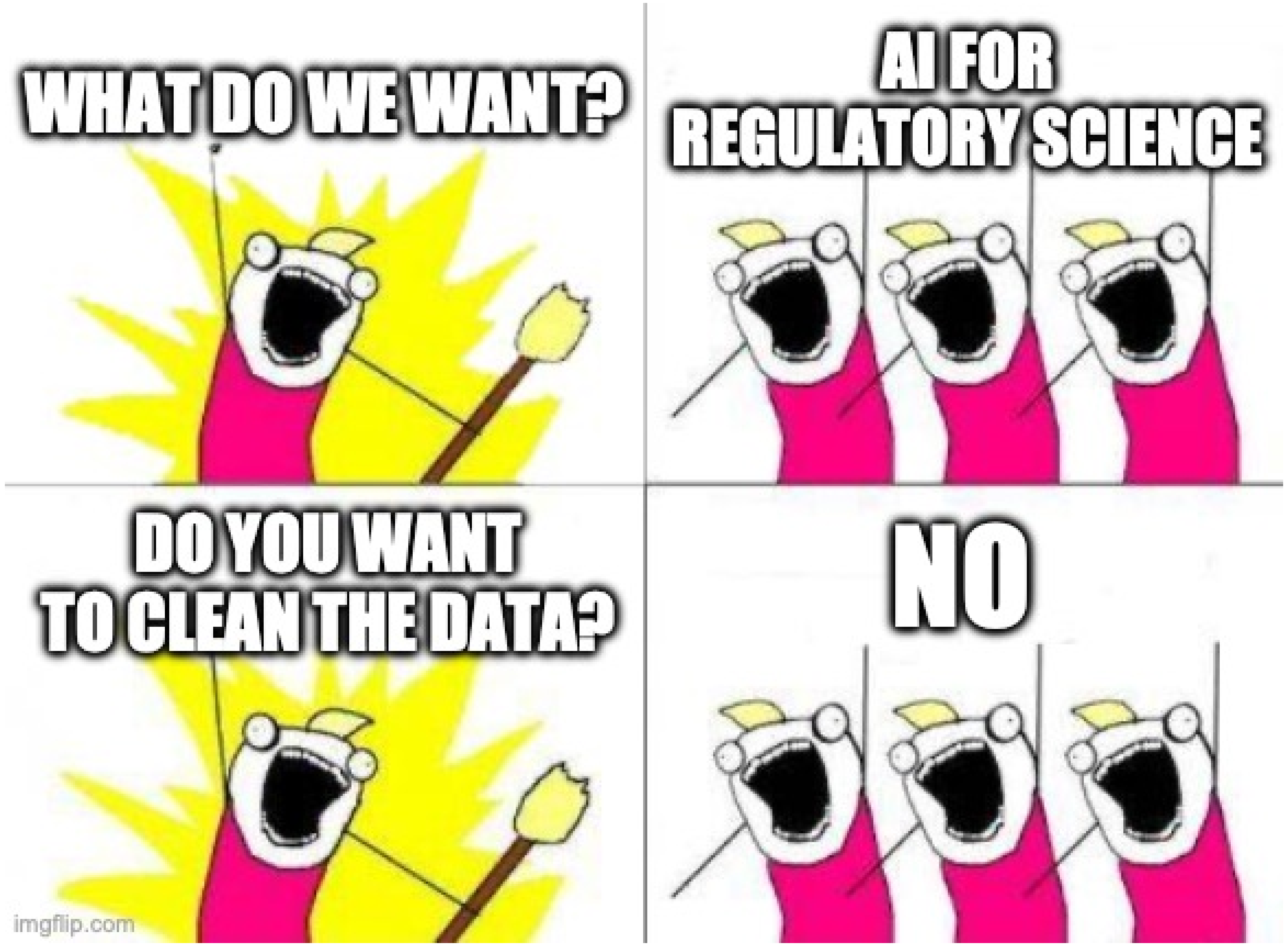
PREREQUISITES FOR DATA ANALYSIS AND AI

ÁKOS BERNARD JÓZWIAK
Digital Food Institute
University of Veterinary Medicine Budapest

INTERNATIONAL SCIENTIFIC CONFERENCE ON "GLOBAL COMMODITY CHAINS FROM A RISK ASSESSMENT PERSPECTIVE"
BERLIN, 27-29 MAY 2024

