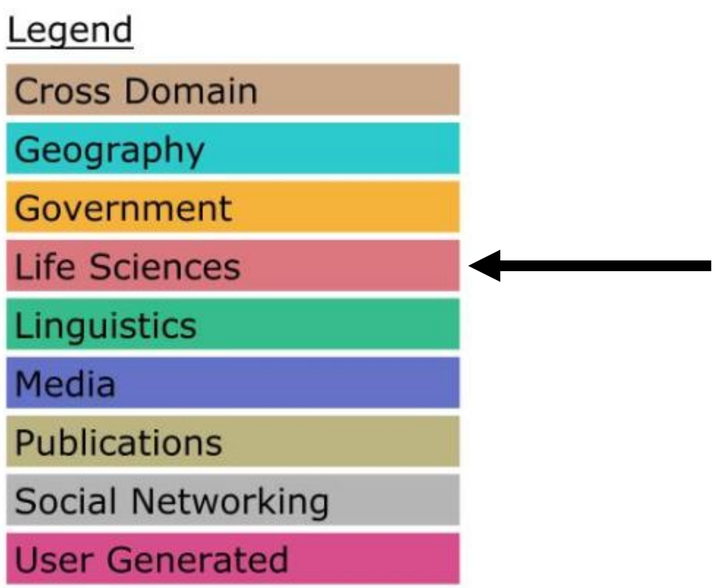
300 business capabilities
>30 business domains
>150 operational processes

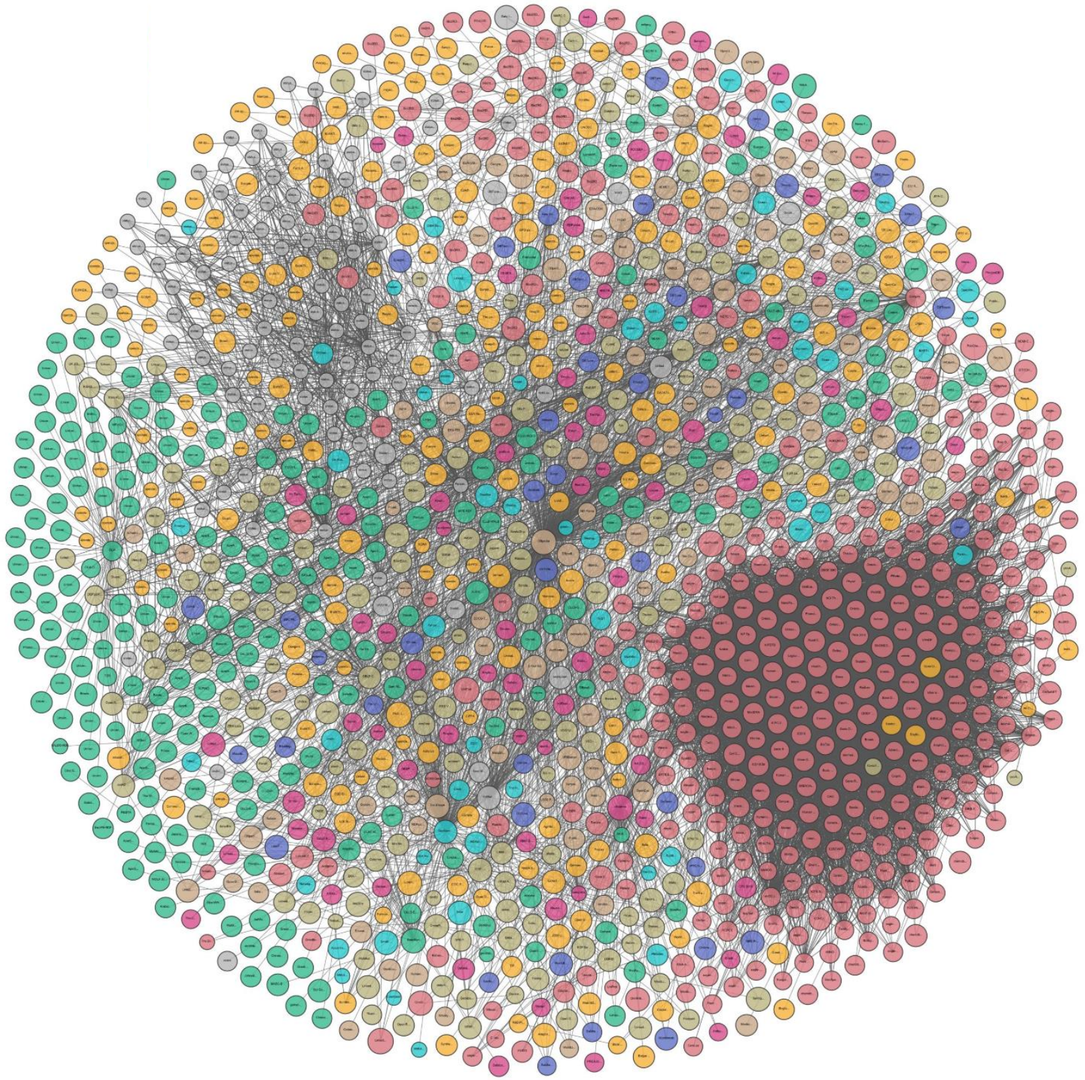
500 to > 6’000
IT systems per “big” pharma
100 petabytes to **exabytes**
of data annually

