Brunel University London

AAAI 2016 Fall Symposium
November 18

Reproducible Science

Dr Larisa Soldatova et al

The AI grand challenge of accelerating science

The AI grand sub-challenge: reproducible science

- The reproducibility crisis: probably the most important problem the scientific community is facing.
- If remains unresolved, the credibility of science could be irrevocably damaged.
- “More that 70% of researchers have tried and failed to reproduce another scientist’s experiments, and more than half have failed to reproduce their own experiments” (1).
- Macleod et al state that in US “85% of research investment is wasted” (2).

1. Nature Editorial (2016) Reality check on reproducibility. *Nature* 533: 437-438.
2. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, *et al.* (2014) Biomedical research: increasing value, reducing waste. *Lancet* 383: 101–104.

Reasons for the non-reproducibility

- the complexity of scientific methods,
- poor experimental design,
- the non-availability of raw data, code, etc.
- **the use of natural language** i.e. English

“The argument is simply that by the word 'experiment' we refer to a situation where we can tell others what we have done and what we have learned and that, therefore, the account of the experimental arrangement and the results of the observations must be expressed in unambiguous language …” (2)

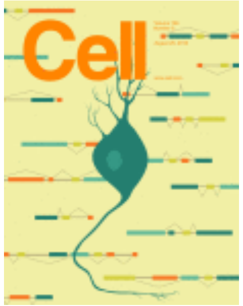
Bohr N in Albert Einstein Philosopher Scientist ed. P.A. Schilpp, 1949.

Formal knowledge representation

- ~500 Ontologies in BioPortal, ~40 MI (Minimum Information for…), other standards, knowledge bases, …
- The main focus is on declarative knowledge – not enough for the reproducibility!

“Recent calls to … the research community, and funding agencies to improve rigor and reproducibility in science clearly point to the need to take a new approach to communicating not just **the ‘what’ but the ‘how’ of science”**

Leading Edge Editorial (2016) A STAR Is Born. *Cell* 166, 1059-1060.



Formal representation of procedural knowledge

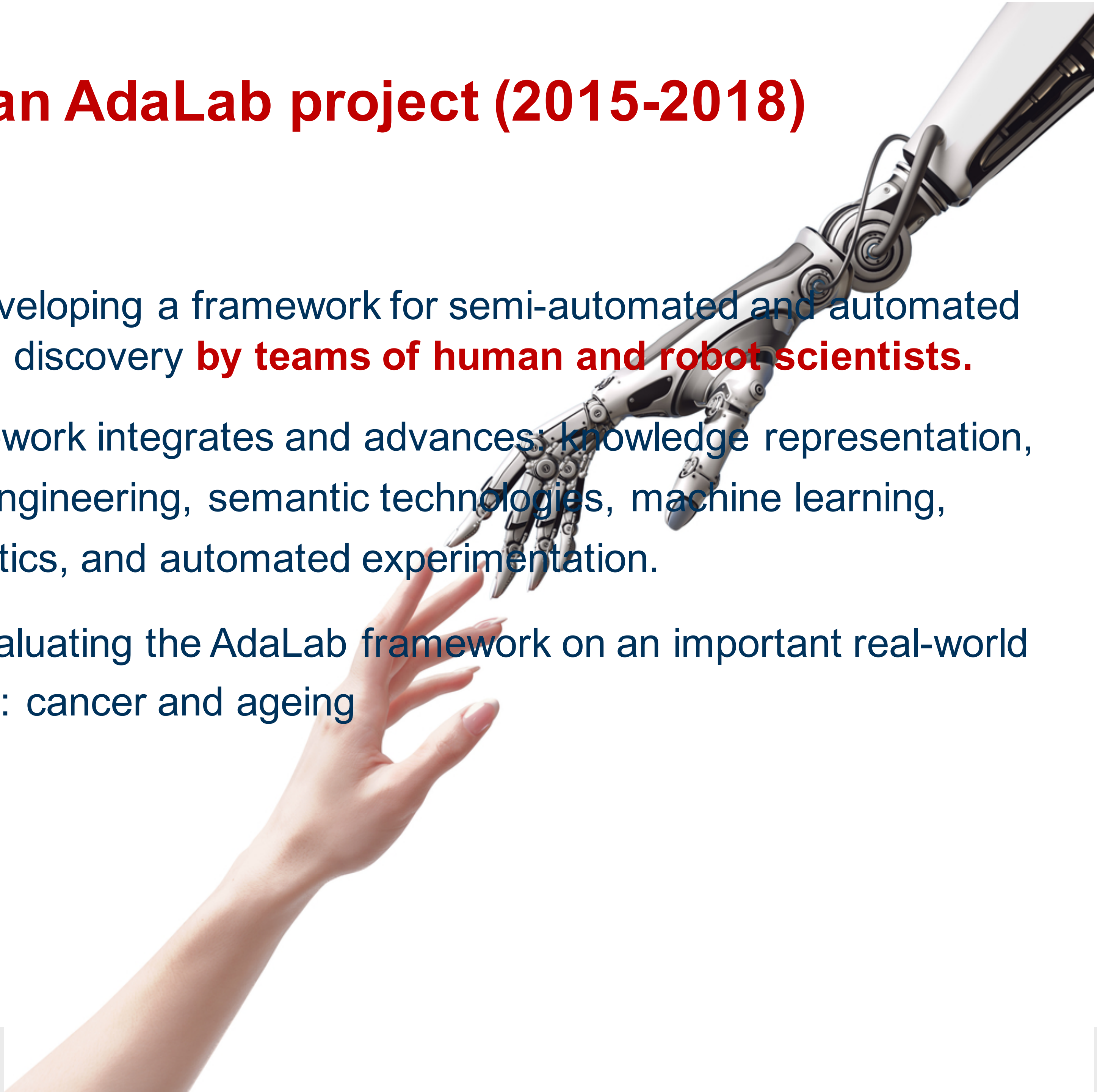
- OBI (the Ontology for Biomedical Investigations): e.g. *OBI: data transformation, OBI: injection*
- EFO (Experimental Factors Ontology): a systematic description of experimental variables for capturing experimental designs
- SMART (SeMAntic RepresenTation for Experimental Protocols): provenance, objectives,…
- EXACT (Experimental ACTions) ontology: definitions of typical actions and their properties