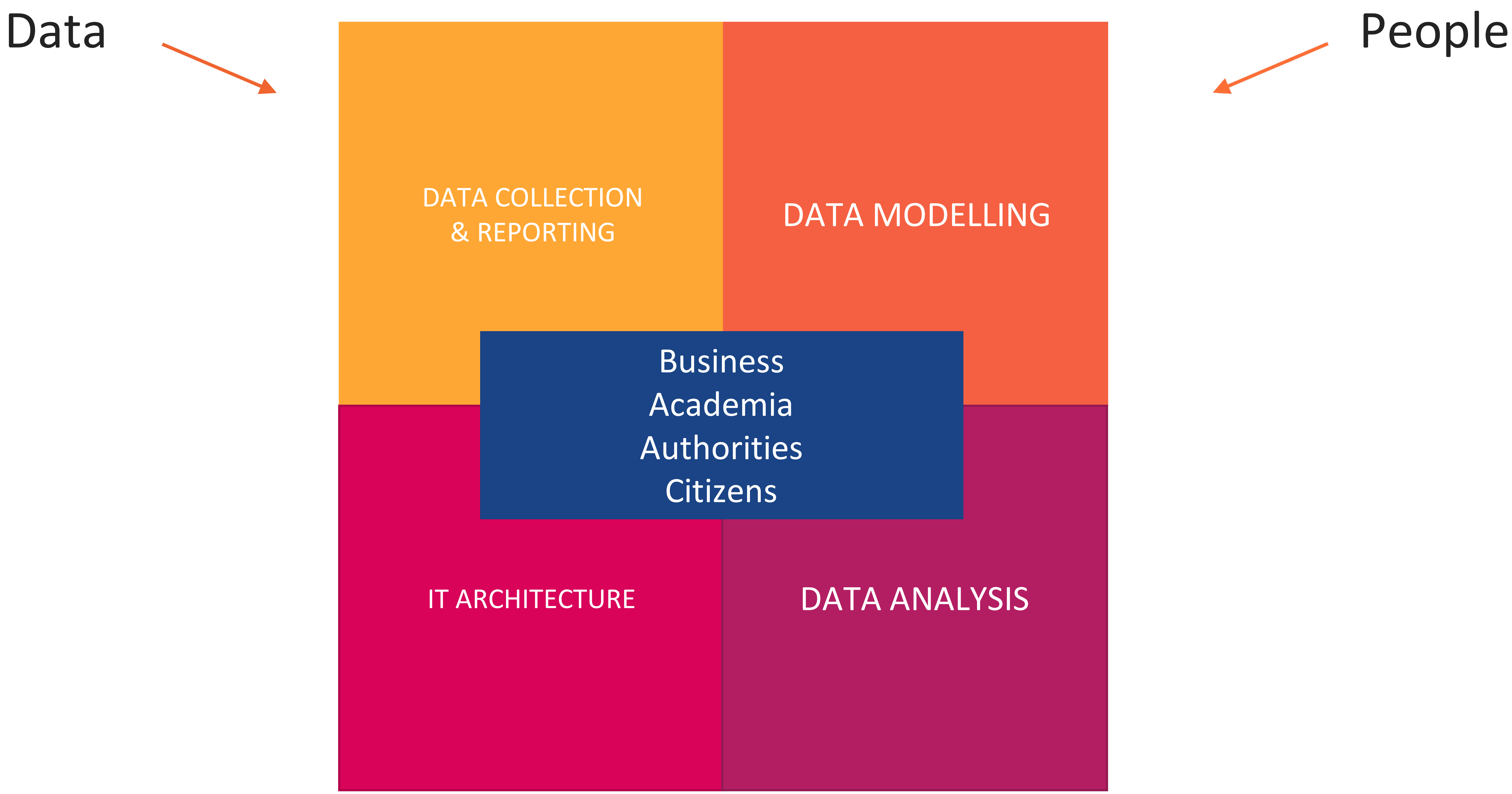
OUTLINE

OUTLINE



PREREQUISITES

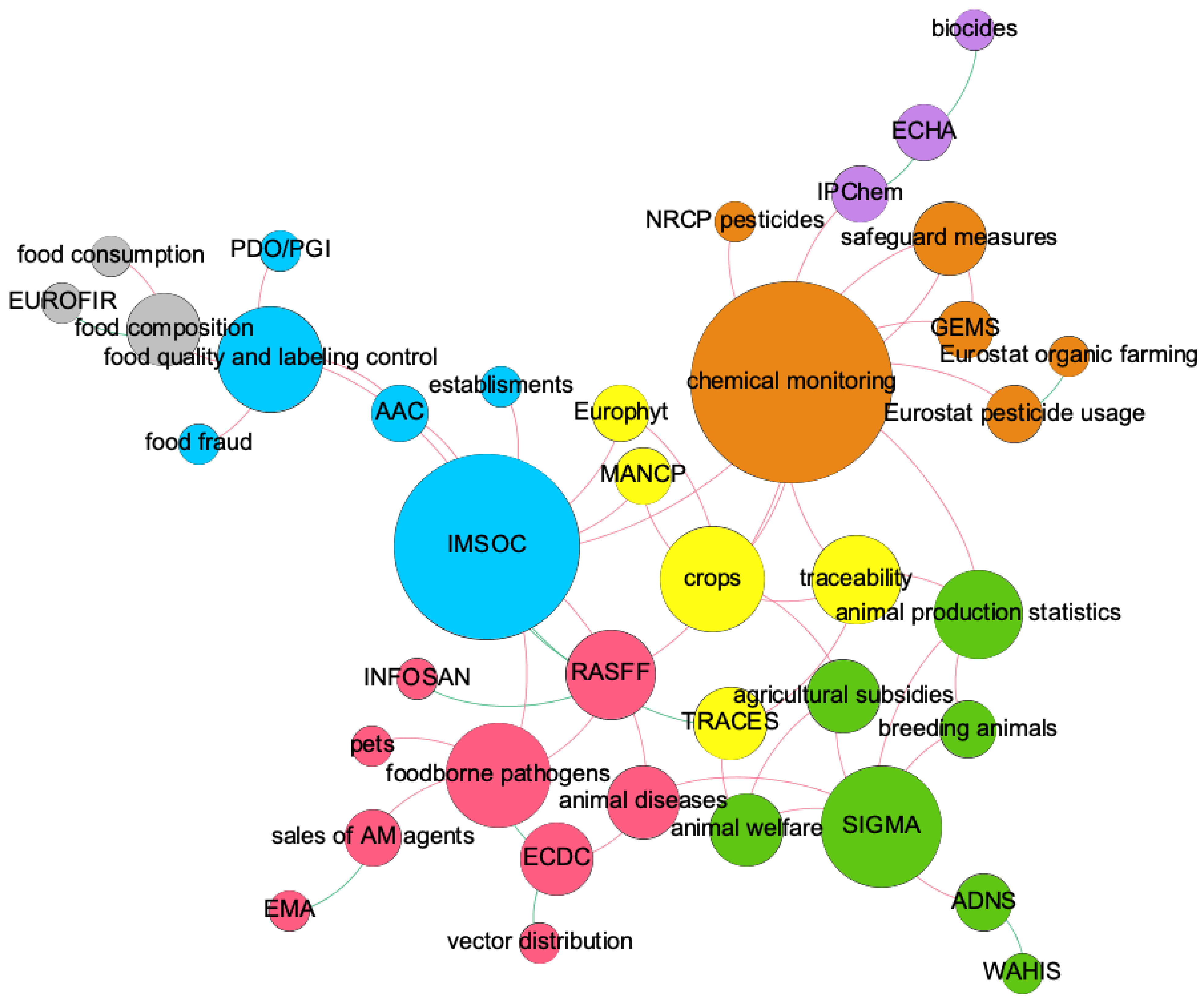
WHAT DO WE NEED?



EFSA ADVISORY GROUP ON DATA

EFSA ADVISORY GROUP ON DATA

- Act as a governance body providing recommendations
- Act as a Think Tank providing input on project ideas
- Act as a channel providing access to knowledge, expertise, competencies and staff in Member States
- Provide strategic input on and oversight of alignment of EFSA’s data roadmap