> 200
distinct data standards

Life Science Open Data

The Linked Open Data Cloud





https://lod-cloud.net/

2016 - 2019

Cornell University

We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors. Donate

arXiv > cs > arXiv:2311.02082

Search... All fields Search

Help | Advanced Search

Computer Science > Artificial Intellige

[Submitted on 20 Oct 2023 (v1), last revised 23 Nov 2

Semantic Modelling of Or
Governance 4.0 -- Applic

Miguel AP Oliveira, Stephane Manara, Br
Kulkarni, Pralipta Jagdev, Cedric R. Berg

Individuals and organizations cope with an
management process yielding data quality
related to multiple usages. Common data g
data. Yet, harnessing this complexity is nec
including generative artificial intelligence tr
framework that enables metadata-driven, a
this framework to integrate 25 years of clin
methodologies and technologies leveraging
context, including governance principles. M
FAIRification, lifecycle management, definit
metadata model is the keystone to data gov
manner to adapt governance constraints to

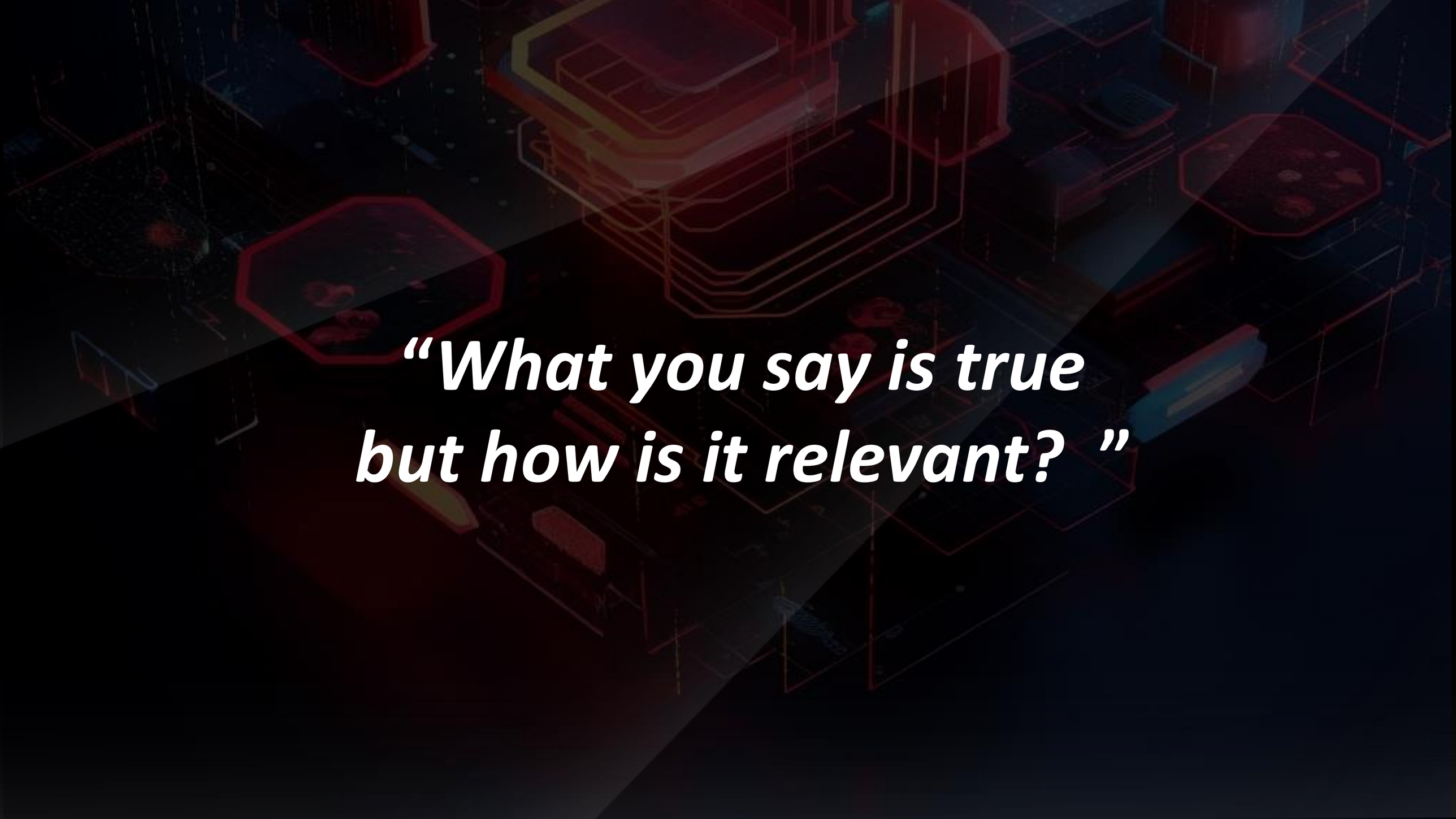
Enterprise Upper Ontology
Metadata governance: upper ontology, master and reference data management