Not Enough!

Example: a Robot Scientist EVE

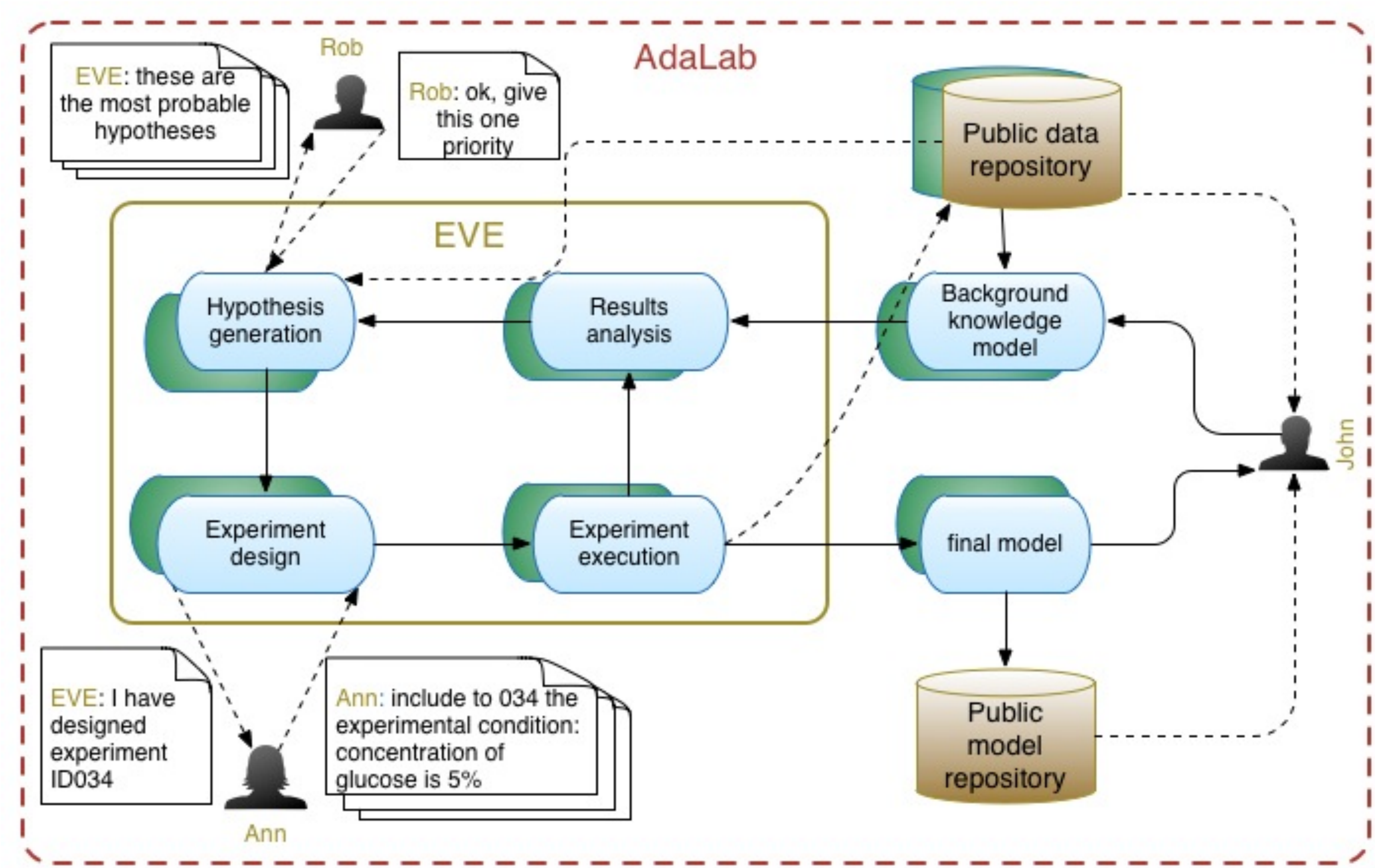
- Fully autonomous robotic system for drug (lead) discovery
- All aspects of scientific studies are formally recorded
- There are dedicated ontologies for Eve: equipment ontology, Eve (typical experiments), HELO (hypotheses), EXACT (protocols), UNO (uncertainties)
- Eve moved from Aberystwyth to Manchester
- Eve could not reproduce previous drug screening experiments
- The reason: a mode of shaking – it took **two months** to find out
- The level of granularity of the representation is
  - **OR too low (for equipment)**
  - **OR too high (for humans)**

European AdaLab project (2015-2018)