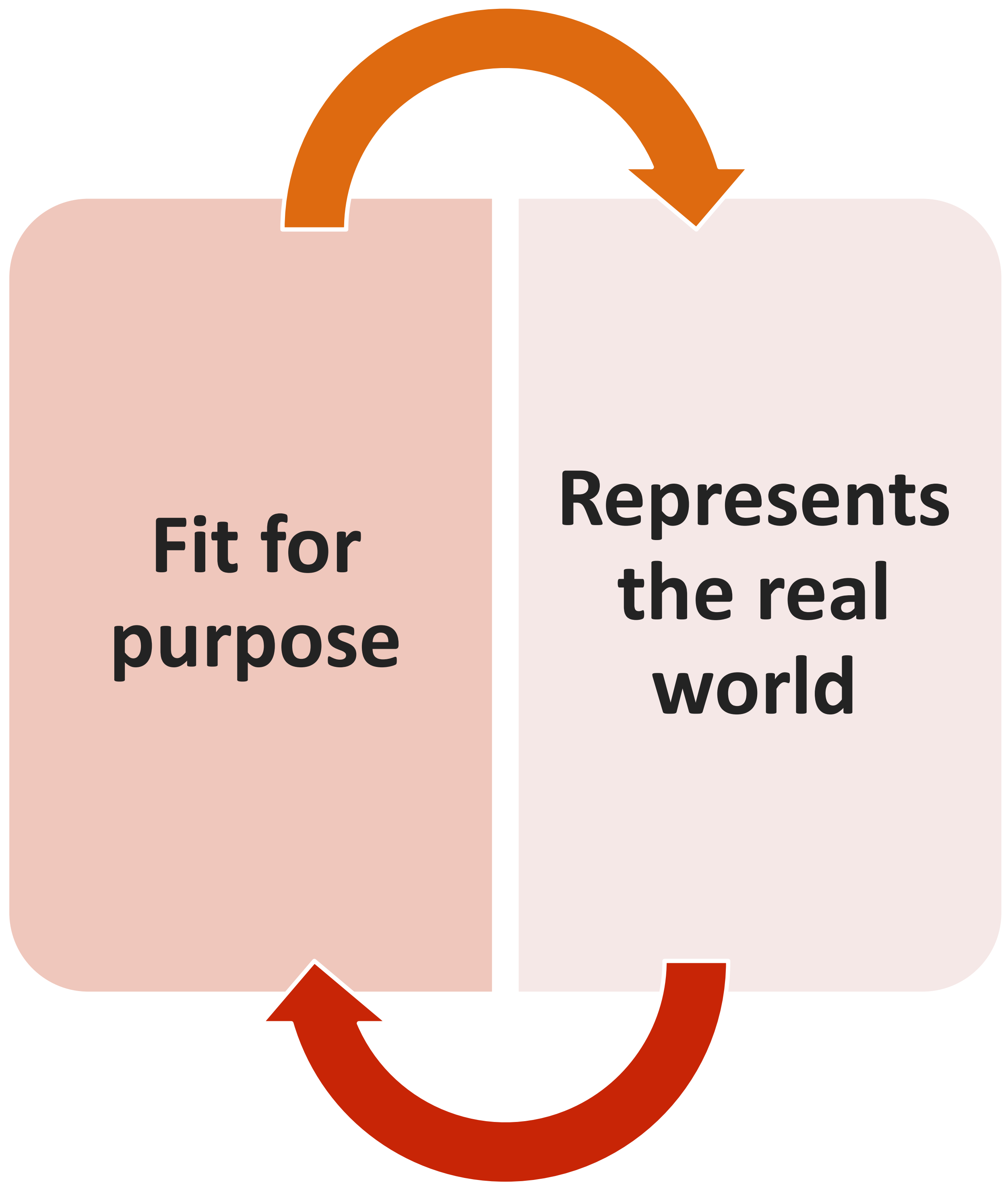


https://doi.org/10.2903/sp.efsa.2020.EN-1901

PREREQUISITES

DATA

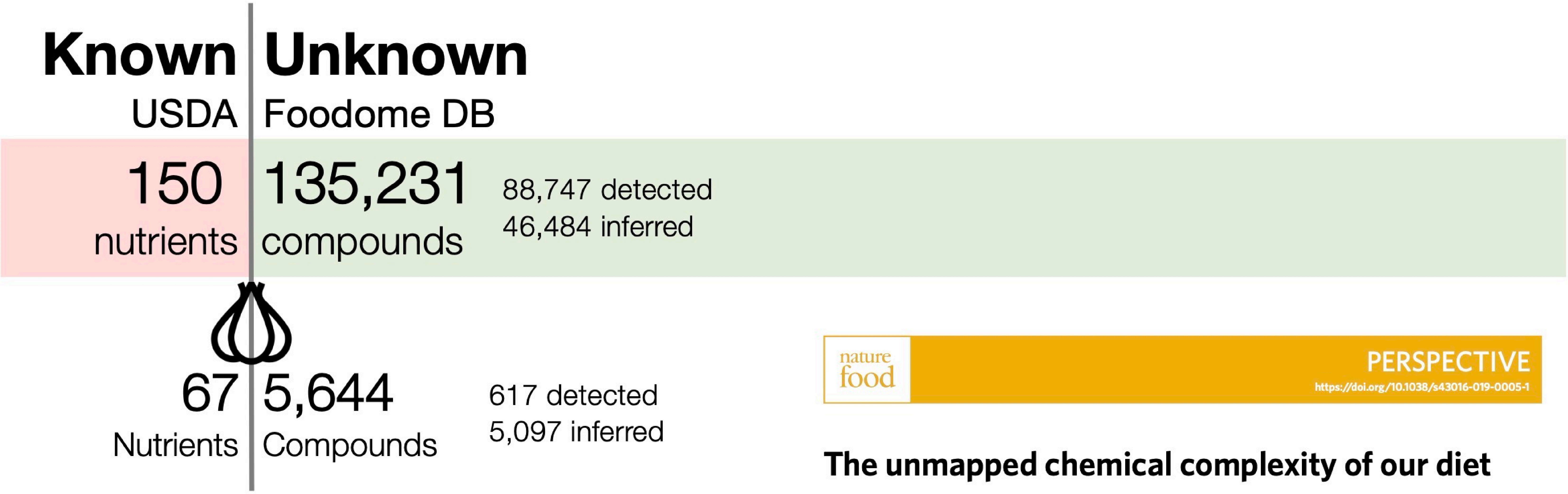
- Quality
  (Completeness, Validity, Uniqueness, Timeliness, Consistency, Accuracy)
- Quantity?
- Granularity?
- Representativity?
- Interoperability?
- ...