Governance Ontology
Principles, policies, laws, rules, checks, classifications, artefacts, roles, pointers to documents, URI builder, processes

Domains and Concepts	Inventories and Catalogues	FAIRification Metadata	Taxonomies Ontologies	Models	Use Cases
Cross-industry business concepts grouped by domains in a one-to-many manner	• Data Sets, Products • Systems, Applications • Roles • Processes and Tasks • Organization units • Technologies • Projects • Products, Brands • Customers • ...	Buckets of metadata to capture FAIR aspects • Identification • Lifecycle • Controls • Provenance, lineage • Licensing • Anonymization • Controls, Access rights	Enterprise-specific cross-domain (OBO, BioPortal, MedDRA, ATC,...)	External (CDISC, IDMP, OMOP) and internal (Conceptual, Logical, Physical) structuring models	Project: • Problem need, OKRs • Proposed solution • KPIs Domain: • Operationalized solution • Benefit mapping
			Reference Lists Enterprise-specific domain-specific flat lists	**Master Data Schemas**	**Utils** Maintenance, demos

Semantic basis
Enhanced dictionary featuring terms, syno- anto- hyper- hyponyms, one-to-many definitions and sources

https://arxiv.org/abs/2311.02082v2


“What you say is true
but how is it relevant? ”

Methodology

A graph- and AI-based approach
for efficient, fit-for-purpose problem-solving in life sciences
that users can trust.

EaaC

(Everything as a Code)