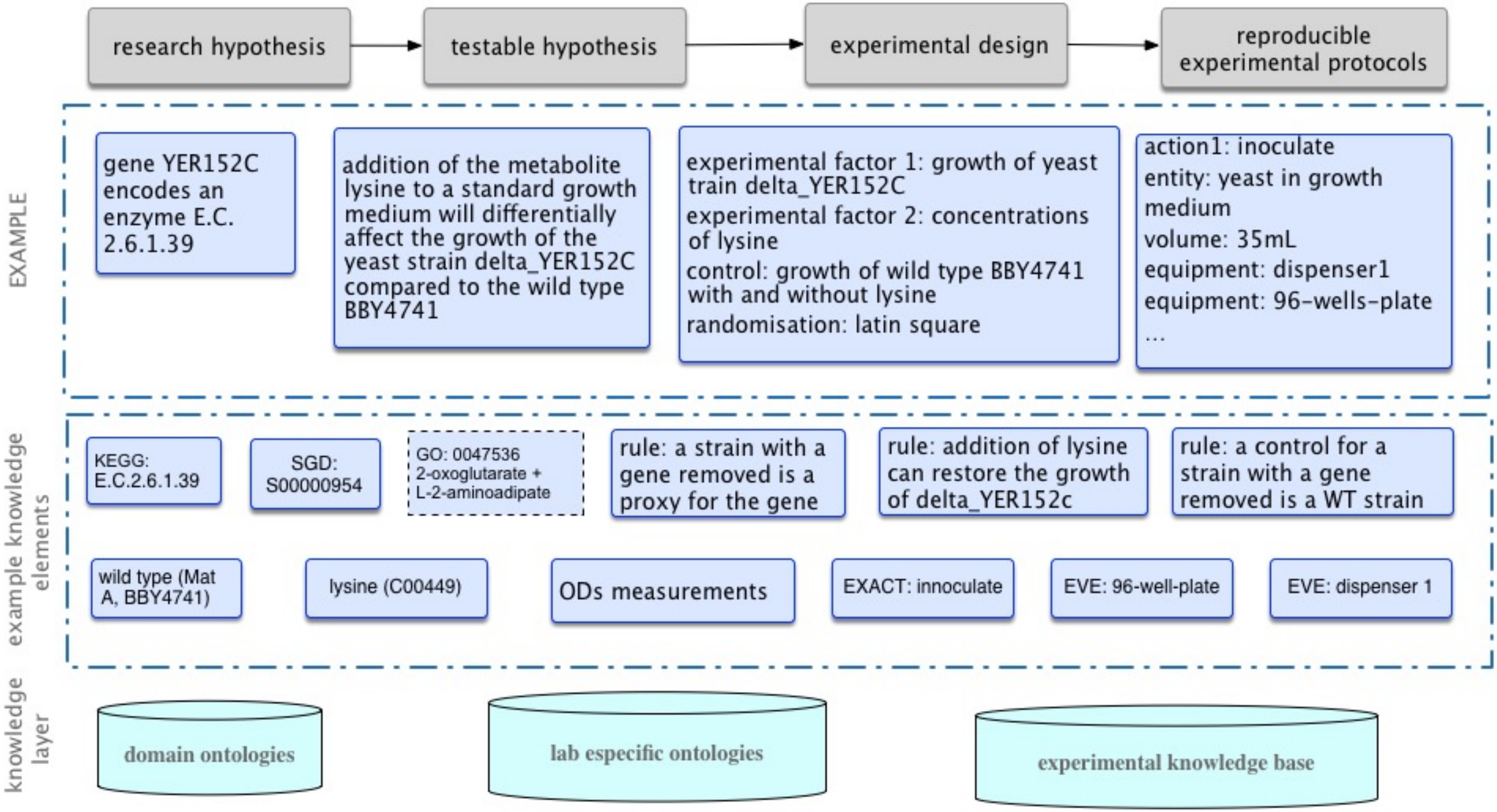


- We are developing a framework for semi-automated and automated knowledge discovery **by teams of human and robot scientists.**
- This framework integrates and advances: knowledge representation, ontology engineering, semantic technologies, machine learning, bioinformatics, and automated experimentation.
- We are evaluating the AdaLab framework on an important real-world application: cancer and ageing