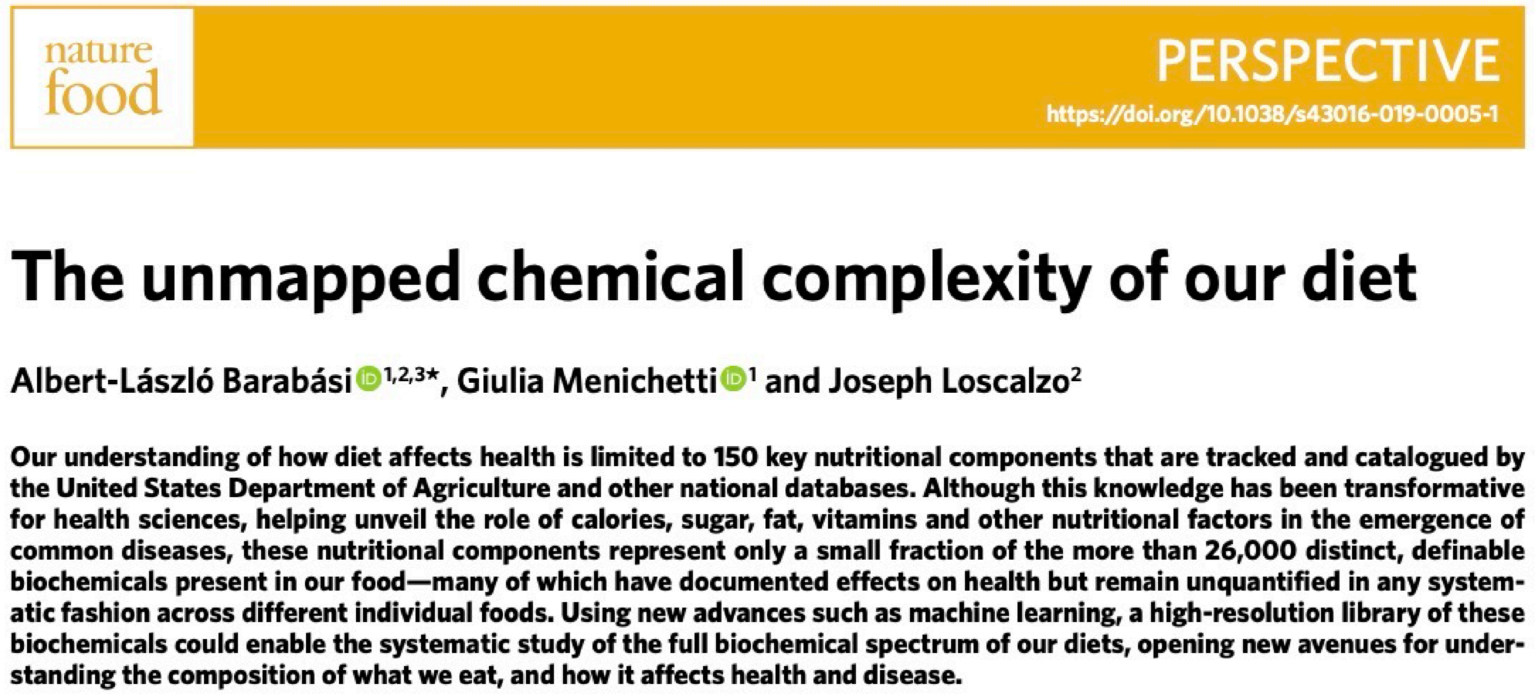


EXAMPLES

MISSING DATA

- Molecular level food composition data
- Project Foodome





nature food
PERSPECTIVE
https://doi.org/10.1038/s43016-019-0005-1

The unmapped chemical complexity of our diet

Albert-László Barabási[1,2,3]*, Giulia Menichetti[1] and Joseph Loscalzo[2]

Our understanding of how diet affects health is limited to 150 key nutritional components that are tracked and catalogued by the United States Department of Agriculture and other national databases. Although this knowledge has been transformative for health sciences, helping unveil the role of calories, sugar, fat, vitamins and other nutritional factors in the emergence of common diseases, these nutritional components represent only a small fraction of the more than 26,000 distinct, definable biochemicals present in our food—many of which have documented effects on health but remain unquantified in any systematic fashion across different individual foods. Using new advances such as machine learning, a high-resolution library of these biochemicals could enable the systematic study of the full biochemical spectrum of our diets, opening new avenues for understanding the composition of what we eat, and how it affects health and disease.

EXAMPLES

REPRESENTATIVITY PROBLEMS

- Sampling strategies:
  - objective (i.e., random)
  - selective (i.e., risk-based)
  - suspect
  - (convenient)

  } *statistically limited interpretability / biased results*
- Challenges:
  - Central level random sampling plan, executed on a risk basis locally: what strategy is reported then?
  - Many questionable reporting practices, inconsistencies
  - E.g.: Veterinary drug residues sampling programmes

EXAMPLES

MISSING / MISALIGNED STANDARDS

- Can we automatically link RASFF data with EFSA contaminants and consumption data?
- Not yet (although EFSA and COM are working on it)