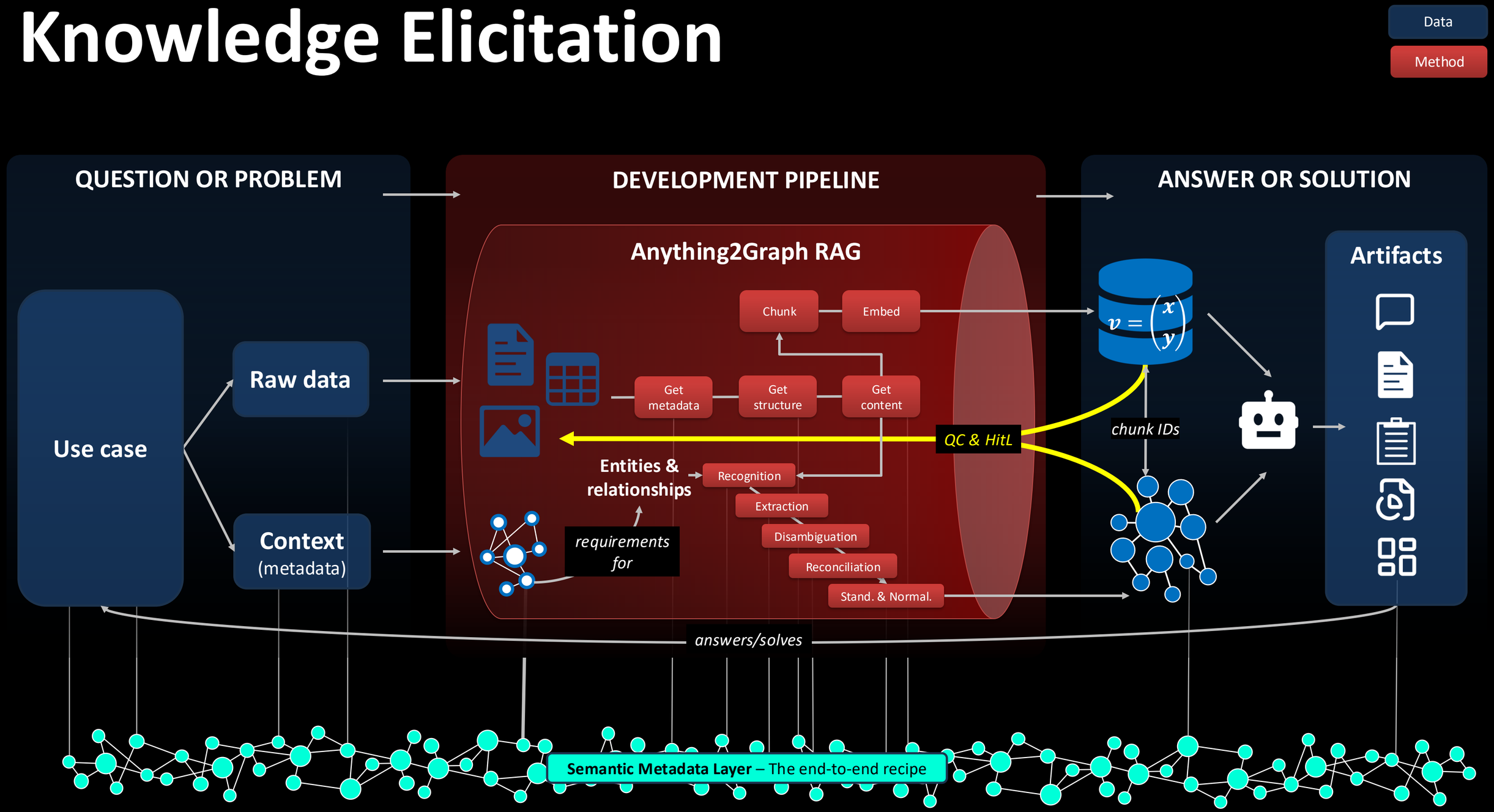
- Git-based
- Modular
- Share and reuse
- FAIR by design

KaaS

(Knowledge as a Service)

- Metadata first
- Use case specific
- Focus on added value
- Agile

Knowledge Elicitation



Knowledge Flow

Data

Method

New solution

Business

(KaaS)

New use case

Graph-Enabled Knowledge Elicitation

Business Analysis
- 5Ws
- CQs
- Sources
- Success/KPIs

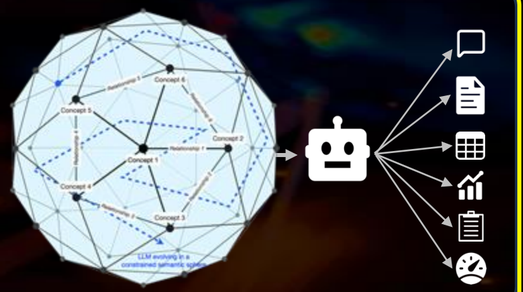
Define **Terminologies**
- for model
- for instances

Top-Down modeling
Business truth

Bottom-Up modeling
IT systems truth

Create knowledge model

Operate knowledge model



Git

Use Case | use-case-o | use-case-t | use-case-i | use-case-sh | use-case-m | use-case-sp