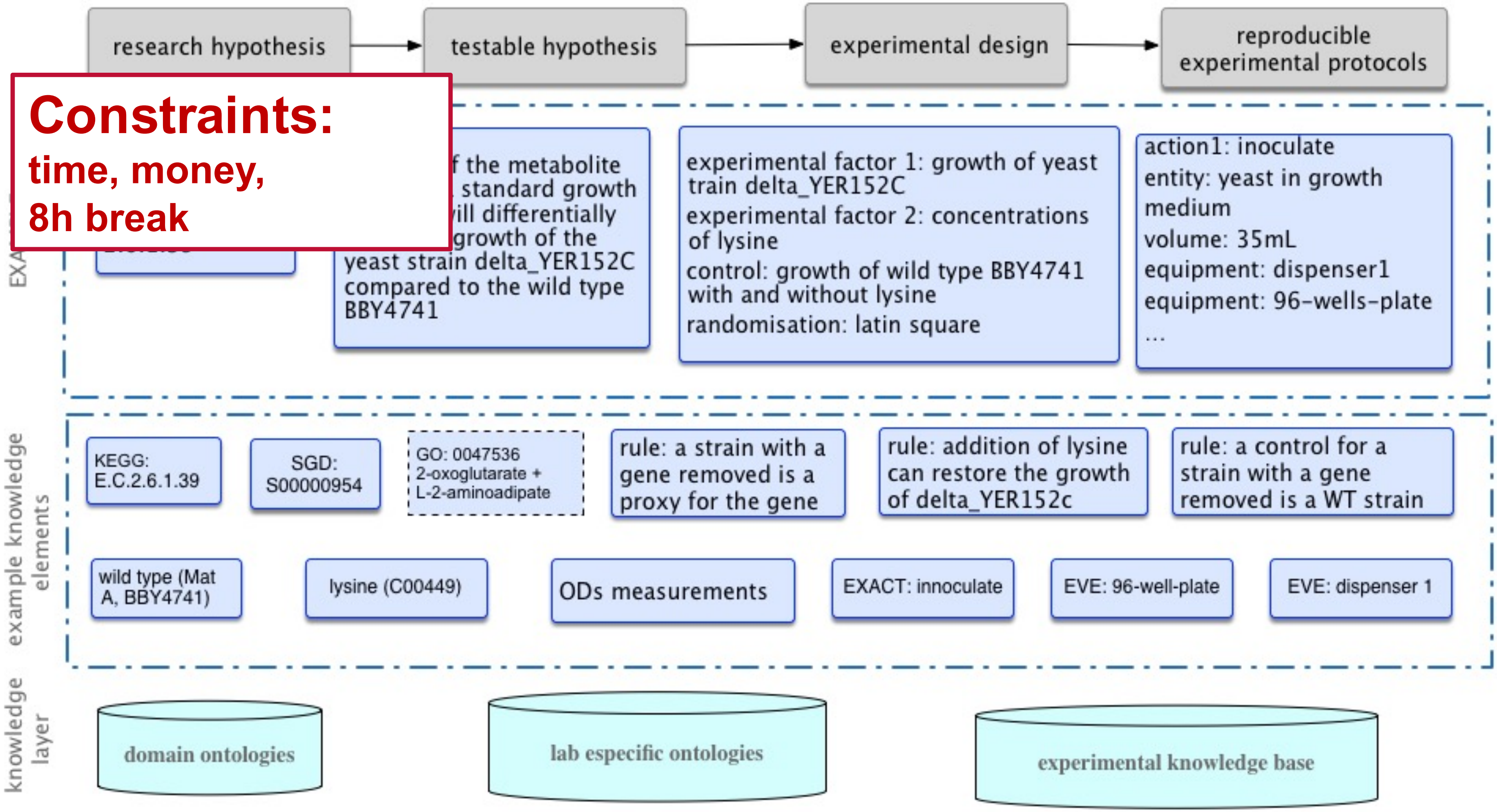
Overview of AdaLab