RASFF (own) catalogues on food categories and hazards

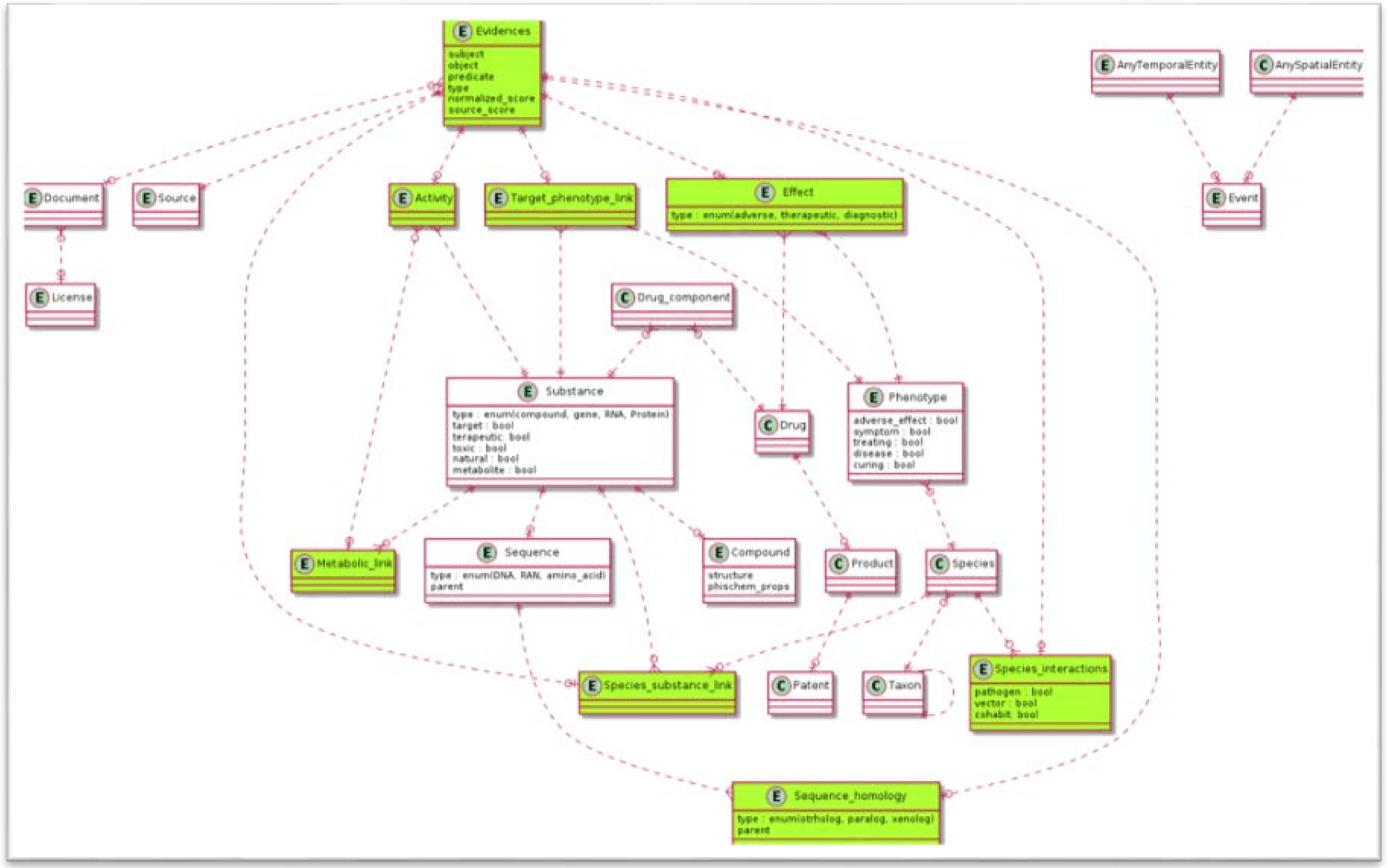
≠

EFSA catalogues on food categories (FoodEx2) and hazards (PARAM)

EXAMPLES

CHALLENGES IN CONNECTING DIFFERENT DATA SOURCES