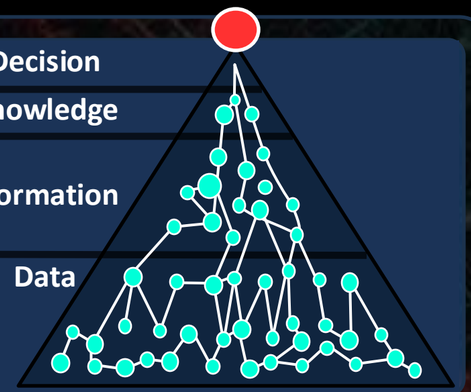
Business-specific assets

Life Science

Semantic Metadata Layer

Housekeeping assets

WHO, MeSH, CDISC, USDM, EMA/FDA, omop, MedDRA, SNOMED, dct, bfo, togaf-archimate, skos/skos-xl, gist, prov-o, bbo, dgov-o, biz_analysis-o, biz_archictecture-o, dcat, dprod, doco, pattern, data-prof-o, data-mod-o, schema.org



Tech

(EaaC)

DevOps

Sources

Entity & relationship extraction

Resolution

Linking

Checks

Human

Machine

Rule-based

LLM-based

QC & HitL

- Complexity on demand
- Gamification
- KG & bot sync
- Personalization

Guiding Principles

People – Ways of Thinking

Graph-
System-
Scientific-
Design-
thinking

Data – The New Oil

- *Sobriety*
- *Sovereignty*
- *Metadata-governed*
- *Linked*

Process – Heart of Business

- *Missing granularity*
- *Changing by the day*
- *Indispensable to AI*
- *Main effort required*

Tech – Pumps & Pipes

- *Value depends on use*
- *Vendor lock-in*
- *Increases data complexity*
- *Raises tech debt*
- *Battle horse for IT careers*

KM – The Why

- *About context that provides meaning*
- *Collaborative/git*
- *Event-driven*
- *Plurality: many2many*

Users yes, but not decision-makers

- *Value/Wii4me*
- *Gamification*
- *Personalization*
- *On-demand*

Org – Culture eats strategy at breakfast

- *Sense & fit the culture*
- *Biz. vs. digital strategy*
- *No one-size-fits-all*
- *Dedicated OCM*

AI – Churning out new stuff from old ones, in new forms, with noise

- *CSA: risk-based validation*
- *Human in the loop*
- *Beware abuse, complacency and resources consumption*

USE CASES

“Increasing entropy enables emergence”

Use Cases Catalog

1. Data Governance - Standards, data quality, FAIRification, provenance, RACI,…

2. Modelling and Linguistics: master and reference data management - Concept and vocabularies integration across domains

3. Business architecture - Business and IT ecosystem description and integration

4. Portfolio management - Molecules, studies, projects, program, applications, collaborations…

5. Inventories and cataloging - Any business and IT assets

6. Data mesh and productization - Data set, domains, roles, systems, customers…

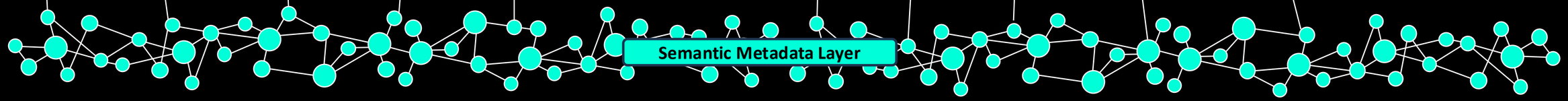
7. Clinical data mapping and integration - 25 years of legacy clinical study variables mapped with 20’000 semantic rules

8. Material supply management - Material, grade, packaging, supplier, trade name, unit price

9. Digital batch tracing and release - Raw material, API, bulk, semi-finished, pack., finished products

10. Monitoring business value delivery - Projects, success, KPIs, roles and responsibilities

11. Delivery resources optimization and simulation – Skills (from **p**eople and agents) management for optimal demand matching and delivery team creation



Integrating 25 years of Clinical Data

Problem:

Data collected upon clinical studies is subject to heavy regulation and constraints.