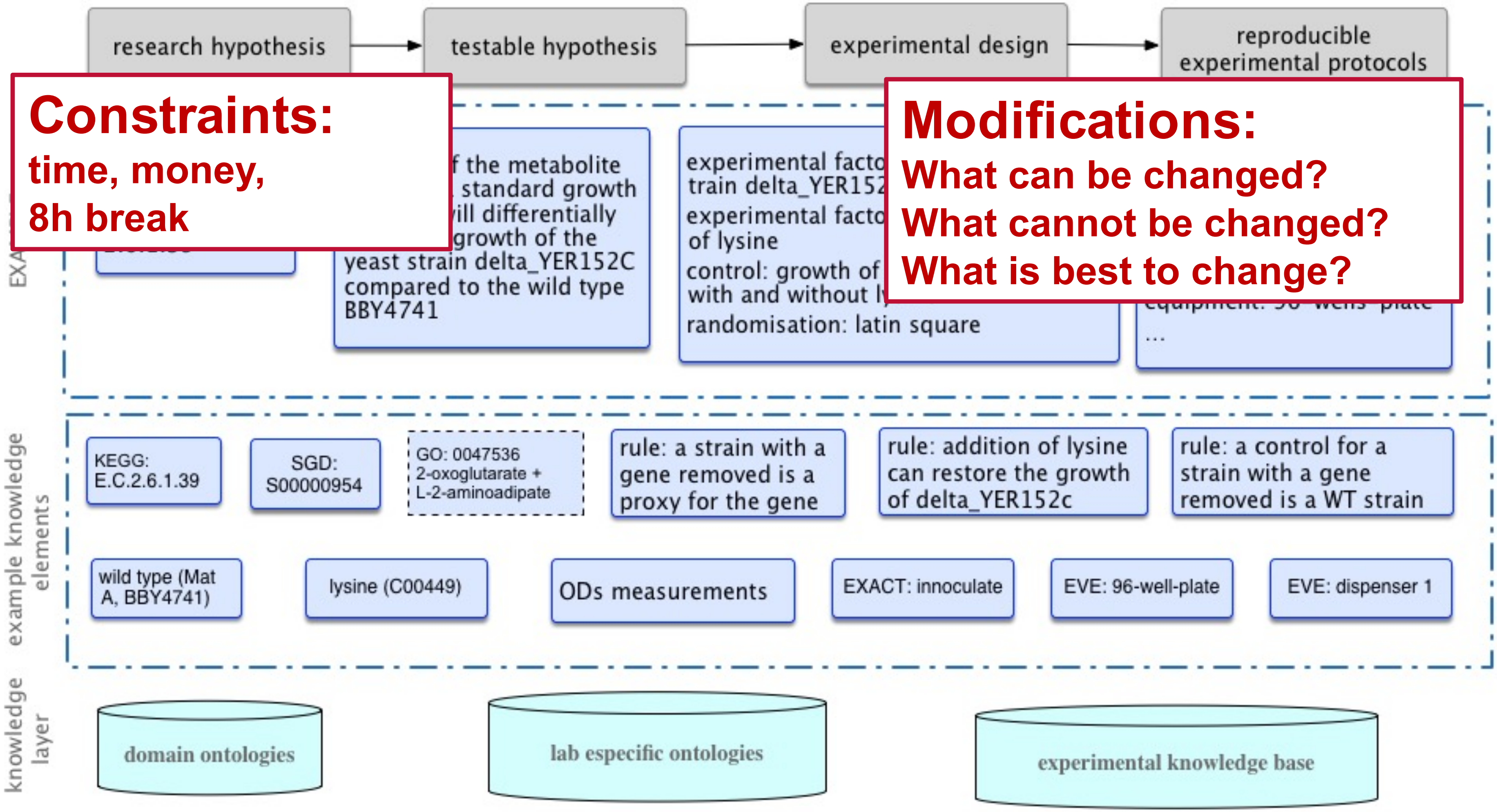
Generation of reproducible experimental protocols