- Can we connect different data sources easily (automatically)?
- No
- There are massive opportunities in using agricultural, customs, trade, business, meteorological, user-generated, … data



PREREQUISITES

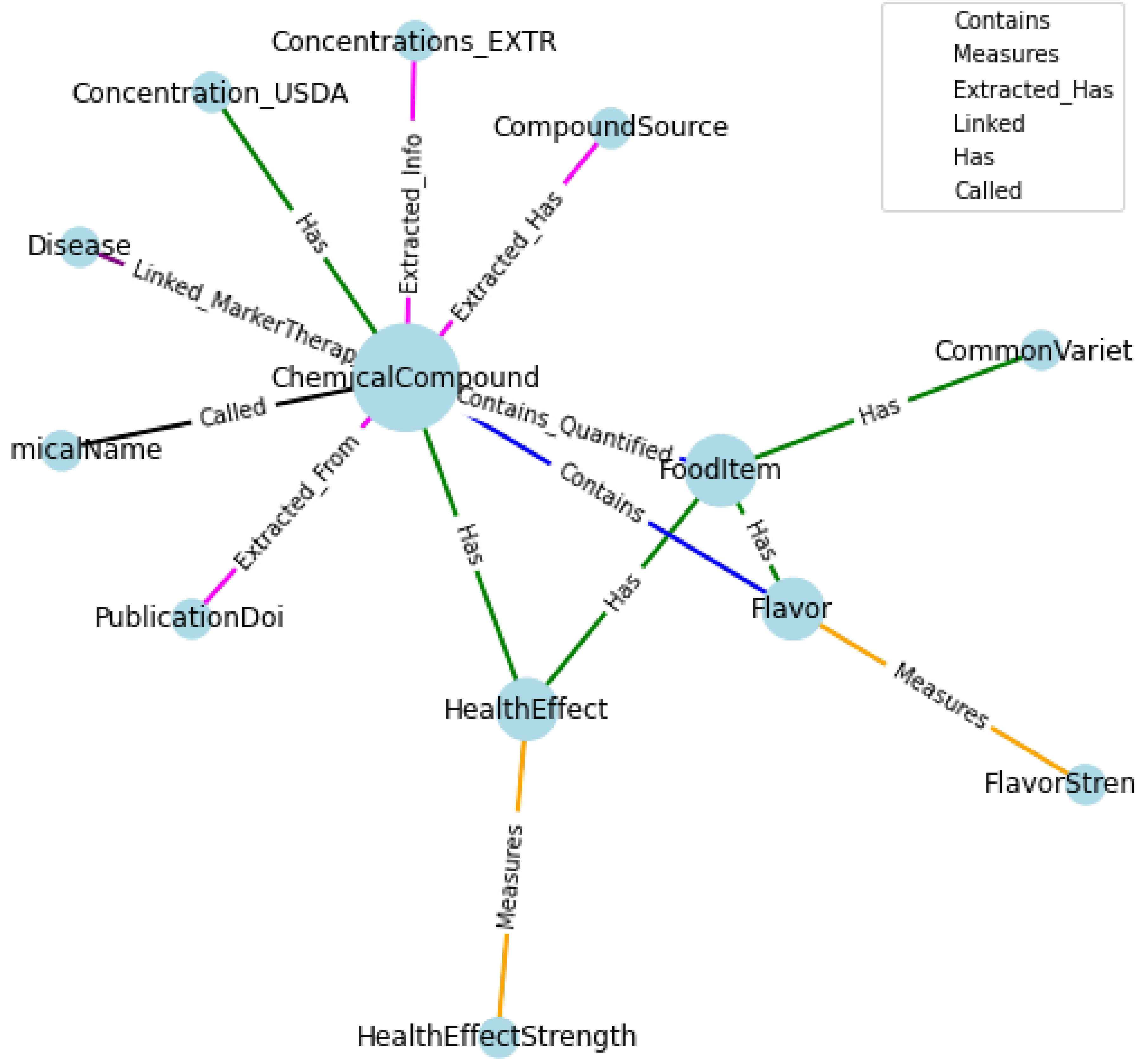
CAN'T WE JUST USE MACHINES FOR THAT?