Clinical study data sets must be conformed to multiple predefined standards,

e.g. CDISC
- SDTM (Study Data Tabulation Model)
- ADaM (Analysis Data Model)

Need:

to
1) share the same understanding
2) manage changes
3) automatize

complex transformations between different clinical study data stages

Success:

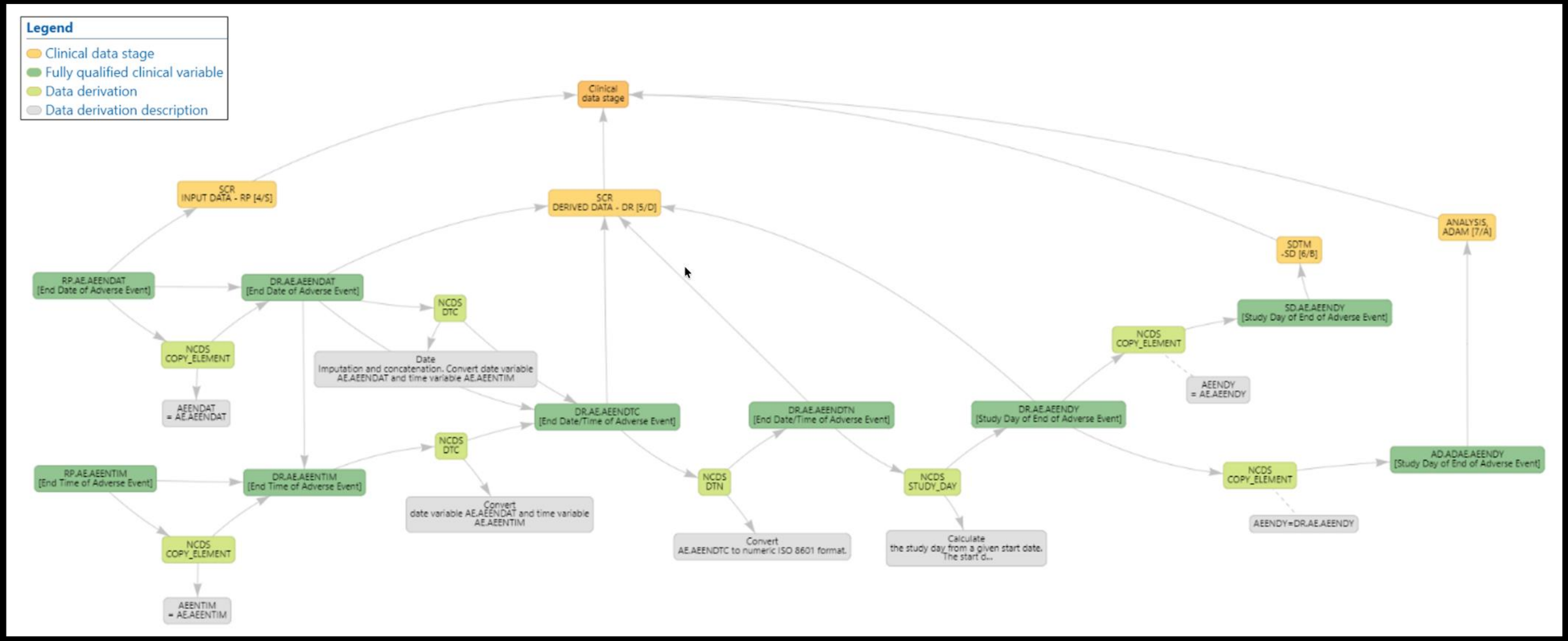
A human- and machine-readable mapping of clinical study data across multiple predefined formats and corresponding transformation.

Rich semantic annotation explaining these transformations using predefined rich metadata set.

KPIs:

- Reduced transformation error rate
- Increase transformation efficiency
- Prediction/simulation of data transformation
- Faster to market by pre-configuring, pre-populating IT systems on critical path

Integrating 25 years of Clinical Data (cont.)



Medical practitioners collect adverse event data using date [RP.AE.AEENDAT] and time [RP.AE.AEENTIM] variables. These are copied to the Derived stage, combined into a single datetime value, converted to a numeric format, and used to calculate the study day of the adverse event's end [DR.AE.AEENDY]. This derived variable is then standardized for CDISC SDTM submission and further transformed for CDISC ADaM analysis use.