Generation of reproducible experimental protocols



Generation of reproducible experimental protocols



Constraints:
time, money,
8h break

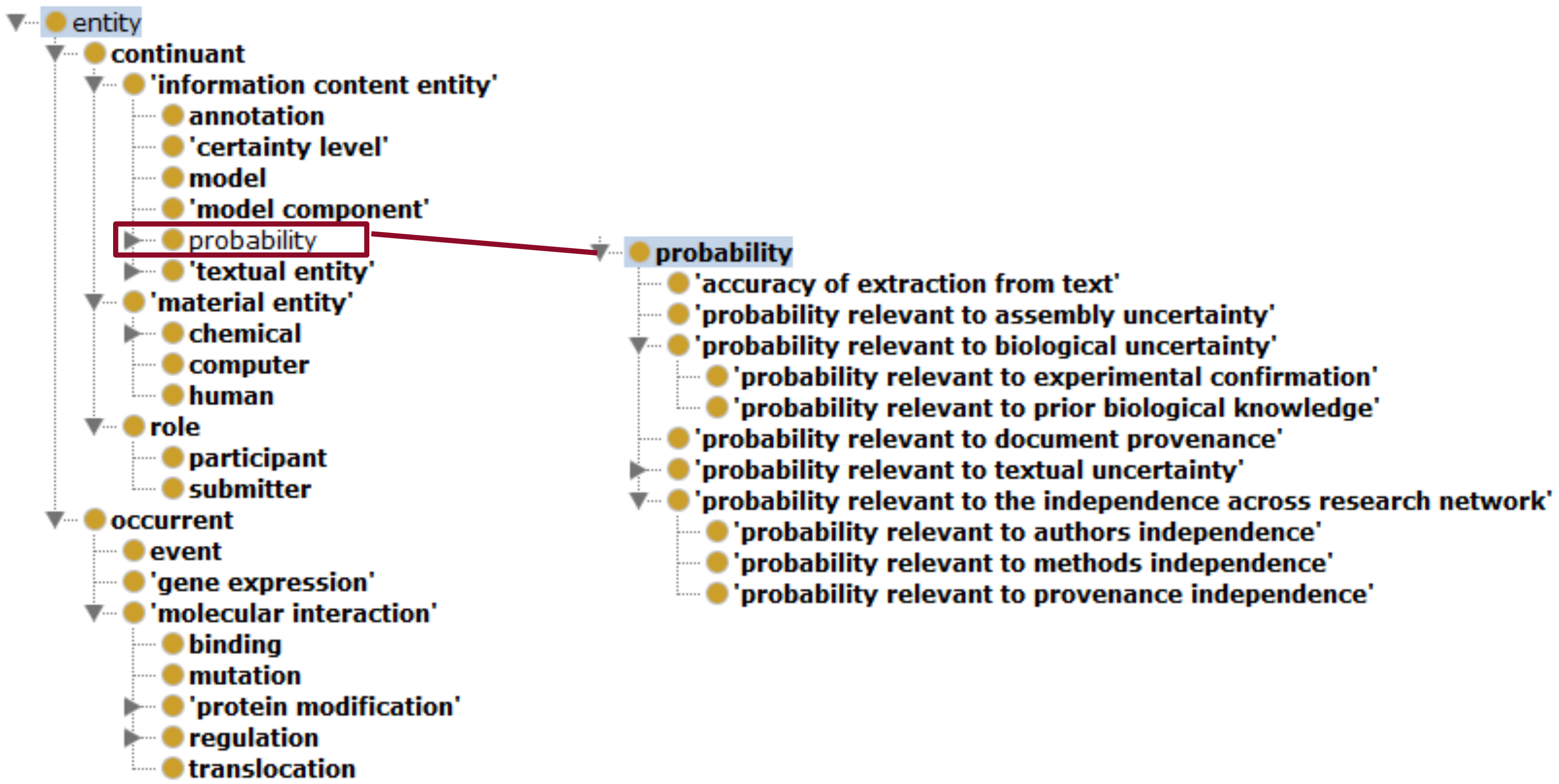
Modifications:
What can be changed?
What cannot be changed?
What is best to change?

From: ACCELERATING SCIENCE: THE VALUE PROPOSITION

- Construct a computational model, e.g., a network of genes that orchestrate a specific biological process of interest, that make experimentally testable predictions.
- Design and prioritize, orchestrate, and execute experiments.
- The task of designing an optimal experiment that provides the most valuable information at the lowest cost to help answer a chosen scientific question requires a careful exploration of the space of possible experiments, their relative cost, risk, and feasibility, in the context of all that is known.

UNO: uncertainties ontology

OWL ontology covering: event-related concepts, metadata concepts and probability types.



PANDA: probabilistic knowledge assembly framework

```
event(ev).
0.7::supported(X) :-      event(X),
                          statement(Y),
                          represents(Y, X),
                          hasTruthValue(Y, true),
                          combinedProb(Y).

combinedProb(Y) :-        extractionProb(Y),
                          provenanceProb(Y).
```

support for an event *= disjoint sum of (combined) probabilities of different supporting statements, with each statement weighted by 0.7*

combined probability = product of all probability scores

```
statement(s1).
represents(s1, ev).
hasTruthValue(s1, true).
0.8::extractionProb(s1).
0.7::provenanceProb(s1).
```