- Yes, but we need few things:
  - Technological Feasibility and Data Availability: abundant, high-quality data
  - Economic Viability: (market) need with a high return on investment
  - Ethical and Social Considerations: decision accountability (clear legal, ethical rules)
  - Technical Complexity and Safety Concerns: advanced and reliable AI systems
  - Cultural and Social Acceptance: demystifying AI
  - Research and Development Focus: research need and funding

EXAMPLES

MACHINE READABLE DATA?

- Building knowledge graphs for research and/or control purposes
  - Need for interoperable, connected ontologies
  - Easy to access data (Repositories, direct database access, API, …)
  - FAIR (Findable, Accessible, Interoperable, Reusable)
- Do we have that?



EXAMPLES

IMPORTANCE OF ONTOLOGIES

- Ontology: a generalized, semantic data model
- Research projects aiming for utilising data for better food systems safety: connecting various (open source) data with the help on ontologies and common identifiers
- Need for standardised, interoperable ontologies
  - Food classification: FoodON is the one used by the research community, not FoodEx2.
    Is it fine for EFSA, COM, MS authorities?
  - Inter-agency exchange of chemical contaminants data: which ontology to choose?
  - No common international ontology of animal diseases
  - …

PREREQUISITES

PEOPLE

- Creation and development of (big) databases is not only an IT problem
- The ability to analyse and evaluate ***input data*** and ***results***: high-level knowledge of food systems science is needed enabling interpretation and validation
- Data literacy
  - Basic statistics is in the food safety risk assessment curricula
  - But data science is not (or very rare)
  - Future (or current) risk assessors need data generation, retrieval, manipulation and analysis knowledge

PREREQUISITES

CAN'T WE JUST USE MACHINES FOR THAT?

- Yes, but we need few things:
  - Technological Feasibility and Data Availability: abundant, high-quality data
  - Economic Viability: (market) need with a high return on investment
  - Ethical and Social Considerations: decision accountability (clear legal, ethical rules)
  - Technical Complexity and Safety Concerns: advanced and reliable AI systems
  - Cultural and Social Acceptance: demystifying AI
  - Research and Development Focus: research need and funding

OUTLOOK

WHAT CAN WE DO?

- Invest in data generation
- Build ontologies
- Share tools, standards, data, models
- Use open data standards
- Educate
- Manage changes
- Explore ‘lighthouse’ ideas/projects for AI
- Build networks and partnerships

EFSA ADVISORY GROUP ON DATA

THANK YOU FOR YOUR ATTENTION

CONTACT

Ákos Józwiak
Research Director I Digital Food Institute, University of Veterinary Medicine Budapest
Head of Food and Nutrition Science I Syreon Research Institute
jozwiak.akos@univet.hu
LinkedIn: akosbernardjozwiak