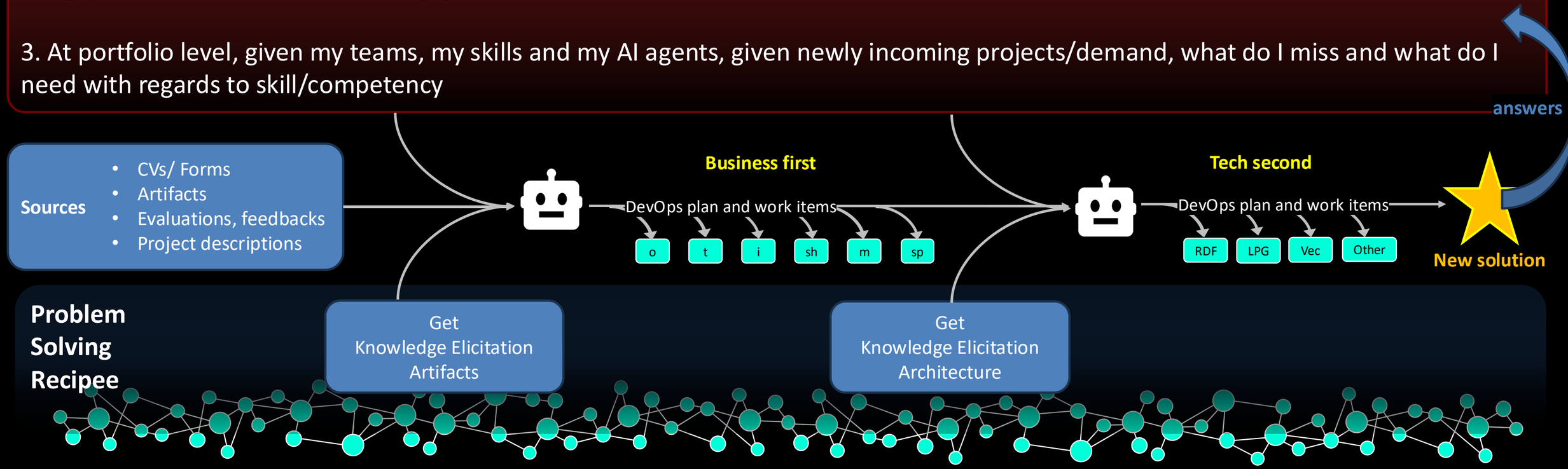
Knowledge Ops

Needs:

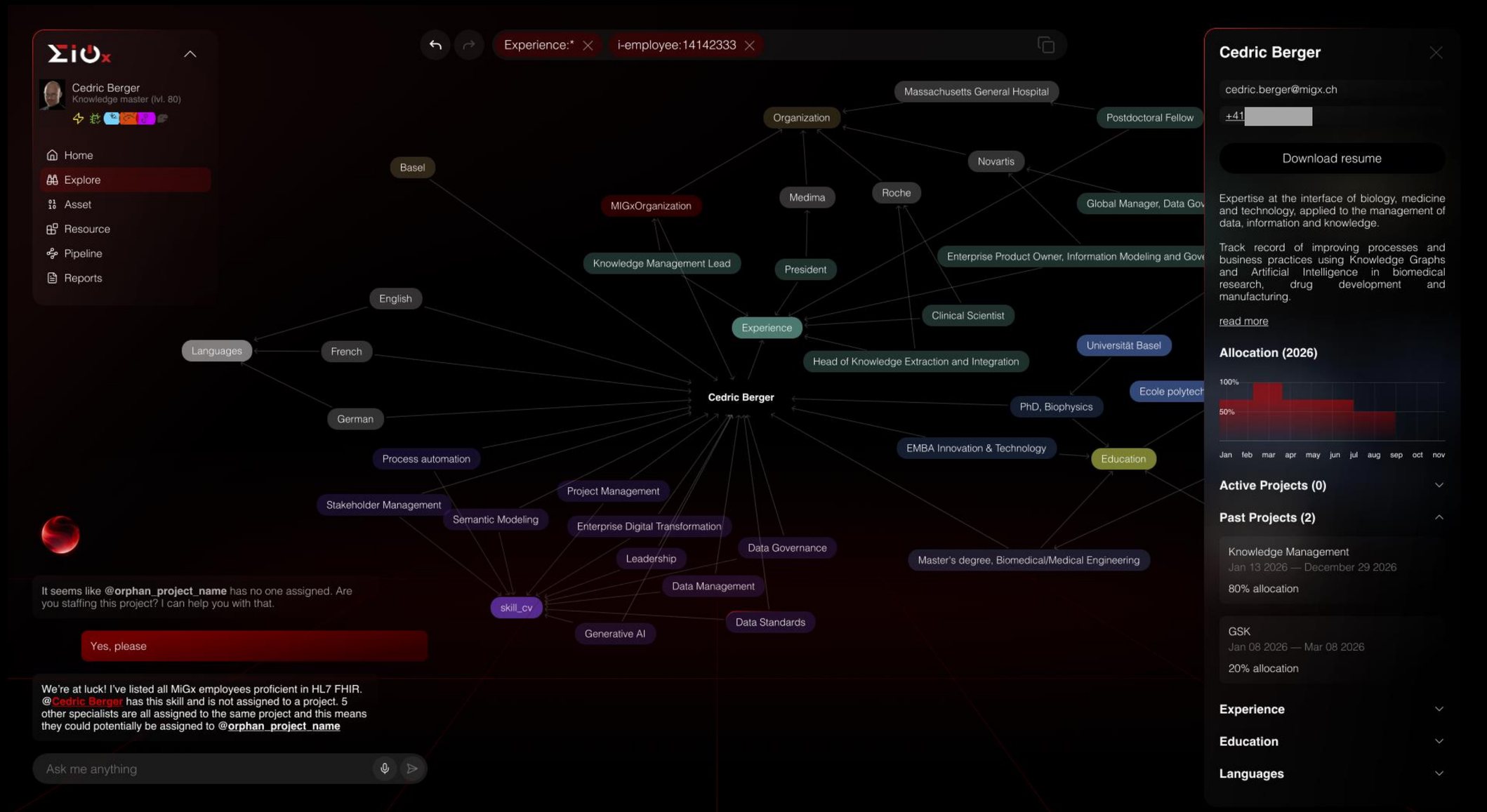
1. Given a named skill/competency, I want to know if we have it in the team? At what are level (1,2,3)? Who are individual(s) possessing it? Which project are they currently assigned to? and what is their current capacity/bandwidth?