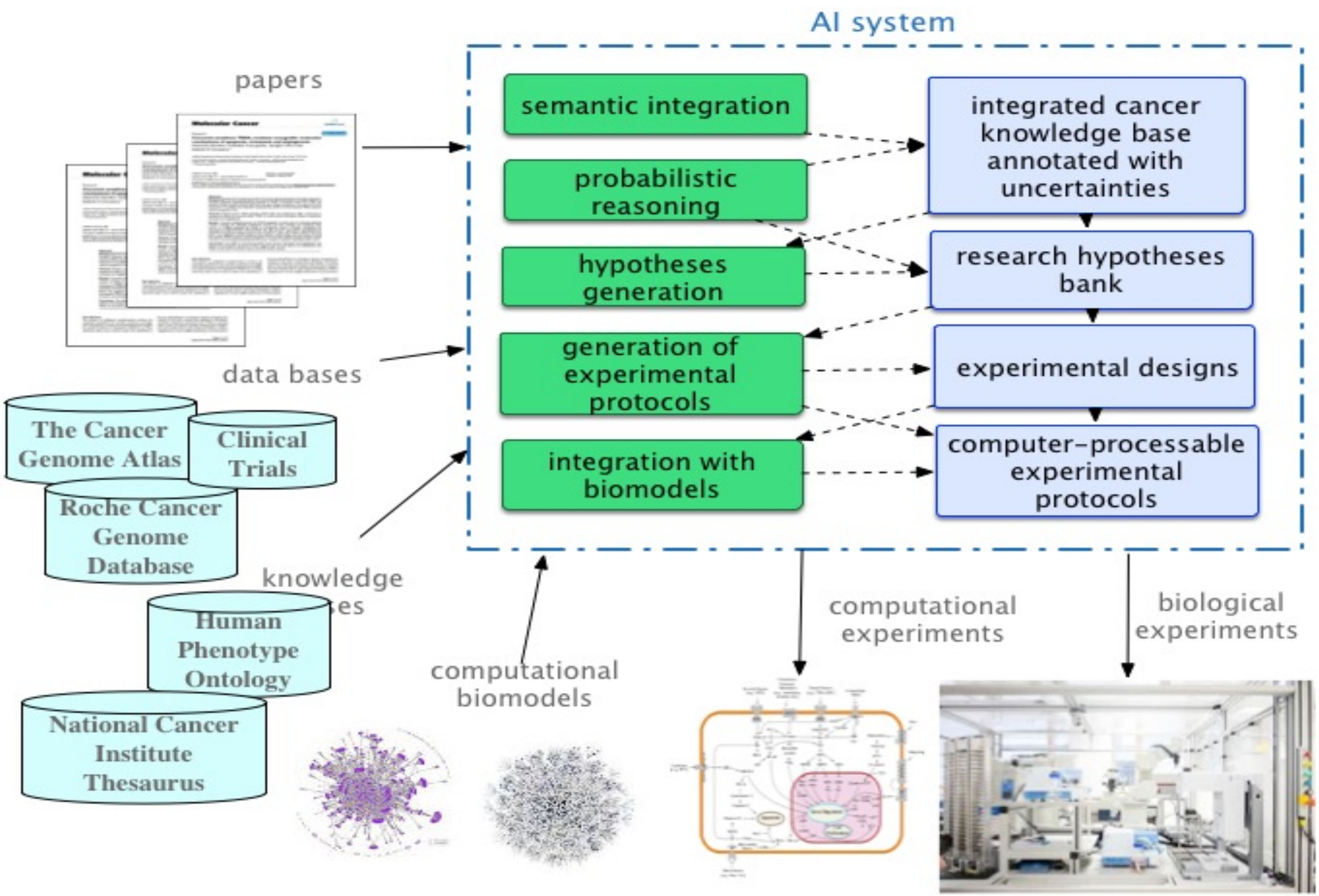
```
statement(s2).
represents(s2, ev).
hasTruthValue(s2, true).
0.7::extractionProb(s2).
0.6:: provenance Prob(s2).
```

Note the increase in the probability of corroborated(ev) *on adding the second supporting statement*

supported(ev) → 0.392

supported(ev) → 0.570752

Our Vision



AI system
papers
semantic integration
probabilistic reasoning
hypotheses generation
generation of experimental protocols
integration with biomodels
integrated cancer knowledge base annotated with uncertainties
research hypotheses bank
experimental designs
computer-processable experimental protocols
data bases
The Cancer Genome Atlas
Clinical Trials
Roche Cancer Genome Database
Human Phenotype Ontology
National Cancer Institute Thesaurus
knowledge
computational biomodels
computational experiments
biological experiments

Acknowledgements

AdaLab:

The University of Manchester, UK

University Paris-Nord, France

University of Évry-Val-d'Essonne , France

Katholieke Universiteit Leuven, Belgium

Big Mechanism:

The University of Chicago, IL

ISI (Information Sciences Institute), CA

Microsoft

The University of Manchester, UK

Previously: the University of Cambridge, Aberystwyth University, Cardiff University,…

Thank you

Questions?