2. Given skill/competency and capacity/bandwidth and given requirements for one new incoming request/project, what would be the ideal (virtual) team (humans and AI agents) to deliver the solution? What would be the best team based on existing resources? What are the gaps? What is needed to fill the gaps?

3. At portfolio level, given my teams, my skills and my AI agents, given newly incoming projects/demand, what do I miss and what do I need with regards to skill/competency



Knowledge Ops – Solution Screenshot



Conclusion

“The object is the horizon
where the subject disappears”

A Problem-Solving Methodology

We propose here **a metadata-driven problem-solving methodology** leveraging **knowledge graphs** and **generative AI** based on two core principles: **EaaC** and **KaaS**

EaaC – Technical side

- **Data & logic = code:** oil & pumps/pipes are two sides of a knowledge coin
- **Code is less ambiguous than NL**:
  - better quality of Gen-AI outcomes
  - used to summarize and re-base Gen-AI conversations at specific control points
- **Collaborative, git/branched:** 1+1=3, knowledge economy, use-case and event driven

KaaS – Business side

- **In a world of oblivion**, the digital layer disconnects people from reality
  People with true knowledge (i.e. representing reality with high fidelity) will thrive
- **Knowledge flows**, is multi-faceted, use-case specific, volatile as fast-changing reality
- **We don't know what we don't know**; acknowledging it is knowledge

Call for Action to Graph-Thinkers

Empower graph-thinkers to enable system-thinking and

- account for the complexity of our world and reveal the connections, dependencies, and feedback loops that are too complex for the human brains to identify and avoid simplistic, b&w thinking
- account for diversity, cognitive richness and human plurality that dominant techno-messianic trend, where providers act as both judges and parties, flattens out and steal meaning

Manage pluralism through ontologies

Acknowledge that anyone's knowledge (represented as a graph) carries uniqueness, subjectivity and shared (usually high-level) concepts enabling disambiguation and reconciliation of different points of view and interests.

```
"""
Integrate Cedric's and John's knowledge about "Product" from a business regulatory point of view and market commercialization point of view
"""

import ISO identification of medicinal product
from CedricKG import Product
from JohnKG import Product


def parse_product_file_names(ttl_file):
"""Parse product_category-t.ttl to extract category URIs with labels and altLabels."""
category_map = {}
```

Thanks for your Attention

which is not all we need but is a rare, hence precious thing nowadays…

https://www.linkedin.com/in/cedricberg/